Project Report

Group 30 :

Kuljot Singh Chadha, Tanay Gawande, Jhanvi Kasundra, Tejas Wandile

Title : Customer Segmentation Using RFM Analysis.

Abstract:

In this project, we help certain retailer or a business who wants to increase the count of its multi time shoppers. By Conducting exploratory data analysis on the sales data set (retailer past sale) and build a predictive model to convert less frequent customers to more frequent customers. Using these Number and segregating the customers. The Retailers can build marketing and sales strategy and increase the sale over the time.

Introduction:

It is an old business adage: that 80% of your sales come from 20% of your customers. We are in business largely because of the support our customer base, "Your best customers". From the perspective of marketing or direct marketing, the best customers or the right customers are most likely to respond to the strength of your business.

Customer Segmentation is where customers are divided into groups based on their attributes - demographics, age group, purchasing power, etc. This helps the businesses target specific customers in a way that can benefit the business. We will be using predictive analytics and visualization techniques to understand their behavior. Our dataset for the same will be online retail dataset.
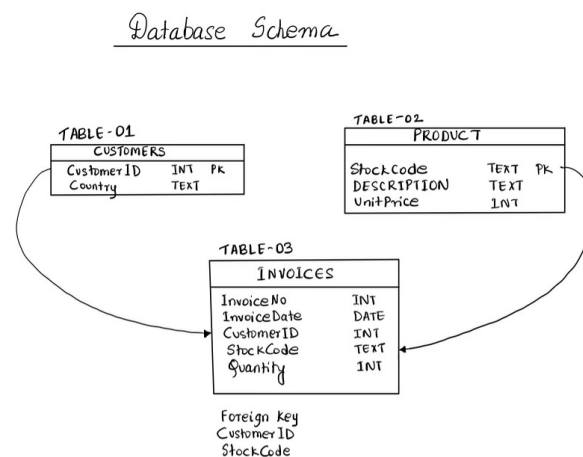
Model:

Our model is a predictive model that increases the count of multi time shopper by performing RFM segmentation, deriving variables like frequency, recency, latency. Then, create target variables followed by customer segmentation using the same.

- **FREQUENCY** : How often do they make purchase? (gives you the count of customers)
- **RECENCY** : How recently did the customer purchase the product.
- **MONETARY** : How much do they averagely spend per purchase?
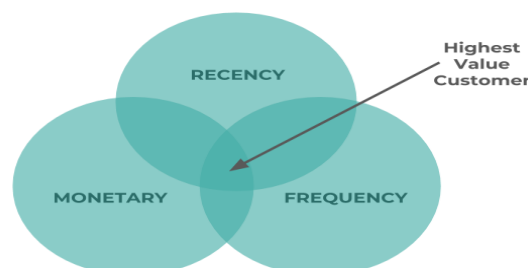
Data:

The data that we are using is Online Retail Dataset, that comprises of 8 columns and 541910 entries. The dataset includes the transaction details of all the customers across the world, it includes information such as invoice date, stock date, invoice number, customer_id, description of the different products, purchase_price(unitprice), quantity & the country of the customer.

Database Schema

TABLE-01			
CUSTOMERS			
CustomerID	INT	PK	
Country	TEXT		

TABLE-02		
PRODUCT		
StockCode	TEXT	PK
DESCRIPTION	TEXT	
UnitPrice	INT	

TABLE-03	
INVOICES	
InvoiceNo	INT
InvoiceDate	DATE
CustomerID	INT
StockCode	TEXT
Quantity	INT

Foreign key
CustomerID
StockCode

Analysis:

*Figure 1:* **Recency of Customer.**

(Calculated the recent purchase date of customer. Created the recency data frame using invoice date, recent purchase date and CustomerID. Recency is calculated by comparing the current date and recent purchase date.
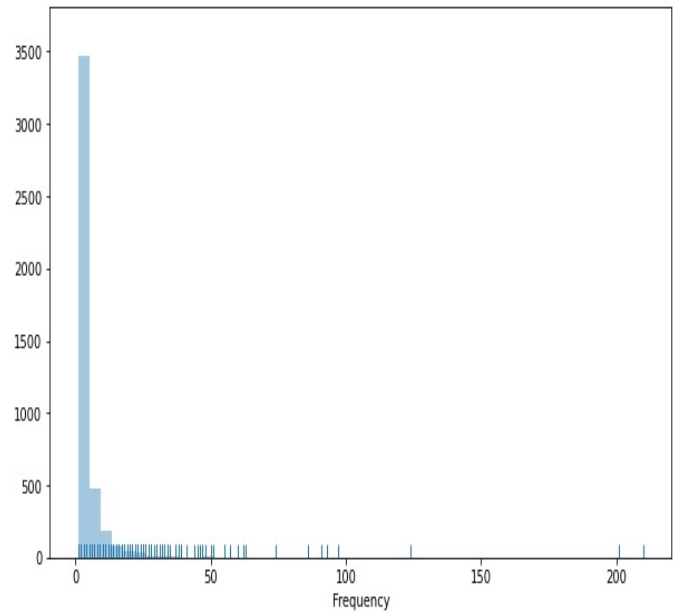
Figure 2: **Frequency of Customer**

(Removed the duplicate customer and invoice IDs. Created frequency using the unique customer and invoice ID.)

