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

Executive Summary

The more effectively we understand our customers, the more likely it is that we will raise brand/product awareness, sales, and ROI. This report is focused on segmenting our client base into categories that, among other things, reflect their purchasing power and pinpointing key commercial prospects.

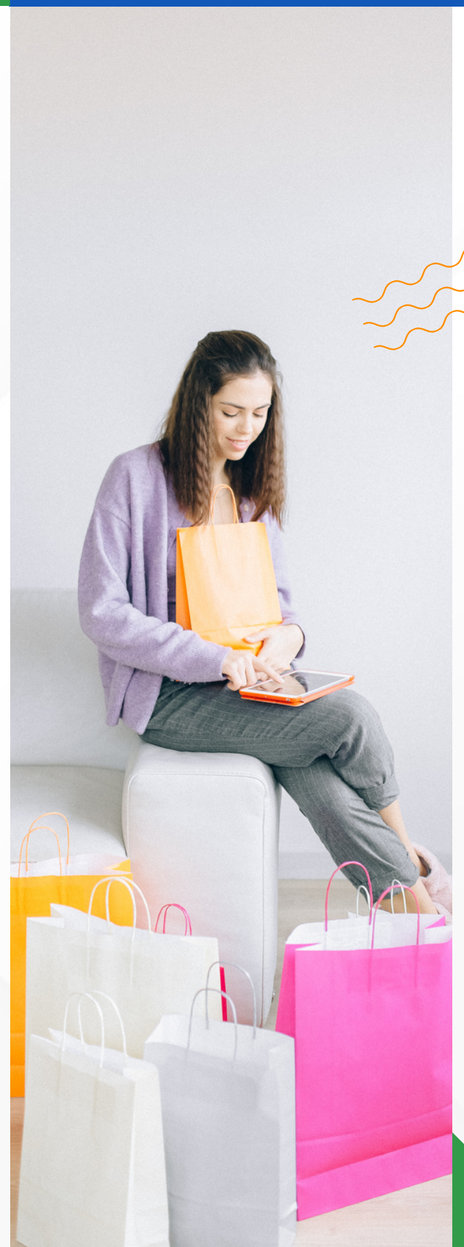
PROBLEM STATEMENT

By segmenting their consumers into **Low, Mid, and High value** segments, I will help them target their customers effectively and maintain customer acquisition and retention, i.e. avoid higher churn rate. This data set was provided by a UK-based and registered non-store online retailer.

INSIGHTS

Upon segmenting the dataset into 5 clusters of scores 0 to 4, leaving a customer with an overall score of 12, the following insights were discovered

- The overall score of a customer is the sum of the recency score, frequency score, and monetary score.
- 50% of their customers had an overall score ranging from 3 to 6, making them Mid-values customers while 0.49% had an overall score ranging from 0 to 2 making them Low-value customers. 0.01% were High-value customers ie the money bags, with overall scores ranging from 7 to 12.
- The customer satisfaction rate is excellent as only 2.2% of customers refunded their order.
- The High-value customers visited the website frequently as opposed to Low and Mid value customers. This makes sense as they are kept abreast of all the new deals and special offers hence they spend more. The same applies to the recency and the monetary value of the customers.



A REPORT BY ODIDI NAOMI TRACY

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

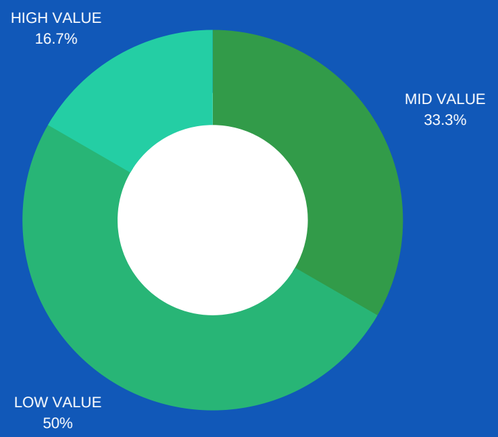
It shows how much a customer usually spends. Monetary value helps to recognize big spenders from bargain hunters. Not all customers have the same purchasing power. The aim is to win customers’ trust to spend as much as possible without pushing them away.

