

# Capstone Project - 4

## Online Retail Customer Segmentation

### Team Member

Vinay Kumar

Nikhil Lamje

# Let's Predict the Demand

1. Defining Problem Statement
2. Exploratory Data Analysis
3. Feature Engineering
4. Preparation of Dataset for Modeling
5. Applying Model
6. Model Validation and Selection



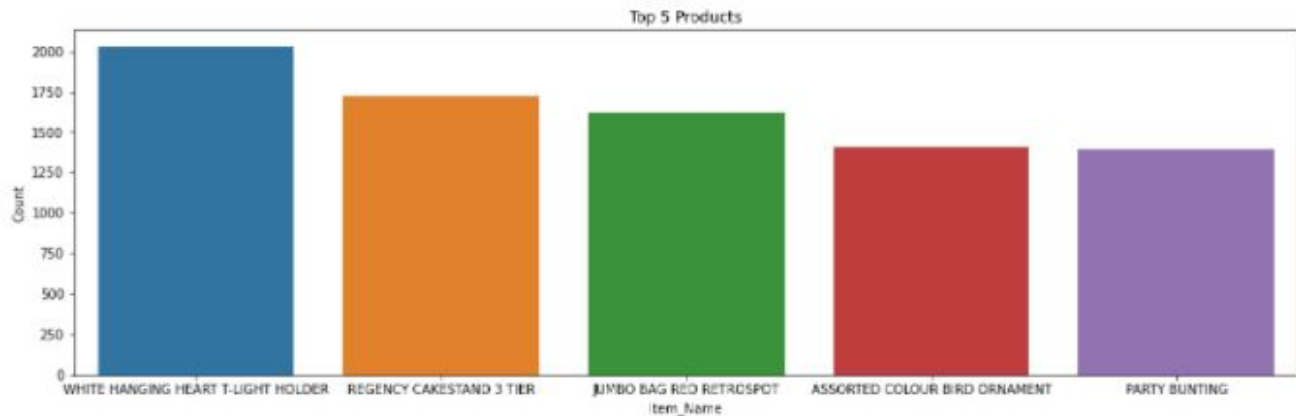
# What is the need



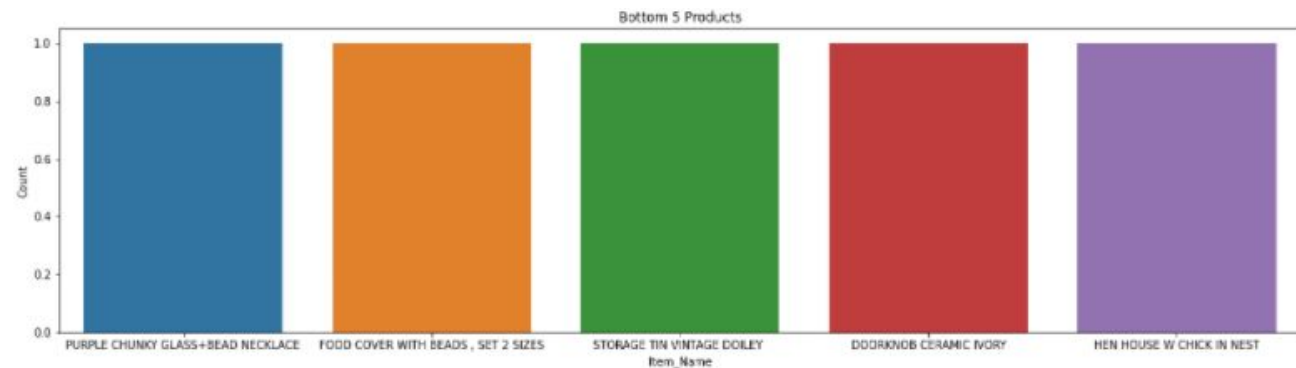
- Customer segmentation insights are used to create personalised marketing campaigns and to plan the overall marketing strategy.
- Finding consumers who are dormant or at high risk of churning and addressing their complaints are equally crucial. Companies employ the method of client segmentation for this.
- In business terms, we may say that 20% of clients account for 80% of a company's overall revenue. That is why it is critical to locate this group of people.

# Data Summary

1. **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.
2. **StockCode:** Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.
3. **Description:** Product (item) name. Nominal.
4. **Quantity:** The quantities of each product (item) per transaction. Numeric.
5. **InvoiceDate:** Invoice Date and time. Numeric, the day and time when each transaction was generated.
6. **UnitPrice:** Unit price. Numeric, Product price per unit in sterling.
7. **CustomerID:** Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.
8. **Country:** Country name. Nominal, the name of the country where each customer resides.



