



Unleashing the Potential in Ambassador Programs

*Identifying Concrete Measures to Improve the Efficiency of Ambassador
Programs*

Thomas Garås Gulli and Frida Rasch

Supervisor: Steffen Juranek

Master thesis, Economics and Business Administration

Major: Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible – through the approval of this thesis – for the theories and methods used, or results and conclusions drawn in this work.

Acknowledgements

This thesis was written as part of a major in Business Analytics at the Norwegian School of Economics (NHH). We have had the privilege of researching a highly relevant business problem in the tech world due to our collaboration with SafetyWing.

During the process of writing this thesis, we have been able to both apply several methods learned from various courses and broaden our horizons with new tools.

We would like to thank SafetyWing, Inc. for allowing us to explore a real-life business problem and for sharing their data and good advice along the way.

We want to extend gratitude to professor Gisle Andersen who advised us on methods for textual analysis. Furthermore, we want to thank assistant professor Maximilian Rohrer for solid advice within the field of topic modeling. Finally, we want to express our sincere gratitude to our supervisor, associate professor Steffen Juranek, for excellent support and guidance along the way. We are also very thankful for Steffen always reminding us to take a step back and keep it simple.

Norwegian School of Economics

Bergen, June 2021



Thomas Garås Gulli



Frida Leiren Rasch

Abstract

This thesis aims to analyze how SafetyWing can improve the efficiency of its ambassadors and what the implications are for ambassador programs in general. Ambassador programs are an expanding part of digital marketing for businesses. Briefly, ambassador programs are when a business awards another business a commission for sending traffic their way. SafetyWing, a Norwegian startup operating on a global basis within the insurance industry, is a company that relies on ambassadors for digital marketing.

We collected user data from SafetyWing, as well as traffic data describing traffic entering the SafetyWing webpage. In addition, we collected both traffic data and textual data from the web pages of SafetyWing's ambassadors. We have used various methods such as textual analysis, topic modeling, correspondence analysis, regression, and visualizations to analyze the data. In addition, we created two new variables, namely category (1-4) and topic (1-6). To analyze profitability, we divided total sales per blog post by total traffic to the blog.

The results showed that blogposts that focus extensively on SafetyWing, category 4, are much more efficient than those focused elsewhere, because of the focus. Moreover, blog posts about digital nomads, the target group of SafetyWing, are relatively unprofitable in category 4. Lastly, we found that blog posts that focus somewhat or extensively on SafetyWing, benefit from also including one or a few competitors.

Our conclusion is shaped as a recommendation to SafetyWing. Namely, we recommend SafetyWing advise ambassadors to write a category 4 blogpost and use this actively as a referral source from other blog posts. Additionally, the post should include one or a few of the competitors of SafetyWing.

Keywords – Ambassador Program, Affiliate Program, Marketing, Topic Modelling, SafetyWing, Insurance

Contents

1	Introduction	1
2	Data and Methodology	4
2.1	Data and Data Collection	4
2.1.1	The SafetyWing Database	4
2.1.1.1	Outliers and Missing Values	6
2.1.2	Google Analytics	8
2.1.2.1	Identifying the Web Pages of Ambassadors	8
2.1.2.2	Data Description	10
2.1.3	Web Scraping of Ambassadors Blogposts	12
2.1.4	Alexa Web Information Service	13
2.1.5	Data Filtering	14
2.2	Unit of Measurement	15
2.3	Extracting Statistics from Blogposts and Categorization	17
2.4	Topic Modelling	19
2.5	Correspondence Analysis	21
2.6	Limitations	22
2.6.1	Implications of Covid-19	22
2.6.2	Google Analytics is Limiting Traffic Data from Social Media	23
3	Analysis and Results	25
3.1	Examining the Ambassador's Audience	25
3.1.1	Niche Markets	25
3.1.2	Regions	30
3.1.3	Age Groups	32
3.1.4	Genders	32
3.1.5	Diversification Is a Necessity	33
3.2	How does Category and Topic Correlate with Sales	34
3.2.1	Identifying Distinguishable Combinations of Category and Topic	34
3.3	Are SafetyWing's Competitors a Threat or an Opportunity?	37
3.3.1	Is the Presence of Competitors Profitable?	38
3.4	Investigating SafetyWing's Vulnerability to Market Fluctuations	40
3.5	Identifying Recommendations for Improving Ambassadors	45
4	Conclusion	47
	References	50

Appendix	52
A SafetyWing's Competitors	52

List of Figures

2.1	Entity Relationship Diagram of the relevant tables and columns from the SafetyWing database	5
2.2	Histogram over the number of ambassadors with X number of users referred. The last column, 100, illustrates ambassadors with 100 or more users referred. The binwidth is 1 user.	6
2.3	Boxplot for observations of total amount paid	7
2.4	Histogram showing the number of ambassadors with X number of blogposts. The binwidth is 1 blogpost.	10
2.5	Left: Histogram showing the number of blogposts with X total sessions. Binwidth is 10 sessions. Right: Histogram showing the number of blogposts with X total sales. Binwidth is 1 sale.	11
2.6	Histogram showing the number of blogposts with X days of lifetime, meaning the number of days a blogpost have existed. Binwidth is 10 days.	11
2.7	Histogram showing the number of blogposts with X page views per million. This plot is based on the average daily page views per month data and binwidth is 0.05 page views.	14
2.8	Histogram showing the number of blogposts with X in adjusted sales. The binwidth is 0.5 sales. Posts with a value higher than 20 is given with 20 as adjusted sales.	17
2.9	Left: Histogram showing the number of blogposts with X SafetyWing mentions. Binwidth is 1 mention. Right: Histogram showing the number of blogposts with X competitors mention, including World Nomads. Binwidth is 1 mention.	18
2.10	Frequency of category occurrences	19
3.1	The topic proportions of each topic, sorted from largest to lowest	27
3.2	Correspondence analysis between ambassador and topic proportions. The color indicates which revenue percentile the ambassador belongs to. Ambassadors are identified by a dot, while a triangle identifies mean topic proportions.	29
3.3	The continent proportion of users referred by ambassadors, based on the ambassador's most used topic.	30
3.4	Gender proportion of users referred by ambassadors, based on the ambassador's most used topic	33
3.5	Category VS Topic by the adjusted sales measure	35
3.6	Topic VS Average Mention of Competitors VS Only World Nomads per post by average sale per post	38

3.7	Topic VS Category by Average Mention of World Nomads or All Competitors	39
3.8	Topic proportions and Sales/Session ratio over time from May 2019 to March 2021	41
3.9	Proportions of the different categories over time, from May 2019 to March 2021	42
3.10	The barplot illustrates the number of sales each month, with the y-axis to the left. The black line represents the average sale per post, with the y-axis on the right. Due to limitations that led to missing data in October 2020, this month should be ignored.	43
3.11	The average lifetime of blog posts within each category. The lifetime is based on the duration between the first and the last time a blog post has referred traffic to SafetyWing.	44

List of Tables

2.1	Description of the relevant tables in the SafetyWing database	5
2.2	The filtering steps taken to arrive at blogposts successfully matched to an ambassador	15
2.3	Description of the categorization of ambassadors blogposts	18
2.4	The topics introduced from topic modeling with their respective top 10 keywords. The name in italic is our custom name for the topic, and the percentage is the proportion of the texts with the respective topic.	21
3.1	Number of ambassadors distributed on the topics	26
3.2	Mean topic proportions for the different most popular topics. The most popular combination is highlighted with green, the second most with blue, and the least popular with red.	28
3.3	31
3.4	The proportions of ambassadors with proportion of users from different regions. The columns illustrate the number of ambassadors with X %, e.g., <i>Low</i> between 0 - 20 %, of users from the region specified on the row. For example, 16.5 % of all ambassadors have a proportion of users from Asia, making up 80 - 100 % of their user base. Similarly, 2.4 % of ambassadors have a proportion of users from Oceania of 40 - 60 %.	31
3.5	The average age of users referred by ambassadors depending on their most popular topic.	32
3.6	Sessions driven from categories of blogposts per page view	45
A0.1	Most Relevant SafetyWing Competitors	52

1 Introduction

The concept of modern-day marketing started developing during the Industrial Revolution (Hardy, J., 2016), and since numerous innovators have found new ways to market their business or product. The various ways of marketing have particularly exploded along with the increased use of the internet and tech. Another market that has exploded in the same era, is that of bloggers and website owners. In order to make a living, they can rely on advertising or marketing other businesses.

In this thesis, we discuss and analyze the efficiency of ambassadors programs, also called affiliate programs, as a marketing tool. An ambassador program is a tool of marketing involving, among others, websites marketing other businesses and their products. The ambassador handles all the risk and takes care of promotion and creating the material. In return, the ambassador typically gets a commission from the business. In other words, an ambassador program can be compared to having a large, independent sales force purely living off commission (Duffy, 2005). Together with our partner company SafetyWing, we are empirically investigating their ambassador program.

SafetyWing is a Norwegian startup founded in 2017 operating on a global basis within the insurance industry. Their vision is to remove the role of geographical borders as a barrier to equal opportunities and freedom for everyone (SafetyWing, 2021b). In their series A round in January 2021 they raised \$8 million (Crunchbase, 2021) to expand their product offerings, primarily a global retirement product (Keane, J., 2021). SafetyWing currently offers two main insurance products: Nomad Insurance (NI) and Remote Health (RH). While NI has several similarities with classical travel insurance, RH is a global health insurance with coverage in the insurance holder's home country as well as abroad. Furthermore, RH, in contrast to NI, targets businesses in addition to private persons. NI has been active since February 2018, while RH was launched in March 2020 (Brusnahan, P., 2020). As a result, data collected is far less extensive for RH and is therefore excluded from our analysis.

In order to reach their target audience, their ambassador program is crucial as they rely only on ambassadors and organic growth. A variety of businesses, such as travel websites, partner up with SafetyWing and promote their insurance product through their own

business. In this thesis, we have limited the ambassadors included to only ambassadors writing a blog, to simplify data collection. The ambassadors gain a commission based on the customers they recruit (SafetyWing, 2021a). By the 1st quarter of 2021, traffic driven by ambassadors to the SafetyWing website consisted of roughly 10 %. Meanwhile, income generated by ambassadors consists of more than 30 % of SafetyWing’s total income. SafetyWing has historically focused on reaching out to travel blogs, explicitly digital nomads (see fact box). They consider their current reach to be primarily digital nomads and, to some extent, backpackers, mainly located in South East Asia between 25 and 35 years old. A part of their new strategy is also recruiting visa services and online travel agencies (OTAs) as ambassadors.

Digital nomad definition

The definition of a digital nomad can be broken into two main categorizations.

1. Location-independent
2. Remote, technological job

Hayes, A. (2021)

In order to analyze SafetyWing’s ambassador program, we have collected a wide range of data and used various methods. The data includes SafetyWing’s database, traffic data from Google Analytics and Alexa, and the textual blogposts ambassadors have written to promote SafetyWing. Throughout the work with this thesis, we have conferred with several of SafetyWing’s employees. This has been necessary to ensure that the data is handled correctly and lawfully, and that our mission coincides with a realistic business problem. In order to analyze the data, we have used a range of methods such as textual analysis, topic modeling, correspondence analysis, regression, and visualizations.

As of May 2021, SafetyWing has almost 1000 registered ambassadors and about 350 active ones. However, the total income for the three most successful ambassadors constitutes a whopping 36 % of all ambassador income. Meanwhile, roughly 50 % of the active ambassadors earn less than \$10 per month. In other words, the differences between the ambassadors in the SafetyWing ambassador program are enormous. The purpose of this thesis is to look closely at SafetyWing’s ambassador program in order to identify *why* some ambassadors are more successful than others and how SafetyWing can advise individual ambassadors to perform better. Another perspective we will look closer at is how the insights learned by looking at SafetyWing’s ambassador program can be applied

to ambassador programs in general. The benefits of having ambassadors are prominent as companies can outsource the marketing and the risk to external actors. Outsourcing is, however, limiting the control and overview and making it hard to understand why some ambassadors are effective and others not. To investigate this further and help make ambassador programs easier to follow up, we have formulated the following research question, which will be examined in this thesis:

How can SafetyWing improve the efficiency of its ambassadors, and what are the implications for ambassador programs in general?

2 Data and Methodology

This section describes the data sources and the methods used in this thesis. We will start by describing each data source and the steps taken to collect, preprocess and explore them. Then we will look at how we created a unit of measurement which made every ambassador comparable before looking at retrieving statistics and categorizing each blogpost. The next step is to introduce and describe the essential methods in this thesis, topic modeling, and correspondence analysis. Lastly, we will describe limitations that are important to consider when reading our analysis in the next section.

2.1 Data and Data Collection

Various data sources are crucial to understanding differences between ambassadors and best practices in ambassador programs. By only looking at SafetyWing’s existing data, one would find limited amounts of new insight that could improve their ambassador program and uncover insights for ambassador programs in general. Therefore, it was necessary to broaden the horizon and collect data from external sources. This subsection will go through the different data sources and the steps taken to collect and preprocess them, making them ready for analysis. We will start by investigating the SafetyWing database before jumping over to traffic data from Google Analytics and the Alexa Web Information Service. The subsection ends with describing how we collected and preprocessed textual data from ambassadors’ blog posts, and finally, a summary of which data filtering steps we performed.

2.1.1 The SafetyWing Database

SafetyWing supplied us with a snippet of their database which described user and ambassador activity from February 2018 to April 2021. The original snippet consisted of 29 anonymized tables in SQL format. However, all tables and columns were not relevant, making it possible to filter out 24 of the provided tables. In the remaining five tables, we kept 33 out of the 86 columns. These remaining tables and columns describe the characteristics of insurance contracts, payment data, users, and ambassadors. They are illustrated in the ER diagram in Figure 2.1 and described in Table 2.1.

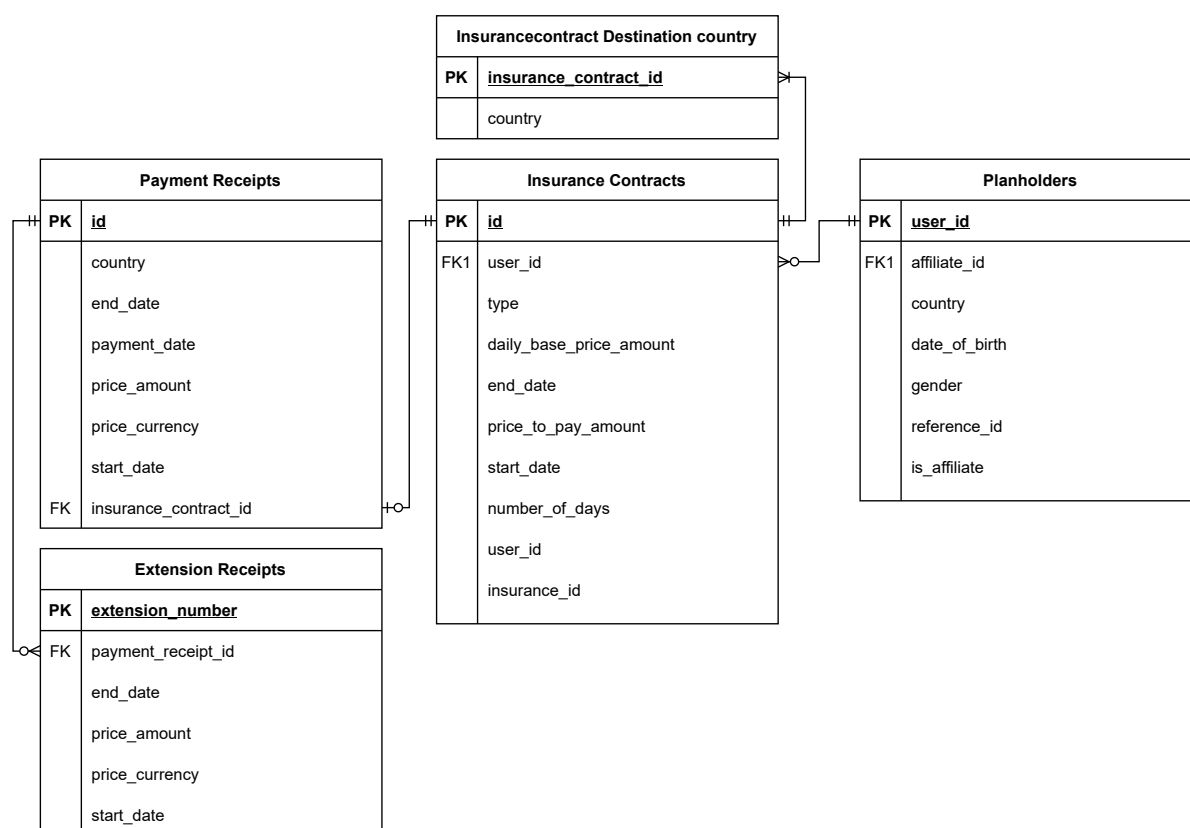


Figure 2.1: Entity Relationship Diagram of the relevant tables and columns from the SafetyWing database

Database table	Description
Planholders	All users registered, both regular users and ambassadors. If ambassador, the boolean <i>is_affiliate</i> is <i>true</i>
Insurance Contracts	All unique insurance contracts
Insurance Contract Destination Country	The destination country an insurance contract was primarily bought for
Payment Receipts	Payment information for the first payment in a subscription and one-time payments
Extension Receipts	Payment information for subsequent payments in a subscription

Table 2.1: Description of the relevant tables in the SafetyWing database

The first step of preprocessing the data snippet was to convert the SQL file with the relevant tables and columns to RData-format. This is consistent with the R-programming language and allowed for seamless integration with R Studio. Then, it was necessary to

classify each column as either numeric, factor, date, or character, depending on what the data represented. The last step was to set missing values to NA (Not Available) to avoid sources of error in the analysis.