Figure 3: **Frequency and Monetary Values customers are segregated into different clusters.**
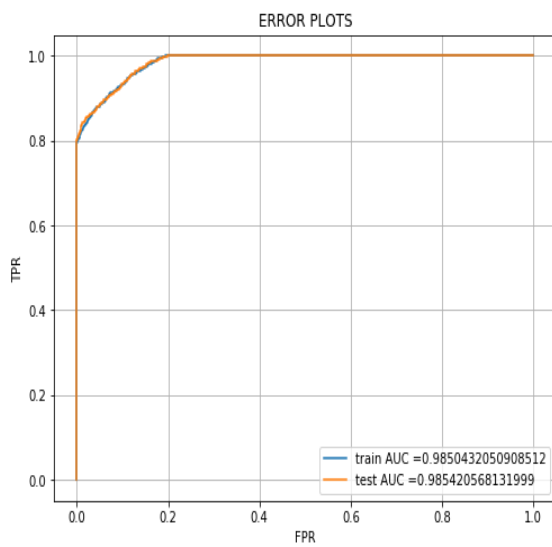
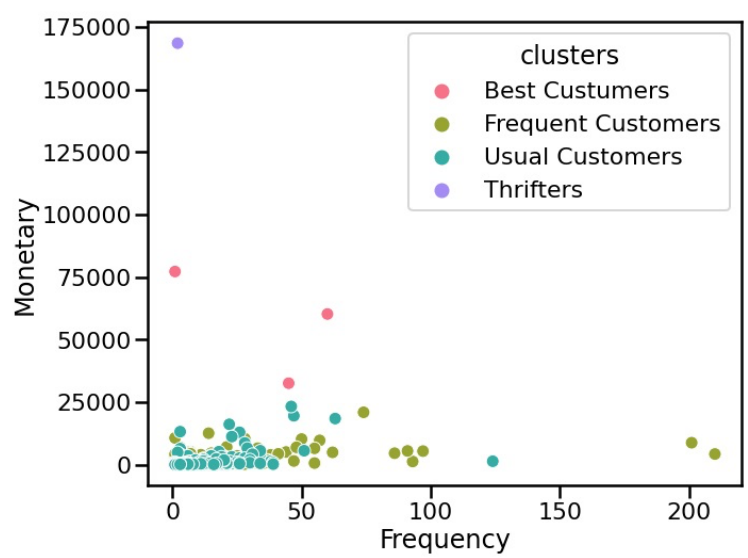*Figure 4:* **ML model Fitting Curve**
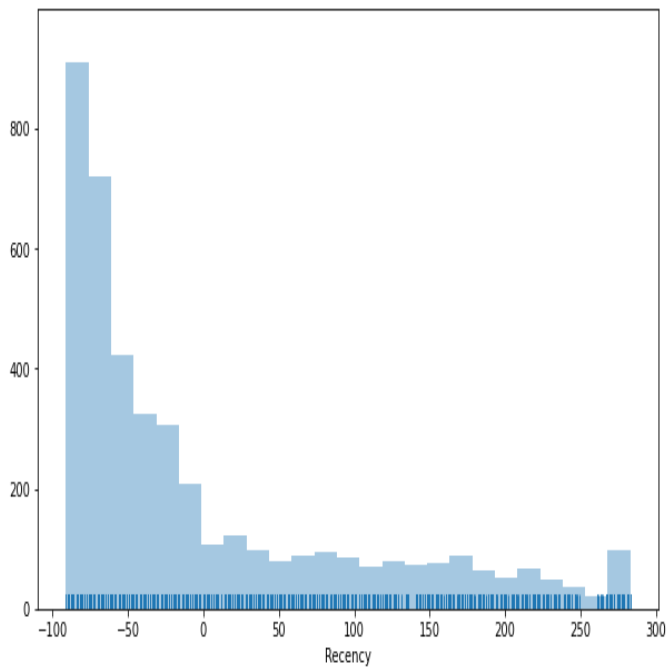
-FIGURE 1-

-FIGURE 2-

-FIGURE 3-

-FIGURE 4-

Results:

- Performed Logistic Regression and got the accuracy of 0.91.
- Customers are classified based on their recent and quantity of purchase.
- Most valuable customers are defined.

Conclusion:

- Customers who aren't visiting frequently can be given incentives also Loyalty plans can be implemented for the loyal customers.
- We can offer long term contracts to the customers to retain customers.

Future Research Directions:

- Customer Segmentation
- Considering More Variable to Draw Segmentation
- Trying Different Model
- Large Dataset (Sales Record)