

Data-DRIVEN MARKETING USING GOOGLE ANALYTICS

Industry Project Final Report

December 8, 2018

Absolute Analytics Ltd.

14/16 Maid stone street, Grey Lynn

By: Rutika Chavan

Supervised By: Dr. Fadi Fayez



Table of Contents

[**EXECUTIVE SUMMARY** 4](#_Toc532025222)

[**INTRODUCTION** 5](#_Toc532025223)

[**DIGITAL MARKETING AND GOOGLE ANALYTICS** 6](#_Toc532025224)

[**Digital Marketing** 6](#_Toc532025225)

[**Google Analytics** 6](#_Toc532025226)

[**CURRENT SYSTEM** 7](#_Toc532025227)

[**PROBLEM WITH THE CURRENT SYSTEM** 7](#_Toc532025228)

[**RECOMMENDED SOLUTION** 7](#_Toc532025229)

[**MEASURABLE ORGANISATIONAL VALUES** 8](#_Toc532025230)

[**SCOPE AND OBJECTIVES** 8](#_Toc532025231)

[**Scope** 8](#_Toc532025232)

[**Objectives** 8](#_Toc532025233)

[**METHODOLOGY** 8](#_Toc532025234)

[**PROJECT** **TIMELINE** 9](#_Toc532025235)

[**DATASET DESCRIPTION** 10](#_Toc532025236)

[**DASHBOARDS** 11](#_Toc532025237)

[MATHEMATICAL TECHNIQUE USED FOR PREDICTIVE ANALYSIS 15](#_Toc532025238)

[**Experimental Settings** 15](#_Toc532025239)

[**Learning Algorithm** 15](#_Toc532025240)

[**Evaluation Measures** 15](#_Toc532025241)

[**Performance study of the model** 16](#_Toc532025242)

[**Model results** 16](#_Toc532025243)

[**STATUS & DELIVERABLES** 18](#_Toc532025244)

[**RECOMMENDATIONS** 18](#_Toc532025245)

[**Predict Conversions in Google Analytics** 18](#_Toc532025246)

[**CONCLUSION** 18](#_Toc532025247)

[APPENDIX 19](#_Toc532025248)

[Gantt chart 19](#_Toc532025249)

[19](#_Toc532025250)

[Dashboard1 20](#_Toc532025251)

[Dashboard 2 21](#_Toc532025252)

[References 22](#_Toc532025253)

[Figure 1: Digital Analytics Methodology 5](#_Toc532018789)

[Figure 2: No of sessions in previous one-week vs current week. 7](#_Toc532018790)

[Figure 3 : Top 3 regions having maximum number of sessions 7](#_Toc532018791)

[Figure 4: Sessions and Pageviews by channel 8](#_Toc532018792)

[Figure 5: Engagement by Age and Gender 8](#_Toc532018793)

[Figure 6 : Top 5 products by revenue and average price 9](#_Toc532018794)

[Figure 7 : Product revenue and E-commerce conversion rate by source 9](#_Toc532018795)

[Figure 8: No. of sessions by Revenue and by device. 10](#_Toc532018796)

[Figure 9: Logistic regression model in R. 12](#_Toc532018797)

[Figure 10: Output of logistic regression model. 13](#_Toc532018798)

[Figure 11: Threshold curve 14](#_Toc532018799)

# 

# **EXECUTIVE SUMMARY**

This report starts with the introduction of the company and its background, providing information about the organisation, current processes used by the organisation for deliverables and goals. Structure of the organisation is necessary to achieve the business objectives which ae outlined in the Measurable Organisational Values(MOVs). The identification of problem areas and opportunities provide a clear direction to which process should focus on. I have identified that the lack of a proper system in place to answer the business questions by clients and a predictive analytics process to be implemented in their work overview to ensure they meet the client requirements. This was done upon the review of the existing processes, from which key performance indicators (KPIs) were identified, forming the basis of the model and implementation of RStudio Logistic regression model and increasing few KPI’s in the Google Analytics Dashboard. All changes in business processes were clearly documented in Microsoft word, with details contained within the process document. I have adopted the Digital Analytics methodology in this project. Gantt chart was used to manage my time and task for the project effectively. As I was not provided with user access to the Google Analytics API database of the client, I have extracted the data manually to demonstrate the full potential of the logistic regression working model. The project is concluded with a Predictive Analytics Logistic regression model and Google Data Studio Dashboards.

# **INTRODUCTION**

Absolute Analytics is a Digital Analytics Company specialist in advanced web analytics and Google Analytics. They provide services like Initial Google Analytics Audit, which involves set-up for their clients who are new into business and Google Analytics Audit for existing GA users. The have different packages as per client requirement wherein different services which help businesses to derive measurable results.

The purpose behind approaching Absolute Analytics for Industry project is to explore Digital Analytics and apply machine learning on the data extracted from Google Analytics. One of the service which absolute analytics provides is the e-commerce setup and tracking for their clients. E-commerce set-up help businesses to track their transactions, see online traffic sources and provide a thorough analysis to establish a path which will lead to maximize their ROI.

Google Analytics is a Digital marketing platform, which tracks and reports website traffic. Google Analytics is extensively used web analytics service on the web. Google Analytics also provides service known as Google Analytics for Mobile Apps which collects data from iOS and Android apps.

For this project, I have extracted data for one of the clients of Absolute Analytics and apply machine learning on it to determine the probability of conversion.

# **DIGITAL MARKETING AND GOOGLE ANALYTICS**

## **Digital Marketing**

Digital Marketing encloses all the marketing efforts that are done using online resources such as electronic device or internet. Business are making use of all the available social media channels, search engines, email and their own website to engage their existing as well as potential customers.

The term “Digital Marketing” was first coined in 1990. With the growing era of marketing on web from the 90’s to 2012, it has helped companies to design their marketing campaigns as per customer segmentation, by launching multi-channel campaigns and provide personalized information for customers. The growth of social media has played a vital role in development of Digital marketing. Consumers became highly dependent on their smart phones, tablets and laptops etc. and use social media platforms like Facebook, YouTube, Instagram and LinkedIn etc. Hence it became necessary to provide them with impeccable user experience for all the different channels where they were searching product information.

