**Customer Segmentation for a Marketing Campaign**

**Group Members   
Mohammed Ismail siddiqui – MIT221487**

**Jonnagiri Samson Kumar - MIT221110  
Syed Imran - MIT221306**

**Table of Contents**

[Introduction 2](#_Toc123898751)

[Data Analysis Steps 2](#_Toc123898752)

[Data collection 2](#_Toc123898753)

[Cleaning the data 3](#_Toc123898754)

[Analysing the data 3](#_Toc123898755)

[Assumptions 3](#_Toc123898756)

[Customer segments identification 4](#_Toc123898757)

[RFM method 4](#_Toc123898758)

[K means 5](#_Toc123898759)

[Limitation in Analysis 6](#_Toc123898760)

[Conducting marketing campaign 7](#_Toc123898761)

[Conclusion 7](#_Toc123898762)

[References 8](#_Toc123898763)

# Introduction

The report aims to conduct a marketing campaign for the clothing retailer in Australia by identifying the customer segments. The data analysis step is conducted by using python Jupyter notebook and the past sales transaction data of the clothing retailer is analyzed. The main goal of conducting data driven analysis is to attract the customers with targeted promotion and drive them to buy in the company websites. Therefore, with the analysis, the customer segments will be discovered, identify limitation in analysis and explain how the marketing campaign can be conducted to achieve the company’s goal.

# Data Analysis Steps

## Data collection

The significant step is data analysis is to collect the data and for this report, the excel data file “A2Sales.clcs” is read using python script. The rows and columns in the excel data is displayed from header to tail. The attributes of the datasets are:

* DateTime\_Of\_Purchase: - This attribute gives the date and the time of purchase of the product made at the retail clothing store website.
* Customer\_ID: - Provides the unique customer ID for each customer who have made purchase at the company website.
* Network\_Address: - Network address gives the IP address of the system or computer from which the customer has purchased the product.
* Product\_Name\_List: - Gives the list of clothing name such as printed pants, salsa dress, draped dress, mini dress, jackets, tights, etc.
* Product\_Code\_List: - For the specific product list, the product cost list is provided.
* Number\_Of\_Items\_Per\_Product: - For each product the number of items or quantity bought by the customer is given.
* Coupon\_If\_Used: - Any promo codes or coupons used by the customer during buying the product is provided with the coupon name.
* Suburb: -Showing the sub urban area where the customers have purchased the products.
* State: - State of Australia such as NSW, QLD, VIC, etc from where the customers purchased.
* DOB: - Showing the customers date of birth so that we can find the age of customers and their preferences in product purchase.
* Purchase\_Amount: - The amount of purchase made by the customer at the clothing centre is given.

## Cleaning the data

For the sales transaction dataset the data analysis sis performed that includes inspecting, cleaning, transforming and then modelling the given data so as to discover the hidden pattern and information. Here, for the different customer ID there are same values showing duplicate data and they are removed at this stage to ensure consistent result. Another data pre-processing technique used is “StandardScalar” to remove the missing values “NaN” s in the coupon attributes. These missing values are disregarded in fit and affect the model’s performance and hence removed. The standard scalar is referred as “z-score” normalization or the standardization where it transforms the data making the data to be zero and the SD (standard deviation) as one (Azevedo, 2023).

## Analysing the data

***Company’s sales performance***

Once the data is collected, they are analysed and in this study the company’s sales performance is analysed. The purchase year of the product is 2017 and it analyses average sales made by the company in that specific year. For the year 2017, the purchase amount of the products by the customers is “135.9167”.

***Products that are likely to be purchased together***

To find products which are likely to be purchased together are analysed using date time of purchase, product name list. During the year 2017, the product that was most likely to be sold together are “modern cotton bralette” and “modern cotton thong”. In the company website, these two products are mostly purchased together.

# Assumptions

It is assumed that in the beginning of the dataset there were no missing values or noise found. So, the dataset was selected with first 80 rows and found there were same values depicting duplication and it is removed with data preprocessing. Data pre-processing is performed to give better outcome.

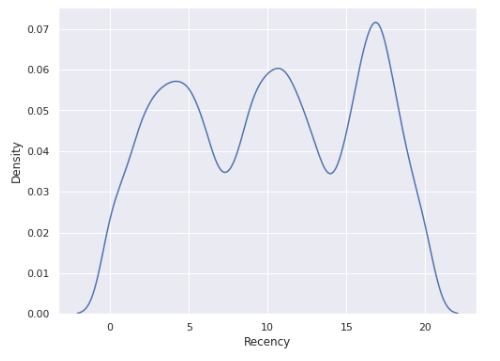
The method “RFM (Recency, Frequency, Monetary)” is used as the dataset has both integer and string values. The RFM analysis is assumed to target the specific clusters of customers with the communication which are relevant to their behavior and they can generate high response rate with increased loyalty. For better segmenting customers, RFM method performs wells and this is used for the clothing retailer.

