THE EMPIRICAL APPROACH IN CUSTOMER DECOMPOSITION USING ML TECHNIQUES

Dr.Sumanth V[1]*, Dr. Shashidhar T M[2], PRASHANTH KUMAR .S .P[3], Dr. Manujakshi B C[4]

[1]*Associate professor, Department of Computer science and engineering, Acharya institute of technology, Bengaluru, ssumanth6@gmail.com
[2]Associate Professor, Department of Computer Science and Engineering, Acharya Institute of Technology, Bengaluru, shashilara@gmail.com
[3]Assistant professor, Department of computer science and Engineering, Acharya Institute of technology, Bengaluru, prashantheshwar2010@gmail.com
[4]Associate Professor, School of Computer Science and Engineering, Presidency University, Bengaluru. manujakshibc@gmail.com

***Corresponding Author: -** Dr.Sumanth V
*Associate professor, Department of Computer science and engineering, Acharya institute of technology, Bengaluru, ssumanth6@gmail.com

DOI: 10.48047/ecb/2023.12.si5a.0245

1 INTRODUCTION

1.1. Project Introduction

Over the years, the business sector has seen major changes in business growth. Businesses are setting new goals and therefore a competitive environment is developing in the business sector. As a result, many competing business sectors will not succeed. The reason is "Companies fail to teach their customers". Any company can achieve success and growth, but the analysis and understanding of customers and the market is where it fails. The answer to this problem is to understand what customer decomposition is. Customer decomposition is the segregation of the market into multiple groups of consumers who share similar characteristics. As customer decomposition is one the most important applications of unsupervised learning. Using clustering techniques, companies can identify the several segments of customers allowing them to target the potential user base. K-Means Clustering Algorithm along with RFM analysis is used in this project to segment the customers. Companies that deploy customer decomposition are under the notion that every customer has different requirements and require a specific marketing effort to address them appropriately. Companies aim to gain a deeper approach of the customer they are targeting. Therefore, their aim has to be specific and should be tailored to address the requirements of each and every individual customer. Furthermore, through the data collected, companies can gain a deeper understanding of customer preferences as well as the requirements for discovering valuable segments that would reap them maximum profit. This way, they can strategize their marketing techniques more efficiently and minimize the possibility of risk to their investment.

1.2. Problem Description

Customer decomposition is essential in today's market as the market is growing at a very high rate and so are the customers. The importance of customer decomposition includes the ability to modify marketing strategies to suit each customer segment and product identification with respect to each customer segment and product offerings are managed, identifying a valuable customer base, and finding solutions.

2 LITERATURE REVIEW

2.1. General Introduction

Literature research is an important activity that we must do in gathering information about a certain topic. It will help us get the necessary information or ideas for work. The following paragraphs discuss related work and issues in Customer Segmentation Analysis Using Machine Learning Algorithm.

2.2. Literature Survey

In the paper titled - "Benefit-based consumer segmentation and performance evaluation of clustering approaches: evidence for data-driven decision making” [1], the authors Arunachalam and Kumar has described the process of clustering. Clustering is the process of dividing a set of physical or abstract objects into groups of similar objects. The K-means algorithm, as one of the most popular clustering algorithms, was first used by Macqueen in 1967 and has been widely used in various fields including data mining, statistical data analysis, and other business applications. A straight-to-the-science article by Deepak Arunachalam and Niraj Kumar shows that one of the main applications of K-means is customer segmentation. The K-means algorithm is widely used to effectively identify valuable customers and develop relevant marketing strategies.

The paper titled – “A hybrid soft computing approach based on clustering, rule mining and decision tree analysis for customer segmentation problem” [2], the author Khalili Damghani investigated that after analyzing and performing K-means, it is noticed a similar relationship between customers and the results suggest that, that the model they designed using k-means is an ideal way to segment the customer and analyze the value of the customer in the market. This proposed method is used to predict new inputs and these inputs become necessary for companies to establish business relationships with customers. This is based on a method like customer clustering and other methods like rule extraction and decision tree. A prediction of the customer's future transactions can be made based on his previous purchase history. Methods such as decision-making and feature selection are used for filtering.

The paper titled - “Retail Analytics: Segmenting Customer Visits Using Market Basket Data” [3], the author Anastasia Griva explored that segmenting customers based on their store visits is done using data analytics. The method used can be feature selection where the input can be taken as product details and the output can be considered as different customer segments.

The paper titled - "RFM ranking - an effective approach to customer segmentation" [4], the authors Priyadarshini Umamamakeswari, A. Neyaa has done the identification of customers on their previous purchase is done by running an RFM model on the proposed data set and then analyzing it using transaction data that contains transactions based in United Kingdom. The dataset contains 541909 instances, of which only 10,000 samples are used. The attributes of this dataset are Invoice Number, Stock Code, Product Description, Product Quantity, Date, Customer ID, Country, Unit Price.

The paper titled – "Segmentation of tele communications customers based on cluster analysis. The K means" [5], the author Jayant prescribed the approach to customer classification states that in customer segmentation, the division of customers into several different groups where each customer of one group has similar interests from a marketing perspective.

S.N	PAPER TITTLE & PUBLICATION DETAILS	NAME OF THE AUTHORS	TECHNICAL IDEAS / ALGORITHMS USED IN THE PAPER & ADVANTAGES	SHORTFALLS/DISADVANTAGES & SOLUTION PROVIDED BY THE PROPOSED SYSTEM
1	Benefit-based consumer segmentation and performance evaluation of clustering approaches: evidence for data-driven decision making	Arunachalam and Kumar	K-Means Clustering Algorithm	
2				
3				
4				
5				

3 REQUIREMENTS

3.1. FUNCTIONAL REQUIREMENTS:

Functional requirements are basically the requirements stated by the user which one can see directly in the final product.