Digital marketing is also referred to as “online marketing” or “web marketing”. Increasing growth of digital marketing also raised concerns for Online Behavioural Advertising of Consumer privacy and data protection.

## **Google Analytics**

Google Analytics is a tool which helps us understand the number of website visits and why have the users visit your website. After the configuration, website is connected to Google Analytics account with the implementation of tracking code. Once the tracking tags are implemented correctly, it tracks the number of visitors to page, create goals to see how visitors are converting into subscriptions, track e-commerce transactions etc.

Adding Analytics code to website is the first step towards tracking your website. Google Analytics gathers the following data:

* Visitors data: How many people visited your website, number of pageviews per day, which operating system the user has used and the region from which user is accessing the account.
* Traffic Sources: traffic sources determine from which online source the traffic is coming to your website. Also, the keyword used by the user can be tracked.

Now, we will have a look at few basic metrics of Google Analytics(GA).

* Sessions: Sessions column shows the sum of interactions by user on website within a given timeframe (GA has 30 minutes as one session)
* Pageviews: A pageview hit in GA is an instance of page being reloaded in browser. It gives the total number of pages viewed.
* Goals: It helps you to track specific user interaction on your website. The user interaction can be anything including button click, link click, page submission etc.
* Conversion Rate: The conversion rate shows the sum of overall conversion rate on your website.

# **CURRENT SYSTEM**

* Absolute Analytics do not have any system to preform predictive analysis. Currently, they are using Google Analytics Premium account, Google Analytics 360 to get insights as per client business questions (Increase web traffic).
* Google Data Studio to prepare Dashboard based on historical data from Google Analytics Account.
* Web Analytics Consulting and providing strategy solutions which help businesses to take their digital offering to next required level.

# **PROBLEM WITH THE CURRENT SYSTEM**

* No specific process to do some detailed predictive analytics. It is important for businesses to understand the trend of data from descriptive analytics and to determine the probability based on conversions.
* Need to define a process which will help their clients to determine the probability of conversion of the next user.

# **RECOMMENDED SOLUTION**

The project started off with looking for clients who were interested in determining the probability of conversion for the next user.

Initially, as per suggestion from the director of Absolute Analytics, we were to get the read access from Manukau Institute of Technology, but it could not be carried forward as the approval was delayed by our project co-ordinator. Hence after that, we decided to approach one of the clients from Absolute Analytics. I attended several client trainings to pitch in the idea, to provide them with predictive analytics service as an add on to the existing service they had purchased with Absolute Analytics for free. After approaching several clients, we got the approval to extract data from Countdown, but they did not give access to their Google API, so we had to extract the data manually from the freemium account from one of the properties of Countdown Google Analytics account.

# **MEASURABLE ORGANISATIONAL VALUES**

|  |  |
| --- | --- |
| **Area of Impact** | **Explanation** |
| Customer Focused | Achieved by ensuring that prospective consumers are increased by mapping the solutions to their marketing strategies. |
| Financial Impact | Increase in revenue by increasing the probability of conversions |

# **SCOPE AND OBJECTIVES**

## **Scope**

Limited on to working to improve the conversion rate using Google Analytics data from Countdown account.

## **Objectives**

* Use Customer behaviour information from their websites and re-engage with them to help them take a decision.
* Deliver better experience for users through marketing activities.

# **METHODOLOGY**

I have chosen to work on this project using the Digital Analytics methodology. Digital Analytics is the analysis of qualitative and quantitative data, which is used to improve the online experience of existing and potential customers which leads to achieve the desired offline and online outcomes.

Figure : Digital Analytics Methodology

Digital Analytics run in parallel with other marketing activities providing the business with real-time supportive information. The Analytics cycle consist of four main steps: Measure, Analyse, Report and Test.

**Measure:**

Collect all the data needed for you to understand your audience, both quantitative and qualitative. An Analytics suite should be combined with live chat, user surveys, heatmaps and videos.

**Analyse:**

Extract actionable information and identify reasons for any patterns and anomality’s in your data. This includes segmentation, attribution and competitive analysis. You’ll want to use industry benchmarks to put the data into context.

**Report:**

Condensethe analysis into easy to digest information, with clear explanations of how the data should inform business decisions.

**Test:**

Findthe best solutions to the problems being identified during the analysis. Testing eliminates any preconceptions and biases from the decision-making process.

The analytical cycle should run in parallel with other marketing activities; provide unbiased actionable recommendations, give additional insights into the results of marketing and drive continual improvement.

# **PROJECT** **TIMELINE**

The Gantt chart, attached in Appendix section, was used to track my progress. This documented the breakdown of the tasks within each area, while accounting for any additions or removals to deliverable requirements. Due to changes in the project requirements put forth by the client, such as there was no longer the need for a data warehouse and alerts system, it impacted on the project timeline significantly as I had initially allocated 120 hours to it that were no longer required. However, as I had to populate dummy data for additional dashboard, I spent 35 hours doing so. During the entire duration of the project the Gantt chart served as a very useful visual tool to help me in my time and task management of the project.

# **DATASET DESCRIPTION**

The Dataset was obtained from Google Analytics account from property Countdown-main view. Dataset has 384 instances and one class variable as “conversion rate” with 9 attributes. Some attributes are Numeric, Nominal, Decimal, Boolean and Binary. The dataset is a weekly data from November 10, 2018 to November 19, 2018

The following is the representation of attributes:

|  |  |  |
| --- | --- | --- |
| **ATTRIBUTE** | **DESCRIPTION** | **TYPE** |
| Client Id | Unique user id | Numeric |
| Sessions | No. of sessions by client id | Numeric |
| Medium | Channel by which the product was searched by user | Nominal |
| Keyword | Keyword by which the product was searched by user | Nominal |
| Source | online source by which the user searched for product | Nominal |
| Pageviews | No of page views in a session by user | Numeric |
| Revenue | Revenue generated by each client id | Decimal |
| Transactions | No. of transactions by each Client Id | Numeric |
| Conversion Rate | Transaction was a conversion or not | Binary |

# **DASHBOARDS**

For data visualization of the extracted dataset of Countdown Google Analytics Account. Google Data Studio is used to prepare Dashboards as per the Business questions from clients.