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.

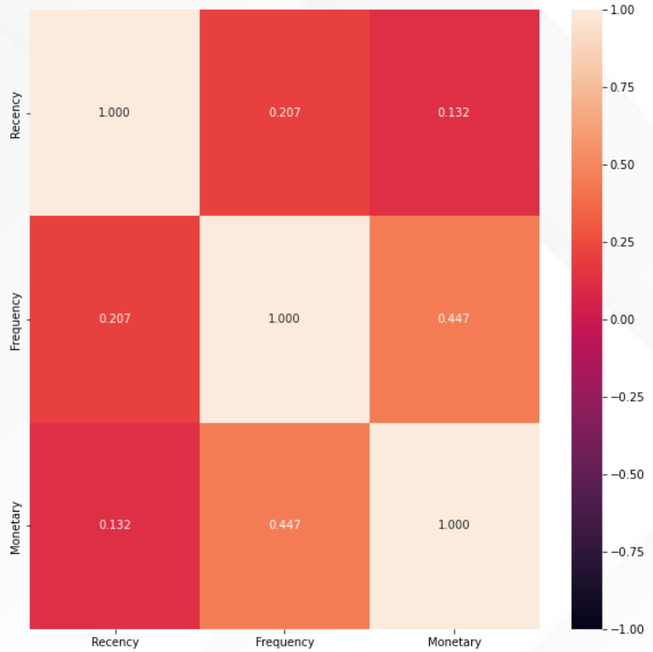
Data Set

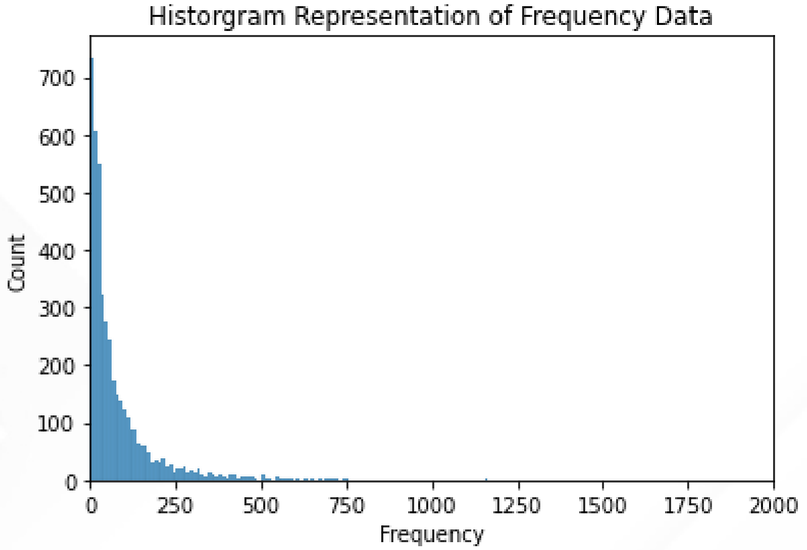
The transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 was used for this analysis. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers. The data set has 8 columns containing InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, & Country and 541909 rows.

