Gokul Ghate

Customer Segmentation

Online Retail Customers

Mind Map Customer Segmentation

1. Use Case
2. Business Understanding
3. Data Understanding
4. Data Preparation
5. Data Cleaning
6. Exploratory Data Analysis
7. Modeling
8. Evaluation
9. Recommendation

Use Case: Customer Segmentation

Objective Statement:

- Get a business insight into how many products are sold every month.
- Get a business insight into how many customers spend their money every month.
- To reduce risk in deciding where, when, how, and to whom a product, service, or brand will be marketed.
- To increase marketing efficiency by directing effort specifically toward the designated segment in a manner consistent with that segment's characteristics.

Challenges:

- The large size of data, can not maintain by an excel spreadsheet.
- Need several coordination from each department.
- Demography data have a lot of missing values and typos.

Methodology / Analytic Technique

- Descriptive Analysis
- Graph Analysis
- Segment Analysis

Expected Outcome

01

Know how many products sold every month.

02

Know how many customers spend their money every month.

03

Customer segmentation analysis.

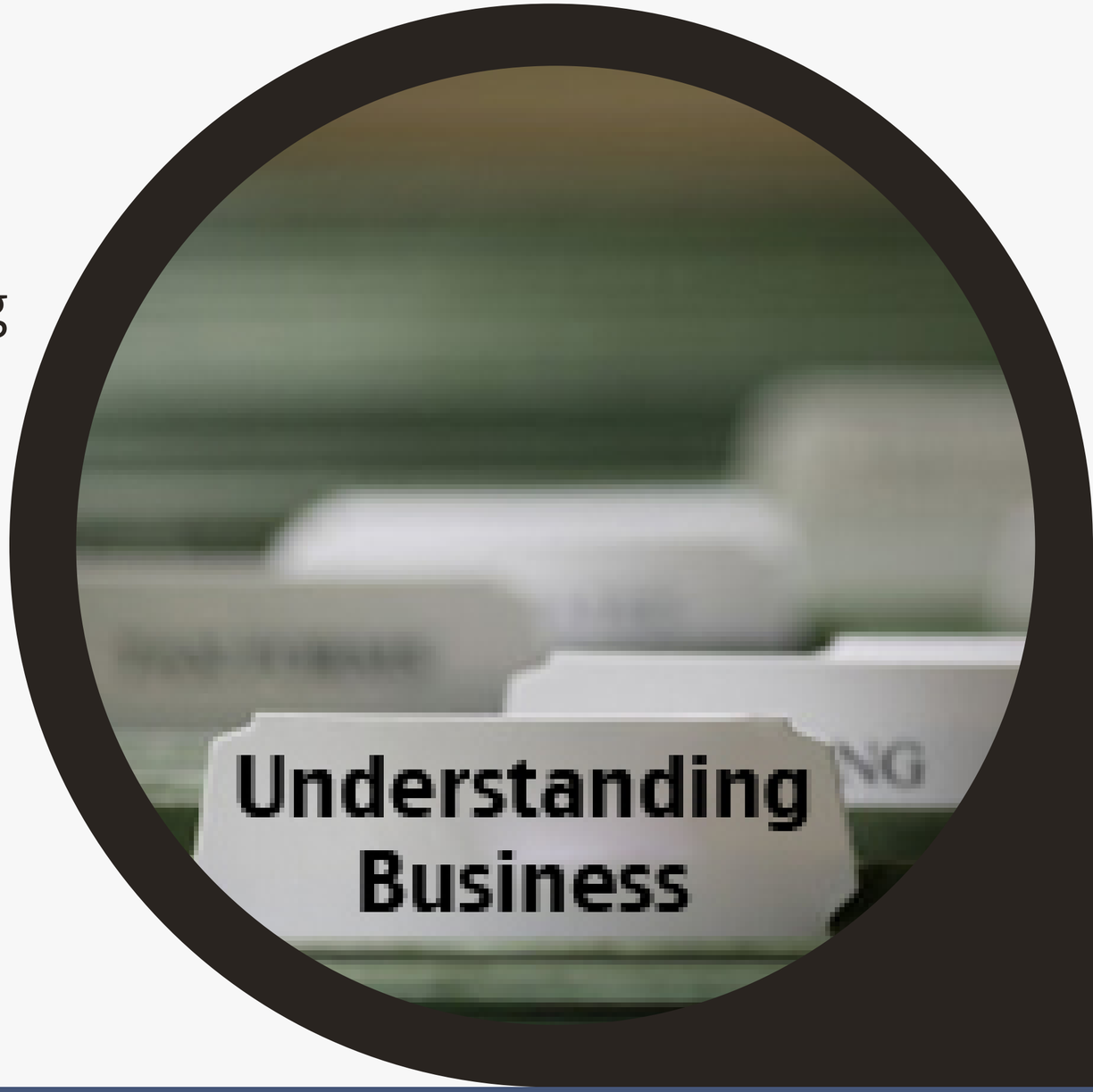
04

Recommendation based on customer segmentation.

Retail is the process of selling consumer goods or services to customers through multiple channels of distribution to earn a profit.

This case has some business questions using the data:

- How many products are sold every month?
- How much does a customer spend their money every month?
- How about Customer segmentation analysis?
- How about recommendations based on customer segmentation?



Data Understanding

- Data of Retail Transaction from 01 December 2010 to 09 December 2011.

- Source Data: Online retail dataset by UCI Machine Learning Library. Customer Segmentation

- The dataset has 8 columns and 541909 rows.

Data Cleansing

- There are about 25% of Null CustomerID in the data. We need to remove them as there is no way we can get the number of CustomerID.

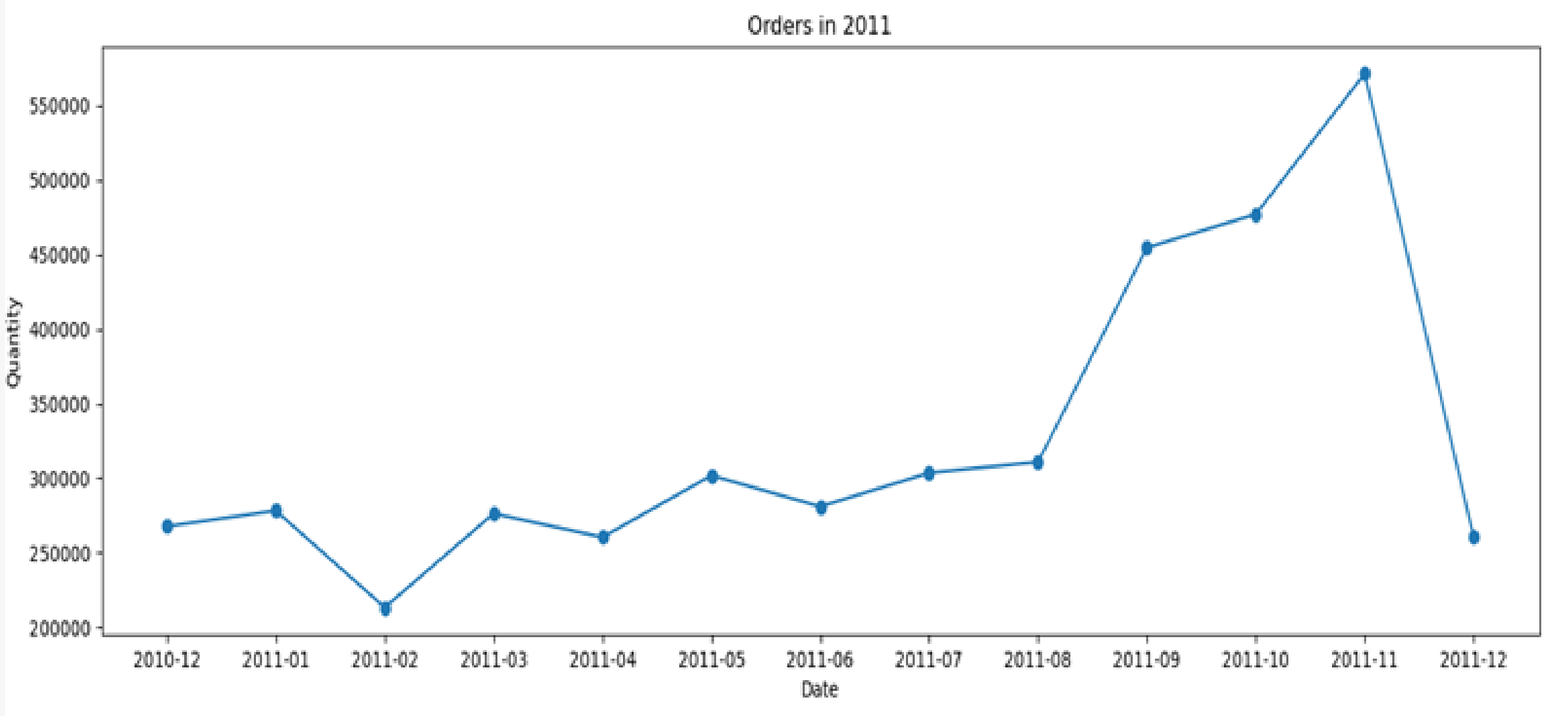
- There are few records with UnitPrice<0 and Quantity<0. We need to remove them from the analysis. This could represent canceled or returned orders.