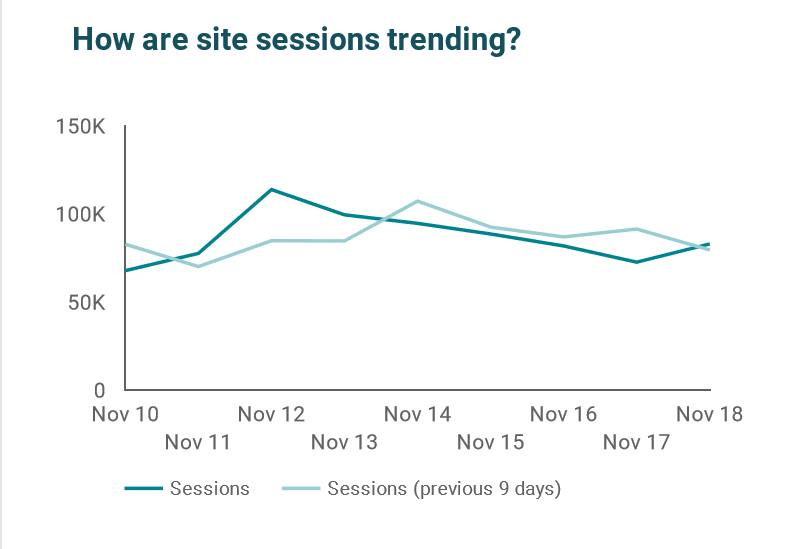


Figure : No of sessions in previous one-week vs current week.

The above chart shows the number of sessions in the current week vs the number of sessions in the previous week. This chart helps to compare the performance of website for every week.

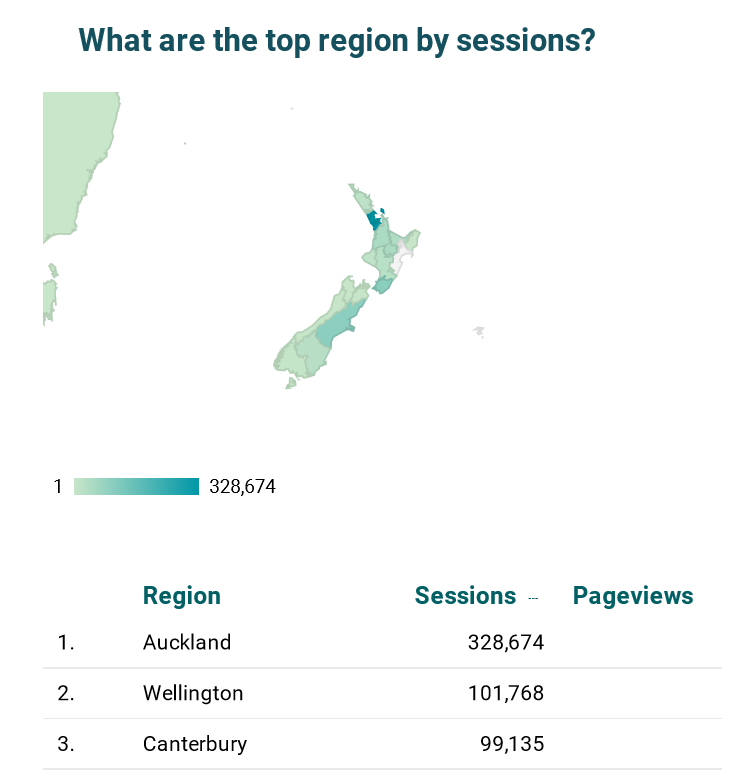


Figure : Top 3 regions having maximum number of sessions

The above chart shows the top three regions where there are maximum number of sessions. This helps to determine the region from which the maximum traffic is coming.

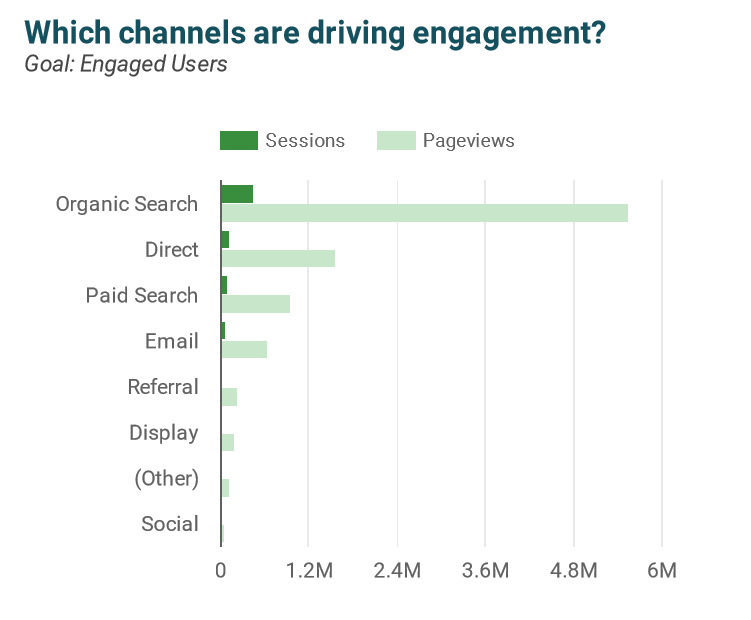


Figure : Sessions and Pageviews by channel

The above chart shows us which channel is responsible for engaging its users to the website. Above chart gives information about engagement by sessions and pageviews. Here Organic search is giving the maximum engagements in terms of page views and sessions followed by Direct search and Paid search.

