

Bachelor's Thesis

Information and Communications Technology

2020

Simranjit Singh

# DATA VISUALIZATION AND ANALYTICS

– using a business intelligence tool to design a  
dashboard

BACHELOR'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Information Communications and Technology

2020 | 29 pages, 24 pages in appendices

Simranjit Singh

## DATA VISUALIZATION AND ANALYTICS

- using a business intelligence tool to design a dashboard

The purpose of this thesis was to find a solution to unify data across platforms and automate the data collection and presentation needs of an organization that sets up a business and technology festival on an annual basis. To achieve this, a data analytics dashboard was developed to collate and present the data in an easy-to-read visual form.

The dashboard was intended to make it easier for the organization to be able to track and follow the changes in the business and use the insights to plan and target their sales and marketing campaigns better. For this purpose, user data was collected from the organization's website and social media channels and then imported, cleaned, and plotted into the dashboard using relevant visualizations. Thereafter, the analysis was carried out to gather insights with the help of these visualizations.

The main result was that there were certain trends and patterns observed concerning user behavior, such as the time of interaction and level of engagement that different social media posts received or the keywords that directed users to the website. These insights were significant because the organization could focus on what users found interesting and publish similar content when the users would be active online. This would provide the business with a greater reach and help create a higher impact.

### KEYWORDS:

data analysis, data visualization, dashboard, business intelligence, Power BI, Google Analytics

## FOREWORD

This thesis stems from a project involving interviews and practical work in the areas of data visualization and analytics. It has been written to meet the graduation requirements for the Information Technology program at Turku University of Applied Sciences (TUAS) and has been simultaneously worked on practically and theoretically.

The thesis project was formulated and commissioned by Mari Männistö, Chief Executive Officer (CEO) of SHIFT Business Festival<sup>1</sup>, where I have volunteered for the past three years. The thesis work was carried out under the supervision of Patric Granholm, Principal Lecturer at TUAS.

The project work was challenging but diligence and use of the right tools allowed for the completion of the commissioned task. The readily available support of Mari and Patric was of great significance as well.

I would like to thank my commissioner and supervisor for their guidance and would also like to thank Poppy Skarli, Lecturer at TUAS, for her support in proofreading and perfecting this thesis.

I would also like to express my gratitude to the SHIFT staff for their valuable cooperation in the project work, especially Sameh Katr for his tremendous assistance in pushing me forward during the long work sessions.

Finally, I would like to thank the rest of my professional network that supported me in one way or another, my friends for bearing with me and my busy schedule and my family for keeping me going and having faith in me.

Thank you.

Simranjit Singh

Turku, November 30, 2019

---

<sup>1</sup> SHIFT is an annually recurring business and technology festival that happens over a two-day period. In 2018, the festival saw more than 3000 visitors, over 500 startups and attendees and was run by close to 200 volunteers.

# CONTENTS

<b>LIST OF ABBREVIATIONS (OR) SYMBOLS</b>	<b>6</b>
<b>1 INTRODUCTION</b>	<b>6</b>
<b>2 BACKGROUND</b>	<b>8</b>
2.1 Data	8
2.2 Data Visualization and Analytics	9
<b>3 METHODOLOGY</b>	<b>10</b>
3.1 Business Intelligence Tool	10
3.2 Interviews	11
<b>4 CASE: MARKETING AND COMMUNICATIONS DASHBOARD</b>	<b>12</b>
4.1 Planning	12
4.2 Designing	13
4.3 First Prototype	14
4.4 Errors and Mistakes	20
4.5 Second Prototype	21
<b>5 RESULTS</b>	<b>22</b>
5.1 Suggestions and Recommendations	25
5.2 Challenges Encountered	26
<b>6 CONCLUSION</b>	<b>27</b>
<b>REFERENCES</b>	<b>28</b>

## APPENDICES

- Appendix 1. Questions Template for Interviews
- Appendix 2. First Prototype
- Appendix 3. Final Prototype

## FIGURES

Figure 1. A mock-up of the Social Media Engagement dashboard report.	14
--	----

## PICTURES

Picture 1. The "Get Data" feature in Power BI.	15
Picture 2. Navigator window within Power BI.	16
Picture 3. Views of the imported data with incorrect and corrected data types.	17
Picture 4. Views of imported data with additional columns before and after they were removed.	17
Picture 5. Views of imported data before and after promoting a header row.	18
Picture 6. A visualization filled in with the respective social media values.	19
Picture 7. A graph visualization with the necessary filters added.	20
Picture 8. Location of active website users.	22
Picture 9. Time when Facebook users are online.	23
Picture 10. ER Model of the first prototype.	1
Picture 11. Social media engagement.	2
Picture 12. Users visiting the website and the session durations.	3
Picture 13. Website users by source.	4
Picture 14. Organic searches by keywords.	5
Picture 15. Details of the newsletter activity.	6
Picture 16. Time when Facebook users are online.	7
Picture 17. Blog and news pageviews on the website.	8
Picture 18. Specific pages by their pageviews.	9
Picture 19. Unique ticket sales.	10
Picture 20. Total ticket sales.	11
Picture 21. Popular paths used to land on the Tickets page.	12
Picture 22. ER Model of the final prototype.	1
Picture 23. User engagement across social media.	2
Picture 24. Website inbound sources and user statistics.	3
Picture 25. Organic searches by keyword.	4
Picture 26. Newsletter statistics.	5
Picture 27. Time when Facebook users are online.	6
Picture 28. Blog and news pageviews on the website.	7
Picture 29. Top 25 webpages by pageviews with a button externally linking to the Customer Path.	8
Picture 30. Unique Ticket Sales.	9
Picture 31. Total ticket sales.	10
Picture 32. Top 30 internal webpages leading to the Tickets page.	11

## **LIST OF ABBREVIATIONS (OR) SYMBOLS**

API	Application Programming Interface
BI	Business Intelligence
ER	Entity Relationship
KPI	Key Performance Indicator
UI	User Interface
URL	Uniform Resource Locator

# 1 INTRODUCTION

In the current age of digitalization, businesses have increasingly moved towards integrating technological infrastructure to their operations. Traditional pen and paper organisational setups seem archaic and cumbersome to work with. Digitalization has made operations very easy due to the automation of manual tasks and reduction of errors that tend to come from humans involved in the work pipeline.

While this move comes with many benefits, there are challenges to be dealt with as well. With different types of business solutions just a simple internet search away, it is crucial for businesses to beat their competition by being more visible to their target customers in order to remain profitable.

There are several options businesses take that may seem to overlap but are important both on their own and when paired together with other solutions since their efficacy increases when combined. Some of these solutions can include search engine optimisation, search engine marketing, data analytics, targeted or personalised advertising and active social media presence among many others.

It is now very common that businesses are increasingly trying to beat competition using data that helps in understanding their customer base and addressing areas of weaknesses within the business. To work with this data, many organisations use tools to help visualize and understand trends in their businesses. This thesis examines and implements one such tool to represent and analyse the commissioning party's data in a visual analytics dashboard.

The first chapter of this thesis introduces the subject of this thesis, defines the problem and lays out the objectives for the commissioned project.

The second chapter provides details and definitions on some of the core concepts involved in the theme of data visualization and analytics. It provides the background information necessary for following along with the work involved in this thesis.

The third chapter delves into the methodologies and practical work that was involved and employed. The work carried out and described in this section is a balance of research, self-study and practical work with the commissioning party.

The fourth chapter explains the actual practical work involved in realising the end product. It takes a case study of one of the teams within the organization and their needs with respect to the thesis objectives and problem definition with the realisation of a Marketing and Communication's dashboard.

The fifth chapter discusses the results from the work done in the previous chapter. It ties the problem and objectives to the final output, as well as provides suggestions and recommendations for the improvement of the end product should more resources be allocated to it.

The sixth chapter concludes the thesis and all the work involved therein and focuses on summarising the problem, solution, work involved and the final result of the project.

The commissioning body for this thesis is an organization that sets up an annual business and technology festival that is attended by thousands of people. The organization uses its website and social media channels to reach out to the general audience and provides different information that potential visitors and participants of the festival might require.

The organization had already been collecting data on user behavior and activity on their website and social media channels but did not have anyone assigned to analyse the data and find insights that could help the business. This meant that there was an abundance of data flowing between different channels, which being left unmonitored, valuable insights that could help increase business and sales were being missed. This also hinted at there being no clear strategy behind the data collection with no specific questions being answered nor was there any significant value being generated by the collected data.

To solve for that, this thesis was commissioned with the objective of:

- Designing a data analytics dashboard to unify the organizational data and provide insights.



## 2 BACKGROUND

Work on this thesis project began with the author becoming familiar with topics such as data analytics, dashboards, business intelligence and the technologies available for the organization, visualization and analysis of data. For this, the author read through online articles, journals and whitepapers published by the leading data visualization software tool providers. It was found out that there are a number of readily available software that can be used for the management of data inside an organization, as well as software for more specific data analytics needs (Chang, 2018). The author also watched videos on the same subjects.

### 2.1 Data

Data comprises raw facts and figures that have no meaning or value on their own (BusinessDictionary, 2019) but in business, means information collected from different sources that is used to plan and run a business (Spacey, 2017).

For a while now, businesses have been increasingly integrating digital infrastructure to their operations due to the tremendous strides in innovation with the advancement of technology. This has made it easier and tempting for both small and larger businesses to make the move. The move is influenced by automation, increased speed, efficacy and safety as opposed to what the now traditional manual methods have to offer. Digitalization also provides additional advantages such as reduction in the running costs of the business, as well as the considerable reduction of human error in the daily operations among many others. (Hall, 2019)

Moving a business to the digital means increasingly working with data at various levels of the business pipeline, and that calls for a need to keep up with the flow of data. There are various tools organizations use and require, and each of these tools require data input in order to process and produce relevant and meaningful information for the business. The different tools being used make it difficult to keep track of the flow of data and information, and for that purpose, many organizations keep business dashboards. (McKown, 2016) (Young, 2017) (Rouse, 2018)

## 2.2 Data Visualization and Analytics

Data visualization is the process of plotting raw data in graphs or charts to make it easier to process input and information from the sourced IT systems. The graphical representation of data makes it easier to get a quick overview into the performance of an activity or organization. (Techopedia, 2019) These visualizations allow users to collect value out of the analysis done on the data and use that to scale up the business or identify areas of weakness. These key findings are known as *insights*. (Marrs, 2016)

The visualizations are often grouped together based on predefined metrics and KPIs and presented on a *Business Dashboard*. A business dashboard is a platform that is used to visually manage, track, analyse, and display key information regarding a specific team or metric in an organization (Klipfolio, 2019). There are different types of dashboards in businesses and they focus on different areas. For example, Management or Executive Dashboards, KPI Dashboards, Project Dashboards, Employee Performance Dashboards and many others (iDashboards, 2019) (Durcevic, 2019). Each of these is a type of a graphical user interface intended to collect and display certain data and information for a specific purpose and for a specific team or department to make use of (iDashboards, 2019). To be useful in providing insights, these dashboards implement some form of data analytics.

Data analytics is the “process of inspecting, cleansing, transforming and modelling raw data” with the intention of finding trends, gaining insights and answering questions from that information. (Masters In Data Science, 2019) (Wikipedia, 2019)

The information resulting from the analysis of data is known as *intelligence* and when this intelligence is used to optimize and increase the efficiency of business systems, processes or departments, it is called *Business Intelligence*.

## 3 METHODOLOGY

After having a clear understanding of the background concepts, it was prudent for the author to pick a data visualization tool that would be used in designing the dashboard.

### 3.1 Business Intelligence Tool

Research was carried out into the BI tools available and their suitability for the commissioned project. Some of the more popular options included:

- Power BI
- Qlikview
- Tableau

Following further research into each of these tools, paired with expert advice from a professional in the Data Science industry, it was decided that Power BI would be the best choice for this project.

Power BI is a business intelligence tool developed and maintained by Microsoft. It is a very powerful tool with several great features that make data visualization and analytics very easy for anyone that learns to use the tool. It is similar to Microsoft Excel but has a very powerful query feature that can handle large quantities of data which is crucial for the processing of the type of data that is collected in businesses and organizations. (Scott, 2019)

Power BI also made a better choice since it is cheaper than the other tools available, and this was important for the commissioning party. The tool, being a Microsoft technology, has a familiar user interface (UI) and, therefore, is slightly easier to learn. It also integrates very well with other Microsoft technologies that the commissioning party uses in their daily tasks. (Molag, 2019)

Some shortcomings of the tool are that the licensing cost goes up as the user-base scales up and that there are other tools that have a better range of visualizations available (Scott, 2019).

QlikSense and Tableau were not selected because the former lacks extensive reporting capabilities while the latter has expensive licensing (Beek, 2019) (Haije, 2019).

The author completed a couple of online courses to learn Power BI over a period of two months. The first course was a thorough and detailed set of free tutorials in Power BI Desktop offered by Microsoft, whereas the second course was “Learning Microsoft Power BI Desktop” provided by the LinkedIn Learning platform.

### 3.2 Interviews

While learning the BI tool, the author simultaneously conducted interviews with the heads of the different departments to collect information required for the project. This was carried out over a period of four weeks after the questions to be asked (see Appendix 1) had been finalized.

The questions focused on learning what the team from that department was already collecting, in terms of data and KPIs and how they were using these in their decision-making. The questions also asked the department heads to explain the reasoning behind their choices.

The department heads were also asked to share what tools they were using in their daily duties, and what data and information they wanted tracked from the different organizational channels and tools that they were using.

The information and answers collected from the interviews were processed and it was easy to see a pattern and similarity in needs across the different departments. This made it easier for the author to start planning and designing the dashboards. For the scope of this thesis project, the author will be using the Marketing and Communications Dashboard as an example case to show and explain the work that was required and put into the making of the dashboard.

## 4 CASE: MARKETING AND COMMUNICATIONS DASHBOARD

The commissioning party is active on different social media and has a growing audience. The social media are used as a tool to keep the community engaged and informed about the activities and progress of the organisation. To better understand the community's needs and how they interact with the information being shared on their social media platforms, it was decided that a dashboard should be created dedicated to the Marketing and Communications team.

### 4.1 Planning

An interview was conducted with the Marketing and Communications team in order to find out their needs and how the collected data could be availed to them in a manner and form that would be useful.

Since Marketing and Communication is widespread throughout the digital, it was not surprising that the team had a broad digital footprint. This resulted in their having a lot of requirements and needs, which in turn, catered for a significant percentage of what the other teams needed for their own dashboards.

It was, therefore, thought best to move forward with the Marketing and Communications dashboard. The idea being that part of the implemented solution would be copied to the other dashboards where relevant.

The team had split their answers to the questions during the interview into three different segments covering social media, the website and the newsletter. During the interview, it was found out that the following KPIs were being tracked by the team for the:

- Social media – the opening and click rate for posts, number of page followers and testing variation of content in posts.
- Website – the new and recurring users, page clicks and duration of sessions and the number of button clicks on the tickets page.
- Newsletter – the testing of content preference using A/B testing.

When asked how these KPIs were used in decision making, the uniform answer for all the different channels in use was that any feature or approach that delivered high engagement would be reused. This also answered the question on how the decisions were being based.

The team generously shared that they would like to track across:

- Social media – the likes, comments, follows, shares, retweets, gender, location, best activity and posting times, and impressions or reach.
- Newsletter – the opening and click rate, the buttons or areas receiving clicks, A/B testing and the number of subscriptions in the list.

Finally, the team provided a lengthy list of KPIs that they wanted to track, out of which only a small number of the more important ones were selected. Some of these included:

- Location of users engaging in social media.
- Customer path on the website.
- Keywords being used to land to the website.

## 4.2 Designing

With the information collected from the interview, work could now proceed to the design phase. The questions and answers from the interview were sorted into groups that addressed a similar, if not the same, bigger issue. This made it easier to structure the elements that would be appearing on the same page of the dashboard.