- User interface for companies to decompose its customers .
- Data entered into the system.
- System must perform the work-flow.
- Outputs given by the model and user interface.

3.2. NON-FUNCTIONAL REQUIREMENTS:

Non- Functional requirements for the proposed project are:

- Reliability -System should be Reliable
- Usability -System Should be user Friendly
- Performance -System Should be able to cluster customers groups faster
- Supportability -System should be easily updatable for future enhancement

3.3. SOFTWARE REQUIREMENTS:

The software requirements for the proposed project are:

1. Operating System: Windows 8 or above
2. IDE: Visual Studio code

3.4. HARDWARE REQUIREMENTS:

The hardware requirements for the proposed project are:

1. Processor: Intel i5 Core Processor or above
2. RAM: 6GB or above

4 PROJECT DESIGN

4.1. Project Design

Project design is one of the important phases of software or system development. Project design can be defined as a method of defining the various modules required for a software or system to fulfil all requirements.

4.1.1. Architecture of the Proposed System

Figure 5.1 shows the overall design of the proposed system, which is shown in the figure. This includes the process of machine learning which includes data collection, data cleaning, processed data set.

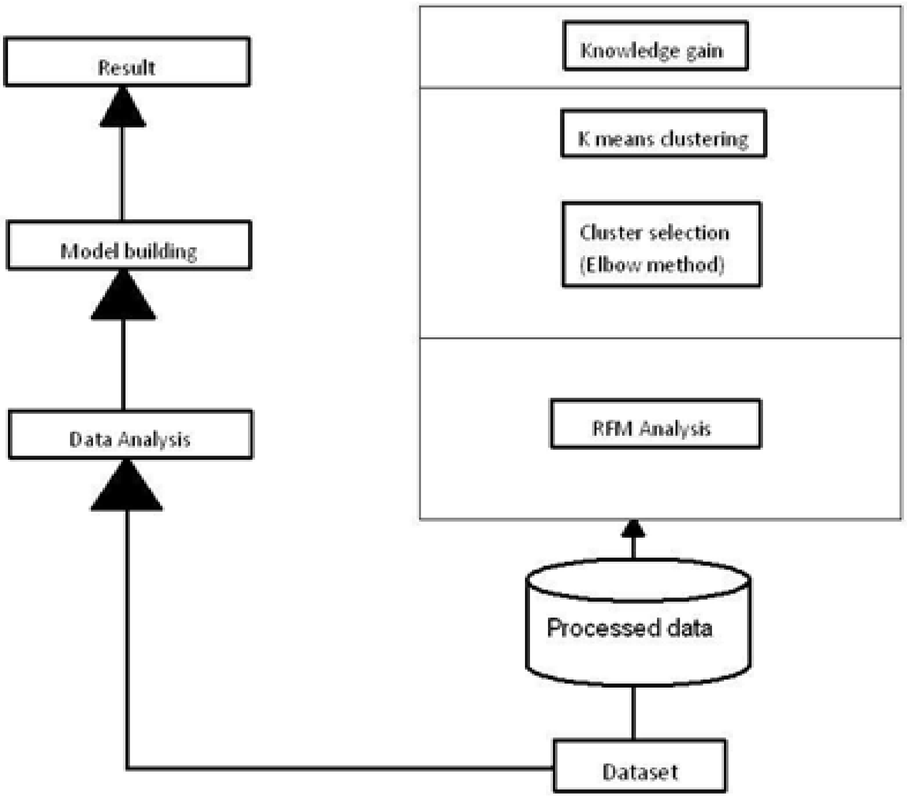


Figure 4.1 Architecture design for the proposed method

5 PROJECT IMPLEMENTATION

5.1. Pseudo code

1. Load the data files
2. Trim the datasets
3. Data analysis
4. Create an RFM table for the given data file
5. Choose the number of clusters using the elbow method
6. Use K stands for Algorithm

   INPUT:

   D= {dl, d2. . . , dn} // set of n data items K // Number of clusters required

   EXIT: STEP:

   Set to clusters

   1. Arbitrarily select K data items from D as the initial centroid.
   2. Repeat

   Assign each item of d1 to the cluster that has the closest centroid.

   Calculate the new mean for each cluster, until the convergence criteria are met.