In total, we had data on 12 781 users who bought a total of 21 783 insurance contracts. Out of these users, 4 438 were referred by 350 different ambassadors, and 7 153 insurance contracts were bought. As seen in the histogram in Figure 2.2, most ambassadors referred below 5 users with 1 as the most prominent column. On average, each of the 350 ambassadors recruits 11.66 users each.

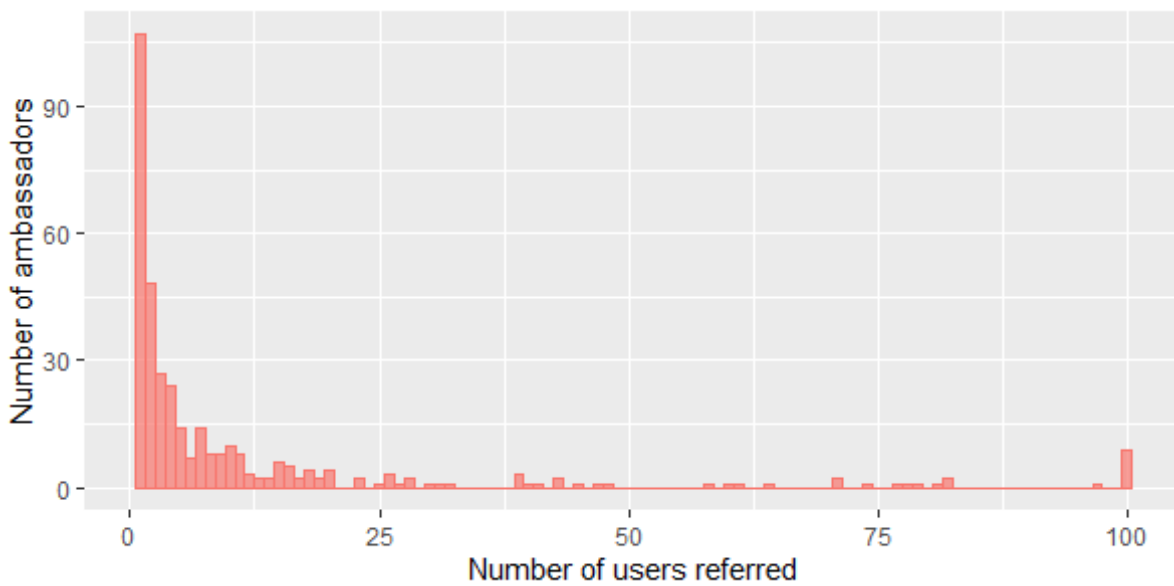


Figure 2.2: Histogram over the number of ambassadors with X number of users referred. The last column, 100, illustrates ambassadors with 100 or more users referred. The binwidth is 1 user.

2.1.1.1 Outliers and Missing Values

After the filtering, which resulted in a dataset of five tables, only one variable could contain outliers, which describes the total amount an insurance holder has paid. This variable is the only value that can contain outliers in our dataset from SafetyWing because the price differs based on the type of insurance and the possibility for customers to purchase insurance for others (e.g., family or friends) in addition to themselves. This means that the value of this variable depends on multiple factors such as age, duration, number of members on the plan, coverage area, and insurance add-ons. For users with multiple plan members for a long duration, one would see a high amount to be paid compared to users

only buying for themselves for a shorter period. This is because there are no quantity discounts and the longer the period, the more has to be paid. As Figure 2.3 shows, the spread of the values for the total amount paid is large. One can see that most of the values fall in the lower segment below about \$ 300, while the tail goes as far as above \$ 3 500. The considerable spread in the payment variable made us look away from this as a tool of measurement. That means that in this thesis, we will only look at *the number* of users referred, not *the value* of users referred.

All other columns have strict requirements for data entered, such as numeric values within a range, categorical values, or custom strings irrelevant to the analysis. Of course, there may be typographical errors in user data entered manually, like age and country for users, affecting the results. However, this is difficult to consider, as we have identified no outliers or strange values in these columns, meaning typographical errors would be within the allowed values.

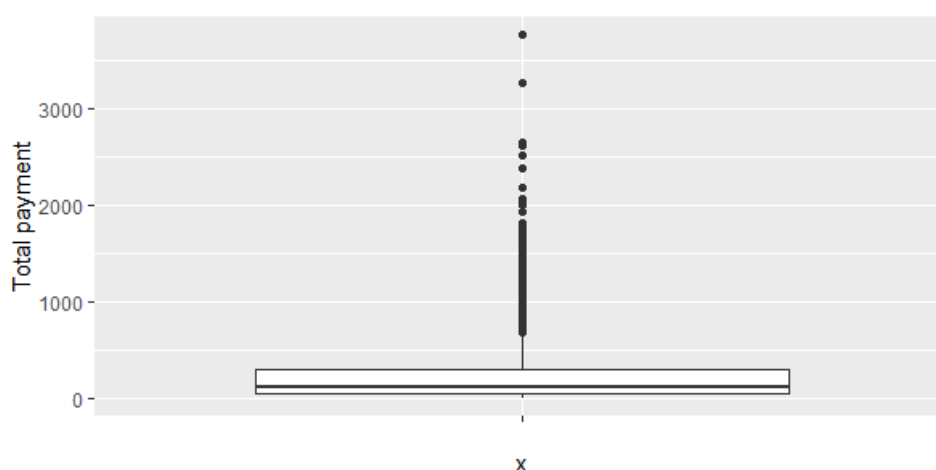


Figure 2.3: Boxplot for observations of total amount paid

As mentioned previously, we have denoted all missing values as NA (Not Available). We have handled these in the following two manners: 1) If there is only one or two NA's per observation, we have considered if the exact variable is crucial. If it is crucial, we have removed the entire observation. 2) If several or all variables per observation are NA's, we have removed the entire observation.

2.1.2 Google Analytics

Google Analytics is a service from Google used to track traffic on websites and apps (Google Analytics, 2021). The purpose of using Google Analytics in this task was to identify the blog and blogposts for each ambassador and get data for sales, user registrations, and total sessions referred to SafetyWing.

To retrieve data from Google Analytics, the R-package *googleAnalyticsR* was utilized. This allowed for a direct connection to the Google Analytics account of SafetyWing and allowed more than a hundred data points to be downloaded. All these data points were however not relevant, and it was filtered down to the following:

- Source – the web page where the user clicked the affiliate link, i.e., `vg.no`
- Full referrer - the full URL to the page where the user clicked the affiliate link, i.e., `vg.no/travel/article`.
- Sessions - number of sessions (clicks) from each full referrer. A session lasts until a user has been inactive for 30 minutes or more (Google Support, 2021). Meaning that no matter how many page views a user leaves behind, it will only be counted *once* - as long as the duration between the page views is less than 30 minutes.
- User registrations - number of users that have created a user account at SafetyWing after being referred by a blog.
- Sales - number of users that purchased an insurance contract after being referred by a blog.
- Landing page - the subpage the user landed on SafetyWing.com. This includes the affiliate ID, ex. `safetywing.com/a/affiliateID`.
- Date - The date in which the session, registration or sale took place .By calculating the number of days between the first and last time a blogpost appeared in the data, it was possible to calculate the lifetime of a blogpost.

2.1.2.1 Identifying the Web Pages of Ambassadors

Having an overview of which URLs ambassadors were using to promote their insurance products was crucial to use blogs as an input in our analysis efficiently. As SafetyWing

had no records of this, we needed to collect this data ourselves. The URLs could be fetched manually, going through each ambassador in the database and looking them up on the Internet. Nevertheless, the identification of the websites could also be made automatically by utilizing the traffic data from Google Analytics. This made us confident that no ambassadors were overlooked and that the blogposts used have driven traffic to SafetyWing.com at least once. Another advantage gained by using the Google Analytics data was that we got the full URL of each blogpost with referred traffic. Meaning we knew precisely which posts had led to a session, sale, or user registration. This made it possible for us to skip making scripts that retrieved the URLs for every blogpost on each blog and then figure out which of these mentioned SafetyWing. Such a method could be complicated due to the different structures on the different blogs and the fact that even though SafetyWing is not mentioned, it may have led to a sale.

The first step in the process of identifying the websites of ambassadors was using regular expressions to retrieve the ambassador ID from the *landing page* variable. This approach was possible because all ambassadors have a unique identifier that has to be used in the URL to SafetyWing.com they promote to earn a commission. This identifier will be visible at the end of the *landing page* from the Google Analytics data on the form `/a/12345678` where `12345678` is the ambassador id. It is then possible to match the *landing page* containing an ambassador ID to the *full referrer* of blogpost that referred the session to SafetyWing.com. The ambassador ID in the landing page is the same ID as in the SafetyWing database, which allowed for connection to further data on the ambassador and the users it has referred.

However, not all blogposts could be automatically matched to an ambassador due to missing ambassador ID in the landing page URL. The cause for this can be multiple things. First of all, the blogpost might not belong to an ambassador, but instead, write about and refer to SafetyWing in general. In some cases, the blogpost might belong to an ambassador, but some errors affected the Google Analytics data blocking the automatic match. To fix this problem, blogposts missing an ambassador ID can be manually investigated and potentially matched. The data for this thesis was, however, extensive enough that we did not see this need.

2.1.2.2 Data Description

In total, we retrieved 12 109 unique web pages and blogposts referring traffic to SafetyWing.com between 1st of January 2019 and 1st of April 2021. However, in these unique records, one will also find irrelevant web pages like Google.com, job advertisement pages, etc. After filtering out these, we had 3 392 records left. However, many of these records had few sessions, and no sales or registrations were referred. This led us to only keep records with at least 5 sessions or 1 sale, ending up with 693 records left. The final step in the data filtering of this part was to filter out blogposts that could not be matched to an ambassador. This accounted for 231 blogposts which made us end up with 462 blogposts written by 127 different ambassadors.

As Figure 2.4 shows, just above half of the ambassadors have only 1 blogpost each. The distribution is also very left-oriented, with only a few ambassadors with more than 10 posts. On average, each ambassador has 3.64 blogposts.

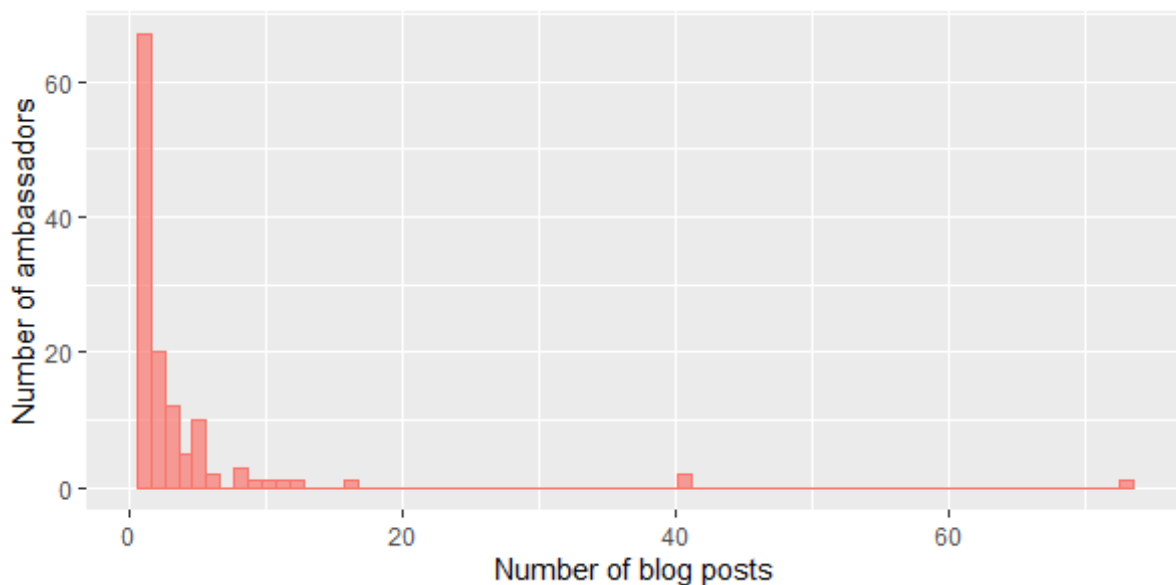


Figure 2.4: Histogram showing the number of ambassadors with X number of blogposts. The binwidth is 1 blogpost.

The number of sessions and sales per blogpost is also in the lower segment when looking at the histograms in Figure 2.5. Most blogposts have fewer than 20 sessions in total and 0 sales, even after filtering out blogposts, so we only kept blogposts with at least 5 sessions or 1 sale.

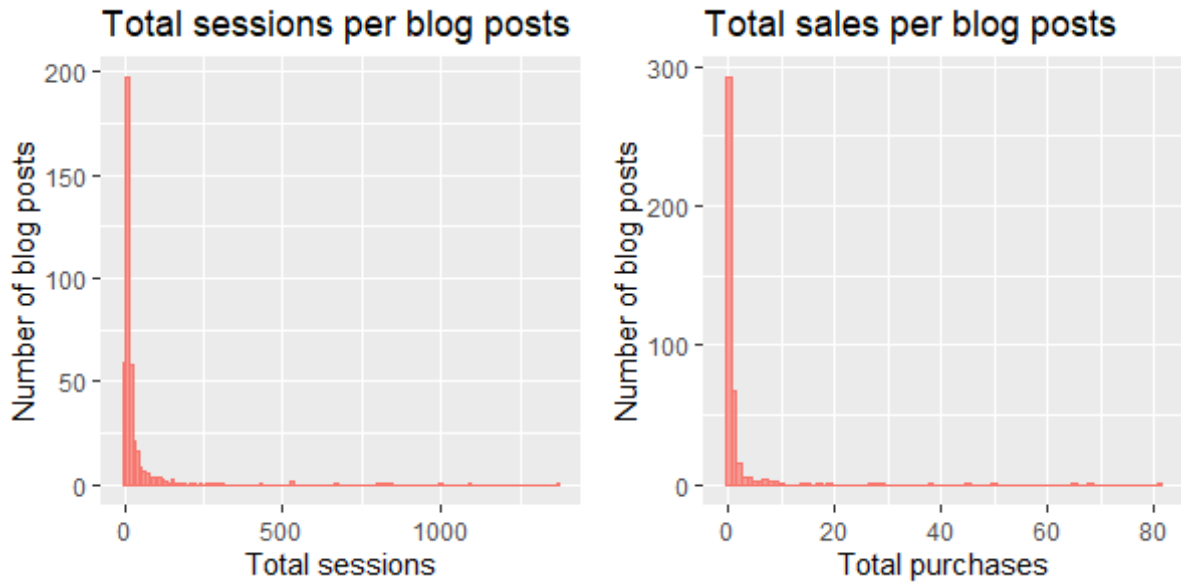


Figure 2.5: Left: Histogram showing the number of blogposts with X total sessions. Binwidth is 10 sessions. Right: Histogram showing the number of blogposts with X total sales. Binwidth is 1 sale.