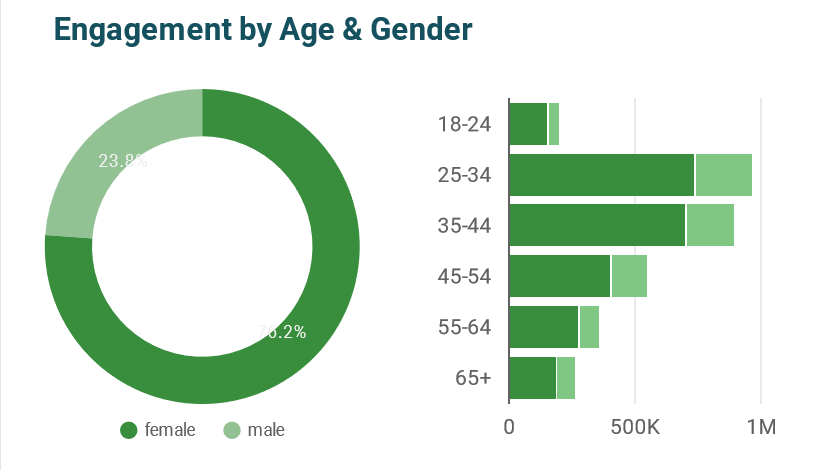


Figure : Engagement by Age and Gender

The above chart shows engagement by Age and Gender. Here the age group 25- 34 shows maximum engagement and the male to female ratio of engagement is 76 % and 24% respectively. Which shows maximum users visiting the website are females.

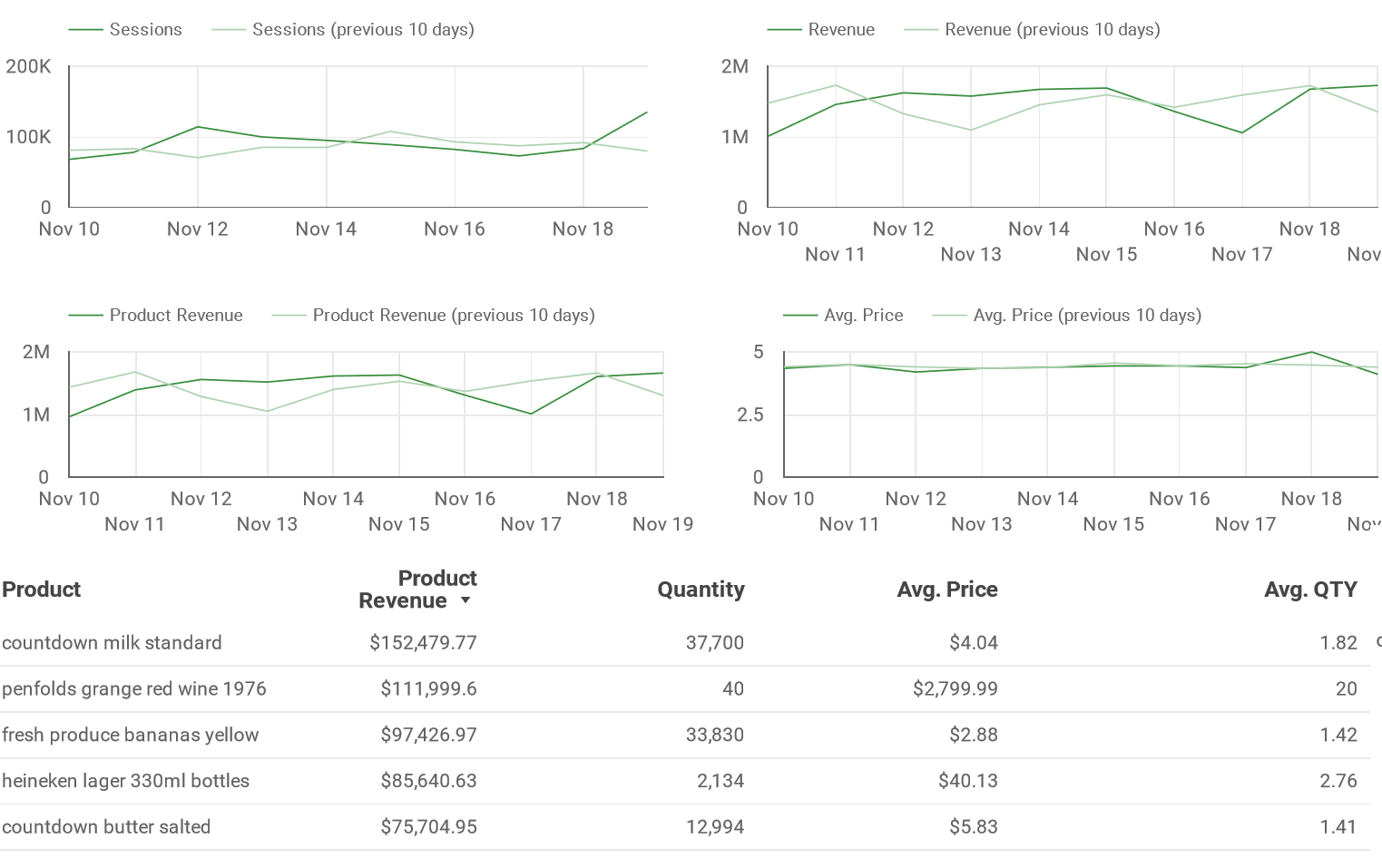


Figure : Top 5 products by revenue and average price

The above chart shows the top 5 products in terms of revenue, the quantity sold for each product in that week. Average price for which the product is sold and the average quantity. Here Countdown milk standard is the highest revenue giving product followed by ‘Penfolds grange red wine’ and ‘fresh produce yellow bananas’.