7. Display the clustered graph.

5.2. Implementation Code

```
import pandas as pd import matplotlib. pyplot
as plt import numpy as np df = pd. read_excel
('Online Retail.xlsx') df = df[df[' Customer ID'].
notna()]
df_fix = df. sample(10000, random_state = 42)
df_fix. shape
df_fix. head()
# Convert to show date only from datetime import
datetime df_fix["Invoice Date"] =
df_fix ["Invoice Date"]. dt. date
# Create Total Sum column df_fix["TotalSum"] =
df_fix["Quantity"] * df_fix["UnitPrice"]
# Create date variable that records recency import
datetime
snapshot_date =max (df_fix.InvoiceDate) + date
time. timedelta(days=1)
# Aggregate data by each customer

customers = df_fix.groupby (['Customer ID']).
Agg ({ 'InvoiceDate': lambda x: (snapshot_date -
x.max()).days, 'InvoiceNo': 'count', 'TotalSum':
'sum'})

# Rename columns
customers. rename(columns = {'Invoice Date':
'Recency', 'InvoiceNo': 'Frequency', 'TotalSum':
'Monetary Value'}, inplace=True) from scipy
import stats

customers_fix = pd.DataFrame() customers_fix
["Recency"] =
stats.boxcox(customers['Recency'])[0] customers
_fix["Frequency"] =
stats.boxcox(customers['Frequency'])[0]
customers_fix["MonetaryValue"] =
pd.Series(np.cbrt(customers['MonetaryValue'])).v
alues customers_fix.tail()
# Import library
```

```
from sklearn.preprocessing import StandardScaler
# Initialize the Object scaler = StandardScaler()
# Fit and Transform the Data scaler.fit
(customers_fix)
customers_normalized =
scaler.transform(customers_fix)

# Assert that it has mean 0 and variance 1

print (customers_normalized.mean(axis = 0).
round(2)) # [0. -0. 0.]

print (customers_normalized.std(axis =
0).round(2)) # [1. 1. 1.] import seaborn as sns
from sklearn.cluster import KMeans sse = {}
for k in range(1, 11):

kmeans = KMeans(n_clusters=k, random_ state=
42) kmeans.fit(customers_normalized) sse[k] =
kmeans.inertia_ # SSE to closest cluster centroid
plt.title('The Elbow Method') plt.xlabel('k') plt.y
label('SSE') sns.pointplot(x=list(sse.keys()), y=lis
t(sse.values()))

plt.show() model = K Means (n_clusters=3,
random_state= 42) model.fit (customers _norm
alized) model. labels_.shape
customers["Cluster"] = model.labels_ customers.
groupby('Cluster').agg({ 'Recency':'mean',

'Frequency':'mean',

'MonetaryValue': ['mean', 'count']}).round(2)

df_normalized = pd.DataFrame (customers_
normalized, columns= ['Recency', 'Frequency',
'MonetaryValue'])df_normalized['ID'] =
customers.index df_normalized['Cluster'] =
model.labels_# Melt The Data

df_nor_melt = pd.melt (df_normalized.reset_
index(), id_vars=['ID', 'Cluster'], value_vars=
['Recency','Frequency','MonetaryValue'],var_name
='Attribute', value_name='Value')df_nor_melt.
head() # Visualize it

sns.lineplot('Attribute', 'Value', hue='Cluster',
data=df_nor_melt)
```

Html home page code:

```
<!doctype html>
{% load static %}
<html lang="en">
<head>
<! -- Required meta tags -->
<meta charset="utf-8">
<meta name="viewport" content="width=device-
width, initial-scale=1">
<link rel="stylesheet" href="{% static 'css/ style.
css' %}">
<link rel="stylesheet" href="{% static 'css/ cards
tyle.css' %}">
</head>
<body>
<section id="header">
<div class="header container">
<div class="nav-bar">
<div class="brand">
<a href="#hero">
<h1><span>M</span>all
<span>C</span>ustomer
<span>S</span>egmentation</h1>
</a>
</div>
<div class="nav-list">
<div class="hamburger">
<div class="bar"></div>
</div>
<ul>
<li><a href="{% url 'home' %}" data-
after="Service">Home</a></li>
<!-- <li><a href="{% url 'about Us' %}" data-
after="Service">About</a></li>
</ul>
</div>
</div>
</div>
</section>
<section id="hero">
<div class="hero container">

<div style="margin-top:50px">
<h1>Connected Data to <span></span></h1>

<h1>Leverage Product-Affinity Driven
<span></span></h1>
<h1>Customer Segmentation
<span></span></h1>
<!-- <a href="Mall_Customers.csv" type=
"button" class="cta">Dataset</a> -->
<a href="{% url 'card' %}" type="button"
class="cta">Explore</a>

</div>
</div>
</section>
<script src="{% static 'projectApp/js/app.js' %}
"></script>
</body>
</html>
```

Html code of " list of malls" page:

```
<!doctype html>
{% load static %}
```

```
<html lang="en">
<head>
<meta charset="utf-8">
<meta name="viewport" content="width=device-width, initial-scale=1">
<link rel="stylesheet" href="{% static 'css/ style.css' %}">
<link rel="stylesheet" href="{% static 'css/ cardstyle.css'  %}">
</head>
<body>
<section id="header">
<div class="header container">
<div class="nav-bar">
<div class="brand">
<a href="#hero">
<h1><span>M</span>all
<span>C</span>ustomer
<span>S</span>egmentation</h1>
</a>
</div>
<div class="nav-list">
<div class="hamburger"> <div class="bar"></div>
</div>
<ul>
<li><a href="{% url 'home' %}" data-after= "Service">Home</a></li>
</ul>
</div>
</div>
</div>
</section>
<head>
<link rel="stylesheet" href="{% static 'css/ cards tyle.css' %}">
</head>
<div class="container">
<div class="grid">
<div class="card">

<input id="card1" type="checkbox"><label class="tgl-btn"
for="card1"><span></span></label>
<div class="tgl-view">
<div class="card-image">
<img src="{% static 'images/mall1.jpg' %} " alt="mall 1">
</div>
<h2 class="card-title"><a href="{% url 'mall1' %}" style="font-weight: bold;">A-Z Supermarket</a></h2>
<p class="card-detail">This dataset contians 541909 instances. This dataset is collected from UCI machine learning repository. <br></p>
</div>
</div>
<div class="card">
<input id="card2" type="checkbox">
<label class="tgl-btn" for="card2"><span></span></label>
<div class="tgl-view">
<div class="card-image">
<img src="{% static 'images/mall 2.jpg' %} "alt="mall 2">
</div> <h2 class="card-title"><a href="{% url 'mall2' %}"style="font-weight: bold;">SHOW OFF</a></h2>
<p class="card-detail">This dataset contians 1115 instances. This is collected from online garment shop<br></p>
</div>
</div>
<div class="card">
<input id="card3" type="checkbox">
<label class="tgl-btn" for="card3"><span> </span></label>
<div class="tgl-view">
<div class="card-image">
<img src="{% static 'images/mall 3.jpeg' %} " alt="mall 3">
</div>
<h2 class="card-title"><a href="{% url 'mall3' %}" style="font-weight: bold;" >JP
ELECTRONICS</a></h2>
<p class="card-detail">This dataset contians 2002 instances.This is collected from online Electronics shop.<br></p>
</div>
</div>
</div>
</div>
</body>
</html>
```

