Norges Handelshøyskole
Bergen, Høst 2021

Datadrevet innovasjon

Er dataintensiv problemløsing en oppgave for økonomen eller teknologen?

Erlend Reksten Amland og Torbjørn Røys

Veileder: Eirik Sjåholm Knudsen

Masterutredning i Økonomi og administrasjon

Hovedprofiler: New Business Development & Business Analytics

NORGES HANDELSHØYSKOLE

Forord

Masterutredningen utgjør det avsluttende arbeidet på studiet Økonomi og administrasjon ved Norges Handelshøyskole (NHH). Avhandlingen er skrevet innenfor fordypningsprofilene New Business Delvelopment & Business Analytics i perioden juli - desember 2021 med støtte fra forskningssenteret Digital Innovation for Growth (DIG) ved NHH.

Arbeidet med oppgaven har vært interessant og intensivt, og vi håper at oppgaven kan gi innsikt og interesse for videre forskning på datadrevet innovasjon.

Takk til familie, venner, kjente og ukjente som har bidratt med å besvare caseoppgaven i masterutredningen, og en ekstra takk til DIG ved NHH som har bidratt med støtte til gjennomføring av caseoppgaven.

Videre rettes en stor takk til Knowit som har stilt verdifull kompetanse, diskusjonspartnere og tid til disposisjon.

Avslutningsvis rettes en stor takk til veileder Eirik Sjåholm Knudsen, som har veiledet gjennom regelmessige møter og lengre sesjoner med god input. Takk for veiledning langt utover det vi kunne forvente.

Norges Handelshøyskole

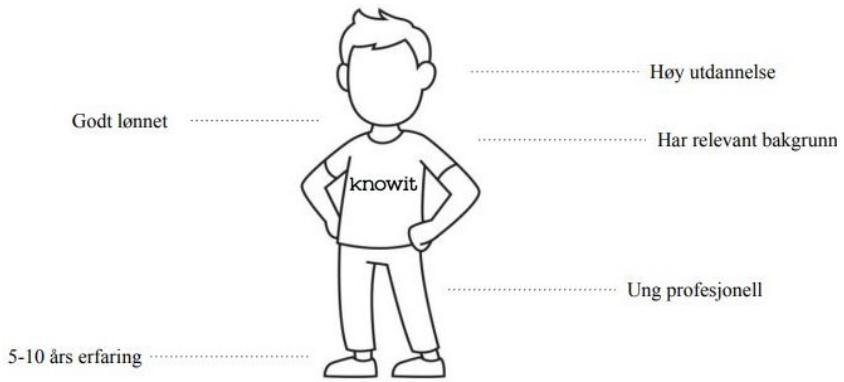
Bergen, desember 2021

Erlend Reksten Amland

Torbjørn Røys

Sammendrag

I samarbeid med DIG-senteret ved NHH og Knowit har vi i denne oppgaven forsket på hvordan humankapital påvirker evnen til datadrevet innovasjon. Et datadrevet forsikringscase ble benyttet til å generere kreative ideer blant 92 respondenter. Ideene ble deretter vurdert av et ekspertpanel fra Knowit langs de fire dimensjonene; kreativitet, gjennomførbarhet, kommersialisering og datainntensivitet. Ved hjelp av ekspertenes vurderinger og demografiske variabler har vi utarbeidet og kjørt ulike statistiske modeller for å belyse hvilke trekk og egenskaper som påvirker evnen til å arbeide med dataintensive problemstillinger. Mer konkret har vi sett på hvordan respondenter med *teknologibakgrunn*, *forretningsbakgrunn*, *innovasjonsbakgrunn* og de *uten relevant bakgrunn* scorer langs de fire dimensjonene. Resultatene indikerer at det ikke handler om én utdanningsbakgrunn eller kompetanseprofil som egner seg for datadrevet innovasjon. Utredningens funn trekker i retning av at svaret på den overordnede problemstillingen er langt mer sammensatt, og det er derfor ikke mulig å besvare den med én stillingstittel eller profesjon. Enkelte signifikante funn, sammen med mønstre og tendenser i datasettet gjør at vi likevel kan tegne et bilde av en profil som har gode forutsetninger for å lykkes med innovasjon og bruk av data. Personen under er likevel et eksempel på en profil som innehar mange av de viktige egenskapene og trekkene som bidrar til å lykkes med datadrevet innovasjon. Denne profilen har høyere utdannelse, på minimum bachelornivå og har kanskje erfaring med innovasjonsmetodikk og innovasjonsprosesser. Sannsynligvis er personen utdannet innen kommersielle, innovative eller teknologiske fag, og omtales gjerne som en ung profesjonell i midten av 30-årene. Det er positivt om vedkommende har mellom 5-10 års arbeidserfaring og har hatt mulighet til å oppnå en god årsinntekt over 600.000 kroner.



Nøkkelord – Datadrevet Innovasjon, Humankapital, Knowit,

Innhold

1 Innledning **1**
1.1 Introduksjon . . . 1
1.2 Problemsstilling . . . 2
1.3 Struktur . . . 4
1.4 Funn . . . 4
1.5 Implikasjoner . . . 5

2 Teori **6**
2.1 Datadrevet Innovasjon . . . 6
2.1.1 Dataverdikjeden . . . 8
2.1.2 Datadrevne beslutninger . . . 9
2.1.3 Big data . . . 10
2.1.4 Analyse av big data . . . 13
2.1.5 Kunstig Intelligens (AI) . . . 15
2.2 Innovasjon i datadrevne virksomheter . . . 16
2.2.1 Innovasjonsbegrepet . . . 17
2.2.2 Behovsdrevet innovasjon . . . 18
2.2.3 Innovasjonsradikalitet . . . 18
2.3 Former for innovasjon . . . 20
2.3.1 Produkt- og tjenesteinnovasjon . . . 20
2.3.2 Prosessinnovasjon . . . 20
2.3.3 Forretningsmodellinnovasjon . . . 21
2.4 Humankapital . . . 21
2.4.1 Kreativitet og innovasjon . . . 22
2.4.2 Utdanning og erfaring . . . 23

3 Metode **25**
3.1 Forskningstilnærming . . . 25
3.2 Forskningsdesign . . . 26
3.3 Forskningsstrategi . . . 26
3.3.1 Business case . . . 27
3.3.2 Spørreundersøkelse . . . 27
3.4 Innsamling av data . . . 28
3.4.1 Utvalg . . . 28
3.4.2 Struktur og utforming . . . 29
3.4.3 Business case og ekspertpanel . . . 30
3.4.4 Spørreundersøkelse . . . 31
3.5 Analyse av data . . . 32
3.6 Evaluering av metode . . . 32
3.6.1 Indre validitet . . . 33
3.6.2 Ytre validitet . . . 33
3.6.3 Reliabilitet . . . 34
3.7 Etiske aspekter . . . 34

4 Analyse **36**
4.1 Datasett . . . 37

4.2 Case . . . 38
4.2.1 Tidsbruk og antall ideer . . . 38
4.2.2 Resultat fra ekspertpanel . . . 39
4.3 Demografiske variabler . . . 40
4.3.1 Kjønn . . . 40
4.3.2 Alder . . . 41
4.3.3 Arbeidsstatus . . . 41
4.3.4 Årslønn . . . 41
4.3.5 Utdanning . . . 42
4.3.6 Erfaring og bakgrunn . . . 42
4.4 Faktoranalyse . . . 43
4.4.1 Er faktoranalyse hensiktsmessig? . . . 44
4.4.2 Antall faktorer . . . 46
4.4.3 Resultat . . . 46
4.4.4 Multippel lineær regresjon . . . 47
4.5 Multippel regresjon . . . 49
4.5.1 Ingen relevant bakgrunn . . . 50
4.5.2 Forretningsbakgrunn . . . 51
4.5.3 Innovasjonsbakgrunn . . . 52
4.5.4 Teknologibakgrunn . . . 52
4.6 Andre funn . . . 53
4.6.1 Ideer og tidsbruk . . . 53
4.6.2 Ideer . . . 54
4.6.3 Tidsbruk . . . 55
4.6.4 Alder og kjønn . . . 55
4.6.5 Arbeidstaker og student . . . 56
4.6.6 Høy og lav lønn . . . 57
4.6.7 Større modeller . . . 58
4.6.8 Beste modell . . . 61

5 Diskusjon **63**
5.1 Utdanning og bakgrunn . . . 63
5.1.1 Utdannings- og erfaringsbakgrunn . . . 64
5.1.2 Arbeidsstatus . . . 66
5.1.3 Lønn . . . 66
5.1.4 Ideer og tidsbruk . . . 67
5.2 Stor modell . . . 67

6 Konklusjon **70**
6.1 Begrensninger . . . 72
6.2 Videre forskning . . . 73

Referanser **74**

Appendiks **80**
A1 Qualtrics rapport . . . 80
A2 Case . . . 85
A3 Korrelasjonsmatrise . . . 92
A4 scree-test . . . 93

A5 Paralellanalyse 94
A6 Faktoranalyse diagram 95
A7 Multippel lineær regresjon total score og faktorer 96
A8 Regresjon ideer og tidsbruk 97
A9 Datasett for større modeller 99
A10 Multippel regresjon Modell 3 100

Figurliste

2.1 Ackoffs kunnskapspyramide (1989) 8
2.2 The Big Data Value Chain (Cavanillas et al., 2016) 9
2.3 Big data processes and terminology (Gandomi and Haider, 2015) 10
2.4 The three Vs of big data (Russom et al., 2011)) 11
2.5 Big Data Dimensions for Value Creation and Capture (Cappa, et. al., 2021) 13
2.6 Rammeverk for datadrevne virksomheter (Berndtsson et al. 2020) 17
A3.1 Korrelasjonsmatrise faktormodell 92
A4.1 scree-test 93
A5.1 Parallell analyse 94
A6.1 Faktoranalyse diagram 95
A7.1 Multippel lineær regresjon total score og faktorer 96
A10.1 Multippel regresjon Modell 3 100

Tabelliste

4.1 Datasett . . . 38
4.2 Gjennomsnitt rangering av ideer . . . 39
4.3 Kjønn . . . 40
4.4 Alder . . . 41
4.5 Arbeidsstatus . . . 41
4.6 Årslønn oppgitt i 1000 . . . 42
4.7 Utdanningsnivå . . . 42
4.8 Bakgrunner . . . 43
4.9 Kaiser-Meyer-Olkin (KMO) . . . 45
4.10 Kaiser-Meyer-Olkin (KMO) 2 . . . 45
4.11 Bartlett's Test of Sphericity . . . 45
4.12 Faktoranalyse . . . 46
4.13 Regresjon - Ingen relevant bakgrunn . . . 50
4.14 Regresjon - Forretningsbakgrunn . . . 51
4.15 Regresjon - Innovasjonsbakgrunn . . . 52
4.16 Regresjon - Teknologibakgrunn . . . 53
4.17 Kjønn gjennomsnitt . . . 55
4.18 Alder gjennomsnitt . . . 56
4.19 Student versus Arbeidstaker - Totalscore . . . 57
4.20 Lønn gjennomsnitt . . . 57
4.21 Kjønn - Høy og lav lønn . . . 58
4.22 Fem større modeller . . . 60
4.23 K-fold CV . . . 62
A8.1 Regresjon ideer og tidsbruk . . . 98
A9.1 Datasett for større modeller . . . 99

1 Innledning

1.1 Introduksjon

I 2018 viste en artikkel i Forbes (2018) til at 90% av verdens totale mengde data ble generert i de to foregående årene fra 2016 til 2018. Denne utviklingen var ventet å akselerere, og halveis inn i 2021 hadde mengden data registrert fra 2018 mer enn fordoblet seg. Oppdaterte estimater antyder at mengden data generert, lagret, kopiert og konsumert i 2025 vil være tredoblet fra 2021-nivået, og den eksponentielle veksten er ikke ventet å avta. Tall fra World Economic Forum estimerte i 2020 at antall bytes - dataens bestanddeler - i det digitale universet var 40 ganger større enn antall stjerner i det observerbare universet (SeedScientific, 2021). Omfanget av data er med andre ord allerede enormt og raskt voksende.

Begrepet *data* brukes i mange sammenhenger og fremstår i forskjellige former, avhengig av situasjon og anvendelse. Data kan forklares som informasjon, observasjoner og statistikk som samles fra ulike kilder og som kan sammenstilles, presenteres og forstås i en kontekst. Data kan både være kvantitativ og numerisk eller kvalitativ og tekstuell. Data kan brukes til å dele informasjon og kunnskap mellom mennesker og bedrifter, og spiller derfor en viktig rolle for for all kommunikasjon, samhandling og beslutningstaking i samfunnet. Foruten dataens åpenbare applikasjoner i hverdagslige aktiviteter som interaksjon og informasjonsinnhenting, er data også en viktig faktor for konkurranse og vekst i markedene. Det er bred konsensus om at selskapene som fremover evner å ta i bruk data i sine verdiforslag eller integrert i sin forretningsmodell vil ha de beste forutsetningene til å vinne frem i konkurransen.

I likhet med det tiltakende fokuset på data og de mange mulighetene anvendelse av data kan føre til, er *innovasjon* et annet populært begrep som stadig trekkes frem som et viktig fokusområde for selskaper. For at selskaper skal opprettеholde og utvikle sin konkurransekraft i markedet er de avhengige av å innovere. Innovasjon betyr i korte trekk problemløsing ved hjelp av nyskaping, og kan helt enkelt forklares ved å *forbedre eller lage noe nytt for å løse et problem"* (Gjelsvik, 2013). Innovasjonsaktiviteter basert på innsikt og kunnskap fra tilgjengelig data omtales derfor som **datadrevet innovasjon (DDI)**.

Ifølge Menon (2019) vil databasert verdiskaping og datadrevet innovasjon kunne utgjøre omlag 300 milliarder kroner, tilsvarende syv prosent av Norges totale BNP, i 2030. Rapporten peker på viktigheten av gode politiske og strukturelle rammebetingelser, men understreker også at virksomheter i større grad enn tidligere må ta i bruk og utnytte data de har tilgjengelig. Insentivene til å anvende data i fremtiden er med andre ord tilstede, for både private- og offentlige aktører.

Utnyttelse av data anses av de fleste som nødvendig for å følge den teknologiske og digitale utviklingen av verden vi lever i. Innsamling og anvendelse av data til verdiskapingsformål er ikke et nytt fenomen, men en viktig distinksjon som skiller fremtiden fra fortiden er introduksjonen av nye muliggjørende teknologier. *Kunstig intelligens(AI)*, *maskinlæring (ML)* og *stordata-analyse* er eksempler på slike teknologier som høyst sannsynlig vil føre til at dataens eksponentielle vekst i omfang og hastighet får enorm påvirkning på verdiskaping, innovasjon og økonomisk vekst i årene fremover.

På tross av at teknologiutviklingen virker å gå stadig hurtigere, samt at forretningsutvikling og innovasjon med all sannsynlighet vil preges av høy dataintensivitet fremover, er det uklart hvilken kompetanse som kreves for å lykkes med datadrevet innovasjon. Et betimelig spørsmål i så måte er om datadrevet innovasjon virkelig *er datadrevet innovasjon*, eller om det i realiteten dreier seg om dataintensiv problemløsing drevet av mennesker.

Hvis antagelsen om at data ikke nødvendigvis er den fundamentale driveren for *datadrevet innovasjon* er korrekt, er det interessant å undersøke *hvordan* og eventuelt *hvilken* type humankapital og demografiske trekk som kan stimulere til dataintensiv innovasjon.

1.2 Problemsstilling

Den enorme mengden data som allerede eksisterer, sammen med den tiltakende mengden som årlig genereres, skaper potensielt store muligheter for innovasjon og utvikling for selskaper. Myndighetene satser i økende grad på offentlig-privat samhandling for å stimulere utviklingen av datadrevet verdiskaping. Antagelsen masterutredningen bygger på er at viktige faktorer som tilgang til datamateriale og kompetanse allerede eksisterer. Det er imidlertid ikke åpenbart om det er dataintensiviteten eller kundebehov som er den viktigste driveren for innovasjon. Problemstillingen handler i forlengelsen av dette om hvorvidt datadrevet innovasjon er en oppgave for forretningsutviklerne med kommersiell

tankegang eller ingeniører med dypere teknisk innsikt. Problemstillingen åpner for at samarbeid mellom forskjellige bakgrunner er mer optimalt i arbeidet med dataintensiv problemløsing.

Forskningsspørsmålet relaterer til og antar at virksomheter besitter kompetent arbeidskraft med god kjennskap til utfordringene i selskapets verdiforslag. Likevel er det ikke åpenbart om dataintensiv innovasjonen i realiteten handler om teknisk problemløsing, kommersiell problemløsing, eller helt andre type egenskaper og humankapital. Det er flere faktorer som kan påvirke hvorfor det eventuelt er slik. Selskapers organisatoriske struktur kan medvirke til at anvendelse av data i innovasjonsaktiviteter hindres, for eksempel gjennom silotenking, hvor avdelingene ikke makter å dele data og kunnskap mellom seg. Andre steder kan utdaterte kjernesystemer og databehandlingsverktøy hemme innovasjonsevnen. For andre selskaper har data andre formål enn innovasjonsaktivitet, og selskapet makter ikke å se mulighetene i data som ellers brukes til kommunikasjon, vedlikehold eller rapportering. Mens noen virksomheter har dedikerte ressurser til å drive innovasjon og databehandling, har andre bare én av delene eller ingen av dem. Uansett struktur, verktøy eller tilgang til riktig humankapital erkjenner vi at det er en vanskelig oppgave å innovere i hele virksomheten gjennom tilstrekkelig kobling mellom den tekniske dataanalysen og det kommersielle kundeperspektivet.

Problemstillingen er dermed formulert som følgende;

Datadrevet innovasjon: Er datadrevet innovasjon en oppgave for økonomen eller teknologen?

Videre stiller følgende forskningsspørsmål som søker å belyser den overordnede problemstillingen;

1. *Hvilke demografiske trekk og humankapital påvirker individers evne til å generere gode ideer?*

For å diskutere forskningsspørsmålet vil vi ved hjelp av datagrunnlaget fra case og spørreundersøkelse se på hvilke utdanning- og erfaringsprofiler som scorer best. Spesifikt ønsker vi å teste om hvorvidt individer med teknologibakgrunn faktisk scorer høyest på dimensjoner som naturlig kan antas å være relatert til deres bakgrunn. På samme måte ønsker vi å teste om individer med forretningsbakgrunn scorer bedre på den kommersielle

dimensjonen. I tillegg til å teste ulike bakgrunner, vil vi utføre en rekke statistiske analyser og modelleringer med utgangspunkt i data fra casebesvarelser og spørreundersøkelse. Dette vil nærmere forklares og utdypes i metodekapittelet.

1.3 Struktur

I arbeidet med masterutredningen har vi samarbeidet med IT- og konsulentselskapet Knowit som jobber i skjæringspunktet mellom teknologi, forretningsutvikling og innovasjon. Vår sammenfallende interesse for bruk av data i innovasjonsarbeid var dermed et naturlig utgangspunkt for diskusjon rundt mulige problemstillinger, og det endelige resultatet vil forhåpentligvis kunne gi Knowit verdifull innsikt som kan brukes i både rekruttering og kundeprosjekter. I følge Østergaard et al. (2011) har mangfold i alder, kjønn og utdanning blant ansatte positiv effekt på innovasjon i et selskap, og ønsket med forskningen er derfor å bidra til å belyse hvilke trekk og egenskaper Knowit bør vektlegge i rekrutteringsprosesser og teamsammensetning på prosjekter.

For å vurdere den valgte problemstillingen og besvare det tilhørende forskningsspørsmålet vil vi benytte eksisterende teori og forskning. Utredningens datagrunnlag baserer seg på en caseoppgave som kartlegger respondentenes evne til idégenerering rundt en dataintensiv problemstilling. Alle ideene vil bli evaluert og scoret av et ekspertpanel fra Knowit langs følgende fire dimensjoner; **(1)** hvor *kreative* er ideene, **(2)** hvor *gjennomførbare* er ideene, **(3)** hvilket potensiale for *kommersialisering* ideene har **(4)**, og i hvilken grad *data* brukes i ideene. I tillegg vil datagrunnlaget baseres på en spørreundersøkelse som kartlegger respondentenes demografiske variabler og syn på de ulike dimensjonene.

1.4 Funn

Utredningens resultater viser at utdannelsesnivå er en klar positiv indikator for idegenerering, og understøtter med det antagelsen om humankapitals betydning for dataintensiv problemløsing. Analysen avdekker av individer med innovasjonsbakgrunn scorer best langs dimensjonene *kreativitet, kommersialisering og dataintensivitet*, mens individer med teknologibakgrunn viser seg å ha de mest *gjennomførbare* ideene.

Motsatt viser resultatene at individer uten relevant bakgrunn presterer dårligere langs

samtlige dimensjoner vi har definert som viktige for dataintensiv problemløsning. Utredningens funn indikerer videre at alder og erfaring er viktige faktorer, som sammen med utdannelsesnivå påvirker individers evne til idegenerering.

1.5 Implikasjoner

Funnene fra analysen vil være av interesse for alle virksomheter som har mål om å optimalisere sine innovasjonsaktiviteter. Særlig for virksomheter som befinner seg i skjæringspunktet mellom teknologi og forretning vil oppgavens funn kunne være interessante perspektiver i å utvikle arbeidet med innovasjon.

Manglende relevant eller feilallokert humankapital, kombinert med silo-tenkning mellom avdelinger er vanlige utfordringer selskaper som arbeider med dataintensive problemstillinger står overfor. Oppgavens funn bidrar derfor til å belyse hvilken humankapital som *faktisk er* relevant for å løse disse problemene i fremtidens datadrevne virksomheter.

Lav kjennskap til potensialet som ligger i nye muliggjørende teknologier og hvordan data kan anvendes til økt innsikt og bedre beslutninger er eksempler på andre utfordringer virksomheter har i dag. Tverrfaglige team og mer kunnskapsdeling vil kunne bidra til å øke virksomheters kunnskapskapital, gjennom optimalisert utnyttelse av de menneskelige ressursene i organisasjonen. Oppgaven adresserer derfor et naturlig startpunkt for å utforske mulighetene som datadrevet innovasjon representerer, nemlig hvilke kompetanse menneskene som skal utforske dem bør ha.

.

2 Teori

I denne delen av utredningen vil vi presentere det teoretiske fundamentet, som søker å belyse data, innovasjon og humankapital hver for seg, og i relasjon til hverandre. Teoridelen har som formål å gi tyngde til den påfølgende analysen og diskusjonen, og vil i sum belyse utredningens overordnede problemstilling og tilhørende forskningsspørsmål.

2.1 Datadrevet Innovasjon

Stadig flere virksomheter betrakter data som en nøkkelressurs og i økende grad som en integrert del av forretningsmodellen. Data har ikke en selvstendig iboende monetær verdi, men mengden kombinert med mulighet for deling og bruk kan utgjøre stor verdi for virksomheter som lykkes med å integrere data i sin forretningsmodell (Bulger et al., 2014). I et samfunnsperspektiv vil næringslivets evne til å nyttiggjøre seg av data som en ressurs også kunne ha stor innvirkning på et lands økonomi og det nasjonale næringslivets totale konkurransekraft (Menon, 2019).

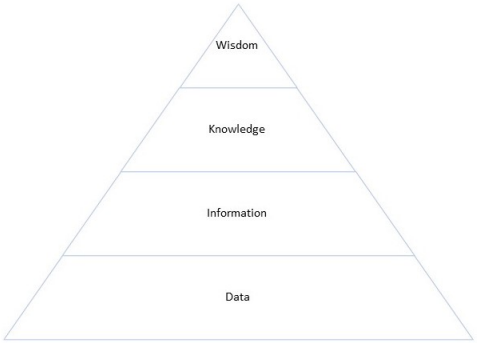
Direkte eller indirekte bruk av data til problemløsning, forbedring eller fornyelse, med formål om å skape verdi kan derfor beskrives som datadrevet innovasjon(DDI). I en rapport fra 2017 viser Organisasjonen for Økonomisk Samarbeid og Utvikling (OECD) til at datadrevet innovasjon er en forutsetning for vekst i dag og vil også være det i fremtiden (OECD, 2017). Globale aktører innen næringsliv og politikk viser økende interesse for DDI og peker på viktigheten av data for å utvikle nye innovative løsninger, som blant annet skal bidra til å lykkes i omstillingen til fornybarsamfunnet, utvikling av vaksiner og automatisering av transportsystemet. I tillegg ligger det store muligheter for finansiell vekst gjennom bedre tilgjengelighet av data, kostnadsreduksjon som følge av billigere teknologikomponenter og utvikling i datakraft. Den teknologiske utviklingen innen dataanalyse og lagring, kombinert med stadig økende dataproduksjon må antas å føre til store endringer i økonomien og samfunnet for øvrig (Jetzek et al., 2014).

"The exponentially growing production of data and the social trend towards openness and sharing are powerful forces that are changing the global economy and society" (Jetzek et al., 2014).

En av de viktigste mulighetene i DDI ligger i utvikling og forbedring av produkter, tjenester og prosesser. Brynjolfsson et al. (2011) viser statistisk at desto mer datadrevet en organisasjon er, desto mer produktiv er den. Videre viser Brynjolfsson et al. (2011) til at bedrifter som adapterer datadrevet beslutningstaking oppnår 5-6% høyere produktivitet og verdiskaping enn forventet. Barua et al. (2012) har undersøkt samme problemstilling ved å se på et større antall virksomheter, og kan dermed si mer om virkningen av effektiv databruk i næringslivet generelt. Analysen av 150 bedrifter viste at små forbedringer innen databehandling kunne utgjøre betydelig effektivitet- og avkastningsvekst (Barua et al., 2012). Sathi (2011) hevder at ved å bedre utnytte data i hele virksomheten skapes nye muligheter for utvikling av nye produkter og tjenester, som kan endre den eksisterende forretningsmodellen. Trenden de overnevnte studiene underbygger handler om overgangen fra passiv innsamling av data, til aktiv utnyttelse i utviklingen av nye produkter og tjenester (Brynjolfsson et al., 2011). Studiene indikerer at det finnes konsensus mellom akademia og OECD om at DDI allerede spiller en vesentlig rolle for vekst og gir anekdotiske bevis for at selskaper som lykkes med data også vil styrke sin konkurransekraft fremover(Barua et al., 2012).

Mulighetene som ligger i datadrevet innovasjon starter med å forstå hva data betyr i kontekst av innovasjon og verdiskaping. **Data** kan forstås som en informasjons-innsatsfaktor selskaper ønsker å tilegne seg, blant annet til anvendelse og beslutningsstøtte for innovasjonsarbeid i organisasjoner. I følge Cronholm et al. (2017) har tilgang og bruk av data det siste tiåret i større grad blitt avgjørende for å overleve i konkurransesituasjonen. Dette har sammenheng med at stadig flere virksomheter evner å bruke innsikten data kan gi til verdiskapende aktiviteter. Databegrepet brukes vilkårlig om ulike typer informasjon, ofte avhengig av kontekst. Rowley (2007) fremhever at data i seg selv ikke har iboende og spesifikk verdi, og må derfor organiseres og prosesseres for å tilføres egenverdi. For å forstå hvordan organisering og prosessering tilfører data verdi kan *Ackoffs kunnskapspyramide* gi nyttig innsikt 2.1.

Ackoff (1989) forklarer ved hjelp av en pyramide hvordan en dataprosess starter med innsamling og anvendelse i bunnen, og gir gjennom stegene en forståelse av hvordan data fungerer som et drivstoff for utvikling og forbedring oppover i pyramiden. 2.1 visualiserer de fire stegene i: *Data, Information*, *Knowledge* og *Wisdom*.



Figur 2.1: Ackoffs kunnskapspyramide (1989)

I det nederste steget finner vi data uten struktur og i sin opprinnelige form. I neste steg prosesseres data, som kan øke dataens nytteverdi. Forskjellen mellom de to første trappene, *data*, og *informasjon* handler derfor om økt funksjonalitet gjennom prosessering og strukturering. Det neste steget i pyramiden - *Kunnskap* - handler om den *riktige* måten å samle inn informasjon på, som gjør *informasjonen* nyttig. Pyramidens øverste steg henviser til hvordan visdom på toppen er nødvendig for å kunne forstå informasjon og kunnskap, som igjen stammer fra data. Mennesker kan bruke sin visdom til å svare på spørsmål om *"hvorfor"* og sin kunnskap til å besvare *"hvordan"*. For virksomheter er det derfor nødvendig å bevege seg oppover i pyramiden og ta beslutninger basert på kunnskap. Oppsummert kan man derfor i Ackoff (1989) sin kunnskapspyramide anse data og informasjon som fundamentet virksomheter bygger sin kunnskapskapital på (Cricelli og Grimaldi, 2008).

2.1.1 Dataverdikjeden

Verdikjeder ble først introdusert av Michael Porter i 1985 og brukes til å forklare de strategiske aktivitetene i en virksomhet (Porter, 1985). Kunnskap og forståelse av verdikjeder er derfor svært viktig for å kunne vurdere ressursallokering og dermed ytelsen til selskapet. Ved å forstå verdikjeden vil beslutningstakere enklere kunne identifisere flaskehalser eller mulighetsrom i organisasjonen, og ved hjelp av data utbedre og utnytte disse (Porter, 1985). En analyse av organisasjonens verdikjede kan derfor gi verdifull innsikt og tegne et bilde av hvordan tilgjengelig data samles inn, analyseres, håndteres og lagres, før det kan brukes til verdiskapende aktiviteter(Cavanillas et al., 2016). Figur 2.2 forklarer og illustrerer stegvis hvordan aktivitetene transformerer verdiløs og uleselig data til prosessert og verdifull innsikt gjennom verdikjeden.

Figur 2.2: The Big Data Value Chain (Cavanillas et al., 2016)

Datainnsamling - Innsamling og filtrering av data som videre blir lagt inn i datavarehus eller andre lagringsløsninger.

Dataanalyse - Utforske, endre og modellere data slik at den kan illustreres og anvendes til videre bruk. Cavanillas et al. (2016) belyser tre viktige egenskaper ved dataanalyse: datautvinning, forretningsanalyse og maskinlæring.

Datakurering - Håndtering av data gjennom hele dens livssyklus for å sikre kvaliteten for videre bruk. Datakurering sikrer at data er tilgjengelig, gjenbrukbar, pålitelig og er riktig til det formålet den er hentet ut for (Pennock, 2007).

Datlagring - Oppbevaring av data stiller særlige krav til tilgjengelighet, struktur og sikkerhet. Med dette menes for eksempel datavarehus i skyen og databasesystemer som strukturerer data på en tilfredsstillende måte og som videre sikrer enkel tilgjengelighet for rapportering, analyse og visualisering (Cavanillas et al., 2016).