The lifetime of the blogposts also varies, calculated by looking at the first and last time a blogpost has referred a session to SafetyWing.com. As Figure 2.6 show, the histogram is almost falling linearly. Most blogposts are in the left section of the histogram, indicating they are younger than the others. Only a portion is older than a year, and on average, a blogpost is 183.03 days old.

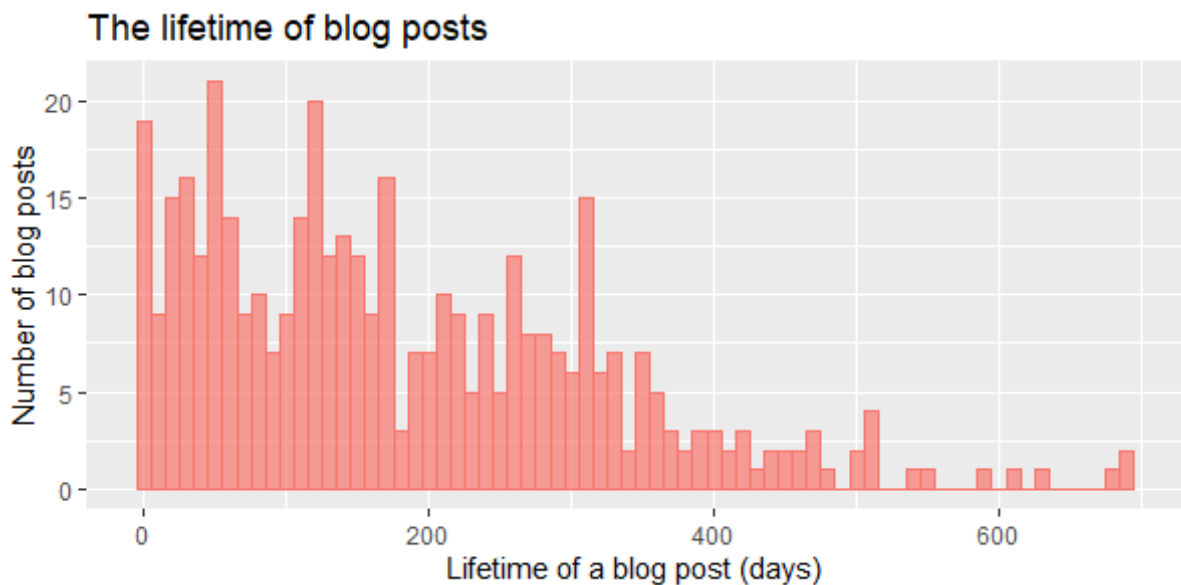


Figure 2.6: Histogram showing the number of blogposts with X days of lifetime, meaning the number of days a blogpost have existed. Binwidth is 10 days.

2.1.3 Web Scraping of Ambassadors Blogposts

Web scraping is a method that allows for effective data collection from websites (Kaur, H., 2020). Simply put, the steps are as followed; Firstly, data in HTML format is collected. Secondly, one may specify which parts to keep by the use of HTML tags. For example, the tag `<p>` represents paragraphs. Therefore, by specifying that one wants to keep all `<p>`-elements, you acquire all, if any, paragraphs from the specific website. Lastly, the data collected is structured in a manner that fits further analysis.

The web scraping was performed by using the `read_html()` function in the `rvest` package to fetch the HTML code which consists of various HTML-tags. HTML-tags are used to organize text into specific elements, such as `<p>` for paragraph and `<header>` for header (Glass, E., 2020), in the hypertext markup language, HTML (Domantas, G., 2019). Overall, a HTML-file consists of two elements - head and body, where body is the relevant part for our analysis. For each blogpost, we extracted data for the following HTML-tags within the body-tag: paragraph (`<p>`), lists (``), titles (`<title>`), headings (`<h1>`-`<h6>`), links (``) and underscores (`<u>`). This data extraction was performed with the following two steps per HTML-tag per blogpost:

- Select content within the specific HTML tag, e.g., `<p>` and `</p>`, with function `xml_find_all()` from the `xml2` package
- Convert HTML to text format with the function `html_text()` from the `rvest` package. This removes all HTML tags and make us end up with plain text.

After the elements from the blogposts had been extracted, we performed a language check on each paragraph element to determine the language of the post. As mentioned above, we only wanted to keep posts in English. The language check was performed with the `textcat()` function from the `textcat` package in R which can identify 26 different languages. The last step in preparing the blogpost data was to pre-process it and prepare it for analysis. This preprocessing removed all punctuation, whitespaces, and stopwords (the most common words in the language such as *and*, *but*, *or*, etc.) as well as stemmed the words (e.g., converting teachers to teacher).

In total it was possible to web scrape 441 of the 462 blogposts provided in the previous section. The reason for this difference of 21 blogposts is two-fold. First of all, not all URLs

to blogposts are working. This may be due to changes in the URL structure, for example, if the ambassador has added or removed a category or date; or that the blog no longer exists. Secondly, we only kept posts written in English. The following two arguments can summarize our reasoning for excluding posts in other languages. Firstly, texts written in different languages are not suitable for direct comparison. Secondly, using a translation tool, such as Google Translate is not entirely reliable.

2.1.4 Alexa Web Information Service

The ambassadors in the SafetyWing Ambassador Program have blogs of different sizes. By comparing the sales figures from the SafetyWing database or Google Analytics, one would be biased towards ambassadors with the most sales. However, the number of sales does not necessarily tell whether an ambassador is efficient or if there is room for improvement. Therefore, to compare ambassadors with fewer biases, retrieving traffic data for each blog was necessary. To efficiently do this, traffic data for each blog was retrieved using the Alexa Web Information Service API. This API allows for connection to retrieve data from Alexa.com - a service by Amazon that, among other things, estimates daily hits on most web pages on the world wide web. As we will describe in subsection 2.2, the retrieval of traffic data will make it possible for us to measure the efficiency of the blogs without biasing more prominent blogs due to possibly higher sales figures.

The traffic measure provided by Alexa is called *page views per million*. This measure is based on pages viewed by users who have the Alexa toolbar or extension installed in their browser. If a web page has a score of 1, it means that 1 page view has been seen in a sample of 1 million users with this toolbar or extension installed; a score of 10 000 would indicate 10 000 page views in the sample (Klipfolio, 2021). The page views are on a daily basis.

It is important to note that the traffic data from the Alexa Web Information Service is just an estimate. As it is calculated based on the web pages users with the Alexa toolbar or extension installed visit, it may not represent the rest of the internet population. For our use case, adjusting the sales figures to a scale that takes the blog's size into account, we find it accurate enough as it was the best tool we found available.

We managed to retrieve traffic data for 420 of the blogposts, lowering the number of posts

for analysis from the 441 as described in the last section. Some of these blogs did, however, have traffic numbers that sky-rocketed in comparison with the other blogs. Looking closer at these, it was evident that these web pages were no regular blogs and irrelevant for this analysis. We, therefore, filtered out all web pages with less than 1 page views per million and ended up with 416 blogposts.

Figure 2.7 represents the distribution of the average daily traffic data among the blogposts. As one can see, most of the blogposts are close to 0. As we have fetched data on a daily basis for each blog, it is important to note that the distribution includes multiple records per blogpost. In order to make it better suited for analysis, we stored the average daily page views per month. This means that the plot below has one record per blogpost per month with data from Google Analytics. The distribution shows us how small the blogs of our ambassadors are, with an average of about 0.192. This means that for every 1 million users with the Alexa toolbar or extension installed, 0.192 page views have been registered daily.

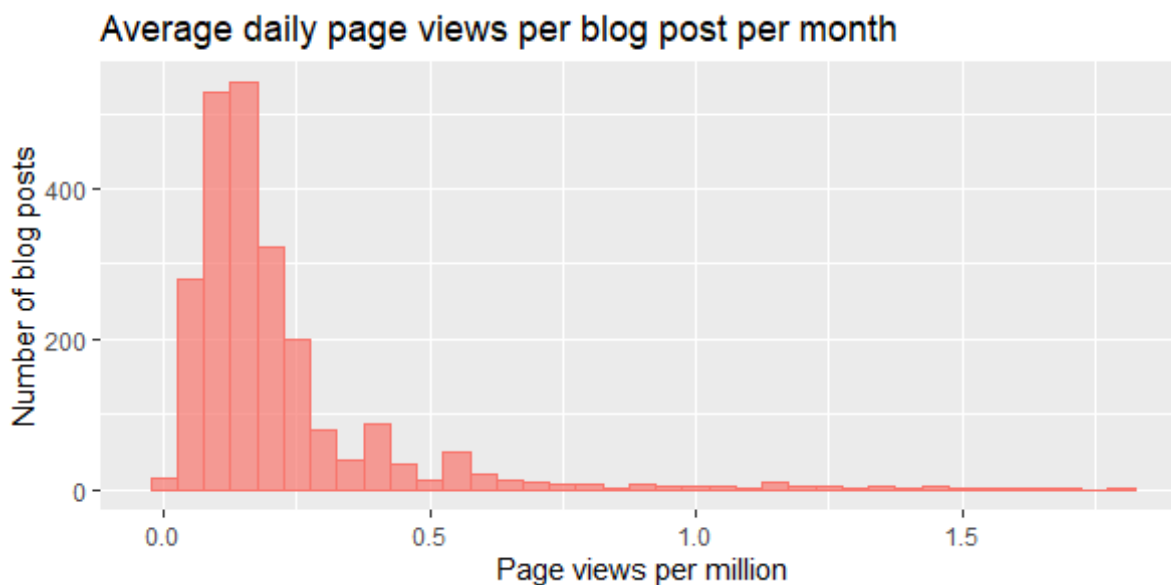


Figure 2.7: Histogram showing the number of blogposts with X page views per million. This plot is based on the average daily page views per month data and binwidth is 0.05 page views.

2.1.5 Data Filtering

In every subsection described above, we have gone through each data source and the steps to preprocess and filter them, making them ready for analysis. We started with 12 109

blogposts from Google Analytics and ended up with 416 blogposts after retrieving traffic data. All the data filtering steps are summed up in Table 2.2. Despite the significant decrease in the number of records, we still believe the data to be representative and relevant for analysis. The ambassadors matched with blogs and used in this thesis make up 52.91 % of the total sales of ambassadors. The remaining ambassadors may have been excluded for multiple reasons. Some reasons may be that the ambassador's blog no longer exists or that the ambassador promotes SafetyWing on social media, newsletters, or closed channels.

Step	Action	Blogposts
1	URLs from Google Analytics between 1st of Januar 2019 and 1st of April 2021	12 109
2	Obviously a blog (filtered out Google.com, Facebook.com, job advertisement webpages and so on)	3 392
3	URLs with at least 5 sessions or 1 sale	693
4	Successfully matched to an ambassador	462
5	Successfully web scrape of the blogpost	441
6	Successfully fetched traffic data	420
7	Webpage with less than 1 mean page views per million	416

Table 2.2: The filtering steps taken to arrive at blogposts successfully matched to an ambassador

2.2 Unit of Measurement

To be able to compare ambassadors, it was therefore necessary to adjust sales figures to a scale which accounted for the size of the blog. This would make it possible for us to identify the features of great ambassadors, and focus on the way they promote SafetyWing, without being misled by the absolute number of sales. To efficiently do this, the traffic data retrieved from Alexa Web Information Services, as described in subsection 2.1.4 was matched with the sales figures from Google Analytics as described in 2.1.2.

The traffic data was retrieved for every link that was successfully webscraped for each month in which we had Google Analytics data. As the main analysis in this thesis is on an aggregated level, it was necessary to also aggregate the traffic in order to make it comparable. This aggregation was done by summing the traffic data fetched for each blogpost, in order to take the life time of the blogpost into account. By not doing this,

a 1-week-old blogpost would have had the same traffic as a 2 year-old blogpost, which could create biases towards older posts if they had higher sales. The traffic used is traffic between the first and the last session referred from a blog post, according to data from Google Analytics.

Data from Alexa also introduced a need for filtering out homepages. The reason for this is that the homepage (i.e. blog.com), of a blog will have more visits than an individual blogpost (i.e. blog.com/blogpost/2021). As Alexa only provides traffic data on the blog as a whole, and not individual blogposts, the measure would be the same for the homepage as for any individual blogpost. This is a disadvantage for our purposes and would be hard to take into account. Nevertheless, we have concluded that by excluding the direct link to the homepages, and using the traffic data on individual blogposts only, we gain a better perspective of the relative sales data.

Income also varies among the ambassadors, which was the last thing which had to be taken into account. The reason for this is that SafetyWing targets people buying a subscription or a short term insurance. In addition, the older you get the more your insurance costs. We are obviously interested in these differences as well, as they can help determine the audience of an ambassador. But, by using revenue figures in the unit of measurement - one would create biases towards ambassadors attracting older people buying a subscription. The solution was therefore to use the sales figures from Google Analytics, focusing on the number of users buying insurance coming from each blogpost instead of revenue in dollars.

The unit of measurement created for each blogpost is calculated by dividing the total sales of a blogpost by the total traffic. It can be explained as *the number of sales per page view per million from Alexa*. If a blogpost has 5 sales and total traffic of 2 (taken the lifetime of the post into account), measured as 2 page views per million from Alexa, it will have a value of 2.5. This means that this blogpost will have 2.5 sales per page view per million from Alexa. In other words, the higher value, the more efficient the post is. Hereafter, we will call this measurement *adjusted sales*. It can be represented in the following equation:

$$\text{Adjusted sales} = \frac{\text{Total sales}}{\text{Total traffic}}$$

Figure 2.8 shows the distribution of the unit of measurement among the blogposts. It seems that most posts have between 0.5 and 2.5 adjusted sales and the average is 4.31.

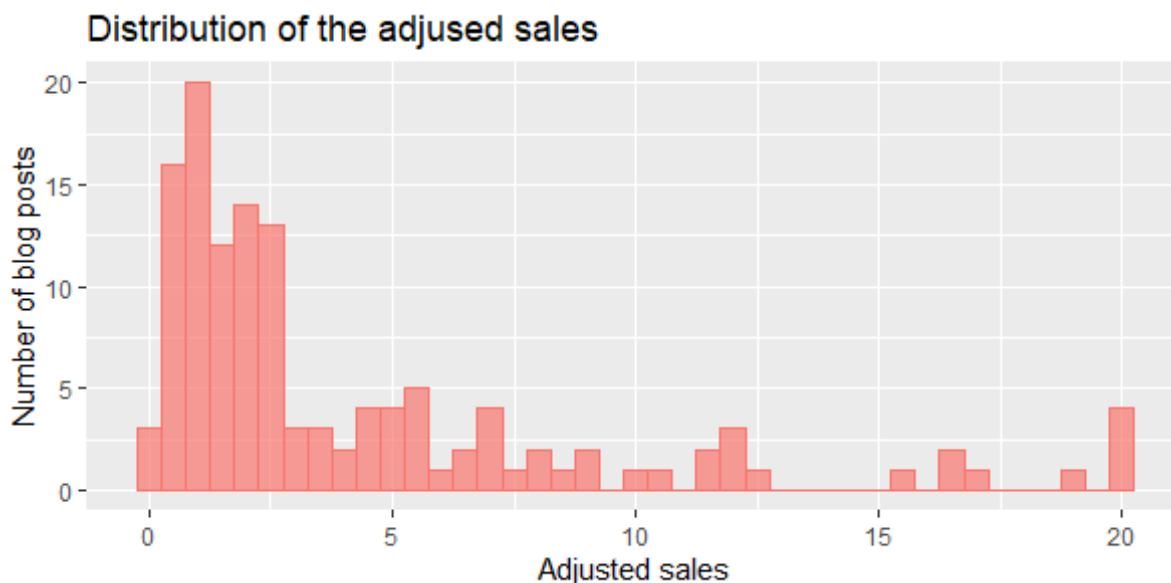


Figure 2.8: Histogram showing the number of blogposts with X in adjusted sales. The binwidth is 0.5 sales. Posts with a value higher than 20 is given with 20 as adjusted sales.

2.3 Extracting Statistics from Blogposts and Categorization

All blogposts refers to SafetyWing in a different manner. While some posts may have an extensive review, others may only mention SafetyWing briefly or not at all. In order to increase the available statistics to be used in the analysis, we retrieved the number of times specific words were mentioned and categorized the blogpost based on the way SafetyWing was mentioned.

The first statistics we collected on the blogposts was a simple count-of-occurrence of 1) SafetyWing, 2) Worlds Nomads, and 3) 18 other competitors, which are listed in the Appendix A. World Nomads was separated as an own count because it was, by far, the strongest competitor when looking at the number of mentions.