Sketches for the Marketing and Communications dashboard were drawn whilst following the notes and instructions obtained from the department's head. Since there were many KPIs requested by the team, it was very difficult to have only one dashboard report. Therefore, to avoid cluttering the dashboard, it was decided that the dashboard would be made up of different reports, each of which would only have data that was related and relevant to the function of that report. Figure 1 below presents one such mock-up of a report:

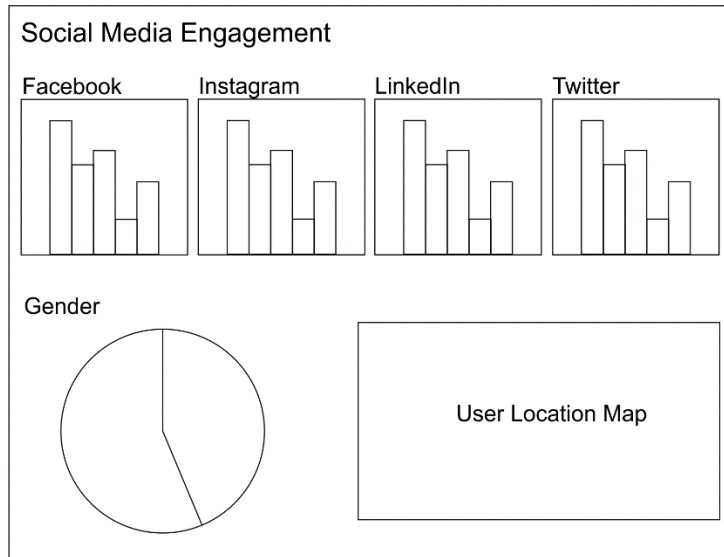


Figure 1. A mock-up of the Social Media Engagement dashboard report.

To avoid confusion and distraction, it was also important to keep the dashboard neat by using as few visualizations and KPIs as possible. The work in this stage also helped identify and plan the different data sources that would be required in the making of the dashboard.

This design process resulted in eleven report mock-ups which would make up the Marketing and Communications dashboard.

#### 4.3 First Prototype

The notes from the interview and the mock-ups sketched were then used to guide the next steps of making the dashboard in Power BI. This followed a clear process as explained below:

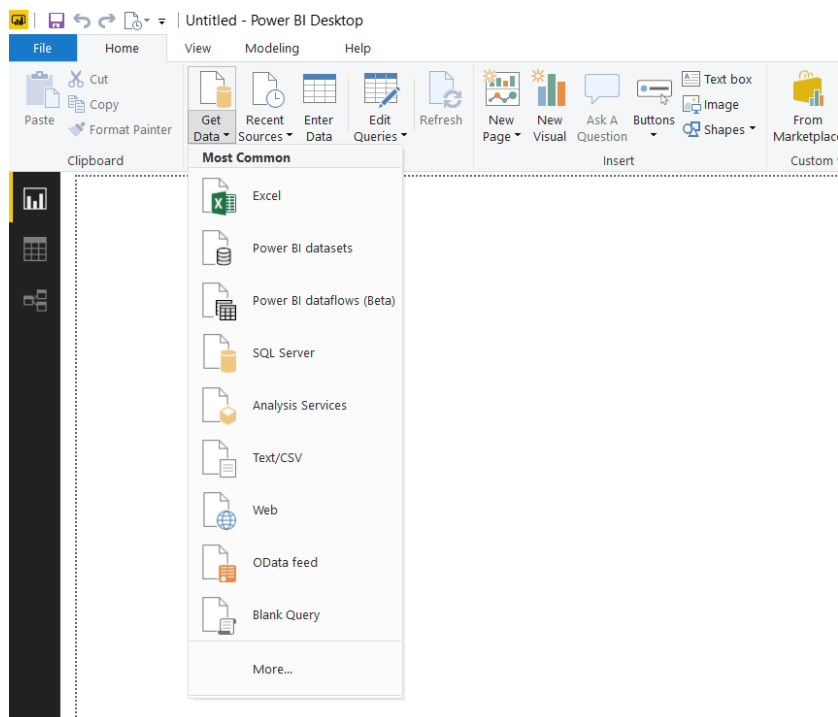
##### Gathering Data

The first step of making the dashboard was to access the data required. From the interview, it was clear where the data that was necessary for tracking the KPIs was being collected and stored. This data was accessible from:

- Google Analytics – All the data from the organization’s website was being collected on this platform. This included user behavior, activity, sessions, clicks, location, device information, customer paths and many more KPIs.
- Social Media – Collected data on user engagement with the organization’s page/channel and posts on that platform.
- Google Drive – The rest of the data had to be sourced from internal organizational files where staff were storing data after manually obtaining it from platforms that either had not automated the export of data or made it difficult to do so. Some of these files included information on user engagement on social media, ticket sales, newsletter subscriptions and more.

The data for the social media engagement was sourced from an Excel sheet from Google Drive. This file was manually filled in with data that had to be scraped from the respective Social Media platforms. This was done so because it was difficult to access the required data straight from within Power BI. Filling in this file was a tedious task that required proper attention to detail in order to avoid human error.

In Power BI, data is sourced by clicking the “Home” tab and selecting “Get Data”. From there one could easily pick the relevant channel from Power BI’s built in options and integrated APIs. Picture 1 below is a visual representation of this.



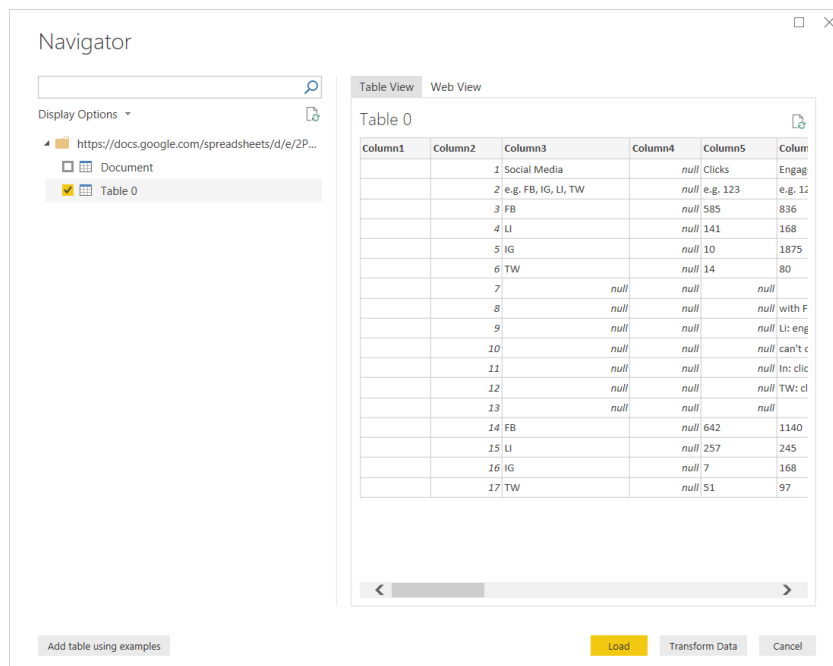
Picture 1. The "Get Data" feature in Power BI.



## Cleaning and Preparing Data

After the data had been accessed through the right channel, it required some *cleaning*. This meant that data had to be prepared for integration to the dashboard without breaking anything in the system.

It was crucial that the imported data sets were uniform and matched where necessary so that the data could be linked across the different sets. This would ensure that the visualizations and the resulting analyses were not affected.

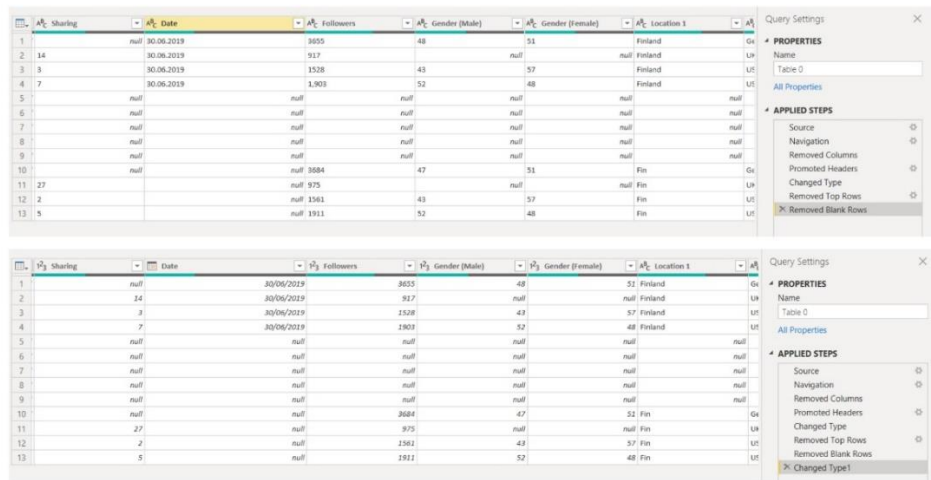


Picture 2. Navigator window within Power BI.

Picture 2 above shows the Navigator window that opens when a data source is selected within Power BI. In this case, the data is being sourced from a URL to Google Sheets. A preview of part of the data set is shown by default.

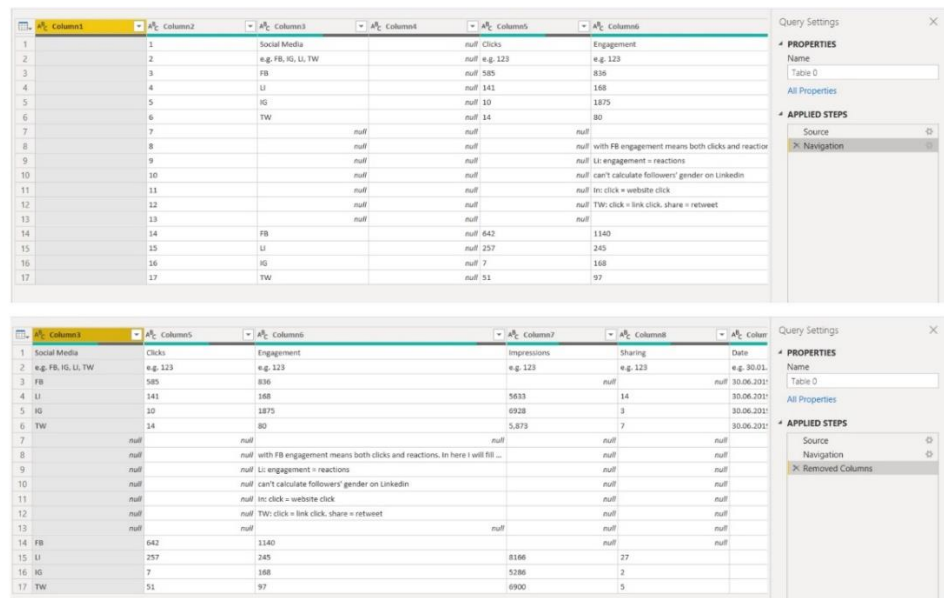
While importing the different data sets, common problems or irregularities encountered along the way included:

- Incorrect data types – some data would be mistakenly represented using an incorrect data type. For example, a set of numerical values was imported as text, a column of dates was imported as numbers and others. Picture 3 below, shows a view of the data being imported with incorrect data types, with another view of the same data with the corrected data types.



Picture 3. Views of the imported data with incorrect and corrected data types.

- Additional rows and/or columns – using Power BI’s integrated API resulted in some data sets having additional rows and/or columns being added to the data set. These were mostly empty and were often attached to either the beginning or end of the data set. Picture 4 below, shows a view of the imported data with two additional columns in the beginning and another view of the same data set after the removal of the column.



Picture 4. Views of imported data with additional columns before and after they were removed.

- Deletion of rows and/or columns – in some cases, there was no option to select the exact data to import from the set, therefore, the entire set was imported and then the unnecessary rows and/or columns had to be deleted. This would sometimes be very time consuming and required keen attention as to which data was unnecessary and could be deleted.
- Promoting headers – a lot of the imported data sets were already stored in a tabular nature, which meant they already had header rows. When the data set was imported, however, these headers were imported as a “normal” row of data. The said row had to then be promoted to be the header row within Power BI.

The image displays two screenshots of the Power BI Query Editor interface, illustrating the process of promoting a header row in a data table.

**Top Screenshot (Before Promotion):**

Column1	Columns	Columns	Columns7	Columns	Column
1 Social Media	Clicks	Engagement	Impressions	Sharing	Date
2 e.g. FB, IG, LI, TW	e.g. 123	e.g. 123	e.g. 123	e.g. 123	e.g. 30.01.
3 FB	585	856		null	30.06.2021
4 LI	141	166	5633	14	30.06.2021
5 IG	10	1079	4928	9	30.06.2021
6 TW	14	80	5,873	7	30.06.2021
7	null	null	null	null	null
8	null	null	with FB engagement means both clicks and reactions, in here I will FB ...	null	null
9	null	null	LI: engagement = reactions	null	null
10	null	null	can't calculate followers' gender on LinkedIn	null	null
11	null	null	IG: click = website click	null	null
12	null	null	TW: click = link click, share = retweet	null	null
13	null	null		null	null
14 FB	643	1140		null	null
15 LI	237	245	8166	27	
16 IG	7	168	5285	2	
17 TW	51	97	4900	5	

**Bottom Screenshot (After Promotion):**

Column1	Clicks	Engagement	Impressions	Sharing	Date
1 e.g. FB, IG, LI, TW	e.g. 123	e.g. 123	e.g. 123	e.g. 123	e.g. 30.01.
2 FB	585	856		null	30.06.2021
3 LI	141	166	5633	14	30.06.2021
4 IG	10	1079	4928	9	30.06.2021
5 TW	14	80	5,873	7	30.06.2021
6	null	null	null	null	null
7	null	null	with FB engagement means both clicks and reactions, in here I will FB ...	null	null
8	null	null	LI: engagement = reactions	null	null
9	null	null	can't calculate followers' gender on LinkedIn	null	null
10	null	null	IG: click = website click	null	null
11	null	null	TW: click = link click, share = retweet	null	null
12	null	null		null	null
13 FB	643	1140		null	null
14 LI	237	245	8166	27	
15 IG	7	168	5285	2	
16 TW	51	97	4900	5	

The 'Applied Steps' pane on the right of the bottom screenshot shows the following steps:

- Source
- Navigation
- Removed Columns
- Promoted Headers
- Changed Type

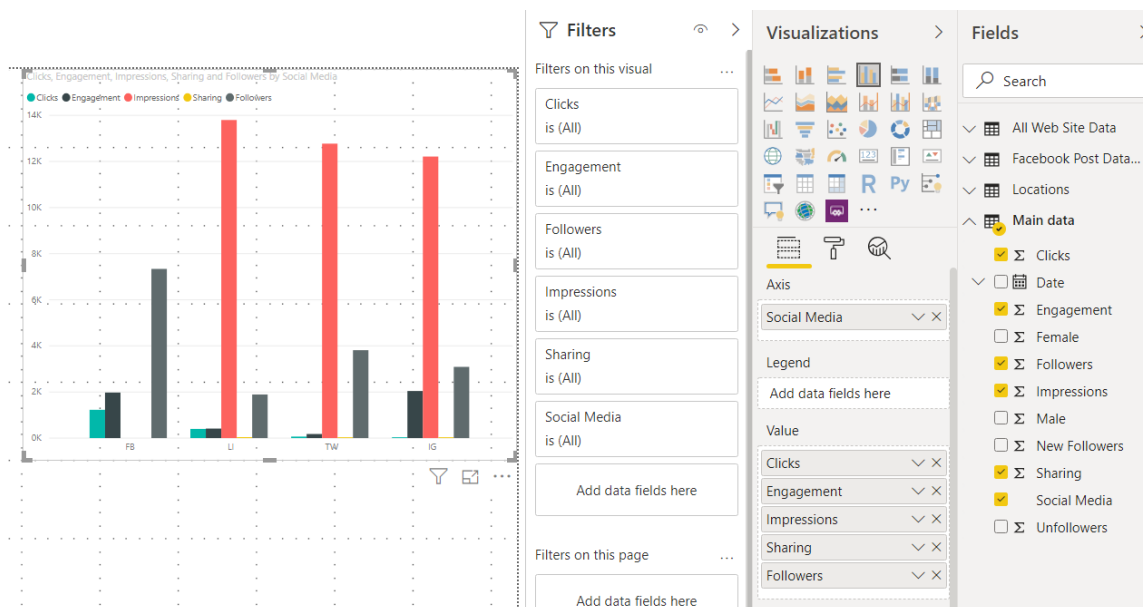
Picture 5. Views of imported data before and after promoting a header row.