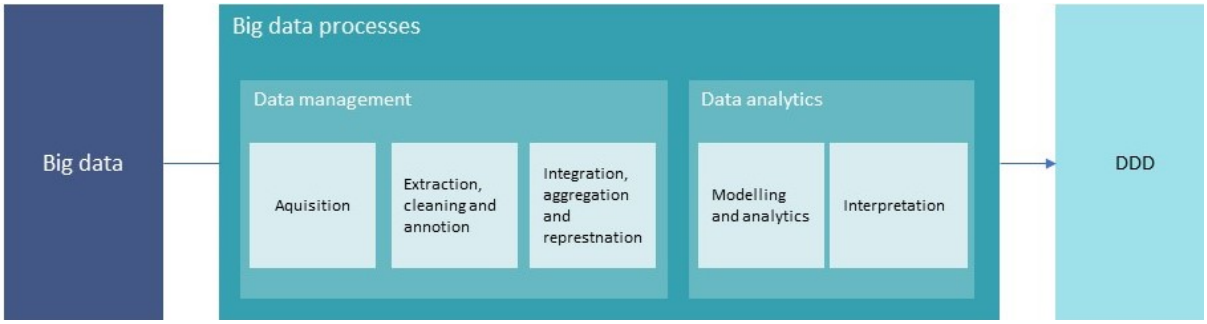
Databruk - Bruk av data for å ta riktige beslutninger kan videre kan gi økt konkurransekraft og lavere kostnader.

2.1.2 Datadrevne beslutninger

Jacobsen og Thorsvik (2013) refererer til Herbert Simon som for over 40 år siden forklarte at informasjonsteknologi har potensiale til å revolusjonere måten forretningsbeslutninger blir fattet på. Videre hevdet Herbert at datamaskinens evne til datainnsamling og databehandling vil kunne heve organisasjoner til et nytt nivå av rasjonalitet (Jacobsen og Thorsvik, 2013). Ved bruk av data kan beslutninger tas basert på analyse istedenfor intuisjon, og til dels vil den menneskelige innflytelsen på beslutningstaking reduseres. Videre vil målet med slike virkemidler være å forbedre beslutningene, ved å kombinere innsikten data gir med menneskelig intuisjon (Provost og Fawcett, 2013). I dag omtaler vi *"Business intelligence"*- verktøy som datamining, datavarehus, big data analytics og bruken av internett som viktige ressurser for at ansatte og ledere kan fatte de beste og

riktige beslutningene (Turban et al., 2010).

Datadrevne beslutningssystemer benyttes i dag til ulike formål som; strategiske forretningsplaner, operasjonelle aktiviteter, prestasjonsmåling i sanntid og håndtering av kunderelasjoner. Power (2008) hevder at ved å benytte systemer som analyserer data i sanntid sammen med historisk data, kan beslutningstakere fatte bedre beslutninger. I 2002 ble det utført en større undersøkelse av McAfee (2012) hvor en så på effekten dataintensive beslutninger har på organisasjoner. Resultatene viste at høy grad av databasert beslutningsstøtte ga en høyere grad av produktivitet og undersøkelsen fant også korrelasjon mellom datadrevet beslutning og finansiell avkastning. (McAfee et al., 2012). Eksemplene føyer seg dermed inn i rekken av argumenter for å øke graden av databaserte beslutninger i virksomheter. Figur 2.3 visualiserer hvordan analyse og prosessering av store datamengder fører til beslutninger hvor menneskelig tolkning er det siste steget før beslutninger fattes (Provost og Fawcett, 2013).



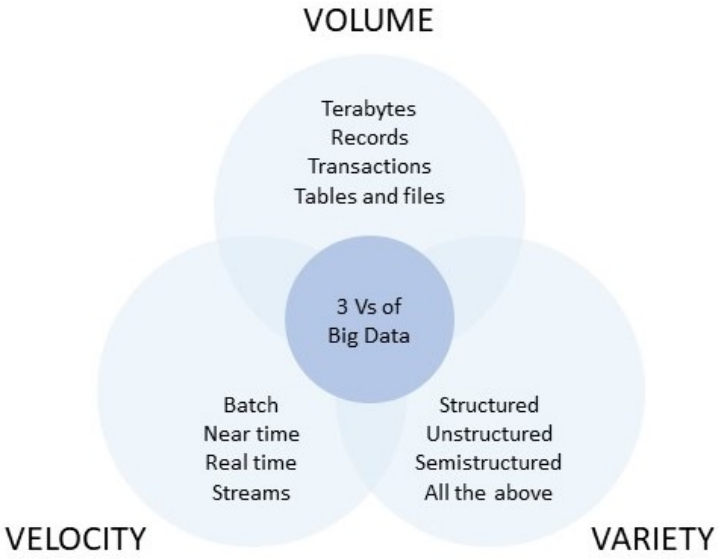
Figur 2.3: Big data processes and terminology (Gandomi and Haider, 2015)

2.1.3 Big data

Big data viser til data som er for stor, mangeartet og ustrukturert for tradisjonelle datainnsamlingsmetoder og som derfor krever eksepsjonell prosessorkraft for å håndteres (Provost og Fawcett, 2013). Informasjon og kunnskap som kan hentes ut fra store datasett har potensiale til å påvirke ytelse og forbedre beslutningstaking i den enkelte virksomhet (McAfee et al., 2012). På samfunnsnivå kan big data muliggjøre nye næringer, prosesser og tjenester, gjennom mer omfattende analyse av for eksempel bevegelsesmønstre, helsedata eller værdata (SNL, 2019).

Big data handler ikke utelukkende om omfanget av data, men også om dataens hastighet og variasjon. Russom et al. (2011) illustrerer de tre komponentene og presenterer trekk

ved hver av de *tre V'ene*.



Figur 2.4: The three Vs of big data (Russom et al., 2011))

Volum - Mengden og omfanget av data er allerede eksepsjonelt, hvilket er en naturlig konsekvens av de siste tiårenes digitalisering av samfunnet. Data produseres i nærmest samtlige interaksjoner og aktiviteter vi foretar oss daglig, og den teknologiske utviklingen sørger for at det blir enklere, rimeligere og mer tilgjengelig for hver dag som går (Davenport, 2014).

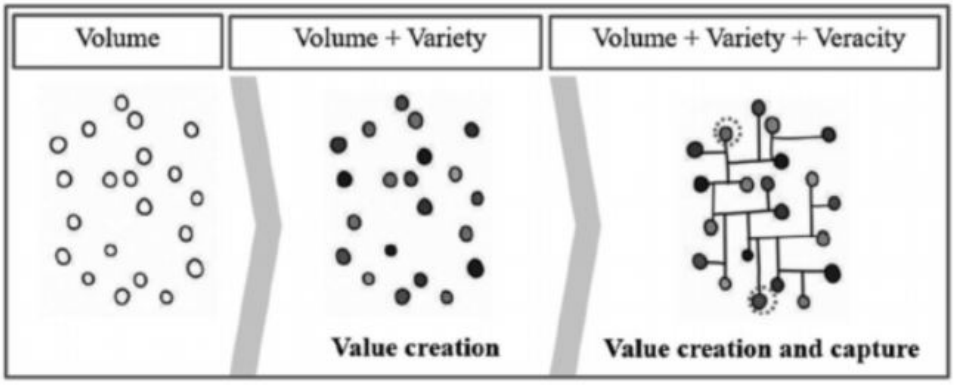
Variasjon - Ettersom dataproduksjonen er stadig tiltakende, er det naturlig at datavariasjonen også øker. Datasett kan inneholde langt mer informasjon enn fødselsdato eller blodtype, og ved å krysse ulike datakilder kan nye typer data skapes. I et forretningsperspektiv kan variert data muliggjøre hensiktsmessige sammenstillinger og bidra til å predikere kjøpsadferd eller avdekke kundebehov. Gjennom tilgang til varierte data kan adferd bli enklere å predikere, for eksempel gjennom søkehistorikk på nett, bevegelsesmønstre og demografiske variabler (Ransbotham et al., 2015). Selv om variasjon og spenn i datakilder har mange positive implikasjoner, argumenterer Huang et al. (2020) for at virksomheter som samler inn mye data må være varsomme, og indikerer at for stor variasjon vil gjøre data uhåndterlig og dermed redusere potensiell gevinstrealisering. George et al. (2014) peker på ledelse, organisering og strategiutforming som essensielt i beslutninger knyttet til datainnsamling, og understreker at det er en menneskelig oppgave å avgjøre hvor bred innsamlingen av data bør være.

Velositet - Fart i denne sammenheng henspiller på hvor raskt og smidig organisasjoner

evner å prosessere innsamlet data. For næringsdrivende er det vitalt at innsamlet data ikke blir utdatert før den har rukket å bli anvendt til forretningsformålet. Dette stiller krav til både verktøy og kompetanse i virksomheten, som kan være avhengig av å analysere data i sanntid for å tilpasse verdiforslag eller implementere verdiskapende endringer. Crié og Micheaux (2006) presenterer tre vanlige forsinkelser som kan bidra til at dataens holdbarhet utløper før den blir anvendt i verdiskapende aktiviteter.

1. **Dataforsinkelser** - Forsinkelser som oppstår i innsamling- eller lagringsprosessen, som for eksempel nettverksbrudd eller serverproblemer.
2. **Analytiske forsinkelser** - Forsinkelser som oppstår under analyse og databehandling, som kan forsinke presentasjonen av data til beslutningstakere.
3. **Ventetid hos beslutningstakere** - Den menneskelige forsinkelsen kan oppstå som følge av de to andre, men kan også selvstendig oppstå hvis beslutningstakere ikke evner eller har vilje til å fatte beslutninger i tide. Det menneskelige aspektet handler både om hvordan data presenteres for beslutningstakere, men også evnen beslutningstakere har til å forstå data som blir presentert.

Forskere virker å være samstemte om at stordata innebærer enorme muligheter for verdiskaping. Samtidig utgjør stordata en betydelig risiko for virksomheter som ikke evner å håndtere dataene på en god måte. Dette kan potensielt innebære store kostnader ved både innsamling, analyse og datakurering. Om organisasjoner ikke håndterer data på riktig måte og raskt klarer å ta de bedre beslutninger enn de ville gjort uten, kan kostnader knyttet til innsamling, analyse og kurering raskt overgå gevinsten. Oppsummert starter derfor optimal utnyttelse av store datamengder allerede i innsamlingsfasen, hvor omtenksomhet i relasjon til volum, variasjon og velositet spiller en avgjørende rolle Russom et al. (2011). Kompleksiteten av de tre V'ene er illustrert i figur 2.5 nedenfor.



Figur 2.5: Big Data Dimensions for Value Creation and Capture (Cappa, et. al., 2021)

2.1.4 Analyse av big data

Begrepet *Big Data Analytics* refereres ofte til som "*Business Analytics*" eller "*Data Analytics*". Begrepene brukes ofte om hverandre, men faller alle inn i analyse-kategorien, hvor en anvender ulike kraftfulle analytiske verktøy for å omdanne data til verdifull innsikt (Saunders et al., 2016). Analyse på dette nivået har tradisjonelt vært brukt til forskningsformål, for eksempel bruk av statistiske programvarer for å finne korrelasjoner eller utføre regresjonsanalyser. Tilsvarende analyseverktøy har de siste årene blitt utviklet til bruk i forretningsøyemed. Økende datamengder kan dermed føre til verdiskaping gjennom predikative analyser av kunders behov, innsikt i kundeadferd og i økende grad generell beslutningsstøtte i organisasjoner (Wamba et al., 2015).

Big Data Analytics er i dag et viktig område som det satses på gjennom nye utdanningsretninger i akademia og opplæring internt i virksomheter. Analyse av store datamengder med kraftige analyseverktøy har påvirket dataintensiviteten i en rekke industrier og markeder, og virksomheter er derfor avhengige av både datatilgang og kompetanse for å omgjøre den til meningsfull innsikt. Garmaki et al. (2016) peker på effektive prosesser som en viktig forutsetning for å lykkes med *big data analytics*, og i tillegg spiller riktig kompetanse en vesentlig rolle i å utvikle disse prosessene.

Big Data Analytics beskriver dermed teknologi og teknikker som benyttes for å innhente verdifull informasjon. En analytisk prosess kan hjelpe med å beskrive hva som har skjedd og hva som kommer til å skje, og benyttes derfor som et beslutningsverktøy (Kwon et al., 2014). Eksempler på analyser som kan benyttes i slike prosesser er; statistisk analyse, tidsserieanalyse, sentimentanalyse, dataforvaltning, og avanserte visualiseringsverktøy

(Kwon et al., 2014).

Manyika et al. (2011) hevder at Big data er i ferd med å endre konkurransesituasjoner gjennom nye prosesser, nye industrielle økosystemer og nye innovasjoner. Analyse av Big data har potensiale til å endre eksisterende bransjer og er et av de heteste områdene innen forskning i dag (Wamba et al., 2015). Analyse av store datamengder er ikke lengre forbeholdt de akademiske kretser. Næringslivet har for lengst skjønt at teknologi kan gi varige konkurransefortrinn, og konsulentselskapet McKinsey & Company peker på at teknologi blir stadig billigere, mer demokratisert og dermed fører til økt produktivitet blant virksomheter (Perrey, 2015). McKinsey, som blant annet spesialiserer seg på å bistå selskaper med innovasjonsprosesser trekker frem tre viktige retningslinjer som virksomheter bør følge for å lykkes med å integrere avansert analyse i forretningsmodellen (Perrey, 2015).

1. **Still de riktige spørsmålene** - Å stille de riktige spørsmålene i starten av analytiske prosesser er kritisk for å kunne håndtere den enorme mengden data som prosesseres. Man ender derfor ofte opp med store datamengder som er vanskelig å strukturere. Klare problemstillinger sammen med riktige spørsmål øker sjansen for at virksomheten evner å identifisere den essensielle informasjonen, og dermed fatte riktige beslutningene basert på relevant innsikt (Perrey, 2015).

2. **Vær kreativ** - Det er ikke alltid slik at mer data gir større verdi, på tross av at større datamengde kan gi et klarere bilde på muligheter og risiko. Det er derfor viktig å være kreativ med data virksomheten har tilgjengeligPerrey (2015).

3. **Gjør analysen lett** - Mye informasjon kan være overveldende og det kreves dermed rapporter og analyser som er lett å forstå. Det er fare for at komplekse rapporter ikke blir tatt i bruk. Det er viktig å tenke på at alle i organisasjonen skal forstå analysene (Perrey, 2015).

Som McKinsey påpeker, handler dataanalyse i stor grad om de menneskelige handlingene som etterfølger datainnsamling, og underbygger viktigheten av å ha kompetente medarbeidere som evner å stille de riktige spørsmålene, være kreative og presentere data på best mulig vis.

2.1.5 Kunstig Intelligens (AI)

Store Norske Leksikon (2020) beskriver kunstig intelligens som datamaskiner som er i stand til å løse problemer og lære av sine egne erfaringer. CEO i Google, Sundar Pichai har hevdet at kunstig intelligens kan være noe av det viktigste mennesker jobber med å utvikle, og sidestiller AI med oppdagelser som ild og elektrisitet (Clifford, 2018). Det man ønsker å oppnå med kunstig intelligens er at datamaskiner etterligner mennesker (Gambus og Shafer, 2018). Man programmerer datamaskiner til å lære av sine erfaringer slik som mennesker gjør, i en prosess som kan beskrives som "læring" (Samuel, 1967).

Kunstig intelligens er basert på maskinlæring (ML) som nesten utelukkende er bygget på *"neural networks"*, eller nevrale nettverk (Gambus og Shafer, 2018). I et datadrevet læringssystem er et nevralt nettverk en læringsalgoritme, som innebærer funksjonalitet som forsøker å forstå input-data og forme det til hensiktsmessig output-data. Disse nettverkene er inspirert av hvordan nevroner i menneskers hjerne fungerer, hvor signaler fra sansene våre registreres og brukes til å fatte menneskelige beslutninger. I dag brukes kunstig intelligens i nevrale nettverk innenfor flere fagfelt, for eksempel som diagnoseverktøy innen helse, eller mer hverdagslige funksjoner som filtrering av uønsket epost og opplåsing av telefoner med ansiktsgjenkjenning (DeepAI, 2021).

Kunstig intelligens handler på samme måte som Big Data om å øke volum, hastighet og variasjon av data. Kobler man kunstig intelligens til Big Data åpner man et mulighetsrom for å identifisere avanserte sammenhenger, kompleks læring og utførelse av databaserte løsninger. Et eksempel på hvordan dette gjøres i dag er innen verdipapirhandel. I dag utføres over halvparten av all handel med verdipapirer på verdens børser ved hjelp av maskinlæringsalgoritmer, som bidrar til å øke hastigheten betraktelig (O'Leary, 2013),

Ved hjelp av kunstig intelligens og Big Data kan ustrukturert data bli fanget opp, lagret og strukturert (O'Leary, 2013). Slik data er ofte svært kompleks og har store variasjoner. Det gjør at den verdifulle informasjonen er vanskelig å analysere uten hjelp av kunstig intelligens.

Rainie og Anderson (2017) peker på bekymringer som knyttes til bruken av kunstig intelligens og hvordan folk flest er engstelige for effekten den kan ha på livene våre. På 80-tallet ble kunstig intelligens presentert som en støtte for beslutninger, som ikke skulle

erstatte mennesker (Silver, 1991). I dag ser vi en annen betydning av begrepet kunstig intelligens, hvor mange beslutninger og prosesser er automatisert med lite innblanding fra mennesker (Markus, 2015). Kolbjørnsrud et al. (2016b) viser til at ledere bruker 54% av tiden sin på administrative oppgaver og bare 10% av tiden på strategi og innovasjon. Det er nettopp disse administrative oppgavene kunstig intelligens og Big Data vil automatisere, noe som igjen kan gi menneskene mer tid til å jobbe med strategi og innovasjon (Kolbjørnsrud et al., 2016a).

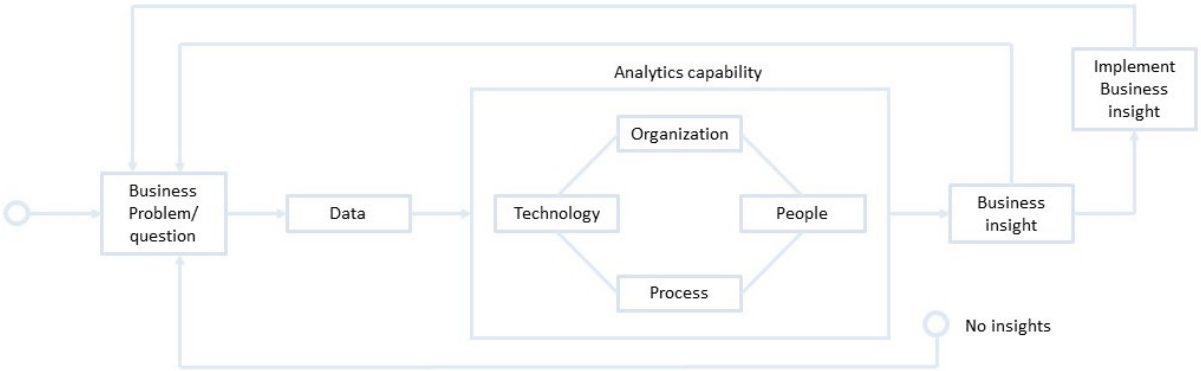
2.2 Innovasjon i datadrevne virksomheter

Selskaper som har tilgang til data og kompetansen til å prosessere den vil kunne bevege seg i en mer datadrevet retning. Berndtsson et al. (2018) peker imidlertid i sin forskning på både utfordringene og mulighetene som ligger i overgangen til å bli en datadrevet virksomhet. Berndtsson et al. (2020) fremhever særlig kompetanse knyttet til analyse- og databehandling som essensielt, og understreker hvor stor påvirkning disse aktivitetene kan ha på beslutningstaking. I tillegg til kompetanse innen analyse- og databehandling peker McAfee et al. (2012) på ledelse og organisasjonsstruktur som viktige forutsetninger for å lykkes som en datadrevet virksomhet. Han indikerer samtidig at det også her finnes potensielle utfordringer. Om ledelsen ikke selv har de riktige insentivene til å bli mer datadrevet, vil heller ikke organisasjonen forøvrig ha den nødvendige støtten til å utvikle de interne prosessene i en datadrevet retning (McAfee et al., 2012).

Berndtsson et al. (2020) presenterer modell 2.6 som visuelt beskriver hvordan selskaper løser forretningsproblemer ved hjelp at data. Analyse-ressursene som beskrives i rammeverket deles inn i *organisasjon*, *humankapital*, *teknologi* og *prosesser*, og utgjø virksomhetens samlede analysekompetanse. Ved å analysere tilgjengelig data i sammenheng med forretningsproblemet er det tre potensielle utfall; hverken ny innsikt eller løsning, verdifull innsikt uten konkret løsning, eller verdifull innsikt som løser problemet (Berndtsson et al., 2020).

I Berndtsson et al. (2020) sitt datarammeverk spiller menneskene en viktig rolle i kraft av å anvende sine egenskaper til å transformere data til løsninger på problemer. Ettersom mange forretningsutfordringer er dataintensive vil selskaper ønske å anvende hele sin analytiske verktøykasse, hvor menneskelige egenskaper utgjør en vesentlig rolle.

Rammeverket illustrerer dermed godt hvordan datadrevet innovasjon kan se ut i en dataintensiv virksomhet.



Figur 2.6: Rammeverk for datadrevne virksomheter (Berndtsson et al. 2020)

2.2.1 Innovasjonsbegrepet

Data i seg selv kan ikke løse problemer, og som tidligere nevnt har ikke data en iboende egenverdi uten bearbeidelse og analyse. Gandomi og Haider (2015) understreker dette ved å peke på at potensiell verdiskaping først skjer når data anvendes som beslutningsstøtte i for eksempel innovasjonsprosesser.

***Big data is worthless in a vacuum. Its potential value is unlocked only when leveraged to drive decision making."* (Gandomi og Haider, 2015).**

I litteraturen finnes en lang rekke definisjoner av innovasjonsbegrepet, ofte påvirket av hvilken kontekst eller bransje begrepet brukes i. Ordet «innovasjon» kan spores tilbake til før vår tidsregning, men først på 1940-tallet introduserte Schumpeter et al. (1939) begrepet i en økonomisk kontekst. Schumpeter et al. (1939) utfordret den ny-klassiske økonomiske tankegangen om vekst som en funksjon utelukkende av tilgangen til kapital og arbeidskraft, og definerte innovasjon som "*ny bruk av kunnskap og ressurser suksessfullt anvendt i markedet*". Den sosiale og nytenkende funksjonen i et forretningsmessig perspektiv var i følge Schumpeter et al. (1939) nyansen som skilte innovasjon fra oppfinnelser.

I følge Cobbenhagen (2000) er det bred enighet om at innovasjon både kan være en kilde til vedvarende konkurransefortrinn og følgelig en viktig driver for vekst. Innovasjon er derfor en svært viktig faktor for selskapets langsiktige sjanser for å overleve i en konkurranseutsatt markedssituasjon (Cho og Pucik, 2005). Motsatt kan innovasjon også være en medvirkende

faktor til selskapers undergang, om de ikke makter å tilpasse seg innovative løsninger fra konkurrenter (Iden et al., 2013).

En rapport fra OECD (2005) peker på at innovasjon typisk foregår i kontinuerlige prosesser, og at initiativer og aktiviteter ofte innebærer inkrementelle endringer i tjenester eller produkter. På grunn av dette peker OECD på at innovasjonsprosesser kan være vanskelig å identifisere isolert sett, men at totaliteten av aktivitetene sett i en kontekst av innovasjon, gjør at man over tid kan se spor av innovasjonsprosesser i organisasjoner.

2.2.2 Behovsdrevet innovasjon

Kritikere av datadrevet innovasjon fremhever at data-analyse alene ikke kan predikere kjøpsadferd eller positivt avdekke kundebehov. Det er altså ikke nødvendigvis en kausal sammenheng mellom mer data og innsikt (Christensen et al., 2016b). Behovsdrevet innovasjon handler om å forstå kundens eksisterende og fremtidige behov, og bruke innsikten aktivt i innovasjonsprosessen (Reitan, 2011). Christensen et al. (2016b) peker på at at kundebehovet må stå sentralt, uavhengig om problemstillingen er dataintensiv. Videre argumenterer han for at problemløsing bør være utgangspunktet for å innovere produkter, tjenester, prosesser eller hele forretningsmodeller.

En av de fremste rammeverkene som forklarer behovsdrevet innovasjon er Christensen et al. (2016b) sitt *job-to-be-done*-perspektiv, som tar utgangspunkt i at en kunde etterspør løsninger på problemer fremfor spesifikke produkter eller tjenester. Om en person kjøper et verktøy på i en forretning, er det ofte ikke fordi personen ønsker å kjøpe produktet i seg selv. Personen ønsker kun produktets hjelp til å løse et problem, og det er derfor tydelig at kundens behov for hjelp er den drivende kraften for kjøpet (Christensen et al., 2016a). *Jobs-to-be-done*-perspektivet utelukker ikke dataintensiv innovasjon, men fremhever kundebehovet som den sentrale driveren innovasjonsaktiviteter og problemløsing. Uavhengig av hvor god tilgang til data eller analyseverktøy selskaper har, er virksomheter avhengig av å identifisere og løse kundenes problemer.

2.2.3 Innovasjonsradikalitet

Innovasjonsradikalitet handler om hvor langt unna innovasjonen er eksisterende kunnskap og teknologi (Fagerberg et al., 2005). I en litteraturstudie om innovasjonsprosesser definerer

Iden et al. (2013) *radikal innovasjon* som *«innovasjon uttrykt i store endringer, som i liten grad er basert på eksisterende teknologi, kunnskap eller aktiviteter, og som ofte er utviklet for å møte behovene til nye og potensielle kunder»*. I den motsatte enden definerer Iden et al. (2013) *inkrementell innovasjon* som *«innovasjon uttrykt i små endringer basert på eksisterende teknologi, kunnskap eller aktiviteter, og som ofte er utviklet for å møte behovene til eksisterende kunder»*. Der inkrementell innovasjon kan påvirke kostnadsstruktur eller funksjonaliteten til en tjeneste eller et produkt, vil en radikal innovasjon kunne medføre en dramatisk endring som transformerer en hel bransje eller skape en nytt marked (Malerba, 2005).

På tross av innovasjonsbegrepets ulike fasetter og nyanser, er det rimelig å stadfeste at innovasjon innebærer elementer av nytenking og kommersialisering. Like fullt er det stor variasjon i hvordan innovasjonsprosesser fremstår på tvers av bransjer og sektorer. Malerba (2005) peker på ulik teknologisk utviklingshastighet, ulikt forhold til kunnskap og stor variasjon i strukturelle drivere som årsaker til at innovasjonsaktiviteter kan ha ulik form . Disse faktorene vil derfor også kunne virke inn på om en sektor preges av store endringer og radikal innovasjon, eller mindre endringer og inkrementell innovasjon (Malerba, 2005).

I en kompetitiv markedssituasjon er selskaper avhengig av å finne balansen mellom å drive radikal og inkrementell innovasjon. Haanæs (1999) peker på to tilnærminger som knytter seg til radikal og inkrementell innovasjon, som forklarer hvordan selskaper kan utøve denne balansegangen og utforme sin strategi for innovasjon.

1. En *utforskende tilnærming* innebærer at man søker å utforske ukjente felt og dermed lete etter nye teknologiske og kommersielle muligheter.

2. I motsatt ende vil et selskap kunne innrette sin innovasjonsstrategi ved hjelp av det Haanæs (1999) beskriver som en *utnyttelse-tilnærming*, som innebærer at man utnytter ideer, kunnskap eller teknologi som allerede finnes i større grad enn tidligere.

Utforskning kan dermed knyttes til radikal innovasjon, mens *utnyttelse* av nåværende ressurser knyttes til inkrementell innovasjon.

2.3 Former for innovasjon

Hvordan individer og selskaper jobber med innovasjon avhenger av hva problemetstillingen handler om, og formen for innovasjon vil derfor kreve ulik tilnærming og kompetanse. For å kunne forstå mulighetene som ligger i datadrevet innovasjon er det derfor hensiktsmessig å se nærmere på ulike måter individer og virksomheter kan jobbe med innovasjon på.

2.3.1 Produkt- og tjenesteinnovasjon

I produkt- og tjenesteinnovasjon står kundeperspektivet og problemløsing sentralt, hvor selskaper gjennom produkt- og tjenesteutvikling kan løse problemer, og på samme tid skape verdi for kunden og selskapet (Tucker, 2002). Pedersen og Nysveen (2010) peker på at produktinnovasjon typisk foregår gjennom formaliserte og velstrukturerte utviklingsprosesser, mens tjenesteinnovasjon ofte baseres på prøving og feiling. Spesielt fraværet av muligheten til å produsere fysiske prototyper trekkes frem som en årsak til at tjenesteinnovasjoner ikke alltid foregår i strukturerte prosesser, i motsetning til produktinnovasjon.

Tjenesteytelser utgjorde i følge Chesbrough (2011) om lag 60% av den globale økonomiske aktiviteten i 2012, og understreker med dette hvor viktig tjenesteinnovasjon er for verdens samlede verdiskaping. Selskaper er nødt til å kontinuerlig forbedre og fornye sine verdiforslag for å nå opp i den globale konkurransen. Chesbrough (2011) anbefaler en åpen innovasjonstrategi for å lykkes med kontinuerlige forbedringer, hvor virksomheter utnytter strømninger og impulser fra både interne og eksterne kilder til å drive innovasjon.

2.3.2 Prosessinnovasjon

OECD (2005) definerer prosessinnovasjon som en implementering av endringer i teknikker og utstyr med målsetting om å forbedre en bedrifts produksjonsprosesser. Ofte vil denne typen innovasjon basere seg på kostnadsreduksjon eller effektivisering av interne prosesser, eller forbedring av metoder for å levere produkter og tjenester til kunder (OECD, 2005). Ifølge Davenport (1993) må prosessinnovasjon være en radikal endring i hvordan et selskap utfører en forretningsaktivitet, og skiller seg dermed fra nyere definisjoner som ikke spesifikt peker på radikale endringer som et vilkår for prosessinnovasjon.