**Top five selling products**



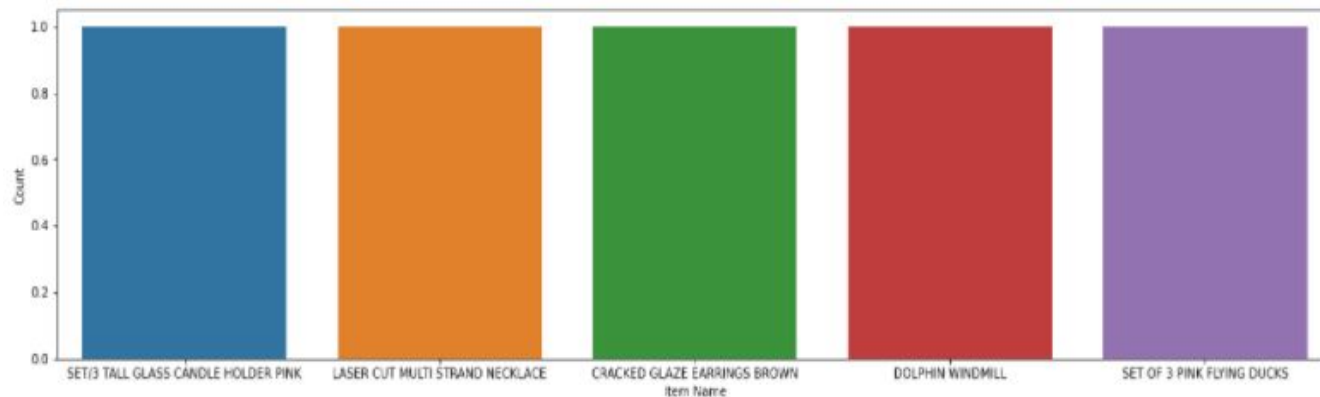
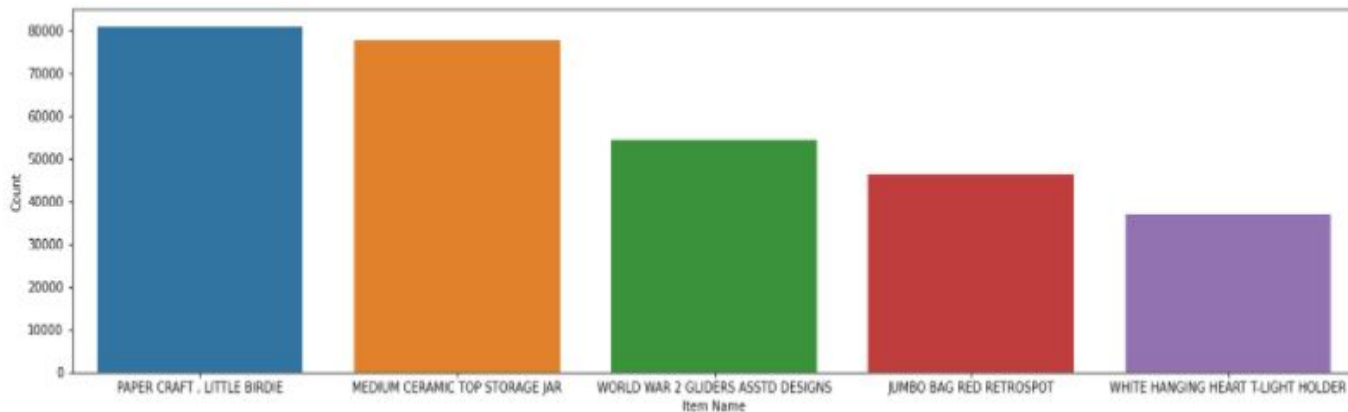
**Bottom five selling products**

# EDA



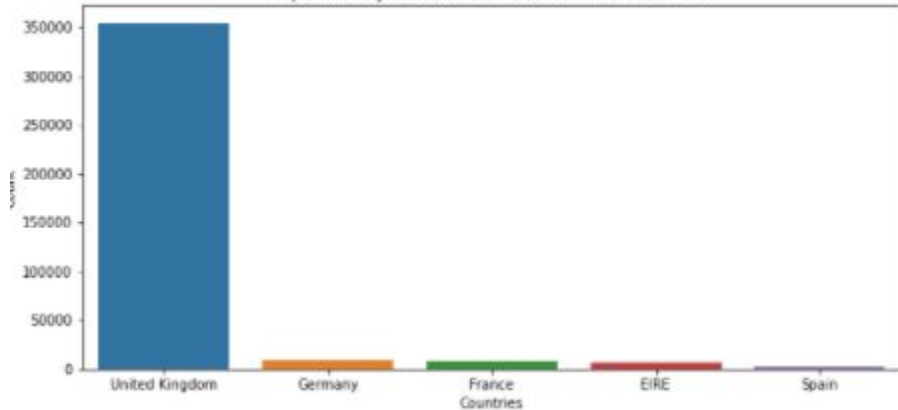
Quantity Wise

Top five selling products



Bottom five selling products

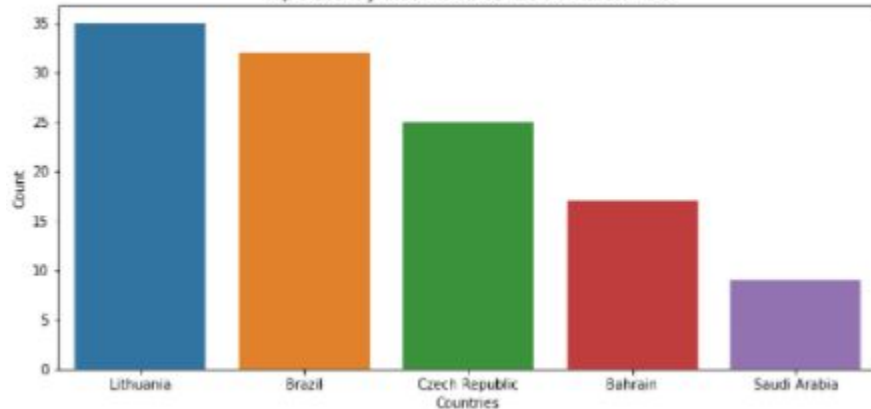
Top 5 Country based on the Most Numbers Customers



We can observe from this plot that the majority of the clients are from the United Kingdom, Germany, France, EIRE, and Spain.

