

# **Capstone Project**

## **Customer Segmentation**

Dharmesh Patel

# Outline:

- **Objective**
- **Data Summary**
- **Data Analysis**
- **Data Cleaning**
- **Feature Engineering**
- **ML model evaluation**
- **Conclusion**

# Problem Statement

- **Given a dataset related to a online retailer based out of the UK, we need to analyse and identify major customer segments using K Means algorithm and also using different verification method to confirm the result.**

# What is Customer Segmentation?

- **Practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests and spending habits.**
- **Allows us to better understand our customers helping us target these customers in a more efficient manner and improve the customer experience.**

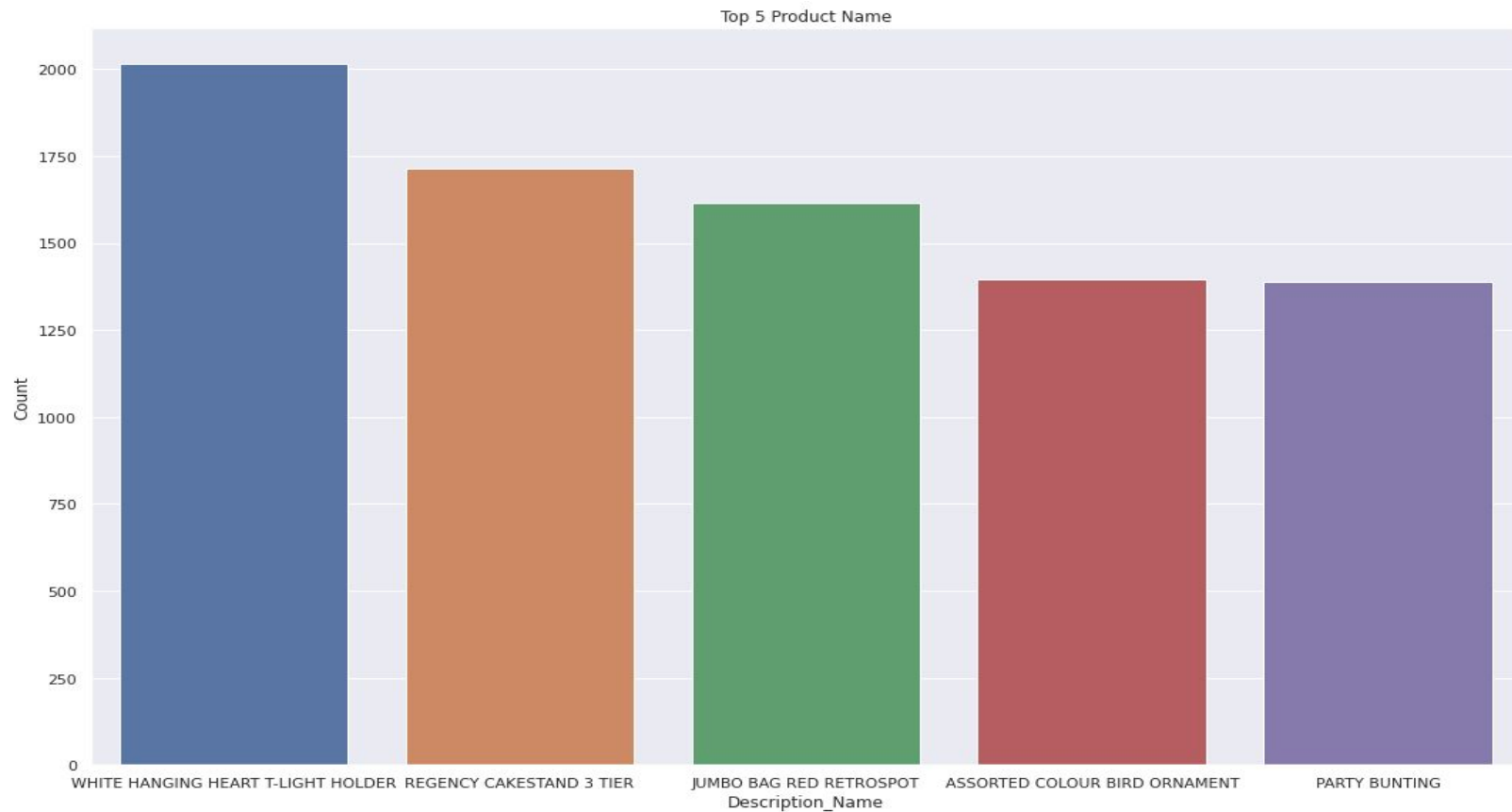
# Data Summary

## Data Description:

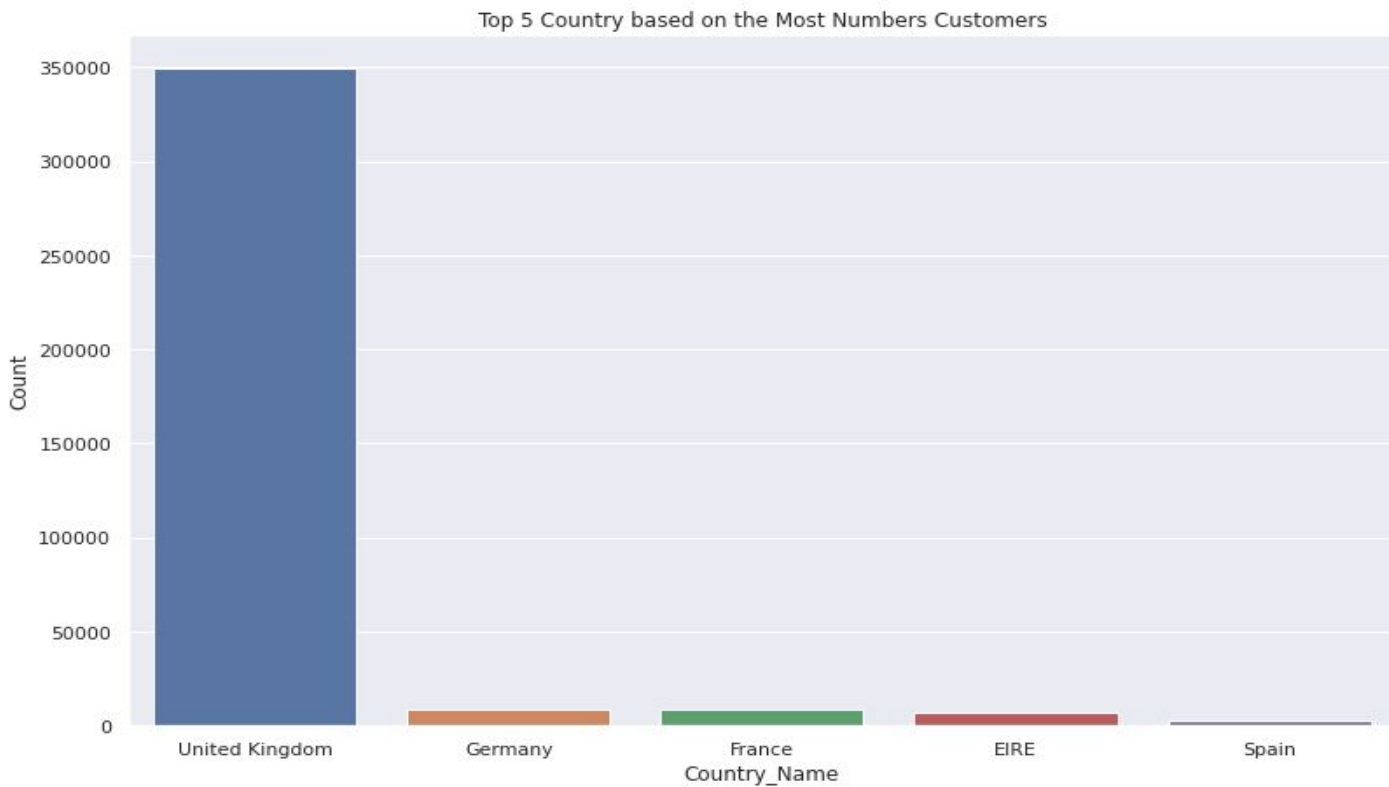
The dataset has 541909 rows and 8 features(columns).

- **InvoiceNo:** Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.
- **StockCode:** Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.
- **Description:** Product (item) name. Nominal.
- **Quantity:** The quantities of each product (item) per transaction. Numeric.
- **InvoiceDate:** Invoice Date and time. Numeric, the day and time when each transaction was generated.
- **UnitPrice:** Unit price. Numeric, Product price per unit in sterling.
- **CustomerID:** Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.
- **Country:** Country name. Nominal, the name of the country where each customer resides.