Html code of Mall:

```
<!DOCTYPE html>
<html lang="en">
<head>
{% load static %}
<title>Customer Dataset</title>
<meta charset="utf-8">
<meta name="viewport" content="width=device-width, initial-scale=1">
<linkrel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstrap/4.5.0/css/bootstrap.min.css">
<script
src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>
<script
src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/1.16.0/umd/popper.min.js"></script>
<script
```

```
src="https://maxcdn.bootstrapcdn.com/bootstrap/
4.5.0/js/bootstrap.min.js"></script>
</head>
<body>
<div class="container">
<h2 class="text-center"><u>First 20 rows of data
set of the mall</u></h2><br>
<table class="table table-dark table-striped">
<thead>
<tr>
</tr>
<th>Invoice No.</th>
<th>Stock Code</th>
<th>Description</th>
<th>Quantity</th>
<th>Invoice Date</th>
<th>Unit Price</th>
<th>CustomerID</th>
</thead>
<tbody>
{% if obj1 %}
{% for i in obj1 %}
<tr>
</tr>
<td>{{i.InvoiceNo}}</td>
<td>{{i.StockCode}}</td>
<td>{{i.Description}}</td>
<td>{{i.Quantity}}</td>
<td>{{i.InvoiceDate}}</td>
<td>{{i.UnitPrice}}</td>
<td>{{i.CustomerID}}</td>
</tr>
{% endfor %}
{% endif %}
</tbody>
</table>
</div>
<img style = "margin-left: 150px; border-radius: 10px; border: 8px solid #343a40;" src="{% static 'output/customer_app1.png'%}" alt="" width=" 550" height="550"><br>
<img style = "margin-left: 850px; margin-bottom: 50px; border-radius: 10px; border: 8px solid #343a40" src="{% static 'output/ customer_ app.png'%}" alt="" width="550" height="550">
</body>
</html>
```

6 TESTING

6.1. System Testing

The purpose of testing is to detect errors. Testing is the process of trying to uncover every conceivable bug or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or the finished product. It is the process of exercising software with the intention of ensuring that the software system meets its requirements and user expectations and does not fail in unacceptable ways. Software testing is the process of checking whether the developed system works according to the original goals and requirements.

Table 6.1: System testing

Test Case Number	Input	Stage	Expected Behavior	Observed Behavior	Status P=Pass, F=Fail
1	The Online Retail dataset	Data Analysis	Most important features are fed into the system	As expected	P
2	The Online Garments dataset	Data Analysis and trimming	Most important features are fed into the system	As expected	P
3	The Online Electronics dataset	Data Analysis and trimming	Most important features are fed into the system	As expected	P
4	K-Means Clustering Algorithm	Result Analysis	Clustered Graph is displayed	As expected	P

7 RESULTS AND DISCUSSION

7.1 Result Analysis

Customers are divided into three groups:

A customer in a cluster that has high recency, low frequency, and monetary value indicates that these customers are the least active customers and are referred to as churned customers. A customer in a cluster that has medium recency, medium frequency, and medium monetary value, indicating that these customers are referred to as new customers who have only recently started the products. A customer in a cluster that has low recency, high frequency, and monetary value, indicating that these customers are the most active customers and are referred to as loyal customers.

7.2 Snapshots

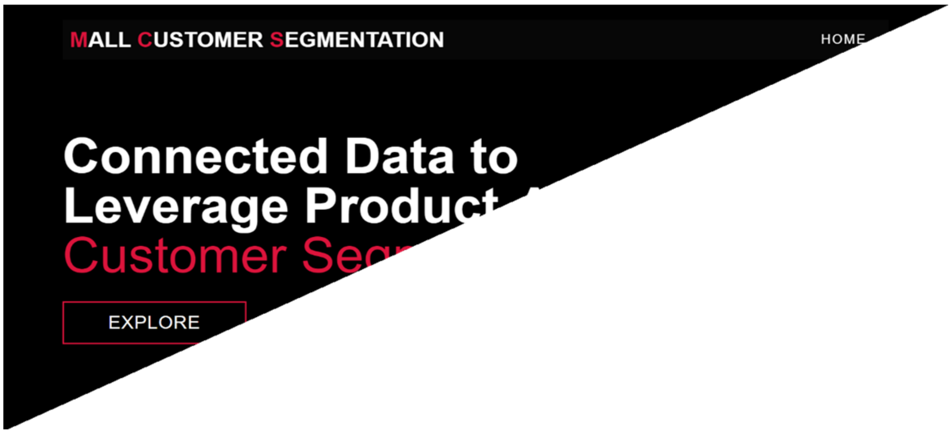


Figure 7.1 Home page

The figure 7.1 is a home page of mall customer segmentation.

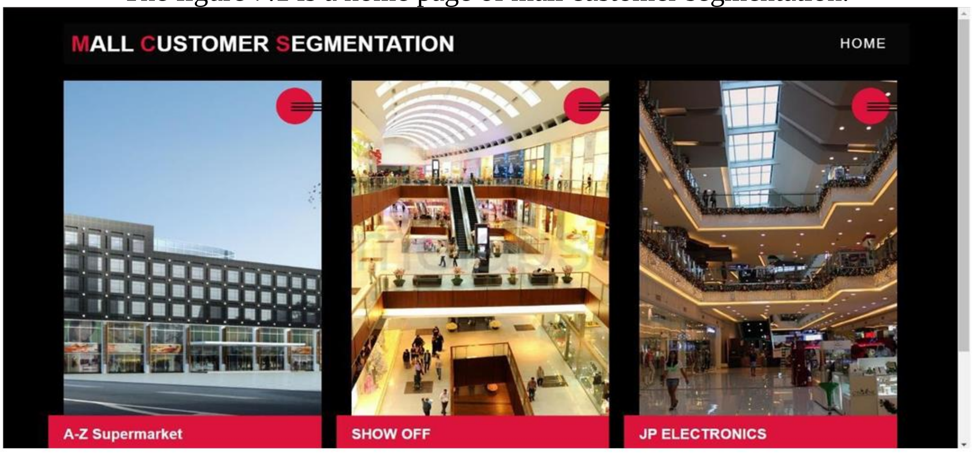


Figure 7.2 List of malls

The figure 7.2 is the page consists of set of malls. There are three different datasets of each mall and they are A-Z supermarket, Show off and JP Electronics.