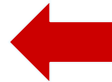
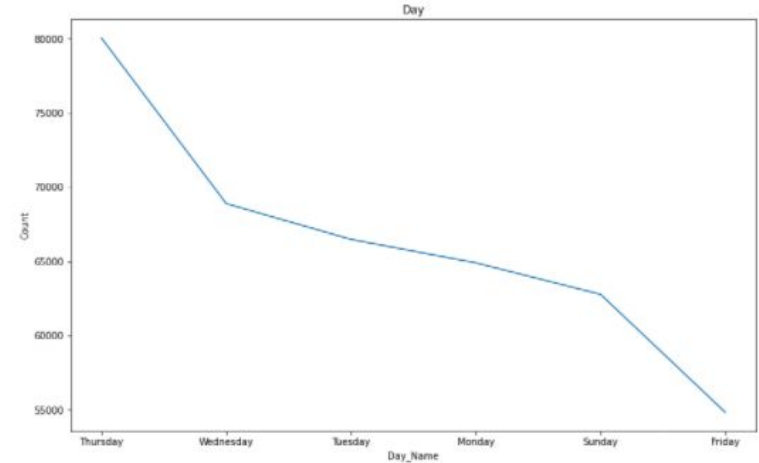
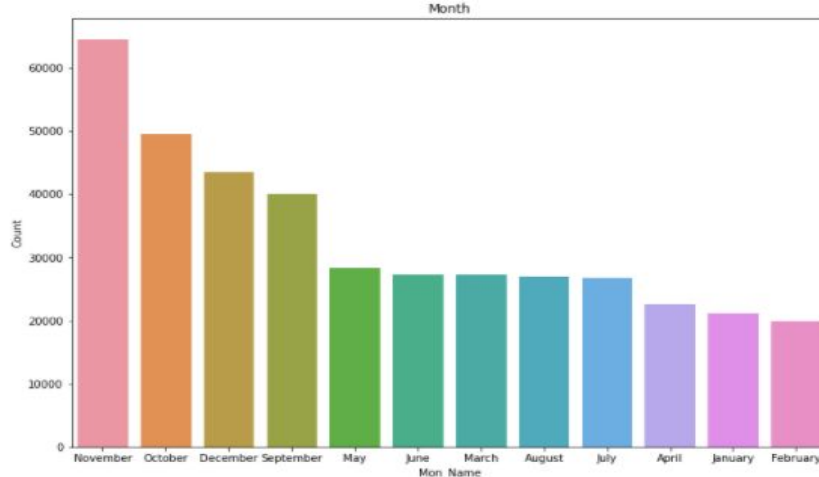
Lithuania, Brazil, the Czech Republic, Bahrain, and Saudi Arabia have the fewest clients.

Top 5 Country based least Numbers of Customers



# EDA

Most of the customers have purchases the items in Thursday ,Wednesday and Tuesday

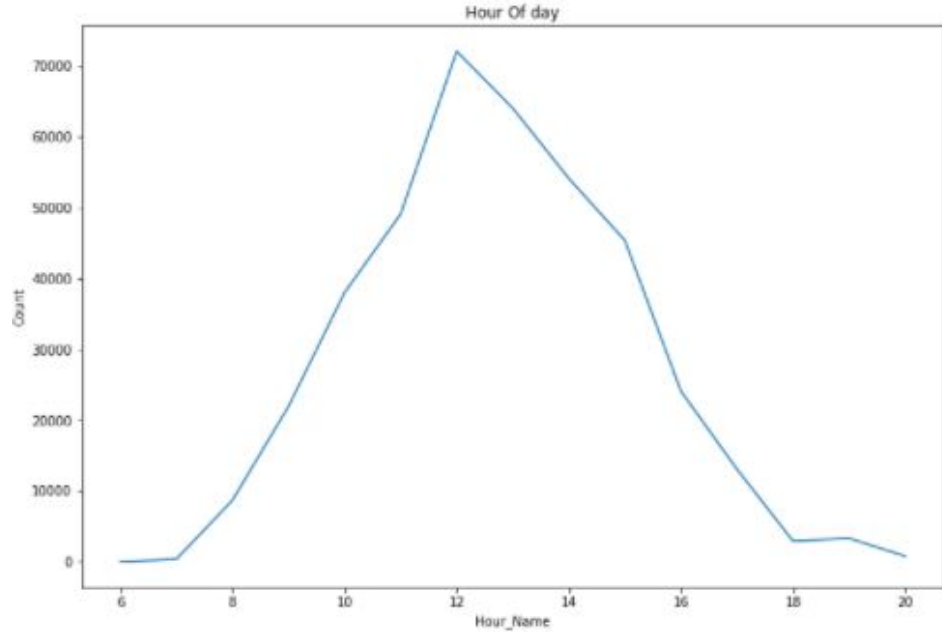


- The majority of customers purchased the gifts during the months of November, October, December, and September
- less numbers of customers have purchased the gifts in the months of April, January, and February.