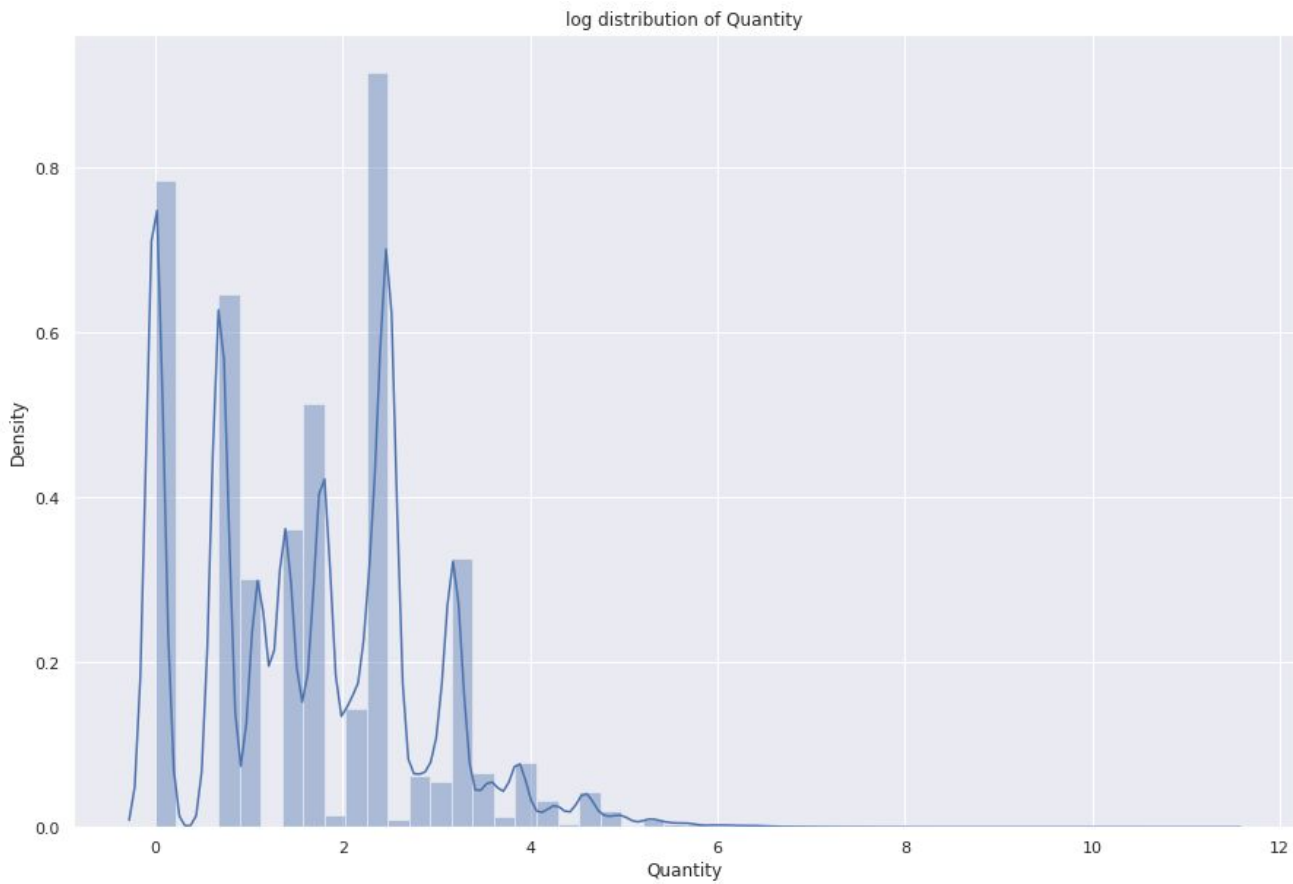
# EDA



## EDA

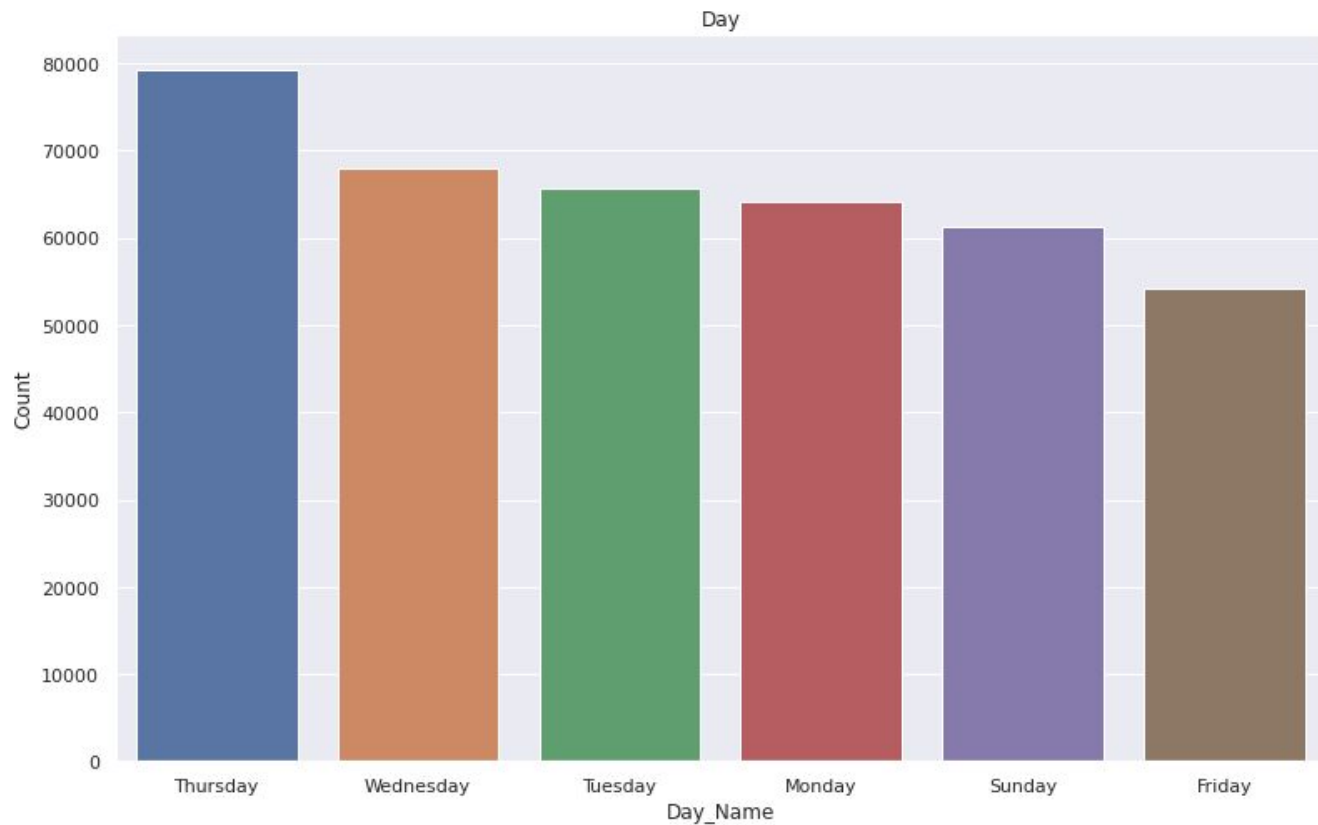


# EDA

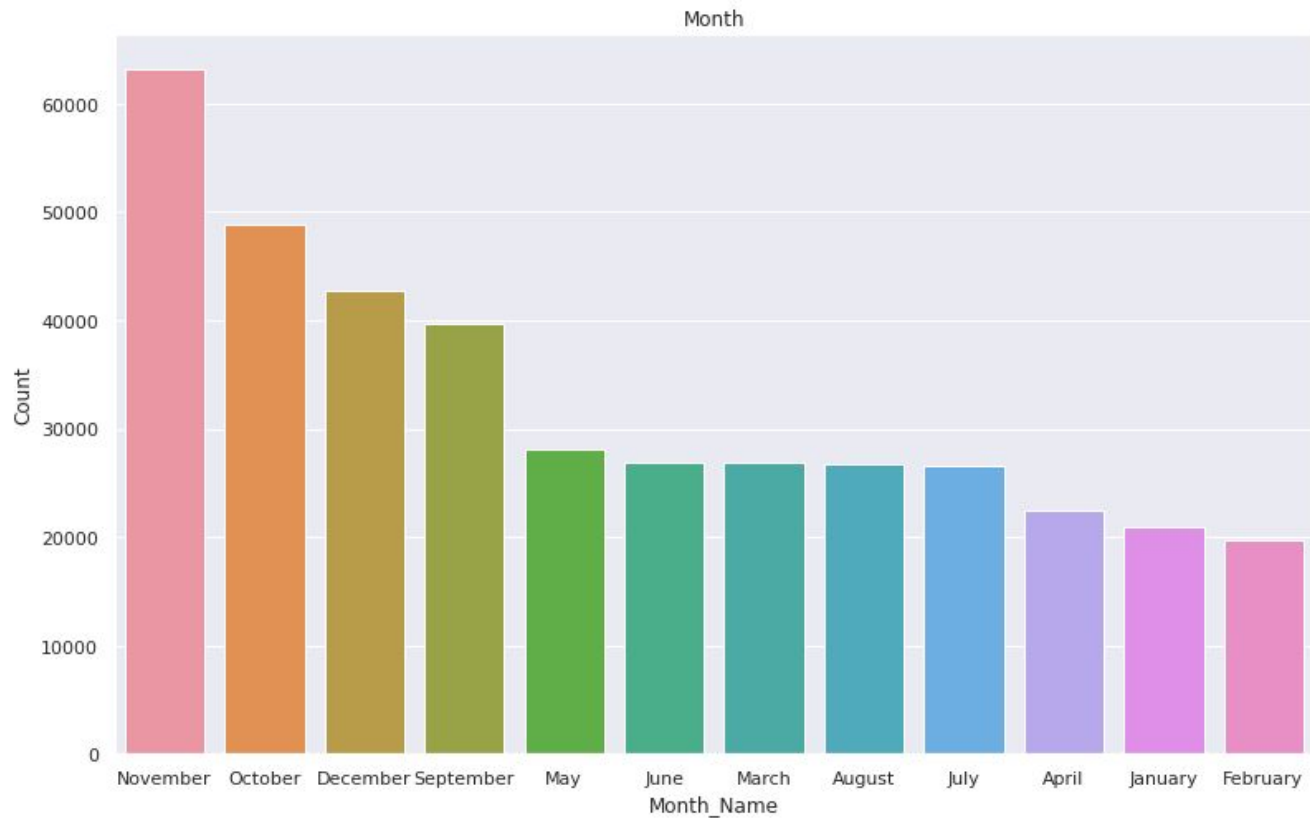