2.3.3 Forretningsmodellinnovasjon

I likhet med Cobbenhagen (2000) oppfatning av det generelle innovasjonsbegrepet som en kilde til konkurransefortrinn, peker Markides og Charitou (2004) på forretningsmodellinnovasjon (BMI) som en mulig måte å skaffe seg vedvarende konkurransefortrinn overfor sine konkurrenter. Ved å endre en forretningsmodell fremfor et produkt, tjeneste eller prosess kan et selskap systematisk og helhetlig utfordre organisasjonens fundament, og dermed innovere (Amit og Zott, 2012). I følge D'Aveni et al. (2004) er det ikke mulig for selskaper å tilegne seg varige kostnads- eller kvalitetsfortrinn, forutenom *evnen* til å innovere nye kostnads- og kvalitetsfortrinn. D'Aveni et al. (2004) peker på at selskapets fokus på innovasjon må være mer helhetlig og rettet mot forretningsmodellen, fremfor tradisjonell produkt- og tjenesteinnovasjon.

Forretningsmodellinnovasjon er imidlertid en krevende øvelse ifølge Christensen et al. (2016a), som peker på manglende ledelse som den fremtredende årsaken til at selskaper ikke makter å forbedre sin forretningsmodell gjennom innovasjon. I tillegg til manglende innovasjonsledelse viser Christensen et al. (2016a) til forskningsfeltets kompleksitet og manglende forskning som to forklaringsfaktorer for hvorfor selskaper ikke lykkes med BMI.

2.4 Humankapital

I rammeverket til Berndtsson et al. (2020) er organisasjonen og menneskene de utøvende driverene blant de analytiske ressursene selskapet besitter. Rammeverket beskriver hvordan data transformeres til løsninger på forretningsproblemer, og illustrerer selskapets analytiske verktøykasse, se figur 2.6. Menneskenes utdanning og erfaring er med andre ord viktige faktorer som påvirker hvordan en dataintensiv virksomhet lykkes med datadrevet innovasjon.

Humankapital eller menneskelig kapital defineres av OECD (2007) bredt som en blanding av enkeltpersoners medfødte evner, sammen med ferdighetene og kunnskapene de erverver seg gjennom utdanning og opplæring. I et forretningsperspektiv betegnes menneskelig kapital hovedsakelig som arbeidsstokkens ferdigheter og talenter i direkte relevans til bedriftens eller bransjens suksess (OECD, 2007).

2.4.1 Kreativitet og innovasjon

For å stimulere til og tilrettelegge for innovasjon i organisasjoner vil kvalitetene og kunnskapene de ansatte besitter være svært viktig (Mumford, 2000). Kreativitet og idégenerering har en naturlig sammenheng, og sammen utgjør disse to det første steget i en innovasjonsprosess (Oldham og Cummings, 1996). I denne fasen er virksomheten avhengig av menneskenes kunnskapskapital, som kan anvendes til å omdanne datapunkter til innsikt, som videre kan bidra til å løse et forretningsproblem. I den neste fasen spiller organisasjonen en viktigere rolle i å implementere løsningen menneskene har utviklet Oldham og Cummings (1996). For å lykkes med kreativt arbeid er derfor rekrutteringsprosesser og utvelgelse av kunnskapsrike ansatte svært viktig, og kan være nøkkelen for a skape de nødvendige forutsetningene for innovasjon (Mumford, 2000).

Å entydig definere kreativitet kan være krevende som følge av begrepets vide forståelse, og dets naturlige tilknytning til oppfatninger av for eksempel fantasi, nytenking, innovasjon og talent. I mangel på en universelt anerkjent definisjon er det hensiktsmessig å se nærmere på begrepets ulike komponenter (Kaufman og Sternberg, 2010). De amerikanske psykologene peker for det første på viktigheten av *novelty*, altså det originale og nyskapende ved ideen. Videre er kvalitetsdimensjonen en viktig faktor som vil bidra til ideens underliggende kreative verdi. Den siste komponenten handler om ideens relevans i forhold til problemet den søker å løse. For å kunne forstå hva kreativitet er, peker Kaufman og Sternberg (2010) på komponenter av kreative ideer som en kobling til forståelsen av det brede kreativitetsbegrepet. Kaufman og Sternberg (2010) oppsummerer dermed selv at en kreativt idé må være ny, god og relevant.

I litteraturen representerer Amabile (1988) en lignende forståelse, som også omtales som det tradisjonelle synet på kreativitet definert som produksjon av ideer eller løsninger som er nye og nyttige». Flere studier fremhever betraktningene til både Kaufman og Sternberg (2010) og Amabile (1988) ved å definere kreative ideer som skjæringspunktet mellom særlig to av komponentene, relevans og originalitet. I forskningen henviser relevans-komponenten til hvor godt den kreative ideen svarer til problemstillingen den søker å løse. Den kreative ideens nytteverdi og effektivitet trekkes frem som underkategorier av relevans-komponenten (Gardner, 1993). Nyhets-komponenten til ideen inkluderer også underkategoriene originalitet og hvor unik den er. Følgelig vil de to overnevnte

komponentene også være særlig viktige i å evaluere en kreativ idé (Moreau og Dahl, 2005).

I tillegg til den teoretiske tilnærmingen til kreativitetsbegrepet som omhandler ulike komponenter og underkategorier, refereres det også i litteraturen til kreativitet som problemløsing som en praktiske applikasjon (Isaksen et al., 2010). **Problemet** i denne konteksten defineres som *«en sak eller situasjon som må løses eller overvinnes ved å identifisere eller oppfinne en ny løsning»* (Joyce, 2009). To mulige fremgangsmåter på problemløsning eksemplifiseres av Moreau og Dahl (2005) ved en hverdagslig aktivitet som å tilberede et måltid; personen kan enten velge å basere måltidet på en kjent oppskrift og gå til innkjøp av nødvendige ingredienser i forveien, eller utforme måltidet basert på tilgjengelige ingredienser, uten å følge en eksisterende oppskrift. Uavhengig om personen velger å følge en gitt oppskrift eller lage en rett basert på tilgjengelige ingredienser vil den kreative prosessen aktiveres, dog på ulikt vis (Joyce, 2009).

2.4.2 Utdanning og erfaring

Et betimelig spørsmål handler om hvorvidt kreativitet kan knyttes til spesifikke fagområder eller erfaringsbakgrunner, og dermed om det finnes en sammenheng mellom personers bakgrunn og evne til å innovere. Sternberg (2003) viser til at analytisk, kreativ og praktisk intelligens er fraskilt fra hverandre, og at man dermed ikke kan si noe om en persons kreative intelligens basert på hvor analytisk eller praktisk anlagt vedkommende er. I forskningen er det heller ikke avklart om kreativitet kan betraktes som en generell ferdighet som kan læres og anvendes på tvers av disipliner. Det er dermed vanskelig å stadfeste en direkte sammenheng mellom utdanningsnivå og kreative evner på individnivå. Salvanes (2014) peker samtidig på utdanning som den sentrale faktoren bak utviklingen av en moderne økonomi som bygger på høy grad av innovasjon, og støtter den rimelige antagelsen om at utdanningsbakgrunn og arbeidserfaring spiller en rolle for innovasjon på individnivå.

Et viktig diskusjonstema i masterutredningen handler om hvilke type humankapital som kan påvirke virksomheters innovasjonsevne, som bygger på antagelsen om at menneskene og deres egenskaper er den essensielle driveren for selskapers innovasjonsaktiviteter. En studie utført av Al-Laham et al. (2011) ser på konsekvensene av blant annet rekruttering og samarbeid for innovasjon, og argumenterer for at nyansatte spiller en viktig rolle i å

føre virksomheter i en mer innovativ retning. Studien peker på at at nyansatte bidrar med oppdatert kunnskap og dermed nye perspektiver som kan støtte virksomheten i utvikling av verdiforslaget. Studien introduserer begrepet "*kunnskapsgrenser*"i sammenheng med selskapets humankapital, og argumenterer for at nyansatte kan bidra til å utforske mulighetene som ligger utenfor virksomheters eksisterende grenser, øke kunnskapskapitalen og dermed stimulere virksomhetens innovasjonsevne (Al-Laham et al., 2011).

I en annen studie viser Knudsen og Lien (2019) til at hvordan bedrifter tar vare på sin kunnskapskapital i nedgangstider vil ha en innvirkning på lang sikt, og trekker nok en sammenheng mellom humankapital og virksomheters innovasjonsevne. Studien ble gjort i etterkant av finanskrisen i 2008, og blant over 1200 bedrifter viste resultatene klare tegn på at bedrifter som prioriterte innovasjon i forretningsstrategien responderte langt bedre under resesjon og krisetider (Knudsen og Lien, 2019).

Det er ikke omstridt at utdanning og kompetanse kan spille en rolle for individers innovasjonsevne. Kompetansen må i neste omgang anvendes på riktig sted i organisasjonen, sammen med de riktige kollegaene og med gode betingelser for at kunnskapskapitalen skal vokse og utvikles. I denne sammenheng er samarbeid og kunnskapsdeling viktige komponenter som sammen kan motvirke at det oppstår siloerrundt avdelinger og menneskelige ressurser, også kalt *silotekning* (Gleeson, 2013).

Mye tyder på at innovasjonsprosesser krever en holistisk tilnærming, hvor organisasjoner må legge til rette for kunnskapsdeling og tverrfaglig samarbeid. Arbeid med dataintensive problemstillinger er intet unntak, og stiller særskilte krav til hvordan organisasjonen anvender sin analytiske verktøykasse, se figur 2.6. Blant virksomhetens ulike analytiske ressurser spiller mennesket en sentral rolle, og som Oldham og Cummings (1996) påpeker starter idégenereringen hos den individuelle medarbeideren, før ideene implementeres av en samlet organisasjon.

3 Metode

Forskningsmetode defineres som de fremgangsmåter eller teknikker som benyttes for å belyse vitenskapelige spørsmål og problemstillinger (Ringdal, 2007). Vi vil i denne delen redegjøre for de metodiske valgene foretatt i undersøkelsen. Hvilke problemstillinger og forskningsspørsmål som er valgt vil derfor ha stor påvirkning på valg av riktig metode (Jacobsen, 2005). Det benyttes kvantitativ metode i denne oppgaven med business case og spørreundersøkelse som datainnsamlingsmetode. Det vil bli redegjort for kvantitativ datainnsamlingsmetode, design og strategi samt utforming av business case og spørreundersøkelse. I tillegg omtales populasjon og utvalg, validitet og reliabilitet og etiske aspekter knyttet til forskningen.

3.1 Forskningstilnærming

Forskningstilnærming blir beskrevet som samspillet mellom metode, data, teori og forskere sin verdi i et gitt studie (Ghauri et al., 2010). Det skilles mellom to ulike forskningstilnærminger, induktiv og deduktiv forskningstilnærming. Induktiv tilnærming blir benyttet for å utforske eksisterende teori videre. Deduktiv tilnærming benyttes for å avkrefte eller bekrefte forskningshypoteser basert på eksisterende teori (Ghauri et al., 2010). Tilnærmingene er ikke utelukkende fra hverandre og vil ofte overlappe. I denne studien benytter vi en overlappende variant av de to nevnte tilnærmingene, som baserer seg på en kombinasjon av kreativitet og kontroll. Først presenteres en overordnet problemstilling, før vi stiller et utdypende spørsmål som konkretiserer problemstillingen. Videre samles data inn for å svare på den aktuelle problemstillingen.

Det skilles mellom kvantitativ og kvalitativ metode. Metodene forklarer hvordan data innhentes og hvilken type data som blir innhentet. Kvantitativ data uttrykkes gjennom tall og mengdeenheter og er kvantifiserbart (Gripsrud et al., 2004). Kvalitativ forskning danner dyp forståelse i form av kontekstuell og ikke tallbasert data ved å møte respondentene i en uformell setting (Ponelis, 2015). I dette studiet ønsker vi å belyse problemstillingen ved å undersøke hvordan humankapital påvirker dataintensiv innovasjon og idegenerering. I den sammenheng ønsker vi å benytte kvantitativ metode for å få et helhetlig bilde på problemstillingen. Bruk av kvantitativ metode kan snevre inn fokuset fra det generelle og

store temaet til det mer spesifikke spørsmålet som skal besvares (Gall et al., 2007).

3.2 Forskningsdesign

For å besvare forskningsspørsmålet gjennom datainnsamling, tolkning og analyse er det viktig å velge riktig forskningsdesign (Saunders et al., 2016). Et godt forskningsdesign er vesentlig for å besvare forskningsspørsmålet på en komplett måte innenfor begrensningene som eksisterer i forskningen og for å redusere risiko for feilaktige svar på forskningsspørsmålet (Saunders et al., 2016).

Forskningsdesign blir i all hovedsak delt inn i eksplorerende, deskriptivt og forklarende design (Saunders et al., 2016). Eksplorerende design kan være hensiktsmessig når det som studeres er ustrukturert, komplekst og relativt nytt. Det motsatte hvor temaet er strukturert, mindre komplekst og man ønsker å beskrive karakteristika eller korrelasjoner kan deskriptivt design være hensiktsmessig. Til slutt ønsker forklarende design å beskrive kausale sammenhenger tilknyttet forskningsspørsmålet.

I dette studiet ønsker vi å få en forståelse for hvordan ulik humankapital påvirker evnen til DDI, hvordan ulike mennesker med ulik bakgrunn og erfaring vektlegger ulike attributter i viktigheten av DDI og hvordan individer er gode eller mindre gode på å generere gode ideer. I den sammenheng vil et eksplorerende forskningsdesign bli benyttet. Det gir og en større grad av fleksibilitet i forskningen ved at vi får muligheten til å utforske data som er ustrukturert og komplekst. Vi kan i større grad stille åpne spørsmål samtidig som vi kan snevre inn forskningen. Det gjør at vi kan kartlegge forskningen og innhente relevant data for å besvare forskningsspørsmålet på en god måte (Saunders et al., 2016). Eksplorativt design gir muligheten til å endre retning dersom forskningen etter hvert skulle trekke i en ny retning (Ghauri et al., 2010).

3.3 Forskningsstrategi

Hvordan man besvarer et forskningsspørsmål kommer an på hvilken forskningsstrategi som benyttes. I et eksplorativt design er case og spørreundersøkelser ulike eksempler på forskningsstrategier som ofte benyttes (Saunders et al., 2016). Denne oppgaven er designet med et datadrevet business case etterfulgt av en spørreundersøkelse. Begge delene henger

sammen i en felles Qualtrics undersøkelse.

3.3.1 Business case

For å kartlegge idegenereringen til ulike individer og samtidig få dypere forståelse av forskningsspørsmålet designer vi ett business case. Studier basert på case kan være fordelaktig for å unngå generalisering i en populasjon, og bidrar samtidig til å utikle eksisterende teori. Derfor er casestudier bedre egnet til å skaffedypere forståelse av tematikken problemstillingen omhandler (Thagaard, 2009). Bruk av case i studier gjør at forskere har fått anledning til å undersøke problemstillingen i dybden, samtidig som caset simulerer en naturlig kontekst for respondenten (Yin, 2014). Grunnet masterutredningens begrensninger med hensyn til tid og ressursbruk, vil vi ikke benytte oss av en fullverdig casestudie. Strategien handler i større grad om å presentere et case til alle respondentene i undersøkelsen, som gir oss muligheten til å simulere starten på en innovasjonsprosess.

3.3.2 Spørreundersøkelse

For å se på hvilke attributter i humankapitalen som påvirker til innovasjon var det naturlig å ta med en spørreundersøkelse i studiet. Definering av problemstilling og forskningsspørsmål var første del av spørreundersøkelsen, videre vil det være viktig å avklare konkrete mål for forskningen og snevre inn fokuset fra det generelle til det spesifikke som ønskes undersøkt (Gall et al., 2007). Spørreundersøkelser inneholder mange respondenter og få variabler. Videre ønsker analysen av undersøkelsen å beskrive karakteristikker, attributter og sammenhenger i datamaterialet, samt til en viss grad å se på årsaker til ulike fenomener (De Vaus, 2002).

Det er flere ulemper ved spørreundersøkelser, for det første gir en spørreundersøkelse færre muligheter til å se kausale sammenhenger og det finnes få muligheter til å få oppklarende eller utfyllende informasjon utover svarene fra undersøkelsen (De Vaus, 2002). Videre omtales dette som en tverrundersøkelse, som vil si at strategien er avgrenset av tid og resultatene kun speiler kunnskap og erfaringer på gitt tidspunkt (Johannessen et al., 2010).

3.4 Innsamling av data

Datainnsamlingen ble gjennomført via en Qualtrics undersøkelse som er et vanlig datainnsamlingsverktøy ved NHH. Denne er utarbeidet i to deler hvor del den første delen består av et business caset og den andre delen er en spørreundersøkelse.

For innsamling av primærdata til kvantitativ metode hevder Jacobsen (2005) spørreskjema med lukkede svaralternativer dominerer. Spørreundersøkelser er en strukturert metode for datainnsamling med svaralternativer og faste spørsmål. Det gjør at man kan se likheter og variasjoner i dataene ved analyse. En slik standardisering kan gi mulighet for generalisering og gjør det enklere å få flere respondenter med kortere tidshorisont. Til slutt kan dataene som er innsamlet analyseres statistisk (Johannessen et al., 2010).

Jacobsen (2005) presenterer tre hovedelementer i planleggingsfasen av en spørreundersøkelse; begrepene som ønskes målt må konkretiseres (operasjonalisering), spørsmålene må utformes slik at de måler det som ønskes målet og hvordan undersøkelsen skal distribueres må avklares.

3.4.1 Utvalg

Alle respondenter fra hele populasjonen som er innenfor de vi ønsker å forske på kalles den teoretiske populasjonen. Denne populasjonen er så stor at vi er nødt til gjøre et utvalg (Jacobsen, 2005). Utvalget ønsker å representere et utsnitt av populasjonen som ønskes undersøkt (Befring, 2007). Teoretisk populasjon i dette studiet er individer med erfaring, kompetanse eller utdanning fra ulike nivå i disse kategoriene; teknologibakgrunn, innovasjonsbakgrunn, forretningsbakgrunn eller ingen relevant bakgrunn.

Prosessen ved å gjøre et utvalg vil være en viktig del av hvor godt generalisering kan gjøres mot populasjonen. Til flere relevante kjennetegn utvalget har mot populasjonen vil si noe om hvor representativt utvalget er. Videre er det et grunnleggende krav for generalisering at utvalget er av en viss størrelse (Jacobsen, 2005). Det er ønskelig med et randomisert utvalg men det er ofte enklere i teorien en i praksis. Med tanke på tid og ressurser så vi fort at vi måtte benytte nettverket rundt oss for å få nok respondenter. Det innebar at hoveddelen av respondentene er fra noen skoler og næringslivsaktører fra Bergen. I litteraturen blir en slik tilnærming ofte omtalt som formålstjenlig utvalg (Befring, 2007).

Hovedutfordringene med en slik tilnærming er at det stor fare for et systematisk skjevt fordelt utvalg og muligheten for generalisering svekkes (Jacobsen, 2005).

3.4.2 Struktur og utforming

Som en del av oppgaven for å måle graden av innovasjon er idegenerering et verktøy som vil bli benyttet. Undersøkelsen er strukturert slik at respondentene vil svare på et business case som representerer datadrevet problemløsing. Respondentene vil notere ned alle ideene de kommer på for å løse problemstillingene i caseoppgaven. Videre vil ideene bli rangert ved ulike suksessfaktorer av et ekspertpanel fra Knowit. Scoren av idegenereringen innenfor de ulike suksessfaktorene, samt total vektet score vil utgjøre de uavhengige variablene som benyttes i videre analyse.

Utformingen i spørreundersøkelsen tar utgangspunkt i problemstillingen og forskningsspørsmålet vi ønsker å besvare. Spørsmålene er laget med hensyn på å være så konkrete som mulig slik at de er lette å besvare, men som samtidig gir den informasjonen som er ønsket. Dette hjelper oss med fortolkningen i analysedelen (Johannessen et al., 2010). Alle andre spørsmål en selve idegenereringen er lukkede spørsmål. Det er sentralt at disse operasjonaliseres slik at vi unngår å være vage eller upresise. Vi har benyttet De Vaus (2002) sin sjekkliste for utforming av spørsmålene. Det vil si at vi har hatt et fokus på å være presise i formuleringene og prøvd å ikke stille ledende spørsmål. For å best mulig gi respondentene mulighet til å svare det som ligger de mest nærliggende benyttes en syvdelt ordinal skala. Rangeringen går fra i svært liten grad til i svært stor grad. Slike spørsmål ble for eksempel brukt for å kartlegge psykologiske karaktertrekk i form av "i hvilken grad kjenner du deg igjen i denne situasjonen eller lignende". Det er omdiskutert om midtkategorien i en slik ordinal skala er fordelaktig. De Vaus (2002) argumenterer for å inkludere denne kategorien ved at man unngår å tvinge respondenter til å svare noe de egentlig ikke har noe særlig mening om. Gjensidige utelukkende kategorier ble også benyttet, særlig i de demografiske variablene.

Alle spørsmål i undersøkelsen er nøye utformet for å svare best mulig på problemstillingen. Videre er spørreundersøkelsen delt inn i 2 deler. Del 1 er selve business caset og idegenereringen som blir gitt en score fra ekspertpanelet. Det er disse scorene som blir de avhengige variablene vi ønsker å analysere. Del 2 inneholder bakgrunnsvariabler som

demografi, utdanning og kompetanse og psykologiske variablene som prøver å kartlegge respondentenes evner og kreativitet.

Spørreundersøkelsen er nøye vurdert av samarbeidsbedrift, veileder og testet på individer fra utvalget av populasjonen. Dette er gjort for å sjekke at spørsmålene er forståelig og konkrete nok til formålet de er laget for. Tidligere forskning og andre masteroppgaver er også brukt som inspirasjon for å utforme spørsmålene på en god måte. I et helhetlig bilde må undersøkelsen være selvinstruerende samtidig som spørsmål er relevante, entydig og formulert på en god måte (Johannessen et al., 2010).

3.4.3 Business case og ekspertpanel

Caset i undersøkelsen er benyttet for å kartlegge idegenerering fra de ulike respondentene. Caset er utviklet basert på egne erfaringer, teori og samtaler med et større forsikringsselskap fra Norge. Ulike typer kategorier med data blir presentert for respondentene før de blir bedt om å komme med nye og innovative forsikrings ideer basert på disse dataene eller plausibel data som de kan tenke seg til. Vi ber respondentene om å skrive ideene på en måte som er mulig for andre å forstå. De kan skrive inn så mange ideer de ønsker og vi sier at cirka 20 minutter anses som tilstrekkelig betenkningstid. Hele caset ligger i appendiks A1.

Ekspertpanelet som skal bedømme de ulike ideene består av fire eksperter fra Knowit. De fire personene har ulike erfaringer og bakgrunner og jobber i alt fra toppledelsen til andre stillinger nedover i organisasjonen. Hver ekspert får utdelt alle ideene hver for seg slik at de ikke skal bli påvirket av hverandre. De svarer i hvert sitt Excel-ark før vi til slutt tar gjennomsnittsverdien av alle svarene som de endelige variablene. Panelet bedømmer de ulike ideene basert på disse fire dimensjonene som representerer de avhengig variablene som benyttes for videre analyse;

1. **Kreativitet** - Vi ønsker å avdekke hvor nytenkende ideene oppfattes, isolert fra om hvorvidt ideen er praktisk gjennomførbar eller realistisk.
2. **Gjennomførbarhet** - Hvor realistisk og gjennomførbar ideen er, men sett bort ifra personvernproblematikk og GDPR.
3. **Kommersialisering perspektiv** - Hvor stort potensiale har ideene for å kunne

kommersialiseres.

4. **Dataintensivitet** - I hvilken grad ideene er basert på data.

Alle ideene blir rangert i en syvdelt ordinal skala som går fra *svært liten grad* til *svært stor grad*. I tillegg til de fire dimensjonene vil vi lage en total score basert på gjennomsnittet fra alle dimensjonene samlet. Det er analyse av disse dimensjonene samt den totale scoren som vil bli benyttet for svare på problemstilling og forskningsspørsmål presentert i introduksjonen.

De tre første dimensjonene er valgt basert på eksisterende teori, og knytter seg til kreativitetsbegrepet, gjennomgått i teorikapittelet. I tillegg har vi valgt å legge til dataintensivitet som en fjerde dimensjon, i lys av caseoppgavens tydelige instruksjon om databruk og oppgavens gjennomgående tematikk og problemstilling.

3.4.4 Spørreundersøkelse

Demografiske variabler som alder, kjønn, utdanning og erfaring blir først presentert i spørreundersøkelsen som etterfølger caset respondentene besvarer. Disse variablene vil benyttes til å kategorisere respondentene innen følgende fire kategorier;

1. **Teknologibakgrunn** - Utdanning og/eller arbeidserfaring innen tekniske fag.
2. **Innovasjonsbakgrunn** - Utdanning og/eller arbeidserfaring innenfor innovasjon.
3. **Forretningsbakgrunn** - Utdanning og/eller arbeidserfaring innenfor forretning.
4. **Ingen relevant bakgrunn** - Ingen utdanning eller arbeidserfaring innenfor tekniske fag, innovasjon eller forretning.

Basert på svarene i første del av spørreundersøkelsen vil vi manuelt kategorisere de respektive respondentene i hver av de fire bakgrunnene. For å knytte case og de ulike dimensjonene til de fire bakgrunnene, vil vi svare på forskningsspørsmålet presentert i innledningen. Videre vil demografiske variabler bli benyttet til videre analyse som del av andre funn presentert i analysen.

Psykologiske variabler vil kartlegge respondentenes kreativitet, grad av kommersialisering og grad av gjennomføringsevne. Vi presenterer ulike påstander hvor respondentene svarer på en syvdelt ordinal skala som går fra svært uenig til

svært enig. Disse variablene kartlegger i hvilken grad respondentene tenker på disse perspektivene når de jobber med ulike ideer og i hvilken grad de er viktige for å oppnå suksess i deres øyne. Vi ønsker her å se på deres subjektive meninger for å kunne knytte variablene opp mot dimensjonene og bakgrunnene. Vi vil basert på de psykologiske variablene gjennomføre en faktoranalysefor å undersøke sammenhenger mellom ulike variabler og for å enklere gjennomføre effektive analyser og regresjoner.

3.5 Analyse av data

Den kvantitative analysen blir utført i RStudio gjennom programmeringsspråket R. RStudio gir oss en stor verktøykasse til å gjøre de relevante analysene, regresjoner, faktoranalyse og visualisering av resultater.

Analysen vil inneholde forskjellige statistiske metoder. Vi begynner med å se på deskriptiv statistikk fra respondentene. Dette er data som antall respondenter, kjønn, alder osv. Videre vil vi gjennomføre en større faktoranalyse hvor vi ønsker å identifisere ulike variabler som kan samles i faktorer. Den mest brukte metoden som benyttes er lineær regresjon. Vi vil kjøre både enkle og multiple regresjoner i analysedelen. Til slutt benytter vi *cross validation* metoder for å optimalisere en modell som kan predikere de avhengige variablene.

3.6 Evaluering av metode

Validitet og reliabilitet er viktige element å diskutere for kvaliteten av forskningen (Johannessen et al., 2010). Validitet handler om i hvilken grad det som undersøkes måler det den er designet for å måle og hvorvidt det er relevant for forskningen. Et grunnleggende mål for kvantitativ forskningen er generalisering fra populasjon til utvalg og validiteten vil si noe om de vi måler i undersøkelsen vil gjelde for en større gruppe (De Vaus, 2002). Jacobsen (2005) deler validitet inn i tre hovedkomponenter; indre validitet, ytre validitet og begrepsvaliditet. Reliabilitet relaterer seg til data som inngår i studien, og handler om hvilken data som er brukt, hvordan de er samlet inn og hvordan de er analysert og benyttet. Reliabilitet sier noe om hvor pålitelig og troverdig forskningen er (Johannessen et al., 2010).

3.6.1 Indre validitet

Sammenhengen mellom variabler og årsaken til disse sammenhengene er omtalt som intern validitet (Saunders et al., 2016). Det handler altså om i hvilken grad det vi tror vi måler, faktisk er det vi måler (Johannessen et al., 2010), og om riktig data er innsamlet fra de riktige kildene (Oates, 2005). En måte å vurdere validitet på er "face validity"-vurdering, som handler om hvor tillitsvekkende forskningen er. Vi har nøye vurdert utvalget fra populasjonen og i stor grad prøvd å få representative respondenter.

Neste metode vil være å eliminere tolkninger som ikke nødvendigvis er sannsynlige eller forenlige med innsamlet data. Det kan øke tilliten og man kan argumentere for en sannsynlig konklusjon (Kleven, 2002). Det er flere mulige årsaker til at individer er gode på DDI, som for eksempel riktig bakgrunn, høy grad av kreativitet eller høye ambisjoner. Selv om det ikke alltid er mulig å finne bevis for kausale sammenhenger, kan en god beskrivelse av resultatene bygge tillit og støtte validiteten.

Siden de fleste respondentene er fra Bergen og i stor grad fra våre egne nettverk, kan dette bidra til å svekke validiteten. Det vil da ikke blitt tatt hensyn til forskjeller i andre deler av landet. For å motvirke dette har vi benyttet vårt nettverk til å få en større bredde av respondenter, samt jobbet aktivt for å finne respondenter som skiller seg fra hverandre. Dette studiet er i stor grad kvantitativt, og vi går ikke like kvalitativt i dybden som en fullverdig casestudie ville gjort. Dette kan dermed bidra til å s svekke validiteten til studien.

3.6.2 Ytre validitet

Den ytre validiteten omhandler i hvilken grad resultatene er overførbare eller gir mulighet for generalisering (Gripsrud et al., 2004). Ved at vi bare utfører undersøkelsen én gang kan det være vanskelig å tilfredsstille den ytre validiteten (Lund, 2002). Det vil altså være vanskelig å si om vi hadde fått de samme resultatene om vi hadde gjort samme undersøkelse en gang til med et annet utvalg. Likevel er målet med en utredning av denne typen å gi et grunnlag for videre forskning for utvikling av en generaliserbar teori. Vi kan anta at vi ville funnet lignende svar i andre utvalg og sett likhetstrekk fra spørreundersøkelsen selv om den ble testen på nytt en annen plass.

3.6.3 Reliabilitet

Reliabilitet kan forklares som i hvilken grad vi kan stole på om resultatene er pålitelige. I hvilken grad tilfeldige, og derfor irrelevante forhold påvirker resultatene (Askheim og Grenness, 2008). Videre omhandler reliabilitet om nøyaktigheten av undersøkelsens data, hvordan den er samlet inn og hvordan den er analysert. Et element som kan påvirke reliabiliteten er uklare formuleringer som gir rom for tolkning, som kan medføre forskjellige svar fra gang til gang (Johannessen et al., 2010).