# EDA

It is likely that the majority of sales take place between the 9 am to 10 a.m. and 4 p.m. Between the evenings of 12 p.m. and 4 p.m., sales are at their peak.



# RFM Modeling

RFM segmentation allows marketers to target specific clusters of customers with communications that are much more relevant for their particular behavior – and thus generate much higher rates of response, plus increased loyalty and customer lifetime value. Like other segmentation methods, RFM segmentation is a powerful way to identify groups of customers for special treatment

1. **Recency:** How much time has elapsed since a customer's last activity or transaction with the brand?
2. **Frequency:** How often has a customer transacted or interacted with the brand during a particular period of time?
3. **Monetary:** Also referred to as “monetary value,” this factor reflects how much a customer has spent with the brand during a particular period of time.

# Feature Selection

	Recency	Frequency	Monetary	R	F	M	RFMGroup	RFMScore
CustomerID								
12346.0	325	1	77183.60	4	4	1	441	9
12347.0	2	182	4310.00	1	1	1	111	3
12348.0	75	31	1797.24	3	3	1	331	7
12349.0	18	73	1757.55	2	2	1	221	5
12350.0	310	17	334.40	4	4	3	443	11

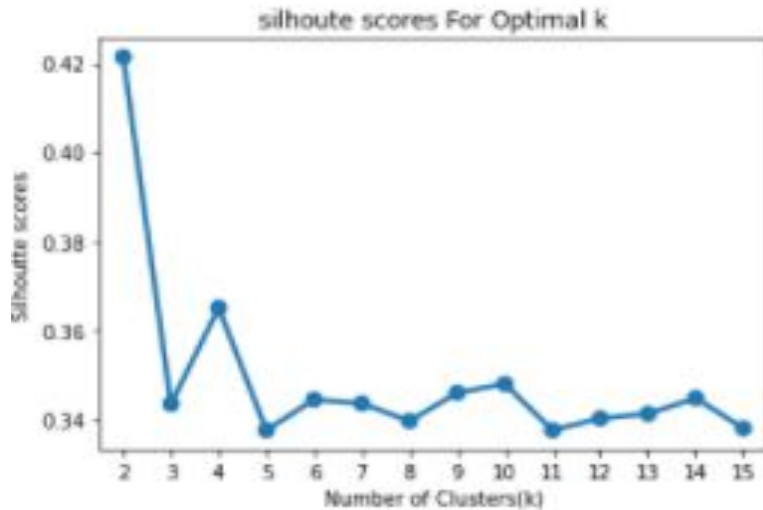
# RFM Modeling

Recency	Frequency	Monetary
R-Tier-1 (most recent)	F-Tier-1 (most frequent)	M-Tier-1 (highest spend)
R-Tier-2	F-Tier-2	M-Tier-2
R-Tier-3	F-Tier-3	M-Tier-3
R-Tier-4 (least recent)	F-Tier-4 (only one transaction)	M-Tier-4 (lowest spend)