First 20 rows of data set of the mall

Invoice No.	Stock Code	Description	Quantity	Invoice Date	Unit Price	CustomerID
546837	85014B	RED RETROSPOT UMBRELLA	3	1300320000000	5.95	13650.0
547651	22293	HANGING CHICK GREEN DECORATION	1	1300924800000	1.45	16904.0
568716	22150	3 STRIPEY MICE FELTCRAFT	1	1317168000000	4.13	None
559691	22089	PAPER BUNTING VINTAGE PAISLEY	12	1310342400000	2.95	13089.0
545682	22975	SPACEBOY CHILDRENS EGG CUP	3	1299369600000	1.25	14701.0
540977	22208	WOOD STAMP SET THANK YOU	2	1294790400000	1.66	None
548156	22652	TRAVEL SEWING KIT	2	1301356800000	1.65	14871.0
580366	23489	VINTAGE BELLS GARLAND	4	1322784000000	2.88	None
543660	22288	HANGING METAL RABBIT DECORATION	1	1297382400000	2.46	None
551891	21402	RED EGG SPOON	24	1304553600000	0.12	17429.0

Figure 7.3 Display of first 20 rows of Mall 1 dataset

The figure 7.3 is the display of the mall 1 supermarket datasets. This dataset contains 541909 instances but first 20 rows of dataset is displayed. This dataset is collected from UCI machine learning repository.

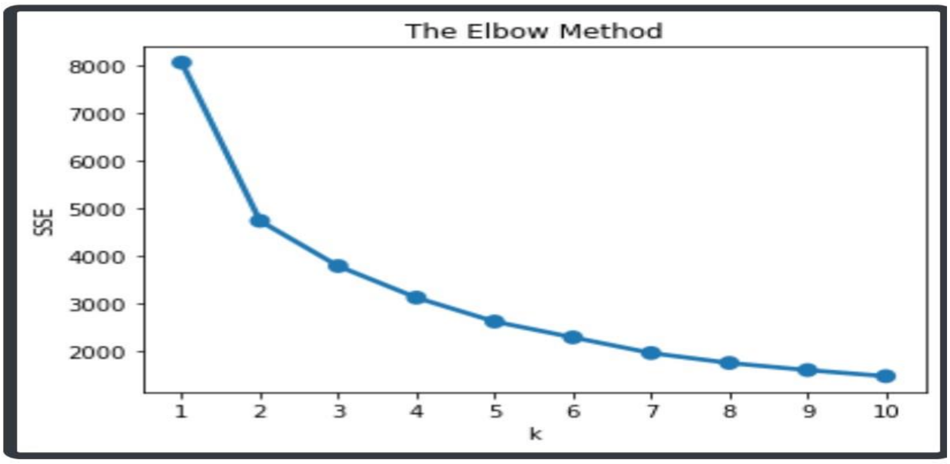
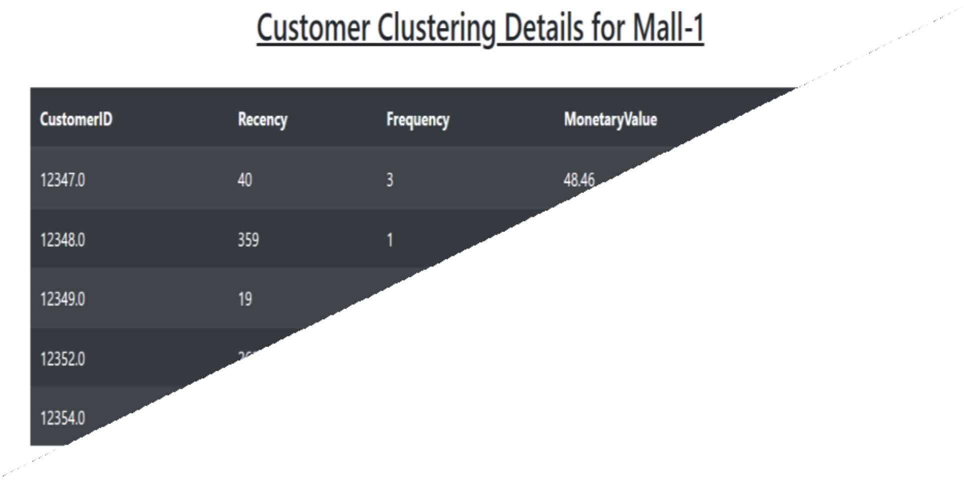


Figure 7.4 Elbow method for Mall 1 dataset

The figure 7.4 is the graphical representation of Elbow method to determine the optimal cluster for Mall 1 dataset. From the observation it is noted that the value of K is 3.



Customer Clustering Details for Mall-1

CustomerID	Recency	Frequency	MonetaryValue
12347.0	40	3	48.46
12348.0	359	1	
12349.0	19		
12352.0			
12354.0			

Figure 7.5 Clustering details for Mall 1 dataset

The figure 7.5 shows the details regarding the clustering in the mall 1 datasets.