- Inconsistent data – some data sets, when imported, did not “behave” as expected. There was an inconsistency and errors in some of the integrated APIs and how they worked in the backend. Where possible, these data sets were carefully examined and set up to work as close to how they were expected to perform.
- Pivoted data – to make visualization easier within Power BI, some data sets had to either be pivoted or unpivoted depending on the visualization and data set in question.

After the cleaning and preparation was completed, the data set was now integrated into the tool and ready to be used to create the dashboard visualizations.

### Making the visualization

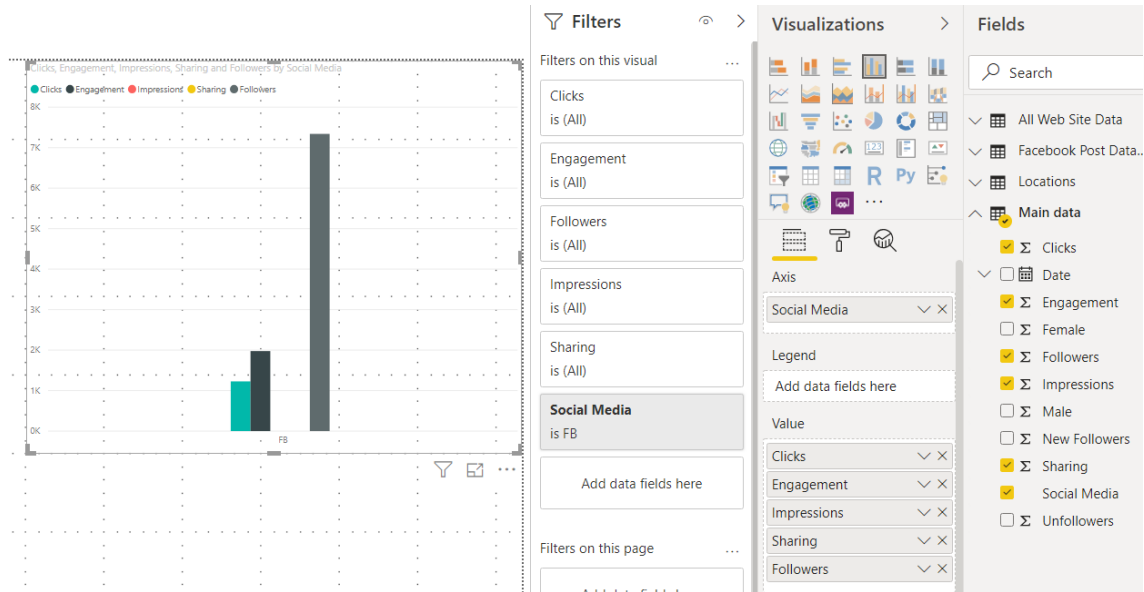
The integrated data was then used with a visualization with reference to the mock-up. This was done by selecting a visualization in Power BI and dragging it into the dashboard and the relevant data being imported to the right axes from the integrated data sets. Picture 6 below shows a clustered column chart with data from four major social media platforms.



Picture 6. A visualization filled in with the respective social media values.

### Adjusting and correcting filters

While Power BI simplified most of the work and automatically filled the visualization, the resulting visual still required some modifications. For this, the filters had to be applied to get the correct data to be output. Filtering allowed for the data set to be narrowed down to specific metrics that were relevant for the question the dashboard visual was intended to answer. Picture 7 below shows a clustered column chart with a filter applied to only show the data from on social media platform.



Picture 7. A graph visualization with the necessary filters added.

The above process was repeated until all the visualizations had been added to the dashboard report. Some formatting was required to make the visuals neater and their values easier to read, for example, sizing the visuals to use enough space in the report as well as increasing the font sizes. See Appendix 2 for the dashboard report of the first prototype.

#### 4.4 Errors and Mistakes

While the errors faced in the cleaning and preparation phase were corrected, they weren't the only issues that presented themselves. As the project grew, both technical and human errors were encountered.

The human error was poor planning through inadequate foresight. There were instances where some important data was not imported into Power BI. This was mainly because the need for said data was not realised during the planning phase, or because the data was not part of a data set that had previously been imported.

The technical issues arose when any file that was a data source was moved or removed from its location. This caused the system to break since Power BI used absolute paths when sourcing data. This meant that once the file was moved, Power BI could not recognize where it was moved to and could not import and/or refresh the data.

Another technical issue was encountered when edits made in the power query didn't run or update correctly. This prevented said query and any other query running after it in chronology from being executed and the task being left incomplete until the issue was resolved. This issue also presented itself when data being refreshed conflicted with the cleaning and preparation done to the imported data sets. For instance, when a data source that had some rows or columns removed during import was refreshed later, Power BI threw an error since it couldn't locate the deleted rows or columns. Solving for this problem was tedious and time-consuming.

The completed dashboard was functional but contained a lot of redundant and unnecessary data sources. Each import meant an additional table was being added to the entity relationship model (see Appendix 2). This made the entity relationship model unnecessarily complex and messy.

#### 4.5 Second Prototype

The completed dashboard served the purpose but had redundancies and other minor issues that would not do well over time. Therefore, there was a need to remake the dashboard with all the necessary information compacted and imported for maximum efficiency. Removing the redundancies would not only make the dashboard's size smaller, but also make the system more robust.

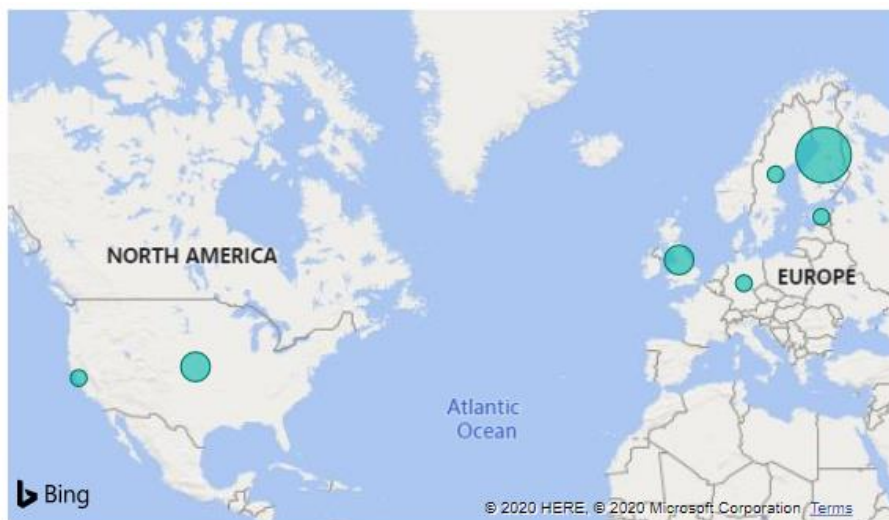
From the issues encountered and the lessons learnt during the first build, a second prototype of the dashboard was made following the same steps. This took lesser time than it took to make the first prototype since there were only minor issues to correct and because the first prototype was there to serve as a guide.

The dashboard now had a neater structure with a simple yet robust entity relationship model. Some new metrics and/or KPIs were introduced in some of the reports as deemed relevant and necessary, with the most significant change being the addition of a date filter on every report view of the dashboard. See Appendix 3 for the dashboard report of the final prototype.

## 5 RESULTS

The completed dashboard (see Appendix 3) was able to meet the requirements and objectives set at the beginning of the thesis project. The user data collected from the different social media channels was now unified and displayed on one platform/report. The graphical nature of the visuals was also immediately effective in showing the different trends in activities and KPIs that the Marketing team wanted to track.

The Social Media platforms were presented in the same dashboard report and it was clear which channel was more popular and how the community of users was, in comparison, reacting and interacting across the other platforms (see Picture 23 in Appendix 3 for more details). The same report also provided insight into the locations of the active users as shown in Picture 8 below.

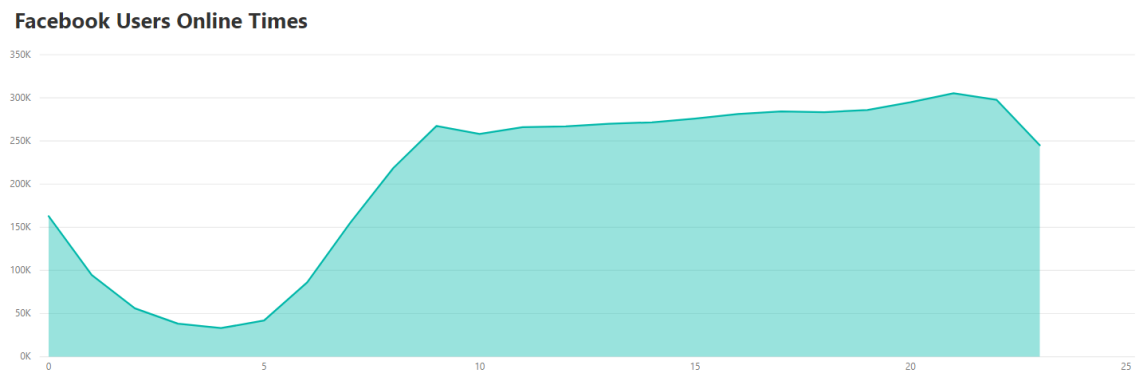


Picture 8. Location of active website users.

While it was expected that Finland would be the location with the most active users, United States of America being in the top three locations was a new find. This insight was immediately considered for the next marketing campaigns, with the intention of making a targeted advertisement to the users from there and increasing the user base by appealing to them by using relevant material in the advertisements. The published content also would be updated accordingly with posts specific for the audience from that location.

The gender metric (see Picture 23) also showed a near enough balance between the organization's audience with a 52.3% female user base and a 47.7% male user base. The teams realized the need to design campaigns and publish content that would reach and benefit the genders equally.

The users' activity on Facebook also provided insight into when the best time to make a new post would be (see Picture 9). The days that the traffic was highest on the commissioning organization's page was also followed (for more details, see Picture 27). The Marketing and Communications team planned their publishing times for future posts with respect to this information in order to reach the maximum audience. The more engagement a post received upon being posted would also boost it in the Facebook algorithms.



Picture 9. Time when Facebook users are online.

The top five Facebook posts with the most reactions were also noted. The content of these posts was analyzed to find out what the audience was interacting well with. Where applicable, the team planned to post similar content and follow along with the same major themes.

Visualizations from the data collected from the website provided a lot of insights as well. The links and sources that brought the users to the organization's website were studied (see Picture 24). This was done to learn where the users come from and what channels are promoting the organization and its contents. The top twenty sources would be focused on in the future marketing campaigns with more content being sent out on those platforms.

Similarly, the top keywords that the users were entering in search engines that landed them to the organization's website were tracked (see Picture 25). The purpose was to



understand what the users were looking for and how to improve the content being put out so that it is easily reachable as well as searchable for the audience.

The data of the users' interaction with the newsletter was also visualized (see Picture 26) and it showed that the newsletter was performing poorly. With only a 20.0% open rate and an even lower 2.9% click rate. This informed the Marketing team on a need to review the content and structure of the newsletter, as well as test out different themes and ways of making the content more engaging. The team considered A/B testing but realized the mailing list is not big enough yet to be able to conduct such a test.

The news and blog post activity was visualized and filtered to show the most popular articles based on the number of views they received (see Picture 28). This was done in effort to learn what content did well and was interesting for the user base. The average time spent on these pages was also monitored.

Apart from the news and blog posts, the top twenty-five pages of the website itself were plotted in a chart with respect to the page views (see Picture 29). This was done in order to make an informed guess on what type of information the users were looking for. It also allowed for the popular pages to be refined and the tailing pages to be improved.

The customer path was also linked to this dashboard report. The customer path showed the flow of users across the pages and was important in learning what page(s) the users entered the website from, what pages they interacted with while in an active session, and finally, what page they left from. This journey of the user across the website provided insight on what the user was looking for and how well the pages were linked for them to find all the information that they needed. The more important thing to deduce from this was why the user left the page and how to improve the webpage(s) to avoid a high drop off rate. Another crucial detail to follow along the customer path was to track how many users ended up on the ticket page, and how many went on to click on the button to buy the ticket. The number of users going through to the tickets page was 46% while the drop-offs were at 54%.

The ticket sales made were also represented in a graph chart and various filters added to the dashboard report to narrow down the choices using the type of ticket as well as the type of ticket campaign. This report allowed for the tracking of ticket sales over time and pointed at the ticket campaigns that did better (see Picture 30 and 31 for more details). The more popular types of tickets also stood out in the data. Tickets from the "Normal" campaign were bought the most with a total of 962 sales. It must be taken into

consideration that the other campaigns span over a relatively short time, with the “2-for-1” ticket campaign being the second-best type with 74 sales. The “Partner Festival” ticket type did the best with 648 sales, followed by the “Partner Student” type with 175 sales.

The top 30 webpages that lead to the tickets page within the website were also tracked and presented in a chart (see Picture 32). A lot of the ticket traffic landed straight to the tickets page, with the next best internal source being the homepage of the website. This information helped the team understand what pages they needed to improve in order to have more users reach the tickets page. The pages performing well would also serve as a lesson for what was working well in sending more users towards the tickets page.

## 5.1 Suggestions and Recommendations

While the resulting dashboard was fully functional, there certainly are ways to improve it moving forward. To limit the scope of the project, some assumptions were made. These included:

- i. The APIs within Power BI work correctly and without any bugs.
- ii. The data imported from a source is correct and retains its integrity.

Steps should be taken to confirm that there are no technical issues with Power BI’s APIs and that all data that is imported to the BI tool is not compromised in the process. This would require thorough understanding of the data sets individually and as a complete relational model.

The current solution works but can be improved in many ways. At present, it is limited to the functionality within Power BI and contains a lot of manual labor. Automating the sourcing of data and its preparation would considerably reduce the workload for the individual in charge of the dashboard.

Instead of using a Power BI tool that confines the user to work within the limits of the tool, it would be advisable to develop a custom dashboard platform for the organization. This would allow for maximum flexibility in functionality and allow for complete customization.

## 5.2 Challenges Encountered

Apart from the technical challenges that were encountered whilst working with and in Power BI, there were some irregularities and inconsistencies that presented themselves during the thesis project timeline and affected it either directly or indirectly. Some of these included:

- i. The organisation's analytics tool was restructured and that affected the data import.
- ii. The organisation's website was redesigned and renewed and had some technical issues that affected the ongoing work.
- iii. The tool used for ticket sales didn't work smoothly and getting data from that source was particularly challenging.
- iv. The Head of Sales and the Head of Marketing were replaced, and work had to be restarted and continued with the staff hired to replace them.

## 6 CONCLUSION

This thesis project aimed to unify and present the data collected across an organization's different channels and present them in an analytics dashboard. This objective was attained by the creation of a dynamic dashboard in Power BI that is hosted on Microsoft's online platform.

The dashboard was used to obtain insights from the visualized data, which in turn would help improve the working methods and rectify shortcomings within the different administrative teams, as well as, strengthen the practices that were successful in meeting the predefined KPIs. Implementing this information to correct or improve the organization's activities would therefore increase business.

The business intelligence tool used requires a license for each user that would like to access an online dashboard and depending on the number of administrative users that need access to it, this is an incremental cost for the organization. The dashboard requires maintenance and updates on regular intervals, and this requires a dedicated employee to manage.