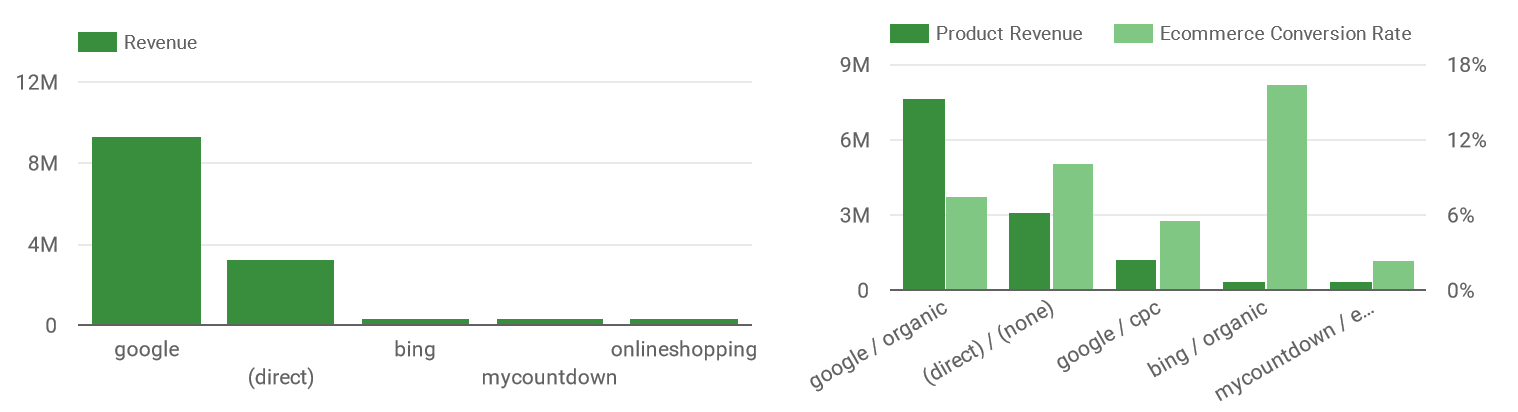


Figure : Product revenue and E-commerce conversion rate by source

The above chart shows revenue by channel, the product revenue and E-commerce conversion rate by channel. Here the highest revenue is from Google Search which is a form of Organic or unpaid search. Similarly, the product revenue is highest for Organic search. But the e-commerce conversion rate is highest for organic search which is through Bing channel.

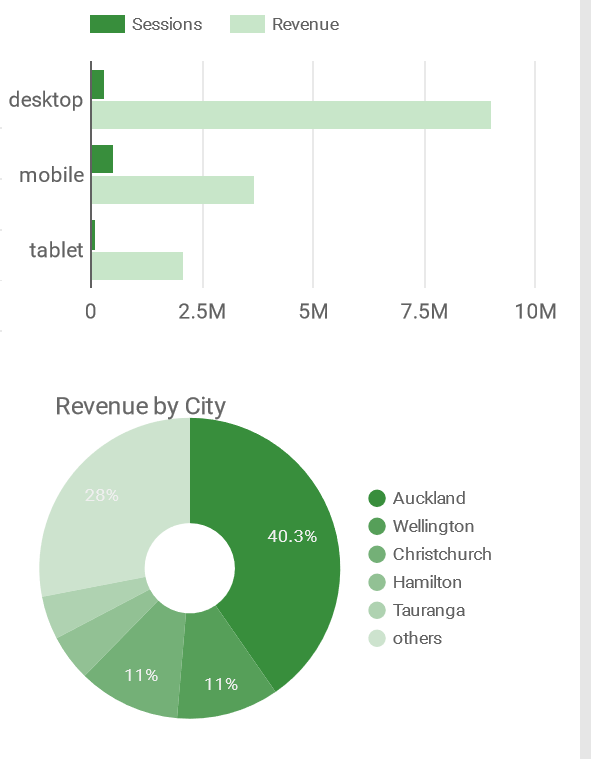


Figure : No. of sessions by Revenue and by device.

The above chart shows revenue by city and the device which are driving maximum users. Here, Auckland is the highest city in terms of revenue followed by wellington and Christchurch, which is a bit obvious looking at the population of the cities. In terms of devices used, maximum revenue is coming from desktop device, but the sessions are maximum for mobile/ smart phone users.

# MATHEMATICAL TECHNIQUE USED FOR PREDICTIVE ANALYSIS

## **Experimental Settings**

For Predictive Analysis, RStudio software is used to do the experiment for machine learning. RStudio is a free and open-source integrated development environment for R, a programming language for statistical computing and graphics. RStudio is a strong software, which helps to apply machine learning techniques for various real-life problems. RStudio provides many packages for data science like ggplot2, dev tools, crate etc. The interface for experimental analysis & results interpretation is also very easy to use and understand.

Machine learning technique, Logistic regression is used for analysis and experimental results are obtained from the Google Analytics dataset for Countdown. For test options, the dataset is divided into training and testing dataset in a split ratio of 75% and 25%. Training dataset is 75% and testing dataset is 25%.

## **Learning Algorithm**

**Logistic Regression:** Logistic regression was developed by David Cox in 1958. In regression analysis, logistic regression is estimating parameters of a logistic model, which is a form of binomial regression. Mathematically, a logistic regression model has a binary dependent variable which has two possible outcomes, such as true/false, yes/no, pass/fail. They are also represented by an indicator variable in which values are represented as “0” and “1”. In a logistic model, the logarithm of odds for the value which are labelled “1” is a linear combination of one or more predictors, predictors can be a binary variable or a continuous variable. The probability of the predictor variable labelled can vary between “0” and “1”. Hence the function which converts the log odds to probability is the logistic function.

Consider a model with three predictors , and respectively. These predictor variables can be continuous or binary taking value between 0 and 1. Then, the model can be represented in general form as:

Where are the coefficients of the model, i = 1,2,3…n. The above model is a linear model hence the log-odds are the linear combination of the independent variables , and respectively.

## **Evaluation Measures**

Models are evaluated under Accuracy, Precision, Recall, F-measure and ROC as evaluation matrices:

1. Precision: It is the Positive Predictive Value. It is the average probability of relevant retrieval.

Precision =

Where is the true positive rate and is the false positive rate.

1. Recall: It is the True positive rate. It is the average probability of complete retrieval.

Recall =

Where the true positive is rate and is the false negative rate.

