Digital Marketing Ad Analysis with Facebook Data

1. Introduction

Over the past decade, the application of data analytics has become ever more prevalent in various sectors of the workforce, increasingly being used as a 'real-time decision-making tool' (Bollier, 2012, p.1), driven by the digitisation of our social data (Aboab, et al. 2016). Similarly, there has been growing interest regarding the benefits of datathons in providing a unique experience for collaboration between data scientists and experts from various sectors, to bring together their varied but potentially complementary skills and knowledge, to address industry specific questions (Aboab, et al. 2016). The term 'datathon' derives from 'hackathon', a successful model for innovation in the business space, which take the formation of short, intense competitions, where teams apply their knowledge to generate innovative solutions (Aboab, et al. 2016). Datathons apply the hackathon model to data analytics, which exposes you to complex problems to work on from different areas of machine learning, artificial intelligence (AI), and data science (Data Science Society, 2018).

This report details the experience of a short-form datathon that we participated in earlier this semester. It explores two areas and is therefore structured as such. The first section discusses the details of the datathon, such as: dataset and variables, process and structure, and group work. The second section analyses the results from the datathon, challenges and concluding thoughts.

2. Datathon Experience

2.1 Dataset and Variables

The project options and group formation for the datathon were decided online, prior to the start date. Our group choose the project which involved applying exploratory data analysis using R to provide actionable insight into a 'example company's' Facebook advertising campaign. The dataset for this project is called 'Sales Conversion Optimisation' and can be downloaded from Kaggle. The objective of this project was to analyse marketing key performance indicators (KPI's) to answer specific research questions for a clients advertising strategy. Our 'example company' took the formation of a premium activewear clothing brand, with three Facebook advertising campaigns. The company had a large target audience, men and women between 30-50 years old, and therefore were interested in identifying which campaigns were the most successful, in terms of conversions, and what insights we can draw from this regarding target audience and where to invest more money.

The research questions for this project were therefore focused on analysing campaign performance and explores this by: Firstly, comparing the campaigns and measuring performance through highest engagement, e.g. number of conversions and clicks on the advert. Once the highest performing campaign has been identified, the group will then investigate further specific demographics, such as age and gender. With the purpose of identifying further which subsection of their wide target audience is interacting the most with the ad most frequently and the relevance between user and ad. This will be measured through specific KPI's including CR and ROAS.

Facebook is an attractive advertising platform as it collects information about its users as they interact with it, enabling advertisers to target specific audiences (Newberry, 2019). This is helpful for our analysis as the dataset already included a variety of useful demographic and KPI variables ready for analysis. The description of all the variables derived from the original dataset source are listed below:

1. Ad_id: Unique ad ID.
2. XYZcampaign_id: Campaign ID.
3. fbcampaignid: Unique ID. Associated with tracking each campaign.
4. Age: Age of who the ad was shown to.
5. Gender: Gender of who the ad was shown to.
6. Interest: A code specifying the category to which the persons interests belong.
7. Impressions: Number of times the advertisement was shown.
8. Clicks: Number of clicks for a specific advertisement.
9. Spent: Amount paid by the example company to Facebook for showing a particular advertisement.
10. Total Conversion: Number of people who enquired about the product after seeing the ad.
11. Approved Conversion: Number of people who bought a product after the ad.

There were, however, some KPI's which would help us perform analysis missing, these variables were created and added to the dataset. They include:

1. **Click through Rate (CTR):** The CTR is the percentage of impressions that resulted in a click, allowing you to identify how relevant users are finding the advertisement. A high CTR signifies high relevance between user and ad. While low CTR signifies that advertisements are less relevant to their audience (Mackey, 2021). A high CTR is crucial for ad relevance and increased exposure (Mackey, 2021).
2. **Total Conversion: The** total conversion is the conversion, which refers to product enquiry, and the approved conversion, number of people who bought a product, added together for each ad. This helps to analyse ad and campaign performance by exploring the number of total conversions. It is also used to calculate the conversion rate.
3. **Conversion Rate (CR):** The CR is the total number of conversions divided by the number of clicks. A conversion can refer to various desired actions that you want the user to make which would be determined by the objective of the campaign. This can range from actions such as but not limited to, purchasing a product, completing a contact form, signing up for a subscription or engaging with the site in some way (Andrus, 2020).

4. **Return on Advertising Spend (ROAS):** The ROAS is a measurement of how much you receive in return for the amount spent on advertising. The ROAS is a useful KPI for evaluating the effectiveness of your advertising strategy and can be used to evaluate specific campaign and ad performance (Andrus, 2020).