# Customer segments identification

## RFM method

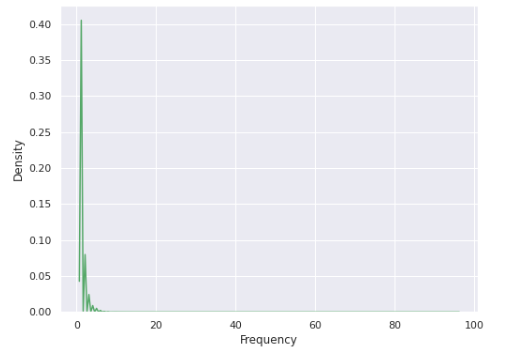
RFM method is significant tool that is used to create targeted marketing campaign which are tailored for targeted customer types. The customer segments that were identified is DateTime\_Of\_Purchase, Product\_Name\_List and Purchase\_Amount. Here we apply the RFM method to segment the customers such as Recency, Frequency and Monetary. To determine which customers are the best for the company, this RFM marketing technique is employed.

*Recency* 🡪 Recency provides details regarding the last purchase of customers or the products that are more likely to be purchased. The recency is measures in days but depending on product it can be measured in years, weeks or even hours.



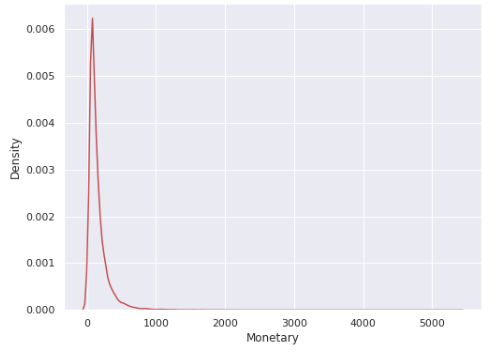
Recency showed the date time of purchase where the count increases to 16 and 17 and decreases suddenly to zero.

*Frequency* 🡪 Frequency shows how often the customers make a purchase in the specific time or period. That is the customers who purchased one time are likely to be purchased next. Hence, first time customers would be a good target for follow up and so they can be converted into regular and more frequent customers.



Product Name List shows the product that are frequently bought. Modern cotton bralette and the modern cotton thong showed increase count (Chandana, 2021).

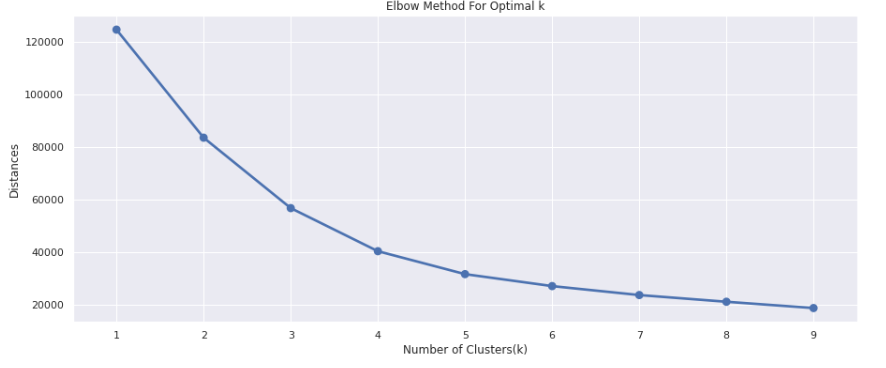
*Monetary* 🡪 Monetary depicts how much the customers spend in buying the product. It helps to predict the customers who are likely to spend money in future and they will have high value to a business.



Product amount increases between 0 to 1000. Monetary was 135 and increases up to 0.006.

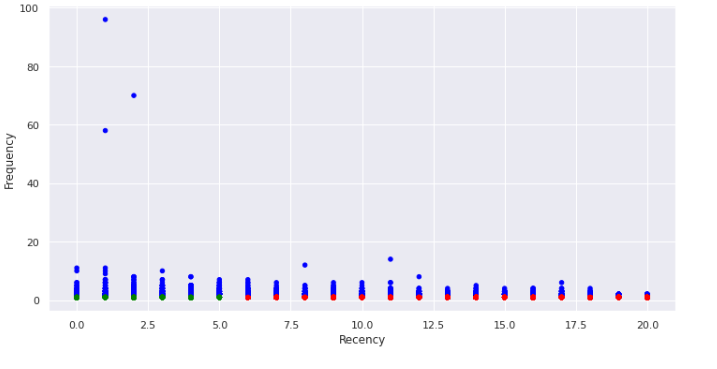
## K means

To find the optimal clustering K, K-means method is used. The objects are divided into clusters which share similarities (Shaik, 2019).