From Figure 2.9 we see a histogram of how many times SafetyWing is mentioned in each blogpost on the left. Most apparent is that the majority, well over 150 blogposts, do not mention SafetyWing at all. Likewise, close to half mentioned SafetyWing once. Frequency is rapidly decreasing from 2 mentions and onward. On the right, we see the number of times competitors, including World Nomads, are mentioned. The same trends as for SafetyWing mentions can be seen here, except that it is a bit more evenly distributed,

and there are some more occurrences of the higher mention counts.

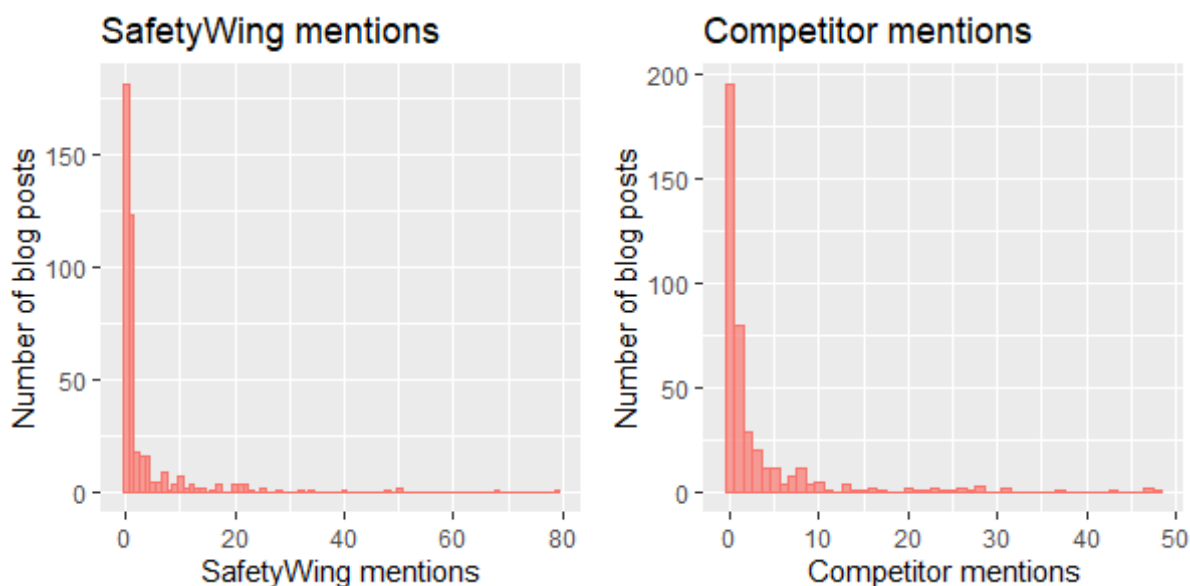


Figure 2.9: Left: Histogram showing the number of blogposts with X SafetyWing mentions. Binwidth is 1 mention. Right: Histogram showing the number of blogposts with X competitors mention, including World Nomads. Binwidth is 1 mention.

Secondly, we divided each blogpost into one of four categories based on where and how often SafetyWing was mentioned in the post. In summary, the categories go from 1, no mention of SafetyWing at all, to 4, SafetyWing is likely the main topic of the blogpost. The categorization can be seen in Table 2.3.

Category	Description
1	SafetyWing is not mentioned
2	SafetyWing is mentioned in paragraphs, but not in the header or subheaders
3	SafetyWing is mentioned in exactly one subheading
4	SafetyWing is mentioned either in the header or in two or more subheaders

Table 2.3: Description of the categorization of ambassadors blogposts

Figure 2.10 represents the previously mentioned categorization and the frequencies associated with it. It is clear that categories 1 and 2 are more used amongst bloggers than categories 3 and 4. This seems logical, as a blogpost where SafetyWing is barely

mentioned or not mentioned is more common than writing, for example, a lengthy review of SafetyWing.

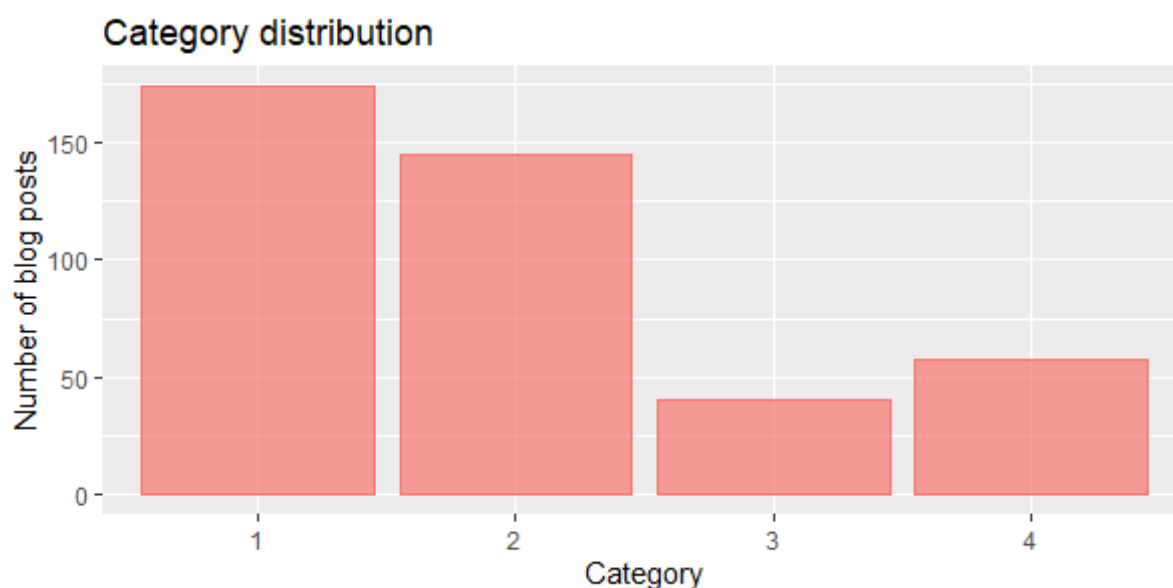


Figure 2.10: Frequency of category occurrences

2.4 Topic Modelling

Topic modeling is a way of discovering the hidden thematic structure in extensive archives of documents. One of the most commonly used topic modeling techniques is Latent Dirichlet Allocation (LDA), with an intuition that topics exhibit multiple topics in different proportions (Blei, 2012).

This method was chosen instead of topic classification, classifying topics based on predefined keywords (Hillard, 2008), as we wanted to see whom SafetyWing reached with their current ambassador program. SafetyWing has no overview of how the ambassadors promote them and whom they are targeting. This means that if we had chosen topic classification as our method - one would be locked in predefined boxes of whom SafetyWing thinks the ambassadors are targeting.

Performing topic modeling on our raw data from the web scraping would, however, produce noisy topics. A study from Martin and Johnson 2015 found that limiting the corpus to only nouns improves topic coherence and reduces model generation time.

The first step in the noise-filtering process was using part-of-speech-tagging (POS-tagging).

POS-tagging is a technique used to automatically assign the part-of-speech, meaning the lexical class, to words using contextual information (Schmid, 1994). Specifically, each word is classified into noun, verb, article, adjective, preposition, pronoun, adverb, conjunction, and interjection based on how the word is used in the sentence. These classifications are again classified into more specific categories, for example, plural and singular for nouns—the purpose of this understanding in which context every single word is used in a text. For example, the word *fish* can have multiple meanings. It can be used as either a singular or plural noun for the aquatic animal *fish* or as the verb *to fish*. Differentiating these different forms is crucial in increasing the accuracy of our filtering.

Our purpose of doing this tagging was to filter out all words but nouns to avoid junk topics - topics lacking coherence. The software used was the TreeTagger software developed by Helmut Schmid at the Ludwig Maximilian University Munich (Schmid, H., 2021). As the POS-tagging in our case filters out all words but nouns, it also handles some of the requirements for preprocessing the corpus before topic modeling. In other words, there was no need to remove stopwords, punctuation, numbers, and whitespace as one usually would have done with a comprehensive corpus. The only necessary steps were to transform the corpus to lower case and stem the words.

Table 2.4 depicts the result from the topic modeling with 6 topics. These show the 10 most frequent words for each topic. From these frequencies, we have manually created a custom name for each topic. For example, topic 4 mentions words which are typical for travel such as *car* and *accommodation*. However, there seems to be a focus on affordable travel such as *hostel* and *budget*. We have therefore concluded to give topic 4 the custom name *Budget Travel*. Lastly, each topic has a percentage indicating how many of the blogposts are classified to each topic.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
<i>Digital Nomads</i>	<i>Travel Knowhow</i>	<i>Insurance</i>	<i>Budget Travel</i>	<i>General Travel</i>	<i>Backpacking</i>
16.79 %	15.03 %	20.29 %	15.89 %	17.34 %	14.66 %
work	visa	coverag	card	citi	backpack
busi	passport	polici	car	beach	life
job	flight	plan	budget	food	road
lifestyl	applic	health	money	island	list
servic	tourist	today	fee	area	dont
life	document	emerg	price	tourist	bag
product	step	hospit	tip	restaur	adventur
hous	test	case	week	tour	water
team	hour	claim	accommod	view	van
communiti	email	provid	hostel	night	bit

Table 2.4: The topics introduced from topic modeling with their respective top 10 keywords. The name in italic is our custom name for the topic, and the percentage is the proportion of the texts with the respective topic.

2.5 Correspondence Analysis

Correspondence Analysis is a technique to explore the relationships among categorical variables. The goal of exploring the relationships, is to find associations among the variables (Sourial, 2010). The input of the correspondence analysis is a contingency table, which in our case consists of topic proportions as columns and ambassadors as rows. The associations between the variables, the rows, and columns, are plotted graphically with two dimensions. The sum of the eigenvalue of each of the dimensions, as given as a percentage of each of the axis on the plot, is the variance in which the plot explains (Abdi, 2014). To measure the relationship between variables, one can look at two things. Firstly, the further away from the origin, the more discriminating. If one row or column variable is at the origin and another in the bottom left corner, the one in the bottom left corner will be more discriminating. Secondly, the smaller the angle connecting a row and a column to the origin, the more associated they probably are. In other words, one measures the angle between a row variable to the origin and then to the column variable

one wants to investigate (Displayr, 2021).

To perform the Correspondence Analysis in R, we used the *CA()* function from the *FactoMineR* package. The input for this function was a table containing ambassador IDs as rows. The columns were the mean topic proportions of all blog posts written by the relevant ambassador, meaning one column per topic and one row per ambassador. The Correspondence Analysis plot itself was plotted with the *fviz_ca_biplot* function from the *factoextra* package and can be seen in Figure 3.2 in the Analysis section.

2.6 Limitations

During the collection and preprocessing of data and general work with the thesis, we have encountered some limitations. As mentioned in the Introduction, we have limited our thesis to only look at the Nomad Insurance product of SafetyWing and only ambassadors writing an ordinary blog. In addition, additional limitations are summarized in the following subsections.

2.6.1 Implications of Covid-19

Covid-19 has, at the time of writing, been a worldwide reality for more than one year. In terms of this thesis, covid-19 has had the following 2 main implications.

Unreliable Data

In the months that followed the beginning of the covid-19 pandemic, the world in terms of travelling came to a halt. People worldwide were urged to return from their travels and even long-term stays abroad. For SafetyWing, this implied an abrupt and significant drop in paying customers of the Nomad Insurance.

Around the same time as the pandemic erupted, SafetyWing launched a new product, namely Remote Health, which supplies health insurance to remote workers. According to Levanon, G. (2020) the share of remote workers in the US is likely to settle at around 20-30% post pandemic, compared to about 5% pre pandemic. The need for insuring ones employees in a remote work space has therefore never been more relevant.

The conclusion to be derived from the previous two paragraphs is this; the data which this thesis is based upon is subject to the extreme situation which the covid-19 pandemic

is. It has therefore resulted in both abnormal lows and highs. Nevertheless, since Remote Health data has been excluded from our analysis, our data has mostly been subjected to unnatural lows. It should, however, be mentioned that whilst the unnatural low of travel restrictions is likely to be a one-time-thing, the high of the remote work force is likely to continue.

The Inclusion of Covid-19 Coverage in Nomad Insurance

As of August 1st 2020, SafetyWing announced that Covid-19 coverage would be included in the Nomad Insurance. Prior to this, SafetyWing, together with most other insurance providers, had an exclusion clause in their insurance about pandemics. Therefore, not surprisingly, sales rised immediately when SafetyWing included this coverage. This can be said to have caused instability to the sales data, and further limits the significance levels derived from this analysis.

2.6.2 Google Analytics is Limiting Traffic Data from Social Media

Google Analytics gave us a lot of valuable data to work with. Nevertheless, in two periods, 15th of March 2019 to 30th of April 2019 and 3rd to 23rd of October 2020, the Google Analytics service was turned off on the SafetyWing webpage. As a result, there exists no web traffic data for these two periods, hence providing a wrongful picture of the referral data as well as artificially low sales figures. To account for this, these periods were filtered out in all aspects of analysis.

Another challenge was a limitation in Google Analytics preventing us from identifying social media accounts referring to SafetyWing with an affiliate link. In these instances we were able to identify the ambassador through the landing page URL. However, the fullReferrer URL would only contain the source URL of the social media, i.e., Instagram.com and Facebook.com. In other words, we got no indication about the ambassador's account or page in these social media channels. Obtaining these connections would require manual identification of social media accounts. Even then, we could not guarantee that we would be able to collect all data, which in turn could create a wrongful representation of SafetyWing's ambassadors.

Including data from social media could give a wrongful picture on the efficiency of the

blog itself. If ambassador A drives 1 000 sessions and 100 purchases from her Instagram account and only 200 sessions and 10 purchases from her blog, the efficiency of the blog would be artificially high. The solution selected was therefore to exclude the social media figures as we were unable to perform an actual analysis on the quality of the content in social media. The reason for this, is that the content can appear in advanced contexts as text on images, Instagram stories, newsletters or in closed groups where we don't have access. We therefore assume that the quality of content in social media and closed channels, is equal of the quality of the content on the respective ambassador's blogs.

3 Analysis and Results

This section will further analyze the findings from our results to answer the research question: How can SafetyWing improve the efficiency of its ambassadors, and what are the implications for ambassador programs in general?

To answer the research question, we will first identify who the ambassador's audience is, look at how vulnerable SafetyWing is to market fluctuations, and inspect whether and possibly when competitors' presence is advantageous. In continuance, we will, in more detail, analyze how topic and category interact and how sales numbers are affected by different combinations of topic and category. At this point, we will have a collection of responses to the research question. To sum it up, we will attempt to identify recommendations for improving ambassadors that can be applied to the ambassadors of SafetyWing and ambassadors in general. This will aid SafetyWing not only to identify the best ambassadors but, more importantly, guide them in finding new and improving existing ambassadors.

3.1 Examining the Ambassador's Audience

What kind of customers buy insurance from SafetyWing through the ambassador program depends on who SafetyWing's ambassadors are and whom they reach. To get an overview of this and determine ambassador audiences, we have looked at the topics introduced by topic modeling and user data from the SafetyWing database.

3.1.1 Niche Markets

The topics introduced by topic modeling will overview which niche markets SafetyWing is reaching with their current ambassadors. As seen in table 2.4, four out of the six topics are travel-related with a total proportion of 62.92 %. The remaining two topics are either digital nomads or insurance-related. In this section, we will use the word *niche* to describe what topic ambassadors are reaching.

As SafetyWing is selling an insurance product, this topic, topic 3, will naturally appear in many of the texts and in combination with other topics. An ambassador with a

high proportion of this topic will not necessarily have insurance as a niche. Table 3.1 shows the distribution of the number of ambassadors that have a specific topic as their most-used topic - meaning the topic that has the highest proportion among all the texts of this ambassador. When looking at the second column, 28 ambassadors, 41.2 % of all ambassadors, have topic 3 as their most-used topic. By comparing the number of blogposts linking to SafetyWing, ambassadors with topic 3 as the most used have on average only 1.86 blogposts linking to SafetyWing.com while other ambassadors have 5.03 blogposts. This can indicate that ambassadors with insurance as the most used topic have focused on an extensive review of SafetyWing, but are not including links in other posts. As they have so few posts, it would be speculation if we said with a high degree of certainty that they were writing in the insurance niche. One can be more confident with the remaining ambassadors as they have a higher number of posts linking to SafetyWing, but the same concept applies here. As we only have used blogposts linking to SafetyWing.com to determine the topics, we can not confidently say what niche an individual ambassador belongs to - only what niche the ambassador's posts promoting SafetyWing.com belong to. To broaden the horizon, one would need to analyze the rest of the ambassadors' blogposts to categorize ambassadors with higher accuracy. This does, however, fall outside the scope of this thesis.