2.2 Process and Structure of the Datathon

The short form datathon was structured into five sections and was participated in via an online four-hour workshop. The five sections comprise of:

- **Initial Presentation:** This section introduced the workshop and our guest speaker, who gave a presentation which defined what a datathon is, how it would be structured within the four-hour schedule, detailed the different roles within the groups and general expectations and ground rules of the datathon.
- **Initial Work:** The initial work for this datathon was laid out in an hour-long presentation by the guest speaker, where the project options with the instructions and expectations for each project were discussed in detail. The project option and the related datasets were shared online, prior to the workshop to maximise time efficiency. This meant that each group had already chosen a project option and the next task was to allocate participants with a specific role within the group. There were five roles defined for this datathon which reflect the reoccurring roles in data science projects more generally, these include: a project sponsor, client, data scientists, data architect and operations (Zumel, Mount, 2019).
- **Exploration:** This section involved the participants exploring the provided datasets, using this data to formulate research questions to drive the focus of the project. This paved the way for various forms of data cleaning to be achieved, and the creation of new variables to enable us to better answer our research questions.
- **Marathon**: This section forms the most intensive, and for many the most excitable, part of this workshop. Each group had to accomplish their chosen task within the given time frame of an hour and 45 minutes. This involved using exploratory data analysis techniques through the software programme R to gain insights regarding the chosen metrics and create visualisations in the form of tables, charts and graphs to interpret the results. This part of the datathon relied exhaustively on group work and collaboration to ensure the limited time was used efficiently.
- **Closure:** Once the time for the marathon was up, each group was required to stop working on their projects, to ensure fairness among groups. The groups were then required to create a presentation which displayed their results in a logical format, as if it were to be presented to a particular client. This again achieved through a collaborative format where each group member with distinct roles participated in the creation of the presentation. Each group then presented their project findings to the rest of the participants and hosts. Once

each group had presented the datathon ended with a collaborative peer evaluation, via an online form, and some final words from our guest host.

2.3 Group Work and Roles

This section will identify the different roles allocated to each team member, the responsibilities attached to each role, and the importance of that role within a data science project. As discussed previously, our project was to investigate an ‘example company’s’ Facebook advertising campaign by measuring the performance of each campaign and particular demographics, to make recommendations for improvement. Each of the previously identified roles were allocated and are detailed below:

- Project Sponsor/Client: This team member represented the business and user interests and were responsible for deciding whether the project is a success (Zumel, Mount, 2019). The role is to define specific goals for the group project and work collaboratively to achieve these. The role of the client and project sponsor were filled by the same person, as this allowed the workload to be spread evenly among the group and prompted active engagement throughout the datathon process. The company we chose to explore had a large target audience, men and women between 30-50 years old. Therefore, the purpose which was formulated by the project sponsor/client was to evaluate campaign performance by comparing the three Facebook campaigns and drawing from this which group demographics have the highest levels of engagement.
- Project Manager: This team member communicated directly with the client to accurately assess their requirements and further direct the data analysis. A framework was then developed to enable the group to work efficiently given the time limitations. The project manager was also heavily involved in the final presentation and discussing the results.
- Data Architect: The data architect role was allocated to two team members, they were responsible for taking the necessary steps to ensure the dataset is ready for analysis, this included exploring the dataset and variables, identifying variables that will be relevant for our project and various forms of data cleaning. Followed by the creation of new KPI’s, calculated using the variables available to us from the dataset, to supplement the project further. The data architect played a crucial role within the project cycle by having a clear understanding of the project, marketing knowledge and data science skills. The utilisation of various skills is considered as a significant benefit for this datathon, and data projects in general.
- Data Scientist: The data scientist role was filled by two team members due to the complexity of skills needed to successfully fill the project needs. They played a decisive role in the proposed techniques for analysis. Due to the project being based primarily in providing insight and measuring campaign performance, it was decided that a data intensive approach would be complimentary to this project (Leonelli, 2012). While broader business goals exist, the purpose of this research project was to provide valuable insight through the application of data analytics, by measuring different definitions of

success through the application of statistical analysis and digital marketing KPI’s.

3. Results

3.1 Findings from the Datathon

The aim of the project was to analyse and compare campaign performance to determine which is performing the most successfully and draw further insight related to factors such as age and gender to determine how relevant their advertising is.

Figure 1. & 1.a.