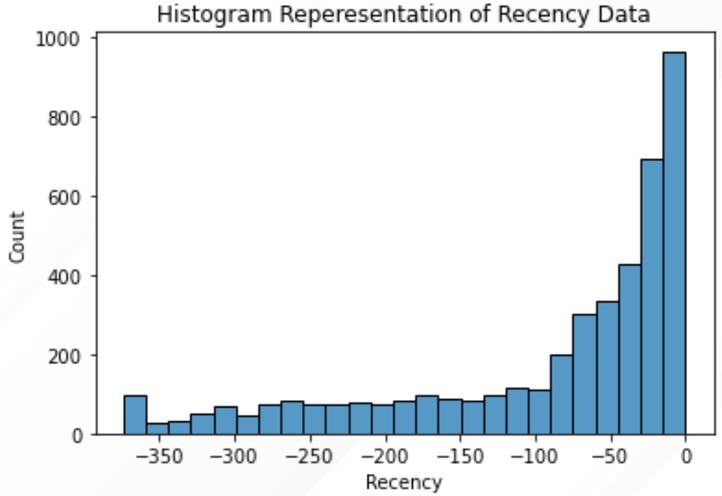


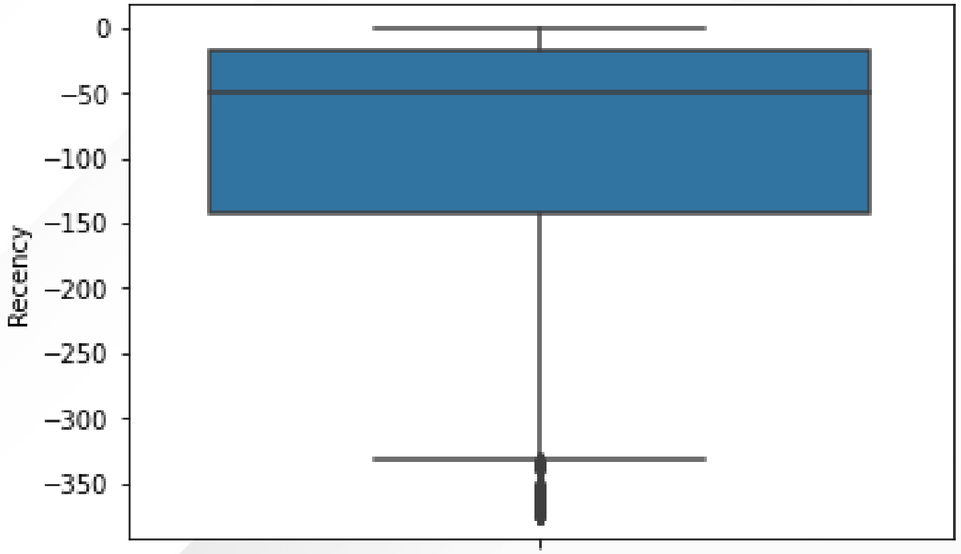
- 4372 unique customers made a total of 541909 transactions.
- 2741 unique customers made a total of 406829 transitions amounting to £8300065.8 in revenue.
- I eliminated 135080 rows that had missing values to leave 406829 rows.
- The statistics revealed that 42 transactions were made on average per customer, their average worth was £648.075, and their average recency was 50 days.
- Of the 406829 transactions made, 8945 of them were returned amounting to a loss of £611342.
- The metrics barely correlate with one another. This suggests that the results of one have little bearing on those of the other.
- There are now 9 columns after the creation of a new one called SaleValue from the product of Quantity and UnitPrice.