God reliabilitet ved en spørreundersøkelse handler i stor grad om respondentene hadde svart det samme om de hadde tatt undersøkelsen en gang til. Det er dermed veldig viktig at spørsmålene er god formulert og at det er brukt riktig ordvalg nettopp fordi respondenter svarer ulikt fra gang til gang (De Vaus, 2002). Vi kan anta at det er lettere å formulere riktige spørsmål helt frem til caset i spørreundersøkelsen. Det vil i en idegenerering fase være vanskelig å få de samme ideene ved en ny gjennomføring av undersøkelsen. Dermed vil det være kvaliteten av disse ideene som vil si noe om reliabiliteten over tid. At respondenter med en gitt bakgrunn eller psykologisk profil vil levere samme kvalitet og legge vekt på de samme egenskapene som tidligere.

3.7 Etiske aspekter

I dette studiet har vi fulgt retningslinjene til NSD (Norsk Samfunnsvitenskapelige Datatjeneste) og NHH retningslinjer for utarbeidelse av masteroppgave. Gjennom samtale med veileder og henvendelse til NSD ble det avklart at denne masteroppgaven ikke trenger forhåndsgodkjennelse. Det er basert på at spørreundersøkelsen ikke inneholder sensitiv informasjon og at IP adressene til respondentene ikke blir lagret.

Gjennom hele prosessen har vi hatt fokus på forskningsetikk og vitenskapelig redelighet. Det handler i stor grad om et sett verdier og normer, juks, forfalskning av data og plagiat. I spørreundersøkelsen har vi et særlig ansvar for å ivareta de etiske aspektene. Vi har i den sammenheng hatt et fokus på frivillig deltakelse, informert samtykke, konfidensialitet og respekt for respondentenes privatliv (De Vaus, 2002).

Rapportering og analyse av resultatene inngår også i det forskningsetiske hensynet. Det legges vekt på at resultatene må presenteres på en god måte, slik at risiko for mistolkning

reduseres. Dette gjenspeiles i en grundig vurdering av riktig metodevalg for å kunne evaluere resultatene (De Vaus, 2002).

4 Analyse

I dette kapittelet presenteres resultatene fra caseoppgaven og spørreundersøkelse. Vi anser det som hensiktsmessig å presentere datasettet med løpende kommentarer, før vi gjenomgår caset respondentene har besvart og til slutt resultatene fra ekspertgruppen. Videre ser vi på de demografiske variablene hvor vi til slutt har kategorisert alle respondenter innenfor våre fire bakgrunner; *innovasjonsbakgrunn*, *teknisk bakgrunn*, *forretningsbakgrunn* og *ingen relevant bakgrunn*. Analyseplattformen Qualtrics ble benyttet til utarbeidelse og gjennomføring av case og spørreundersøkelse.

For dataanalyse benyttes Rstudio med programmeringsspråket R, og det er tre hovedgrunner til at vi har valgt denne programvaren; **(1)** R er et kraftfullt programmeringsspråk som gir muligheten til å gjennomføre et bredt spekter av analyser. **(2)** R lar brukere sende inn sine egne pakker til R-serveren som alle kan bruke. Dette betyr at man har tilgang til et stort utvalg av pakker, som kan gjøre alt fra tunge statistiske analyser til tekstutvinning. **(3)** R er et gratis åpen-kildekode programmeringsspråk, tilgjengelig for de fleste operativsystemer, hvor syntaks og pakkene kan transporteres fra ett system til et annet. Dette kan stimulere til tverrfaglig forskningsamarbeid, samtidig som det legger til rette for at egen forskning er mulig å replikere senere (Venables et al., 2009)

Vi vil ved hjelp av R gjennomføre en faktoranalyse for de psykologiske variabelene for å undersøke om det finnes sammenhenger som kan representeres i ulike faktorer. Vi vil deretter gjennomføre statistiske analyser for å undersøke de ulike bakgrunnere og belyse andre interessante funn. Resultatene vil presenteres og visualiseres ved hjelp av tabeller, tekst, diagrammer og utklipp fra Rstudio.

For ordens skyld gjentas masterutredningens problemstilling og forskningsspørsmål:

Er datadrevet innovasjon en oppgave for økonomen eller teknologen?

Hvilke demografiske trekk og hvilken type humankapital påvirker individers evne til å generere gode ideeer?

4.1 Datasett

Datasettet består av totalt 38 variabler. Variablene *Q1-Q28* er hentet direkte fra case og spørreundersøkelse, og henviser til spørsmål 1 til spørsmål 28. Variabelene er listet i samme rekkefølge som Qualtrics-undersøkelsen og kan finnes i appendiks A1. Variablene *N1-N7* og *F1-F3* er generert i Excel og Rstudio med utgangspunkt i case og spørreundersøkelsen. Disse variablene refereres til som *nye variabler* og blir benyttet for faktor- og regresjonsanalyse. For å vise hvordan alle variablene henger sammen er de sammenstilt i tabell 4.1. Vi vil forløpende presentere variablene etterhvert som de fremkommer i de neste avsnittene.

Variabel	Beskrivelse	Type
Q1 - Case	Respondentenes ideer	Case
Q2 - Tidsbruk	Antall minutter benyttet på case	Case
Q3 - Kjønn	Kjønn	Spørreundersøkelse
Q4 - Alder	Alder	Spørreundersøkelse
Q5 - Arbeidsstatus	Arbeidsstatus	Spørreundersøkelse
Q6 - Uretning	Utdannelsesretning	Spørreundersøkelse
Q7 - Utillegg	Om respondenten har tilleggsutdanning	Spørreundersøkelse
Q8 - Utilleggr	Hvis tilleggsutdanning hvilken retning	Spørreundersøkelse
Q9 - Utilleggrt	Tilleggsutdanning fritekst	Spørreundersøkelse
Q10 - Univå	Utdanningsnivå	Spørreundersøkelse
Q11 - Årslønn	Årslønn	Spørreundersøkelse
Q12 - Bransje	Bransje for arbeidstakere	Spørreundersøkelse
Q13 - Bransjeled	Bransje for arbeidsledig	Spørreundersøkelse
Q14 - EI	Arbeidserfaring Innovasjon	Spørreundersøkelse
Q15 - ET	Arbeidserfaring Teknologi	Spørreundersøkelse
Q16 - EF	Arbeidserfaring Forretningsutvikling	Spørreundersøkelse
Q17 - PKre1	Kreativitet spørsmål 1	Spørreundersøkelse
Q18 - PKre2	Kreativitet spørsmål 2	Spørreundersøkelse
Q19 - PKre3	Kreativitet spørsmål 3	Spørreundersøkelse
Q20 - PKom1	Kommersialisering spørsmål 1	Spørreundersøkelse
Q21 - PKom2	Kommersialisering spørsmål 2	Spørreundersøkelse

Q22 - PKom3	Kommersialisering spørsmål 3	Spørreundersøkelse
Q23 - PGje1	Gjennomførbart spørsmål 1	Spørreundersøkelse
Q24 - PGje2	Gjennomførbart spørsmål 2	Spørreundersøkelse
Q25 - PGje3	Gjennomførbart spørsmål 3	Spørreundersøkelse
Q26 - Vkre	Veklegging av kreativitet	Spørreundersøkelse
Q27 - Vgje	Veklegging av gjennomførbarhet	Spørreundersøkelse
Q28 - Vkom	Veklegging av Kommersialisering	Spørreundersøkelse
N1 - Ide	Antall ideer - Case	Ny variabel
N2 - Kre	Kreativitet score - Case	Ny variabel
N3 - Gje	Gjennomførbarhet score - Case	Ny variabel
N4 - Kom	Kommersialisering score - Case	Ny variabel
N5 - Dat	Datainntensivitet score - Case	Ny variabel
N6 - Tot	Totalscore - Case	Ny variabel
N7 - Bak	Type bakgrunn	Ny variabel
F1	Faktor 1 - Kreativitet	Ny variabel
F2	Faktor 2 - Kommersialisering	Ny variabel
F3	Faktor 3 - Gjennomførbarhet	Ny variabel

Tabell 4.1: Datasett

4.2 Case

Vi starter analysen med å se på resultatene fra casebesvarelsene - *Q1*. Undersøkelsen ble besvart av totalt 92 respondenter og antall ideer varierer fra 1 - 5 per respondent. Noen av svarene er utfyllende besvart og andre består av stikkord og mindre tekst. En oversikt over alle besvarelsene finnes i appendiks A2. I denne delen vil vi se på hvor lang tid respondentene benyttet i gjennomsnitt, hvor mange ideer hver respondent i snitt hadde og til slutt resultatet fra rangeringen ekspertpanelet har gjennomført.

4.2.1 Tidsbruk og antall ideer

Fra Qualtrics-rapporten i appendiks A1 kan vi se at respondentene benyttet i snitt 14,47 minutter per besvarelse, med et standardavvik på 9,84. Vi anser gjennomsnittstiden

benyttet som relativt bra, særlig i lys av de generelle utfordringene som knytter seg til å få respondenter til å besvare tidkrevende caseoppgaver og undersøkelser. Videre ser vi at standardavviket er høyt i forhold til gjennomsnittet, hvilket indikerer at spredningen blant respondentene er stor. Dette innebærer at flere av respondentene har brukt mer enn 14,47 minutter, og mange har brukt langt mindre tid enn gjennomsnittet.

12 respondenter har i casebesvarelsen ikke kommet med konkrete ideer. Dette kan være forårsaket av manglende kunnskap, anledning eller vilje til å gjennomføre caseoppgaven. Disse har skrevet at de ikke kommer på noen ideer og dermed er antall ideer for de respondentene satt til null. Videre ser vi at gjennomsnittlig antall ideer per respondent, inkludert de 12 som er satt til null er 1,55 ideer. Ekskluderer vi disse 12 respondentene fra datasettet ser vi at gjennomsnittet øker til 1,8 ideer per besvarelse. Vi ser at flertallet i utvalget har beskrevet én enkelt ide, hvilket fører til at snittet havner under 2 per respondent. Vi er likevel fornøyd med dette resultatet og vil under andre funn presentere mulige sammenhenger mellom mellom *antall ideer* og andre variabler. Variabelen antall ideer er derfor lagt til i datasettet og heretter henvisst til som *N1 - Ide* (antall ideer).

4.2.2 Resultat fra ekspertpanel

Som forklart i metodekapittelet består ekspertpanelet av tre eksperter fra Knowit med ulike bakgrunner, alder og kjønn. De har rangert alle ideene langs de fire dimensjonene; **(1)** kreativitet, **(2)** gjennomførbarhet, **(3)** kommersialisering og **(4)** datainntensivitet. Dimensjonene er rangert på en likert-skala som går fra i svært liten grad (1) til i svært stor grad (7), og ekspertene har blitt bedt om å gi hver respondent fire individuelle. Datasettet og resultatene er videre bearbeidet i Excel og oversatt fra tekst til den numeriske skalaen (1-7). Dette gjør vi for å enklere kunne behandle kvantifiserbare resultater i Rstudio. I tabell 4.2 presenteres gjennomsnittscoren de tre ekspertene har gitt innen hver dimensjon og gjennomsnittlig totalscore samlet.

Kreativitet	Gjennomførbarhet	Kommersialisering	Datainntensivitet	Total score
3,48	3,57	3,40	4,07	3,63

Tabell 4.2: Gjennomsnitt rangering av ideer

Vi ser at dimensjonene *kreativitet*, *gjennomførbarhet* og *kommersialisering* alle ligger under 4,0, hvilket innebærer at ekspertpanelet har vurdert ideene til litt under middels.

Datainntensivitet er dimensjon som er rangert høyest med et gjennomsnitt på 4,07. Dette resultatet er forventet i lys av hvordan caseoppgavens utforming og informasjon. Dataintensivitet var den eneste dimensjonen som ble presentert for respondentene i forkant av casebesvarelsen, i form av ulike datatyper.

For videre analyse har hver respondent fått 5 nye variabler heretter omtalt som; *N2 - Kre* (kreativitet), *N3 - Gje* (gjennomførbarhet), *N4 - Kom* (kommersialisering), *N5 - Dat* (datainntensivitet) og *N6 - Tot* (totalscore). Disse variablene blir benyttet for analyse i Rstudio i senere avsnitt.

4.3 Demografiske variabler

I denne delen tar vi for oss de demografiske variablene *Q3* til *Q13* fra spørreundersøkelsen, se tabell 4.1. De innledende spørsmålene i undersøkelsen omhandler kjønn, alder, arbeidsstatus, årslønn, utdanning og erfaring. Til slutt viser vi hvordan vi har plassert alle respondentene innenfor bakgrunnene; *forretningsbakgrunn*, *innovasjonsbakgrunn*, *teknologibakgrunn* og *ingen relevant bakgrunn*. Respondentene er kategorisert manuelt basert på de demografiske variablene beskrevet i denne delen. Analysen av de demografiske variablene er hentet fra Qualtrics sin rapportfunksjon og ligger som appendiks A1. Videre er Excel benyttet for de variablene som måtte behandles manuelt, slik som for de aktuelle bakgrunnene.

4.3.1 Kjønn

Q1 kartlegger kjønn for respondentene i undersøkelsen, som består av 32 kvinner og 60 menn - se tabell 4.3. Utvalget har en overvekt av mannlige respondenter, hvilket kan antas å speile nettverket som ble benyttet til å innhente besvarelser. Likevel ser man over 34% kvinnelige respondenter, hvilket gjør at vi har et rimelig utvalg med begge kjønn representert i undersøkelsen.

Kvinne	Mann
32	60
34,79%	65,21%

Tabell 4.3: Kjønn

4.3.2 Alder

Q4 kartlegger utvalgets alder og plasserer dem i fem aldersintervaller. Vi ser i tabell 4.4 at nesten 60% av respondentene er innenfor intervallet 20 til 29, hvilket også er naturlig med tanke på nettverket som er benyttet for å innhente dataene. Likevel observerer vi at 40% av respondentene er i de eldre aldersgrupper, som sørger for at aldersvariasjonen i utvalget også er tilfredsstillende.

20 til 29	30 til 39	40 til 49	50 til 59	60+
55	17	6	8	6
59,78%	18,48%	6,52%	8,70%	6,52%

Tabell 4.4: Alder

4.3.3 Arbeidsstatus

Q5 viser respondentenes arbeidsstatus og er kategorisert som enten student, arbeidstaker eller arbeidsledig/pensjonist, se tabell 4.5. I denne variabelen har vi et godt og variert utvalg relatert til student og arbeidstakere. Det kan nevnes av det ikke var mulig for respondenten å velge *både* student og arbeidstaker, som kan tenkes å være en kategori flere respondenter ville valgt om de ble gitt muligheten. Med hensyn til undersøkelsen var det ønskelig med jevn fordeling mellom arbeidstakere og studenter i utvalget, i lys av utrednings problemstilling. I forhold til respondenter i kategorien arbeidsledig/pensjonist har vi et lavt antall respondenter, som er noe lavere enn ønskelig.

Student	Arbeidstaker	Arbeidsledig
37	53	2
40,22%	57,61%	2,17%

Tabell 4.5: Arbeidsstatus

4.3.4 Årslønn

Q11 viser respondentenes årslønn, se tabell 4.6. Vi ser en god spredning i inntektsnivået, hvilket indikerer at utvalget består av både studenter med deltidsjobb, studenter uten deltidsjobb og arbeidstakere i ulike faser i karrieren. I tillegg har vi 34 respondenter med lønn over 600 000 kroner, hvilket indikerer at utvalget har respondenter med lengre erfaring.

0 til 199'	200' til 399'	400' til 599'	600' til 799'	800' til 999'	1 000'+	Ønsker ikke oppgi
22	11	21	17	7	9	5
23,91%	11,96%	22,83%	18,48%	7,61%	9,78%	5,43%

Tabell 4.6: Årslønn oppgitt i 1000

4.3.5 Utdanning

Q10 viser påbegynt utdanningsnivå for respondentene. Vi har et markant flertall av respondenter på mastergradnivå med hele 66,30% se tabell 4.7. Videre ser vi at over 22% av respondentene ligger i kategorien bachelorgrad og at om lag 10% av respondentene ligger i resterende kategorier. Vi vurderer variasjonen i utvalget tilfredsstillende og forventet, men vedgår at enda høyere variasjon kunne vært fordelaktig.

Grunnskole	Videregående	Bachelorgrad	Fagbrev/Fagskole	Mastergrad	PhD
0	4	21	4	61	2
0%	4,35%	22,83%	4,35%	66,30%	2,17%

Tabell 4.7: Utdanningsnivå

Q6 ser på hvilken utdanningsretning respondentene har. Hele 58,70% av respondentene har utdanning innen kategorien *Økonomi og administrasjon*. Dette er en allsidig utdanningskategori, og respondenter vil ikke uten videre plasseres i kategorien *forretningsbakgrunn* selv om de oppgir denne utdanningsbakgrunnen. Videre ser vi at den nest største retningen er *IT og media* etterfulgt av *Teknologi og ingeniør*. *Q7* kartlegger om respondentene har tilleggsutdanning og hele 38,64% har besvart dette spørmålet positivt. *Q8* spør videre disse respondentene om hvilken tilleggsutdanning de har og vi ser at de største kategoriene er *Økonomi og administrasjon*, *Annet* og *Helse og psykologi*. Tallgrunnlaget presentert i dette avsnittet kan finnes i sin helhet i appendiks A1.

4.3.6 Erfaring og bakgrunn

For å kartlegge de uavhengige variablene vi ønsker å måle på en hensiktsmessig måte har vi plassert alle respondenter innenfor kategoriene; *Innovasjonsbakgrunn*, *Teknologibakgrunn*, *Forretningsbakgrunn* og *Ingen relevant bakgrunn*. Vi har i metodekapittelet beskrevet de ulike bakgrunnene og vurderingene som ligger bak kategoriseringen. For å plassere respondentene i de riktige kategoriene har vi benyttet *Q12* som beskriver hvilken bransje respondentene jobber i, *Q8-Q10* som omhandler hvilken utdanning respondentene har,

samt *Q14-Q16* hvor respondentene oppgir på hvor mye erfaring de har innenfor de ulike kategoriene. I tabell 4.8 ser vi en god spredning i de ulike bakgrunnene hvor *Teknologibakgrunn* og *Forretningsbakgrunn* er de to største kategoriene. Sistnevnte kategori var forventet å utgjøre en betydelig andel med tanke på utvalget og den var på noen områder vanskelig å skille fra *Innovasjonsbakgrunn* som har mange likheter. Vi har derfor brukt skjønn og tilgjengelig data for å best mulig skille de to kategoriene. Bakgrunnene henvises heretter til som *N7 - Bak* (bakgrunner) og lagt inn i datasettet for videre analyse i Rstudio.

Innovasjonsbakgrunn	Teknologibakgrunn	Forretningsbakgrunn	Ingen relevant bakgrunn
19	27	24	22
20,65%	29,35%	26.09%	23,91 %

Tabell 4.8: Bakgrunner

4.4 Faktoranalyse

Målet med en faktoranalyse er å finne den enkleste mulig modellen som forklarer sammenhenger i data, og brukes derfor til å forklare mange ulike variabler ved hjelp av et mindre sett med faktorer (Friborg, 2011). Faktoranalyse gjør i teorien to ting; **(1)** identifisere antallet bakenforliggende faktorer variabelene kan reduseres til, og **(2)** hvordan kan disse variablene kombineres i ulike faktorscorer (Friborg, 2011). I denne delen vil vi analysere variablene *Q17-Q25*. Dette er variabler hvor vi har presentert respondentene for ulike utsagn og bedt dem oppgi i hvilken grad de er enige eller uenige. Variablene er delt i tre deler hvor *Q17-Q19* omhandler kreativitet, *Q20-Q22* omhandler kommersialisering, og *Q23-Q25* omhandler gjennomførbarhet. Målet er å se om vi kan identifisere felles faktorer innenfor disse tre kategoriene, som videre kan gi oss kunnskap gjennom regresjonsanalyse. Utgangspunktet er at vi ønsker å identifisere faktorer som kan forklare et individs vektlegging av kreativitet, kommersialisering og gjennomførbarhet ved idegenerering.

Vi ønsker å undersøke om observerbare variabler $X_1, X_2...X_N$ er lineært relatert til et lite antall uobserverbare (latente) faktorer $F_1, F_2...F_K$, med $K << N$.

$$X_1, = x_{1_0} + w_{1_1 F_1} + ...w_{1_K} + e_1$$

$$X_2, = x_{2_0} + w_{2_1 F_1} + ...w_{2_K} + e_2$$

...

$$X_N, = x_{N_0} + w_{N_1 F_1} + ...w_{N_K} + e_N$$

Hvor e_i er feilleddet, så relasjonshypotesen er ikke eksakt.

Prameteren w_{i_j} kalles *loadings*

Det er følgende forutsetninger:

A1 feilleddet e_i er uavhengige av hverandre $E(e_i) = 0$ og $var(e_i) = \alpha_i^2$

A2 De uobserverbare faktorene F_i er uavhengige fra hverandre $E(F_j) = 0$ og $var(F_j) = 1$

Med *loadings* er det mulig å beregne kovariansen til hvilke som helst av alle to observerte variabler:

$$Cov(X_{i,X_j}) = \sum_{k=1}^{K} w_{i_k w_{j\,k}}$$

og variansen til hver X_i

$$Var(X_i) = \sum_{k=1}^{K} w_{ik} + \alpha_i^2$$

Hvor summen er fellesskapet til variabelen, variansen forklart av fellesfaktorene F_k. Den gjenværende variansen kalles spesifikk varians (João, 2014).

4.4.1 Er faktoranalyse hensiktsmessig?

Vi starter analysen ved å laste inn hele datasettet i Rstudio og reduserer det til variablene vi ønsker analysere, *Q17-Q25.* For å sjekke om faktoranalyse er hensiktsmessig gjennomfører vi tre tester; **(1)** *Korrelasjonsmatrise* som sjekker om det er korrelasjoner mellom variablene, **(2)** *Kaiser-Meyer-Olkin (KMO)* som brukes til å måle prøvetakingstilstrekkelighet, og **(3)** *Bartlett's Test of Sphericity* som sjekker modellens signifikansnivå.

Korrelasjonsmatrise

Appendiks A3.1 viser korrelasjonene mellom variablene for å finne ut om en faktoranalyse er hensiktsmessig. Vi kan lese matrisen ved å se på styrke og forskjell på fargene. Mørkere blå farge antyder en sterkere positiv korrelasjon og en mørkere oransje farge antyder en sterkere negativ korrelasjon. I korrelasjonsmatrisen kan vi se at vi har både positive og negative korrelasjoner. I korrelasjonsmatrisen vil et resultat på 1 indikere perfekt positiv korrelasjon, mens -1 indikerer perfekt negativ korrelasjon mellom variablene. Vi har ingen korrelasjoner over 0.5 og de fleste ligger rundt 0.3.

Kaiser-Meyer-Olkin (KMO)

I følge Kaiser (1974) sine retningslinjer er en foreslått grenseverdi for å bestemme faktorbarheten til prøvedataene KMO $\geq$ 0,6. Den totale KMO er 0,66 - se tabell 4.9, noe som indikerer at vi basert på denne testen sannsynligvis kan gjennomføre en faktoranalyse. Likevel er den akkurat i grenseverdien og vi ser at *Pkom3* og *PGje2* er under 0,6. I tillegg ser vi av korrelasjonsmatrisen at verdiene har en lav korrelasjon og vi velger derfor å fjerne dem fra modellen.

Q17 - PKre1	Q18 - PKre2	Q19 - PKre3	Q20 - PKom1	Q21 - PKom2	Q22 - PKom3	Q23 - PGje1	Q24 - PGje2	Q25 - PGje3
0.75	0.64	0.66	0.63	0.69	0.54	0.64	0.58	0.80
Overall MSA	0.66							

Tabell 4.9: Kaiser-Meyer-Olkin (KMO)

Vi gjennomfører KMO på nytt og ser i tabell 4.10 en ny KMO på 0,7.

Q17 - PKre1	Q18 - PKre2	Q19 - PKre3	Q20 - PKom1	Q21 - PKom2	Q23 - PGje1	Q25 - PGje3
0.74	0.68	0.65	0.68	0.74	0.70	0.72
Overall MSA	0.7					

Tabell 4.10: Kaiser-Meyer-Olkin (KMO) 2

Bartlett's Test of Sphericity

Lave verdier (4,983183e-12 $<$ 0,05) av signifikansnivået indikerer at en faktoranalyse kan være nyttig med benyttet data - se tabell 4.11.

chisq	p.value	df
98.65671	4.983183e-12	21

Tabell 4.11: Bartlett's Test of Sphericity

Vi kan dermed konkludere basert på de tre gjennomførte testene at en faktoranalyse er hensiktsmessig for våre data.

4.4.2 Antall faktorer

I utgangspunktet ønsker vi tre faktorer som forklart innledningsvis. For å sjekke om dette gir mening lager vi en Scree-test og gjennomfører en Parallellanalyse.

Scree-test plotter faktorene som X-aksen og de tilsvarende *eigenvalues* som Y-aksen. Når man beveger seg til høyre, mot senere faktorer, synker *eigenvalues*. Når fallet opphører og kurven gjør en albue mot mindre bratt nedgang, sier Scree-test å droppe alle faktorene etter albuen. Appendiks A4.1 viser Scree-testen utført i denne faktormodellen. Albuen dropper mot to faktorer og igjen med et mindre fall mot tre faktorer. Dette antyder at vi helst burde benytte to faktorer.

Parallellanalyse *Kaiser-Guttman*-regelen sier at vi bør velge alle faktorer med *eigenvalues* større enn 1. Vi ser i Paralellanalysen i appendiks A5.1 at to faktorer er hensiktsmessig. Vi velger likevel å benytte tre faktorer for videre analyse, det er basert på Scree-testen som ligger nær grensen til 3 faktorer og formålet med dataen vi ønsker å utforske.

4.4.3 Resultat

Vi kjører faktoranalysen med *FA-metoden*. Vi benytter *Principal Axis Factor Analysis* (PA) og rotasjon er satt til *Varimax*.

	PA1	PA2	PA3	h2	u2	com
Q17 - PKre1	0.44	0.25	0.37	0.39	0.61	2.5
Q18 - PKre2	0.86	0.12	0.22	0.81	0.19	1.2
Q19 - PKre3	0.41	0.39	-0.03	0.32	0.68	2.0
Q20 - PKom1	0.06	0.80	0.15	0.66	0.34	1.1
Q21 - PKom2	-0.19	-0.41	-0.11	0.22	0.78	1.6
Q23 - PGje1	0.12	0.04	0.59	0.36	0.64	1.1
Q25 - PGje3	0.07	0.13	0.58	0.36	0.64	1.1
	PA1	PA2	PA3			
SS loadings	1.17	1.05	0.90			
Proportion Var	0.17	0.15	0.13			
Cumulative Var	0.17	0.32	0.45			
Proportion Explained	0.37	0.34	0.29			
Cumulative Proportion	0.37	0.71	1.00			

Tabell 4.12: Faktoranalyse

Vi ser i figur 4.12 at *PA1-PA3* er våre *loadings* (faktorer), og vi vil se at variablene med høyest nummer i hver rad går i samme faktor. Videre ser vi *h2* som forklarer hvor

mye faktorene forklarer variabelen. *Pkre2* er variabelen som er best forklart av faktorene. Motsatt viser *u2* hvilken variabel som er minst forklart. Vi ser at *Pkom2* er minst forklart i denne modellen. Vi kan også se at vi har en *cumulative variance* fra 0.17 i faktor *PA1* til 0,45 i faktor *PA3*. Dette tallet antyder hvor stor variansen er i snitt for hver faktor basert på alle variablene i faktoranalysen.

Appendiks A6.1 viser forholdet mellom faktorene og hvordan variablene er fordelt i de tre faktorene vi har dannet. Det er verdt å merke seg at *Q21-PKom2* har en rød strek. Dette betyr at den har en negativ korrelasjon, hvilket vi også kan se under *PA2* i figur 4.12. Det er ikke uvanlig å fjerne negative verdier, men vi har valgt å beholde den for å bedre kunne forstå faktoren.

Faktoranalysen har identifisert tre faktorer som vil bli benyttet for videre analyse. Vi vil heretter henvise til faktorene slik; **(F1)** - Faktor 1 forklarer individets syn kreativitet, **(F2)** - Faktor 2 forklarer individets syn på kommersialisering og **(F3)** - Faktor 3 forklarer individets syn på gjennomførbarhet. Faktorene vil bli lagt inn i datasettet med fortegnelsene *F1-F3* og kan ses i tabell 4.1.

4.4.4 Multippel lineær regresjon

Vi vil i dette avsnittet undersøke om det finnes signifikante verdier mellom faktorene og oppnådd score fra casebesvarelse. Vi vil benytte multippel lineær regresjon til å foreta analysen. For å forstå multippel lineær regresjon vil vi først presentere enkel lineær regresjon.

Enkel lineær regresjon brukes til å forutsi verdien av en utfallsvariabel Y, basert på en eller flere inngangsprediktorvariabler X. Målet med lineær regresjon er å modellere en kontinuerlig variabel Y som en matematisk funksjon av en eller flere X-variable(r), slik at vi kan bruke denne regresjonsmodellen til å forutsi Y når bare X er kjent. Denne matematiske ligningen kan generaliseres som følger;

$$Y = \beta_1 + \beta_2 X + \epsilon$$

Hvor β_1 er skjæringspunktet og β_2 er helningen. Til sammen kalles de regresjonskoeffisienter. ϵ er feilleddet, den delen av Y regresjonsmodellen ikke klarer å forklare (Gareth et al.,

2013).

Multippel lineære regresjonsmodeller er videre definert ved ligningen;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon$$

Den ligner ligningen for enkel lineær regresjon, bortsett fra at det er mer enn én uavhengig variabel $X_1, X_2, ..., X_p$.

Estimering av parametrene $\beta_0, ..., \beta_p$ ved minste kvadraters metode er basert på samme prinsipp som for enkel lineær regresjon, men benyttes på p dimensjoner. Det er altså ikke lenger et spørsmål om å finne den beste linjen som er nærmest punktparene y_i, x_i, men å finne det p dimensjonale planet som er nærmest koordinatpunktene $y_i, x_{i1}, ..., x_{ip}$ (Gareth et al., 2013).

Totalscore

Vi kjører en multippel lineær regresjon for Totalscore (*N6*) som avhengig variabel og faktorene *F1-F3* som uavhengige variabler - se appendiks A7.1.

Residualer: Oppsummerer residualene, avviket mellom prediksjonen av modellen og de faktiske resultatene. Mindre residualer er bedre. Vi ser av regresjonen at medianen er nær null, hvilket er ønskelig. Videre ser vi at spredningen mellom 1Q og 3Q mot medianen og *min* og *maks* mot medianen er nesten lik, som også er ønskelig.