The Elbow method helps in finding the optimal ‘K’ in K means clustering. The k-means algorithm chooses the random points as the algorithm’s centroids and they would iterate them so as to adjust them. The distance decreases when the curve falls below the clusters and the range is set to 1 to 10. The value of k (here at 3), there is improvement in distortion which gets declined and is called elbow. After this it would stop dividing the data into further clusters.

After the dataset has been scaled, the dataset is fitted into K-Means model (Danurisa, 2022).



Frequency vs recency graph is generated to show the cluster region. The cluster is scattered in few regions when the recency is between 0.0 to 2.5. The red colour shows the date time of purchase which is zero. The green colour shows the data time of purchase is 2. The date time of purchase of the product lies between 0.0 to 5.0. After 5.0 the cluster shows red colour which is zero indicating the best customer. That is, they are frequent and recent shoppers.

# Limitation in Analysis

There are some limitations in the analysis that should be noted. The data used for the analysis is limited to one year and may not represent the entire customer base. Additionally, the RFM method is a simple method and may not provide the full picture of customer behavior. Further analysis should be conducted to understand the customer segments in more detail.

by using RFM method valuable insights regarding the customers were obtained but it does not consider other factors in the datasets such as DOB, state, etc. The customer demographics like age, sex, gender and ethnicity are not covered by the RFM analysis. The significant limitation by using RFM analysis is that, it has used only the historic data and they don’t have the capability to predict the future customer activity (Ezenkwu, 2015).

There are some limitations in the analysis that should be noted. The data used for the analysis is limited to one year and may not represent the entire customer base. Additionally, the RFM method is a simple method and may not provide the full picture of customer behavior. Further analysis should be conducted to understand the customer segments in more detail.

# Conducting marketing campaign

The marketing campaign is conducted by using the market research data obtained from the demographic survey. In our case study, we perform market research for the historical data of customer transaction made with the clothing retailer company. For the campaign we can use the demographic data obtained that is DateTime\_Of\_Purchase, Product\_Name\_List and Purchase\_Amount. As we have performed the cluster analysis for the market research, we can interpret the result and promote the clothing brand. In the cluster, we have created many clusters of similarities. Cluster analysis has helped to obtain patterns and that can be used in marketing. The successful marketing campaign for the clothing retailer can be conducted by:

* Identification of the product market – The target customers are women and the price of the product is set to low so that every woman gets an opportunity to buy the product and use it.
* Creating the buyer persona – preference of the customer is to buy modern cotton bralette and modern cotton thong.
* Establishing the unique positioning and the messaging – Using storytelling and brand messaging to describe the benefits of the product.
* Setting product marketing goals - Boosting revenue and increasing the brand’s recognition.
* Determine the offering price for the product – For the clothing brand, value-based pricing is used that is based on products price on the benefits.
* Discussion of final launch plane with internal teams – Planning the timeline for launching the product (Wagner, 2022).
* Launching the product to market – As the main goal is to allow the customer to buy the product in company website, the applicable channel for product launch is “website landing page”.

# Conclusion

In conclusion, the analysis of the sales data has provided valuable insights into the customer segments that are likely to purchase the product. The marketing campaign should be targeted towards the women who purchase the modern cotton bralette and modern cotton thong. The marketing campaign can use demographic data of the customers to perform cluster analysis. Although there are limitations in the analysis, the findings provide a starting point for the marketing campaign.

# References

Azevedo, N. (2023). Data Preprocessing: 6 Techniques to Clean Data. *https://www.scalablepath.com/data-science/data-preprocessing-phase*.

Chandana, D. (2021). Exploring Customers Segmentation With RFM Analysis and K-Means Clustering. *https://medium.com/web-mining-is688-spring-2021/exploring-customers-segmentation-with-rfm-analysis-and-k-means-clustering-118f9ffcd9f0*.

Danurisa, A. R. (2022). Customer Clustering Using the K-Means Clustering Algorithm. *Budapest International Research and Critics Institute-Journal*.

Ezenkwu, C. P. (2015). Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services. *International Journal of Advanced Research in Artificial Intelligence*.

Shaik, I. (2019). K-means Clustering Algorithm Based on E-Commerce Big Data. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*.

Wagner, E. (2022). 6 Steps to Develop a Winning Marketing Campaign. *https://www.i7marketing.com/blog/internetmarketing/6-steps-develop-winning-marketing-campaign*.

Assignment Video Link : <https://youtu.be/nherWZ9RE1o>