

Customer Segmentation of an Online Retail Store

UNSUPERVISED MACHINE LEARNING
USING KMEANS CLUSTERING MODEL TO
UNDERSTAND THE RECENCY,
FREQUENCY AND MONETARY VALUE OF
CUSTOMERS.

A REPORT BY:
ODIDI NAOMI TRACY, DATA SCIENTIST



Objectives

USING DATA TO MAKE STRATEGIC
DECISIONS

I obtained the customer dataset which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail and the objective of this report is to:-

- show how segmenting customers into groups that reflects their monetary power amongst other things can help the business make strategic business decisions.
- proffer business strategist and recommendations ranging from brand awareness to enhanced competition

A REPORT BY ODIDI NAOMI TRACY



Methodology

1. CLEAN DATA & PERFORM EDA

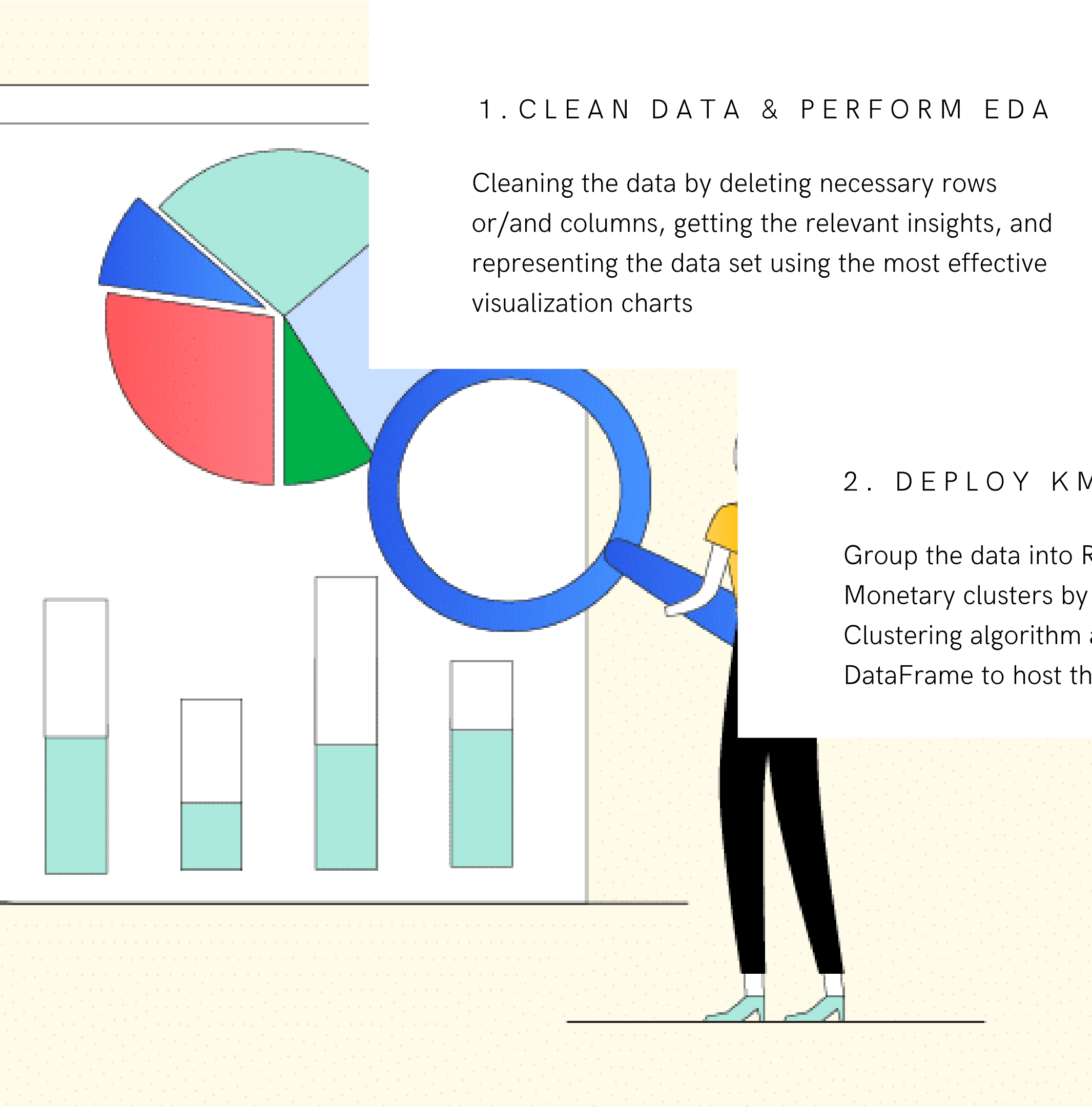
Cleaning the data by deleting necessary rows or/and columns, getting the relevant insights, and representing the data set using the most effective visualization charts

2. DEPLOY KMEANS MODEL

Group the data into Recency, Frequency and Monetary clusters by deploying the KMeans Clustering algorithm and creating a new DataFrame to host the new columns generated.

3. MAKE DEDUCTIONS.

Interpret the results of analysis and proffer business strategies and recommendations tailored to increased revenue and ROI.



Definition of Terms.