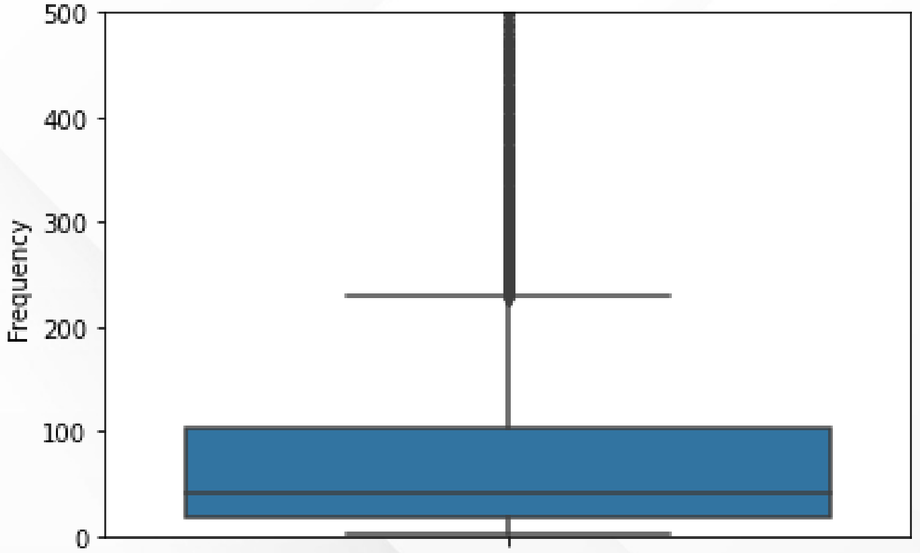
A REPORT BY ODIDI NAOMI TRACY

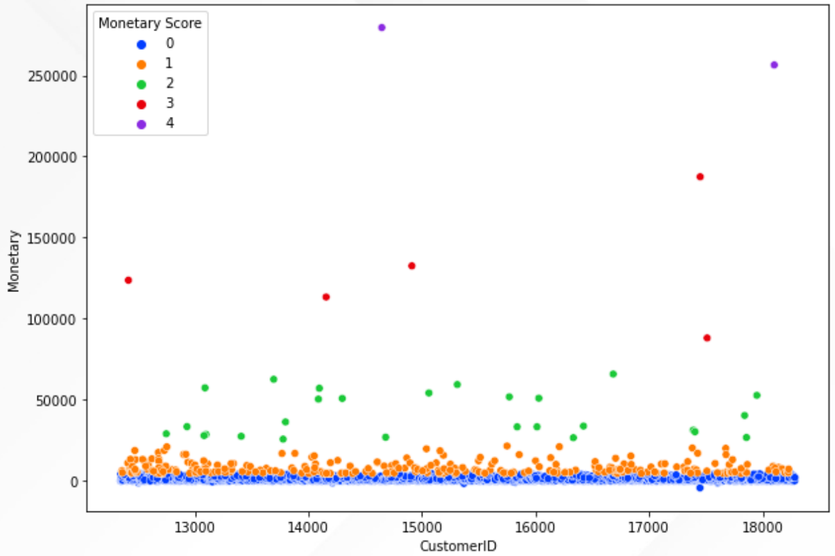

Recency
Frequency
Monetary
1.000
0.207
0.132
0.207
1.000
0.447
0.132
0.447
1.000
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

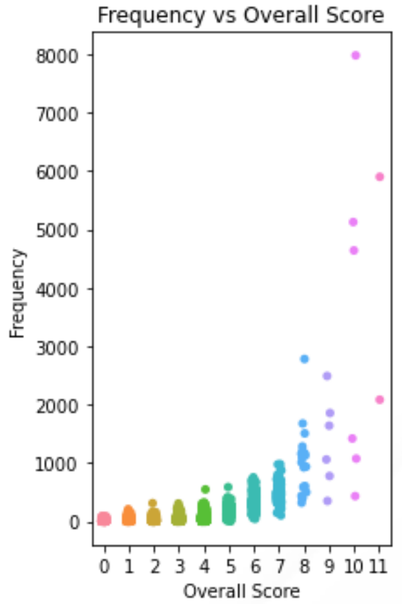

Historgram Representation of Frequency Data
Count
Frequency
0
100
200
300
400
500
600
700
0
250
500
750
1000
1250
1500
1750
2000

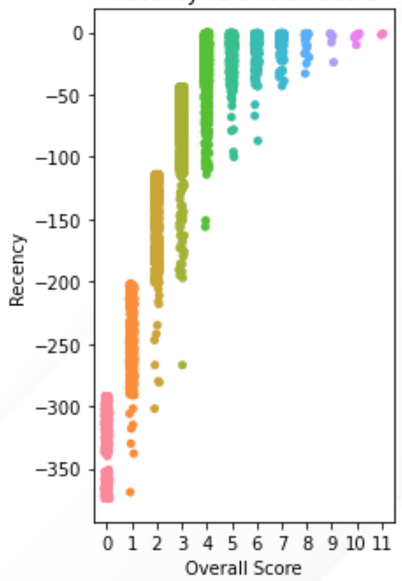

Histogram Reperesentation of Recency Data
Count
Recency
0
200
400
600
800
1000
−350
−300
−250
−200
−150
−100
−50
0