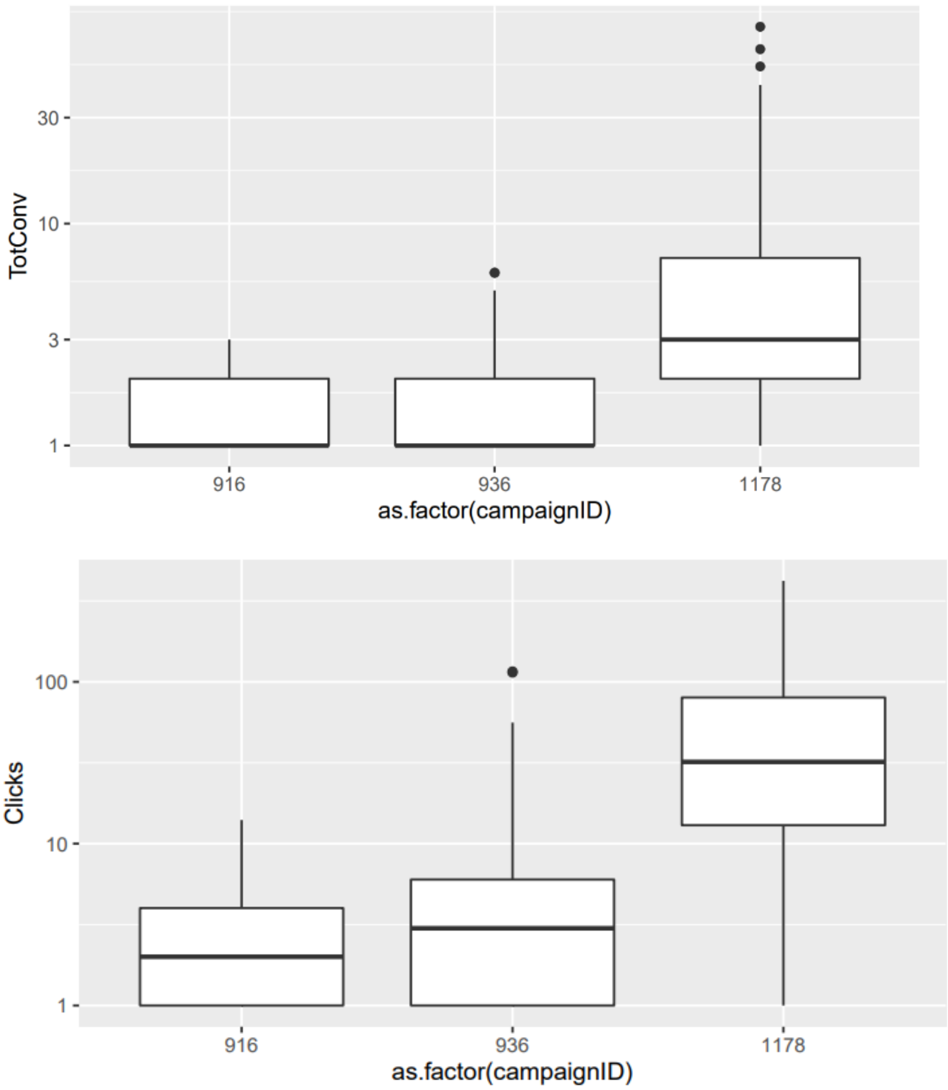
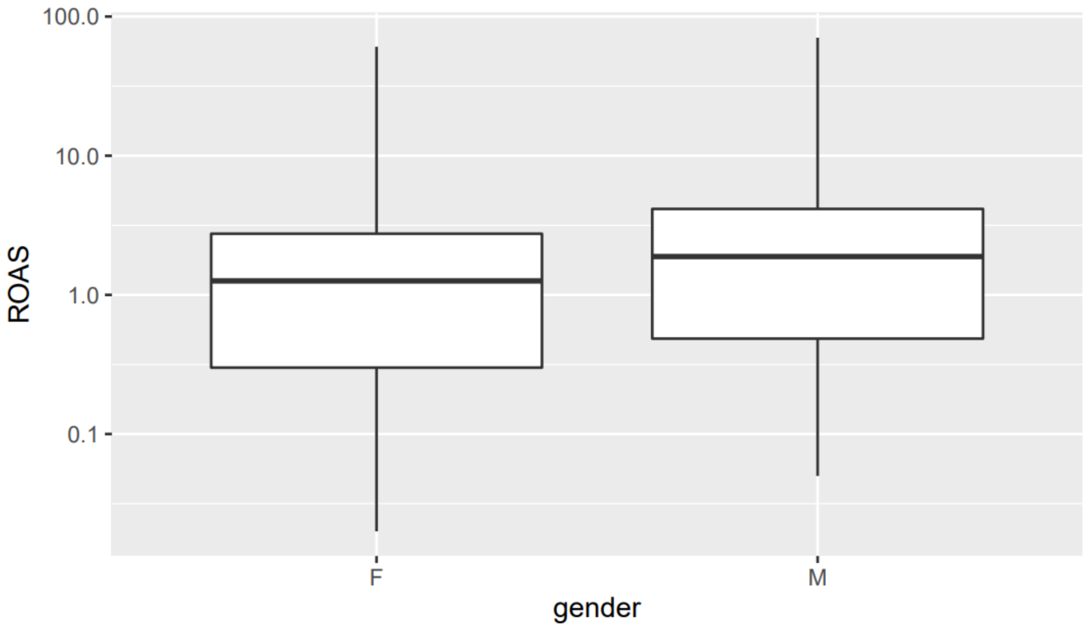