WHAT TO EXPECT



RECENCY

It shows how recently a customer has purchased from the store. The more recently a customer 's purchase, the more likely they will place a new order in the near future. If days pass without them placing a new order, they are more likely to churn.



FREQUENCY

Businesses like repeat purchases. Many businesses are sustained by returning customers. Companies built around subscription models come to mind. This is why many companies reward customer loyalty.



MONETARY VALUE

It shows how much a customer usually spends. Monetary value helps to recognize big spenders from bargain hunters. Not all customers have the same purchasing power. The aim is to win customers' trust to spend as much as possible without pushing them away.



EXPLORATORY DATA ANALYSIS (EDA)

It's the critical process of performing initial investigations on data so as to discover patterns, spot anomalies, test hypothesis and check assumptions with the help of summary statistics and graphical representations.

Data Cleaning & EDA

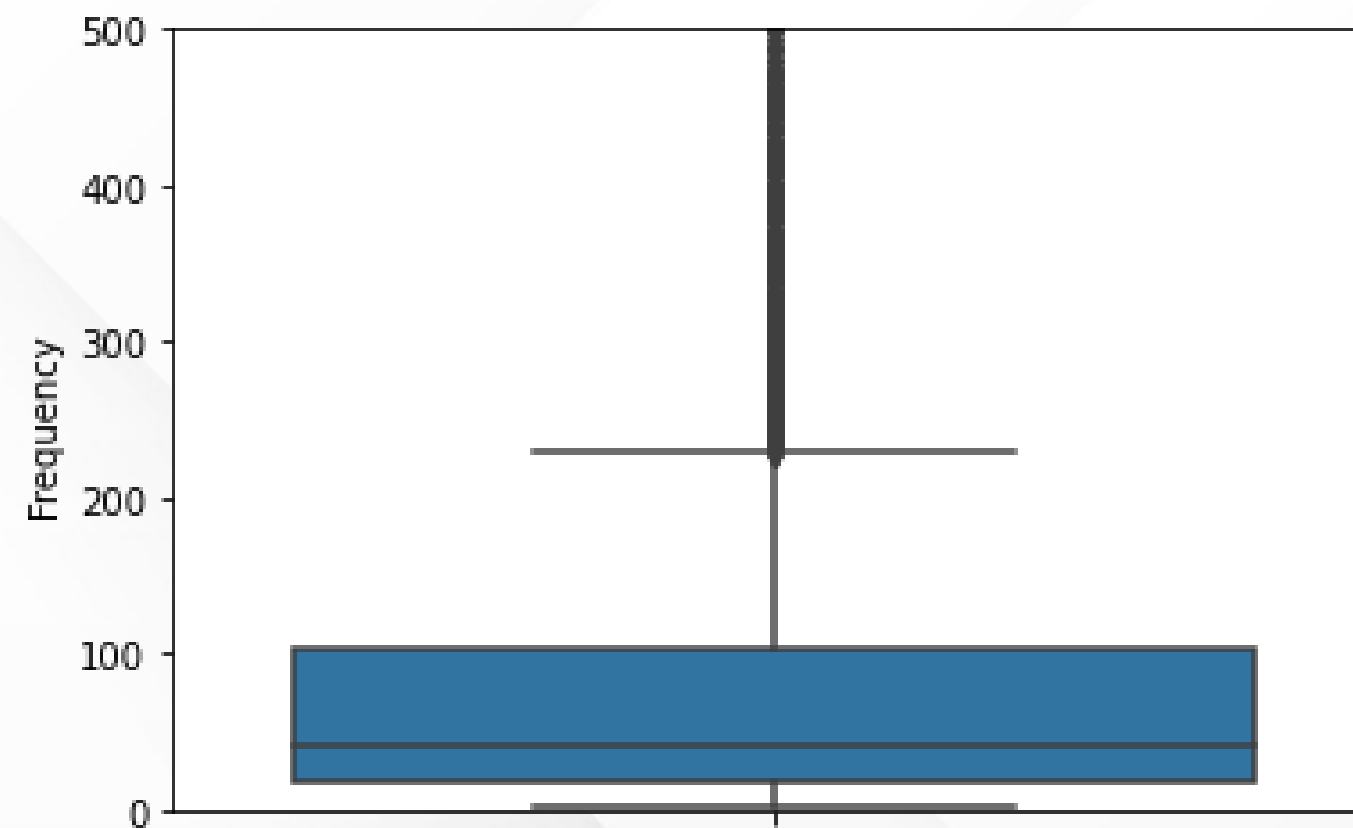
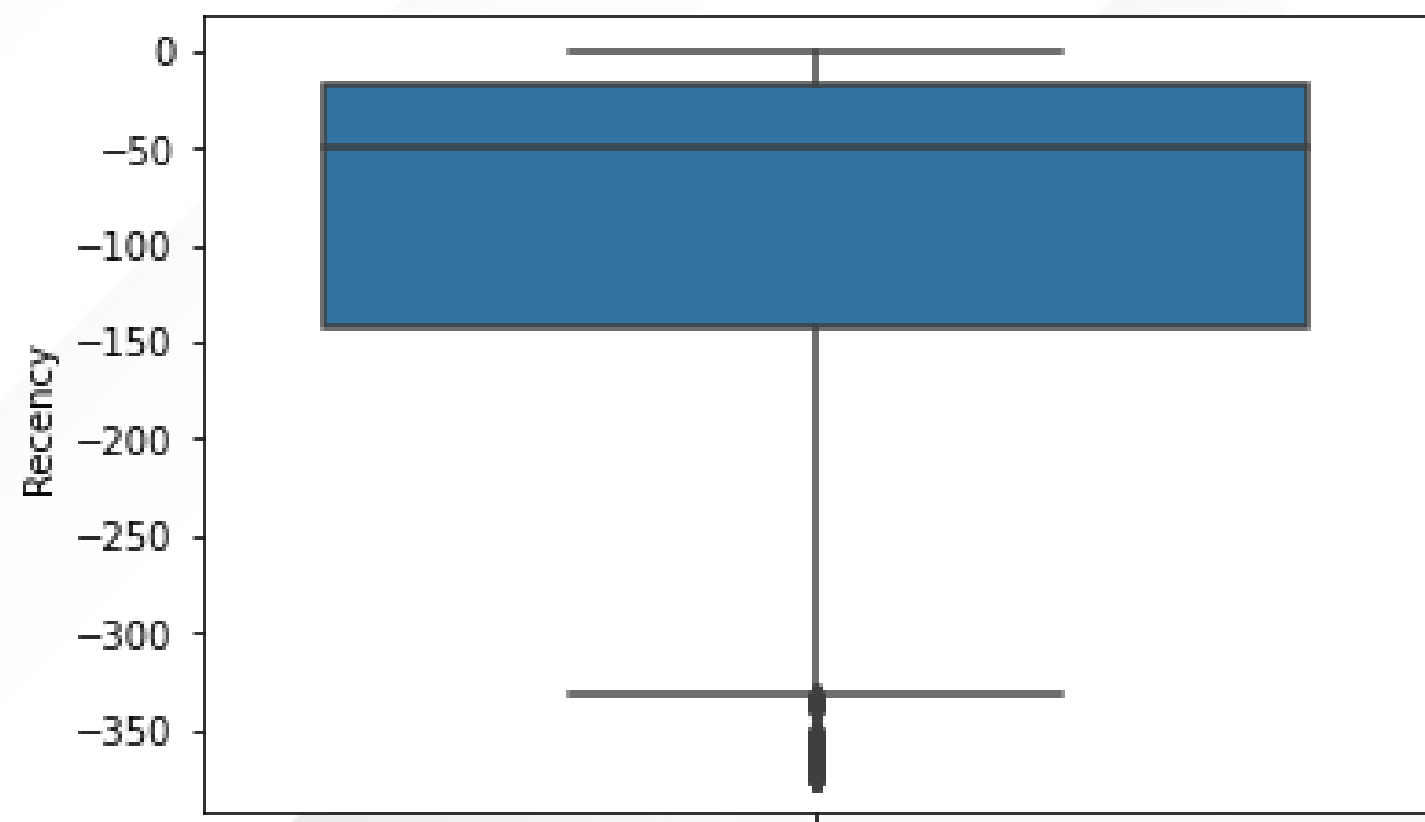
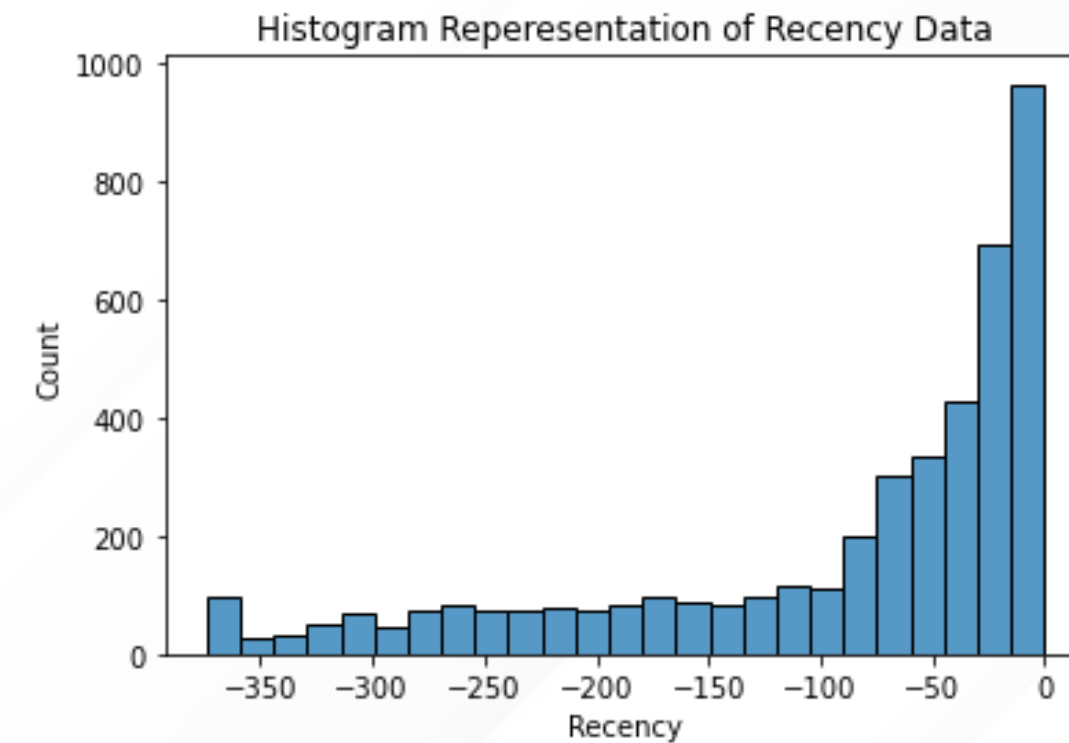
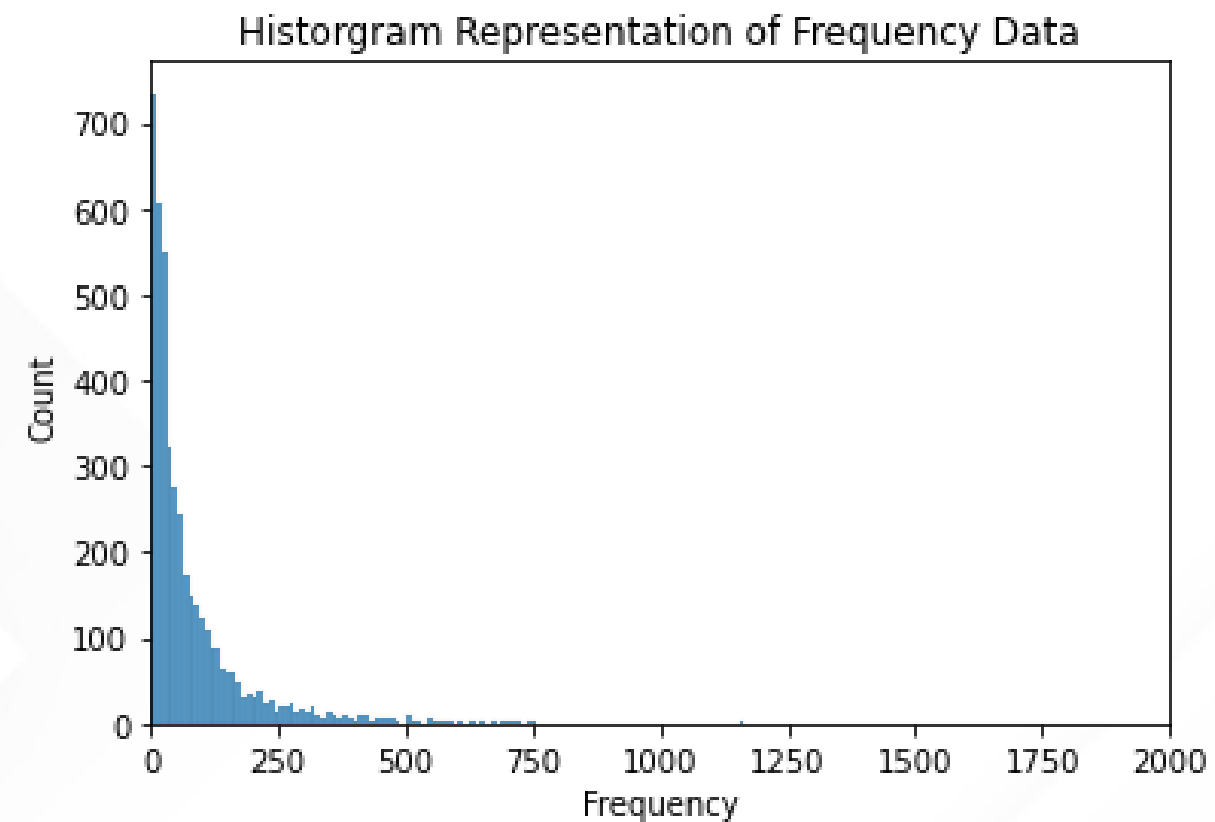
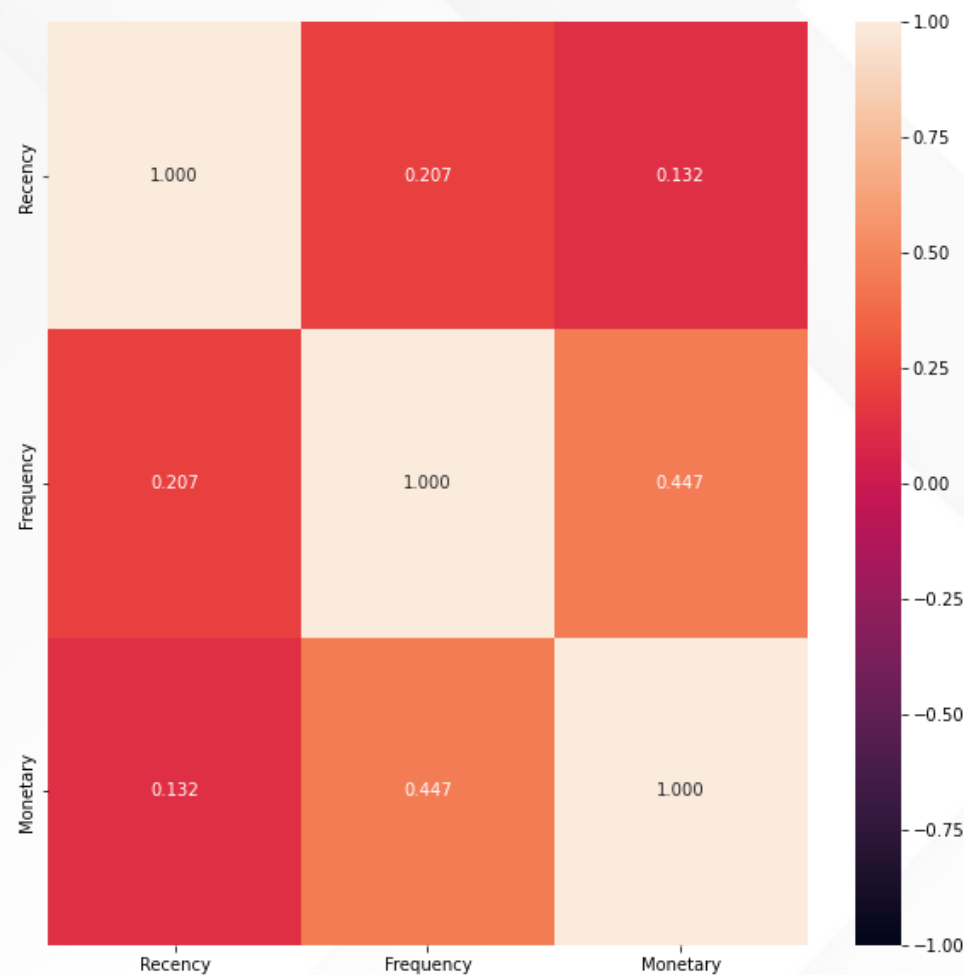
FOR THE TECHNICAL SUMMARY, SEE THE JUPYTER
NOTEBOOK([CLICK HERE](#))