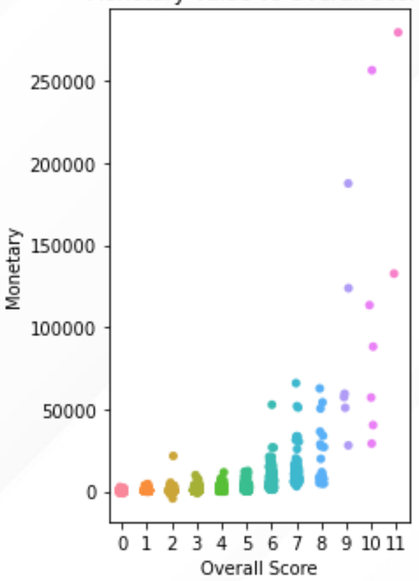

Recency
0
−50
−100
−150
−200
−250
−300
−350

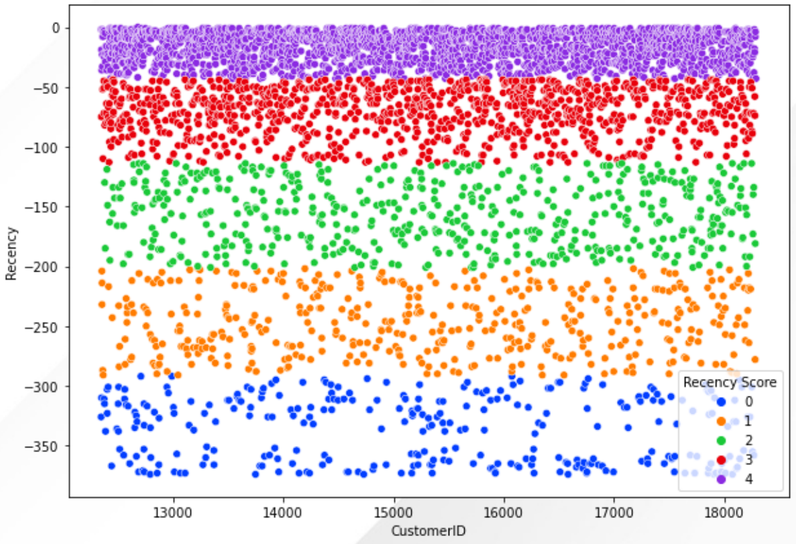

Frequency
0
100
200
300
400
500

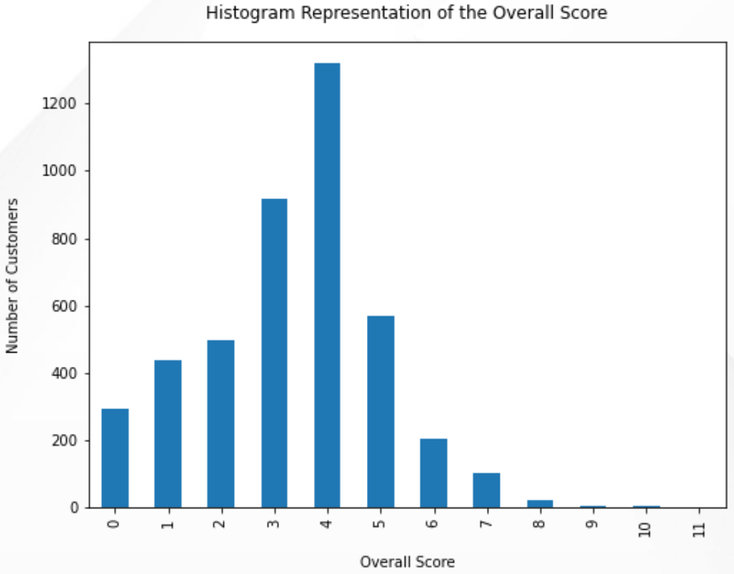

Monetary Score
0
1
2
3
4
Monetary
CustomerID

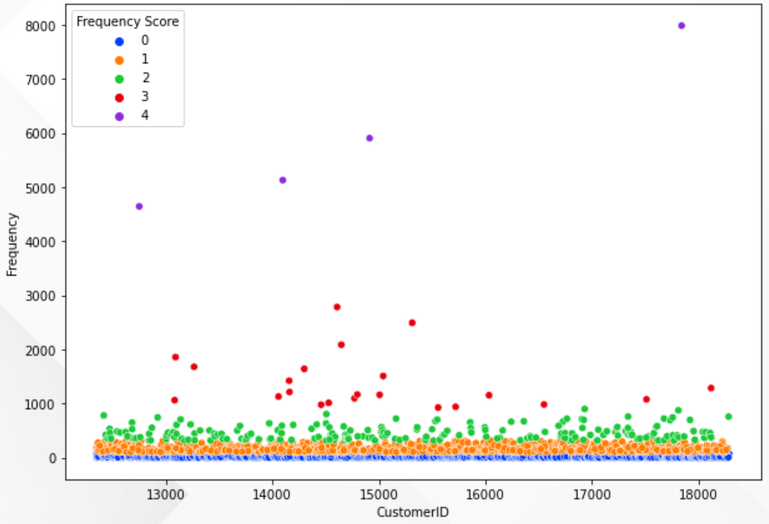

Frequency vs Overall Score
Frequency
Overall Score


Recency vs Overall Score
Recency
Overall Score


Monetary Value vs Overall Score
Monetary
Overall Score


Recency Score
0
1
2
3
4
Recency
CustomerID


Histogram Representation of the Overall Score
Number of Customers
Overall Score


Frequency Score
0
1
2
3
4
Frequency
CustomerID

Recommendations

Having thoroughly examined this online retail store, I would suggest the following:

- I advise the company to hire more staff even though the service is excellent and the return rate is only 2.2 percent in order to guarantee that consumers receive their orders accurately and on time.
- Due to the fact that they account for 99.9% of sales, it is necessary to develop customized offers for clients with Mid and Low scores.
- By boosting website visits and exposing them to targeted bargains and special offers assigned to them, a more successful marketing plan will influence Mid and Low consumers' total scores and ultimately enhance their monetary value.
- High-value clients should be given account managers. They can then be closely watched and given individualized care in this way. Making a personal connection with the money bags is a great concept for retention and acquisition because business is all about people.

A REPORT BY ODIDI NAOMI TRACY

About Me

Hi! My name is Odidi Naomi Tracy.

I am a mathematician with a flair for analysis. My areas of expertise are in data analysis and visualization, data modeling, machine learning, business development, and reporting. I excel at picking up new skills and tools quickly. I currently have experience with GitHub, Excel, PowerBI, and Python libraries like Pandas, Numpy, Matplotlib, Seaborn, and ScikitLearn. My portfolio includes projects like Sentiment Analysis, Customer Segmentation Using RMF Model, Online Fraud Detection, and Loan Prediction using Python and Machine Learning techniques like Clustering, Regression & Classification.

I have a love for using data to develop tech-enabled, strategic business solutions that encourage revenue growth. I am a recipient of the Women Techster Fellowship class of 2023, which is offered by Tech4Dev in partnership with Microsoft and the Central Bank of Egypt. I'm now working on my Actuarial Science master's degree.

Connect with me via LinkedIn or send me a mail. Opportunities in data science, finance, and insurtech are always welcome to me. You can browse my portfolio and essays on medium.