- There is more than 90% of ‘United Kingdom’ customers, therefore we will restrict the data to only United Kingdom customers.

Exploratory Data Analysis

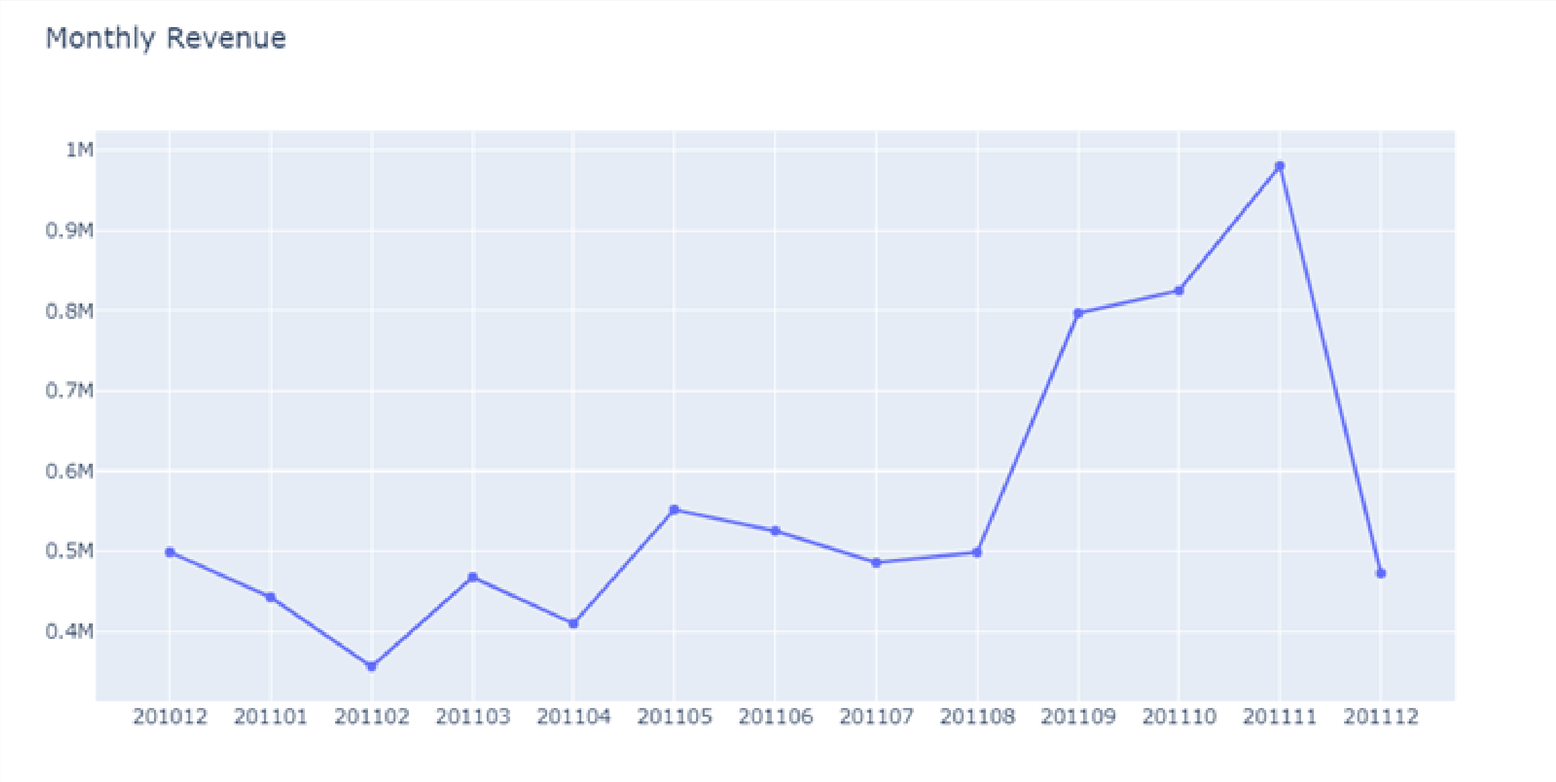
How many products are sold every month?