Jupyter Notebook was used to carry out this project. I discovered the missing cells and made the necessary adjustments by eliminating them after loading the required libraries. I continued by making sure the columns were of the required types, however for this project, the **CustomerID** and **InvoiceDate** columns' data types were modified to achieve our ultimate aim.

I then determined each customer's **Recency**, **Frequency**, and **Monetary Values**. I only counted the customer's most recent purchase as their **Recency**, number of purchases as their **Frequency**, and **Monetary** value as the sum of all of their transactions over time. The new columns were added to a brand-new data frame called **RFM**.

I found patterns like the relationship between the metrics, the highest and lowest values for each metric as well as anomalies like why certain clients had negative monetary value. It turns out that customers were given a straightforward return policy, and those who sent their goods back resulted in a loss of money. There was a weak correlation between each variable which implied that the outcome of one metric didn't necessarily affect the outcome of the other unlike in Bayes' theorem. For further technical information, please check the jupyter file.

I went on to visualize the data using boxplots, bar plots, histograms, box plots, and heatmaps.



Deploying the KMeans Model

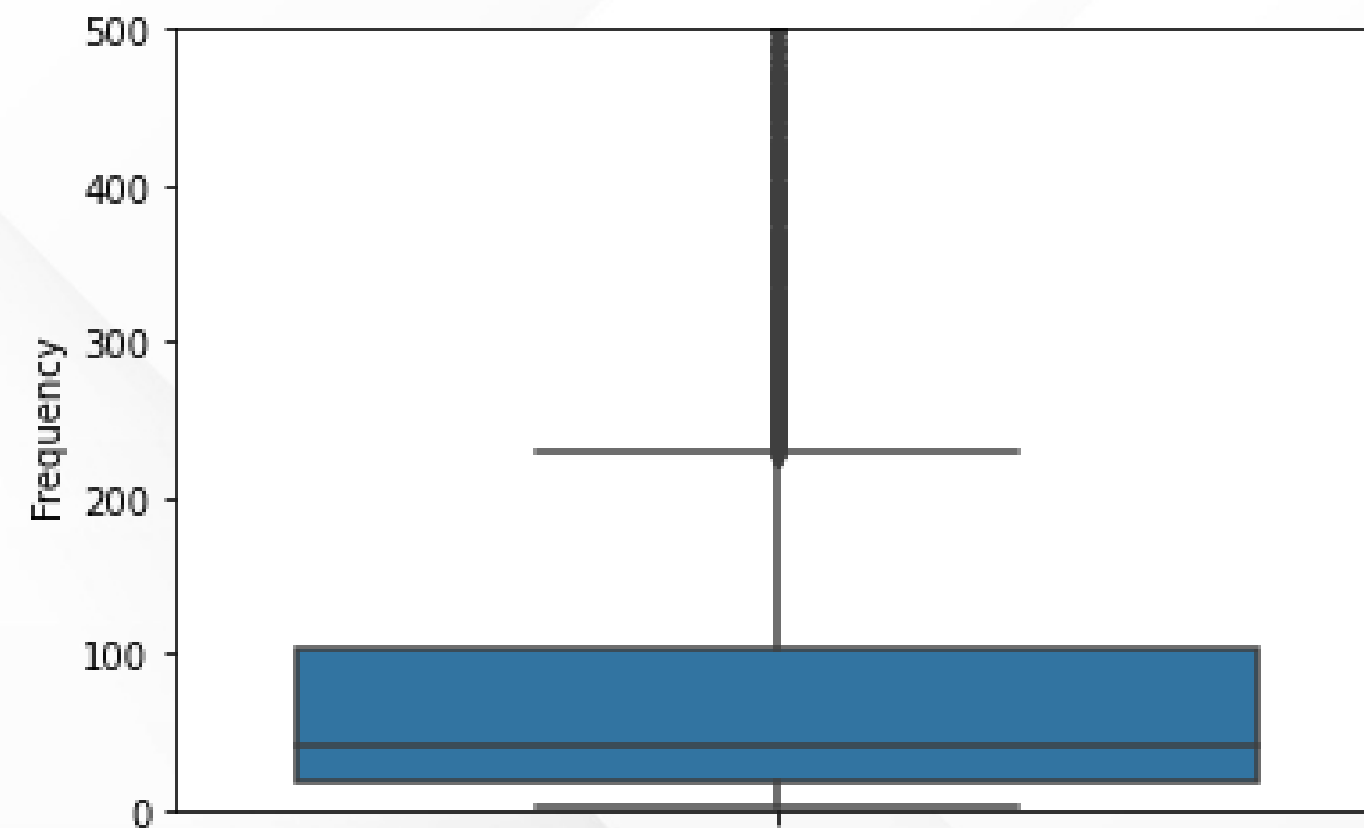
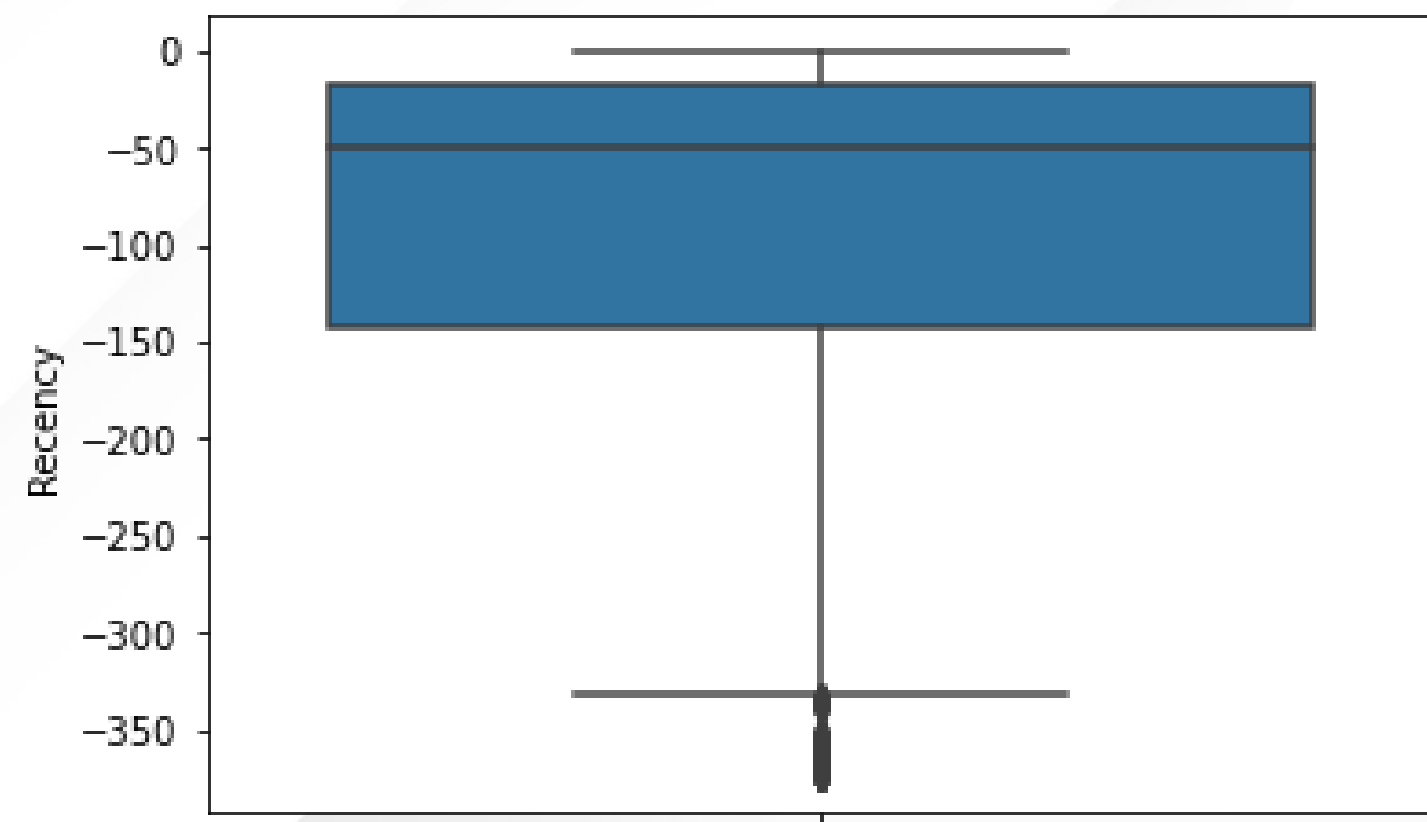
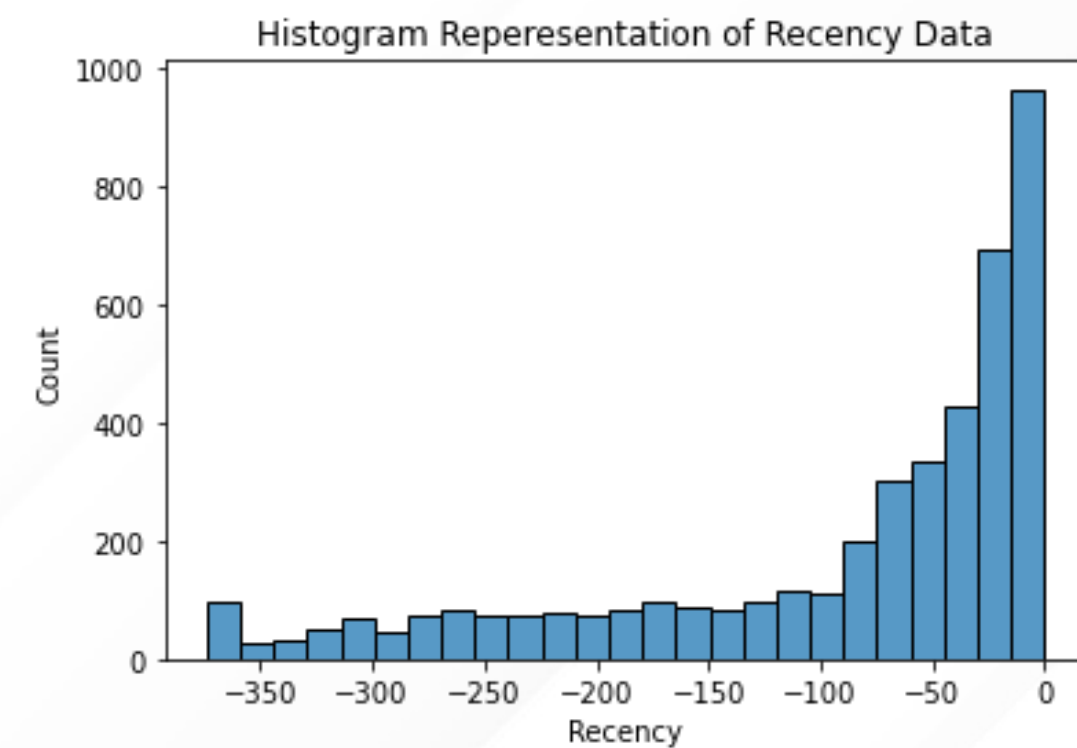
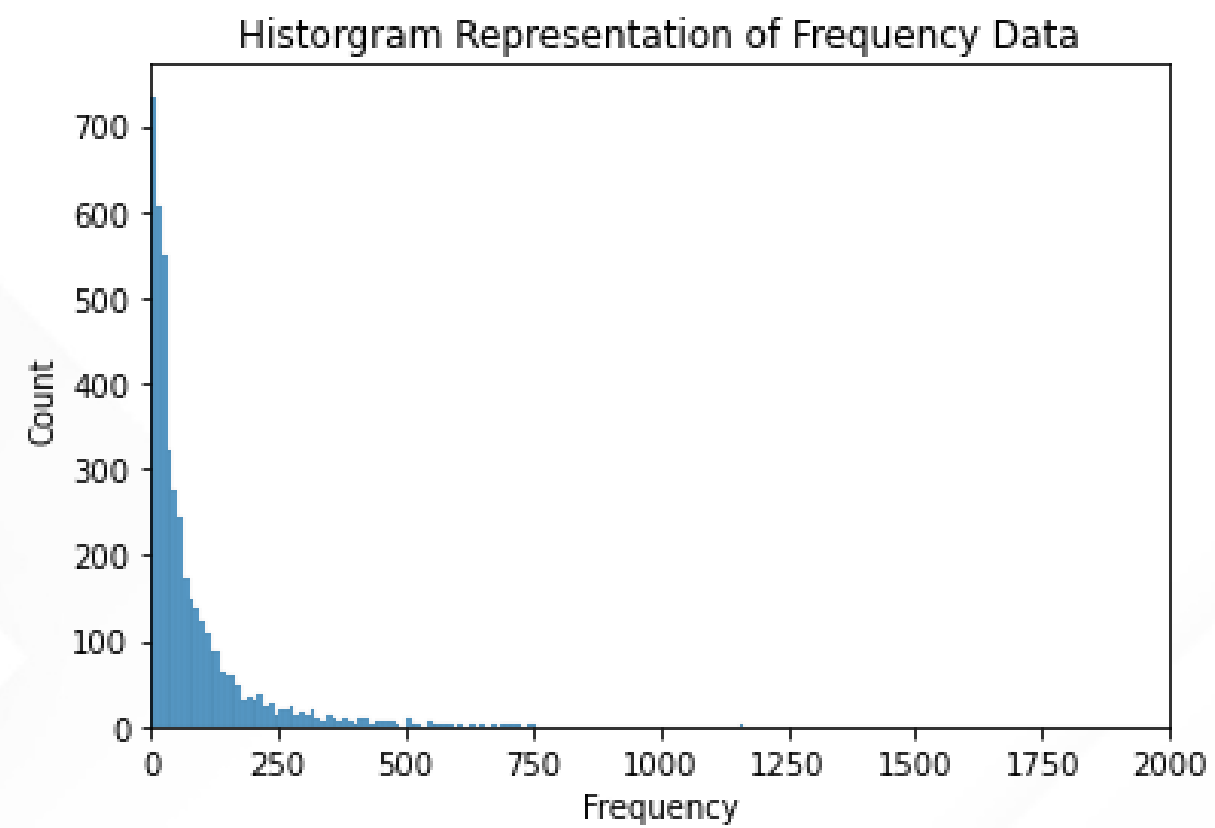
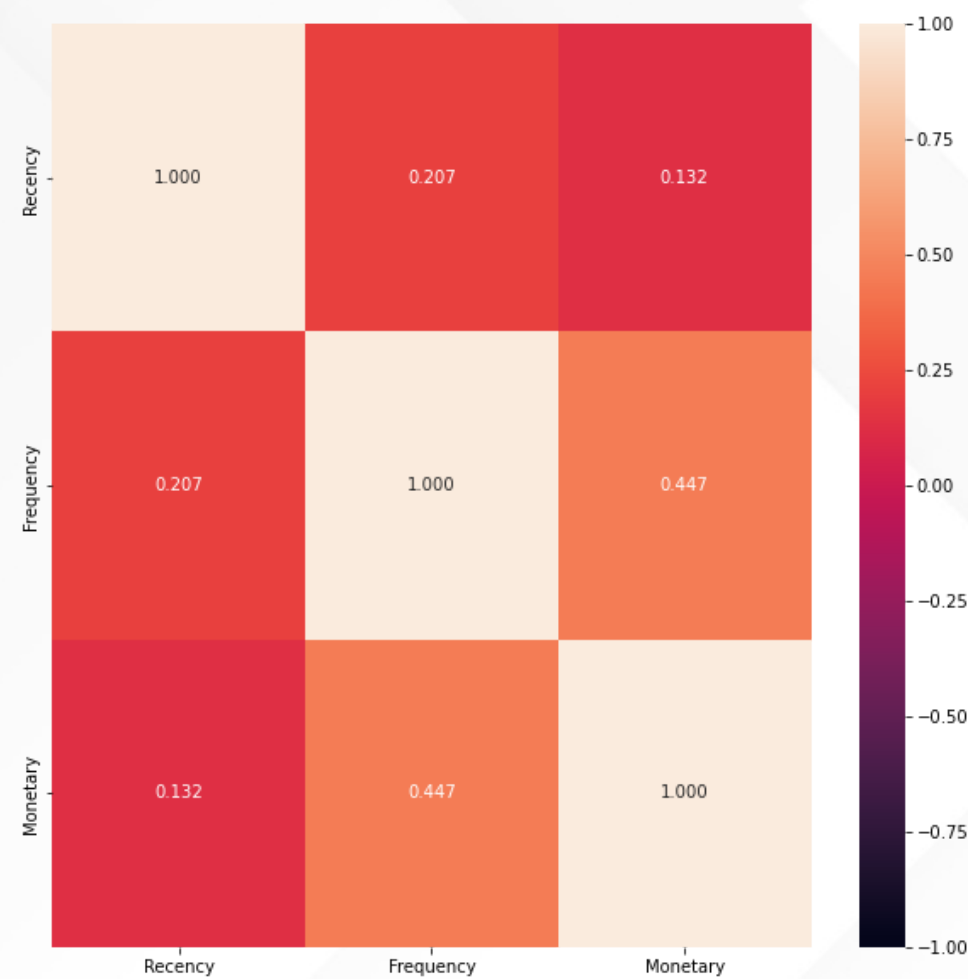
FOR THE TECHNICAL SUMMARY, SEE THE JUPYTER
NOTEBOOK(CLICK HERE)