Figure 1 and 1.a. represent the prelininary exploration of the three campaigns by measuring and comparing performance based metrics such as, number of clicks each ad within the campaigns received and the campaign with the highest number of conversions. It becomes clear that campaign 1178 is performing the best when mesuring these metrics. The median average for the total number of conversions were 3 per ad for campaign 1178, compared to 1 conversion per ad for both campaign 916 and 936. All three campaigns are skewed to the right with significant outliers. These outliers represent particular ad's which performed exceptionally well in comparison to the average. Exploring outliers can provide interesting discoveries, however, refining the dataset in such a way significantly lowers the number of observations, resulting in less confidence in the results being significant or simply noise. The client is interested in campaign performance more broadly and targeting the relevant audiences. The median was chosen as a more accurate measure of central tendancy as outliers can have a substantial impact on the mean, with large values pulling the mean away from the center (Frost, 2018).

Campaign 1178 was then selected for further exploratory data analysis. The dataset provides certain demographic information such as age groups and gender which will be explored further in this particular campaign to identify which group are engageing with the advertising content the most, therefore allowing the company to target relevant audiences more effectively.

Figure 2. ROAS for Campaign 1178 by gender



Overall males and females generate a similar ROAS, this is reflected in figure 2. The median average ROAS for males is 1.88, while females are slightly lower at 1.23. However, the data above reflects a symmetric distribution therefore the mean may reflect the central tendency more accurately (Frost, 2018). The mean ROAS for males is 4.5 compared to mean ROAS females is 2.82. This shows that the campaign is working most effectively for the male audience, in terms of profitability and engagement. A non-parametric test was performed, and the results were significant (p-value = <0.005). These figures show a significant difference and therefore will be explored further to analyse campaign performance and identify where their advertising is working well and areas for improvement.