Product sold in November has the highest quantity that has around **13.41%** product sold from all transaction along 1 year. Therefore the business team can increase sales in this month such as promoting new products to customers in this months.



Exploratory Data Analysis

How much customer spend their money every month?

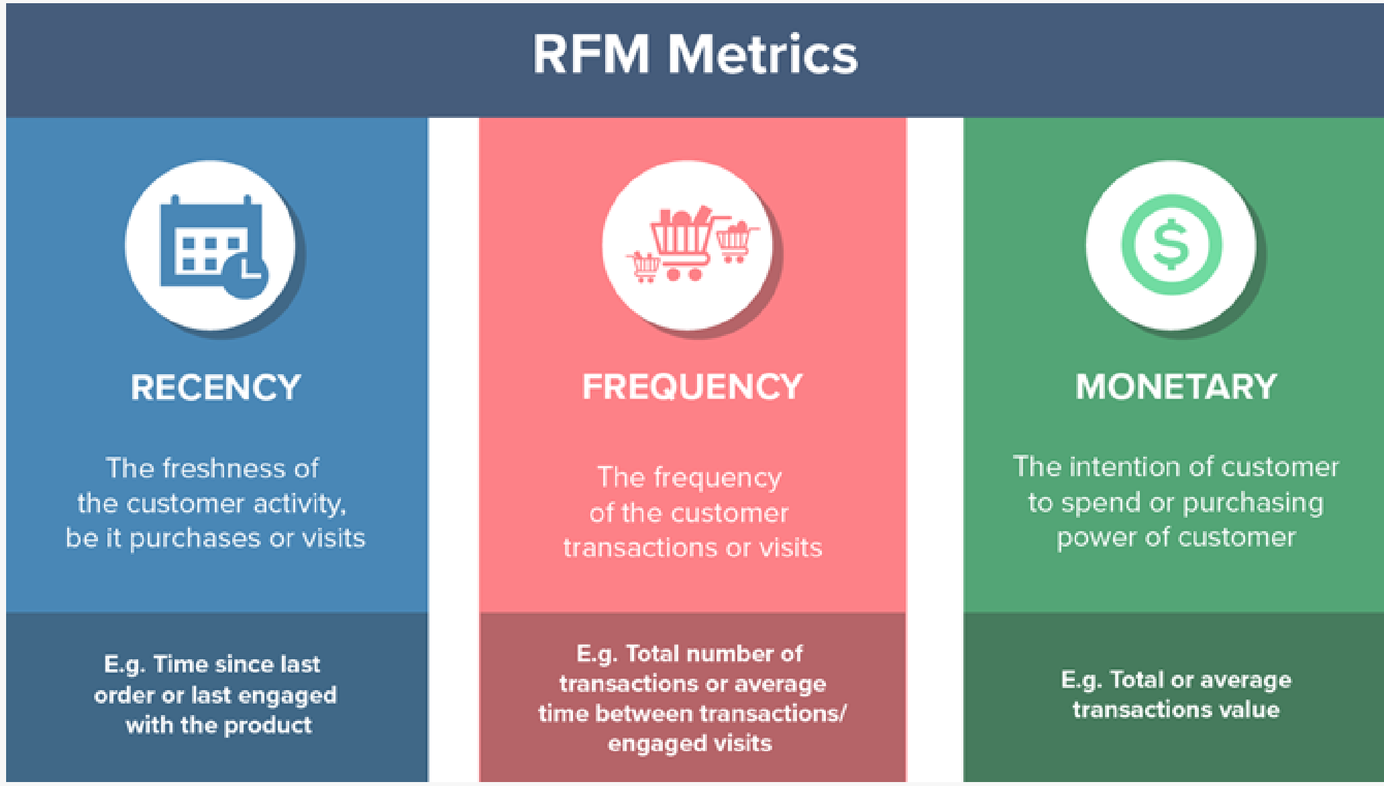


Data Modeling: RFM Quantiles

Recency Frequency Monetary (RFM)

RFM analysis allows you to segment customers by the frequency and value of purchases and identifies those customers who spend the most money.

- Recency – how long it's been since a customer bought something from us.
- Frequency – How often a customer buys from us.
- Monetary value – the total value of purchases a customer has made.