# Day vs Order Count



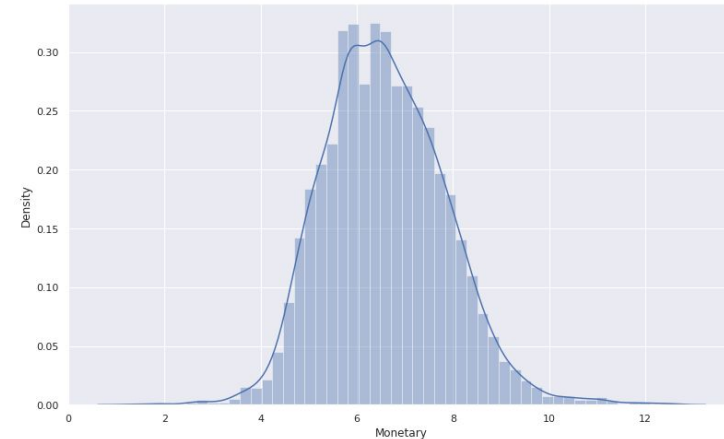
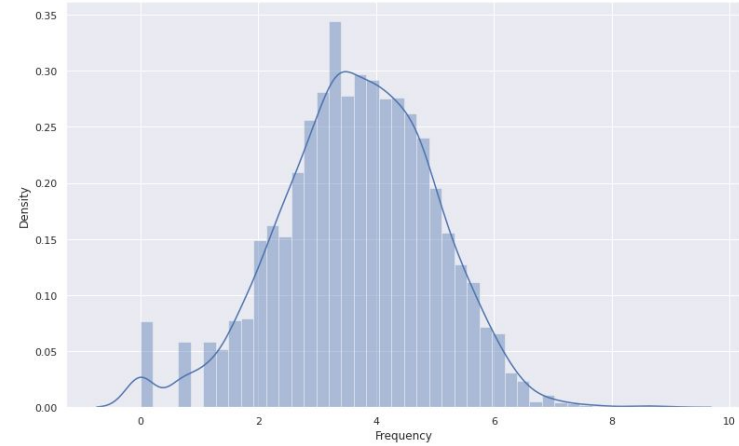
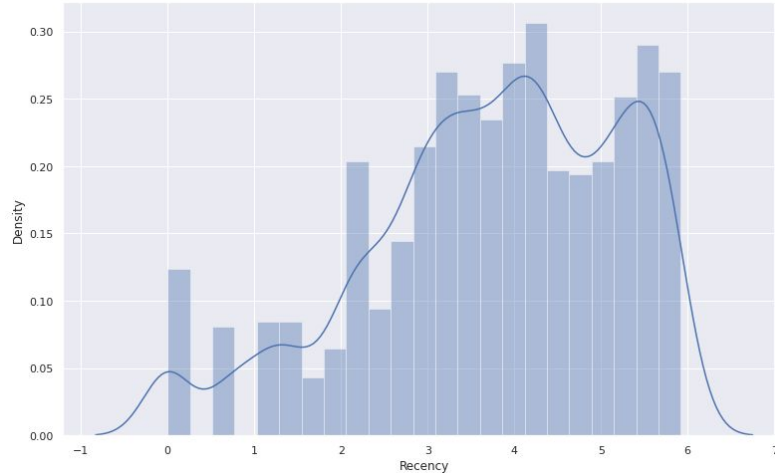
# Month vs Order