Topic	Number of ambassadors
1	13
2	4
3	28
4	5
5	10
6	8

Table 3.1: Number of ambassadors distributed on the topics

In figure 3.1 all topics proportions are sorted descending from the largest to the smallest value within each topic. This means that the largest proportion is to the left and the smallest to the right for each topic. The purpose of this plot is to investigate whether some of the topics stick out with a higher or lower proportion in more blogposts compared

to other topics. By looking at the green line, topic 3, one can see that this topic sticks out. While the other lines follow each other, falling with almost exponential decay, the green line remains relatively strong. This fits well with the insights in table 3.1 as described above, that indices that most ambassadors have this topic as the most popular, as this is the topic with the highest proportion in most texts. Remembering that ambassadors with this topic also had fewer posts in general, it can be natural to think these proportions, in many of the cases, illustrates the proportion of insurance in the posts of ambassadors with this topic as most-used.

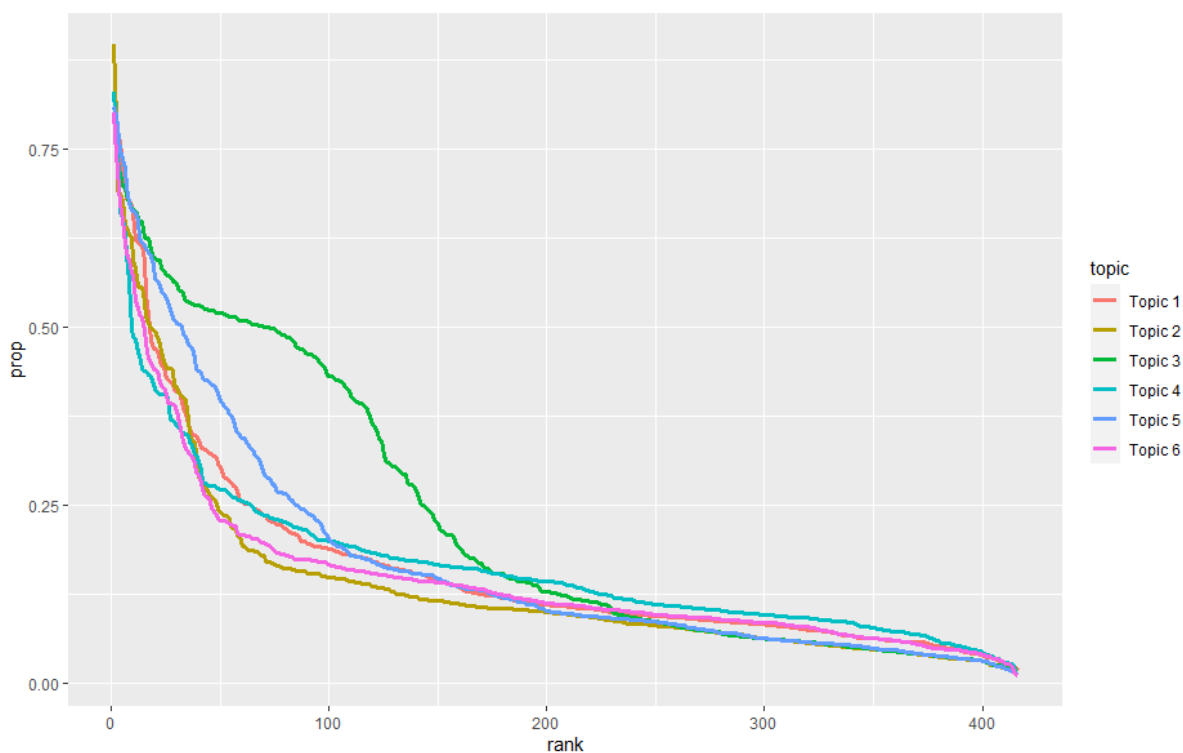


Figure 3.1: The topic proportions of each topic, sorted from largest to lowest

In table 3.2 the most used topics for the ambassadors are in column 1. The remaining columns are the average topic proportions in the posts of the ambassadors, with the topic in column 1 as the most used. For example, ambassadors that mostly use topic 5 have an average topic proportion of 0.091 of topic 1, 0.083 of topic 2, and so on. The insights gained from this table are which topics are used the most together - where green is the largest topic proportion, blue the second largest, and red the smallest. It is pretty interesting that the proportion of topic 4, budget travel, is the highest or second-highest for all topics with a relatively good margin. This may indicate that SafetyWing is described with an economic and budget perspective to recruit potential customers. It is also interesting to

see that topic 1, digital nomads, is the second-highest proportion for topics 2-4. Digital nomads are the cornerstone of SafetyWing's strategy. This topic is so frequently used with the other topics that this type of life, being a digital nomad, is described even in posts classified as travel.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
Topic 1	0.401	0.102	0.114	0.142	0.118	0.123
Topic 2	0.107	0.470	0.101	0.139	0.094	0.090
Topic 3	0.118	0.103	0.477	0.120	0.076	0.108
Topic 4	0.141	0.090	0.142	0.430	0.117	0.114
Topic 5	0.091	0.083	0.064	0.156	0.474	0.133
Topic 6	0.112	0.067	0.080	0.130	0.150	0.462

Table 3.2: Mean topic proportions for the different most popular topics. The most popular combination is highlighted with green, the second most with blue, and the least popular with red.

The correspondence analysis plot between the ambassadors and mean proportion of each topic as shown in figure 3.2 explains 63 % of the variance. This is calculated by summing up the percentage on each dimension, 41.1%, and 21.9 %. The dots in the plot represents ambassadors, while the triangles represent the different topics. As shown in the legend on the right, the gradient color scale illustrates the different revenue percentiles. The green ambassador dots represents the highest revenue percentile (as calculated based on the sales per mean million-page view metric), while the red represents the lowest. To see which values corresponds with each other, there are mainly two metrics: 1) the further away from the origin, the more discriminating, 2) the smaller angle connecting an attribute (mean topic proportions, red triangle) and a value (ambassador, colored dot) to origin the more they are associated.

As one can see in the plot, there is a grouping of green ambassador dots in the bottom left corner. By looking at the angle between these green dots and the *mean_proportion_topic_3* with the origin, one can see that it is smaller than the angle between the green dots and the other mean topic proportions. This indicates that these ambassadors, the ambassadors in the highest revenue percentile, correspond more with a high proportion of topic 3 than other topics. *Mean_proportion_topic_3* is also further

away from the origin than the other mean topic proportions and in a separate part of the plot, which indicates that this attribute is highly discriminating. This fits well with the insights described above that ambassadors writing in topic 3 have, on average fewer posts than ambassadors writing in other topics.

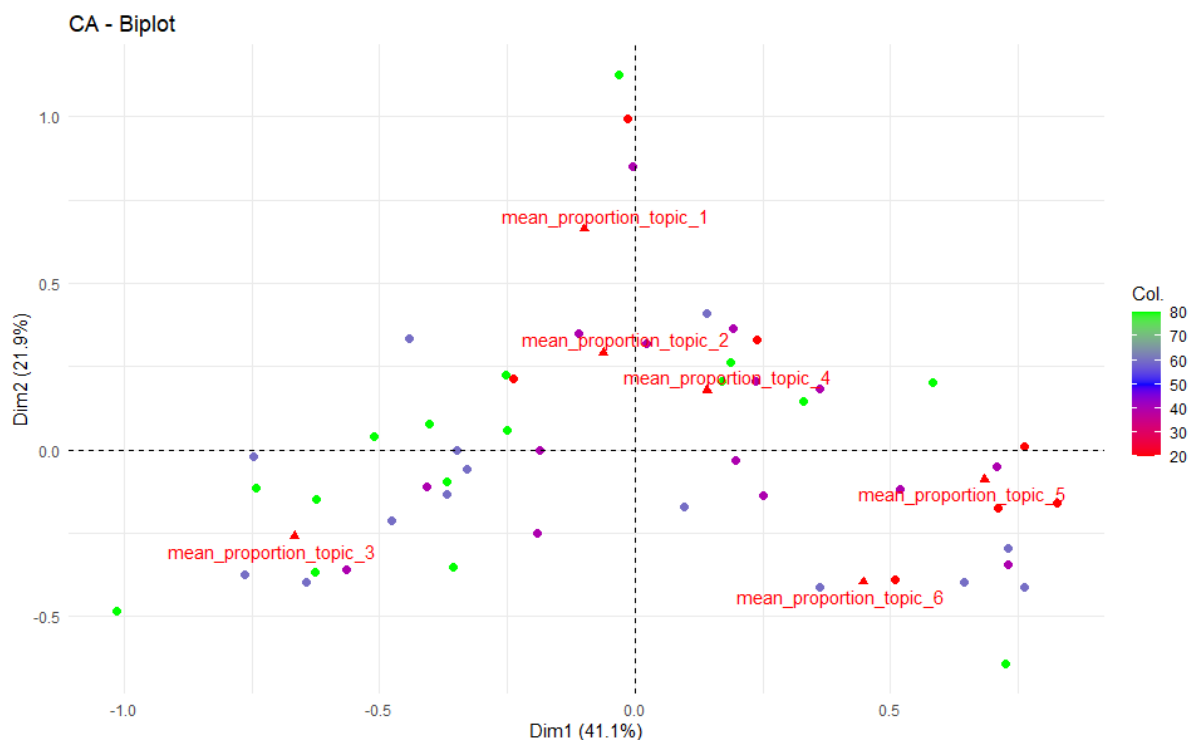


Figure 3.2: Correspondence analysis between ambassador and topic proportions. The color indicates which revenue percentile the ambassador belongs to. Ambassadors are identified by a dot, while a triangle identifies mean topic proportions.

An important sub-note regarding the topics is that they may change over time. This is especially true when the situation with covid-19 stabilizes, and we may see more normal travel again. New niches targeted by SafetyWing also indicate that there may be changes to the topics. It is therefore essential to treat the topics as an indication of the past and not the future. With new niches targeted, new topics can be found by using topic modeling.

To make a more robust and diversified ambassador program, SafetyWing can use the insights gained from the topics to target new niches and change their current strategy. For example, if they have been working on targeting students going on exchange - they can see that they are not reaching these people with their current ambassadors. This is especially important as SafetyWing is not advertising themselves, meaning that they are

dependent on organic growth and ambassadors in the right niches and segments to reach the desired customers.

As mentioned in the background, SafetyWing had historically targeted travel blogs focused on digital nomads before they recently started approaching visa services and online travel agencies. This strategy fits well with the niches indicated by the topics and should not be surprising. Therefore, one can say that SafetyWing has successfully reached the niches they have targeted on an aggregated level. In the following subsections, we will look more into *who* SafetyWing is reaching with their current ambassadors within the different niches.

3.1.2 Regions

In addition to different segments reached, SafetyWing is reaching different regions with the niches. As Figure 3.3 shows, European and Asian customers make up the largest proportion in each topic. However, an interesting aspect is the degree to which this varies. Topic 2, for example, seems to target more the western world, Europe, and North America, while topics 1 and 4 target mostly Asian customers.

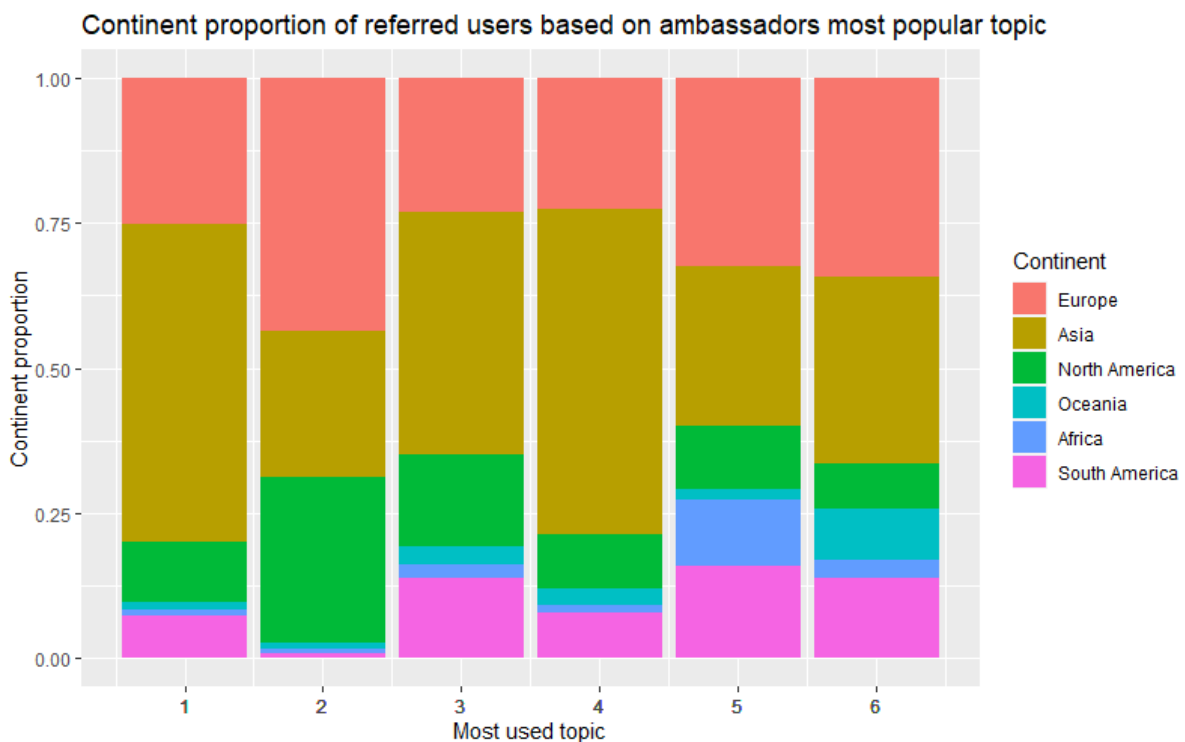


Figure 3.3: The continent proportion of users referred by ambassadors, based on the ambassador's most used topic.

An interesting aspect is that some ambassadors seem to specialize in reaching European and Asian users. However, few or none are specialized on users from the remaining regions as shown in table 3.4. This table shows the distribution of the proportion of continents among ambassadors. A higher percentage in the *Medium High* and *High* columns indicate that ambassadors have a high proportion of customers from these regions. The percentage above the column name indicates the proportion of users from this region. For example, looking at North America, 81.2 % of ambassadors have a proportion of users from North America between 0 and 20 %, column *Low*, among all users this ambassador has referred. Only 3.5 % has a proportion between 80 - 100 %, meaning that only 3.5 % of ambassadors are specialized on users from North America. The exciting aspect is that below 5 % of ambassadors have a proportion of users from Oceania or Africa higher than 20 %. This means that ambassadors reaching users from these regions also reach users from other regions in a similar or higher degree.

Table 3.3

User Share	0-20% Low	20-40% Medium Low	40-60% Medium	60-80% Medium High	80-100% High
North America	81.2%	12.9%	1.2%	1.2%	3.5%
Oceania	96.5%	1.2%	2.4%	0.0%	0.0%
Africa	95.3%	2.4%	2.4%	0.0%	0.0%
Europe	48.2%	25.9%	10.6%	2.4%	12.9%
South America	87.1%	7.1%	0.0%	1.2%	4.7%
Asia	30.6%	21.2%	23.5%	8.2%	16.5%

Table 3.4: The proportions of ambassadors with proportion of users from different regions. The columns illustrate the number of ambassadors with X %, e.g., *Low* between 0 - 20 %, of users from the region specified on the row. For example, 16.5 % of all ambassadors have a proportion of users from Asia, making up 80 - 100 % of their user base. Similarly, 2.4 % of ambassadors have a proportion of users from Oceania of 40 - 60 %.

As a final note, one can say that SafetyWing is reaching the European and Asian market - but struggles with the remaining. According to Laurinavicius, T. (2021) the best destinations for digital nomads in 2021 include, in addition to Asia and Europe, several destinations in South America, as well as North America and Africa. For SafetyWing to reach more potential customers within their target group of digital nomads, one suggestion could be to find ambassadors targeting people in the regions they currently are not

covering.