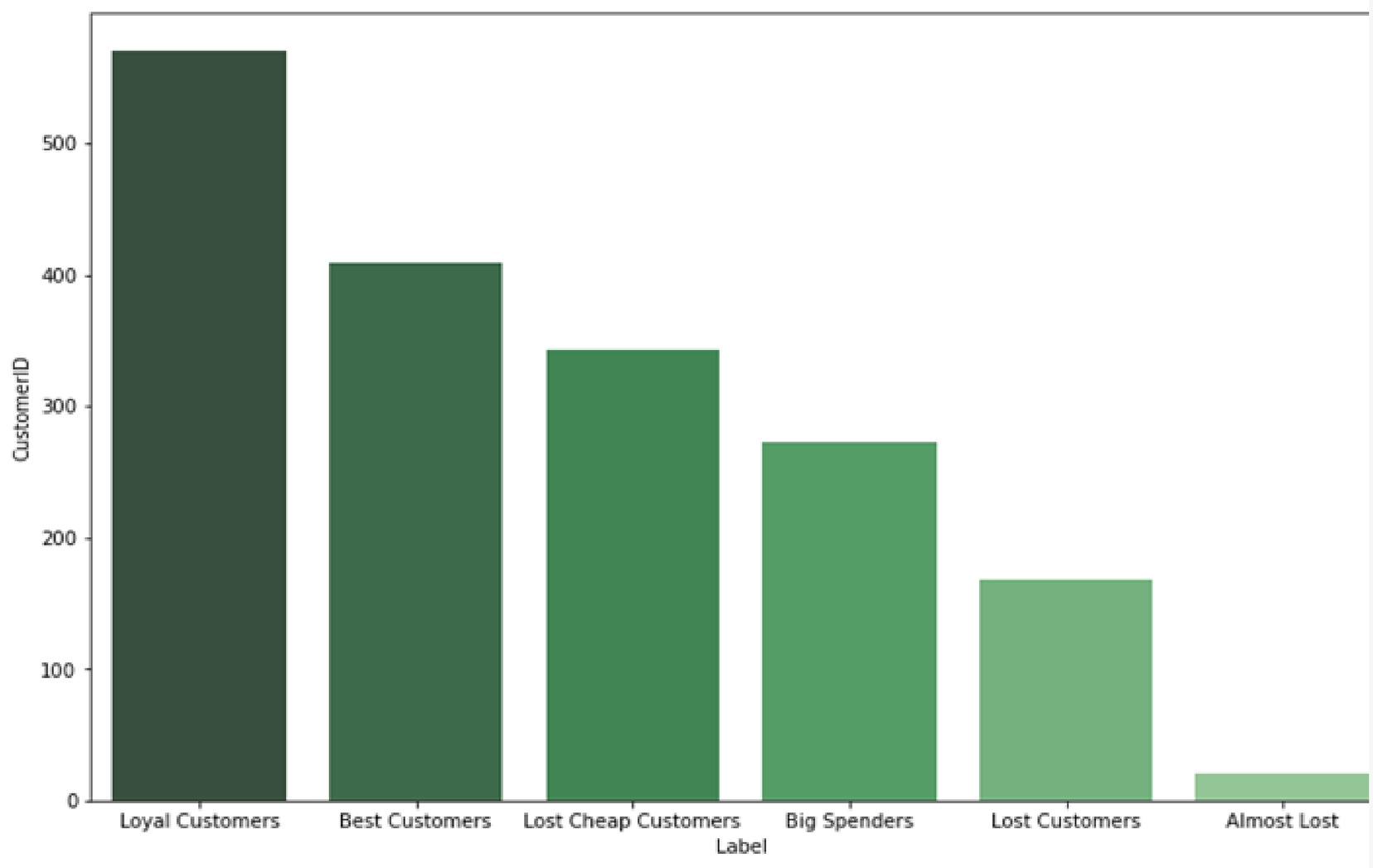
Data Modeling: RFM Quantiles

RFM Quantiles

- Split the metrics into segments using quantiles.
- We will assign a score from 1 to 4 to each Recency, Frequency, and Monetary respectively.
- 1 is the highest value, and 4 is the lowest value.
- A final RFM score (Overall Value) is calculated simply by combining individual FRM score numbers.

Segment	RFM Score
Best Customers	111
Loyal Customers	F=1
Big Spenders	M=1
Almost Lost	134
Lost Customers	344
Lost Cheap Customers	444

Data Modeling: RFM Quantiles



Data Modeling: K-Means Clustering

- K-Means clustering algorithm is an unsupervised machine learning algorithm that uses multiple iterations to segment the unlabeled data points into K different clusters in a way such that each data point belongs to only a single group that has similar properties.