The results were clear and indicative of the user behavior and interests. This can easily be tapped into and converted into leads and generate more business value. Business leads that become lost in the volume of data collected in the organization, when visualized, provide clarity in defining new marketing campaigns and adjusting those that are ongoing.

The dashboard, while completed, still requires constant maintenance and improvements. The solution can be improved by having a custom-designed platform to host and visualize the organization's data. This would cut licensing costs and in the longer run be cheaper for the organization. This platform would also work better with APIs that allow for continuous refreshing of data and automate even more background tasks to reduce the need for involvement from an employee. This would ensure a robust and well-functioning solution that could be packaged and sold as a product or service as well.

## REFERENCES

- Beek, D. v. (2019, October 20). *QlikView*. Retrieved November 12, 2019, from Passionned Group: <https://www.passionned.com/bi/tools/qlikview/>
- BusinessDictionary. (2019, September 27). *data*. Retrieved November 06, 2019, from BusinessDictionary: <http://www.businessdictionary.com/definition/data.html>
- Chang, J. (2018, July 23). *20 Best Data Analytics Software for 2019*. Retrieved November 04, 2019, from FinancesOnline: <https://financesonline.com/data-analytics/>
- Durcevic, S. (2019, June 05). *An Introduction To Data Dashboards: Meaning, Definition & Industry Examples*. Retrieved November 07, 2019, from Datapine: <https://www.datapine.com/blog/data-dashboards-definition-examples-templates/>
- Frankenfield, J. (2019, April 27). *Data Analytics*. Retrieved November 21, 2019, from Investopedia: <https://www.investopedia.com/terms/d/data-analytics.asp>
- Haije, E. G. (2019, November 06). *Top 15 Business Intelligence Tools: An Overview*. Retrieved November 12, 2019, from Mopinion: <https://mopinion.com/business-intelligence-bi-tools-overview/>
- Hall, T. (2019, February 13). *The Rise of Digitalisation in Business*. Retrieved November 06, 2019, from Compare The Cloud: <https://www.comparethecloud.net/articles/business-digitalisation/>
- iDashboards. (2019, September). *What is a Dashboard?* Retrieved November 07, 2019, from iDashboards: <https://www.idashboards.com/guides/what-is-a-dashboard/>
- Klipfolio. (2019, September). *What is a data dashboard?* Retrieved November 07, 2019, from Klipfolio: <https://www.klipfolio.com/resources/articles/what-is-data-dashboard>
- Marrs, M. (2016, May 10). *The Difference Between Data, Analytics, and Insights*. Retrieved November 07, 2019, from Localytics: <http://info.localytics.com/blog/difference-between-data-analytics-insights>

- Masters In Data Science. (2019, May 16). *What is Data Analytics?* Retrieved November 09, 2019, from Masters In Data Science: <https://www.mastersindatascience.org/resources/what-is-data-analytics/>
- McKown, A. (2016, September 26). *What is The Purpose of a Dashboard?* Retrieved November 06, 2019, from iDashboards: <https://www.idashboards.com/blog/2016/09/26/what-is-the-purpose-of-a-dashboard-2/>
- Molag, T. (2019, February 21). *Power BI vs Tableau*. Retrieved November 11, 2019, from Encore Business: <https://www.encorebusiness.com/blog/power-bi-vs-tableau/>
- Rouse, M. (2018, October 30). *Business Intelligence Dashboard*. Retrieved November 06, 2019, from TechTarget: <https://searchbusinessanalytics.techtarget.com/definition/business-intelligence-dashboard>
- Scott, T. (2019, May 30). *Power BI vs Tableau: A Data Analytics Duel*. Retrieved November 11, 2019, from TechnologyAdvice: <https://technologyadvice.com/blog/information-technology/power-bi-vs-tableau/>
- Spacey, J. (2017, October 12). *10+ Types of Business Data*. Retrieved November 06, 2019, from Simplicable: <https://simplicable.com/new/business-data>
- Techopedia. (2019, October 15). *Data Visualization*. Retrieved November 07, 2019, from Techopedia: <https://www.techopedia.com/definition/30180/data-visualization>
- Wikipedia. (2019, October 09). *Data analysis*. Retrieved November 09, 2019, from Wikipedia: [https://en.wikipedia.org/wiki/Data\\_analysis](https://en.wikipedia.org/wiki/Data_analysis)
- Young, N. (2017, July 27). *The Importance of Dashboards*. Retrieved November 06, 2019, from CPA Global: <https://www.cpaglobal.com/cpa-global-blog/the-importance-of-dashboards>

## **Appendix 1. Questions Template for Interviews**

1. What sort of KPIs do you use?
2. How do you use these KPIs in making decisions?
3. How do you use these KPIs in setting up a plan?
4. How do you base your decisions?
5. What kind of data do you use/wish to use to make the right actions/plans?
6. What more should be being tracked?

# Appendix 2. First Prototype

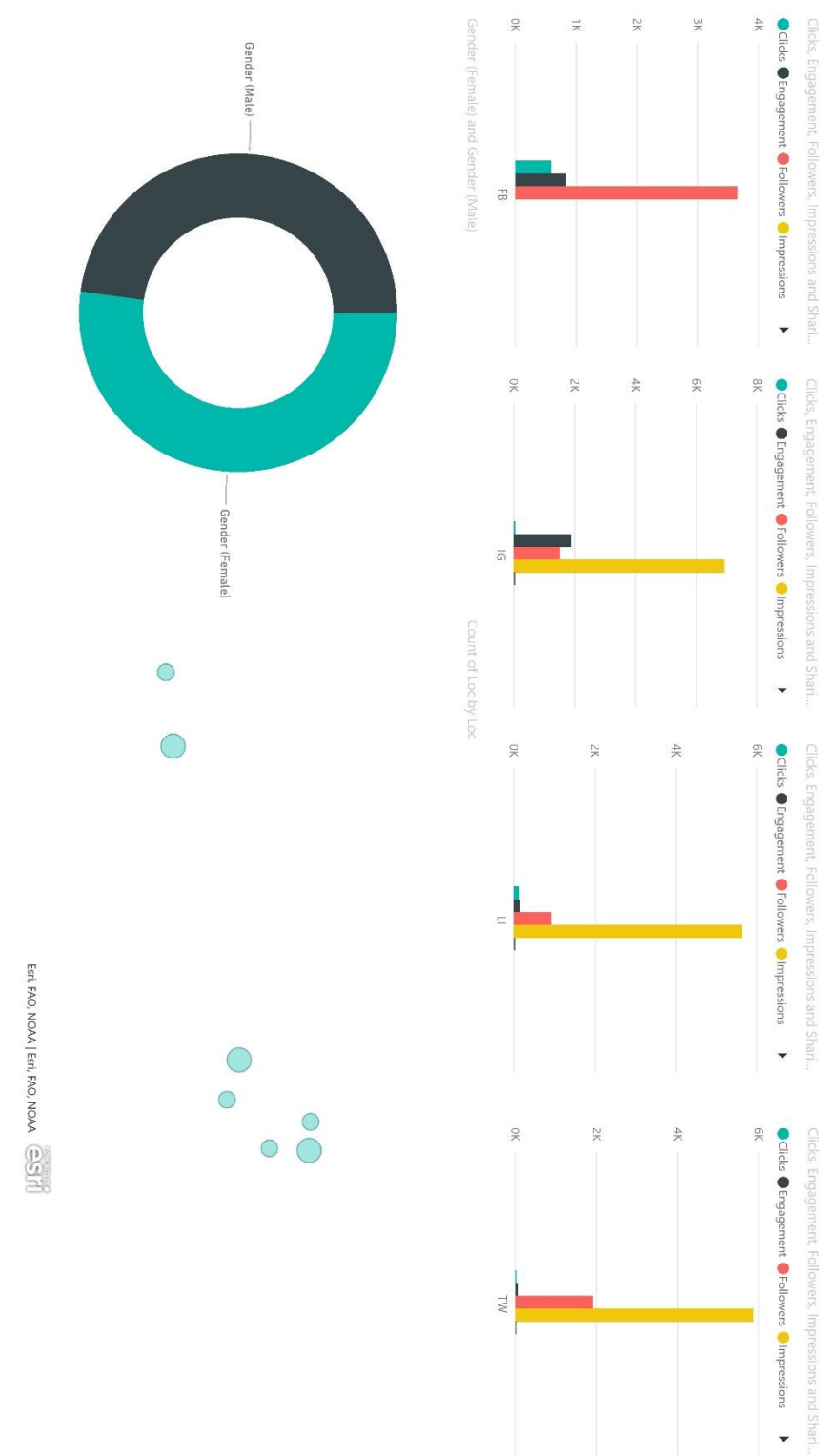
## Entity Relationship Model



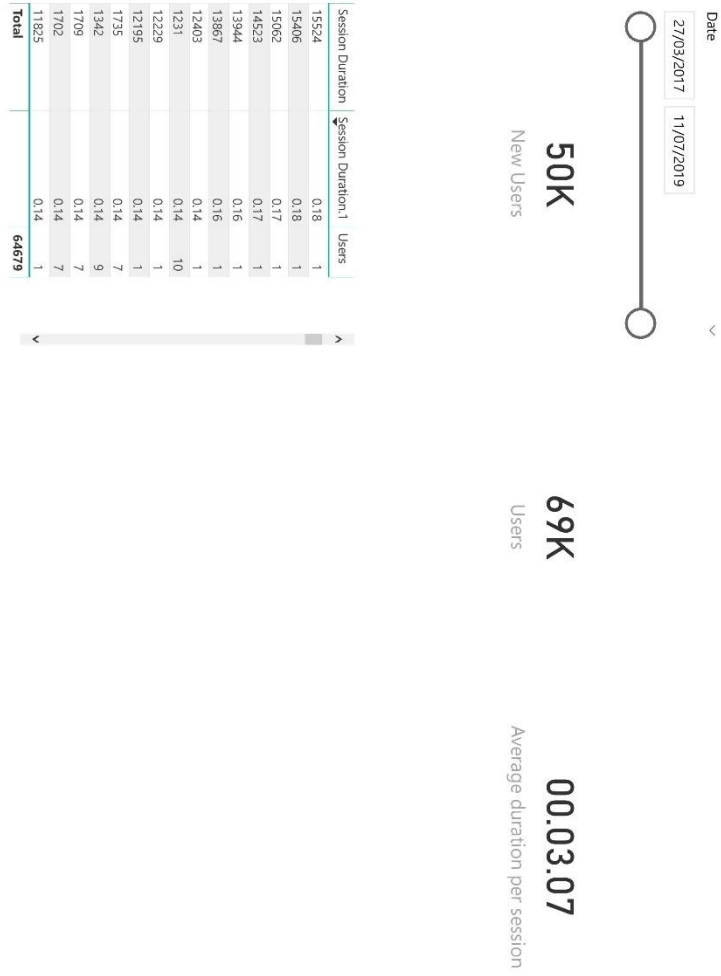
Picture 10. ER Model of the first prototype.



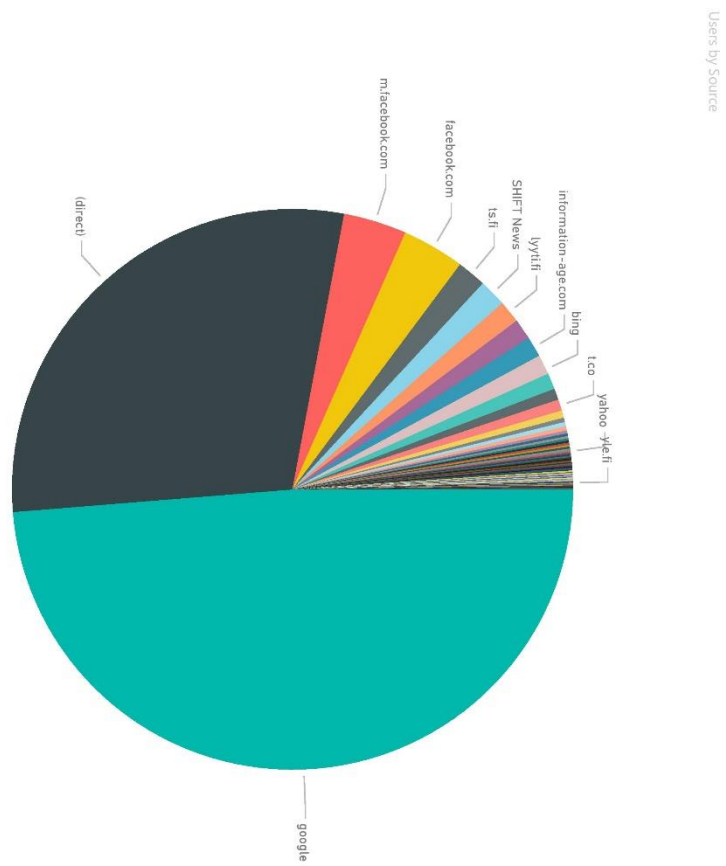
# Dashboard Report



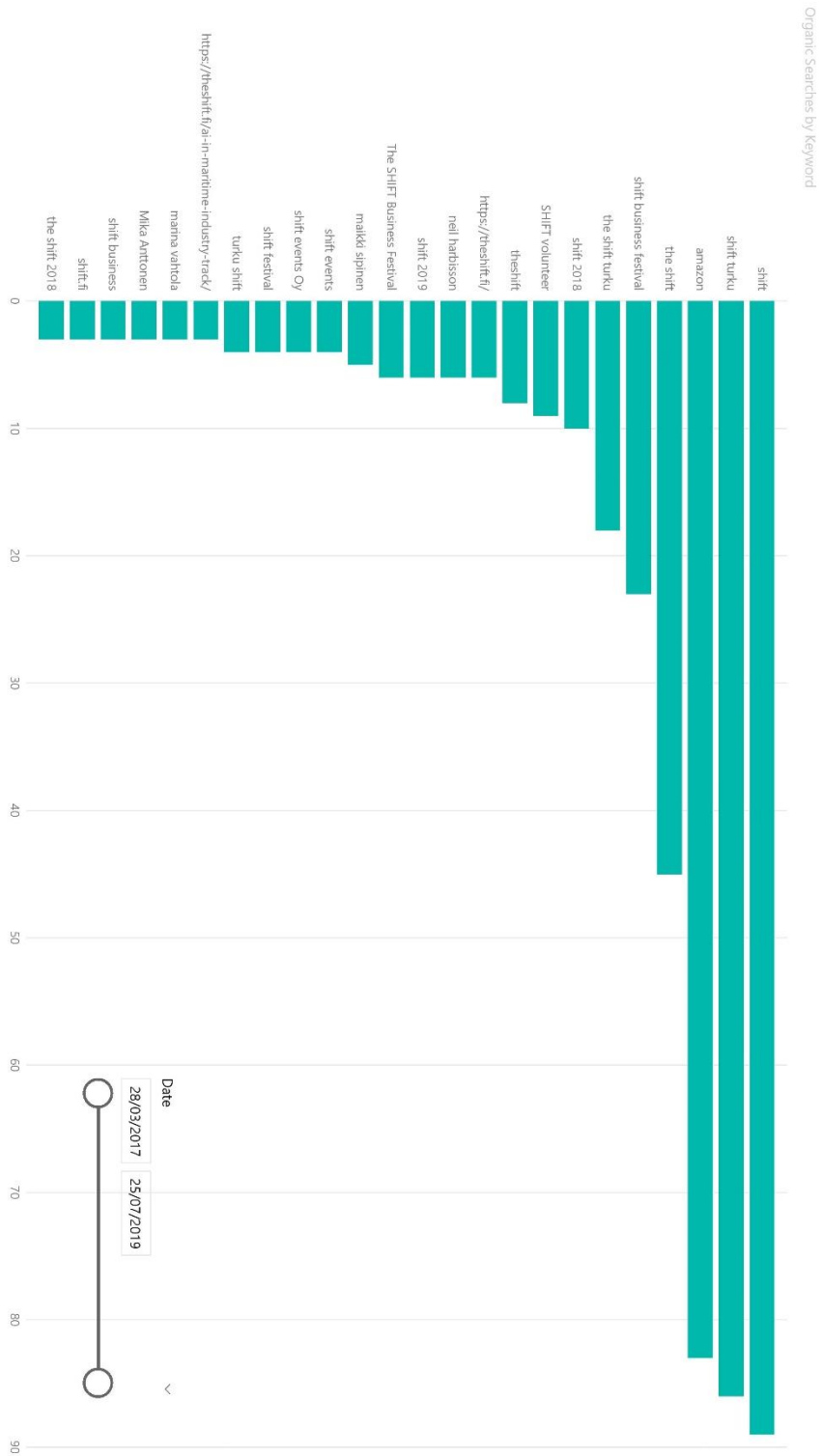
Picture 11. Social media engagement.