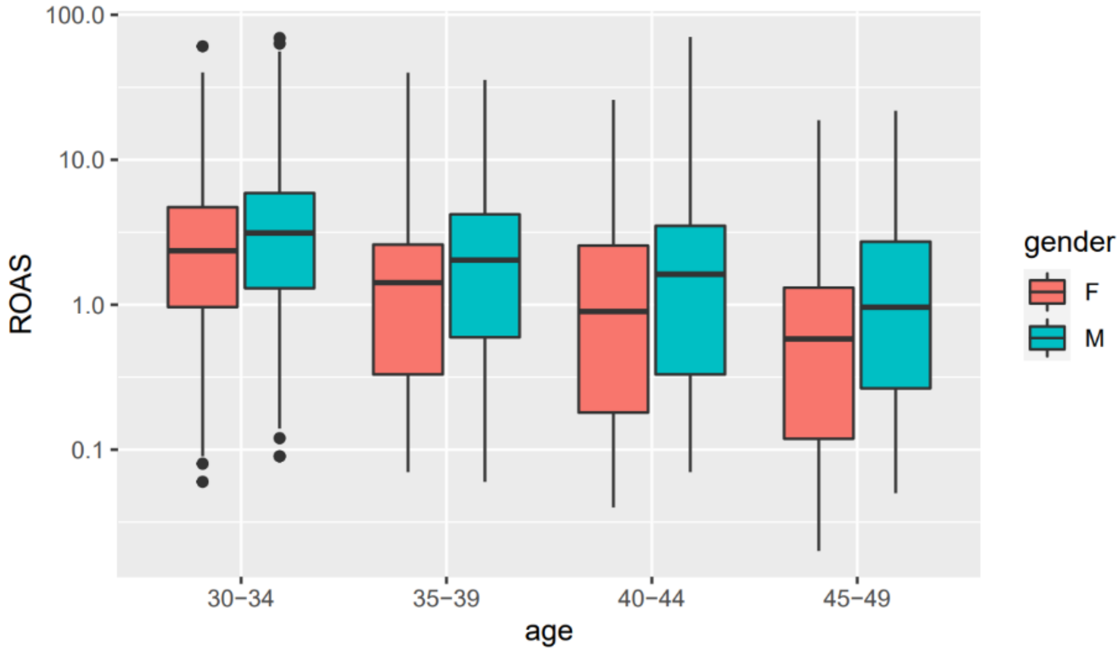


Figure 3. ROAS for Campaign 1178 by age and gender.

Figure 3 explores the ROAS further by visualising gender and age groups. This shows similar patterns relating to gender performance. Overall males in each category generate a higher ROAS, this is not unexpected given the previous insight (figure 2). However, it is interesting to note how the distribution between gender in each age category is evenly skewed, giving us further insight into how the difference in engagement seems to correlate with age. The advertising campaign 1178 is generating the highest ROAS in the age category 30-34, again males in this age group generated a higher ROAS than women, but overall, this demographic represents a relevant target audience who are actively engaging with the advertising content. The table below displays the mean and median ROAS for males and females in campaign 1178.

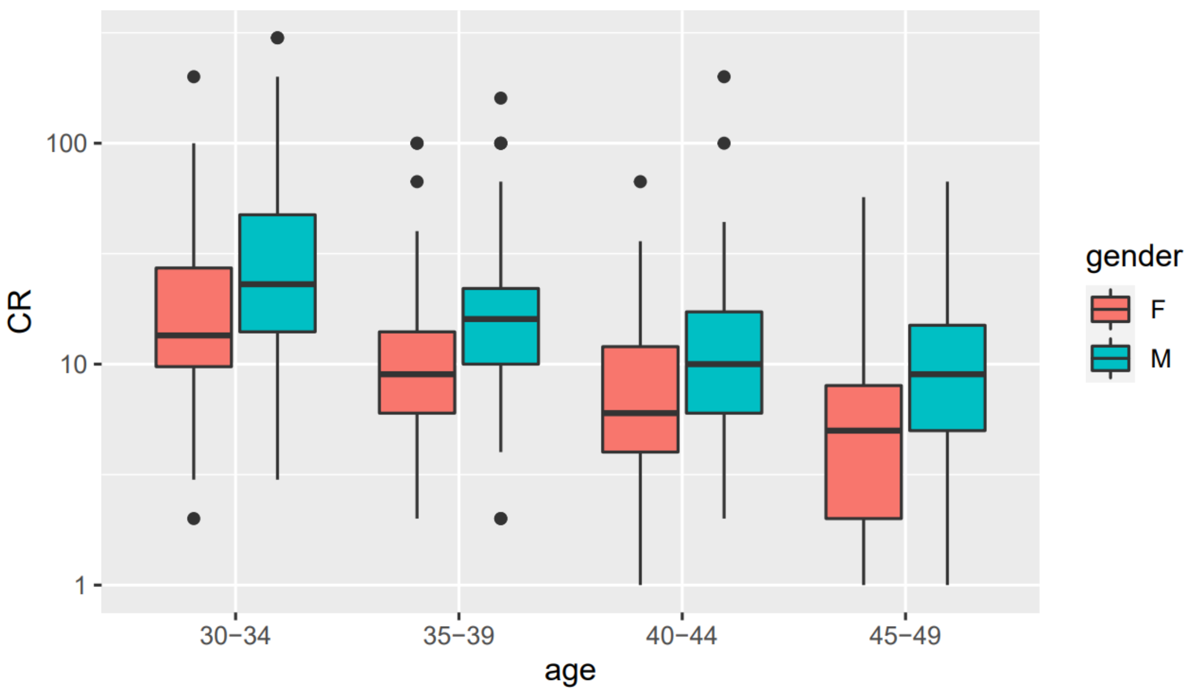
Age	**Gender**	**Median ROAS**	**Mean ROAS**
30-34	Male	3.13	7.257663551
30-34	Female	2.35	4.939294118
35-39	Male	2.025	3.761022727
40-44	Male	1.625	3.903529412
35-39	Female	1.42	2.518245614

45-49	Male	0.96	2.091025641
40-44	Female	0.85	1.957
45-49	Female	0.58	1.229

Table 1. Mean and median average ROAS

Table 1 shows that males in the 30-34 category generate the highest ROAS with a mean average of 7.2, then women in the 30-34 age group with a mean average of 4.9. The ROAS is a metric to evaluate campaign effectiveness by calculating how much you receive in return for the amount spent on advertising. You can see from the table what there are wide differences between age and ROAS, with each group returning a lower mean average, the older the age group. Therefore, it would be beneficial to increase advertising spending for males and females aged between 30-34 while looking at decreasing advertising towards 45-49 year olds, or use this information to perform further research and investigate why levels of engagement are low among this group and see how they can make their advertising more relevant to this target group.