3.1.3 Age Groups

As table 3.5 shows, we can group the topics into two categories based on the average age of their audience. Topic 1, 3, and 5 with an average age between 35.12 and 37.07 and topics 2, 4, and 6 with an average age between 31.88 and 33.54. It seems that younger people are more open to simple, affordable travel and are less experienced in travel. Meanwhile, older people seem to have more experience and are aware of travel necessities such as insurance. They also appear to be more interested and aware of the possibilities of remote work.

Most popular topic	Average age
1 <i>Digital Nomads</i>	37.07
2 <i>Travel Knowhow</i>	33.54
3 <i>Insurance</i>	36.36
4 <i>Budget Travel</i>	31.88
5 <i>General Travel</i>	35.12
6 <i>Backpacking</i>	33.47

Table 3.5: The average age of users referred by ambassadors depending on their most popular topic.

3.1.4 Genders

Figure 3.4 shows a varying gender proportion depending on which topic is the most used per ambassador. An interesting aspect is that ambassadors writing about insurance (topic 3), general travel (topic 5), and travel know-how (topic 2) tend to reach an almost equal proportion of men and women. This is interesting as these are broad topics and not as specified as the others.

According to a report by MBO Partners (2020), 59% of digital nomads are male and 41% female. Compared with the gender proportion for topic 1, *digital nomads*, in figure 3.4, we can see a significant difference. In a simplified world, the gender proportions in the report and of the ambassadors' users should be approximately equal - if SafetyWing has reached a representative group of digital nomads with the ambassadors. More than 72% of users

who buy insurance from blogs writing about digital nomads are men in our findings. This difference in proportion may indicate a market of remote female workers that SafetyWing currently does not reach.

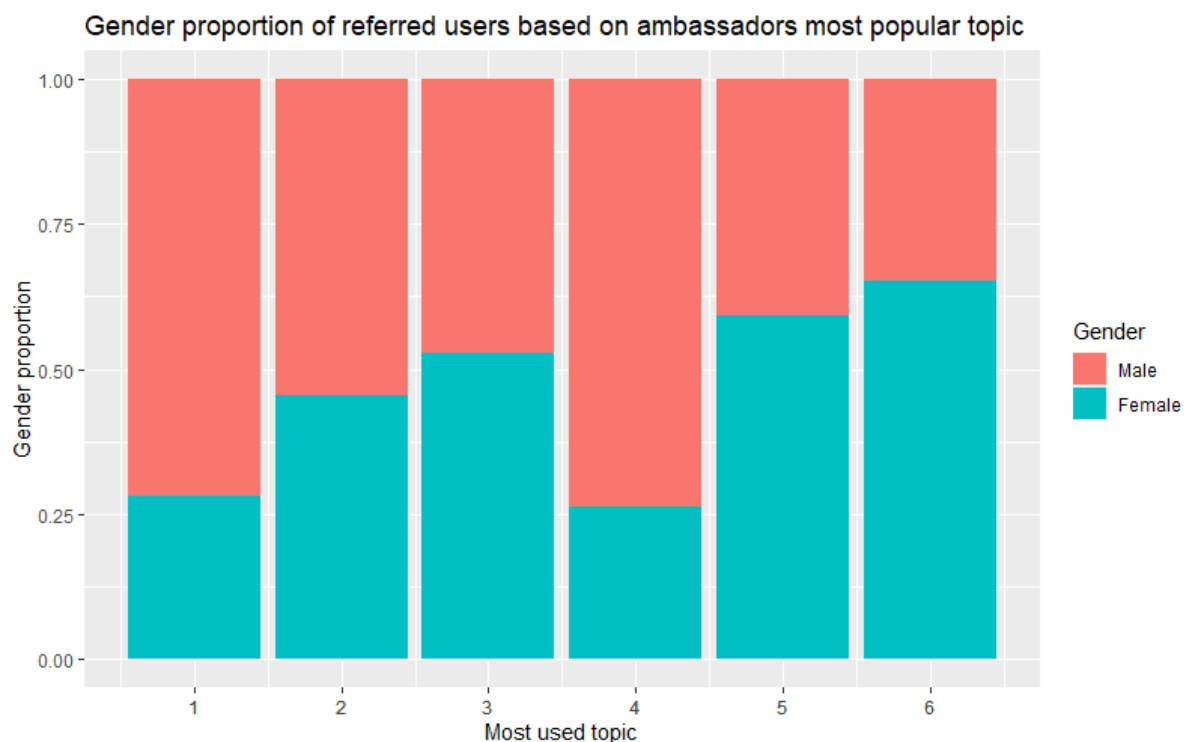


Figure 3.4: Gender proportion of users referred by ambassadors, based on the ambassador's most used topic

3.1.5 Diversification Is a Necessity

Deciding on one segment that is perfect for SafetyWing to target is hard and maybe not meaningful. Based on the arguments and descriptions in this section, one can clearly state that SafetyWing is reaching a relatively narrow audience today. Going forward, SafetyWing could therefore use the intel gathered in this section to focus on recruiting ambassadors in specific niches, targeting specific genders, age groups, and continents. There is great potential in recruiting ambassadors more specialized on users from Oceania and North America and female digital nomads. The intel in this section should help SafetyWing diversify their audience and focus on and evaluate the audiences they see they reach today.

3.2 How does Category and Topic Correlate with Sales

In the following subsection, we will look at different combinations of category X and topic Y. Furthermore, we will again introduce sales per mean million pageviews, hereafter adjusted sales. Adjusted sales bases sales numbers on traffic to the associated blog. This was explained in more detail in 2.2. We will use this metric to investigate how adjusted sales are linked to different categories and topics.

3.2.1 Identifying Distinguishable Combinations of Category and Topic

In Figure 3.5 we see the adjusted sales metric associated with category X and topic Y. Most noticeable is category 4, where at least 4 of the highest conversion ratios are located. As a reminder, category 4 represents blogposts where SafetyWing is mentioned either in the header or in at least two subheaders. In other words, these posts are likely to, at least, have a significant discussion of SafetyWing. The question which remains is whether it seems reasonable that category 4 has such high adjusted sales numbers.

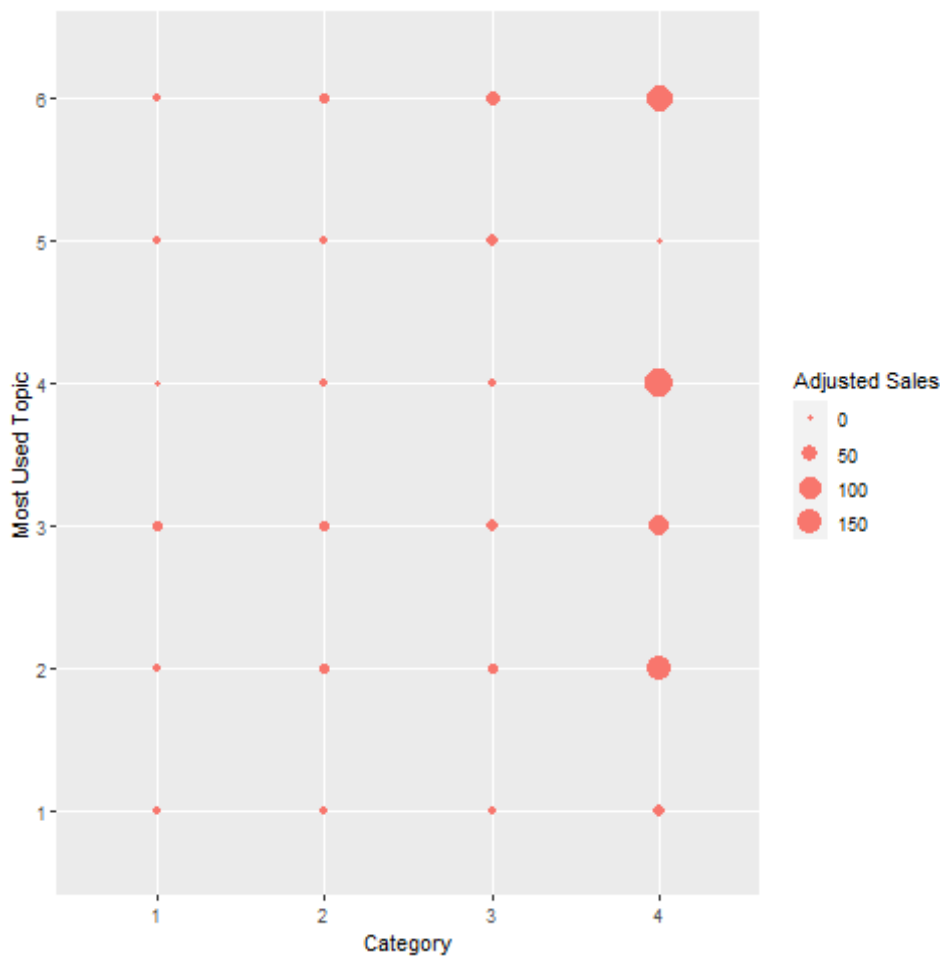


Figure 3.5: Category VS Topic by the adjusted sales measure

For example, the posts in category 4 could be a review of SafetyWing alone or a review of SafetyWing and one or more of their competitors. This implies that, for the readers who press the SafetyWing.com link, many are likely to be aware of where they are headed. Some might even have already decided to buy SafetyWing insurance. Solely from this argumentation, it seems probable that the conversion ratio for these posts is higher. However, the category seems especially correlated with 4 of the 6 topics, namely topics 2, 3, 4, and 6. Topic 3, which represents *insurance*, seems reasonable to be combined with a review on SafetyWing and other insurance providers/products. Meanwhile, topics 2, 4, and 6 relate to travel somehow, and the connection needs to be investigated further.

Topic 2, *travel knowhow*, represents posts where it is explained and given tips on how to travel and what one needs to think about. For example, if one is going to visit another country, one might need a visa. Another such necessity might be insurance. Since category 4 only requires either a heading mention of SafetyWing or at least two

subheading-mentions, it could be probable that these posts reach a similar reader group as those in topic 3. Hence, it appears to be reasonable that topic 2 has a higher adjusted sales metric than average.

Furthermore, topic 4, *budget travel*, also has a high adjusted sales metric. Although travel insurance is often thought of as expensive, many find it necessary, even budget travelers. From reading some of the reviews on SafetyWing and its competitors, we have found that one of their primary diversification is their affordable price. Hence, from similar reasoning as for topic 2, it also seems reasonable that topic 4 both focuses on SafetyWing in their posts and that their readers are genuinely interested when clicking the link to SafetyWing.com.

The last topic to discuss concerning category 4 is topic 6, *backpacking*. Backpackers can be said to have different reasons for choosing that exact form of travel. Some might do it simply because it may be considered an affordable way of travel. For those, the argumentation from above about topic 4, *budget travel*, can be said to apply. For others, backpacking could be a way of living a simpler life, perhaps even in combination with working remotely. For those, a similar argumentation as that of topic 2 may apply. However, their motivation may be more towards the remote aspect that is SafetyWing.

The reason for the low adjusted sales metric for topic 5, combined with category 4, is perhaps not so evident at first glance. Likewise, as with some of the other topics, topic 5 is about travel. However, more specifically, it covers *general travel*. To put it another way, there is no specific aspect of traveling that explicitly connects to SafetyWing. Therefore, all types with general travel interests in joint may be readers of these blogs. It, therefore, seems reasonable that the adjusted sales metric is lower in this instance.

Lastly, but not to be dismissed, is topic 1 combined with category 4. This topic revolves around *digital nomads*, which coincides with the target group of SafetyWing. The low metric may indicate that SafetyWing does not fully reach this target group. The reasons for this would be pure speculation. However, the low metric remains and indicates that SafetyWing should consider how they approach these kinds of ambassadors. For example, they might actively track down and reach out to these ambassadors. Another option could be to more actively collaborate with existing ones.

From the three remaining categories, the differences are more subtle. We see that for category 3, topics 2, 3, and 6 are somewhat more prominent than the remaining topics. Since category 3 is, to some extent, similar to category 4, we would assume that the reasonings would be similar as well. The lower extent of the metric might be since SafetyWing is likely less visible and less mentioned than in category 4.

For categories 1 and 2, topic 3 stands out. Since SafetyWing offers insurance products, and topic 3 revolves around insurance, this is not a surprise. Posts about insurance do not necessarily have to be a review. It could also be a story of how insurance was necessary at some point in time, or perhaps a post about other insurance types that SafetyWing does not offer. However, people reading these blogs are likely to have some interest in insurance, given that they are reading about it to some extent. However, since these posts are not about SafetyWing to a considerable extent, it can also be deemed unlikely that there is strong potential for increasing the adjusted sales metric.

In conclusion of this subsection, category 4 is most prominent in terms of adjusted sales. Most noticeable are topics 2, 3, 4, and 6, which already perform well on their own, and we, therefore, do not recommend an increased focus from SafetyWing. Topic 5 is weak in terms of the adjusted sales metric. However, as argued above, it seems unlikely that there is potential for an improved metric. Topic 1, however, seems to have an undeserving low metric and hence has potential for increasing their sales relative to traffic. For categories 1-3, we found there to be minor variation in the adjusted sales metric, and hence we did not find it fitting to draw any specific conclusions.

3.3 Are SafetyWing's Competitors a Threat or an Opportunity?

In the following section, we will consider how the presence of competitors in blogposts affects sales and compare the effect of the different topics and categories. Before we continue, we find it necessary to establish who SafetyWing's competitors are. Firstly, as a travel insurance provider, there are, in theory, countless competitors. Nevertheless, for their target group of digital nomads, the possibilities are few. As mentioned in section 2.3, we have chosen to base SafetyWing's main competitors on the competitors mentioned

in their ambassadors' blogposts. From this, we have found that World Nomads is, by far, SafetyWing's strongest competitor. Furthermore, we have collected a list of 18 other competitors who have been considered in our competitor analysis below. These are listed in Appendix A.

3.3.1 Is the Presence of Competitors Profitable?

According to (Ellicott Development, 2018), proximity to competitors is profitable in some circumstances. One example is when the product one is offering differentiates itself from the competitor's product in a way the market responds. The example mentioned in the article is about two car dealerships that offer different car models. The customer is likely to enjoy having options. However, more importantly, they are likely to visit a location where they know cars are for sale when they are in the market for a car.

The article from the previous paragraph discusses the traditional scenario of physical markets in "brick-and-mortar establishments". Although our scenario is somewhat different, we assume that similar traits are applicable.

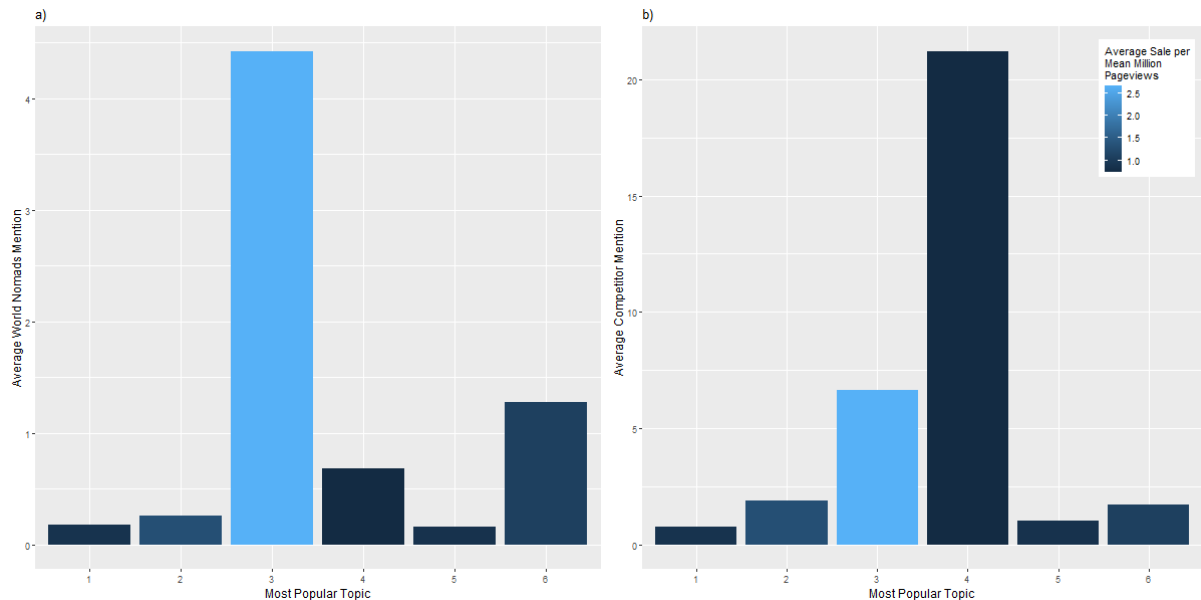


Figure 3.6: Topic VS Average Mention of Competitors VS Only World Nomads per post by average sale per post

By considering graph a) in Figure 3.6, we see the interaction between topic and the average number of times World Nomads is mentioned per blogpost. Furthermore, the bars are filled with a lighter shade of blue, indicating a higher adjusted average sale (both

the blog's traffic and lifetime have been taken into account). Topic 3 stands out. The conclusion from this graph would be that mentioning World Nomads in a blogpost is highly profitable.