1. Accuracy: It is the proportion of correct classifications.

Accuracy =

Where the false positive is rate and is the false negative rate.

1. F-measure: It is the harmonic mean of precision and recall.

F-measure = 2.

1. ROC: The ROC curve is created by plotting the true positive rate against the false positive rate at various threshold settings.

## **Performance study of the model**

The model was evaluated with several configurations and presenting some of results in this section. From these models, the accuracy is calculated and analysed. 8 attributes were used for learning the model. Performance is determined on the Precision, Recall, Accuracy. F-measure and ROC.

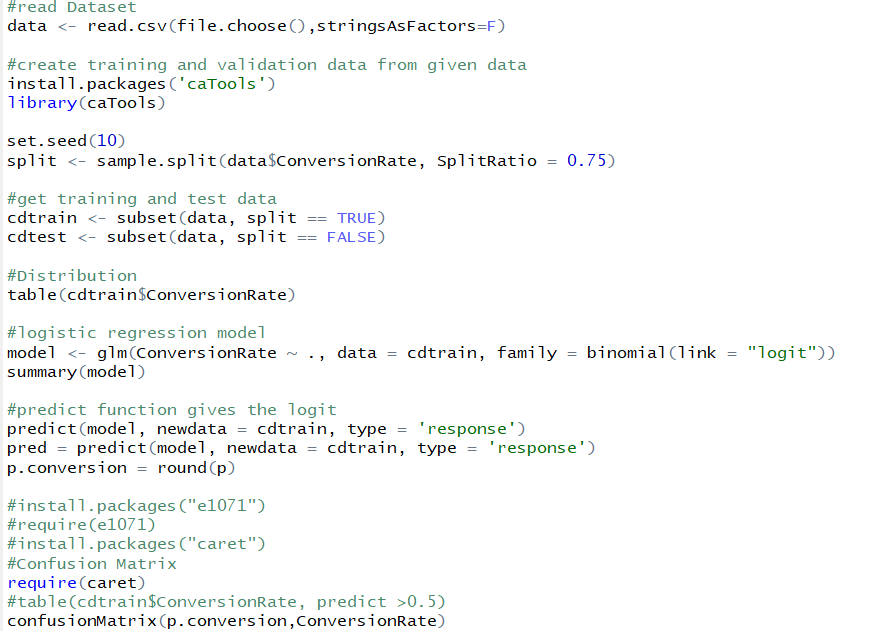


Figure : Logistic regression model in R.

## **Model results**

Following were the model results obtained from the experiment.

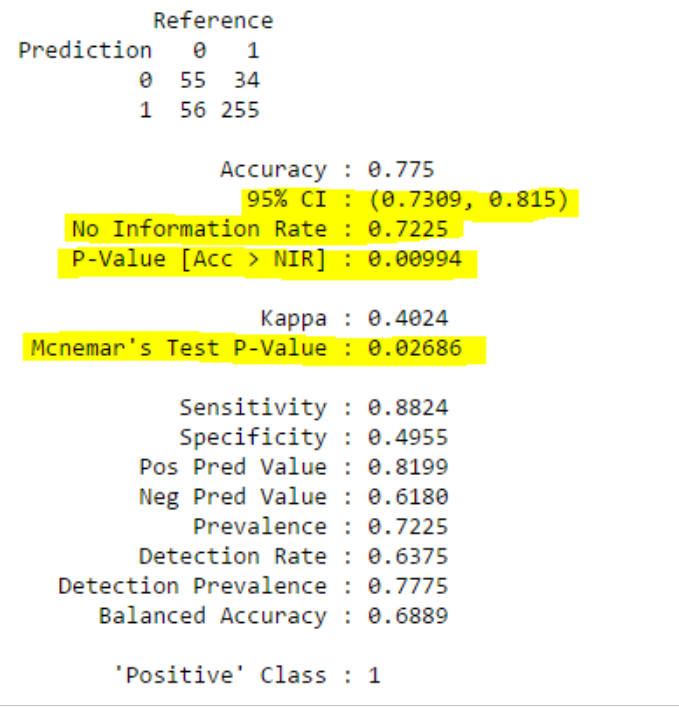


Figure : Output of logistic regression model.

The Table below summarizes the model results obtained from the experiment. We can see that the Accuracy of the model is 77.5%. Accuracy of the model can be improved by adding more independent variables to the dataset which will help in predicting the dependent variable more accurately. Some of the independent variables which can be included are age, gender, region , average time of session, device used etc.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CLASSIFIER** | **ACCURACY** | **SENSITIVITY** | **SPECIFICITY** | **AUROC** |
| Logistic Regression | 77.5% | 0.8824 | 0.4955 | 0.8978 |

Table : Accuracy and performance measure of the algorithm.

The following is the ROC (Receiver Operating Characteristic) curve for the Google Analytics dataset. The area under the curve is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. The ROC curve describes the inherent validity of diagnostic tests. The ROC curve can be interpreted as the probability that a randomly chosen diseased subject is rated or ranked a more likely to be diseased than a randomly chosen non-diseased subject. The maximum ROC =1 means that the diagnostic test is perfect in the differentiation between the conversion and the non-conversion. This happens when the distribution of test results for the conversion and the non-conversion do not overlap. ROC =0.5 means the chance discrimination that the curve located on diagonal line in ROC space. The minimum ROC should be considered a chance level i.e. ROC = 0.5 while ROC =0 means test incorrectly classify all subjects with conversion as negative and all subjects with non-conversion as positive that is extremely unlikely to happen in e-commerce Practice.