Koeffisienter: For hver variabel og skjæringspunktet produseres det attributter som standardfeil, en t-testverdi og signifikans.

1. **Estimat** - dette er vekten som gis til variabelen. Den viser altså for hver enhet hvor mye de ulike faktorene påvirker. I denne modellen ser vi at hvor hver enhet 1 høyere i F2 tilføres modellen 0,02.

2. **Standardavvik** - forteller hvor nøyaktig anslaget ble målt. Det er egentlig bare nyttig for å beregne t-verdien. Til nærmere null denne er desto bedre, vi ser i denne modellen av vi har et relativt lavt standardsavvik for alle variabler.

3. **T-verdi og** $Pr(>[t])$ - T-verdien beregnes ved å ta koeffisienten delt på standardavvik. Den brukes deretter til å teste om koeffisienten er signifikant forskjellig

fra null. Hvis den ikke er signifikant, tilfører ikke koeffisienten noe til modellen og kan droppes eller undersøkes videre. $Pr(>[t])$ er signifikansnivået. Når t-verdien er høyere en 1.96 vil P-verdien være lavere enn 0,05. I denne modellen er det ingen signifikante nivåer. Det vil si at vi beholder null hypotesen for F1, F2, og F3 om at variablene ikke har noen påvirkning på modellen.

Ytelsesmålinger: tre sett med mål er gitt.

1. **Residual standardavvik** - dette er standardavviket til residualene, desto mindre jo bedre. I denne modellen er målingen på 1.339.
2. **Multiple / Justert R-squared** - for en variabel spiller ikke skillet noen rolle. *R-squared* viser mengden varians som er forklart i modellen. *Justert R-Squared* tar hensyn til antall variabler og er mest nyttig for multippel regresjon. I denne modellen ser vi på *justert R-squared* som ligger på -0.03. Denne vil vi ha så høy som mulig og siden den her er negativ kan vi si at vi ikke klarer å måle noe av variansen i Y. Modellen forklarer altså veldig dårlig predikasjonen av Y og vi ville ikke benytte denne modellen for videre analyse.
3. **F-statistikk** - sjekker om vekten til minst én variabel er vesentlig forskjellig fra null. Dette er en global test for å hjelpe med å vurdere en modell. Hvis p-verdien ikke er signifikant (f.eks. større enn 0,05) gjør modellen i hovedsak ingenting. Vi ser at vi har en total P-verdi i denne modellen 0.9927 som vil si at vi ikke har en signifikant modell.

Vi har videre kjørt regresjon for alle score verdiene *N2-N6*. Vi har ikke funnet noen signifikante verdier og modellene egner seg i liten grad til å predikere Y. Det vil altså si at vi kan beholde hypotesene om at faktorene ikke har noen påvirkning på variabel Y.

4.5 Multippel regresjon

De fire bakgrunnene; *Ingen relevant bakgrunn*, *Forretningsbakgrunn*, *Innovasjonsbakgrunn* og *Teknologibakgrunn* er våre uavhengige variabler som vi i dette avsnittet vil analysere nærmere. Vi ønsker å se om det er forskjell mellom de ulike bakgrunne og om vi kan finne signifikante verdier som belyser hvilke bakgrunner som er best på de ulike dimensjonene; *kreativitet*, *gjennomførbarhet*, *kommersialisering* og *datainntensivitet*, representert som de

avhengige variabler.

Flere variabler er omgjort til mindre kategorier eller binære variabler. *Q4 - Alder* er gjort binær ved at de under 40 år er kategorisert som *Yngre* og respondenter over 40 år er kategorisert som *Eldre*. *Q10 - Utdanningsnivå* er gjort binær ved at utdanningsnivå lavere enn bachelorgrad er kategorisert som *Lav* og fra bachelornivå og høyere er representert som *Høy*. *Q6 - Utdanningsretning* er delt inn i tre kategorier; *kommersiell utdanning*, *teknisk utdanning* og *annen utdanning*. *Q11 - Årslønn* er gjort binær ved at respondenter som har oppgitt under 400 000 kroner er representert som *Lav*, mens respondenter over 400 000 kroner kategoriseres som *Høy*. Endringene av variablene er gjort for å danne et datasett som er enklere å jobbe med og hvor vi får flere respondenter innen hver kategori. I resten av analysen vil den nye kategoriene benyttes.

4.5.1 Ingen relevant bakgrunn

Vi kjører fem lineære regresjoner hvor vi tester mot samtlige dimensjoner, sammen med totalscore som avhengige variabler. *Ingen relevant bakgrunn* er satt som referanseverdi og vi ønsker her å se om de andre bakgrunnene statistisk er bedre eller dårligere langs de ulike dimensjonene.

	N2 - Kre	N3 - Gje	N4 - Kom	N5 - Dat	N6 - Tot
N7 - Bakgrunn (Teknologibakgrunn)	0.91(***) 0.38 (SE)	1.21(***) 0.33 (SE)	1.02(**) 0.33 (SE)	1.32(**) 0.41 (SE)	1.10(**) 0.33 (SE)
N7 - Bakgrunn (Forretningsbakgrunn)	0.80(**) 0.38 (SE)	1.11(**) 0.33 (SE)	1.16(***) 0.33 (SE)	1.33(**) 0.41 (SE)	1.09(**) 0.33
N7 - Bakgrunn (Innovasjonsbakgrunn)	1.41(**) 0.44 (SE)	1.13(**) 0.38 (SE)	1.28(**) 0.37 (SE)	1.73(***) 0.46 (SE)	1.38(***) 0.38 (SE)
Konstant	2.80(***) 0.25 (SE)	2.79(***) 0.22 (SE)	2.64(***) 0.21 (SE)	3.12(***) 0.26 (SE)	2.85(***) 0.21 (SE)
N.	92	92	92	92	92
F-statistic	4.04(**)	6.11(***)	6.31(***)	6.54(***)	6.62(***)
Rsquared	0.12	0.17	0.17	0.18	0.18
Adjusted Rsquared	0.09	0.14	0.15	0.15	0.16

Tabell 4.13: Regresjon - Ingen relevant bakgrunn

Tabell 4.13 viser resultatene fra regresjonene som er utført. I tabellen ser vi de ulike bakgrunnene som representerer våre uavhengige variabler i første kolonne. I midten av modellen finner vi konstantleddet som forklarer estimatet til de avhengige variablene om referanseverdien legges til grunn. Første tall i andre rad, i andre kolonne er estimatet og

videre representerer (*) - (***) signifikansnivået fra en p-verdi under 0.05 til under 0.001. Tallet i raden under viser standardavviket. Dette oppsettet og måten en leser tabellene på er identisk for de resterende tabellene i fortsettelsen. Det vi kan trekke ut fra tabellen er at i forhold til *Ingen relevant bakgrunn* scorer de tre andre bakgrunnene signifikant høyere langs alle dimensjoner. Dette viser at respondenter i uten relevant bakgrunn scorer statistisk dårligere på idegenerering.

4.5.2 Forretningsbakgrunn

Vi kjører fem lineære regresjoner hvor vi tester mot samtlige dimensjoner, sammen med totalscore som avhengige variabler. *Forretningsbakgrunn* er satt som referanseverdi og vi ønsker her å se om de andre bakgrunnene statistisk er bedre eller dårligere langs de ulike dimensjonene.

	N2 - Kre	N3 - Gje	N4 - Kom	N5 - Dat	N6 - Tot
N7 - Bakgrunn	-0.80(*)	-1.11 (**)	-1.16(***)	-1.33(**)	-1.09(**)
(Ingen relevant bakgrunn)	0.38 (SE)	0.33 (SE)	0.33 (SE)	0.41 (SE)	0.33 (SE)
N7 - Bakgrunn	0.11	0.10	-0.14	-0.01	0.01
(Teknologibakgrunn)	0.41 (SE)	0.36 (SE)	0.35 (SE)	0.44 (SE)	0.35 (SE)
N7 - Bakgrunn	0.61	0.02	0.12	0.40	0.28
(Innovasjonsbakgrunn)	0.46 (SE)	0.40 (SE)	0.39 (SE)	0.49 (SE)	0.40 (SE)
Konstant	3.60(***)	3.91(***)	3.81(***)	4.46(***)	3.94(***)
	0.29 (SE)	0.25 (SE)	0.25 (SE)	0.31 (SE)	0.25 (SE)
N.	92	92	92	92	92
F-statistic	4.04(**)	6.11(***)	6.31(***)	6.54(***)	6.62(***)
Rsquared	0.12	0.17	0.17	0.18	0.18
Adjusted Rsquared	0.09	0.14	0.15	0.15	0.16

Tabell 4.14: Regresjon - Forretningsbakgrunn

Tabell 4.14 viser resultatene fra regresjonene utført. Vi har signifikante negative verdier på *Ingen relevant bakgrunn* langs samtlige dimensjoner, hvilket som representerer at *Forretningsbakgrunn* gir en høyere forventet score sammenlignet med *Ingen relevant bakgrunn*. For resterende bakgrunner er det ingen observerbare signifikante verdier og vi kan dermed ikke statistisk hevde at *Forretningsbakgrunn* gjør at man ventes å scorer bedre eller dårligere langs noen av dimensjonene. Vi kan likevel se tegn i resultatene på at respondenter med *Forretningsbakgrunn* scorer dårligere enn respondenter med teknisk- og innovasjonsbakgrunn på dimensjonene *kreativitet* og *gjennomførbarhet*, i tillegg til samlet *totalscore*. Det er også tegn til at *Forretningsbakgrunn* scorer dårligere en

Innovasjonsbakgrunn på *kommersialisering* og *datainntensivitet*. Videre ser vi tegn til at *Forretningsbakgrunn* scorer bedre en *Teknologibakgrunn* innen *kommersialisering* og *datainntensivitet*.

4.5.3 Innovasjonsbakgrunn

Vi kjører fem lineære regresjoner hvor vi tester mot samtlige dimensjoner, sammen med totalscore som avhengige variabler. *Innovasjonsbakgrunn* er satt som referanseverdi og vi ønsker her å se om de andre bakgrunnene statistisk er bedre eller dårligere langs de ulike dimensjonene.

	N2 - Kre	N3 - Gje	N4 - Kom	N5 - Dat	N6 - Tot
N7 - Bakgrunn (Forretningsbakgrunn)	-0.61 0.46 (SE)	-0.02 0.40 (SE)	-0.12 0.39 (SE)	-0.40 0.49 (SE)	-0.28 0.40 (SE)
N7 - Bakgrunn (Ingen relevant bakgrunn)	-1.41(**) 0.44 (SE)	-1.13(**) 0.38 (SE)	-1.28(**) 0.37 (SE)	-1.73(***) 0.46 (SE)	-1.38(***) 0.38 (SE)
N7 - Bakgrunn (Teknologibakgrunn)	-0.49 0.46 (SE)	0.08 0.40 (SE)	-0.26 0.39 (SE)	-0.41 0.49 (SE)	-0.27 0.40 (SE)
Konstant	4.22(***) 0.36 (SE)	3.93(***) 0.31 (SE)	3.93(***) 0.31 (SE)	4.86(***) 0.38 (SE)	4.23(***) 0.31 (SE)
N.	92	92	92	92	92
F-statistic	4.04(**)	6.11(***)	6.31(***)	6.54(***)	6.62(***)
Rsquared	0.12	0.17	0.17	0.18	0.18
Adjusted Rsquared	0.09	0.14	0.15	0.15	0.16

Tabell 4.15: Regresjon - Innovasjonsbakgrunn

Tabell 4.15 viser resultatene fra regresjonene utført. Vi har signifikante negative verdier på *Ingen relevant bakgrunn* langs alle dimensjoner som indikerer at *Innovasjonsbakgrunn* gir en høyere forventet score enn *Ingen relevant bakgrunn*. Videre ser vi tegn fra resultatene på at *Innovasjonsbakgrunn* forventes å score høyere enn de andre bakgrunnene på nesten samtlige dimensjoner. Det eneste unntaket er at det er tegn til at *Innovasjonsbakgrunn* scorer dårligere enn *Teknologibakgrunn* på dimensjonen *gjennomførbarhet*.

4.5.4 Teknologibakgrunn

Vi kjører fem lineære regresjoner hvor vi tester mot samtlige dimensjoner, sammen med totalscore som avhengige variabler. *Teknologibakgrunn* er satt som referanseverdi og vi ønsker her å se om de andre bakgrunnene statistisk er bedre eller dårligere langs de ulike dimensjonene.

	N2 - Kre	N3 - Gje	N4 - Kom	N5 - Dat	N6 - Tot
N7 - Bakgrunn	0.49	-0.08	0.26	0.41	0.27
(Innovasjonsbakgrunn	0.46 (SE)	0.40 (SE)	0.39 (SE)	0.49 (SE)	0.40 (SE)
N7 - Bakgrunn	-1.11	-0.10	0.14	0.01	-0.01
(Forretningsbakgrunn)	0.41 (SE)	0.36 (SE)	0.35 (SE)	0.44 (SE)	0.35 (SE)
N7 - Bakgrunn	-0.91(*)	-1.21(***)	-1.02(**)	-1.32(**)	-1.10(**)
(Ingen relevant bakgrunn)	0.38 (SE)	0.33 (SE)	0.33 (SE)	0.41 (SE)	0.33 (SE)
Konstant	3.72(***)	4.01(***)	3.66(***)	4.44(***)	3.96(***)
	0.29 (SE)	0.25 (SE)	0.25 (SE)	0.31 (SE)	0.25 (SE)
N.	92	92	92	92	92
F-statistic	4.04(**)	6.11(***)	6.31(***)	6.54(***)	6.62(***)
Rsquared	0.12	0.17	0.17	0.18	0.18
Adjusted Rsquared	0.09	0.14	0.15	0.15	0.16

Tabell 4.16: Regresjon - Teknologibakgrunn

Tabell 4.16 viser resultatene fra regresjonene utført. Vi har signifikant negative verdier på ingen relevant bakgrunn på alle dimensjoner som representerer at tekniskbakgrunn gir en høyere score en ingen relevant bakgrunn. Videre ser vi tegn i dataen på at tekniskbakgrunn scorer høyere en forretningsbakgrunn på kreativitet, gjennomførbarhet samt høyere en innovasjonsbakgrunn på gjennomførbarhet. På resterende verdier er det tegn til at *Teknologibakgrunn* scorer dårligere en de andre bakgrunnene.

4.6 Andre funn

I denne delen av analysen presenteres andre relevante funn fra resultatene. Vi begynner med å kjøre to multiple regresjoner med antall ideer generert og antall minutter benyttet på casebesvarelsen som avhengige variabler. Videre betrakter vi ulike sammenhenger i resultatene, hvor vi trekker ut og legger til variabler for å grave dypere i datamaterialet. Her ser vi på variabler som *Alder*, *Kjønn*, *Arbeidsstatus*, og *Årsinntekt*. Mot slutten lager vi fem regresjonsmodeller av ulik størrelse og med ulike kontrollvariabler for å se videre på sammenhenger i datamaterialet. Helt til slutt ser vi nærmere på modellen vi mener predikerer Y best og gjennomfører en *10-fold cross validation* (Gareth et al., 2013).

4.6.1 Ideer og tidsbruk

I dette avsnittet har vi kjørt to multiple regresjoner med *ideer* og *tidsbruk* som avhengige variabler - se appendiks A8.1. Bakgrunne er de uavhengige variablene hvor *Ingen relevant*

bakgrunn er satt som referanseverdi. De øvrige referanseverdiene i modellen er satt som følger;

- Q3 - Kjønn = Kvinne
- Q4 - Alder = Eldre
- Q5 - Arbeidsstatus = Student
- Q6 - Utdanningsretning = Kommersiell utdanning
- Q7 - Tilleggsutdanning = Nei
- Q10 - Utdanningsnivå = Lav

Tabell A8.1 viser i andre rad, andre kolonne regresjonens estimat og signifikansnivå med asterisk (*) - (***). Tallet under viser standardavviket og under statistikk finner vi antall observasjoner, *F-statistic*, *Rsquared* og *Adjusted Rsquared*.

4.6.2 Ideer

I den andre kolonnen i appendiks A8.1 presenteres regresjonen for *Ideer*. Vi ønsker altså å undersøke hva som påvirker *antall ideer* respondentene har produsert i caseoppgaven. Vi har signifikante nivåer på alle bakgrunnen og *Innovasjonsbakgrunn* skiller seg ut med et høyt estimat på 1.53. Vi ser imidlertid at alle bakgrunner har positive estimat, hvilket antyder at referanseverdien *Ingen relevant bakgrunn* predikeres til å gi et lavere antall ideer. Videre ser vi tegn i dataene som indikerer at kvinner har et høyere predikert estimat. *Tilleggsutdanning* og *kommersiell* utdanningsretning tyder virker også indikere et høyere estimat. Resultatene som viser at *Tilleggsutdanning* kan ha en positiv påvirkning på antall ideer må sees i sammenheng med at også de eldre respondentene har et positivt estimat. Det er også tegn i resultatene som trekker i retning av at arbeidsledige produserer færre ideer. Til slutt ser vi at modellen i sin helhet ikke er signifikant og har en *Adjusted Rsquared* på 0.12. Modellen forklarer altså 12% av variasjonen i antall ideer generert.

Modellen gir interessante resultater og indikerer interessante funn, men det er likevel tydelig at mange sammenhenger forklares i liten grad og at mer data og flere variabler antagelig kunne gitt bedre estimater.

4.6.3 Tidsbruk

I tredje kolonne i appendiks A8.1 finner vi regresjonen for tidsbruk. Her ønsker vi å se på hva som eventuelt påvirker tiden respondentene har benyttet på caseoppgaven. Denne modellen er signifikant og har en høyere *Adjusted Rsquared* enn for ideer på 0.36. Vi har signifikante nivåer på *Forretningsbakgrunn* og arbeidsstatus *Arbeidsledig*. Vi ser at alle bakgrunner har positive estimater, men at *Forretningsbakgrunn* og *Innovasjonsbakgrunn* har et langt høyere estimat enn *Teknologibakgrunn*. Det er altså tegn i dataen på at respondenter i kategoriene *Teknologibakgrunn* og *Ingen relevant bakgrunn* bruker mindre tid på caseoppgaven. Videre ser vi tegn til at respondenter i kategoriene *Kvinner* og *Yngre* bruker lengre tid, mens respondenter i kategoriene *Student* og *Høy utdannelsenivå* studenter bruker mindre tid. I forbindelse med at at de med teknologibakgrunn bruker relativt kort tid på idegenerering, ser vi også at respondenter med utdannelse innen teknisk retning har et negativt estimat.

4.6.4 Alder og kjønn

I tabellene 4.17 og 4.18 har vi valgt å fjerne respondentene i kategorien *Ingen relevant bakgrunn*. Dette har vi gjort for å fremheve forskjellene i *Alder* og *Kjønn* i relasjon til score utvalget ved å fjerne respondenter uten relevant bakgrunn. Verdiene i tabellen viser gjennomsnittsscore for de ulike dimensjonene; *kreativitet*, *gjennomførbarhet*, *kommersialisering*, *datainntensivitet*, samt *totalscore*. Vi begynner med å se på gjennomsnittet til alder og kjønn når vi fjerner respondenter innen *Ingen relevant bakgrunn*. Videre fjerner vi respondenter i kategorien *Lav* på utdannelse, og sjekker om gjennomsnittetene påvirkes. Til slutt fjernes respondenter som ikke har *Tilleggsutdannelse* og vi sjekker på nytt hvordan dette påvirker gjennomsnittscore.

Score	Fjernet Ingen relevant bakgrunn		Høy utdannelse		Tilleggsutdanning	
	Kvinner	Menn	Kvinner	Menn	Kvinner	Menn
Kreativitet	3.81	3.79	3.86	3.85	3.66	3.80
Gjennomførbarhet	3.90	3.97	3.92	3.98	3.73	3.88
Kommersialisering	3.37	3.96*	3.37	3.99	3.33	3.82
Datainntensivitet	4.42	4.61	4.49	4.66	4.26	4.80
Total score	3.87	4.08	3.91	4.12	3.75	4.07

Tabell 4.17: Kjønn gjennomsnitt

Tabell 4.17 viser at det finnes variasjon blant kvinner og menn med ulikt utdanningsnivå det gjelder idegenerering og score. Dataene gir kun ett signifikant resultat, markert med én asteriks ("*"). Resultatet viser at menn uten relevant bakgrunn scorer tydelig høyere langs kommersialiserings-dimensjonen, målt mot kvinner uten relevant bakgrunn. Vi ser en generell trend av resultatene som tyder på at menn scorer noe bedre en kvinner på gjennomsnittsverdien til de ulike dimensjonene i idegenereringen, uten at resultatet er signifikant. Dette gjelder både når vi fjerner respondenter uten relevant bakgrunn, lav utdanning og uten tilleggsutdannelse.

Score	Fjernet Ingen relevant bakgrunn		Høy utdannelse		Tilleggsutdanning	
	Yngre	Eldre	Yngre	Eldre	Yngre	Eldre
Kreativitet	3.84	3.66	3.87	3.77	3.88	3.5
Gjennomførbarhet	3.97	3.89	3.97	3.94	3.78	4.00
Kommersialisering	3.84	3.58	3.85	3.63	3.64	3.83
Datainntensivitet	4.59	4.43	4.63	4.55	4.73	4.50
Total score	4.06	3.89	4.08	3.97	4.01	3.95

Tabell 4.18: Alder gjennomsnitt

Tabell 4.18 gir ingen signifikante funn, men en observerbar tendens i resultatene indikerer at yngre respondenter scorer høyere enn eldre respondenter langs alle dimensjonene. Vi har definert kategoriene *Yngre* til å omfatte respondenter fra 20 til 39 år og *Eldre* fra 40 til 60+ år, og kan dermed ikke peke på en mer spesifisert aldersgruppe som er ventet å score bedre enn andre. Videre kan vi ikke se at skillet mellom gjennomsnittscoren til Eldre og Yngre endres markant når vi fjerner respondenter i kategoriene *Lav* på utdanningsnivå og uten *Tilleggsutdannelse*.

4.6.5 Arbeidstaker og student

Tabell 4.19 viser gjennomsnittsverdien til *N6 - Totalscore* hvor vi ser på utdanningsretning mot arbeidsstatus. De to respondentene som var arbeidsledige er fjernet fra datasettet i denne analysen, ettersom vi ønsker å se nærmere på forskjeller mellom studenter og arbeidstakere. Vi finner ingen signifikante resultater, men vi ser imidlertid av dataene at arbeidstakere med teknologisk utdannelse virker å oppnå høyest gjennomsnittscore. Vi har ingen studenter med *teknisk* utdanningsretning i datasettet og kan dermed ikke sammenligne resultatet mot studenter. Videre ser vi at gjennomsnittet for *kommersiell* utdanning er nokså lik både for arbeidstakere og studenter. Blant respondentene med

utdanningsretningen *Annet* ser vi en tendens til at studentene scorer høyest, men utover det er det vanskelig å lese noe mer konkret av resultatet.

	Arbeidsstatus	Utdanningsretning	Gjennomsnitt
Totalscore	Arbeidstaker	Annet	3.13
		Kommersiell	3.69
		Teknisk	4.03
	Student	Annet	3.72
		Kommersiell	3.66

Tabell 4.19: Student versus Arbeidstaker - Totalscore

4.6.6 Høy og lav lønn

Tabellen 4.20 viser flere signifikante resultater mellom oppnådd score og lønnsnivå. Markert med én asteriks ("*") viser resultatene markante forskjeller i oppnådd score mellom høyt- og lavtlønnede. Vi har definert Lav lønnfra 399 999 kroner og lavere, mens "Høy lønnfra 400.000 kroner og høyere. Resultatet kan ved første øyekast indikere at lønnsnivå spiller en rolle for idegenerering, men det er rimelig å tenke seg at variabelen fanger opp ulike uobserverbare egenskaper. Dette kan for eksempel være hvor flink respondenten er i sitt daglige virke, lønnsnivået i bransjen vedkommende jobber i og respondentens alder. Verdiene i tabellen viser gjennomsnittet for *N6 - Totalscore* og dataen viser at høy lønn har et markant høyere gjennomsnitt på samtlige dimensjoner.

	Lav lønn	Høy lønn
Kreativitet	3.40*	4.02*
Gjennomførbarhet	3.77	4.05
Kommersialisering	3.54	3.92
Datainntensivitet	4.15*	4.78*
Total score	3.71*	4.19*

Tabell 4.20: Lønn gjennomsnitt

Tabellen 4.21 viser gjennomsnittet langs de ulike dimensjonene når vi sammenligner lav og høy lønn mellom kjønn. Vi har ingen signifikante resultater, men vi ser en tendens til at menn scorer høyere en kvinner både ved lav og høy lønn. Det vi kan merke oss av resultatet er at differansen i gjennomsnittscorene mellom menn og kvinner øker når vi går fra lav til høy lønn. Vi ser altså tegn i dataen til at differansen mellom kvinner og menn blir større når vi går fra lav til høy lønn, hvilket også trekker i retning av at det finnes en del uobserverbare egenskaper ved lønns-variabelen.

Score	Lav lønn		Høy lønn	
	Kvinne	Mann	Kvinne	Mann
Kreativitet	3.60	3.74	3.95	4.33
Datainntensivitet	3.75	3.33	3.83	4.13
Kommersialisering	3.58	3.81	4.00	4.09
Datainntensivitet	3.16	3.62	3.42	4.20
Totalscore	3.91	4.20	4.57	4.90

Tabell 4.21: Kjønn - Høy og lav lønn

4.6.7 Større modeller

I denne delen gjennomfører vi fem multiple regresjoner presentert gjennom fem ulike modeller, hvor vi legger til flere kontrollvariabler for hver modell - se tabell 4.22. Vi ønsker her å se det store bildet og undersøke hvilke variabler som påvirker Y. Appendiks A9.1 viser hvilke variabler som benyttes i modellene. Vi benytter *N6 - Totalscore* som avhengig variabel, bakgrunnene som uavhengige variabler og de resterende variabler som kontrollvariabler. Siste kolonne i appendiks A9.1 viser hvilke referanseverdi som ligger til grunn for regresjonene. Modell tre viser seg å forklare mest av variansen i Y og vi vil til slutt gjennomføre en dypere analyse med den modellen hvor vi gjennomfører *cross-validation* for å predikere Y.

Modell 1

I den første modellen i tabell 4.22 har vi ikke tillagt noen kontrollvariabler. Vi finner signifikante nivåer på alle bakgrunner og i modellen samlet sett. *Innovasjonsbakgrunn* ser vi har en større effekt på totalscoren, relativt til de øvrige bakgrunnene. Vi har i modellen en *Adjusted Rsquared* på 0.15, som vil si at modellen forklarer 15% av variansen i Y. Vi kan dermed konkludere med at modellen forklarer en del av variansen i Y, samtidig som det er mye uobserverbart i den avhengige variabelen.

Modell 2

I modell 2 i tabell 4.22 har vi lagt til kontrollvariablene *Q3 - Kjønn* og *Q4 - Alder*. Vi finner ingen signifikante nivåer på kontrollvariablene, og av resultatet kan vi ikke konkludere med at noen av kjønnene scorer høyere. Vi ser imidlertid antydninger til at yngre respondenter scorer høyere enn eldre. Modellen er fortsatt signifikant og vi har en *Adjusted Raquared* på 0.15. Modell 2 forklarer dermed ingenting ekstra av variansen i Y i forhold til den første

modellen (Modell 1).

Modell 3

I modell 3 fra tabell 4.22 har vi lagt til kontrollvariabelen *Q10 - Utdanningsnivå* i tillegg til variablene i modell 2. Utdanningsnivået til respondentene gir et signifikant resultat, hvor vi ser at høyt utdanningsnivå fører til en høyere totalscore. Funnet indikerer at høy utdanning gir 0.44 ekstra i score-predikasjonen av Y. Modellen er signifikant og har en *Adjusted Rsquared* på 0.21, som viser at inkludering av utdanningsnivå forklarer mer av variansen i Y, sammenlignet med de øvrige modellene. Modell 3 er så langt den beste modellen i å kunne predikere hvordan respondentene vil score.

Modell 4

I modell 4 har vi lagt til *Q6 - Utdanningsretning* og *Q7 - Tilleggsutdanning* i tillegg til variablene i modell 3 - se tabell 4.22. De nye variablene viser ingen signifikante verdier, men resultatene indikerer likevel at både utdanningsretning *Annet* og *Teknisk* gir en høyere score. Det er også indikasjoner på at *Tilleggsutdanning* påvirker oppnådd score positivt.. Modellen er signifikant og vi har en *Adjusted Rsquared* på 0.21. Denne modellen forklarer dermed mindre av variansen i Y enn modell 3, som i tillegg har færre variabler inkludert. Mange variabler i modellen øker sannsynligheten for at ulike variabler korrelerer med hverandre, og dermed påvirker resultatet av modellen.

Modell 5

I tillegg til variablene i modell 4 har vi lagt til kontroll variablene *Q5 - Arbeidsstatus* og *Q11 - Årslønn* i modell 5 - se 4.22. De nye variablene viser ingen signifikante verdier, men resultatene tyder på at høy årslønn predikerer en høyere totalscore. Som tidligere nevnt kan dette ha sammenheng med uobservere egenskaper i variabelen, og bør derfor ikke tillegges for mye vekt. Modellen er signifikant, men vi ser at *Adjusted Rsquared* har falt til 0.18. Dette betyr i korte trekk at inkludering av alle variablene ikke nødvendigvis gir en modell som gir en bedre predikasjon av Y.