Figure 4. Conversion rate for Campaign 1178 by age and gender



You can see from figure 4 that the conversion rate, when measured by age and gender, tend to follow a similar pattern as the ROAS. However, you can see that males in age category 35-39 perform very well when measuring this metric, with a higher median conversion rate than women in the 30-34 category (15.5% compared to 13%). As you can see below when measuring the median, the two categories are pretty equal (22.2% for males between 34-39 compared to 22.8% for females ages 30-34). Figure 4 represents a fairly symmetrical distribution of the datapoints, however there are a lot of outliers for all age groups apart from the 45-49 age group, which may pull the mean away from the central tendency, for that reason the median will used to determine the most accurate conversion rate.

Age	Gender	Median CR	Mean CR
30-34	Male	23	43.514019
35-39	Male	15.5	22.159091
30-34	Female	13	22.882353
40-44	Male	10	17.161765
35-39	Female	9	13.824561
45-49	Male	9	12.961538
40-44	Female	6	9.15
45-49	Female	5	6.6428571

Table 2. Mean and median average CR

Overall, males perform the best in terms of conversion, which tells us that the advertising campaign is currently more engaging towards men. Therefore, it would be beneficial to continue advertising spending in this area, particularly among men between 30-39 years old. While taking further steps to either reduce spending among females aged 40-49 as they return a low conversion rate compared to all the other groups under analysis, with a mean average of 6% for female aged 40-45 and 5% for females aged 45-49.