From the above experimental results done in RStudio, we have obtained threshold curve as 0.8978 which says that the model is predicting 89.78 % yes class, which means that the test is differentiation between the conversion and non-conversion perfectly 89.78%

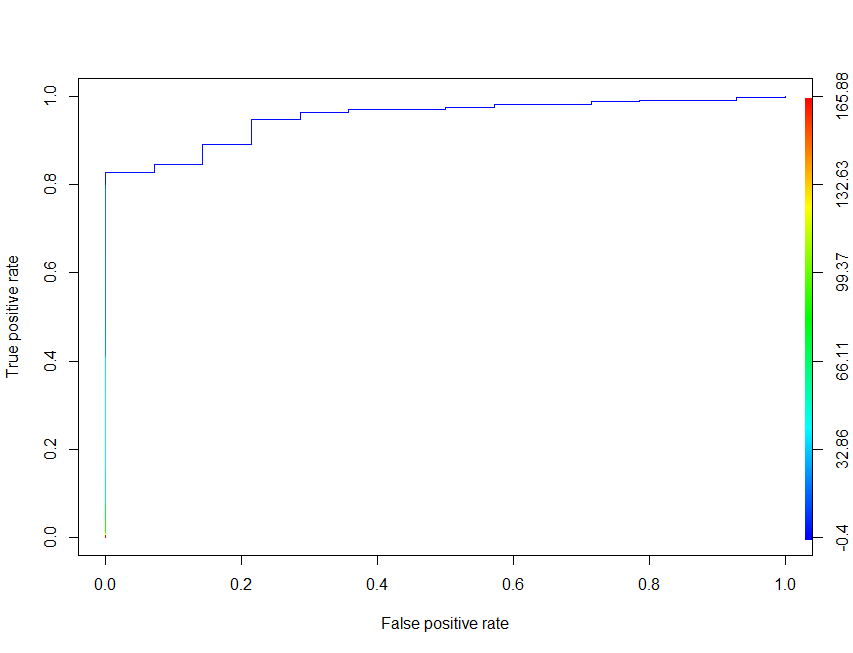


Figure : Threshold curve

# **STATUS & DELIVERABLES**

All the model results obtained from the experimental results performed in RStudio and the Dashboard which answers the business questions based on their historical data in Google Data Studio were send to client as a softcopy via e-mail.

Google Data Studio and RStudio files have been completed and their access is given to Absolute Analytics via the Google Analytics Account created under their company domain.

# **RECOMMENDATIONS**

## **Predict Conversions in Google Analytics**

**PAST: VISIT IN LAST 7 DAYS**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Client ID | Sessions | Medium | Keyword | Source | Pageviews | Revenue | Transactions | Conversion Rate |
| 26098705 | 9 | Organic | Oranges | Facebook | 2 | NZD89 | 3 | 0 |
| 28874887 | 5 | referral | Milk | YouTube | 1 | NZD56 | 5 | 1 |

**PRESENT: A NEW VISIT IN PROGRESS (EXAMPLE)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2378765 | 3 | referral | Wine | Instagram | 2 | NZD93 | 2 | ? |

A logistic regression model which will help their client to predict the probability of conversion for the next visitor to their website based on the previous consumer buying behaviour.

# **CONCLUSION**

This project has helped Absolute Analytics to define a new process for predictive analytics for their clients and help them determine the probability of conversion for the next visitor of their website. Based on the Current processes used to answer the business questions from their clients, team members will need to take an ownership to make their potential clients understand the need to increase the conversion rate which will in turn help them to increase the revenue.

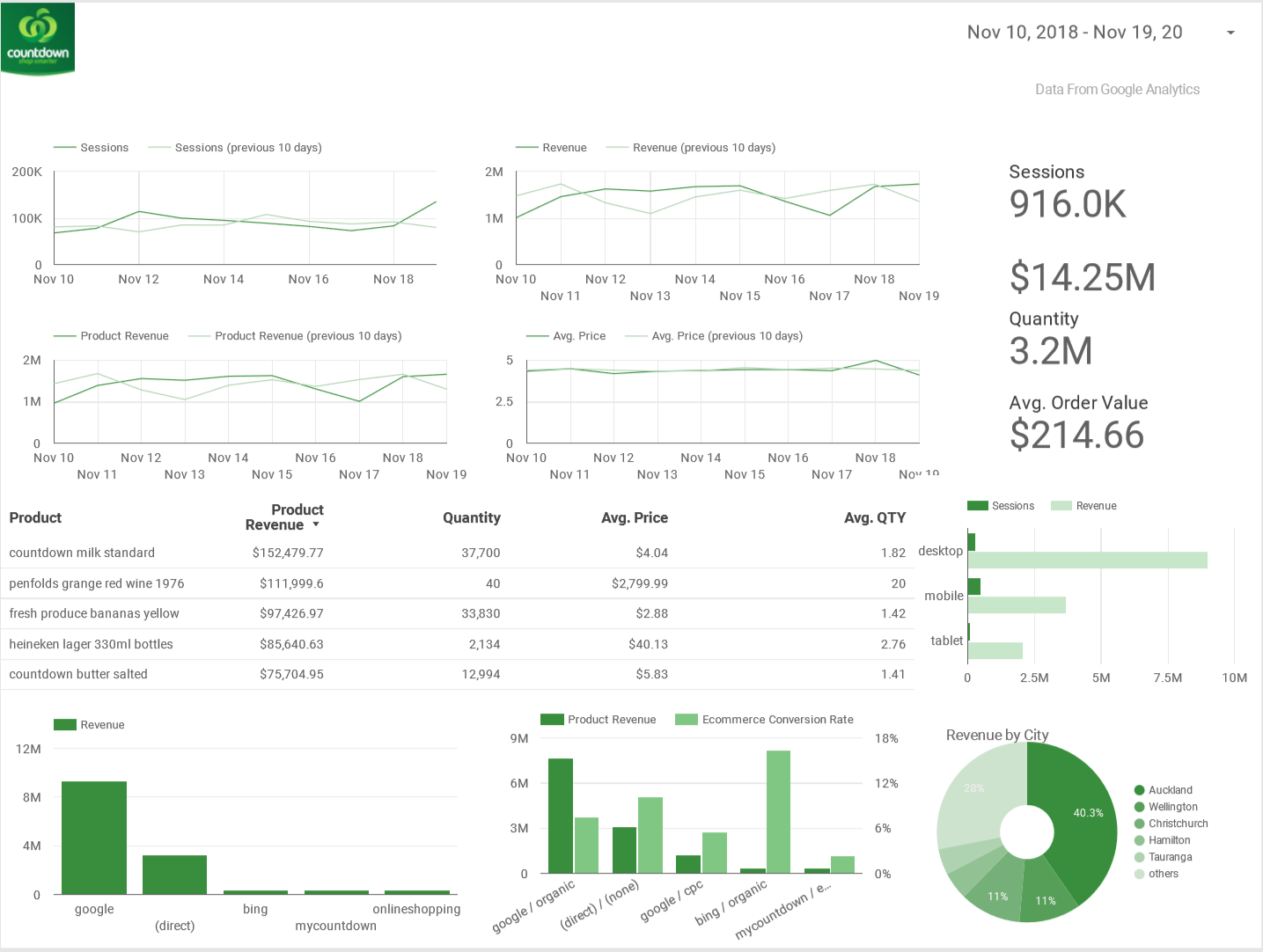
The Predictive analytics model based on logistic regression has been shared with the team which will be helpful to them in strategic planning of the process.