Uavhengige variabler	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
N7 - Bakgrunn	1.09(**)	1.07(**)	0.92(**)	1.09(**)	1.03(**)
(Forretningsbakgrunn)	0.33 (SE)	0.34 (SE)	0.33 (SE)	0.36 (SE)	0.39 (SE)
	1.38(***)	1.36(***)	1.15(**)	1.32(**)	1.29(**)

N7 - Bakgrunn					
(Innovasjonsbakgrunn)	0.38 (SE)	0.39 (SE)	0.38 (SE)	0.40 (SE)	0.41 (SE)
N7 - Bakgrunn	1.10(**)	1.17(***)	1.06(**)	1.09(**)	1.06(**)
(Teknologibakgrunn)	0.33 (SE)	0.34 (SE)	0.33	0.37 (SE)	0.38 (SE)
Kontrollvariabler					
Q3 - Kjønn		-0.12	-0.04	0.06	0.07
(Mann)		0.27 (SE)	0.26 (SE)	0.29 (SE)	0.30 (SE)
Q4 - Alder		0.31	0.24	0.22	0.24
(Yngre)		0.32 (SE)	0.31 (SE)	0.32 (SE)	0.35 (SE)
Q5 - Arbeidsstatus					-0.11
(Arbeidstaker)					1.02 (SE)
Q5 - Arbeidsstatus					0.10
(Student)					0.96 (SE)
Q6 - Utdanningsretning				0.52	0.50
(Annet)				0.39 (SE)	0.41 (SE)
Q6 - Utdanningsretning				0.07	0.06
(Teknisk)				0.41 (SE)	0.43 (SE)
Q7 - Tilleggsutdanning				0.20	0.19
(ja)				0.27 (SE)	0.28 (SE)
Q10 - Utdanningsnivå			1.24(**)	1.52(**)	1.48(**)
(Høy)			0.44 (SE)	0.50 (SE)	0.53 (SE)
Q11 - Årslønn					0.29
(Høy)					0.50 (SE)
Konstant - Totalscore	2.85(***)	2.67(***)	1.64(**)	1.06	0.96
	0.22 (SE)	0.38 (SE)	0.52 (SE)	0.65 (SE)	1.01 (SE)
Statistikk					
N.	92	92	92	92	92
F-statistic	6.62(***)	4.18(***)	5.06(***)	3.57(***)	2.63(**)
Rsquared	0.18	0.19	0.26	0.28	0.28
Adjusted Rsquared	0.15	0.15	0.21	0.20	0.18

Tabell 4.22: Fem større modeller

4.6.8 Beste modell

I dette avsnittet ser vi nærmere på resultatet fra modell 3, som blant de fem modellene var den beste. Videre gjennomfører vi en *10-fold cross-validation*, før vi undersøker godt modellen er egnet til å predikere Y.

Resultat Vi kjører en multippel lineær regresjon og får følgende resultat i R - se appendiks A10.1. Median til residualene ligger nærme 0 og intervallene er nokså like på hver side av medianen. Vi har en justert R-squared på 0.211 og modellen kan derfor forklare ca 21% av variansen til Y. Videre har vi en p-verdi for modellen på 0.0001762 vi kan dermed forkaste nullhypotesen og stadfeste modellen har påvirkning på Y. Vi ser signifikante nivåer på alle bakgrunnene og kontrollvariabelen *Q10 - Utdanningsnivå*.

Modell 3 kan formuleres på følgende måte;

$$\begin{aligned} Y_{Totalscore} = 1.64 + 0.92X_{For} + 1.15X_{Inn} + 1.06X_{Tek} + \\ 1.24X_{Univ} - 0.04X_{Mann} + 0.25X_{Yngre} \end{aligned} \tag{4.1}$$

K-fold cross-validation

Denne tilnærmingen innebærer en tilfeldig *k-fold Cross-validation* som deler opp settet med observasjoner i k grupper, eller folder, på ca lik størrelse. Den første folden behandles som et valideringssett, og metoden passer på de resterende $K-1$ foldene. Den gjennomsnittlige kvadratiske feilen, $MSE1$, er deretter beregnet på observasjonene i den utholdte folden. Denne prosedyren er gjentatt k ganger; hver gang behandles en annen gruppe observasjoner som et valideringssett. Denne prosessen resulterer i k estimater av testfeilen, $MSE1, MSE2, ..., MSEk$. K-fold CV-estimat beregnes ved gjennomsnittsberegning av disse verdiene (Gareth et al., 2013). Vi har kjørt en *10-fold cross validation*, er repetert 3 ganger. Matematisk kan *K-fold cross validation* skrives på følgende måte:

$$CV_{(k)} = 1/K \sum_{i=1}^{K} MSE_i$$

Tabell 4.23 viser resultatene fra valideringsmetoden. **Root Mean Squared Error (RMSE)** er kvadratroten av den gjennomsnittlige kvadratiske forskjellen mellom den

faktiske verdien og den anslåtte verdien til målvariabelen. Det gir den gjennomsnittlige prediksjonsfeilen laget av modellen (Gareth et al., 2013). Vi har i denne modellen en *RMSE* på 1.18, som kan anses som en relativ høy verdi. Subtraherer vi minimumsverdien fra maksverdien til *N6 - Totalscore* og dividerer *RMSE* på dette tallet får vi 0.25. Tallet representerer en verdi mellom 0 og 1, vi ønsker så lav verdi som mulig. I denne modellen er det tydelig at *RMSE* er relativ høy og vi har dermed en høyere verdi av prediksjonsfeil i modellen en vi ønsker.

Rsquared gir en idé om hvor stor prosentandel av variansen i den avhengige variabelen som forklares samlet av de uavhengige variablene - se tabell 4.23. Det reflekterer med andre ord forholdsstyrken mellom målevariabelen og modellen på en skala fra 0 – 100% (Gareth et al., 2013). Modellen forklarer dermed 28% av variansen i den avhengige variabelen Y. Resultatet blir som regel brukt til å si hvor godt regresjonsmodellen passer til de observerte dataene. 28% er relativt lavt og vi skulle ønskelig hatt en $Rsquared > 0.6$.

Mean Absolute Error (MAE) er den absolutte forskjellen mellom de faktiske verdiene og verdiene forutsagt av modellen for målvariabelen - se tabell 4.23. Hvis verdien av *outliers* ikke har mye å gjøre med nøyaktigheten til modellen, kan *MAE* brukes til å evaluere ytelsen til modellen (Gareth et al., 2013). En lav verdi er ønskelig ettersom vi da treffer bedre med predikasjonene mot den faktiske verdien. I snitt bommer altså modellen med 0.96 når den prøver å predikere *N6 - Totalscore*. Med tanke på at totalscore går fra 1 til 5.66 vil 0.96 være en relativt høy verdi som ønskelig skulle vært lavere.

RMSE	Rsquared	MAE
1.18	0.28	0.96

Tabell 4.23: K-fold CV

5 Diskusjon

5.1 Utdanning og bakgrunn

Ekspertpanelet fra Knowit vurderte ideene forholdsvis likt, med total gjennomsnittscore på 3,63. En forklaring kan derfor være at utvalget hadde god variasjon og et bredt spekter av ulik kunnskapskapital. Den jevne gjennomsnittscoren langs dimensjonene kan også ha sammenheng med at de ulike profilene har vurdert de ulike dimensjon likt, og har dermed vektlagt dem på samme måte. Dette illustrerer også den overordnede problemstillingen om hvem dataintensiv innovasjon er en oppgave for, ettersom resultatene ikke peker tydelig i én retning.

På tross av at masterutredningen i all hovedsak er basert på kvantitativ data og metode, ligger det kreative og interessante funn bak tallene. Casebesvarelsen fikk nær hundre respondenter, og med det en lang rekke kreative ideer som fortjener å benyttes anekdotisk i en drøftelse om innovasjon. Selv om totalscore langs alle dimensjonene var relativt like, finnes det enkelte ideer som i seg selv kan være interessante å se nærmere på. Følgende idé scoret høyt på alle dimensjonene, og trekkes derfor frem som en av de beste ideene.

Data som trengs: Sensor på bil, og denne dataen er anskaffet fra en persons tidligere kjørehistorikk. Forsikringsselskapet inngår samarbeidsavtale med en bilforhandler. Bilforhandleren selger bilen med forsikring, men forsikringen er bakt inn i prisen. Det vil si at en trygg sjåfør får en billigere bil enn en utrygg sjåfør. Bilforhandleren gir så forsikringsselskapet «kick-back» på «forsikringsdelen» av salget (f.eks forsikring som er en fastpris, varer 2 år). Dermed unngår man forsikringsselskapers store problem med fallende informasjonsasymmetri, og kan igjen oppnå høye priser, fordi det inngår i prisen på bilen..."

Respondenten presenterer ideen med utgangspunkt i tilgjengelig data, hvilket er rimelig å forvente med tanke på caseoppgavens utforming. Ideen presenterer også en tydelig relasjon mellom dataintensivitet og kommersialisering, ettersom respondenten i klartekst forklarer hvordan selskapet skal distribuere og tjene penger på den. All teknologi som kreves for å realisere ideen må antas å eksistere, og respondenten adresserer derfor implisitt gjennomførbarhets-dimensjonen på en måte som falt i smak blant ekspertene. Det kreative

aspektet ved ideen kan tenkes å ligge i utformingen av tjenesten, og hvordan et bi-produkt av tjenesten skaper verdi for selskapet gjennom redusert informasjonasymmetri.

Den høye scoren langs samtlige dimensjoner indikerer ingen spesifikk utdanning- eller erfaringsbakgrunn, men ideens tydelige fokus på inntjening gjennom en provisjonsmodell og potensielt økt verdikapring trekker i retning av at respondenten kan ha forretningsbakgrunn. Ved å spore ideen tilbake til respondenten bekreftes mistanken om forretningsbakgrunn, men interessant nok oppgir den mannlige respondenten at kommersielt potensiale er *svært uviktig* ved idegenerering. Respondenten faller inn i det yngre aldersintervallet, er høyt utdannet på masternivå og oppgir en årsinntekt i den høye kategorien. I sum passer dermed respondenten godt inn i den demografiske profilen som er ventet å oppnå en høy totalscore. Som analysen og tabell 4.14 viser kan man ikke statistisk hevde at forretningsbakgrunn gir bedre eller dårligere score langs noen av dimensjonene. Eksempelet står dermed igjen kun som et anekdotisk argument for at forretningsprofiler kan spille en viktig rolle i dataintensive innovasjonsaktiviteter.

5.1.1 Utdannings- og erfaringsbakgrunn

Den klareste trenden i analysen indikerer at profiler med innovasjonsbakgrunn scorer høyere enn de andre bakgrunnene totalt sett, og på nær samtlige enkelt-dimensjoner, se tabell 4.15. Dette funnet kan ved første øyekast betegnes som overraskende, med tanke på at flere av dimensjonene rimelig kan knyttes til spesifikk teknisk- eller kommersiell kompetanse. En mulig forklaring kan ligge i utvalget av respondenter med innovasjonsbakgrunn, som kan antas å inneholde et bredt spekter av profiler, med innslag av både teknisk og kommersiell humankapital. I lys av dette resultatet er det rimelig å hevde at individer med innovasjonsbakgrunn spiller en viktig rolle i innovasjonsarbeid, hvilket i seg selv ikke er et overraskende funn. Det er likevel et interessant resultat som kan bidra til å besvare deler av masterutredningens overordnede problemstilling, nemlig *om datadrevet innovasjon er en oppgave for økonomer eller teknologer.*

Det er også interessant at innovasjonsprofilen har en høyere estimert score langs dimensjonene som man kan anta at teknologer og økonomer bør ha sine respektive fortrinn innen, se tabell 4.15. En mulig forklaring kan være at erfaring og kjennskap til innovasjonsprosesser har gitt respondenter med innovasjonsbakgrunn et utslagsgivende

fortrinn. Dette må sees i lys av at caseoppgaven konkret oppfordrer respondentene til idégenerering, og søker dermed å etterligne den første fasen i en innovasjonsprosess. Forskningen understøtter at erfaring med rammeverk og innovasjonsmetodikk kan påvirke evnen til nytenkning, og det kan tenkes at det er en slik effekt vi ser tendensene av her, særlig i relasjon til kreativitetsdimensjonen (Christensen et al., 2016a). I tillegg ser vi at individer med innovasjonsbakgrunn har flere ideer og bruker kortere tid på å generere dem, hvilket må betegnes som et interessant funn som underbygger indikasjonen om at innovasjonserfaring er en fordel i idégenereringprosessen.

At respondenter med innovasjonsbakgrunn oppnår en høyere score på kommersialiseringsaspektet enn de øvrige profilene indikerer at tverrfaglig kompetanse og evne til å vekte ulike dimensjoner spiller en vesentlig rolle ved idegenerering. Resultatene blant respondentene med tilleggsutdanning understøtter denne påstanden, hvor analysen avdekker at tilleggsutdanning virker å være en faktor i å lykkes med idegenerering. At en person har tilleggsutdanning illustrerer godt bildet av hybrid-fagperson, som analysen viser oppnår høy score.

Høy score på dataintensivitet kan ha en lignende forklaring som for kommersialisering, i at innovasjonsprofilene i større grad evner å se mulighetsrommet som ligger i bruk av data, uten å nødvendigvis inneha omfattende teknisk innsikt. I sum tegner resultatene et bilde av hybrid-fagpersonen som svært verdifull i innovasjonsprosesser, hvilket underbygger påstanden om at innovasjon er komplisert og krever tverrfaglig kompetanse (Christensen et al., 2016a). Det er også i tråd med forskningen som fremhever mangfold som en positiv faktor på innovasjon, eksemplifisert ved at innovasjonsbakgrunn-kategorien antas å inneholde et bredt spekter av utdanning- og erfaringsbakgrunner (Østergaard et al., 2011).

Den eneste dimensjonen hvor analysen viser tegn til at respondenter med innovasjonsbakgrunn ikke scorer høyest er gjennomførbarhet. Regresjonsanalysen viser i dette tilfellet en marginalt bedre score blant respondenter med teknologibakgrunn. Det er knyttet stor usikkerhet til dette funnet, men en kan likevel tillate seg å spekulere i om respondenter med innovasjonsbakgrunn kan ha større tilbøyelighet til å overvurdere hvor gjennomførbar en idé er. Teknologibakgrunnen virker derimot å ha bedre evne til å forkaste ideer med lav gjennomførbarhet på et tidligere tidspunkt, og scorer dermed

høyest på *gjennomførbarhet*.

5.1.2 Arbeidsstatus

Ved å sammenligne resultatene blant respondenter med og uten arbeidserfaring kan vi si noe om hvordan *mengden* humankapital respondentene har eventuelt påvirker evne til å generere gode ideer, se tabell 4.5. Resultatene viser ingen signifikante forskjeller mellom de to kategoriene, og det understrekes derfor at det ikke er mulig hevde at arbeidsstatus spiller en vesentlig rolle for individers evne til å generere gode ideer. Det er likevel interessant å drøfte mulige årsaker til at respondenter med teknisk utdanning som oppgir å være arbeidstakere oppnår en høyere totalscore i gjennomsnitt. Ved å referere til utredningens teoretiske fundament, kan man i lys av Berndtsson et al. (2020) sitt rammeverk for datadrevene virksomheter peke på fordelene ved å være virksomhetens samlede analytiske verktøykasse. Det er likefullt viktig å påpeke at selv om selskaper har anledning til å utnytte bredden av sine analytiske ressurser til innovasjonsformål, er det til syvende og sist individene som må tenke ut løsninger og finne på nye ting. Mangelen på signifikante forskjeller mellom de to kategoriene er i så måte et forventet funn.

5.1.3 Lønn

Et tema som i denne utredningen ikke belyses i særlig grad handler om kompensasjon og avlønning i relasjon til innovasjonsevne. Vi valgte likevel å be respondentene oppgi sin brutto årsinntekt i datainnsamlingen, se tabell 4.6. Analysen viser signifikante resultater for at lønn over 400.000 kroner har en positiv effekt på dataintensiv idegenerering. Dette kan ha sammenheng med at blant respondentene i utvalget i denne kategorien finnes høyt etterspurt arbeidskraft, som for eksempel utviklere og konsulenter. Dette må antas ha sammenheng med at tekniske- og økonomiske utdanning- og erfaringsbakgrunner i mange tilfeller vil være aktuelle kandidater for disse stillingene. Funnet er også i tråd med forskning som viser at belønning basert på langsiktige resultater gir mennesker rom til å prøve og feile, og dermed gir en positiv effekt på innovasjon (Ederer og Manso, 2013). Skal man tro forskningen og funnene i undersøkelsen, vil det derfor kunne være en god strategi for virksomheter å belønne humankapital som bidrar til at selskapet lykkes med innovasjon, ettersom avlønning i kombinasjon med friheten til å utforske muligheter virker å ha en selvforsterkende positiv effekt på selskapets innovasjonsevne.

5.1.4 Ideer og tidsbruk

Modellen egner seg ikke til å si så mye om hva som eventuelt kan påvirke antall ideer respondentene i utvalget har avgitt i sine casebesvarelse, illustrert ved lav forklaringskraft og få signifikante resultater, se appendiks A8.1. Man ser imidlertid antydninger til at kvinnelige respondenter har flere ideer enn menn, hvilket ut i fra den øvrige analysen er vanskelig å finne en begrunnelse for. En av få signifikante sammenhenger finner vi blant respondenter med innovasjonsbakgrunn, som i snitt har flest ideer. Uten at vi legger for mye vekt på det aktuelle funnet, er det naturlig å påpeke at dette er i tråd med de øvrige resultatene,som viser at innovasjonsbakgrunn gir gode forutsetninger for idegenerering.

I likhet med funnene knyttet til antall ideer, kan tiden respondenten har brukt til å utarbeide sine casebesvarelser være interessante. Heller ikke her har analysen gitt signifikante resultater, hvilket vi mistenker har sammenheng med at datagrunnlaget er for snevert. Resultatene som ikke kan tillegges nevneverdig vekt indikerer imidlertid at kvinner bruker mer tid enn menn, og at yngre respondenter bruker mer tid enn de eldre respondentene.

5.2 Stor modell

I arbeidet med å besvare forskningsspørsmålet om hvilke demografiske trekk og humankapital som kan påvirke i idegenerering, forsøkte vi å utvikle en modell som kunne predikere hvilke profiler som ville være best egnet til å generere gode ideer. Modellen som forklares nærmere i analysedelen søker å predikere hvilken totalscore et individ er ventet å oppnå med sine ideer, basert på individets human kapital og demografi, se tabell 4.22.

Modellens forklaringskraft på 28% viser at det er en beskjeden andel av resultatene som kan forklares ved hjelp av den, og at modellen predikerer totalscoren svært unøyaktig, tyder på at mengden data ikke er tilstrekkelig, samtidig som variasjonen i dataene kunne vært høyere. En åpenbar svakhet er altså modellens avhengighet av data, som dermed må antas å være en medvirkende årsak til dens svake predikasjon. Et større og bredere utvalg blant respondentene ville bidratt til høyere forklaringskraft, og antagelig ville en utvidelse gjennom å legge til ytterligere variabler kunne påvirket modellens predikative evne. Svakheten ved modellen illustrerer godt et gjennomgangstema i utredningen, nemlig

tilgangen og variasjonen i datagrunnlaget.

Motsatt kan en argumentere for at modellen faktisk treffer med 75% av prediksjonene, og gir med det et godt utgangspunkt for å analysere indikasjonene en kan lese ut av modellen. I tillegg til å øke mengde og variasjon i datasettet kan en tenke seg at hverken forskningsmetodikk eller datainnsamling-strategien var utformet optimalt for modellens resultat. Mer tid til dypere utvikling av case og spørreundersøkelse og en annen praktisk gjennomføring av casebesvarelse kan tenkes at ville hatt positiv innvirkning på modellens styrke.

På tross av modellens svakheter ser vi indikasjoner som sammenfaller godt med masterutredningens øvrige funn. Ved å sammenstille resultatene fra analysen og diskusjonsmomentene beskrevet over, kan en tegne et bilde av en profil som vil ha gode forutsetninger for å score høyt langs de fire dimensjonene utredningen baserer datadrevet innovasjon på. Vi ser klare tendenser i de demografiske resultatene som tyder på at grad av utdannelse, alder og muligens erfaring spiller vesentlige roller for evnen til å generere ideer langs de fire dimensjonene. I tillegg viser de enkle regresjonene sammen med den multiple regresjonsmodellen at individer med bakgrunn og erfaring med innovasjon oppnår signifikante bedre score ved idegenerering.

Totalt sett gir ikke resultatene fra analysen et klart svar på masterutredningens overordnede problemstilling, om hvorvidt datadrevet innovasjon er en oppgave for økonomene eller teknologene. Funnene peker imidlertid i retning av at svaret er langt mer sammensatt og tvetydig enn hva problemstillingens utforming åpner for. Diskusjonen underbygger den underliggende antagelsen om at problemstillingen ikke kan besvares med en spesifikk stillingstittel eller utdanningsprofil og bidrar til å underbygge forskningen som viser at mangfold har en positiv påvirkning på innovasjon (Østergaard et al., 2011).

Istedenfor å peke på økonomen eller teknologen som den optimale innovatøren, viser resultatene at varianten av humankapital en person har tilegnet og opparbeidet seg ikke nødvendigvis er den viktigste determinatoren for evne til å innovere. Graden av humankapital i kombinasjon med ulike demografiske trekk virker derimot å være en langt bedre indikator for evne til å innovere, og vi kan oppsummere det hele ved å stadfeste at arbeidet med datadrevet innovasjon er sammensatt og komplisert og bør deles mellom ulike disipliner og fagfelt. I tillegg viser masterutredningen at det finnes utallige faktorer

som enkeltindividet ofte ikke kan påvirke, som spiller viktige roller for virksomheters innovasjonsevne.

6 Konklusjon

Formålet med masterutredningen har vært å undersøke og belyse hvilken påvirkning humankapital har på datadrevet innovasjon. Konklusjonen er basert på resultatene fra innsamlet data, i form av casebesvarelser og spørreundersøkelser.

Masterutredningens analyse viser tydelige tegn på at humankapital og demografiske trekk spiller en viktig rolle for å lykkes med dataintensiv innovasjon og dermed datadrevet verdiskaping. Høy utdannelse og relevant bakgrunn viser seg å være viktige faktorer for å predikere hvordan et individ vil prestere i innovasjonsarbeid. I tillegg til resultatene fra analysen har det teoretiske fundamentet belyst hvordan en helhetlig datastrategi og kompetanse til å behandle analyseverktøy er viktige forutsetninger for virksomheter som ønsker å bli mer datadrevet.

Funnene fra analysen vil kunne være av interesse for virksomheter som søker å bli mer datadrevet, enten gjennom rekruttering av kompetanse eller utvikling av medarbeidere. Særlig for virksomheter som opererer i skjæringspunktet mellom teknologi og forretning vil funnene kunne gi innsikt om hvordan de bør disponere sin analytiske verktøykasse. Manglende relevant eller feilallokert humankapital og for lite kunnskapsdeling i virksomheter er utfordringer som masterutredningen adresserer.

Analysen har i stor grad tatt for seg de demografiske variablene og sett på ulike mønstre i datasettet. Selv om noen funn er uklare har vi identifisert faktorer som gjennomgående gir høyere score langs dimensjonene. Under presenteres derfor de viktigste faktorene som besvarer utredningens forskningsspørsmål.

1. **Utdannelse** - Høyt utdanningsnivå scorer signifikant bedre på idegenerering. Vi ser også tegn til at tilleggsutdanning er en indikator for høy gjennomsnittscore.

2. **Kjønn** - Langs de ulike dimensjonene har menn den høyeste gjennomsnittscoren. Vi finner imidlertid ingen signifikante resultater som kan bekrefte om det finnes forskjeller mellom kjønnene. Kvinnene brukte i snitt mer tid på caseoppgaven og genererte i tillegg flere ideer.

3. **Lønn** - Individer med høy lønn scorer signifikant bedre på dimensjonene *kreativitet* og *datainntensivitet*. Det er imidlertid vanskelig å si hva som er kausaliteten, ettersom

variabelen sannsynligvis vil korrelere med flere av de andre kontrollvariablene. Det er dermed uklart hvilken effekt lønn har for individers evne til å arbeide med dataintensiv problemløsing.

4. **Erfaring** - Vi har ikke avdekket tegn til at arbeidstakere scorer bedre en studenter. Vi ser likevel at personer noen års erfaring scorer høyere, men vi kan ikke hevde at dette er en nødvendig forutsetning for å lykkes med datadrevet problemløsing.

5. **Alder** - Vi ser at den respondenter i den yngre kategorien scorer best på idegenerering. Om vi går dypere inn i datagrunnlaget, ser vi at aldersintervallet 30-39 har den høyeste gjennomsnittscoren.

For å besvare utredningens overordnede problemstillinger, tar vi for oss den andre delen av forskningsspørsmålet som baserer seg på humankapital, og dermed de ulike bakgrunnene vi har definert og diskutert i oppgaven. Vi presenterer de viktigste funnene for hver av de fire kategoriene, slik at vi bedre kan forstå om datadrevet innovasjon er en oppgave for økonomen eller teknologen.

1. **Innovasjonsbakgrunn** - Respondenter med innovasjonsbakgrunn scorer bedre enn de andre bakgrunnene på *kreativitet*, *kommersialisering*, *dataintensivitet*, samt på *totalscoren*. Innovasjonsprofilen brukte i snitt lengst tid på å gjennomføre caset og kom i snitt med flest ideer.

2. **Teknologibakgrunn** - Respondentene scoret best på dimensjonen *gjennomførbarhet*, men dårligere en respondentene med innovasjonsbakgrunn på de øvrige dimensjonene. Respondentene med teknologibakgrunn brukte i snitt minst tid på å gjennomføre caset og kom i snitt med færrest ideer.

3. **Forretningsbakgrunn** - Respondentene med forretningsbakgrunn scoret dårligere enn respondentene med innovasjonsbakgrunn på samtlige dimensjoner, men scoret bedre enn teknologibakgrunn på *kommersialisering* og *datainntensivitet*.

4. **Ingen relevant bakgrunn** - Respondentene uten relevant bakgrunn scoret statistisk signifikant dårligere enn respondenter med en relevant bakgrunn på samtlige dimensjoner.

Svaret på masterutredningens problemstilling og forskningsspørsmål er med andre ord

kompleks og sammensatt, og resultatene fra analysen trekker i retning av at datadrevet innovasjon er en oppgave for *både* økonomene og teknologene - antageligvis i tverrfaglige samarbeid resten av organisasjonen.

6.1 Begrensninger

Masterutredningen søker å belyse et sammensatt og komplekst tema som er aktuelt både i dag og i fremtiden. Vi erkjenner at problemstillingene knyttet til innovasjon og bruk av data krever tverrfaglig innsikt og kunnskap, og at det derfor vil være aspekter ved utredningen som ikke belyses i tilstrekkelig grad. Videre er vi innforstått med at en masterutredning på langt nær vil kunne gi utfyllende svar og konklusjoner, men isteden vil være et bidrag og forhåpentligvis en katalysator for videre forskning.

Regulatoriske rammeverk som GDPR innebærer utfordringer for alle virksomheter, både i relasjon til interne rutiner og prosesser, behandling av personsensitiv data, og i utviklingsprosesser. En begrensning med masterutredningen i så måte er valget om å utelate personvernproblematikken fra både caseoppgaven og fra det teoretiske grunnlaget. Dette har vi gjort for å rette fokuset på mulighetsrommet, og la undersøkelsen i størst mulig grad handle om innovasjonsfaktorer som virksomheter selv kan påvirke.

Underveis i arbeidet med masterutredningen har vi hatt praktiske utfordringer knyttet til forskninsdesignet, caset og koronasituasjonen. På grunn av lokale restriksjoner og nasjonale anbefalinger gjennomførte vi all datainnsamling gjennom en Qualtrics undersøkelse. Dette skapte utfordringer knyttet til hvilken grad vi kunne kontrollere forskningsomgivelsene. Vi fikk tilbakemeldinger på at caset var vanskelig for noen og vi har sett i Qualtrics-rapporten at det er et stort variasjon i hvor lang tid som ble benyttet. Ved å gjennomføre studien fra hjemmekontoret har vi derfor opplevd begrensninger i forhold til fleksibilitet. Vi kunne hatt mer rom for å skape diskusjoner, stille oppfølgingspørsmål, og få tilbakemeldinger under datainnsamling og ved ekspertpanelets vurdering av casebesvarelsene om dette ble gjennomført fysisk.

I arbeidet med datasettet har vi underveis identifisert ulike variabler som kunne vært beriket utredningen, men som ikke var en del av den opprinnelige spørreundersøkelsen. Spesielt variabler som omhandler ledererfaring og kompetanse med teknologi, samt variabler som kartlegger i hvilken grad respondenten jobber med innovasjon til daglig. De indentifiserte

variablene som ikke ble inkludert i masterutredningen har likevel indirekte vært en del av våre vurderinger, og foreslås som interessante elementer i den videre forskningen.

6.2 Videre forskning

Gjennom funnene i denne oppgaven har vi identifisert flere demografiske variabler som kan påvirke hvordan individer presterer på idégenerering og datadrevet innovasjon. Videre har vi observert at individer som scorer høyt langs de definerte dimensjonene kan inneha svært ulik bakgrunn, erfaringsprofil og utdannelse.

I faktoranalysen identifiserte vi tre ulike faktorer som omhandler psykologiske dimensjoner innen kreativitet, gjennomførbarhet og kommersialisering. Fra denne analysen fant vi ingen konkrete resultater og utredningen fikk derfor et mindre fokus på hvordan psykologiske faktorer påvirker evne til datadrevet innovasjon. For videre forskning ser vi på dette som et svært interessant område. En utvidelse av datainnsamling rundt psykologiske spørsmål og et større teoretisk grunnlag kunne vært et startpunkt for å undersøke det det psykologiske aspektet nærmere.

Selv om det er innovasjonsprosessen starter med enkeltindivider, foregår prosessen naturlig som en del av en organisasjon og ofte i sammensatte team. Derfor er teamsammensetting og perspektiver rundt smidige arbeidsmetoder, innovasjonskultur og organisasjonsstruktur spesielt interessant for videre forskning. Vi ser på dette som et viktig område for å få en større forståelse for datadrevet innovasjon og hvordan disse prosessene foregår i en bedrift.

Referanser

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, 16(1):3–9.

Al-Laham, A., Tzabbar, D., og Amburgey, T. L. (2011). The dynamics of knowledge stocks and knowledge flows: innovation consequences of recruitment and collaboration in biotech. *Industrial and Corporate Change*, 20(2):555–583.

Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in organizational behavior*, 10(1):123–167.

Amit, R. og Zott, C. (2012). Creating value through business model innovation. *2012*.

Askheim, O. G. A. og Grenness, T. (2008). *Kvalitative metoder for markedsføring og organisasjonsfag*. Universitetsforl.

Barua, A., Mani, D., og Mukherjee, R. (2012). Measuring the business impacts of effective data. *Report accessed at http://www. sybase. com/files/White_ Papers on Sep*, 15:2012.

Befring, E. (2007). *Forskningsmetode med etikk og statistikk*. Samlaget.

Berndtsson, M., Forsberg, D., Stein, D., og Svahn, T. (2018). Becoming a data-driven organisation.

Berndtsson, M., Lennerholt, C., Svahn, T., og Larsson, P. (2020). 13 organizations' attempts to become data-driven. *International Journal of Business Intelligence Research (IJBIR)*, 11(1):1–21.

Brynjolfsson, E., Hitt, L. M., og Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN 1819486*.

Bulger, M., Taylor, G., og Schroeder, R. (2014). Data-driven business models: challenges and opportunities of big data. *Oxford Internet Institute. Research Councils UK: NEMODE, New Economic Models in the Digital Economy*.