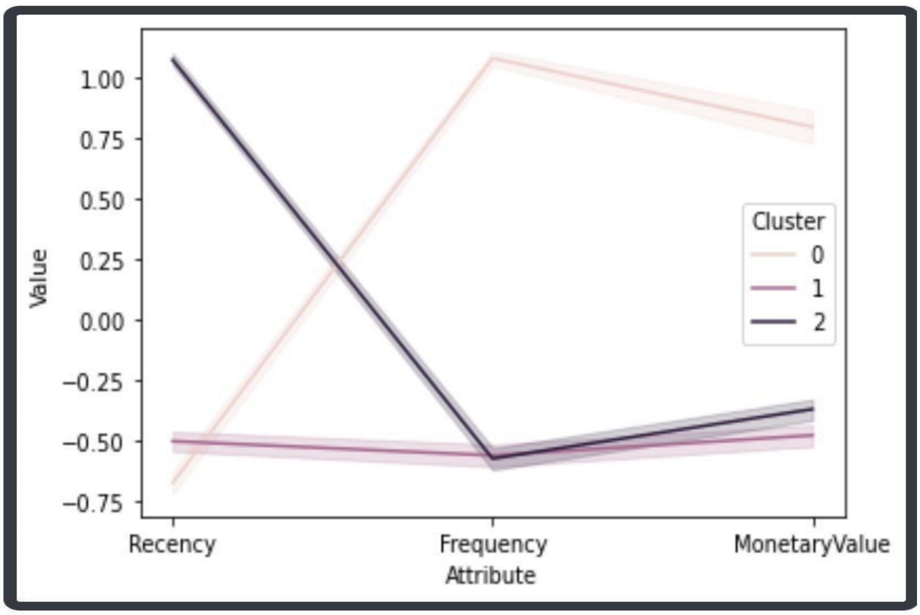


Figure 7.6 Final graph of clusters based of RFM for Mall 1 dataset

The figure 7.6 is the final resultant graph with respect to RFM is displayed.

First 20 rows of data set of the mall

Invoice No.	Stock Code	Description	Quantity	Invoice Date	Unit Pri
574306	22086	Insignia - Fixed TV Wall Mount For Most 40-70" TVs - Black NS-HTVMF0C"	6.0	132027	
574321	71477	Alpine CDESXM145BT Advanced Bluetooth CD / SiriusXM Receiver			
574298	21544	Details About Sony Kd70x690e 70inch 4k Ultra Hd Sma (2017 Model)			
574311	21931	PYLE - 8 Single-Voice-Coil 4-Ohm Amplifier - Black"			
574306	35953	SAMSUNG QN			
574304					

Figure 7.7 Display of first 20 rows of Mall 2 dataset

The figure 7.7 is the display of the mall 2 show off datasets. This dataset contains 1115 instances but first 20 rows of dataset is displayed. This dataset is collected from an online garment shop.

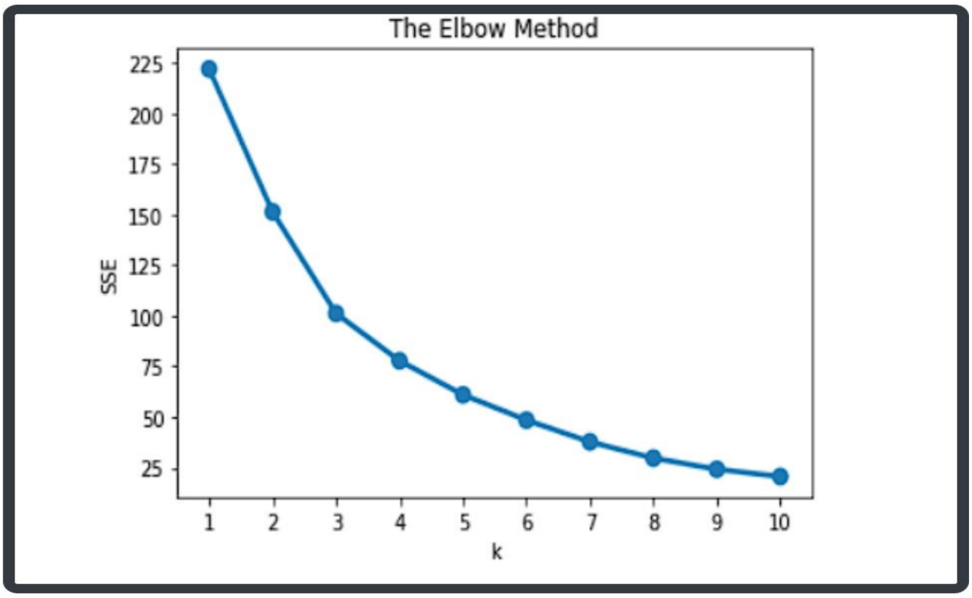


Figure7.8 Elbow method for Mall 2 dataset

The figure 7.8 is the graphical representation of Elbow method to determine the optimal cluster for Mall 2 dataset. From the observation it is noted that the value of K is 3.

Customer Clustering Details for Mall-2

CustomerID	Recency	Frequency	MonetaryValue	Cluster
12431.0	360	121	3214.3	0
12583.0	360	152	6593.66	0
12662.0	360	55	953.1	1
12748.0	360	16	79.2	1
12791.0	360	29	1898.4	1

Figure 7.9 Clustering details for Mall 2 dataset

The figure 7.9 shows the details regarding the clustering in the mall 2 datasets.