# Recency & Monetary Models

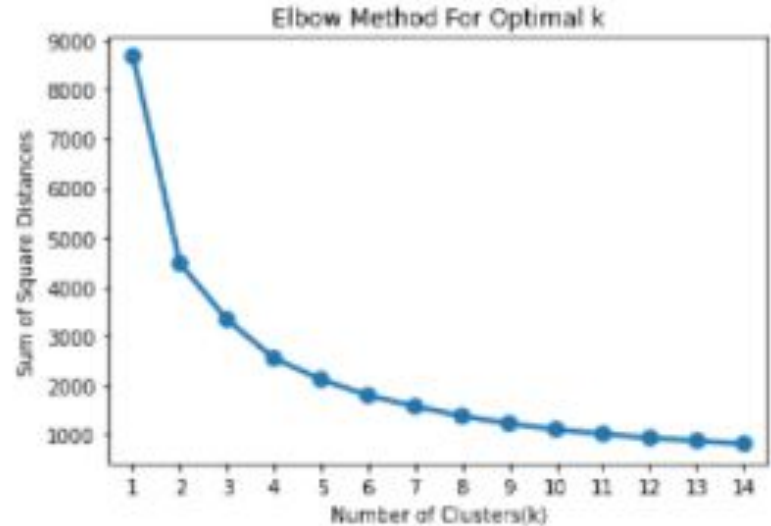
## Silhouette Score

No. of Clusters = 2



## Elbow Method

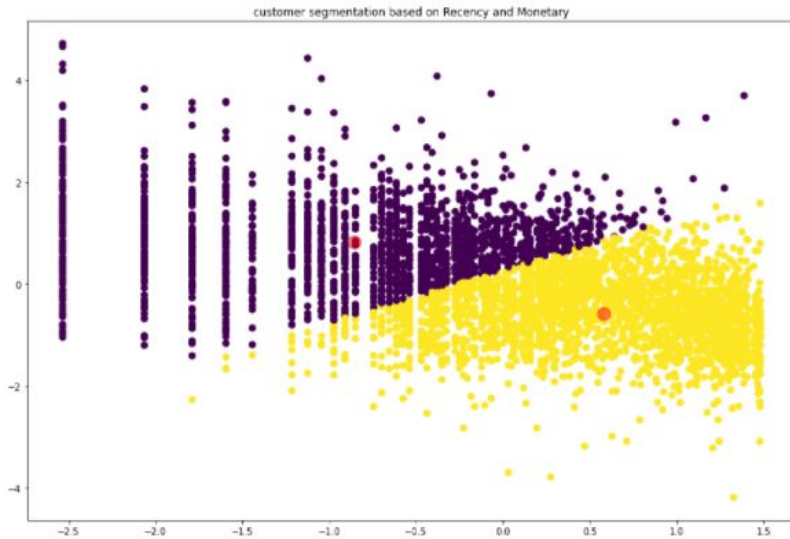
No. of Clusters = 2



# Recency & Monetary Models

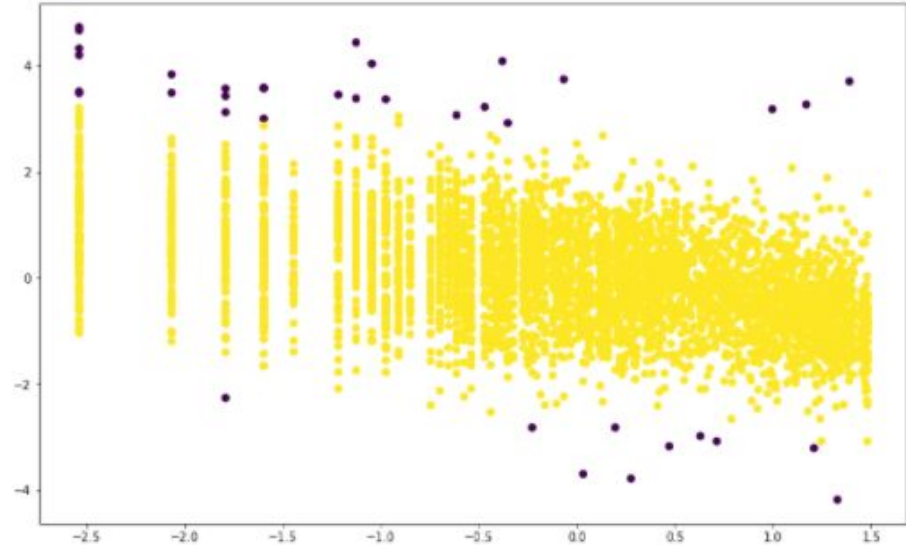
## K-Means

No. of Clusters = 2



## DBSCAN

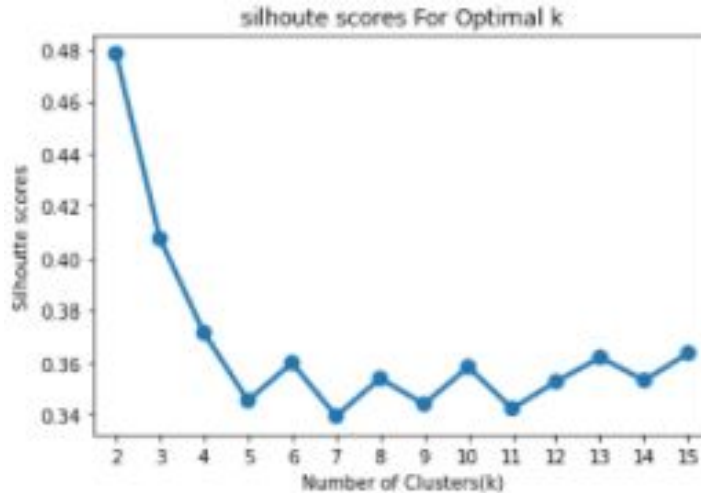
No. of Clusters = 2



# Frequency & Monetary Models

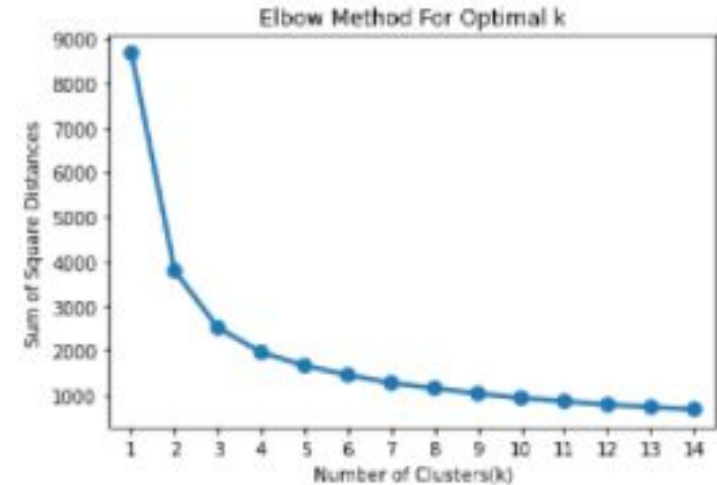
## Silhouette Score

No. of Clusters = **2**



## Elbow Method

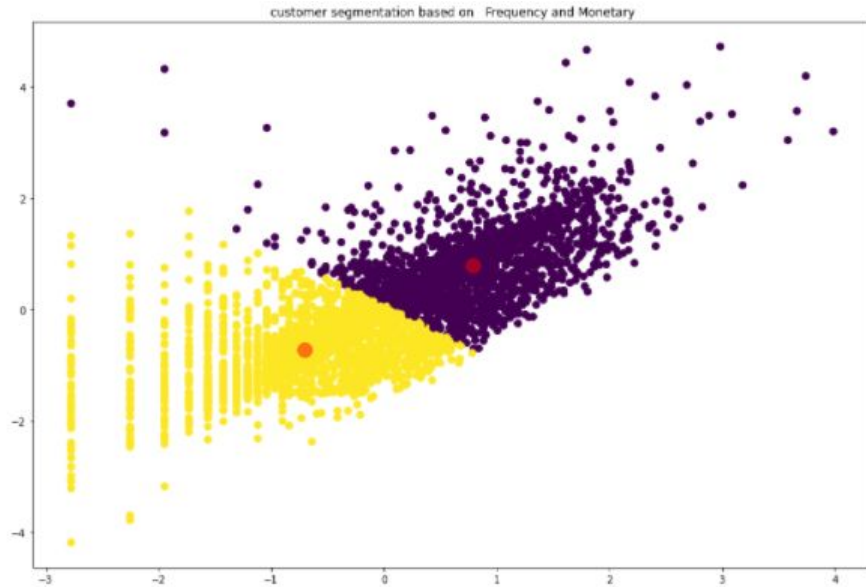
No. of Clusters = **2**