Cavanillas, J. M., Curry, E., og Wahlster, W. (2016). *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe*. Springer Nature.

Chesbrough, H. (2011). *Open services innovation: Rethinking your business to grow and compete in a new era*. John Wiley & Sons.

Cho, H.-J. og Pucik, V. (2005). Relationship between innovativeness, quality, growth, profitability, and market value. *Strategic management journal*, 26(6):555–575.

Christensen, C. M., Bartman, T., og Bever, D. v. (2016a). The hard truth about business model innovation.

Christensen, C. M., Hall, T., Dillon, K., og Duncan, D. S. (2016b). Know your customers' jobs to be done. *Harvard Business Review*, 94(9):54–62.

Clifford, C. (2018). Life with a.i.: Google ceo: A.i. is more important than fire or electricity. https://www.cnbc.com/2018/02/01/google-ceo-sundar-pichai-ai-is-more-important-than-fire-electricity.html. Lastet ned: 01.12.2021.

Cobbenhagen, J. (2000). *Successful innovation: towards a new theory for the management of small and medium sized enterprises*. Edward Elgar Publishing.

Cricelli, L. og Grimaldi, M. (2008). A dynamic view of knowledge and information: a stock and flow based methodology. *International Journal of Management and Decision Making*, 9(6):686–698.

Crié, D. og Micheaux, A. (2006). From customer data to value: What is lacking in the information chain? *Journal of Database Marketing & Customer Strategy Management*, 13(4):282–299.

Cronholm, S., Göbel, H., og Rittgen, P. (2017). Challenges concerning data-driven innovation. I *The 28th Australasian Conference on Information Systems, Hobart Australia, December 4-6, 2017.*

D'Aveni, R. A., Ravenscraft, D. J., og Anderson, P. (2004). From corporate strategy to business-level advantage: Relatedness as resource congruence. *Managerial and Decision Economics*, 25(6-7):365–381.

Davenport, T. H. (1993). *Process innovation: reengineering work through information technology*. Harvard Business Press.

Davenport, T. H. (2014). How strategists use "big data" to support internal business decisions, discovery and production. *Strategy & Leadership*.

De Vaus, D. A. (2002). Surveys in social research. wyd.

DeepAI (2021). What is a neural network? https://deepai.org/machine-learning-glossary-and-terms/neural-network. Lastet ned: 05.11.2021.

Ederer, F. og Manso, G. (2013). Is pay for performance detrimental to innovation? *Management Science*, 59(7):1496–1513.

Fagerberg, J., Mowery, D. C., Nelson, R. R., et al. (2005). *The Oxford handbook of innovation*. Oxford university press.

Friborg, O. (2011). Hva er faktoranalyse?

Gall, M., Gall, J., og Borg, W. (2007). *Educational Research*. Pearson education.

Gambus, P. og Shafer, S. L. (2018). Artificial intelligence for everyone. *Anesthesiology*, 128(3):431–433.

Gandomi, A. og Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2):137–144.

Gardner, H. (1993). Creating minds: An anatomy of creativity seen through the lives of freud. *Einstein. Picasso. Stravinsky. Eliot. Graham, and Gandhi. New York: Basic Books*.

Gareth, J., Daniela, W., Trevor, H., og Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.

Garmaki, M., Boughzala, I., Wamba, S. F., et al. (2016). The effect of big data analytics capability on firm performance. I *PACIS*, side 301.

George, G., Haas, M. R., og Pentland, A. (2014). Big data and management.

Ghauri, P., Gronhaug, K., og Kristinslund, I. (2010). Research methods in business studies. 4. painos. *Essex: Pearson Education Limited*, sider 977–979.

Gjelsvik, M. (2013). *Innovasjonsledelse: ledelse av innovasjon og internt entreprenørskap*. Fagbokforlaget Vigmostad Bjørke.

Gleeson, B. (2013). The silo mentality: How to break down the barriers. https://www.forbes.com/sites/brentgleeson/2013/10/02/the-silo-mentality-how-to-break-down-the-barriers/?sh=3b1206e38c7e. Lastet ned: 10.12.2021.

Gripsrud, G., Olsson, U. H., og Silkoset, R. (2004). Metode og dataanalyse–med fokus på beslutninger i bedrifter.

Haanæs, K. (1999). Innovasjon som strategisk utfordring. *Magma*, 3:1999.

Huang, C.-K., Wang, T., og Huang, T.-Y. (2020). Initial evidence on the impact of big data implementation on firm performance. *Information Systems Frontiers*, 22(2):475–487.

Iden, J., Andestad, M., og Grung-Olsen, H.-C. (2013). Prosessledelse og innovasjon: en litteraturstudie.

Isaksen, S. G., Dorval, K. B., og Treffinger, D. J. (2010). *Creative approaches to problem solving: A framework for innovation and change*. Sage Publications.

Jacobsen, D. og Thorsvik, J. (2013). Hvordan organisasjoner fungerer.[how organizations functions].(4. utg).

Jacobsen, D. I. (2005). Hvordan gjennomføre undersøkelser? innføring i samfunnsvitenskapelige metode.(utgave 2). kristiansand: Høyskoleforlaget as.

Jetzek, T., Avital, M., og Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of theoretical and applied electronic commerce research*, 9(2):100–120.

Johannessen, A., Tufte, P. A., og Christoffersen, L. (2010). *Introduksjon til samfunnsvitenskapelig metode*, volume 4. Abstrakt Oslo.

Joyce, C. K. (2009). *The blank page: effects of constraint on creativity*. University of California, Berkeley.

João, N. (2014). Factor analysis. http://www.di.fc.ul.pt/~jpn/r/factoranalysis/factoranalysis.html. Accessed: 2021-11-25.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1):31–36.

Kaufman, J. C. og Sternberg, R. J. (2010). *The Cambridge handbook of creativity*. Cambridge University Press.

Kleven, T. (2002). *Ikke-eksperimentelle design*. Unipub forlag.

Knudsen, E. S. og Lien, L. B. (2019). Hitting the gas or the brake? recessions and firms' knowledge investments. *Managerial and Decision Economics*, 40(8):1000–1015.

Kolbjørnsrud, V., Amico, R., og Thomas, R. J. (2016a). How artificial intelligence will redefine management. *Harvard Business Review*, 2:1–6.

Kolbjørnsrud, V., Amico, R., og Thomas, R. J. (2016b). The promise of artificial intelligence. *Accenture: Dublin, Ireland*.

Kwon, O., Lee, N., og Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management*, 34(3):387–394.

Lund, T. (2002). Generaliseringsproblematikk. i t. *Lund (red.), Innføring i forskningsmetodologi*, sider 125–140.

Malerba, F. (2005). Sectoral systems of innovation: a framework for linking innovation to the knowledge base, structure and dynamics of sectors. *Economics of innovation and New Technology*, 14(1-2):63–82.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A., et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Markides, C. og Charitou, C. D. (2004). Competing with dual business models: A contingency approach. *Academy of Management Perspectives*, 18(3):22–36.

Markus, M. L. (2015). New games, new rules, new scoreboards: the potential consequences of big data. *Journal of Information Technology*, 30(1):58–59.

Marr, B. (2018). How much data do we create every day? the mind-blowing stats everyone should read.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., og Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10):60–68.

Menon (2019). Verdiskapning med data. https://www.menon.no/wp-content/uploads/2019-88-Verdiskaping-med-data.pdf. Lastet ned: 12.09.2021.

Moreau, C. P. og Dahl, D. W. (2005). Designing the solution: The impact of constraints on consumers' creativity. *Journal of Consumer Research*, 32(1):13–22.

Mumford, M. D. (2000). Managing creative people: Strategies and tactics for innovation. *Human resource management review*, 10(3):313–351.

Oates, B. J. (2005). *Researching information systems and computing*. Sage.

OECD (2005). Guidelines for collecting and interpretting innovation data. *The Measurement of Scientific and Technological Activities*.

OECD (2007). Human capital: How what you know shapes your life. *OECD Insights*.

OECD (2017). *OECD Digital Economy Outlook 2017*.

Oldham, G. R. og Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of management journal*, 39(3):607–634.

O'Leary, D. E. (2013). Artificial intelligence and big data. *IEEE intelligent systems*, 28(2):96–99.

Østergaard, C. R., Timmermans, B., og Kristinsson, K. (2011). Does a different view create something new? the effect of employee diversity on innovation. *Research policy*, 40(3):500–509.

Pedersen, P. E. og Nysveen, H. (2010). Service innovation challenges at the policy, industry, and firm level: A qualitative enquiry into the service innovation system.

Pennock, M. (2007). Digital curation: A life-cycle approach to managing and preserving usable digital information. *Library & Archives*, 1(1):1–3.

Perrey, J., S.-D. . U. A. (2015). Smart analytics: How marketing drives shortterm and long-term growth. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Marketing%20and%20Sales/Our%20Insights/EBook%20Big%20data%20analytics%20and%20the%20future%20of%20marketing%20sales/Big-Data-eBook.ashx. Lastet ned: 25.10.2021.

Ponelis, S. R. (2015). Using interpretive qualitative case studies for exploratory research in doctoral studies: A case of information systems research in small and medium enterprises. *International Journal of Doctoral Studies*, 10(1):535–550.

Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. Simon and Schuster.

Power, D. J. (2008). Understanding data-driven decision support systems. *Information Systems Management*, 25(2):149–154.

Provost, F. og Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1):51–59.

Rainie, L. og Anderson, J. (2017). The future of jobs and jobs training. *Pew Research Center*.

Ransbotham, S., Kiron, D., og Prentice, P. K. (2015). *Minding the analytics gap*, volume 56. MIT Sloan Management Review.

Reitan, J., S.-T. F. P. N. O. H. K. . R. M. (2011). Behovsdrevet innovasjon. 10 steg til innovasjon i helsesektoren. *SINTEF, InnoMed, versjon 1*.

Ringdal, K. (2007). Enhet og mangfold. fagbokforlaget.

Rowley, J. (2007). The wisdom hierarchy: representations of the dikw hierarchy. *Journal of information science*, 33(2):163–180.

Russom, P. et al. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, 19(4):1–34.

Salvanes, K. G. (2014). Humankapital og omstilling?

Samuel, A. L. (1967). Some studies in machine learning using the game of checkers. ii—recent progress. *IBM Journal of research and development*, 11(6):601–617.

Sathi, A. (2011). *Customer experience analytics: The key to real-time, adaptive customer relationships*. MC Press.

Saunders, M., Lewis, P., og Thornhill, A. (2016). *Research methods for business students (5th edition)*. Pearson education.

Schumpeter, J. A. et al. (1939). *Business cycles*, volume 1. McGraw-Hill New York.

SeedScientific (2021). How much data is created every day. https://seedscientific.com/how-much-data-is-created-every-day/. Accessed: 2021-09-02.

Silver, M. S. (1991). *Systems that support decision makers: description and analysis*. John Wiley & Sons, Inc.

SNL (2019). Stordata. https://snl.no/stordata. Lastet ned: 10.12.2021.

SNL (2020). Kunstig intelligens. https://snl.no/kunstig_intelligens. Lastet ned: 18.09.2021.

Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. Cambridge University Press.

Thagaard, T. (2009). *Systematikk og innlevelse: en innføring i kvalitativ metode*, volume 3. Fagbokforlaget Bergen.

Tucker, R. B. (2002). *Driving growth through innovation: How leading firms are transforming their futures*. Berrett-Koehler Publishers.

Turban, E., Sharda, R., og Delen, D. (2010). Decision support and business intelligence systems (required). *Google Scholar*.

Venables, W. N., Smith, D. M., Team, R. D. C., et al. (2009). An introduction to r.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., og Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165:234–246.

Yin, R. (2014). Case study research: Design and methods 5. dallas: edn.

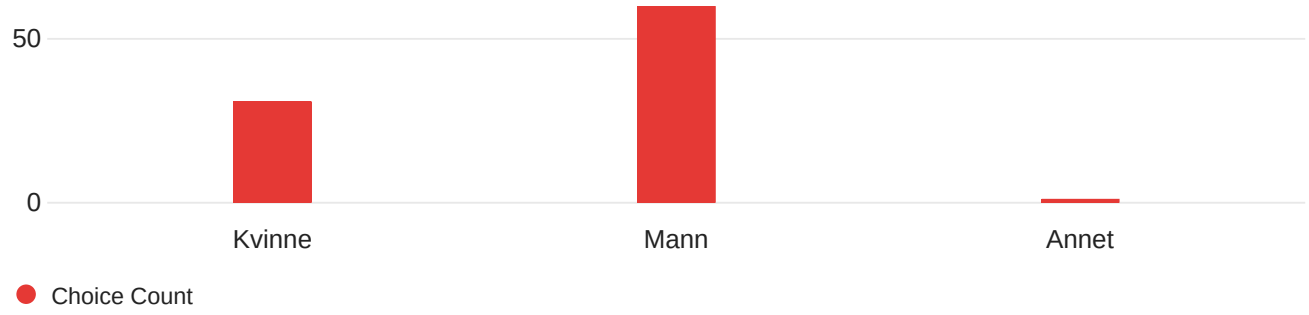
Appendiks

A1 Qualtrics rapport

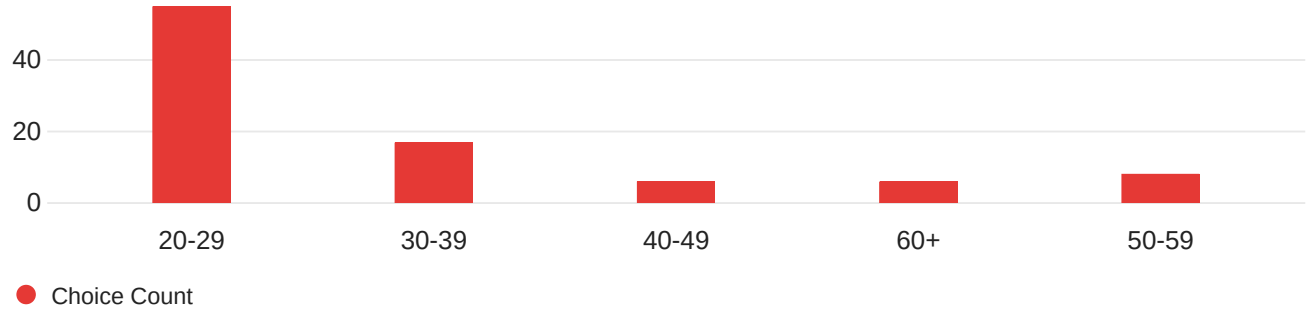
Q2_1 - Estimert tidsbruk

Field	Min	Max	Mean	Standard Deviation	Variance	Responses	Sum
Estimert tidsbruk	0.00	60.00	14.47	9.84	96.81	92	1331.00

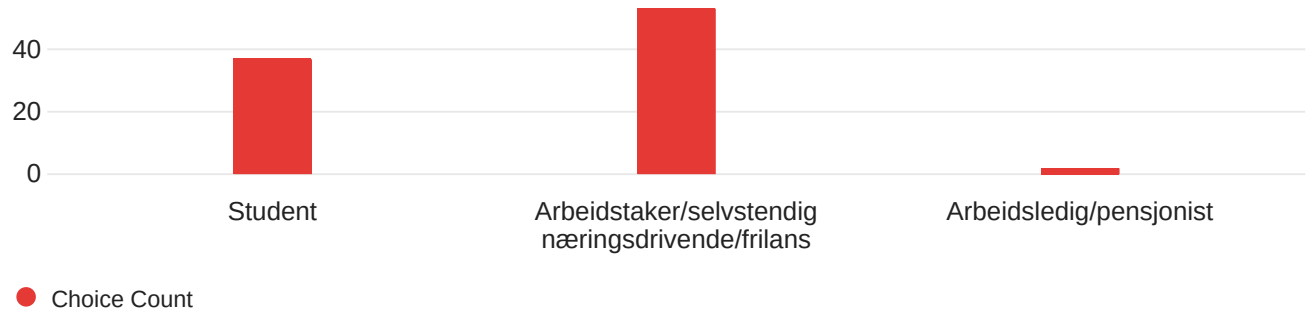
Q3 - Kjønn:



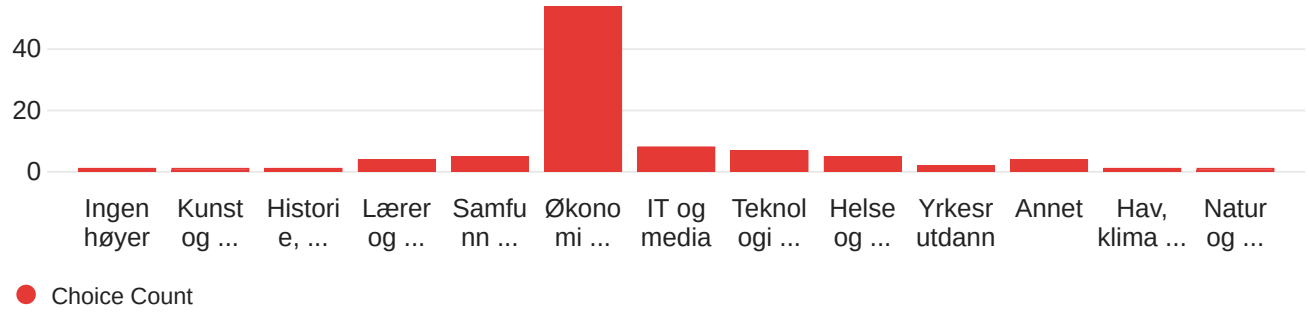
Q4 - Alder:



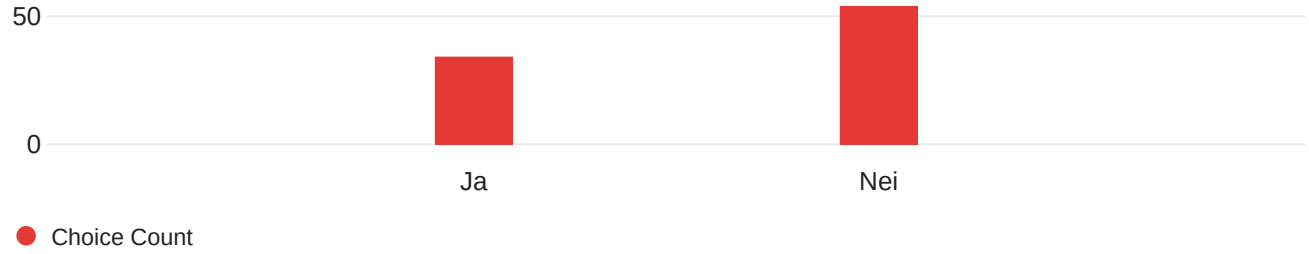
Q5 - Arbeidsstatus (velg kategorien som passer best):



Q6 - Velg retningen du mener passer best basert på din utdannelse.

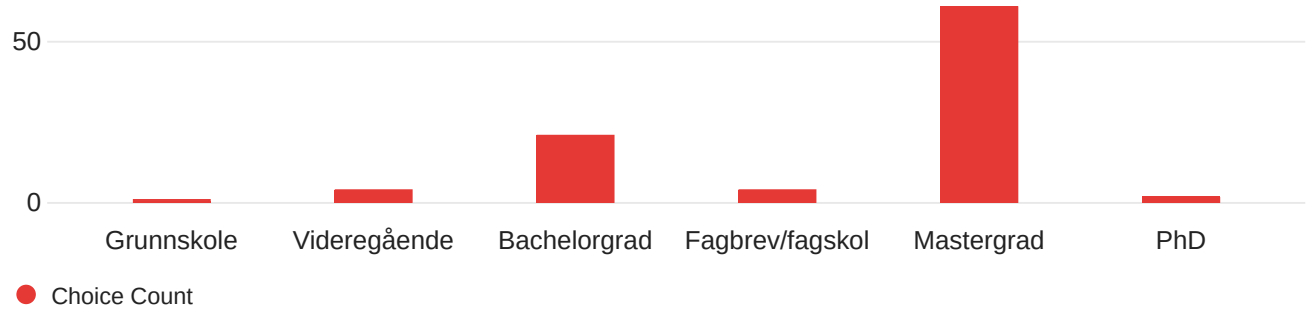


Q7 - Har du tilleggsutdanning innenfor samme eller andre felt enn det du oppga i forrige spørsmål?

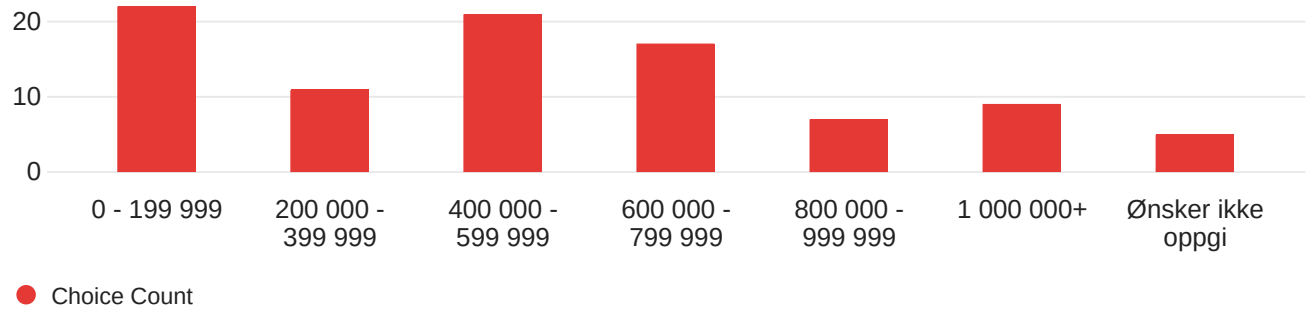


Q8 - Kryss av hvilke retninger du har tilleggsutdannelse innen. - Selected Choice

Q9 - Utdanningsnivå (velg høyeste påbegynte):



Q10 - Årslønn(før skatt):



Q13-Q15 - Ta stilling til de ulike påstandene og vurder utifra din egen arbeidserfari...

Field	Min	Max	Mean	Standard Deviation	Variance	Responses	Sum
Jeg har mye erfaring med innovasjonsarbeid	1.00	7.00	4.27	1.85	3.41	51	218.00
Jeg har mye erfaring med teknologi	1.00	7.00	4.56	1.76	3.09	52	237.00
Jeg har mye forretningserfaring	1.00	7.00	4.65	1.82	3.33	51	237.00

Q16-Q25 - Her kommer noen påstander hvor vi er ute etter dine subjektive meninger. Sv...

Field	Min	Max	Mean	Standard Deviation	Variance	Responses	Sum

Jeg synes kreativ tenking og ideutvikling er gøy	2.00	9.00	7.13	2.20	4.83	92	656.00
Jeg er mer iderik enn snittet på min arbeidsplass	1.00	9.00	5.97	2.18	4.77	92	549.00
Jeg er blant de første som tar i bruk ny teknologi i min omgangskrets	1.00	9.00	4.71	2.33	5.45	92	433.00
Blant kollegaer/studievenner er jeg ofte den som fokuserer på progresjon og gjennomføring i prosjektarbeid	1.00	9.00	6.48	2.08	4.32	89	577.00
Jeg vegrer meg for å gå i gang med prosjekter hvor resultatet er uvisst	1.00	9.00	3.26	1.75	3.05	89	290.00
I prosjektarbeid vurderer jeg alltid hvilket resultat man kan vente seg i andre enden	2.00	9.00	6.21	2.05	4.19	89	553.00
Potensiell verdiskaping er en viktig faktor når jeg evaluerer en ny ide	1.00	9.00	6.84	2.21	4.90	89	609.00
Det spiller ikke så stor rolle for meg om et prosjekt har lavt inntjeningspotensiale, så lenge ideen er god	1.00	9.00	4.10	2.29	5.26	89	365.00
Jeg har en realistisk tilnærming til nye ideer og prosjekter	1.00	9.00	6.90	2.07	4.29	89	614.00

Q26-Q29 - Hvordan vekter du viktigheten av disse tre perspektivene når du vurderer en...

Field	Min	Max	Mean	Standard Deviation	Variance	Responses	Sum
Hvor kreativ ideen er	9.00	20.00	16.98	2.94	8.66	88	1494.00
Hvor gjennomførbar ideen er	10.00	20.00	19.00	1.54	2.36	89	1691.00
Hvor stort kommersielt potensiale ideen har	9.00	20.00	18.50	2.36	5.59	88	1628.00

A2 Case

Caseoppgaven

Introduksjon

Denne spørreundersøkelsen handler om datadrevet innovasjon. Under kan du lese litt mer om datadrevet innovasjon og hvordan vi har definert det.

Når man ser på data og innovasjon sammen og hvordan data kan brukes til å skape innovasjon snakker vi om datadrevet innovasjon. Fremover er det ventet at datadrevet innovasjon vil være en viktig driver for økonomisk vekst, både for samfunnet, selskaper og forbrukere. Sammensmeltning av flere trender sammen med en økende migrasjon av sosioøkonomiske aktiviteter til internett og reduksjon i kostnader ved datainnsamling, lagring og behandling fører til økende tilgang og bruk av store datamengder. Slike datamengder blir ofte omtalt som “big data” og har potensiale til å muliggjøre nye næringer, prosesser og produkter. Innovasjoner som baserer seg på innsikt og bruk av slike data forstås dermed som **datadrevet innovasjon.**

I den første delen av denne undersøkelsen blir du bedt om å besvare en praktisk caseoppgave hvor vi ønsker dine ideer. Du kan bruke den tiden du trenger, men ca. 20 minutters tenketid anses for å være tilstrekkelig.

Andre del av undersøkelsen er en spørreundersøkelse som kartlegger ulike demografiske variabler, samt dine subjektive meninger om kreativitet. Denne delen tar ca. 4 minutter å gjennomføre.

Om du ønsker kan du til slutt oppgi din epostadresse for å være med i trekningen av et gavekort på Eplehuset til 2000,- kroner. Vi trekker to vinnere.

Undersøkelsen er anonym, og casebesvarelsen kan ikke knyttes til deg.

Vi takker for deltakelsen.

Del 1 - Databasert caseoppgave

NB: Caset er et tenkt eksempel, ideene som måtte fremkomme vil ikke bli brukt til annet enn forskningsformål. Caset gjøres ikke på oppdrag fra et selskap eller andre.

Du jobber med innovasjon i et norsk forsikringsselskap. Forsikringsselskapet har en stor kundeportefølje og dermed også store mengder data. For å nå opp i konkurransen i bransjen og fra nye aktører trenger forsikringsselskapet hjelp med å komme på mulige bruksområder for dataen de har tilgjengelig. Selskapet er helt i startfasen og har ikke klart for seg hvilket forretningsområde de ønsker å satse på. Det eneste de vet er at de **ønsker å anvende dataen til å utvikle nye produkter og tjenester mot privatmarkedet.**

Under finner du eksempler på ulike type data forsikringsselskapet har tilgang til:

Persondata:
For eksempel data om kundenes alder, kjønn og adresse, samt inntekt, formue og gjeld.

Kunde- og skadehistorikk:
For eksempel data om kundeforhold, skadehistorikk, og utbetalinger.

Bevegelse og helseinformasjon:
For eksempel GPS-data, vitale tegn, aktivitetsnivå, søvndata fra smarte enheter (telefon, klokke, andre wearables). Ved tegning av helseforsikring er det vanlig å måtte gjennomføre en helseattest.

Sensordata:
Data som for eksempel temperatur og fukt fra sensorer montert i hus, bil og båt. Data i sanntid fra eksempelvis røykvarslere, fuktmålere og alarmsystemer kan bidra til å monitorere, og dermed forutse og forhindre skader.

Data fra samarbeidspartnere: Innsikt og data fra samarbeidspartnere, som for eksempel fagforeninger og industrikunder sine medlemmer, som kan gi forsikringsselskapet mulighet til å utvikle løsninger med utgangspunkt i data, skreddersydd til medlemmenes/kundenes ønsker og behov.

I tillegg til data som selskapet har tilgang til i dag, finnes det mye data som kunden frivillig kan velge å dele med forsikringsselskapet. Du står fritt til å foreslå andre typer datakilder enn de som er nevnt.

Oppgave

Med utgangspunkt i informasjonen over, **ber vi deg komme med konkrete forslag til nye og innovative produkter og tjenester relatert til forsikring for privatmarkedet, som tar utgangspunkt i nevnte data forsikringsselskap enten har tilgjengelig i dag, eller som det er plausibelt at de kan skaffe seg i fremtiden.**

Du kan bruke ta den tiden du trenger, men ca. 20 minutters tenketid anses for å være tilstrekkelig. Beskrivelsen trenger ikke være lang, men prøv å beskriv ideen på en måte som gjør det mulig for andre å forstå hva du tenker. Du trenger ikke ta hensyn til personvern.

Vennligst beskriv din ide/dine ideer i boksen under.

Oppgave Med utgangspunkt i informasjonen over, ber vi deg komme med konkrete forslag til nye og innovative produkter og tjenester relatert til forsikring for privatmarkedet, som tar utgangspunkt i nevnte data forsikringsselskap enten har tilgjengelig i dag, eller som det er plausibelt at de kan skaffe seg i fremtiden. Du kan bruke ta den tiden du trenger, men ca. 20 minutters tenketid anses for å være tilstrekkelig. Beskrivelsen trenger ikke være lang, men prøv å beskriv ideen på en måte som gjør det mulig for andre å forstå hva du tenker. Du trenger ikke ta hensyn til personvern. Vennligst beskriv din ide/dine ideer i boksen under.

Kommer ikke på noen gode forslag

Har ingen ideer til nye forsikringsprodukter. Men kanskje de kan selge informasjonen om hvem som eier båt og hytter for eksempel, til andre som selger produkter knyttet til båter og hytter.

Bedre vilkår på boligforsikring dersom det er registrert enheter for å avdekke farer (fuktmålere, brannvarslere etter gjeldende anbefalinger..). Kan også gjelde innboforsikring. Tilsvarende med hytte, bil, båt - lavere forsikringspremie dersom eier har installert alarmer osv. for å avdekke farer og skade. Kan fastsette en høyere forsikringspremie på bil dersom sensordata oppfatter at kjører opptrer uaktsomt i trafikken (og motsatt). Kjørelengden kan utgjøre grunnlaget for pris på bilforsikring eller summen som dekkes i en forsikringssak. Ved å dele helsedata fra klokker og wearables kan man oppnå billigere forsikring om man trener, konsumerer sunnere mat og følgelig får en sunnere livsstil. Stiller dog spørsmålstegn ved det etiske aspektet til en slik tjeneste! Reiseforsikring basert på hyppighet av reiser, og kanskje til hvilke land man reiser. Høyere forsikringspremie ved reise til mer kriminelle destinasjoner.