If we now consider graph b) in Figure 3.6, we see the same x-axis and color fill. However, the y-axis now represents all competitors (see Appendix A). For topic 3 we witness how the average mention has only increased by approximately 1. Meanwhile, topic 4 reveals an incredible increase of about 20 mentions. For this topic, the average sale is very low. One possibility could be that since a high mention of World Nomads increases sales, lower mentions decrease sales. However, another scenario could be that the total mention of other competitors is too high. In order to investigate this further, we will add category as a factor to the discussion.

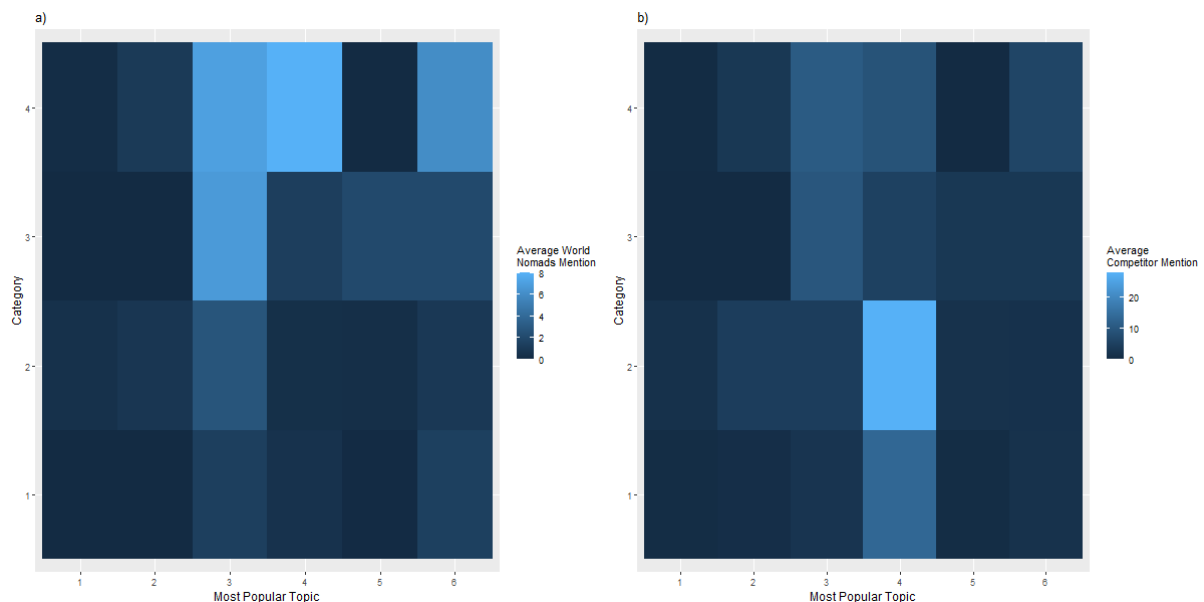


Figure 3.7: Topic VS Category by Average Mention of World Nomads or All Competitors

Graphs a) and b) in Figure 3.7 divides the blogposts into topic X and category Y. Moreover, it provides the average mention of either World Nomads (a)) or all of SafetyWing's competitors (b)). A lighter shade of blue indicates a higher number of mentions. Topic 3 shows how there are more mentions of World Nomads correlating with a higher level of category. This could, for example, indicate that blogposts where SafetyWing is in focus also have World Nomads in focus. In other words, they could be an extensive post about the differences and similarities of the two competitors. If we now again look at 3.6, we are reminded that topic 3 is the most effective in terms of adjusted sales. By reviewing

all arguments, we find that there is a high possibility that by comparing SafetyWing to its main competitor, World Nomads, sales increase. In other words, as mentioned in the intro to this subsection, proximity to competitors can be profitable.

If we again consider graphs a) and b) from Figure 3.6, we see that topic 4 has low adjusted sales. Furthermore, they have a higher high mention of competitors. Graph b) from Figure 3.7 also reveals that most of these blogposts are either category 1 or 2, with a focus on category 2. Together, this is likely to imply that many of the topic 4 blogposts with low adjusted sales are when SafetyWing is mentioned in a list, or similar, along with many other competitors. There are several possibilities as to why this is disadvantageous. Firstly, it could simply be too many competitors to choose from. Secondly, SafetyWing could be mentioned too far down so that others are preferred. Thirdly, there could be too little written about the companies that seem credible. These are all possibilities that should be investigated further in a more detailed analysis.

What we have found in the previous paragraphs can be summarized in the following way. There is a high likelihood that blogposts with the main focus on SafetyWing, such as a category 3 or 4 blogpost, benefit from comparing their main competitors, World Nomads. However, posts which focus on a large variety of insurance providers are not as likely to be profitable. This insight might be used by SafetyWing when advising ambassadors of beneficial and profitable methods.

3.4 Investigating SafetyWing's Vulnerability to Market Fluctuations

With mainly two niches reached with the ambassador program, SafetyWing can be vulnerable to market fluctuations. In this subsection, we will look closer at how the different topic proportions, niches, have developed over time, as well as the sales/sessions ratio.

As Figure 3.8 shows, there has been a shift in which topics dominates. The figure is based on the topic proportions of the blogposts that have generated traffic to SafetyWing each month and not the publishing date of blogposts. The reason for this is that we consider blogposts to have a long lifetime due to, among others, search engine optimization driving

traffic, as well as references from other blogs and blogposts.

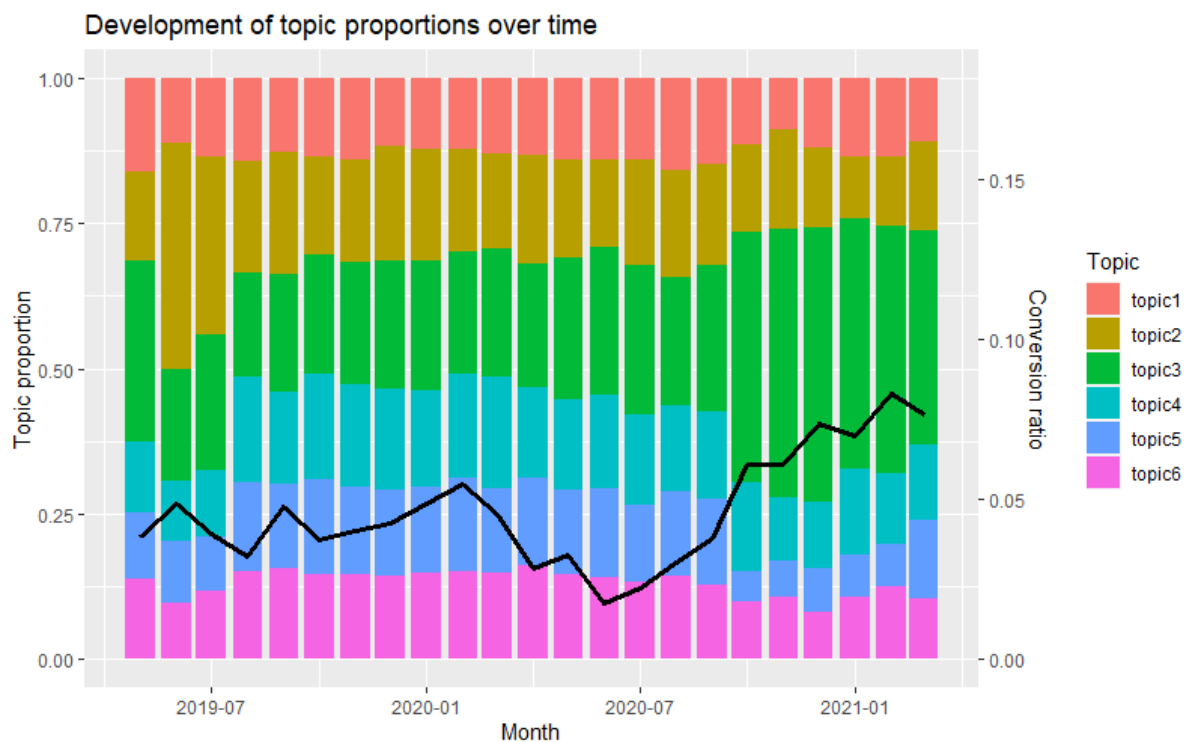


Figure 3.8: Topic proportions and Sales/Session ratio over time from May 2019 to March 2021

Between August 2019 and September 2020, topic proportions appeared to be stable and approximately equal. Nevertheless, in October 2020, the proportion of blogposts within topic 3 - *insurance* - nearly doubled in size and subsequently claimed the position as the dominant topic. The interesting insight gained here is that this growth mainly comes by stealing proportions from one of the niches identified in subsection 3.1 - the travel topics, topics 2,4,5 and 6. Topic 1, digital nomad, has been relatively stable throughout the period indicating steady demand compared to the other topics.

We also see that the Sales/Sessions ratio started decreasing in March 2020, represented by the black line-graph in Figure 3.8, before reaching a low point in June 2020. The changes in this ratio can be explained by people still wanting to read blogs during the pandemic and check out relevant products, like SafetyWing, but they do not need it. The sessions driven to SafetyWing.com are therefore not as efficient as they once were. This started climbing in together with the rise of the insurance topic proportion.

Compared with the insights discussed in section 3.2 above, the insurance topic tends to

appear together with category 4. This trend can be confirmed in Figure 3.9 which shows an increasing trend of category 4, similar to that of topic 3 as seen in Figure 3.8. It is important to note that this plot is not showing new blog pots written but the proportion of the categories of the blog posts that have led to a sale in this period. Meaning, older blog posts that have led to a sale recently will be a part of the most recent data. It is interesting to see that both category 3 and 4 have gained terrain at the same point as topic 3, at the expense of mostly category 1 and partly category 2. Categories 1 and 2 have gone from being the most dominant categories to being almost eradicated in the last months.

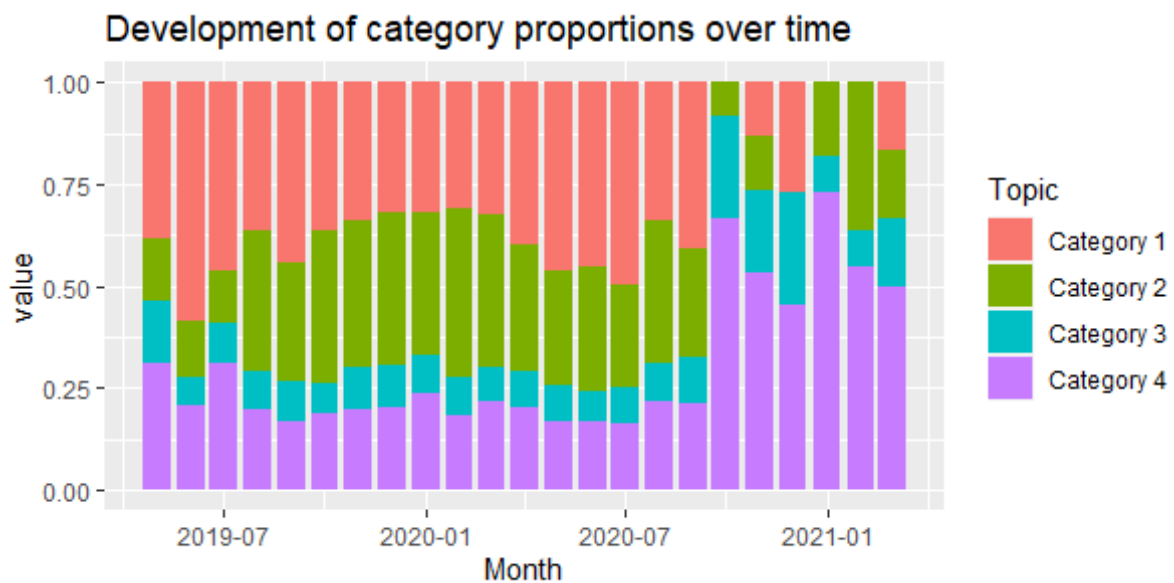


Figure 3.9: Proportions of the different categories over time, from May 2019 to March 2021

There may be multiple causes for this shift in topics and categories. First of all, as mentioned in the introduction, SafetyWing introduced Covid-19 coverage in its insurance products in August 2020. Therefore, one hypothesis could be that the increase of topic 3 and category 3 and 4 is due to the modification of the insurance product. Figure 3.10 shows us the number of sales per month, and we can see the implications of this inclusion in the significant increase in the number of sales in August and September. October 2020 is included in the graph, but due to the limitation described in 2.6.2, it should be ignored as the service was partly turned off. The way ambassadors have described SafetyWing may have focused more on the coverage and details surrounding it and therefore also increasing the proportion of category 3 and 4. One possibility could be that ambassadors were not

sufficiently notified about the change, and therefore did not know how to increase sales. However, the increase in sales in August does not fit with when the topic proportions shifted towards topic 3 and category towards category 3 and 4. The sale increase happens 2 months earlier, in August, which destroys the idea of a shift due to the modification of the insurance product.

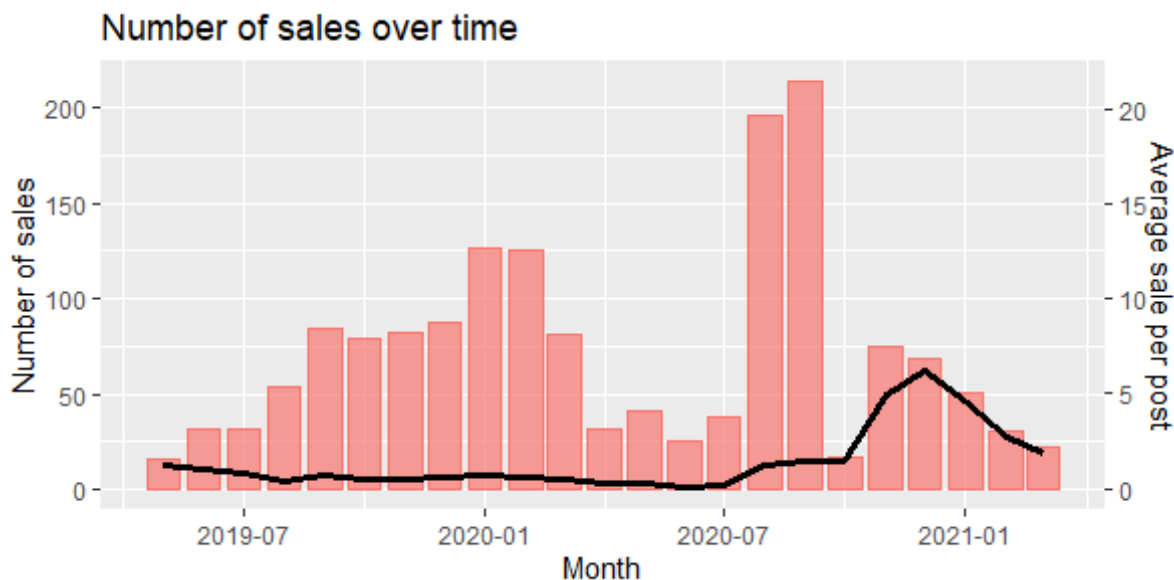


Figure 3.10: The barplot illustrates the number of sales each month, with the y-axis to the left. The black line represents the average sale per post, with the y-axis on the right. Due to limitations that led to missing data in October 2020, this month should be ignored.