Figure 5. Conversion rate by interest identifier

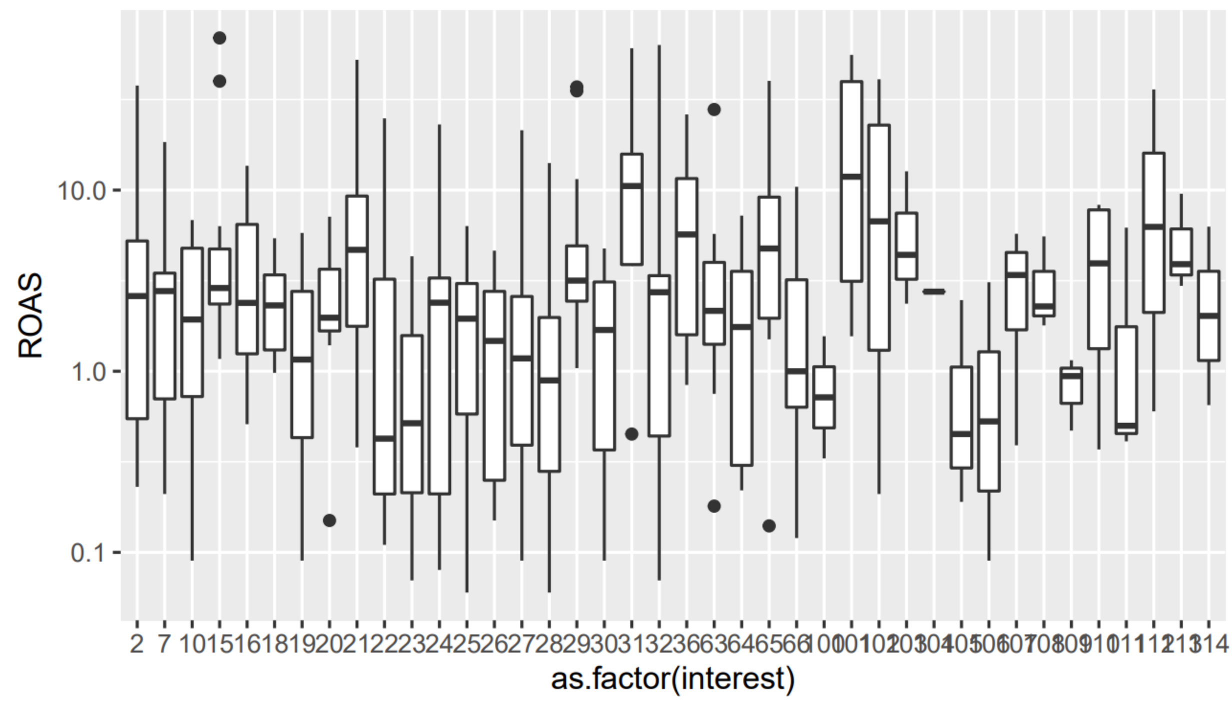
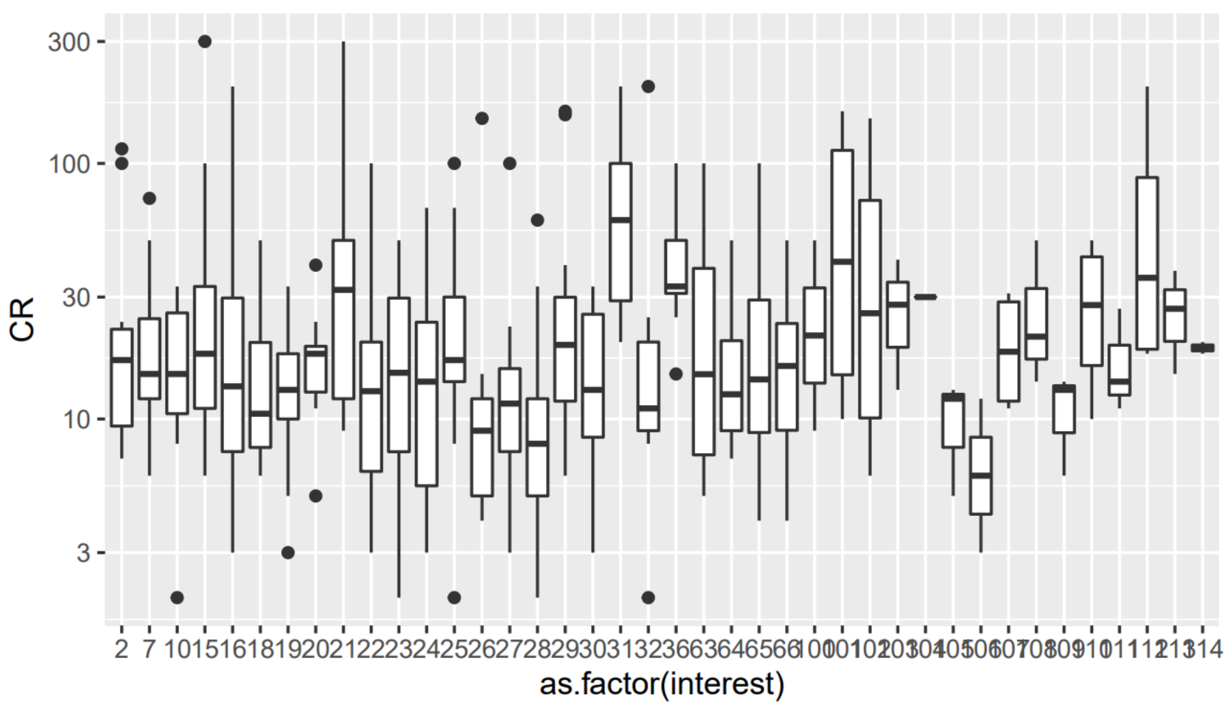


Figure 6. ROAS by interest identifier



Due to the high levels of performance from the groups aged between 30-40. We then filtered the data further to look into the particular interest ID's within this age group which were performing the most effectively. The boxplots above (figure 5 and 6), show the distribution of the conversion rate and ROAS when looking at particular interest identifiers. The interest identifiers tell us which particular group a user belongs to. What determines your interest ID are demographic values, collected from

your Facebook profile. This is what makes Facebook data particulary useful to perform data analysis on, as they collect a lot of user information which can draw insight relating to your business needs.
As you can see there are significant differences in the distribution of the interest identifiers. Some have a very large distribution of data, with significant outliers skewing the results, while others have a shorter distribution of data. To investigate these differences we investigated further interest ID performance. The table below shows the interest ID's which had the highest conversion rate and ROAS average.