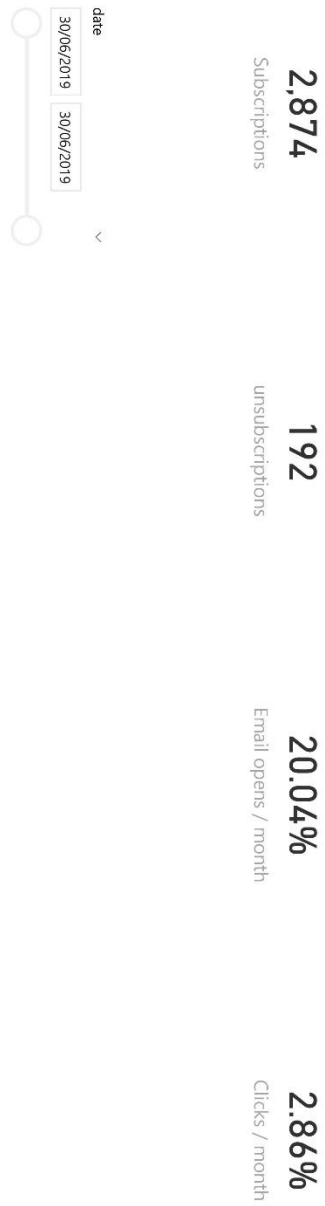
Picture 12. Users visiting the website and the session durations.



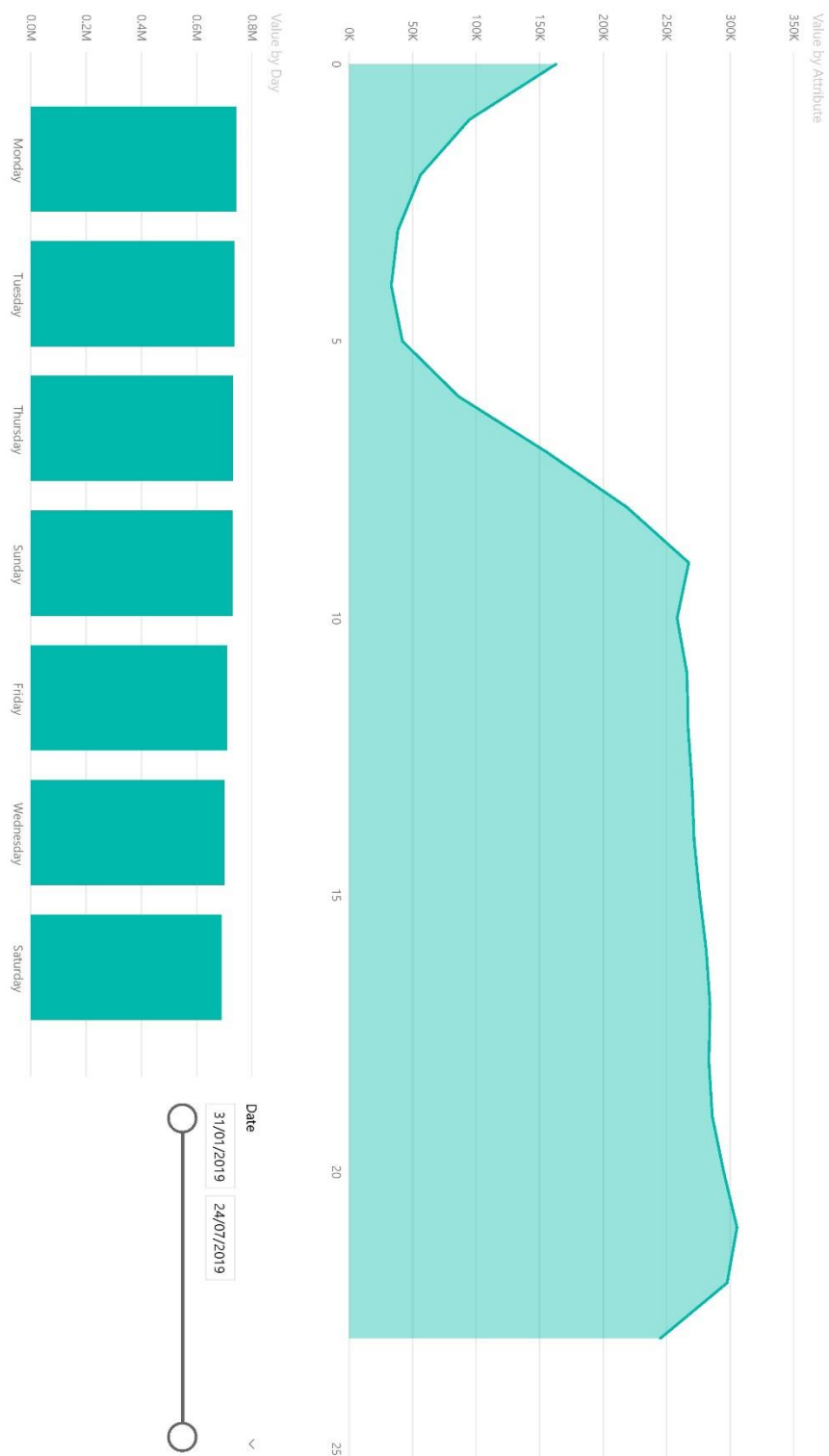
Picture 13. Website users by source.



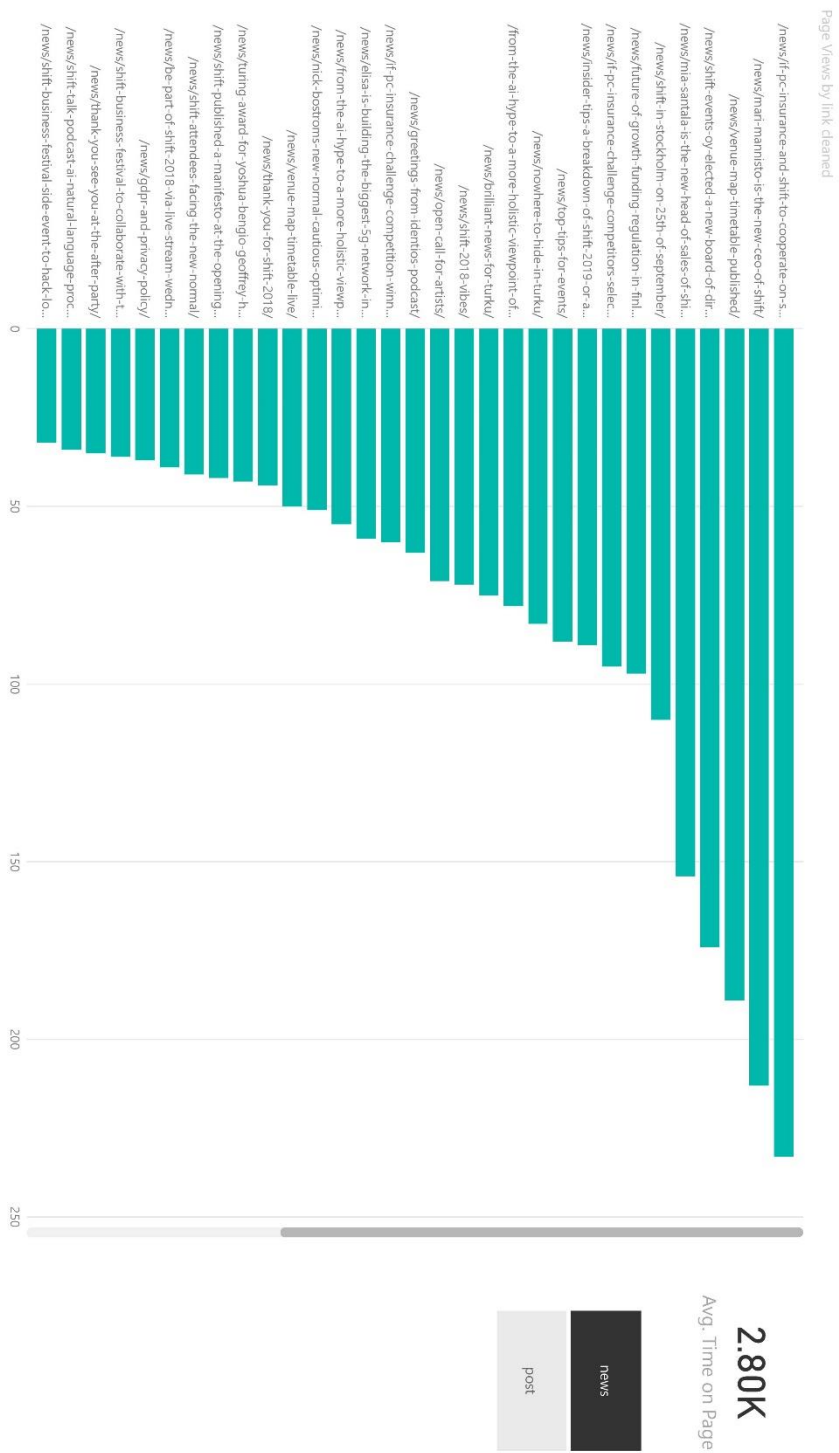
Picture 14. Organic searches by keywords.



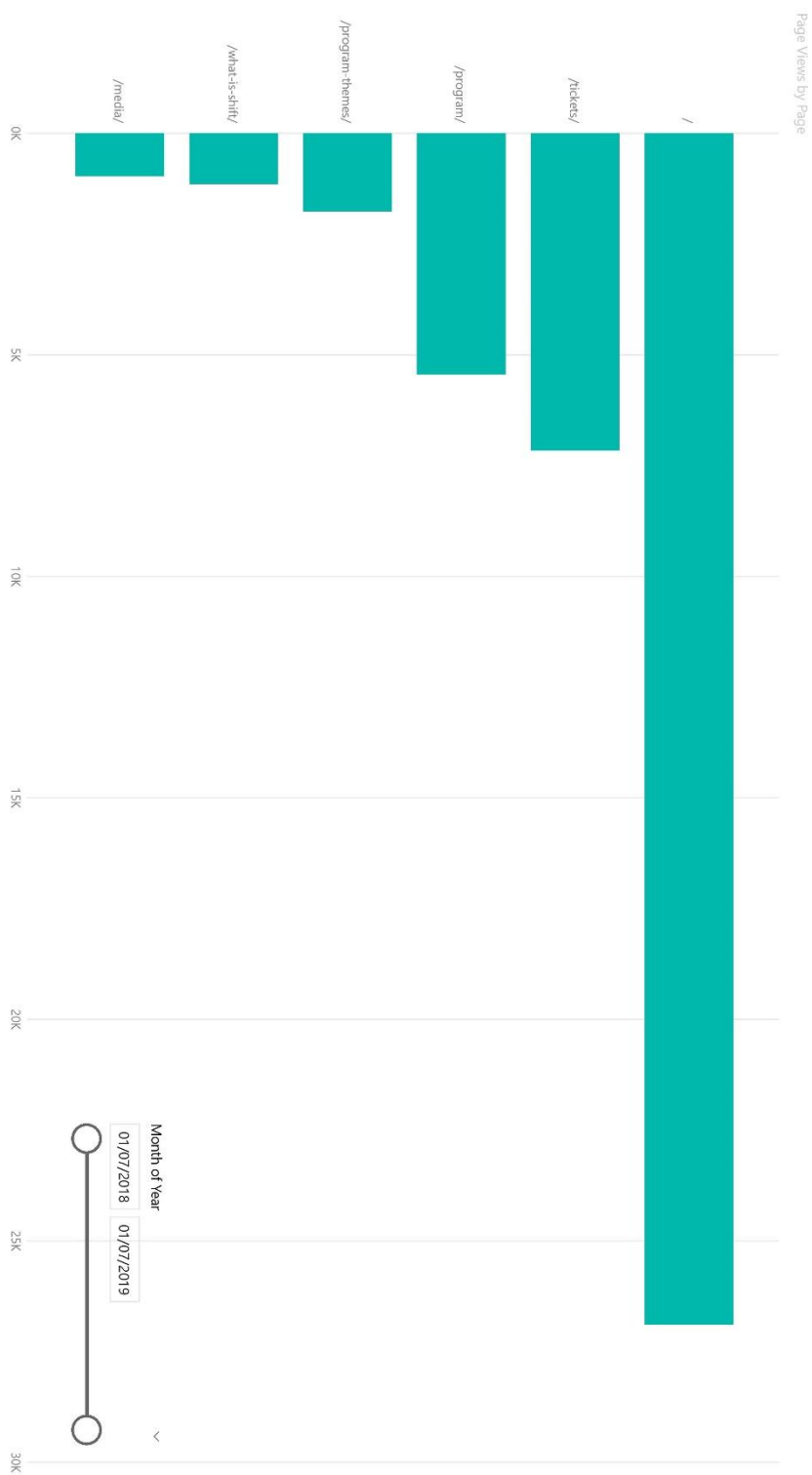
Picture 15. Details of the newsletter activity.



Picture 16. Time when Facebook users are online.

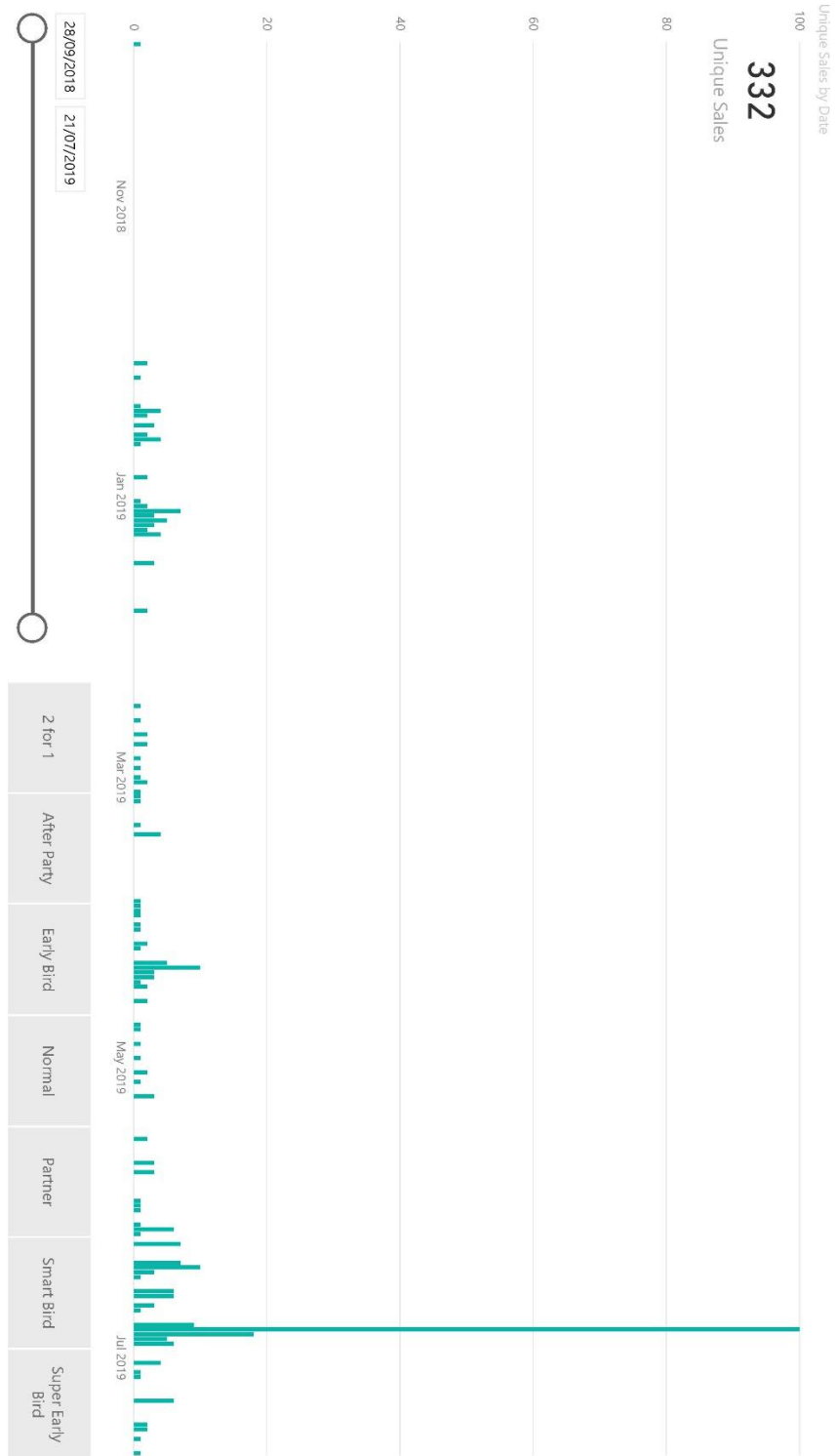


Picture 17. Blog and news pageviews on the website.