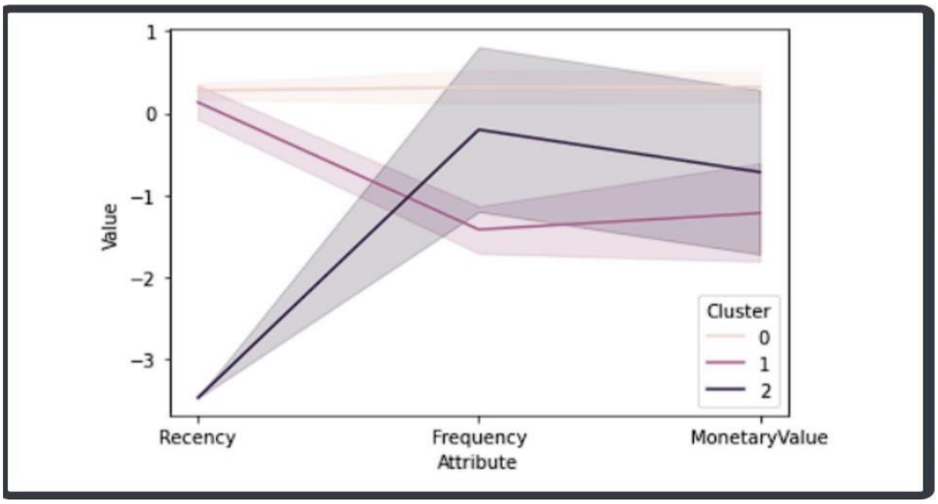


Figure 7.10 Final graph of clusters based of RFM for Mall 2 dataset
The figure 7.10 is the final resultant graph with respect to RFM is displayed.

First 20 rows of data set of the mall

Invoice No.	Stock Code	Description	Quantity	Invoice Date	Unit Price	CustomerID
10012825	72818	SEJ by Nisha Gupta Set of 6 Mustard & Pink Printed Table Placemats	1	1291161600000	0.85	17968.0
10019701	20974	Bitiya by Bhama Girls Mustard Embellished Top with Palazzos	3	1291161600000	0.65	14078.0
10027957	37370	HERE&NOW Women Black Solid Biker Jacket	2	1291161600000	1.25	15012.0
10006003	37444C	ID Men Grey Solid Thong Flip-Flops	1	1291161600000	2.95	15311.0
10017021	85014B	Campus Sutra Men Blue Solid Round Neck T-shirt	3	1291161600000	5.95	14307.0
1001061	84509A	Blackberrys Black Single-Breasted Slim Fit Formal Blazer	2	1291161600000	3.75	15862.0
10017899	85099C	Parx Men Yellow & Off-White Slim Fit Checked Casual Shirt	10	1291161600000	1.95	14688.0
10016715	22041	Vishudh Women Black & Maroon Embroidered Straight Kurta	48	1291161600000	2.1	16210.0
10003809	21533	Homesake Gold-Toned & White Solid Handcrafted Table Lamp	1	1291161600000	4.95	15311.0

Figure 7.11 Display of first 20 rows of Mall 3 dataset\
The figure 7.11 is the display of the mall 3 JP Electronics datasets. This dataset contains 2002 instances but first 20 rows of dataset is displayed. This dataset is collected from an online electronics shop.

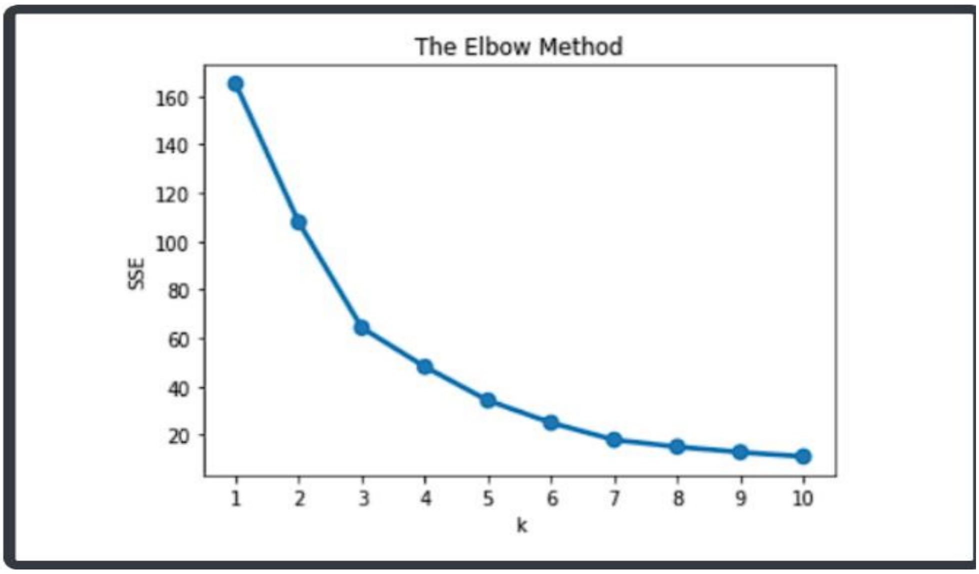


Figure 7.12 Elbow method for Mall 3 dataset
The figure 7.12 is the graphical representation of Elbow method to determine the optimal cluster for Mall 3 dataset. From the observation it is noted that the value of K is 3.

Customer Clustering Details for Mall-3

CustomerID	Recency	Frequency	MonetaryValue	Cluster
12352.0	22	68	1439.1	0
12362.0	21	72	2246.38	0
12397.0	1	289	5995.64	2
12415.0	21	2	-525.46	1
12544.0	22	101	1877.58	0

Figure 7.13 Clustering details for Mall 3 dataset
The figure 7.13 shows the details regarding the clustering in the mall 3 datasets.