Interest	Median ROAS	Mean ROAS
101	19.715	24.1925
31	10.53	18.27
102	10.605	15.5775
112	7.715	12.9925
15	2.88	10.91461538
65	4.775	10.38833333
21	4.68	9.432307692
36	5.69	8.142222222
32	2.73	7.741818182
29	3.165	7.681875

Table 3. Average ROAS by Interest ID

Interest	Median CR	Mean CR
31	60	81.8
112	43	76
101	58.5	71.75
102	34	56
21	32	53.461538
15	18	51
36	33	43.777778
29	19.5	36.25
2	17	31.181818
16	14	31.166667

Table 4. Average CR by Interest ID

The tables above shows clear similarities in the interest ID's which tend to return the highest ROAS and CR. Interest ID's 101, 31, 112 and 102 are the top four for both KPI's. Interest ID 101 averages the highest mean ROAS of 24.2 and would therefore be beneficial to increase advertising spending here as for every £1 spent on advertising it is returning over £24. It would also be beneficial to increase brand awareness by increasing ad spending for interest ID 31 as the CR is very high, which tells us that our advertising campaign is very relevant to audiences of this interest group. As identified previously, broadly the disparities between males and females for both ROAS and CR is relatively small, and therefore we would suggest increasing advertising spend on Facebook for males and females aged between 30-40 years, this would be beneficial in terms of profitability and increasing the overall number of conversions by targeting audiences who have high levels of relevance.

3.1 Discussion of datathon experience

Overall, the group worked together cohesively, fostering a collaborative atmosphere which enabled idea generation throughout and collective group work when facing implications. When allocating roles we all discussed our previous experience, skills and which roles we would prefer. Given our varied educational backgrounds, with some team members more skilled in data science skills and others in project management and marketing insight, allocating roles was fairly straight forward. Every team member worked colloboratively but also independantly on the final presentation, adding the relevant information from their role. While it was the task of the project manager to then finalise the presentation by editing and proof reading, ensuring fluidity. The challenges we faced during this datathon are primarily placed on the time constrictive nature of the datathon and the overall lack of expertise in marketing ad analysis and calculating KPI’s. A given amount of time was allocated to researching the relevant KPI’s we should use for analysis and improving our result accuracy through a short trial and error period, where we performed preliminary anaysis on the dataset to decide which variables to focus on. Given a wider time frame or a deeper insight into the topic would have enabled us to use our time more efficiently, generating further insight and more accurate results. Given that this was the first datathon each team member had participated in, the team performed well and were flexible within their roles, with each team member communicating well with one another to ensure the project ran smoothly and were efficent in multi-tasking, taking on responsibilities outside of their pre-defined roles to alliviate the workload when required.

4. Conclusion

The implementation of a short-form datathon was overall successful in exposing students to a new, collaborative way of carrying out a data science project. The competitive nature promotes quick thinking, originality and idea generation throughout. While, creating an environment which focuses on effective team work to promote successful completion. The time restraints limit the potentially for disputes and foster a decisive attitude among team members, as constant back and fourth would have been timely and impacted the overall process in a negative sense. It can therefore, be defined as both a challenge and a benefit to the project. Overall, the datathon served as a exciting prospect for research methods, it is advantageous for inspiring high levels of efficiency towards problem solving, while providing a competitive but inclusive environment to learn new skills and improve individually.

Bibliography

Andrus, A. (2020). What is ROAS? The Complete Guide to Using Return on Ad Spend. Disruptive Advertising. [online] February 7. Available at: https://www.disruptiveadvertising.com/marketing/roas-return-on-ad-spend/

Aboab,J. Anthony, C. Charlton, Mengling, F. Mohammad, G. Dominic, C. Marshall. Mayaud, L. Naumann, T. McCague, N. Paik, K. Pollard, T. Resche-Rigon, M. Salciccioli, J. Stone, D. (2016). A “datathon” model to support cross-disciplinary collaboration. [online] Available at: Science Journals — AAAS (sciencemag.org)

Bollier, D. (2012) The Promise and Peril of Big Data. [online] Available at: https://scholar.google.co.uk/scholar?q=David+Bollier,+(2012)+The+Promise+and+Peril+of+Big+Data&hl=en&as_sdt=0&as_vis=1&oi=scholart

Frost, J. (2018). Measures of Central Tendency: Mean, Median, and Mode. [online] Available at: https://statisticsbyjim.com/basics/measures-central-tendency-mean-median-mode/#comments

Mackey, M. (2021). What Is Click-Through Rate & Why CTR Is Important. Search Engine Journal. [online] February 16. Available at: https://www.searchenginejournal.com/ppc-guide/click-through-rate-ctr/#close

Newberry, C. (2019). The Facebook Pixel: What It Is and How to Use It. Hootsuite. [online] January 14. Available at: https://blog.hootsuite.com/facebook-pixel/

Zumel, N. Mount, J. (2019). Practical Data Science with R.