# Frequency & Monetary Models

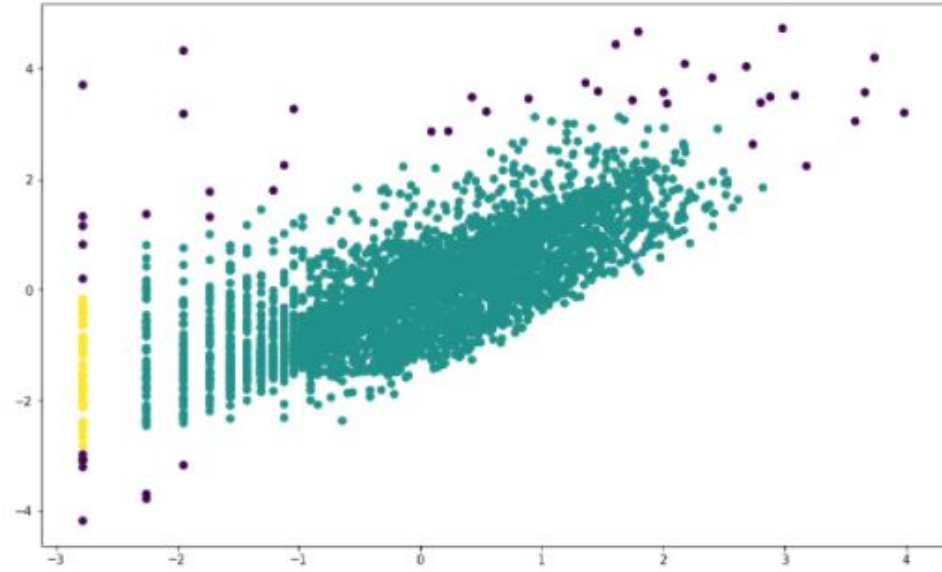
## K-Means

No. of Clusters = **2**



## DBSCAN

No. of Clusters = **3**

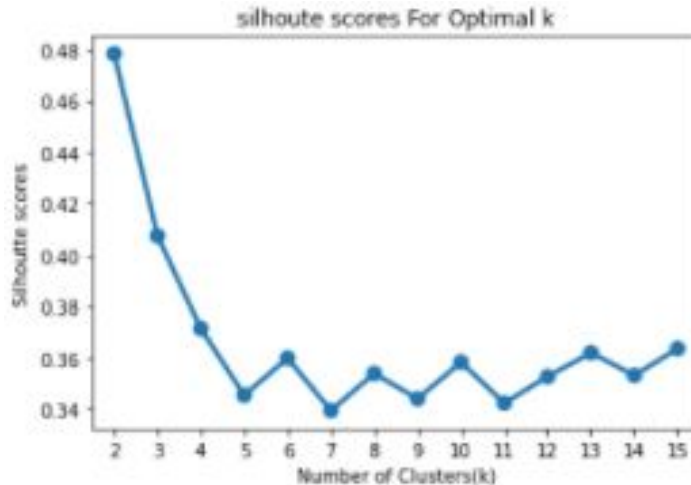




# Frequency & Recency Models

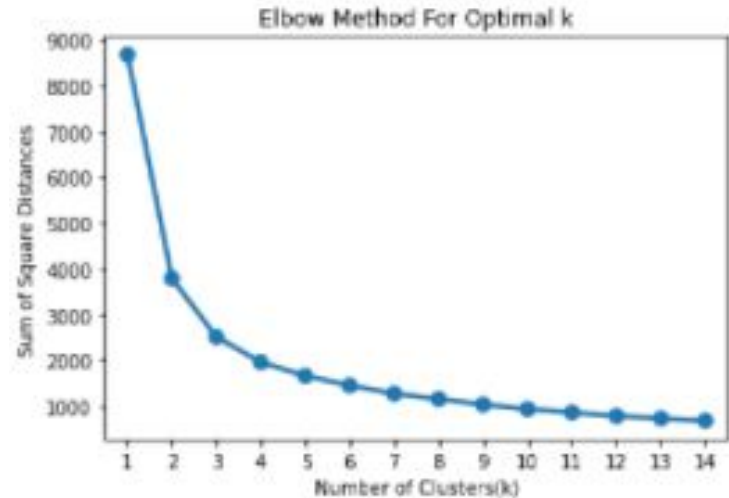
## Silhouette Score

No. of Clusters = 2



## Elbow Method

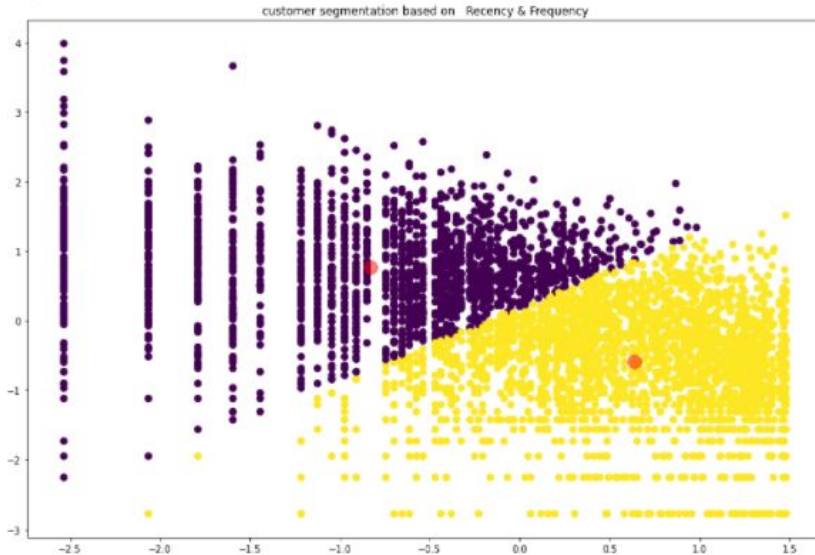
No. of Clusters = 2



# Frequency & Recency Models

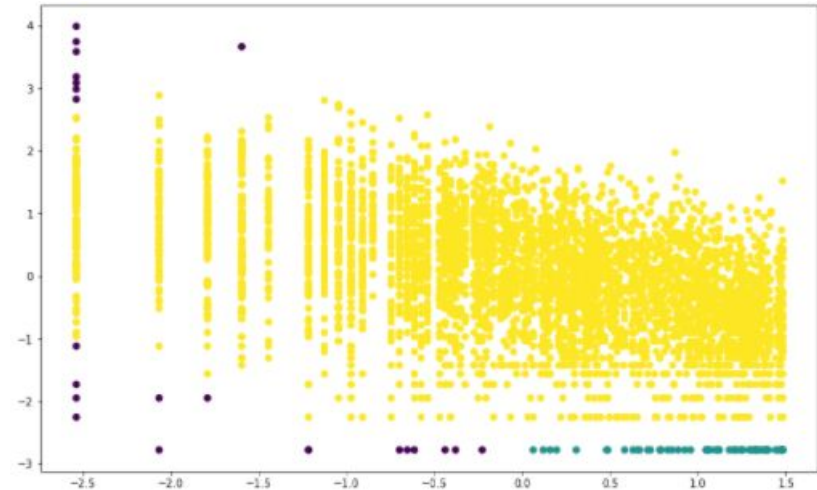
## K-Means

No. of Clusters = 2



## DBSCAN

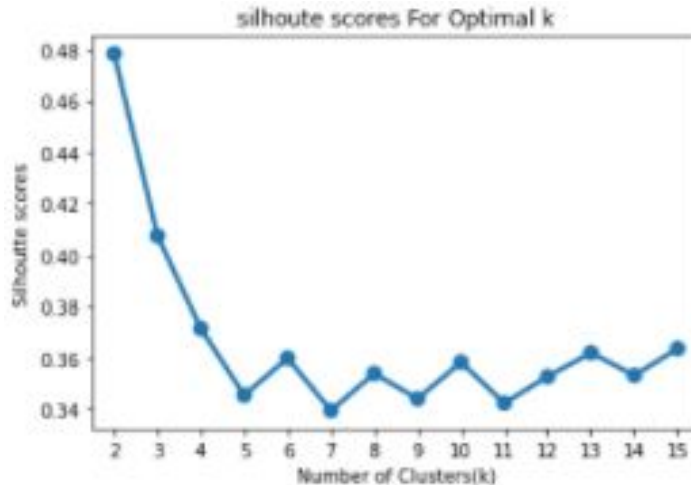
No. of Clusters = 3



# Recency, Frequency & Monetary Models

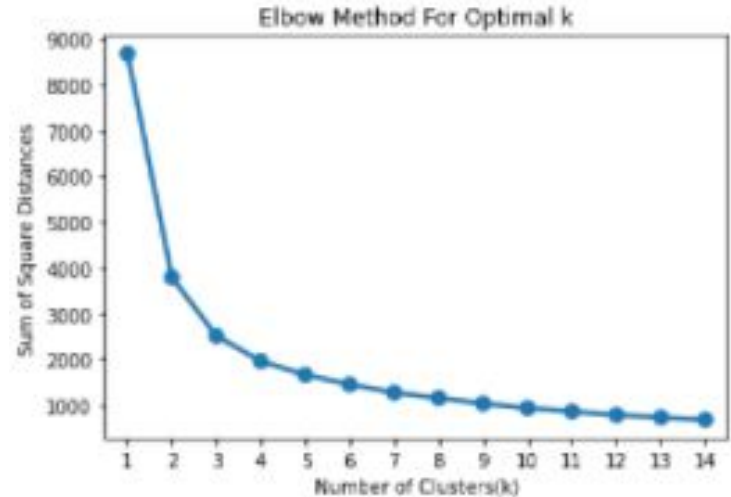