It's, therefore, necessary to look closer at the number of posts and the average sale per post. As seen in Figure 3.10, the average number of sales increased at the same time as the topic 3 proportion increased. As the average increases significantly while the sales only dropped slightly, the number of posts referring to SafetyWing must have decreased. This may have multiple reasons. First of all, posts belonging to other categories might have referred fewer than 5 sessions and were therefore filtered out. Secondly, people's interests may have shifted away from the other topics. When reading a post about travel, one might be hesitant to click on a SafetyWing travel insurance link because travel will not happen due to covid-19. The last argument is purely speculation and hard to verify but still a point that should be considered. Another interesting fact is that blog posts of category 4 have a longer lifetime than those of other topics, as seen in Figure ???. As argued previously, topic 3 and category 4 often appear together. A longer lifetime means

that posts in category 4 refer traffic, and possibly sales, to SafetyWing for a longer period than other posts. This means that the proportions of category 4, and possibly topic 3, in Figure 3.9 and Figure 3.8, consists of older posts than the other categories and topics. Therefore, ambassadors have to produce fewer posts in category 4 to still impact on the topic proportions as these posts will contribute to the statistics for a longer period.

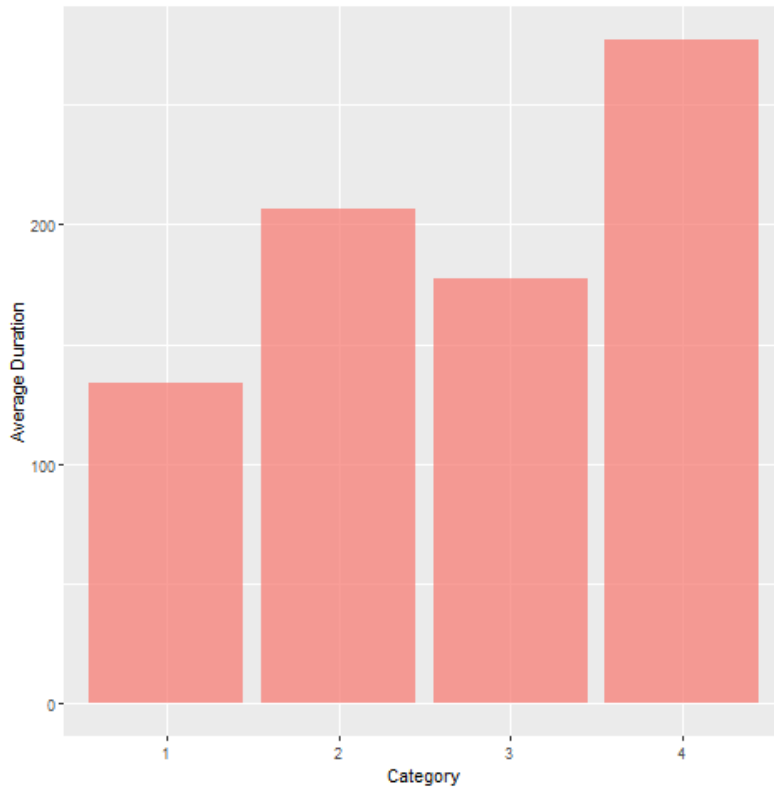


Figure 3.11: The average lifetime of blog posts within each category. The lifetime is based on the duration between the first and the last time a blog post has referred traffic to SafetyWing.

To round this subsection up, one can conclude that SafetyWing is vulnerable to market fluctuations, which the covid-19-pandemic has shown. While the digital nomad niche stayed stable throughout the period, the travel segment lost proportions to general posts about insurance. This shift also was evident in regards to the category, where categories 3 and 4 gained terrain at the expense of categories 1 and 2. Interestingly, the efficiency of each click on a SafetyWing link increased together with the insurance topic. This means that ambassadors in this period also have managed to readjust a new market situation.

3.5 Identifying Recommendations for Improving Ambassadors

The less traffic a blog has, the less traffic it will drive to SafetyWing. As blogs participate in the SafetyWing Ambassador Program, it's naturally to think they do so to earn money on sales. But, the less traffic referred to SafetyWing the less chance of earning money. This raises a natural question, *how will the size of the ambassador's blog affect the way the ambassador should promote SafetyWing?*

Blogposts categorized as category 1 or 2, focuses to a small extent on SafetyWing and therefore gives SafetyWing less of the readers attention. The less SafetyWing is mentioned and used in the post, the less of the users attention is given to SafetyWing. As seen in table 3.6, the calculated sessions per million page view globally varies a lot between the categories. As seen, category 1 and 2 are substantially lower than 3 and especially 4. What these numbers mean, is that a blogpost falling in category 1 can expect to drive 0.5837 % *sessions*, clicks on a link to SafetyWing, per page view. The conversion ratio, meaning *total sales / total sessions* varies with category and topic, as discussed in part 3.2.

Looking at table 3.6 the difference between the categories becomes substantial. Not only do the click ratio increase per category, the conversion ratio also increases. This means that you will get more clicks and a higher ratio of the people actually clicking will purchase the insurance. With other words, ambassadors writing in category 1 have to wait 45.86 times as long as ambassadors writing in category 4 to make a sale.

Category	Click ratio	Conversion ratio	Page views per sale
1	0.5837 %	1.9256 %	8897.19
2	0.5825 %	2.0175 %	8509.12
3	3.025 %	5.5736 %	593.20
4	10.282 %	4.9961 %	194.67

Table 3.6: Sessions driven from categories of blogposts per page view

How an ambassador chooses to write about SafetyWing, depends on the overall income

strategy of the ambassador. Some ambassadors may have a diversified income strategy consisting of multiple partners, but others may be exclusive to SafetyWing. A speculation can be that blogs with multiple partners may be hesitant to write a category 4 post, as it may affect other partners negatively or the blog may be perceived as advertisement-heavy.

A final note is also that an ambassador can't have only category 3 and 4 blogposts, as this may interfere with the ambassador's audience's impression of the blog. People are visiting blogs that give them value, and only writing category 3 and 4 posts about SafetyWing will probably give visitors little value. A combination of the categories and referring to a more extensive review, can therefore be a good solution for ambassadors. Investigating the effect of referring to a category 3 or 4 post in another blogpost, is however outside the scope of this thesis.

As discussed in 3.1, ambassadors with topic 3 as the most popular topic has few blogposts linking to SafetyWing. SafetyWing could investigate this further to see what recommendations these ambassador can get. If the traffic of these blogs are low, SafetyWing should probably ignore them for now - but if the traffic is high a recommendation could be to include a SafetyWing widget or a link to SafetyWing in other blogposts as well. This would classify the blogpost as category 1 or 2, and as table 3.6 shows, this require a huge amount of traffic to drive sales.

48.5 % of the ambassadors analysed in this thesis do not have at least one category 3 or 4 blogpost. These ambassadors have on average 3.94 sales in total, compared with 30.14 for ambassadors with such a post - not adjusted for the size of the blog. One clear recommendation for SafetyWing is therefore to *assure* that ambassadors have at least one blogpost that falls into category 3 or 4. If not, they should use little or no time focusing on these ambassadors - as resources then is more well spent elsewhere.

4 Conclusion

Our mission with this thesis has been to identify how SafetyWing can improve the efficiency of its ambassadors and what the implications are for ambassador programs in general. To answer this research question, we have collected data from various sources, including SafetyWing's user database. All the data has then been cleaned and restructured to the same format for the purpose of analysis. Our analysis methods have been, primarily, textual analysis, topic modeling, correspondence analysis, regression, and visualizations. In the following paragraphs, we will summarize our findings from these analyses.

By examining the ambassador's audience, we analyzed which topics SafetyWing reach, and how these topics covariate with the demographic variables region, age group and gender. To summarize, Safetywing covers six topics, namely *1: digital nomads*, *2: travel knowhow*, *3: insurance*, *4: budget travel*, *5: general travel* and *6: backpacking*. While customers from topics 1, 3, and 5 have an average age of approximately 36.3 years, the average age for customers from topics 2, 4, and 6 is 33. Conclusions drawn from this is that younger people prefer affordable travel and have less experience than older people. Furthermore, topics 1 and 4 are dominated by men, and topic 6 is dominated by women. Meanwhile, topics 2, 3, and 5 reach an approximately equal share. The tendency implies that more specific topics reach either male or female readers primarily. However, broader topics are equally preferable to both genders. The analysis also refers to a survey which shows that 59% of remote workers are men, while SafetyWing reaches 72% men in this topic. Based on the assumption that the two groups, remote workers and digital nomads, mostly overlap, there has been identified potential for SafetyWing to reach more female digital nomads. Lastly, there is an overwhelming share of ambassadors with an audience mainly from Europe and Asia. This, however, excludes an exception of North America, where topic 2, *travel knowhow*, is somewhat popular.

Our analysis on how category and topic correlate with sales revealed the upcoming findings. First of all, category 4 is strongly noticeable in terms of adjusted sales. Specifically, in 4 out of 6 topics, category 4 has a significantly higher adjusted sales number than any other category and topic combination. However, we found that all these four combinations were already performing well independently and did not have any prominent areas of

improvement. Topic 5, combined with category 4, was found to have a low adjusted sales metric. However, due to the topic at hand, there seemed to be unlikely that this metric could be noticeably improved. Lastly, we found that the only topic with potential was topic 1, specifically in combination with category 4.

From our analysis on whether competitors are a threat or an opportunity for SafetyWing, we found the following. There seem to be two main segments of competitors. The first segment includes SafetyWing's main competitors. From our research, we found this to be only one, namely World Nomads. Our analyses revealed a high probability that blogposts that mainly focus on SafetyWing, such as a category 3 or 4 blogpost, benefit from comparing their main competitors, World Nomads. The second segment includes all remaining competitors, for example, standard travel insurance providers. Our analysis was based upon 18 of the most relevant ones (see Appendix A). Our analysis found that blogposts, where many of these competitors were mentioned, had lower profitability. In conclusion, we found this to be insight that may be used to advise their ambassadors about how to promote SafetyWing regarding their competitors.

Looking at how vulnerable SafetyWing is to market fluctuations, one can conclude that the vulnerability is present. During the covid-19-pandemic one have seen a shift from travel related topics towards topic 3, which describes insurance in general. The interesting aspect is that topic 1, digital nomads, has remained stable throughout the period before and during the pandemic. With this shift, one has also seen increasing efficiency of each click - with conversion ratio nearly double compared with the period before covid.

From the summaries we have presented above, we have found it natural to shape our conclusion as an X-part recommendation to SafetyWing. Firstly, we recommend SafetyWing advise all existing and future ambassadors to write at least one category 3 or category 4 post and actively use this as a source of referral from other blog posts. By nudging ambassadors to write a post type proven to be efficient, SafetyWing may see increased profitability and efficiency of its ambassadors. Suppose users refuse to write an extensive review. In that case, SafetyWing should still include them in the program - but time and energy should be focused on other ambassadors as there is little to gain unless the traffic is substantial. Secondly, when recruiting future ambassadors, prioritize those who write in the topic of *digital nomads*. A bonus is if they can advertise with reaching

particularly female readers.

Some of the findings and conclusions directed at SafetyWing specifically can also function as general implications for ambassador programs. The main ones are summarized here. First, as we have seen, the target audience and audience reached do not always overlap. It is therefore beneficial to always keep track of the demographics describing the audience of the ambassadors. Findings may indicate gaps that need to be filled or that the target audience does not exist. Either way, action must be taken to improve efficiency. An effective way of determining the niches of ambassadors is to conduct topic modeling on relevant blog posts. The topic modeling should happen regularly to catch new topics entered or changes in the existing topics. It has to be viewed as a description of the past and present, not a view long into the future. Secondly, keep track of which types of posts your ambassadors write. There should be at least one post that extensively covers what your business is about. Third, last, and related to the previous point, do not be afraid of competitors, especially in review-type posts. As long as the product is good, competitors may increase sales as customers may feel they are presented with relevant information.

References

- Abdi, H., . B. M. (2014). Correspondence analysis.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.
- Brunnahan, P. (2020). Safetywing launches first global health insurance plan for remote workers. Accessed April 15th 2021 at <https://www.lifeinsuranceinternational.com/news/safetywing-health-insurance-remote-workers/>.
- Crunchbase (2021). Safetywing, company financials. Accessed March 8th 2021 at https://www.crunchbase.com/organization/safetywing/company_financials.
- Displayr (2021). How to interpret correspondence analysis plots (it probably isn't the way you think). Accessed May 24th 2021 at <https://www.displayr.com/interpret-correspondence-analysis-plots-probably-isnt-way-think/>.
- Domantas, G. (2019). What is html? the basics of hypertext markup language. Accessed May 26th 2021 at <https://www.hostinger.com/tutorials/what-is-html>.
- Duffy, D. L. (2005). Affiliate marketing and its impact on e-commerce. *Journal of Consumer Marketing*, 22(3):161–163.
- Ellicott Development (2018). Is proximity to your competitors working for or against your business? Accessed May 21th 2021 at <https://www.ellicottdevelopment.com/is-proximity-to-your-competitors-working-for-or-against-your-business/>.
- Glass, E. (2020). What is an html tag? Accessed May 26th 2021 at <https://www.digitalocean.com/community/tutorials/what-is-an-html-tag>.
- Google Analytics (2021). Google analytics overview. Accessed April 8th 2021 at <https://marketingplatform.google.com/about/analytics/>.
- Google Support (2021). Session. Accessed May 23rd 2021 at <https://support.google.com/analytics/answer/6086069?hl=en>.
- Hardy, J. (2016). The history of marketing: From trade to tech. Accessed May 21th 2021 at <https://historycooperative.org/the-evolution-of-marketing-from-trade-to-tech/>.
- Hayes, A. (2021). Digital nomads. Accessed April 15th 2021 at <https://www.investopedia.com/terms/d/digital-nomad.asp>.
- Hillard, D., P. S. . W. J. (2008). Computer-assisted topic classification for mixed-methods social science research. *Journal of Information Technology Politics*, 4(4):31–46.
- Kaur, H. (2020). What is web scraping and how to use it? Accessed March 12th 2021 at <https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/>.
- Keane, J. (2021). Safetywing raises \$ 8 million to expand its health insurance for digital nomads. Accessed March 8th 2021 at <https://www.forbes.com/sites/jonathankeane/2021/01/22/safetywing-raises-8-million-to-expand-its-health-insurance-for-digital-nomads/?sh=33dff71c776b>.
- Klipfolio (2021). Page views per million users. Accessed May 24th 2021 at <https://www.klipfolio.com/metrics/marketing/page-views-per-million-users>.

- Laurinavicius, T. (2021). 52 best digital nomad destinations to go in 2021. Accessed May 24th 2021 at <https://tomaslau.com/blog/digital-nomad-destinations>.
- Levanon, G. (2020). Remote work: The biggest legacy of covid-19. Accessed March 22nd 2021 at <https://www.forbes.com/sites/gadlevanon/2020/11/23/remote-work-the-biggest-legacy-of-covid-19/?sh=73cfac3a7f59>.
- Martin, F., . J. M. (2015). More efficient topic modelling through a noun only approach. *Proceedings of the Australasian Language Technology Association Workshop*, pages 111–115.
- Partners, M. (2020). Covid-19 and the rise of the digital nomad. Accessed May 19th 2021 at <https://s29814.pcdn.co/wp-content/uploads/2020/10/MBO-Digital-Nomad-Report-2020-Revised.pdf>.
- SafetyWing (2021a). Our partners. Accessed March 8th 2021 at <https://safetywing.com/partners>.
- SafetyWing (2021b). Safetywing - insurance for nomads, by nomads. Accessed March 8th 2021 at <https://safetywing.com/>.
- Schmid, H. (1994). Part-of-speech tagging with neural networks. *arXiv preprint cmp-lg/9410018*.
- Schmid, H. (2021). Treetagger - a part-of-speech tagger for many languages. Accessed May 10th 2021 at <https://www.cis.lmu.de/~schmid/tools/TreeTagger/>.
- Sourial, N., e. a. (2010). Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *Journal of clinical epidemiology*, 63(6):638–646.

Appendix

A SafetyWing's Competitors

Table A0.1 represents the 19 most relevant SafetyWing competitors. These were collected through a combination of concordance analysis in Sketch Engine, as well as searching through relevant blogposts.

Table A0.1: Most Relevant SafetyWing Competitors

1	True Traveller / truetraveller
2	IMG / IMG Travel Health Insurance
3	Allianz
4	N26
5	Revolut
6	Travelinsurance.com
7	AIG
8	April international
9	AXA
10	Generali Global Assistance
11	Global Trip Protection
12	John Hancock insurance agency
13	Arch Roam Right / Arch RoamRight
14	Roamright
15	C&F Travel Insured International / C&F
16	Travel Insured International
17	Travelex
18	Trawick / Trawick international
19	World Nomads / WorldNomads