Jupyter Notebook was used to carry out this project. I discovered the missing cells and made the necessary adjustments by eliminating them after loading the required libraries. I continued by making sure the columns were of the required types, however for this project, the **CustomerID** and **InvoiceDate** columns' data types were modified to achieve our ultimate aim.

I then determined each customer's **Recency**, **Frequency**, and **Monetary Values**. I only counted the customer's most recent purchase as their **Recency**, number of purchases as their **Frequency**, and **Monetary** value as the sum of all of their transactions over time. The new columns were added to a brand-new data frame called **RFM**.

I found patterns like the relationship between the metrics, the highest and lowest values for each metric as well as anomalies like why certain clients had negative monetary value. It turns out that customers were given a straightforward return policy, and those who sent their goods back resulted in a loss of money. There was a weak correlation between each variable which implied that the outcome of one metric didn't necessarily affect the outcome of the other unlike in Bayes' theorem. For further technical information, please check the jupyter file.

I went on to visualize the data using boxplots, bar plots, histograms, box plots, and heatmaps.



Conclusion



The forecasting methods were compared against each other using MAD, MSE, and MAPE. Of the three accuracy measurements, MAPE provides the most accurate and fair comparison of the forecasting methods. Using the Naive Approach is the best fit predicting method for our data set.

	NA	MA	ES	SLR	FS
MAPE	98%	95%	96.7%	57%	69.7%



About Me

I obtained the customer dataset which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail and the objective of this report is to:-

- show how segmenting customers into groups that reflects their monetary power amongst other things can help the business make strategic business decisions.
- proffer business strategist and recommendations ranging from brand awareness to enhanced competition

A REPORT BY ODIDI NAOMI TRACY

