Online Retail

MACHINE LEARNING- UNSUPERVISED
MINI PROJECT

GROUP 1

Our Team

DJEREMIAH CHRISTOFEL

ANNISA DWI ARI

SALWA AHLA AMANIA

FARHAN IMAM NAUFAL

MIKHAEL ADITHA SEMBIRING MELIALA

Outline

01 THE OBJECTIVE

02 DATA PREPARATION

03 ANALYSIS & FINDING

04 CONCLUSION

01

THE OBJECTIVE

Background

E-commerce has become a new channel to support businesses development. Through e-commerce, businesses can get access and establish a wider market presence by providing cheaper and more efficient distribution channels for their products or services. E-commerce has also changed the way people shop and consume products and services. Many people are turning to their computers or smart devices to order goods, which can easily be delivered to their homes.

Goals

Gather useful intrinsic information by combining RFM model and some unsupervised learning model.

For each segments, analyze the data, and try to answer the following questions

- How was the sales trend over the months?
- What are the most frequent purchased products?
- How many products does the customer purchase in each transaction?
- How does each cluster/group contribute to the company's revenue?
- What are the interval date between transactions?
- Based on your findings, what strategy could you recommend to the business to gain more profit?

Methodology

METHODOLOGY

- Data Preparation
- Preprocessing
- RFM Model Building
- Clustering
- Predictive Model Building
- EDA

02

DATA PREPARATION

Initial Data Quality Report

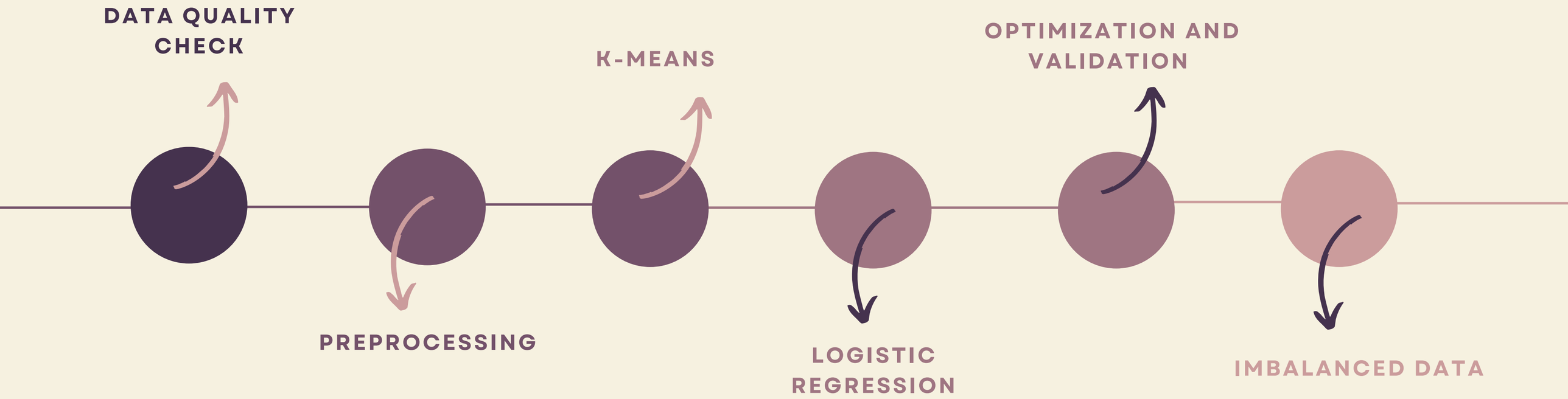
Total of 541,909 Data

Total of 8 Columns

2 Columns with Missing Value

136,534 missing values

Preparation Steps & Process

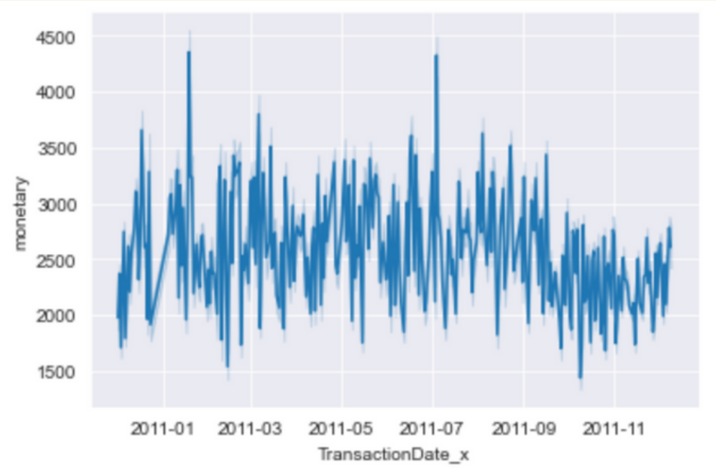


Analaysis

AND FINDING

Sales Trend

CLUSTER 0



For c0 (Cluster 0) it tends to have the lowest monetary value at 1500 Pounds. The highest monetary value was found in January 2011 and July 2011 ranging from 4400 pounds.

CLUSTER 1

For c1 (Cluster 1) has the highest monetary reach 280000 Pounds. Monetary on c1 tends to be evenly distributed, it can be concluded that customers on c1 shop regularly.

CLUSTER 2

For c2 (Cluster 2) it has a fairly high monetary range above 120,000 pounds. Monetary on c2 is almost the same as c1, but on c2 experienced several decreases in Feb 2011, May 2011 and Oct 2011.

CLUSTER 3

Untuk c3 (Cluster 3) memiliki monetary cukup rendah dibandingkan dengan c2 berkisar di antara 10000-40000 Pound.

Most Sold

CLUSTER	PRODUCT	QUANTITY
0	WORLD WAR 2 GLIDERS ASSTD DESIGNS	28968 UNIT
1	CREAM HEART CARD HOLDER	5946 UNIT
2	POPCORN HOLDER	11624 UNIT
3	WORLD WAR 2 GLIDERS ASSTD DESIGNS	23376 UNIT

Customer Purchase in each Transaction

CLUSTER	UNIT AVERAGE
0	9 UNIT/TRANSCACTION
1	114 UNIT/TRANSACTION
2	28 UNIT/TRANSACTION
3	18 UNIT/TRANSACTION

Conclusion

In the clustering that has been done, c1 becomes a cluster that provides a revenue of 258 Pounds for each customer. While the other clusters have a much smaller average revenue of 52 pounds for each customer. But on the other hand, the average cancellation rate in each cluster is twice the average sales.

c1 has the lowest recency compared to other clusters.

Strategy

To be able to generate maximum profit, companies can increase the number of sales by reducing the number of product cancellations. Product cancellations generally occur because the product is not ready in stock, thus encouraging the customer to cancel the product.

Thank you
CHUR