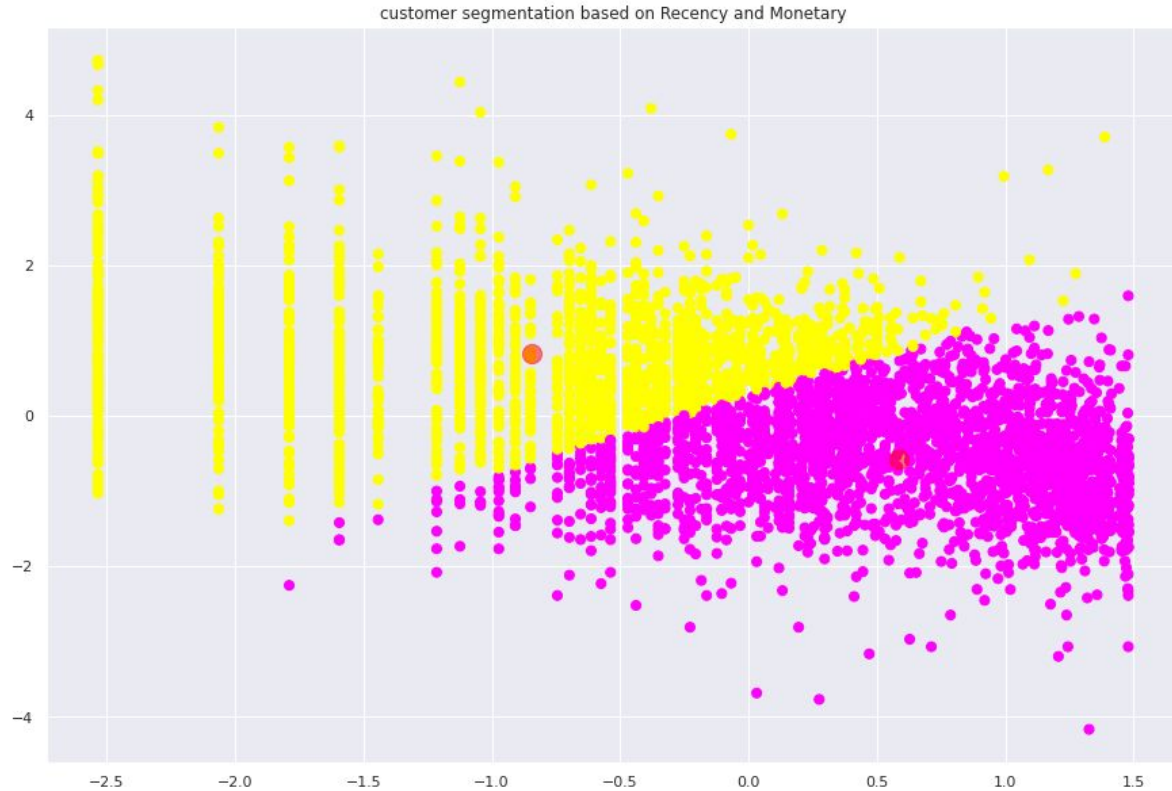
# RFM

**Recency, frequency, monetary value is a marketing analysis tool used to identify a company's or an organization's best customers by using certain measures.**

- 1. How recently a customer has made a purchase**
- 2. How often a customer makes a purchase.**
- 3. How much money a customer spends on**