- K-means gives the best result under the following conditions:

1. Data’s distribution is not skewed
2. Data is standardized

- The data is highly skewed, therefore I will perform log transformations to reduce the skewness of each variable and I standardized the data.

Data Modeling: K-Means Clustering

- K-Means clustering algorithm is an unsupervised machine learning algorithm that uses multiple iterations to segment the unlabeled data points into K different clusters in a way such that each data point belongs to only a single group that has similar properties.

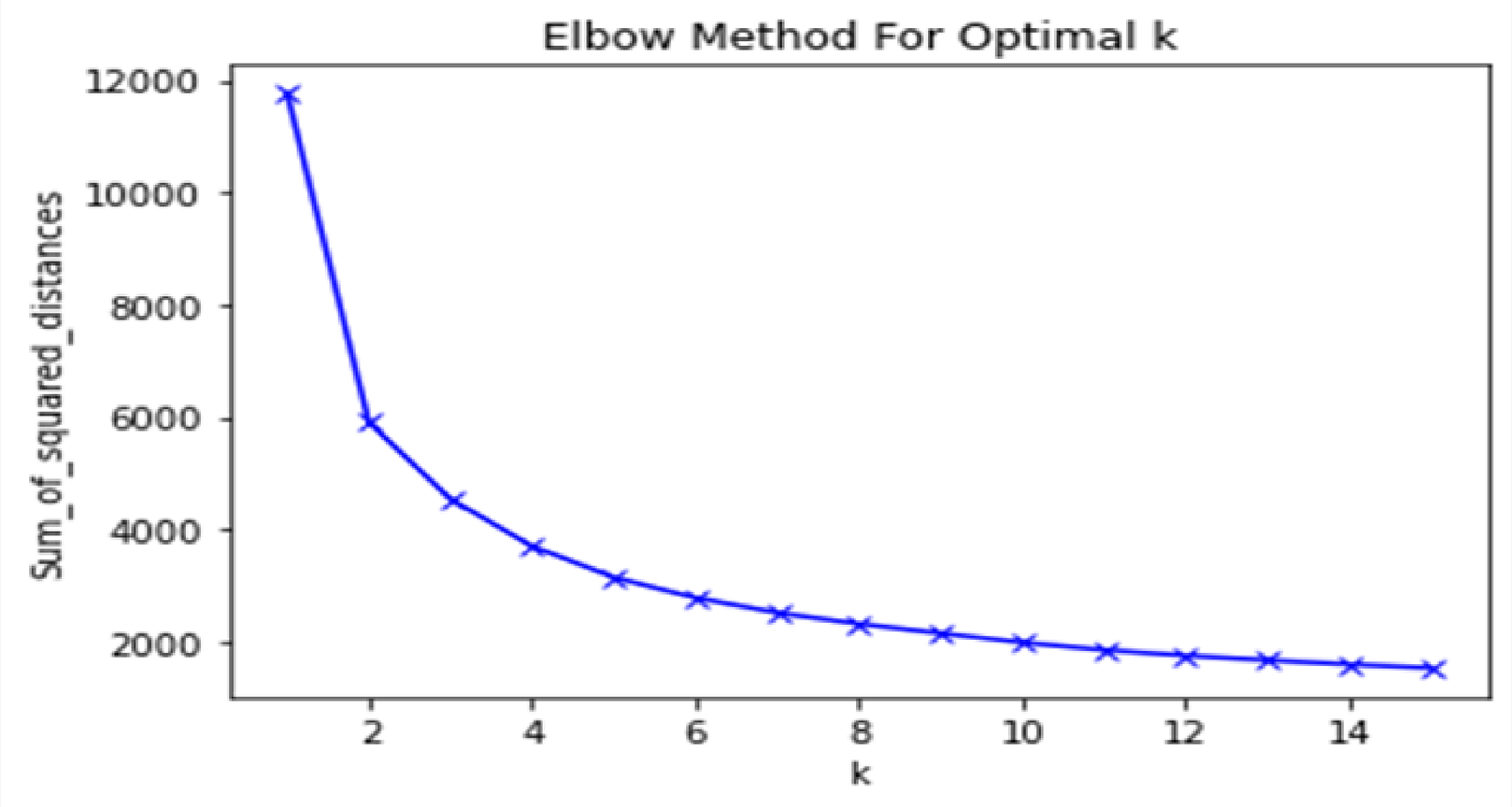
- K-means gives the best result under the following conditions:

1. Data’s distribution is not skewed
2. Data is standardized

- The data is highly skewed, therefore I will perform log transformations to reduce the skewness of each variable and I standardized the data.

Data Modeling: K-Means Clustering

Finding the optimal number of clusters (Finding K) using the Elbow Method.



Evaluating Model: K-Means Clustering

Davies Bouldin Score is a metric for evaluating clustering algorithms.

The smaller Davies Bouldin Score is the more optimal the cluster is.

K-Means 4 clusters have the lowest Davies Bouldin Score than another cluster. Therefore the optimum cluster is 4.

K-Means Cluster	Davis Bouldin Score
3	1.119
4	1.065
5	1.067

Interpretation of clusters formed using k-means:

- **“Cluster0”** has **29%** customers. It belongs to the **“Loyal Customer”** segment as they haven’t purchased for some time, but used to purchase frequently (F=2) and spent a lot.

- **“Cluster 1”** has **20%** customers. It can be interpreted as **“Almost Lost”**. They purchase recently (R=2). However, they do not purchase frequently and do not spend a lot.

- **“Cluster 2”** has **30%** customers. It can be interpreted as **“Lost Cheap Customers”**. Their last purchase is long ago (R=4), purchased very few (F=4) and spent little (M=4).

- **“Cluster 3”** has **21%** customers. It belongs to the **“Best Customers”** segment which we saw earlier as they purchase recently (R=1), frequently buyers (F=1), and spent the most (M=1).

Recommendation

Recommendation for “Best Customers” segment:

- Focus on increasing customer purchases therefore it is necessary to form a cross/Up-Selling Strategy.

Recommendation for “Loyal Customers” segment:

- The business team must optimize the budget campaign and the time campaign for this customer segment to maintain their loyalty and increase their value.

Recommendation for “Almost Lost” segment:

- This customer segment is very at risk for churn, so focus on activating customers and making repurchases by forming a Reactivation Strategy, Retention Strategy.

Recommendation for “Lost Cheap Customers” segment:

- This customer segment has churned, so the focus of the campaign is to reactivate the customer by forming a Reactivation strategy.

Gokul Ghate

Thank you!

gg.gokulghate@gmail.com

https://github.com/gokul-insights/datascience

Gokul Ghate - https://www.linkedin.com/in/gokulghate/