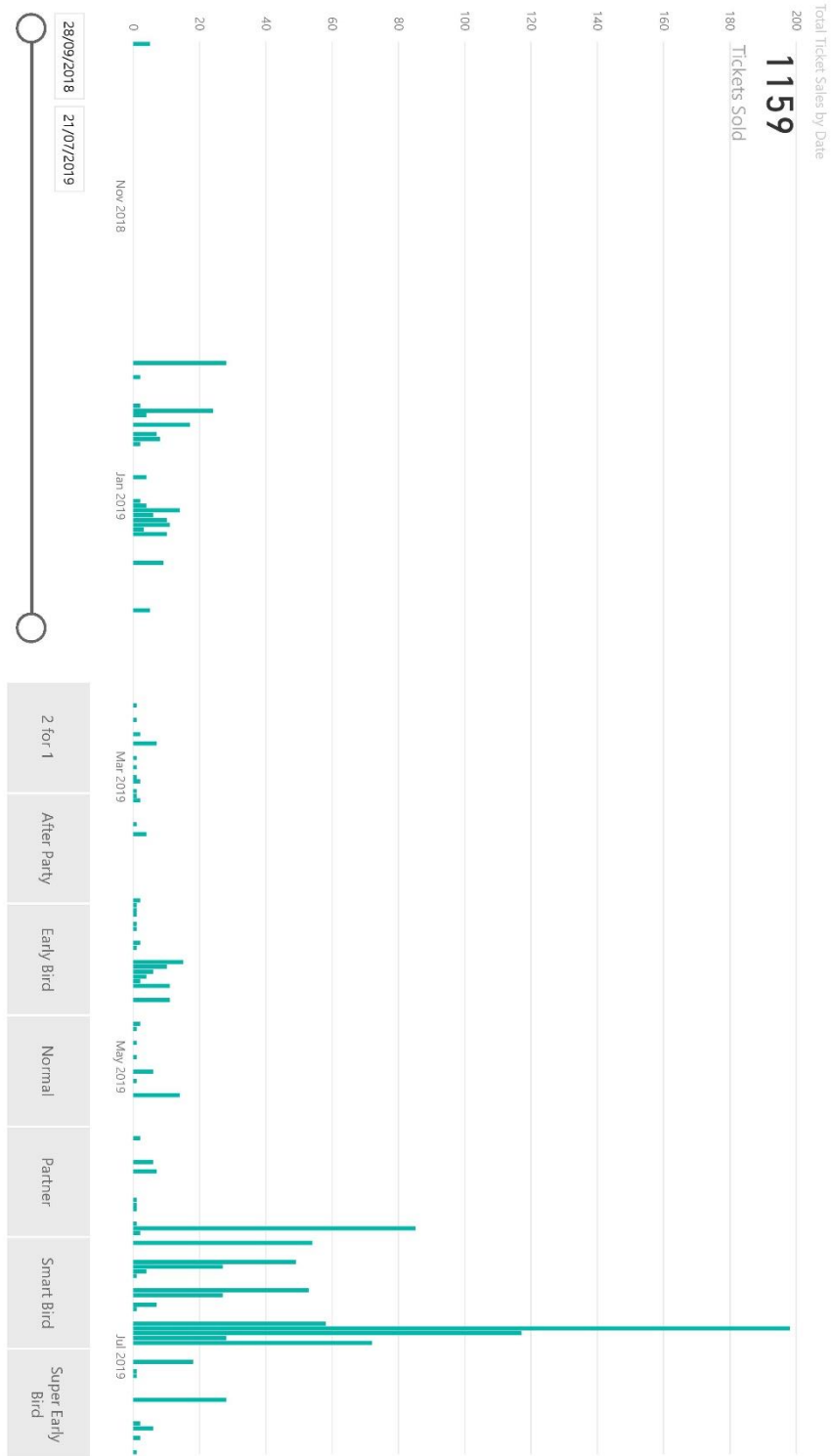


Picture 18. Specific pages by their pageviews.

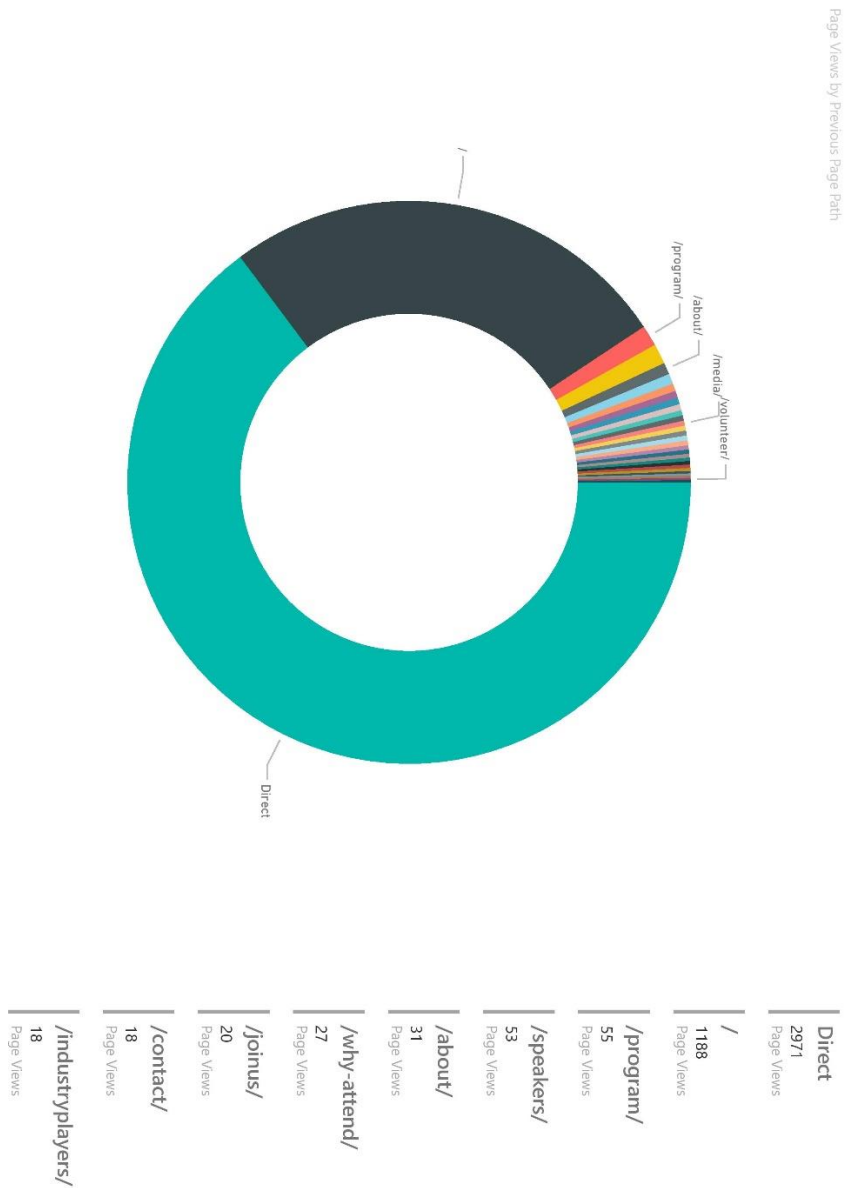




Picture 19. Unique ticket sales.



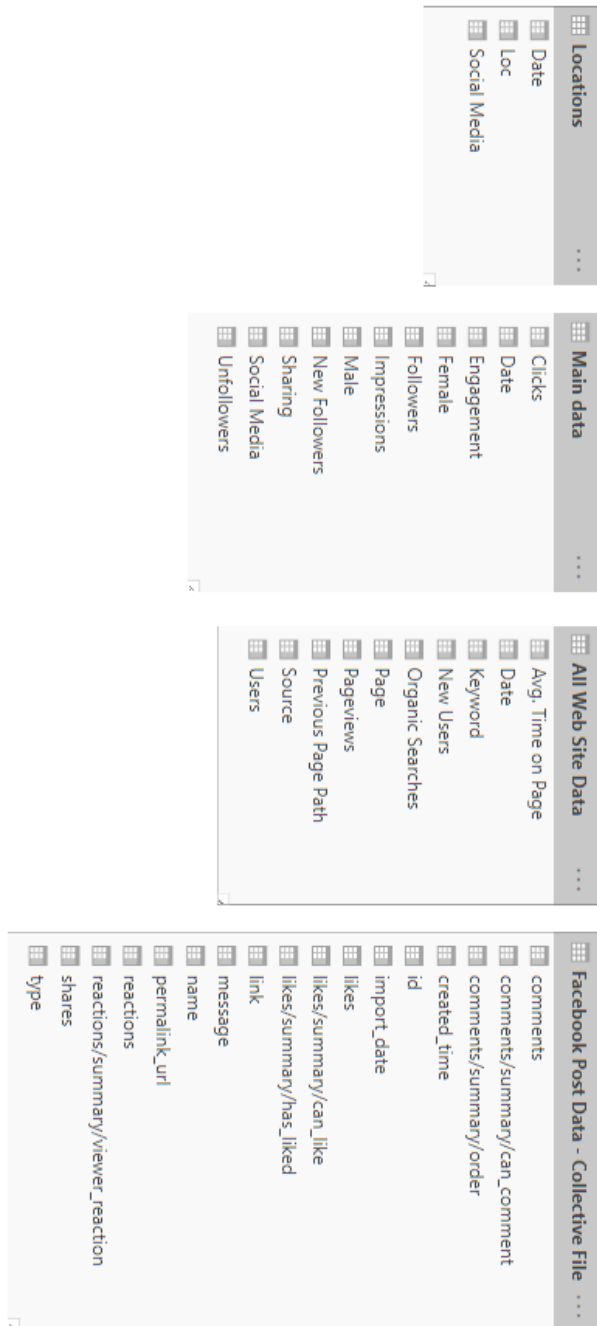
Picture 20. Total ticket sales.



Picture 21. Popular paths used to land on the Tickets page.