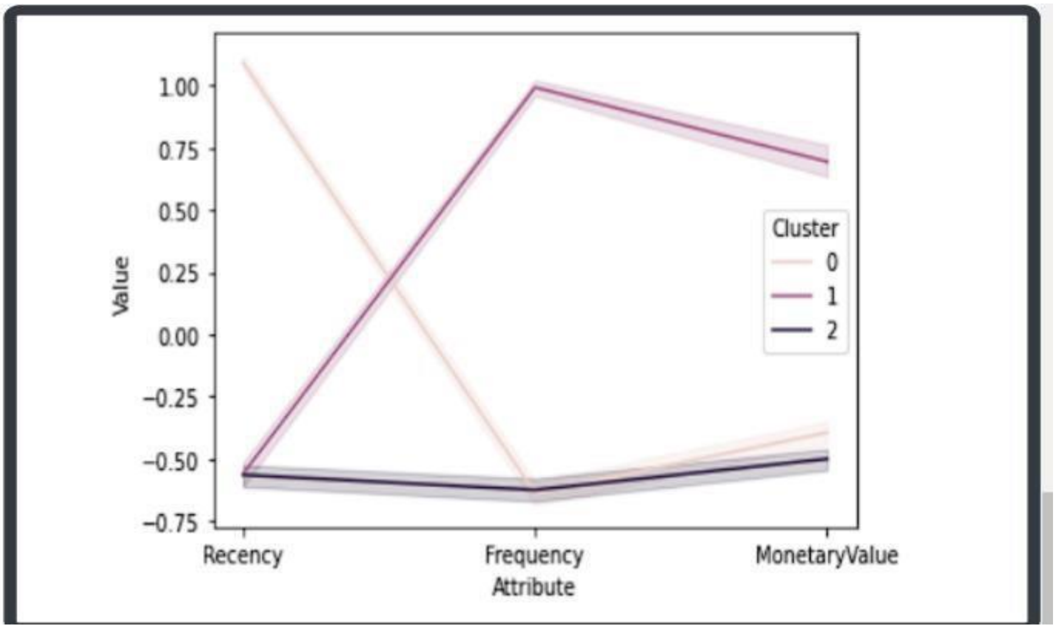


Figure 7.14 Final graph of clusters based of RFM for Mall 3 dataset
The figure 7.14 is the final resultant graph with respect to RFM is displayed.

7.3 Summary

Since loyal customers are the most active, the company can focus more on these customers to plan some schemes to encourage them and improve their marketing strategies. The company also needs to focus on new customers because they too may buy more products in the future.

8 CONCLUSION & FUTURE SCOPE

8.1 Conclusion

Customer decomposition can have a positive impact on business if done correctly. Customer decomposition involves grouping customers together to better meet their needs while maintaining economies of scale. Therefore, ML techniques are used for effective customer decomposition. The K-Means clustering algorithm is used in this project to decide how to relate to customers in each segment to maximize the value of each customer to the business. Customer decomposition can have a positive impact on business if done correctly. Customer decomposition involves grouping customers together to better meet their needs while maintaining economies of scale. Therefore, ML techniques are used for effective customer decomposition.

The K-Means clustering algorithm is used in this project to decide how to relate to customers in each segment to maximize the value of each customer to the business. K Means analysis is extremely beneficial because it can be used for large datasets unlike other clustering algorithms. Also, with the generation of a large number of customer data given that the number of businesses is rapidly increasing and so is the number of customers obtaining services from the business it becomes extremely difficult to keep track of what services to provide to whom and what changes to be made in a business this is where this analysis comes into use.

RFM Analysis is a marketing framework that is used to understand and analyze customer behavior based on the above three factors Recency, Frequency and Monetary. The RFM Analysis will help the businesses to segment their customer base into different homogenous groups so that they can engage with each group with different targeted marketing strategies. Instead of examining the whole client base, it's smarter to segment them into clusters, comprehend the qualities of each gathering, and engage in them with relevant deals. One of the most famous, simple to-utilize and successful division strategies to empower advertisers to break down client behavior is RFM with K-Means decomposition.

8.2 Scope for Future Work

It is very important for a company to know its valuable customers. In the future, the model can be trained on different attributes for better decomposition. The user interface can be made more interactive by adding features such as uploading a dataset and visualizing the graph in a better way. It is very important for a company to know its valuable customers. In the future, the model can be trained on different attributes for better decomposition. The user interface can be made more interactive by adding features such as uploading a dataset and visualizing the graph in a better way. Individual analysis of each attribute can be done for better marketing strategies and analyzing the performance of those groups can improve the marketing, sales, and customer service efforts.

REFERENCES

1. Arunachalam and Kumar, 2018 Benefit-based consumer segmentation and performance evaluation of clustering approaches: evidence of data-driven decision-making. Expert Syst. Appl., 111 (2018), pp. 11-34.
2. T. Jayant Tikmani, Sudhanshu Tiwari, Sujata Khedkar "Telecom customer segmentation based on cluster analysis. An Approach to Customer Classification using k-means", IJIRCCE, Year: 2018.
3. Kaveh Khalili -D-Damghani, Farshid Abdi, Shaghayegh Abolmakarem "Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem". Appl. Soft Compute., 73 (2018), pp. 816-828.
4. Anastasia Griva, et al. "Retail business analytics: customer visit segmentation using market basket data". Expert Syst. Appl., 100 (2018), pp. 1-16.
5. A. Christy, A. Joy, L. Priyatharsini Umamakeswari, A. Neyaa "RFM ranking–an effective approach to customer segmentation "J. King Saud Univ.-Comput. Inf. Sci. (2018).
6. D. Aloise, A. Deshpande, P. Hansen, and P. Popat, "The Basis of Market Segmentation" Euclidean sum-of-squares clustering," Machine Learning, vol. 75, pp.245-249, 2009.
7. S. Dasgupta and Y. Freund, "Random Trees for Vector Quantization," IEEE Trans. on Information Theory, vol. 55, pp. 3229-3242, 2009.
8. Li, Youguo & Wu, Haiyan. (2012). A Clustering Method Based on K-Means Algorithm. Physics Procedia. 25. 1104-1109. 10.1016/j.phpro.2012.03.206.
9. Yei Wei, Jo-Ting & Lin, Shih-Yen & Wu, Hsin-Hung. (2010). African Journal of Business Management December Special Review. 4. 4199-4206.
10. Minghua Han. "Customer segmentation model based on retail consumer behavior analysis". In: Intelligent Information Technology Application Workshops, 2008. IITAW'08. International Symposium on. IEEE. 2008, pp. 914–917.