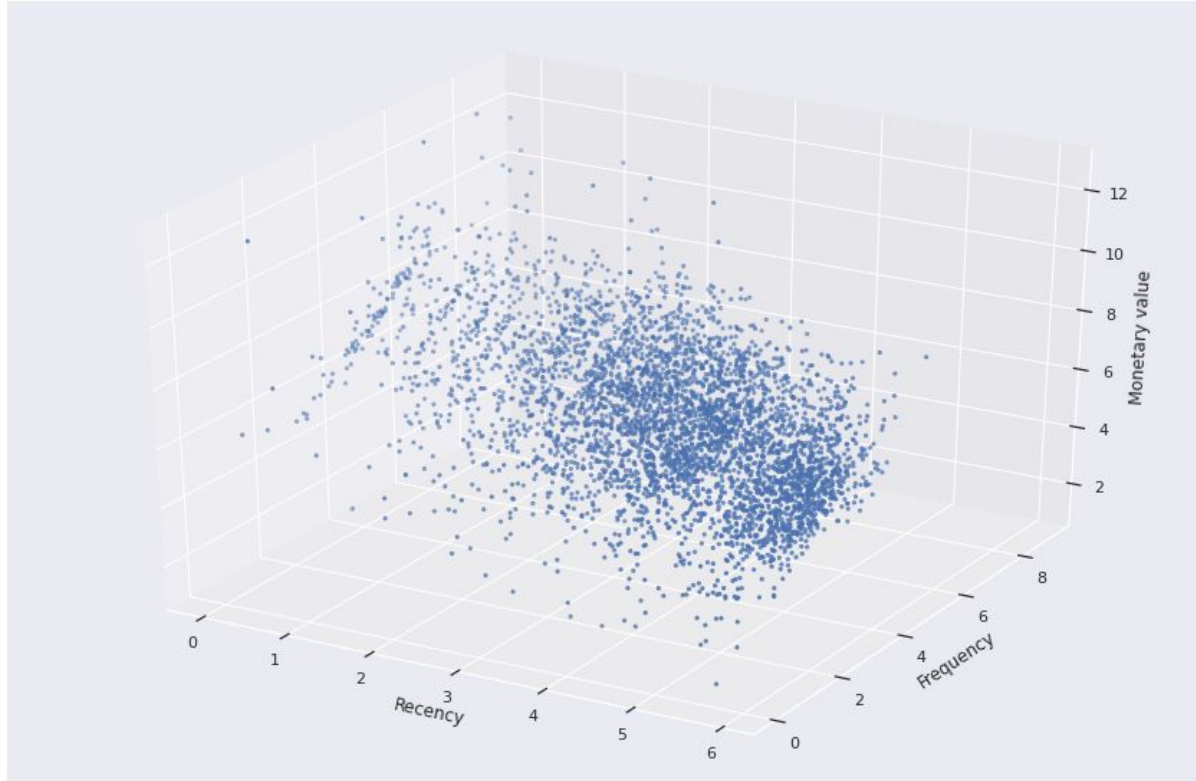
# Based on Recency and Monetary



# Based on Frequency and Monetary

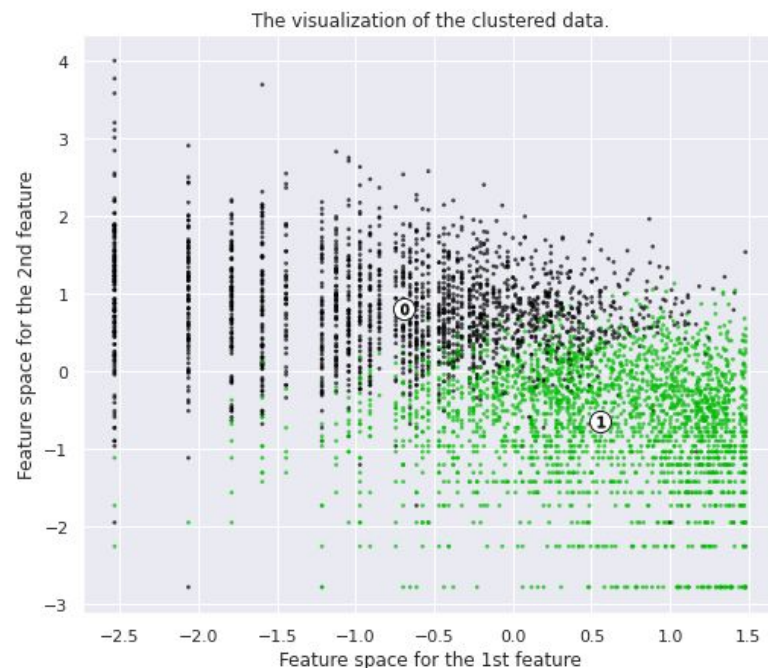