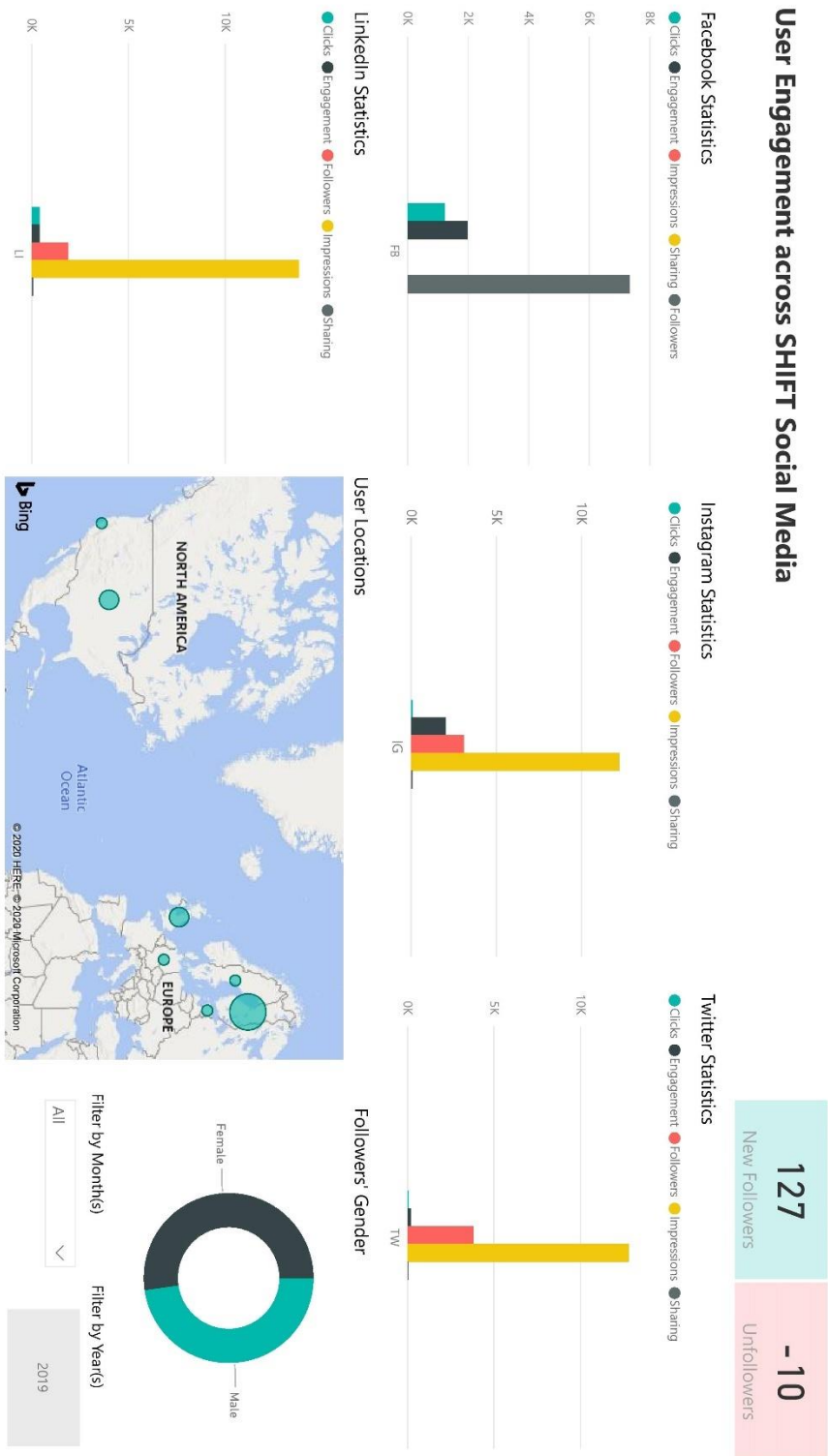
## Appendix 3. Final Prototype

### Entity Relationship Model

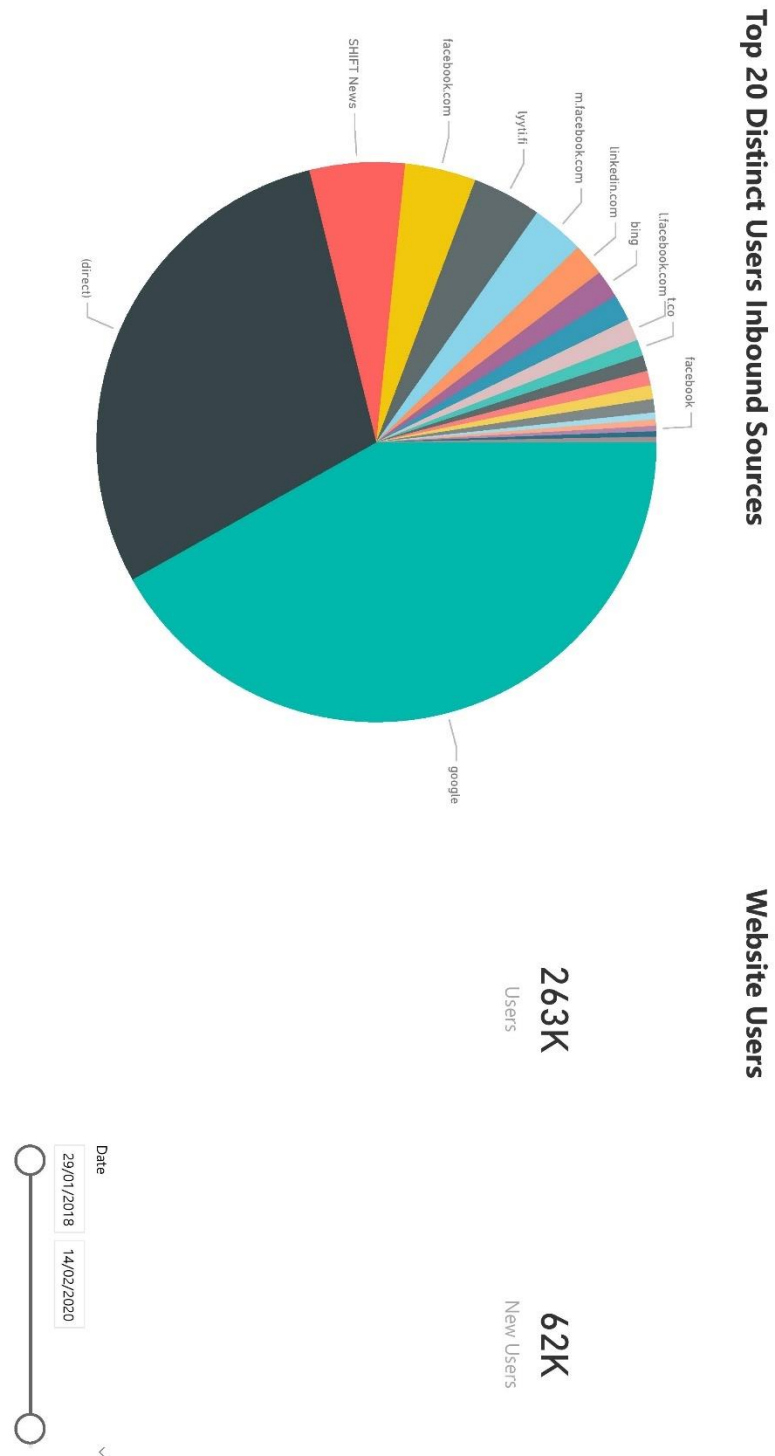


Picture 22. ER Model of the final prototype.

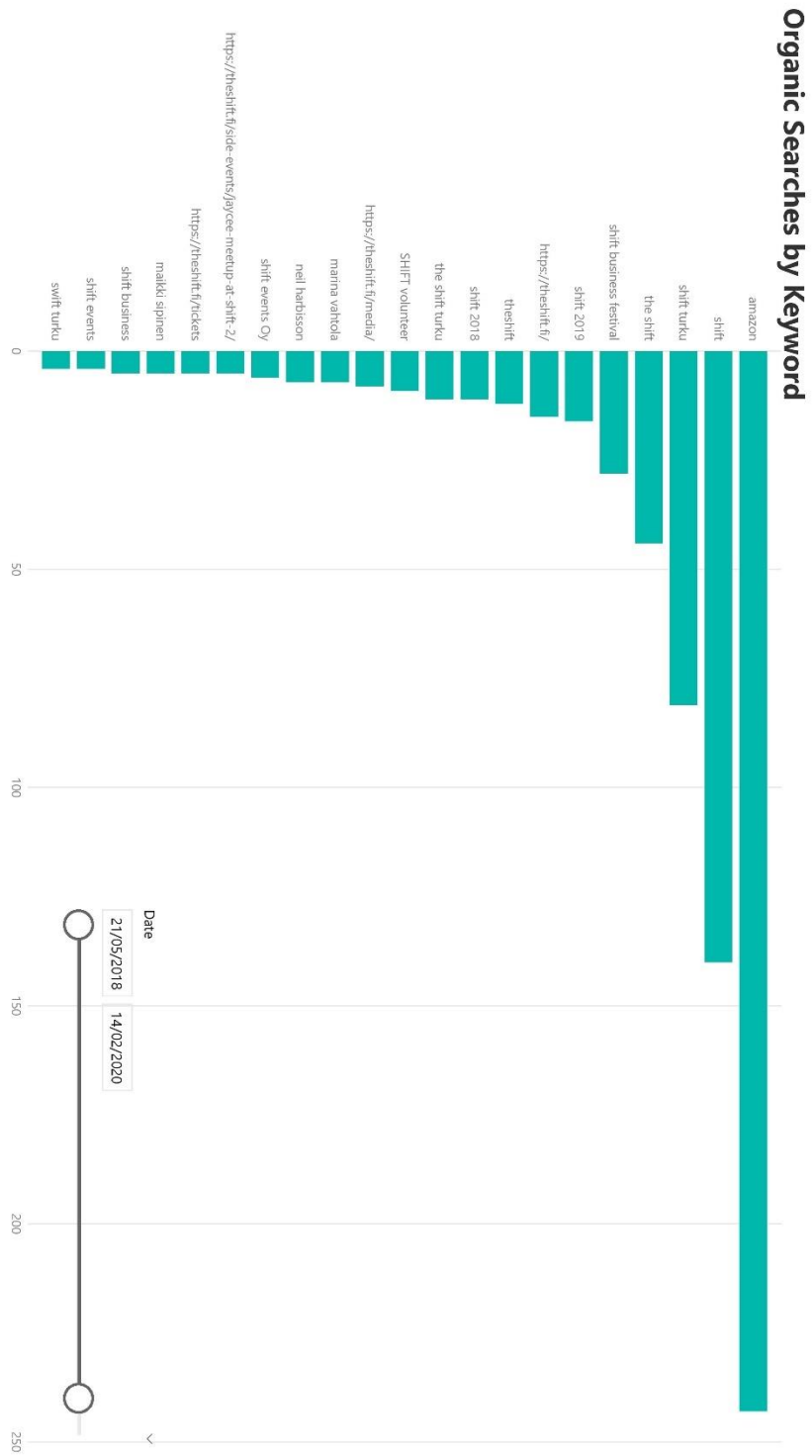
# Dashboard Report



Picture 23. User engagement across social media.

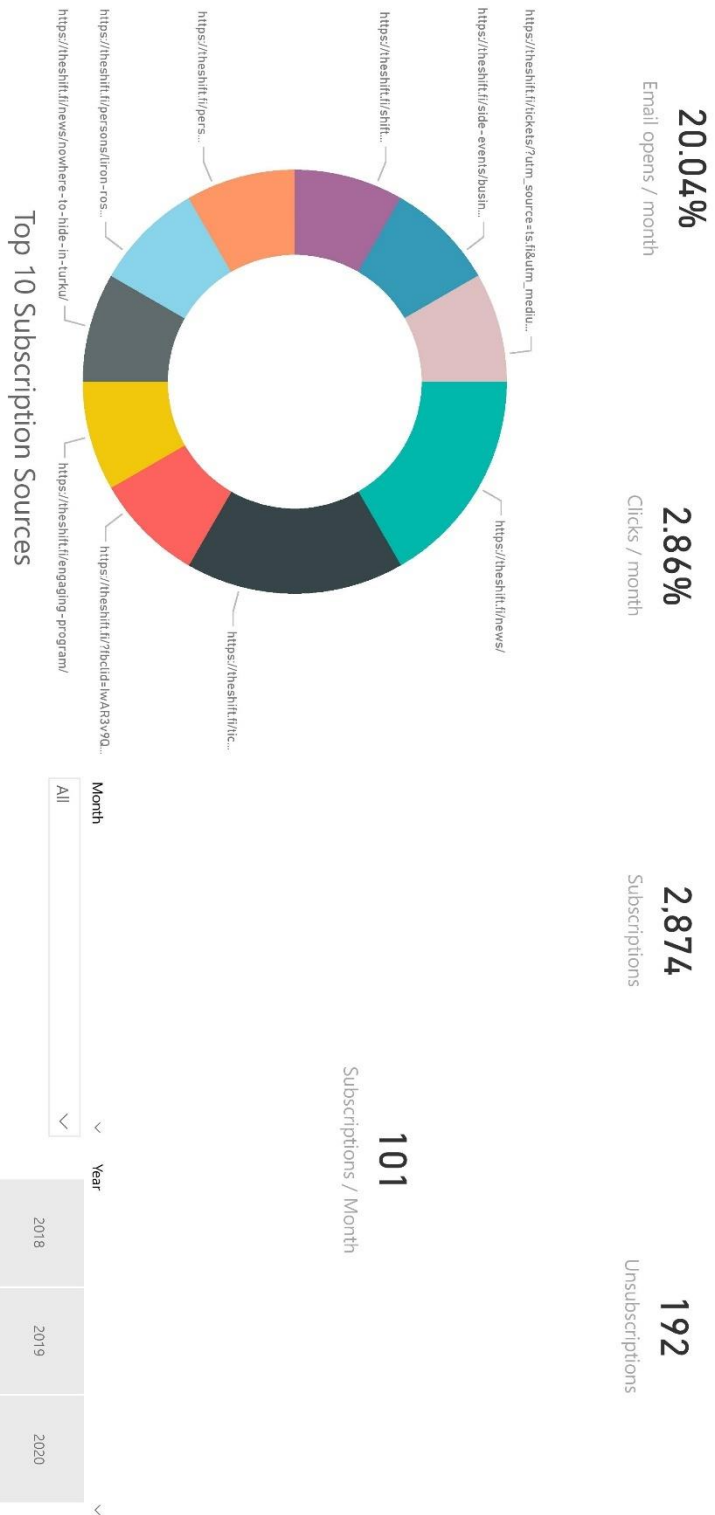


Picture 24. Website inbound sources and user statistics.



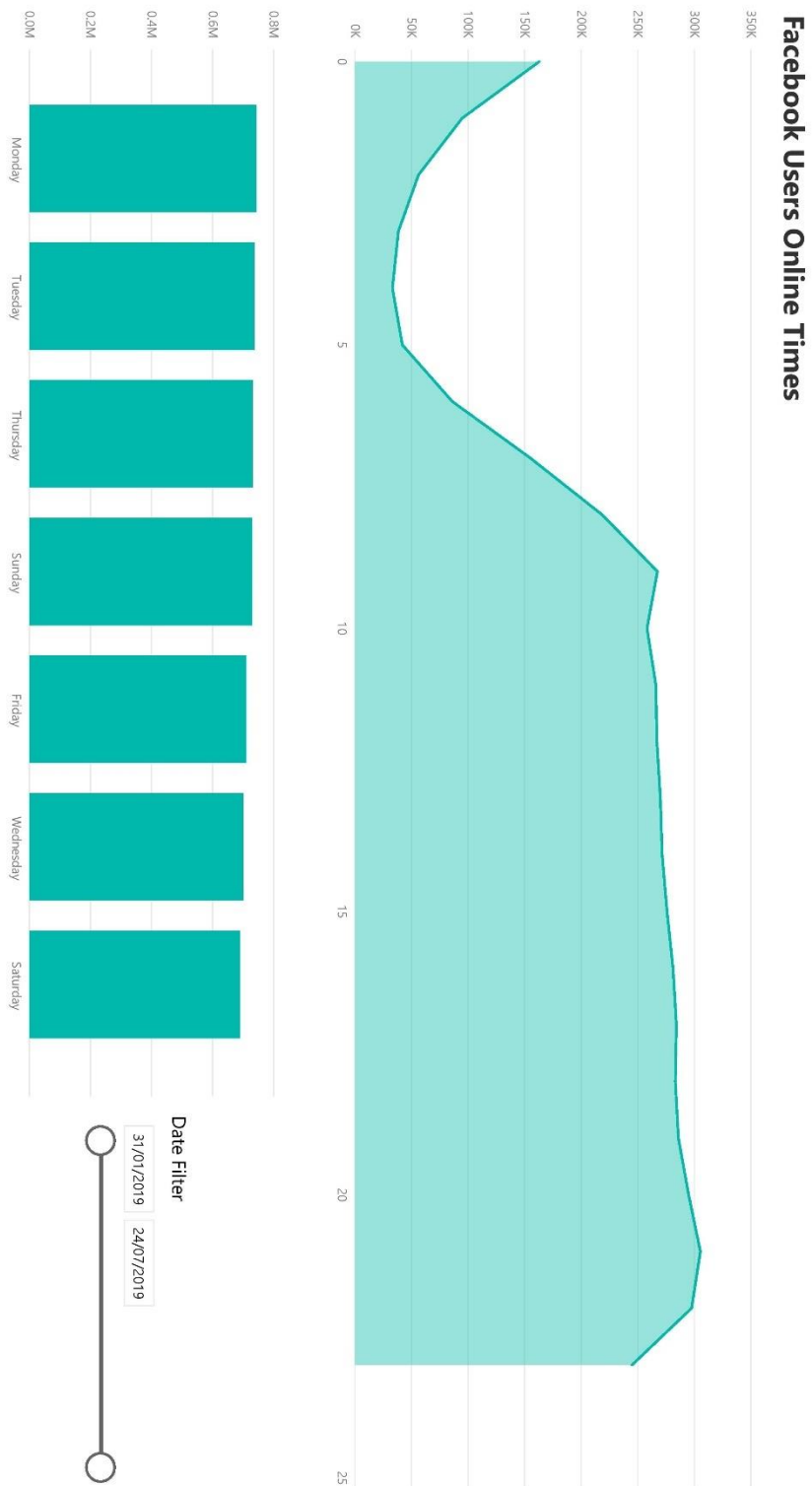
Picture 25. Organic searches by keyword.