# APPENDIX

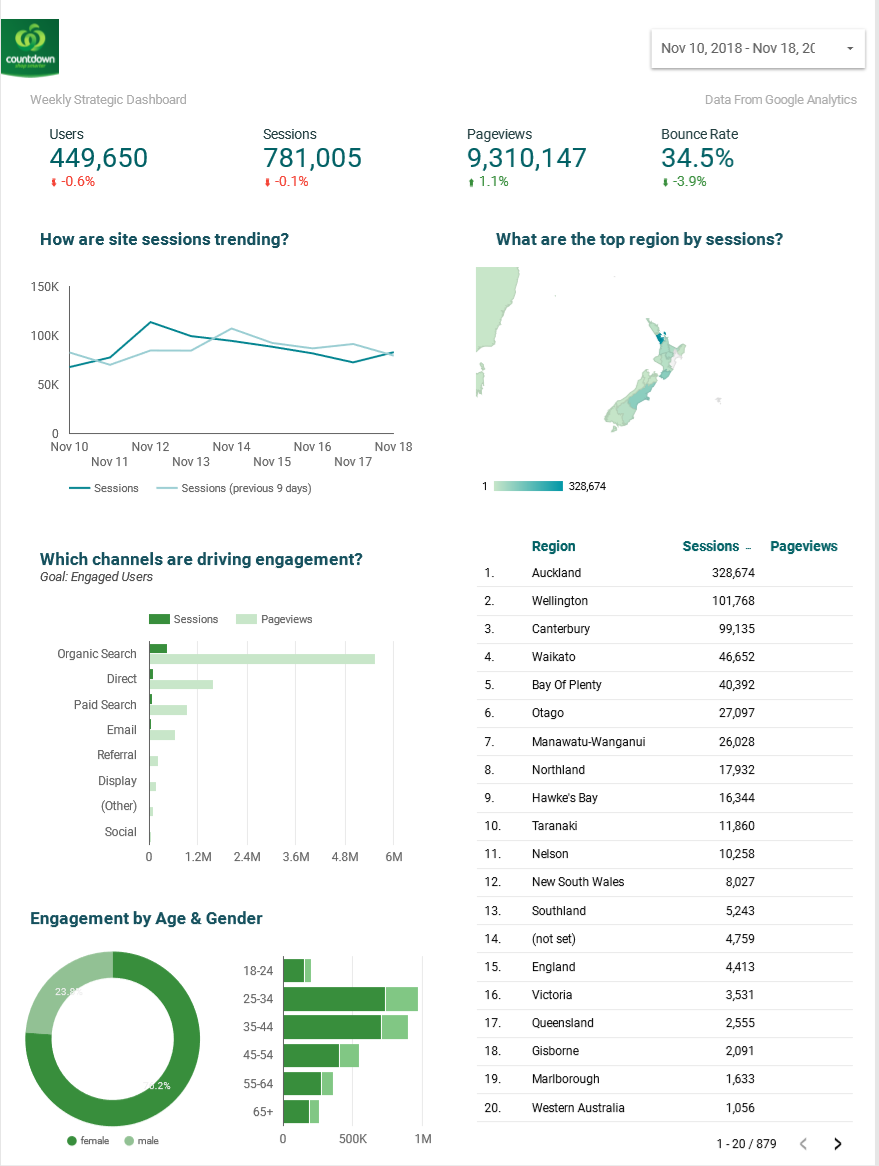
## Gantt chart

## 

## Dashboard1



## Dashboard 2



# References

A.Pop, M. L. (2011). Learning web analytics: A tool for strategic communication. *Science Direct*, 536-543.

Ainscough, T. L. (1996). The Internet for the rest of us: Marketing on World Wide Web. *Journal of Consumer Marketing*, 36-47.

AkiyukiSekiguchiaKazuhikoTsuda. (2014). Study on Web Analytics Utilizing Segmentation Knowledge in Business to Business Manufacturer Site. *Science Direct*, 902-909.

BeatrizPlaza. (2011). Google Analytics for measuring website performance. *Science direct*, 477-481.

By Jerri L. Ledford, J. T. (2010). *Google Analytics.* Google books.

By Sebastian Tonkin, C. W. (2012). *Performance Marketing with Google Analytics: Strategies and Techniques for maximizing ROI.* Google Books.

ConditFagan, J. (2014). The Suitability of Web Analytics Key Performance Indicators in the Academic Library Environment. *Science Direct*, 25-34.

Cutroni, B. J. (2012). *Google Analytics: understanding Visitor Behaviour.* Google Books.

H.PakkalaaK.PresserbT.Christensen. (2012). Using Google Analytics to measure visitor statistics: The case of food composition websites. *Science Direct*, 504-512.

JoelJärvinenHeikkiKarjaluoto. (2015). The use of Web analytics for digital marketing performance measurement. *Science Direct*, 117-127.

Probets, L. H. (2009). Using Google Analytics to Evaluate the Usability of E-Commerce Sites. *SpringerLink*, 697-706.