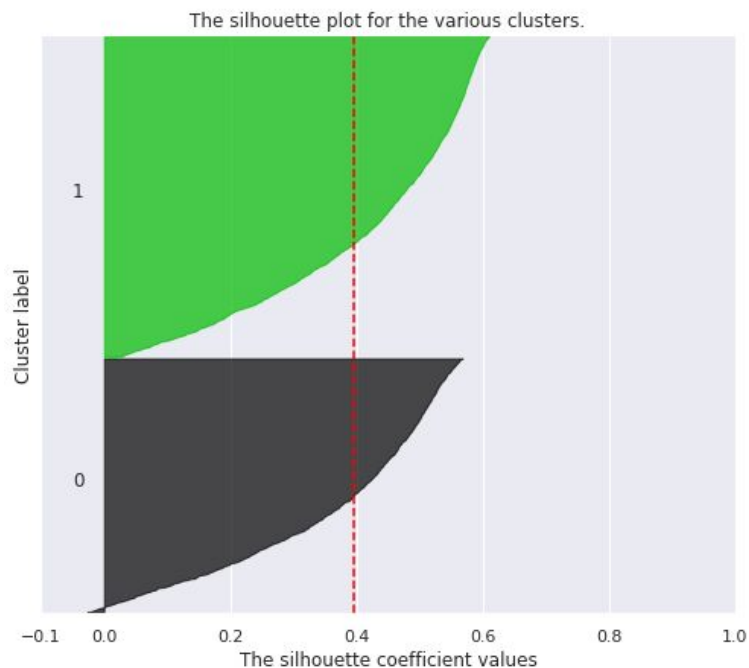


# 3D visualization



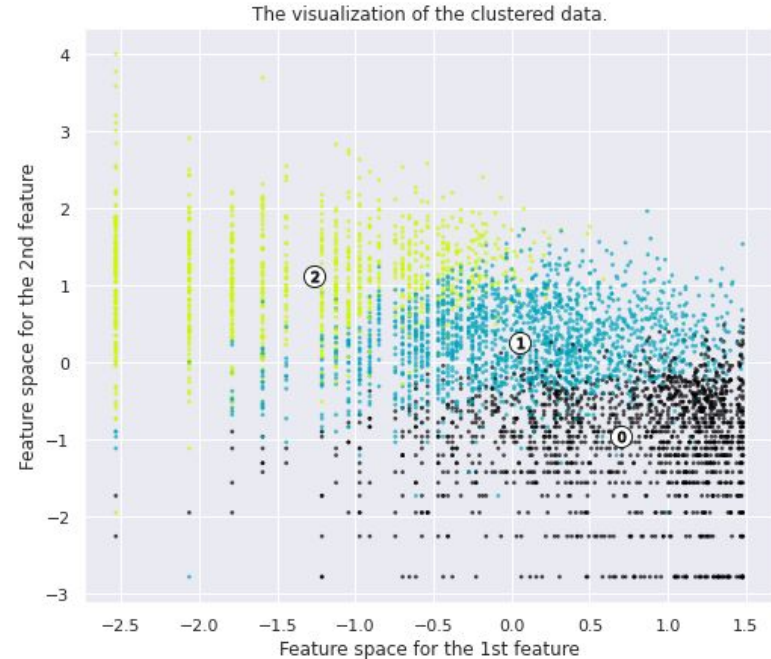
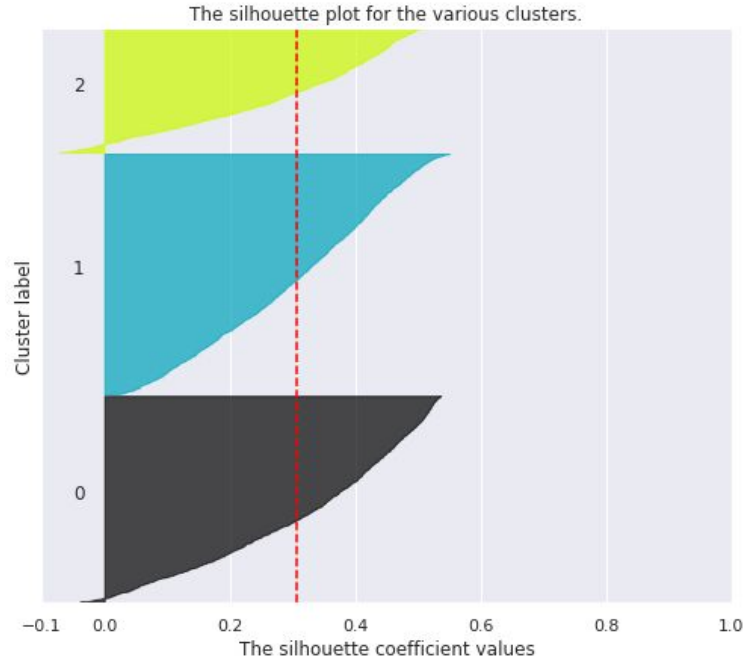
# Silhouette method on R,F,M