### MailChimp Stats

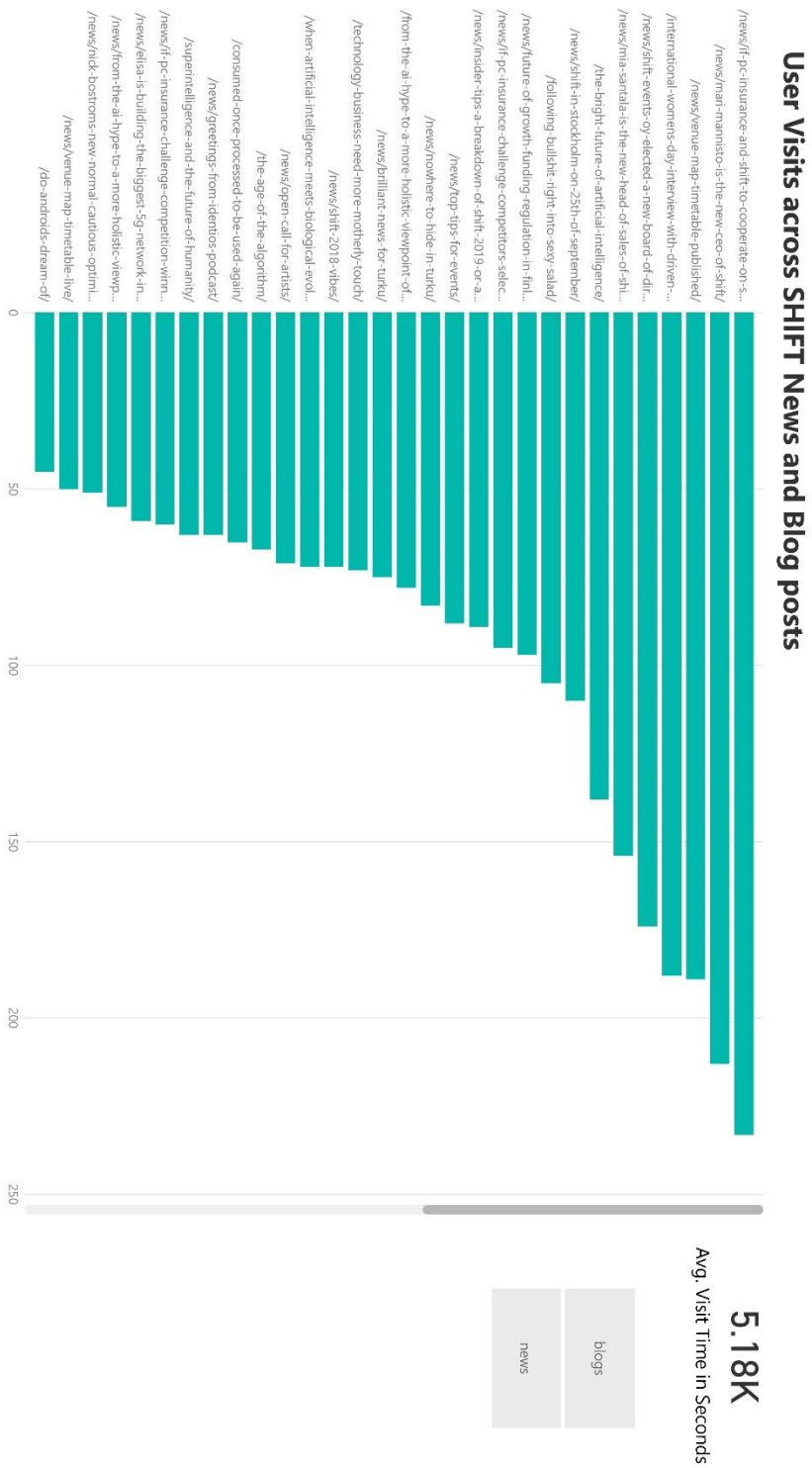


Picture 26. Newsletter statistics.

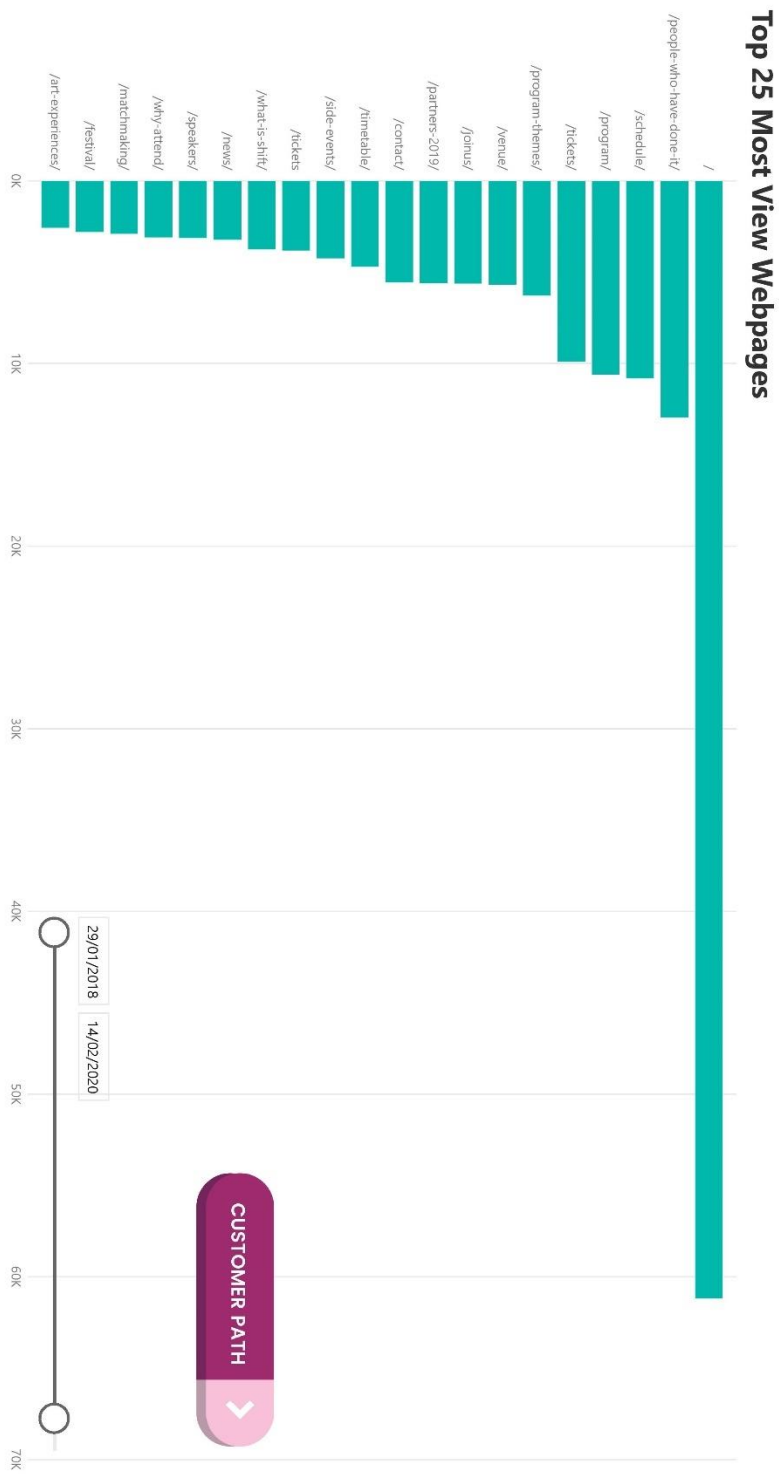




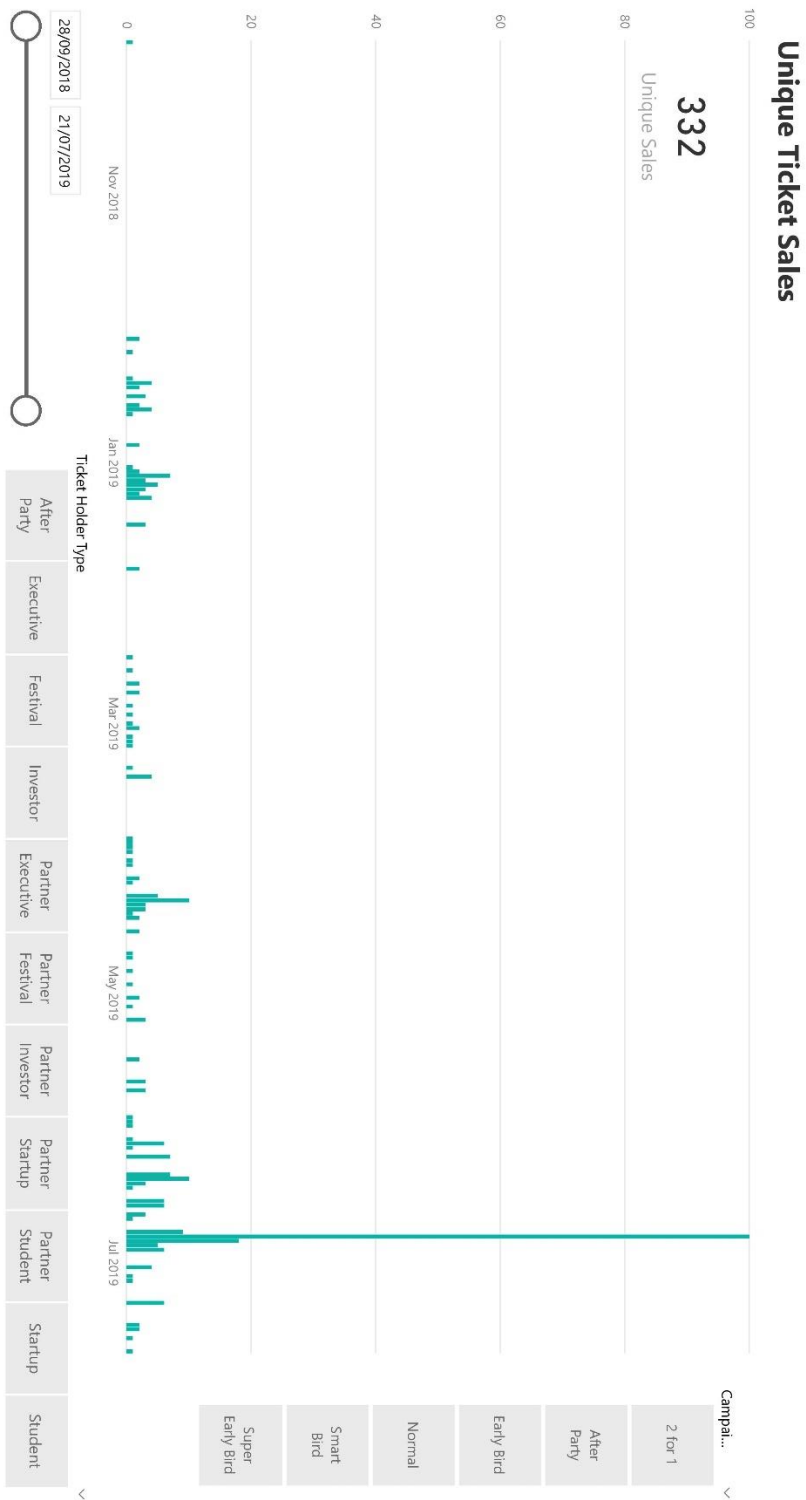
Picture 27. Time when Facebook users are online.



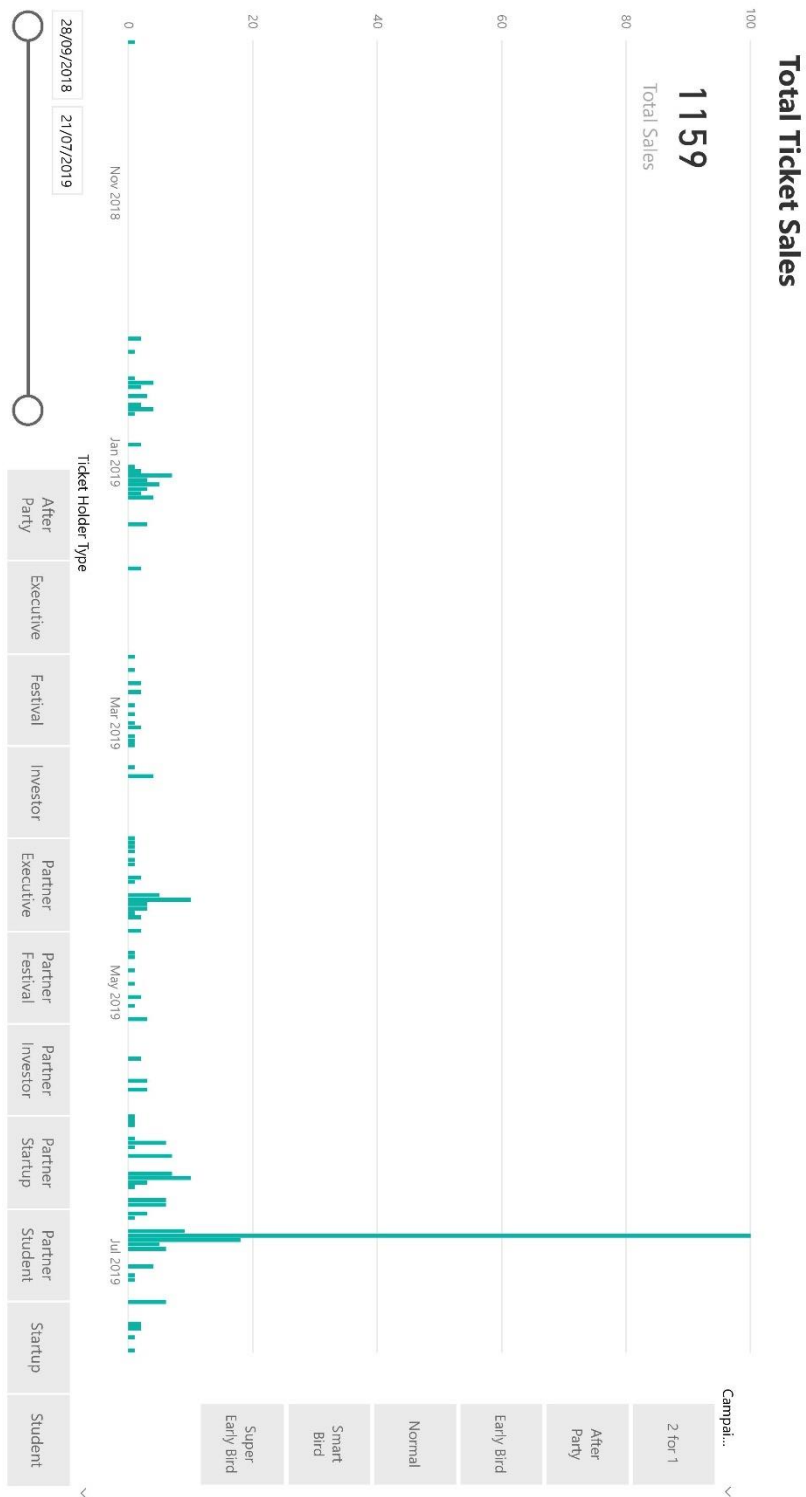
Picture 28. Blog and news pageviews on the website.



Picture 29. Top 25 webpages by pageviews with a button externally linking to the Customer Path.

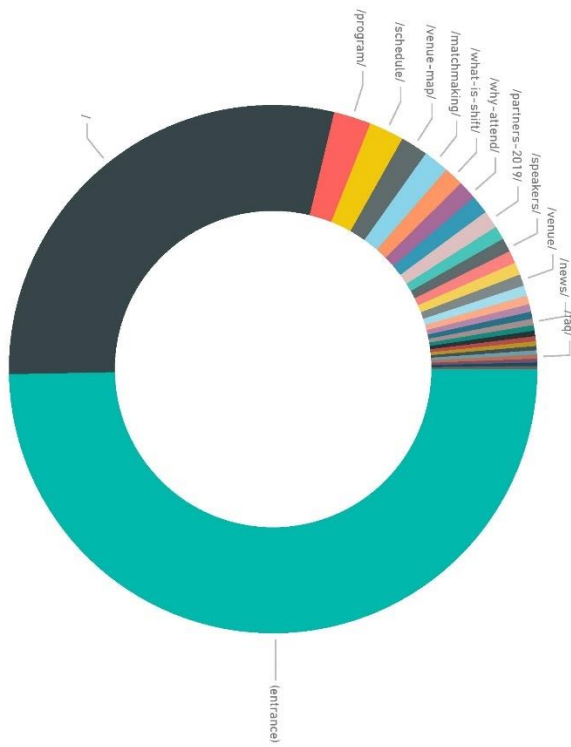


Picture 30. Unique Ticket Sales.



Picture 31. Total ticket sales.

### Top 30 Pages Leading to Tickets Page



(entrance)	6162	Pageviews
/	3598	Pageviews
/about/	79	Pageviews
/after-party-2/	43	Pageviews
/art-experiences/	53	Pageviews
/contact/	104	Pageviews
/faq/	31	Pageviews
/festival/		
Date	29/07/2018	14/02/2020

Picture 32. Top 30 internal webpages leading to the Tickets page.