Forsikringselskapet kan gi eksisterende- og potensielle kunder en mulighet til å dele sin vaksine-journal fra "helseregisteret", og utifra hvilke vaksiner vedkommende har kan forsikringsselskapet skreddersy helseforsikring ved å belønne personer som er registrert med div. vaksiner (det være seg covid-vaksine eller andre kritiske vaksiner).

Produkter retter mot å imøtekomme stadig økende krav til bærekraft. Bruke overnevnte data for skreddersydde produkter rettet mot forebygging av skader- samt helserelatert forebygging. Utvikle tjenester innenfor forsikringsselskapets eierskap både i forebygging av potensielle skader og uehelse- samt i oppfølgingen av ev. skade og uehelse (for slik å bedre dokumentere for kunden bærekraft i alle ledd- samt tilby brukervennlige og enkle tjenester som er viktig i konkurranse med nye aktører). Feks mtp helse- produkter rettet mot helsefremming. For å skreddersy- bruke persondata, bevegelses- og helseinformasjon, samt vil kunder muligens være villige til å oppgi mer informasjon om sin livsstil/ livsbetingelser - utover informasjon som er vanlig å innhente i helseattest -som er rettet mye mot historikk og uhelse. Utvikle/ tilby produkter som feks tilgang til app (søvn, psykisk helse, trening, ernæring osv), lavterskel tilgang på helsefremmende tjenester (før uhelse har oppstått). Tilsvarende tenkning innenfor skadeforebygging- tilby produkter rettet mot å forebygge-ev. fange opp før skaden skjer/ blir alvorlig. Slike produkter vil også kunne gi tilgang på mer data (som sensorer etc).

1. at man innstallerer sensorer i hus for å lettere oppdage feks fuktskader

Som forsikringsselskap: Ny teknologi og nye datainsamlingsmetoder bør bidra til at kunder får redusert risiko for ulykker samtidig som det senker kost for forsikringselskapet. All typer data beskrevet over vil kunne bidra til dette. Redusert forsikringspremie bør da gis til kunder som bidrar med mer data. For eksempel bør redusert premie gis til kunder som har installert fuktmålere og andre sensorer hjemme.

Min ide til forsikringselskapet er å bruke dataen til å prisdiskriminere på innboforsikring. Jeg tenker da at en mulighet er å bruke sensordataen i huset til å måle ting som fukt nivå og antall ganger alarmer går av, til å vurdere hver enkelt husholdning. Hus hvor det er målt høy fuktighet over lang tid vil få høyere priser da de har høyere sannsynlighet råte skader i huset. Denne ideen om å prisdiskriminere basert på risiko hos kunden kan da også benyttes hos andre forsikringer. Dette kan være dersom forsikringsselskapet tilbyr bilforsikring eller lignende og de har data på kundens skadehistorikk.

Billigere forsikring på bakgrunn av hva du tillater forsikringsselskapet å kontrollere. Mer data om bruker og for eksempel personens hjem kan tillate forsikringsselskapet å ta større risiko.

1. Grønne forsikringsprodukter Jo mindre strøm og energi du bruker jo biligere forsikring får du. Dersom du velger grønne valg ofm med hus bil etc.. vil du kunne forsikre disse produktene biligere. 2. En transparent app som gir deg bonus/minus poeng ut ifra adferd. En samlet app. Dersom du får bankkortet ditt stjålet på ferie, så får du minus poeng. Dersom du ikke har hatt noen ulykker/skader på lang tid så får du pluss poeng. Folk elsker generelt når livet blir et spill man kan vinne. Dersom forsikringsselskapene er transparente om kundene sin profil, vil kundene ha større insentiver til å følge opp kriteriene som gir poeng. Eksempelvis, dersom jeg visste at jeg fikk poeng for hver måned som jeg klarer å bevise at jeg klarer å ta vare på ting, da hadde jeg kanskje vært enda mer forsiktig. I hvert fall hvis jeg fikk vite at forsikringen min kan bli billigere med xx antall kroner dersom jeg oppnår en viss poengsum. 3. En app som gir deg tips om hvordan man kan få billigere forsikring. Eks. "Dersom du bytter låsen på døren din med en xxxx lås, kan du spare opptil xx kr per måned" "Vi anbefaler deg å bytte bildekk. Dersom du gjør dette kan vi tilby deg en redusert pris på bilforsikringen din i fem måneder" Dette vil gi folk insentiver til å handle forsvarlig, og å sikre seg for ting som kan føre til farer.

En sensor som måler gjennomstrøm av vann i utvendige avløpsrør på gamle bygg, og dersom et rør begynner å tette seg, varsler den (f eks borettslaget) om dårlig/svak vannstrøm. Dette kan forhindre store skader ved store nedbørsmengder

Boligforsikring med progressiv bærekrafts-rabatt basert på hvor smart boligen din er. Helseforsikring med prisintervaller basert på hvor mye du trener. Boliglån basert på hvor godt du er pensjonsforsikret.

Det kan foretas en analyse av hvor stor risiko kundens daglige arbeid innebærer. For eksempel de som jobber i risikoutsatte grupper som politi, brannvesen, forsvar. For en brannmann kan det være aktuelt å medregne hva forsikringen skal dekke av senskader som følge av innånding av giftige gasser/røyk fra brann. Det er gjerne ikke lett å påvise skader underveis, men kan oppstå helseproblematikk senere i livet. Hvordan denne typen skader kan dekkes av forsikringstjenester, kan være aktuelt.

Med bakgrunn i all infoen forsikringsselskapet har i dag, vil det være naturlig å anta at de har en relativt stor oversikt over hvilke forskjellige type enheter og eiendeler kundene har. Ved å innhente ytterligere data om enheter og eiendeler som kunden har, vil selskapet kunne kartlegge verdien på de ulike enhetene, og dermed kan de bedre tilpasse en egenandel dersom uhellet først skulle være ute, og kunden ønsker en utbetaling fra forsikringsselskapet. Dette vil være til stor hjelp for kunden, som til en hver tid kan se hvor mye egenandelen vil være, og gjør at kundens tillit økes til selskapet (som opererer i en bransje hvor tillit er viktig for at kunde skal betale for tjenestene). Det kan virke urettferdig å måtte betale samme egenandel på f.eks. reparasjon av en MacBook fra 2011 som på en reparasjon av en MacBook fra 2020, og prisdifferensiering kan ha flere fordeler, som at selskapet enklere kan rettferdiggjøre en høyere egenandel ved ulike typer produkter. En tilleggstjeneste kan være at egenandelen reduseres på enten enkelte, eller alle eiendelene til kunden ved å øke forsikringspremien.

Til tross for strenge regler for personvern og bruk av persondata, ligger forholdene til rette for å individualisere forsikringsprisene. Med dataen kundene allerede har, kan forsikringsselskapene følge mye av den samme modellen som de sosiale platformene; De får et tilbud om et gode, i bytte mot at personlig informasjon samles inn av tilbyderen når godet anvendes. I forsikringsbransjen kan økt samtykke til å anvende personlig data imøtekommes med rimeligere forsikringer, da forsikring i dag stort sett er priset etter generisk personlig informasjon (alder, bosted, utdannelse, skadehistorikk etc.). Et eksempel kan være å samle inn helsedata fra smartgadgets i bytte mot rimeligere livsforsikring, data fra Teslaen i bytte mot rimeligere bilforsikring, etc. I dag gir mange forsikringsselskaper rabatt om kunder samler flere av sine forsikringer i ett selskap. Med stordata kan det være lettere å tilby spesifikt mersalg som treffer bedre, evt. bundle ulike og spesifikke forsikringer. I forlengelsen av punktet over, kan bruk av persondata og sensordata gi økte muligheter for (enda mer) spesifikke forsikringer av ulike formuesgjenstander, slik som vi f.eks. har sett at bunadsforsikring i senere tid har blitt populært; golfutstyrforsikring, og forsikring av annet utstyr som kan komme ut for skade/tyveri. Et annet et eksempel, spesielt vha sensorer er f.eks. brukerbasert forsikring for gjenstander. Etksempelet i dag er bil og km, men hva med å bruke sensordata til å gi forsirking basert på tidsbruk, evt. bykjøring vs. kjøring på landet(gps-data).

Kommer dessverre ikke på noe

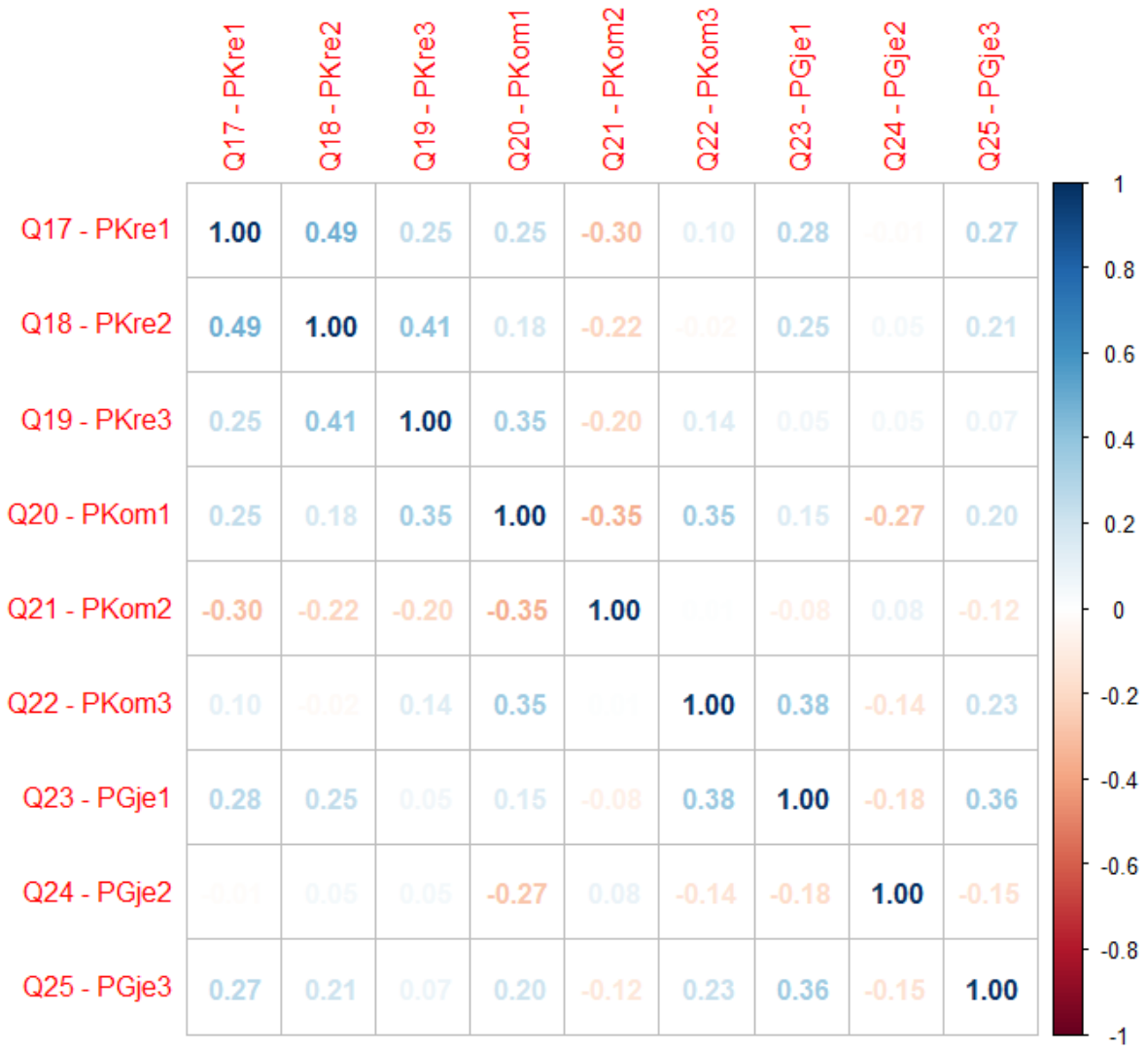
Bruk av værdata til å predikere nedbørsmengder/skredfare/hetebølger som igjen kan benyttes til å utvikle en varslingstjeneste for berørte områder. Ved utstrakt bruk av sensorer i samfunnet vil dataene etterhvert bli bedre, og man kan muligens ta seg betalt for å "være medlem" av tjenesten. Det må antagelig skapes mer verdi enn bare varslinger, og inntjeningen for forsikringsselskapet virker noe fjern. Cyberforsikring kan utvikles videre med en tjeneste som baserer seg på hjelp av 50+, som stadig vil falle lengre bak den rivende utviklingen. På samme måte som man tilbyr hjelpelinjer for psykisk syke, kan man etablere en hotline for datahjelp som et tilbud til privatpersoner, basert på alder, geografisk lokasjon og andre datakilder man har tilgang til. Kunden kan f.eks melde inn hvilke dingser de eier, og dermed få tilpasset support/tips + gode tilbud på forsikring for produktene. En "all-inclusive" pakke, hvor man kan "melde inn" så mange enheter man vil i ordningen, og samtidig alltid ha en man kan ringe for hjelp.--> en slags digital verdigjenstandforsikring.

All data er relevant for targeted marketing, forsikringsselskapet kan selv bruke dataen de sitter på til å nå ut til nåværende kunder med markedsføring som er skreddersydd mot kunder for nye produkter, altså mersalg. Alternativt kan dataen bli solgt videre til de fleste selskaper i verden som selv kan velge å markedsføre effektivt til forsikringsselskapets kunder. Dermed blir salg av data det enkleste produktet som kan skapes fra data. Dataene kan brukes til å utvikle intelligente chatbots som "vet hvem de snakker med" og hvordan kunden trenger hjelp for å øke kundetilfredsheten. Dog er det nok varierende hvor godt mottatt en slik innovasjon blir mottatt av kunden. Slik data kan og være relevant for selskaper som utvikler chatbots og kundeservice. Sensordata kan spesifikt brukes til å lage spesifikke forsikringer rettet mot for eksempel rom i hus som bad, kjøkken eller våtrom som kan anses som "høy risk". Hvis fuktsensor registrerer lav fuktighet og lav risiko for skade kan pris justeres i forhold til risiko. Det samme gjelder for kunder som bor i bygninger og godtar installasjon av sensorer, avhengig av hva som kan måles kan prisen være lav eller høy basert på risiko. I.e. hvis et bygg har sprinkleranlegg eller solid alarmsystem er risiko for henholdsvis stor brannskade og innbrudd lavere og pris kan justeres. Ved å bruke sensordata som f.eks fuktmåler eller luft/røyksensor kan forsikringsselskapet lage en tjeneste eller produkt som anbefaler kjøp basert på hva som trengs. F.eks hvis inneklima er dårlig kan kunden anbefales å kjøpe en luftrenser eller luftfukter/avfukter og potensielt kan det være forsikringsselskapet som selger disse produktene selv. Det samme gjelder helsedata som kan brukes til å anbefale produkter rettet mot aktivitet og helse for å senke risiko for dødsfall/sykdom og øke insentiv for å komme i bedre form. F.eks smartscales, sportsklokker, elektriske tannbørster, eller selv vitaminer eller reseptfrie medisiner osv. Intensiver som lavere pris på livsforsikring ved kjøp av helseprodukter kan bidra til mersalg.

Bil - smart tracking Starte samarbeid med bildelingstjenester Få direkte avtaler med billeverandører, eks Volvo Sensorteknologi

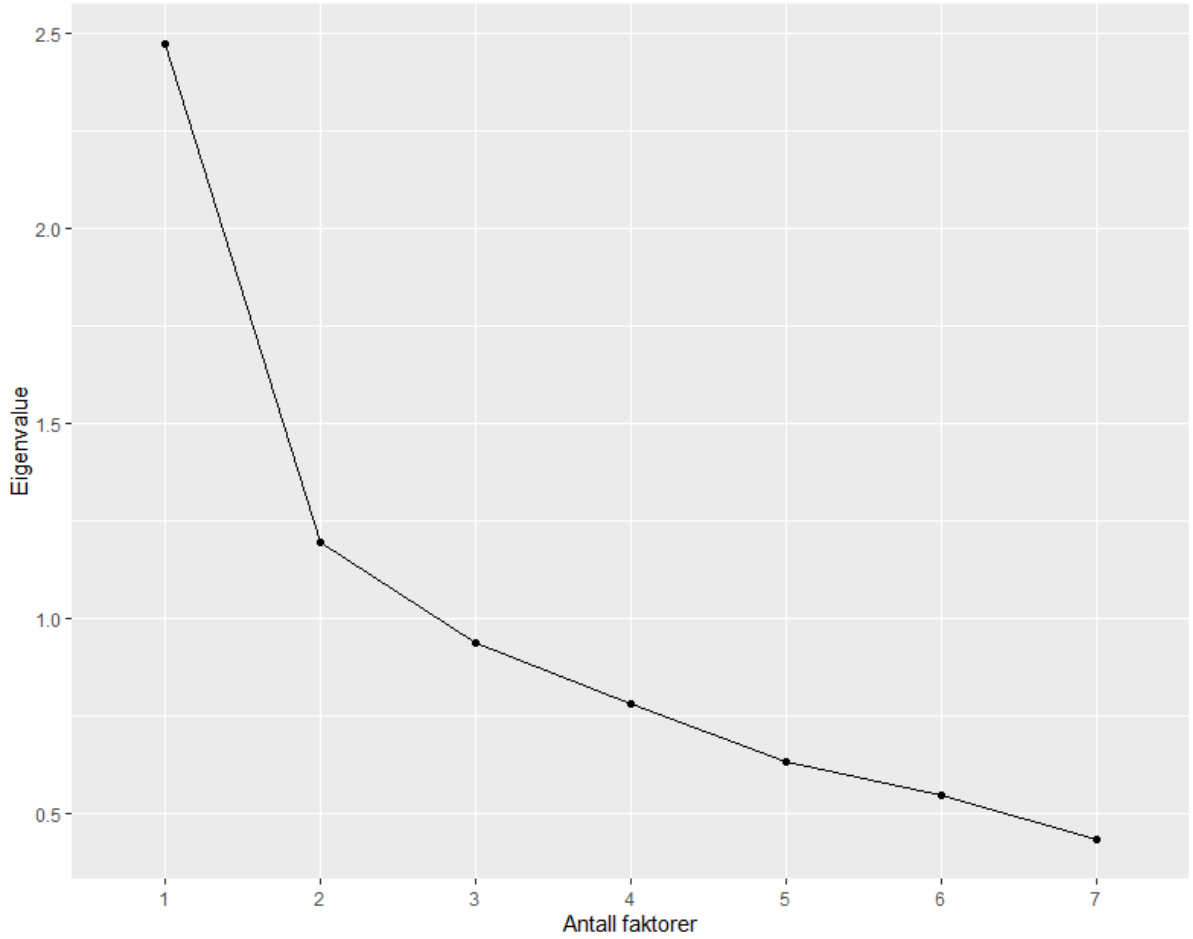
Bruk av sensordata kan være med å bestemme prisen på innbo, villa og bil. Gitt at kunden velger å benytte seg av sensorer slik at sannsynligheten for å oppdage feil/skader burde premien kunne bli noe lavere enn hos kunder som ikke vil benytte seg av sensorer. Det gir også kunden muligheten til å oppdage feilen før det blir en større skade-noe som resulterer i at det mindre sansynlighet for at selskapet må betale dyre dommer for. GPS-data kan også bli brukt for å kartlegge indivets kjøremønster ved tegning av bilforsikring. I USA eksisterer det en applikasjonen som tracker kjøremønster og registerer antall bruk av brems, blinking ved endringsretning og vedlikeholdelse av fartsgrenser. En felles platform hvor kunden får en oversikt i form av nettsted og/eller applikasjon hvor brukeren får tilgang til innsikt i egen data i et gamification-system. Ved innlogging kan kunden få oversikt over sin forsikring og hvilke data som er med å bestemme pris. Kjøp av diverse sensorer som gir et tryggere hjem kan gi bonuspoeng i gamification-systemet som gjør neste kjøp billigere eller gir en billigere pris på den aktuelle forsikringen neste måned. Ettersom det eksisterer så mye data kunden kan gi fra seg kan også en innføring av dynamiske priser være aktuelt. Gitt at brukeren ikke har meldt inn skade på lang tid blir en definert som en mindre risiko kunde og burde få en lavere pris som er tilpasset vedkommende. Det samme vil gjelde høy-risiko kunder vil få en høyere pris ved hyppig innmelding av skade. For oppsummere: 1. Platform med oversikt sine forsikringer og hvilke sensorer en kan benytte seg av for å få et tilbud som er skreddersydd vedkommende sin atferd og behov. 2. Enkel tilgang til kjøp av sensorer som kan styrke tryggheten for brukeren 3. Ved kjøp får brukeren "poeng" og ved oppnåelse av en viss "poengscore" vil en få tilbud på andre sensoriske produkter eller lavere pris på forsikringspremien. 4. Dynamiske priser kan gi brukeren insentiver for installasjon av produkter og/eller endre sitt kjøremønster feks. Oversikt og forståelse for prissetting av forsikring er viktig fra kundens perspektiv, og muligheten for å kunne påvirke avtalene en har med forsikringsgiveren. 46% av norges befolkning er åpne for å dele egen data for å bedre egne avtaler, dvs at det er store rom for at summen som betales hver måned er skreddersydd etter kundens atferd, behov og ønske om installasjon av produkter om skaper trygghet for alle involverte parter.

A3 Korrelasjonsmatrise



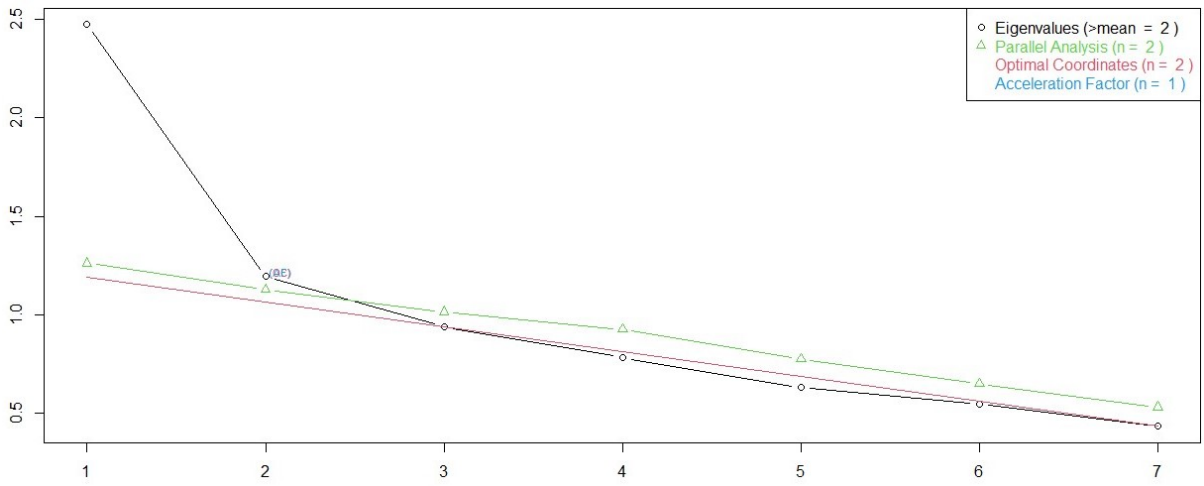
Figur A3.1: Korrelasjonsmatrise faktormodell

A4 scree-test



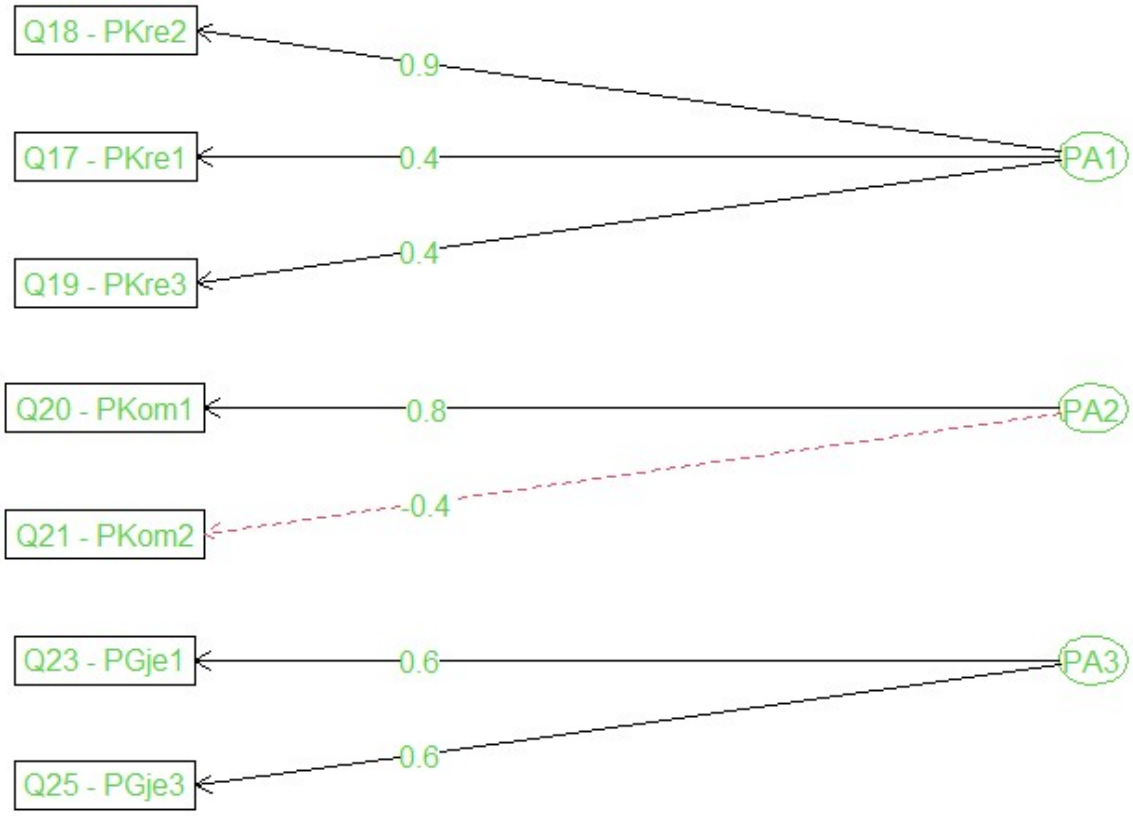
Figur A4.1: scree-test

A5 Paralellanalyse



Figur A5.1: Parallell analyse

A6 Faktoranalyse diagram



Figur A6.1: Faktoranalyse diagram

A7 Multippel lineær regresjon total score og faktorer

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.7282 -0.5775  0.2362  0.8852  2.0917

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.63134    0.13958  26.015   <2e-16 ***
F1          -0.03256    0.16116  -0.202    0.840
F2           0.02145    0.17154   0.125    0.901
F3          -0.03393    0.19582  -0.173    0.863
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.339 on 88 degrees of freedom
Multiple R-squared:  0.001048,  Adjusted R-squared:  -0.03301
F-statistic: 0.03077 on 3 and 88 DF,  p-value: 0.9927
```

Figur A7.1: Multippel lineær regresjon total score og faktorer

A8 Regresjon ideer og tidsbruk

Uavhengige variabler	Ideer	Tidsbruk
N7 - Bakgrunn (forretningsbakgrunn)	0.80(*) 0.37 std	4.73(*) 2.30 std
N7 - Bakgrunn (innovasjonsbakgrunn)	1.53(***) 0.41 std	4.92 2.62 std
N7 - Bakgrunn (tekniskbakgrunn)	0.88(**) 0.39 std	0.80 2.46 std
Kontroll variabler		
Q3 - Kjønn (mann)	-0.25 0.29 std	-1.27 1.93 std
Q4 - Alder (yngre)	0.10 0.35 std	3.11 2.23 std
Q5 - Arbeidsstatus (arbeidstaker)	0.00 0.32 std	2.64 2.07 std
Q5 - Arbeidsstatus (arbeidsledig)	-1.01 0.32 std	46.85(***) 6.05 std
Q6 - Utdanningsretning (annet)	-0.18 0.37 std	-0.66 2.64 std
Q6 - Utdanningsretning (tech)	-0.10 0.43 std	-1.96 2.75 std
Q7 - Tilleggsutdanning (ja)	0.47 0.28 std	
Q10 - Utdanningsnivå (høy)		-0.86 3.39 std
Konstant	0.85 0.50 std	9.36 4.71 std
Statistikk		
N.	92	92
F-statistic	2.23	6.99(***)

Rsquared	0.21	0.46
Adjusted Rsquared	0.12	0.39

Tabell A8.1: Regresjon ideer og tidsbruk

A9 Datasett for større modeller

Variabel	Variabel type	Data type	Referanse variabel
N6 - Tot	Avhengig variabel	Nummerisk	-
N7 - Bak	Uavhengig variabel	Faktor - 4 kategorier	Ingen relevant bakgrunn
Q3 - Kjønn	Kontroll variabel	Binær - mann og kvinne	Kvinne
Q4 - Alder	Kontroll variabel	Binær - eldre og yngre	Eldre
Q5 - Arbeidsstatus	Kontroll variabel	Faktor - 3 kategorier	Arbeidsledig/pensjonist
Q6 - Uretning	Kontroll variabel	Faktor - 3 kategorier	Business
Q7 - Tillegsutdanning	Kontroll variabel	Binær - ja og nei	Nei
Q10 - Univå	Kontroll variabel	Binær - høy og lav	Lav
Q11 - Årslønn	Kontroll variabel	Binær - høy og lav	Lav

Tabell A9.1: Datasett for større modeller

A10 Multippel regresjon Modell 3

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.23875 -0.67795  0.04598  0.77596  2.59566

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.64930    0.52488   3.142  0.00232 **
`N7 - Bak`Forretningsbakgrunn  0.95919    0.34713   2.763  0.00703 **
`N7 - Bak`Innovasjonsbakgrunn  1.18776    0.39452   3.011  0.00344 **
`N7 - Bak`Tekniskbakgrunn      1.03510    0.34002   3.044  0.00311 **
`Q10 - Univå`Høy               1.23842    0.44992   2.753  0.00724 **
`Q3 - Kjønn`Mann              -0.02355    0.27180  -0.087  0.93117
`Q4 - Alder`20-29              0.18328    0.34541   0.531  0.59708
`Q4 - Alder`Yngre              0.35103    0.39723   0.884  0.37938
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.175 on 84 degrees of freedom
Multiple R-squared:  0.265,     Adjusted R-squared:  0.2037
F-statistic: 4.326 on 7 and 84 DF,  p-value: 0.0003931
```

Figur A10.1: Multippel regresjon Modell 3