Silhouette analysis for KMeans clustering on sample data with  $n\_clusters = 2$



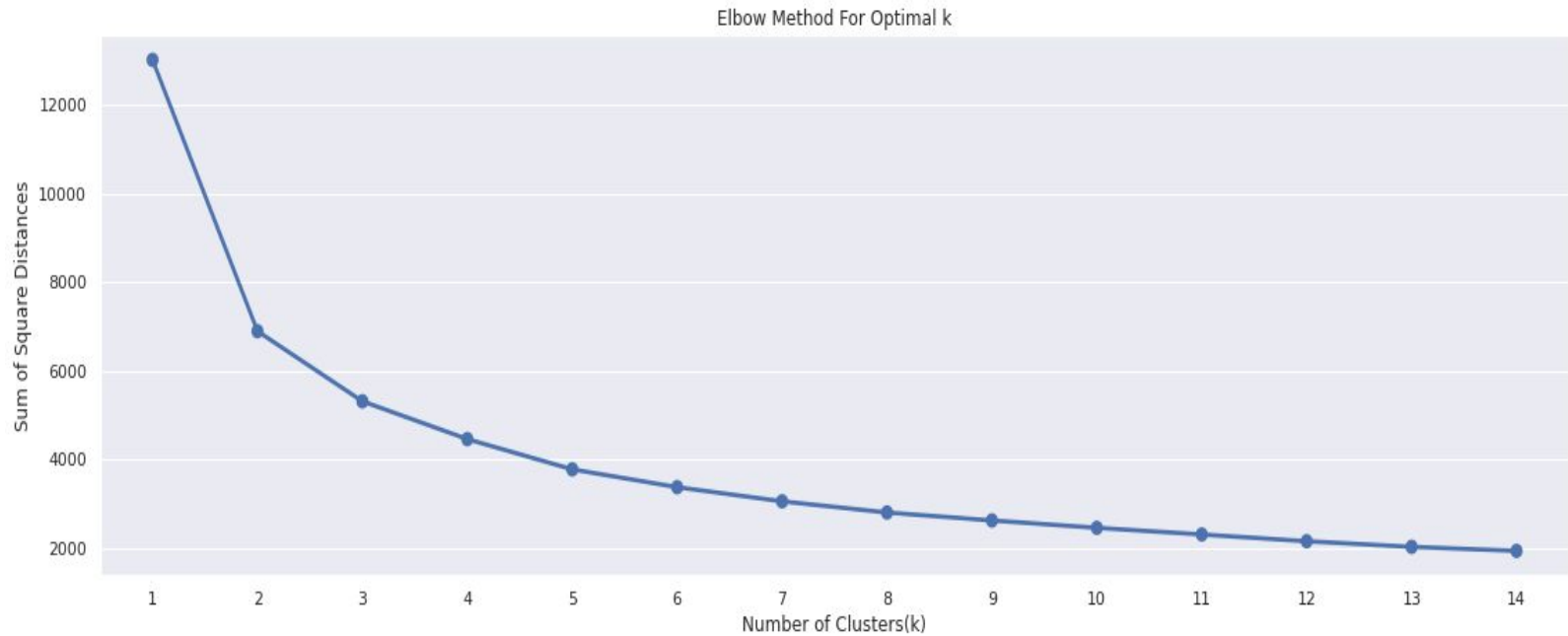
# Silhouette method on R,F,M

Silhouette analysis for KMeans clustering on sample data with  $n\_clusters = 3$

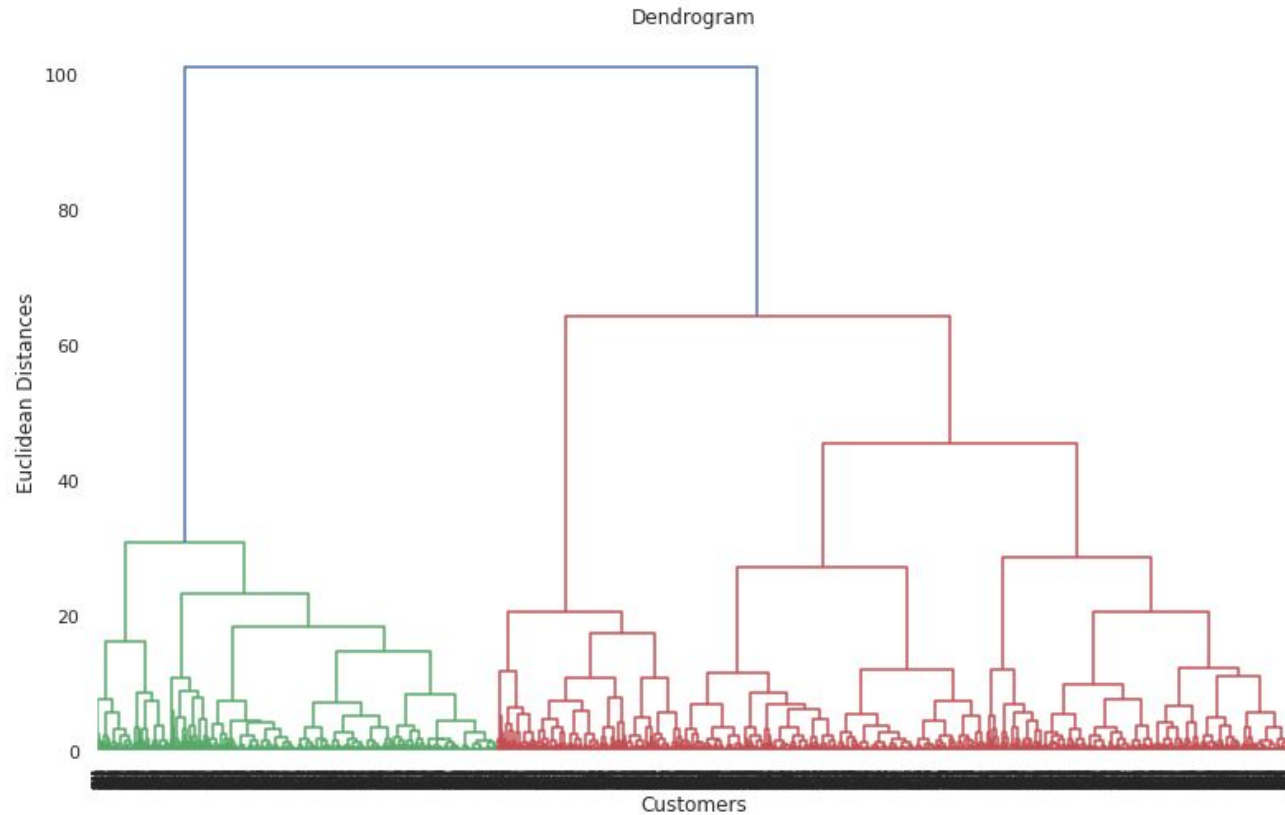




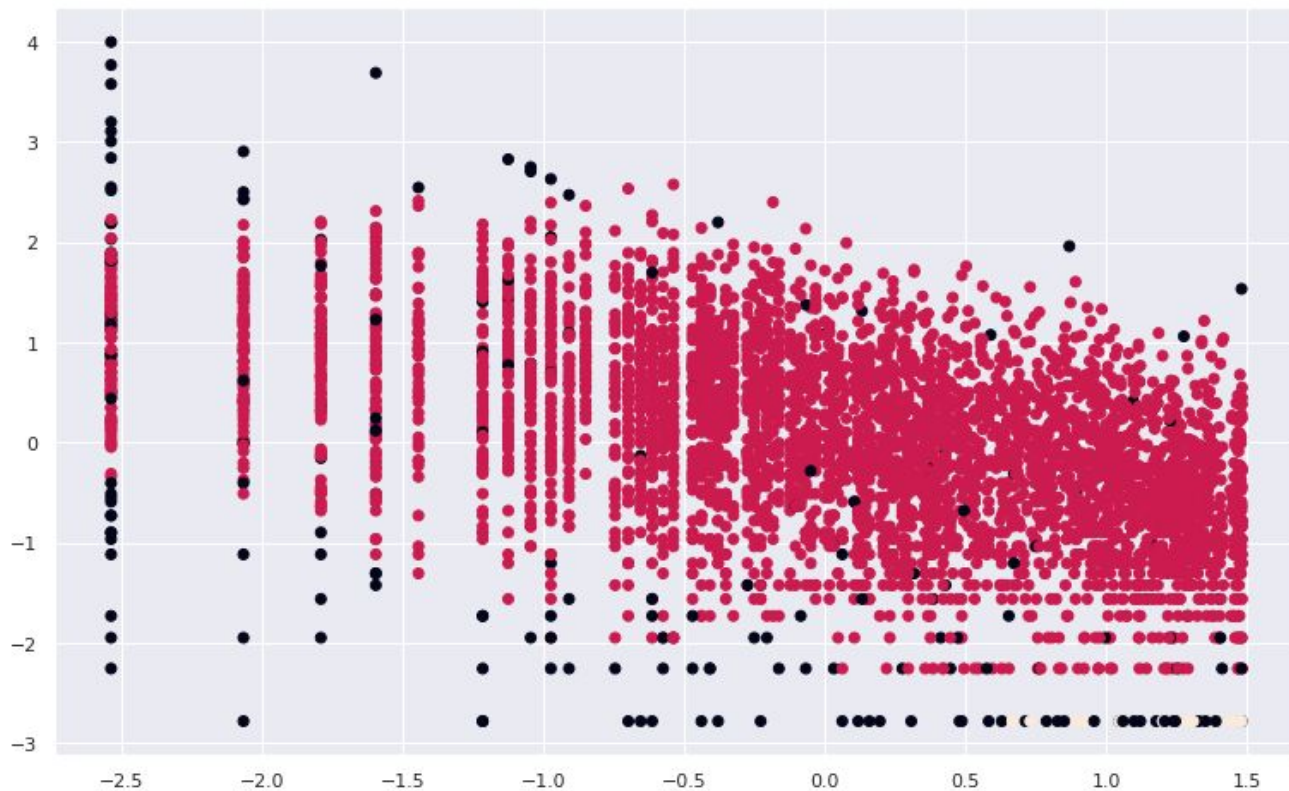
# Elbow Method



# Dendrogram



# DBSCAN on RFM



# Final Modelling table

No	Model Name	Data	Optimal number of cluster
1	K-Means with silhouette_score	RM	2
2	K-Means with Elbow method	RM	2
3	DBSCAN	RM	2
4	K-Means with silhouette_score	FM	2
5	K-Means with Elbow method	FM	2
6	DBSCAN	FM	2
7	K-Means with silhouette_score	RFM	2
8	K-Means with Elbow method	RFM	2
9	DBSCAN	RFM	3

# Challenges

- **Large Dataset to handle**
- **Lot of Nan values and cancelled order**
- **Feature Engineering**

# Conclusion

- **RFM analysis can segment customers into homogenous group quickly with set of minimum variables. Scoring system can be defined and ranged differently.**
- **RFM analysis can help in answering many questions with respect to their customers and this can help companies to make marketing strategies for their customers, retaining their slipping customers and providing recommendations to their customer based on their interest.**
- **We used the K-means algorithm to segment our customer in various clusters having similar similarity.**

**Thank You**