## Silhouette Score

No. of Clusters = 2



## Elbow Method

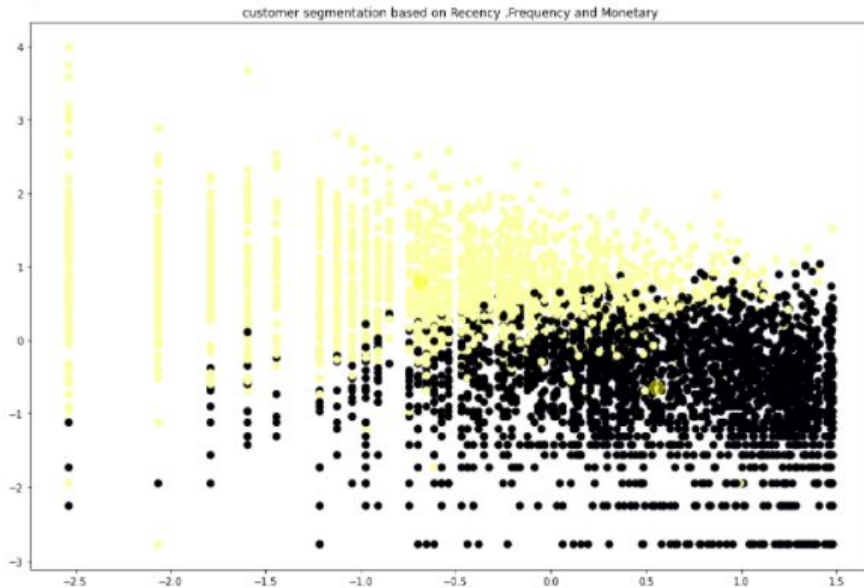
No. of Clusters = 2



# Recency, Frequency & Monetary Models

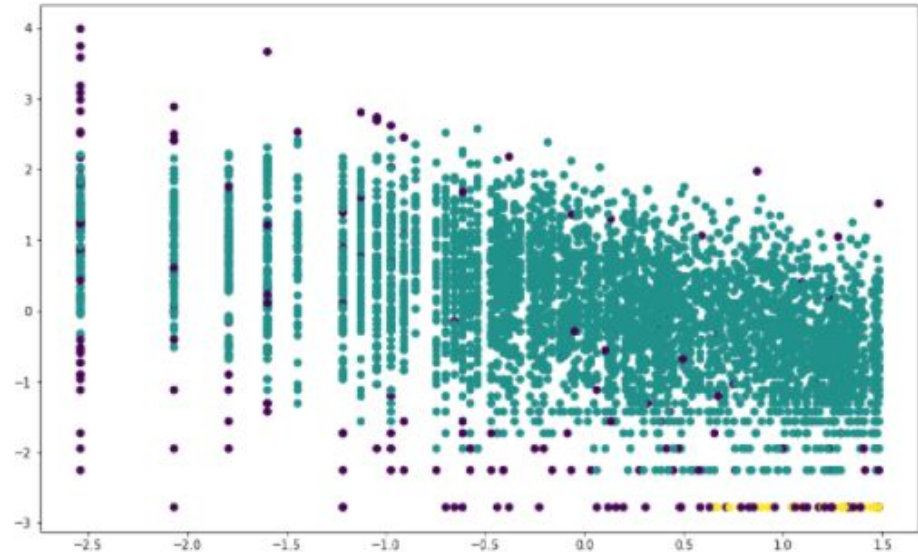
## K-Means

No. of Clusters = 2



## DBSCAN

No. of Clusters = 3



# Result

**RM**



Model	Ideal Number of Clusters
1. K-Means with silhouette_score	2
2. K-Means with Elbow method	2
3. DBSCAN	2

**RF**



Model	Ideal Number of Clusters
1. K-Means with silhouette_score	2
2. K-Means with Elbow method	2
3. DBSCAN	3

**FM**



**RFM**



# Conclusion

- This study proposed the use of machine learning techniques to identify the number of segments for online retail customers.
- Indeed, selecting the appropriate features to get the intended outcome was crucial. Here, we performed the RFM (Recency, Frequency, and Monetary) technique to provide features to our model.
- During the model development, we found the optimum number of clusters from the Elbow method and Silhouette Scores.
- So we ran DBSCAN on the dataset, and it yielded a clear number of clusters, namely three.
- Indeed, selecting the number of clusters varies from business to business case.