**Business Analytics with R Project BUAN 6356**

**Ecommerce Customer Segmentation**

Shape

Description automatically generated

In partial fulfillment of the Master of Science degree

**Under the supervision of**

**Prof. Shujing Sun**

**Naveen Jindal School Of Management**

**Group Members:**

1.Haritha Nallam

2.Akanksha Sharma

3.Varun Vinod

4.Pravin Adep

5.Natasha Mehta

**Business Context**

Ecommerce retailers are operated for selling their products to customers through their websites. The most difficult task for retailers is understanding the customers and identifying their purchase behavior and preferences. Ecommerce retailers’ understanding of the characteristics of their consumer base is crucial for improving their targets and to increase the company shares in the market. They face

intense competition with other online retailers, so to remain competitive in the market, ecommerce retailers need to continuously improve their offerings.

**Problem Statement**

Customer segmentation is to enable businesses to create targeted and personalized marketing campaigns that are tailored to each specific group. This approach can enhance customer satisfaction, increase loyalty and ultimately lead to improved business performance. Also business has to focus on developing their sales strategies and marketing ways in order to acquire new customers and retain existing ones.

**Data Description**

**Data Source**

We extracted the data from the Kaggle.

Source Link: [https://archive.ics.uci.edu/ml/datasets/online+retail#](https://archive.ics.uci.edu/ml/datasets/online+retail)

**Data Challenges**

* Due to insufficient exposure to the marketing domain knowledge, we have encountered

difficulties in understanding certain variables and linking them to correlated variables.

* Data size was large – handling huge amount of data is a resource intensive, time consuming

Process.

**Structure of the Data**

The Online retail Dataset consists of 440,374 observations of 8 variables. The

variables InvoiceNo, StockCode, Description, Country and CustomerID are nominal.

Quantity, UnitPrice and InvoiceDate are numeric.

**Missing/Abnormal Data**

* We found NA in Customer ID -135,080 rows were removed.
* Description of the product had abnormal data instead of product name.

**Data Cleaning**

**Deletion:**

* Out of 440,374 records, the missing record in CUSTOMERID variable - 135,080 deleted as they cannot be imputed or calculated. This step is mandatory to ensure accuracy.

**Imputation:**

* For 1470 missing values for Product Descriptions, instead of deleting - we have imputed values for descriptions by identifying the same StockCodes from other documents.
* For Unitprice “zero” values, we have imputed 2352 values, by taking the mean value for that particular product which counts to nearly 0.5% of data.
* We calculated a column, amount the product of Quantity and Unitprice for our analysis purpose and better representation.

Chart, bar chart

Description automatically generated**Data Analysis**

* We plotted a graph, to see the most

popular product amongst our

customers. Item verses Quantity

purchased. And we found that

MEDIUM CERAMIC TOP STORAGE JAR

has the highest total quantity of 80,000.

* Above said product is popular, and we

should make sure that our inventory is well

managed for top 5 products (in-stock)

Chart, bar chart

Description automatically generated

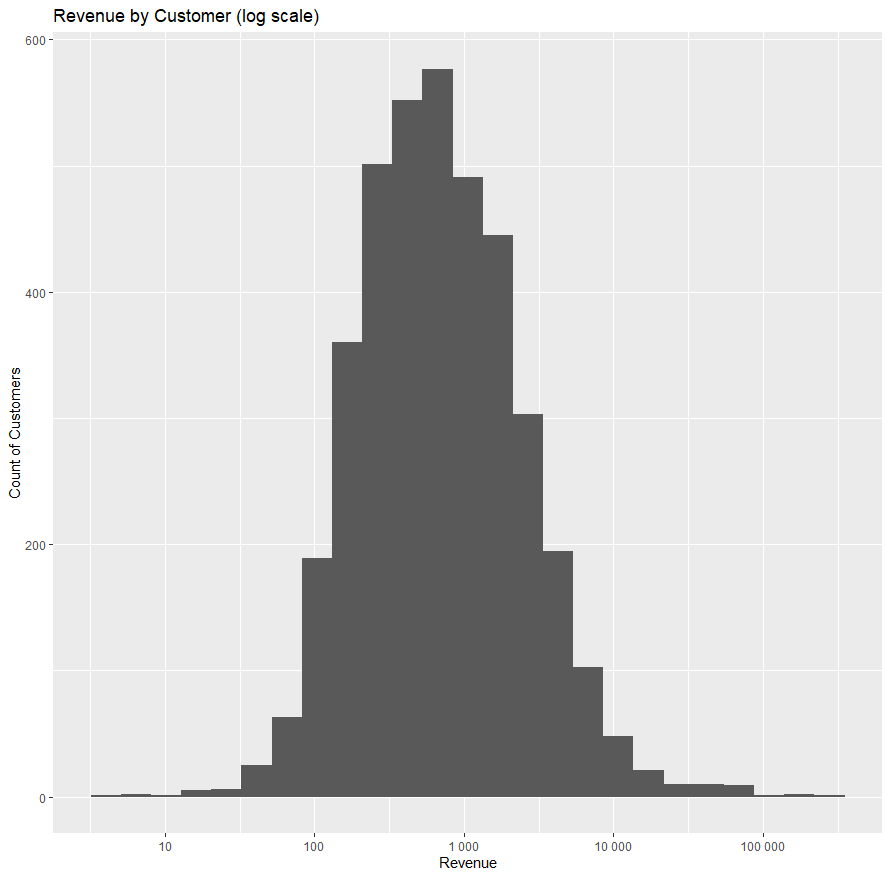
* After analyzing the revenue data for each day of the week, we have determined that Thursday generates the highest revenue with a total of USD 1,500,000.
* In order to increase revenue for other days, we have to attract the customers by offering with discounts on that particular days.

Chart

Description automatically generated

Here we are analyzing the day wise transactions with the box plot and we identified that the number of transactions are high on Sunday.

As Sunday is leisure day, customers show interest in shopping. So, we need to ensure that our website can handle the Sunday traffic.



* Based on this histogram graph, we can conclude that the

distribution is nearly normal and that a larger number of

customers are inclined towards purchasing products of

less price.

* As a result, we must increase our marketing efforts for

high-value items by offering such as gifts or vouchers

that can be redeemed on our website.

**Clustering Customers:**

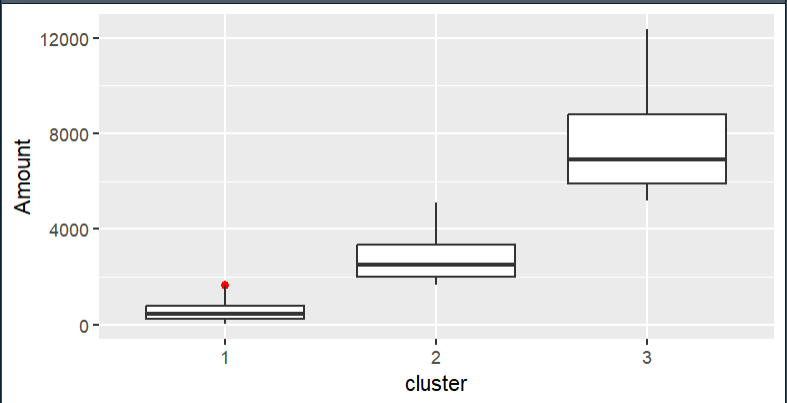
* Using elbow and silhouette plot, we figured that the best fit could be three clusters.
* There are 2943 customers in cluster 1 which is almost 80% of the data.

Chart, line chart

Description automatically generatedGraphical user interface, text, application

Description automatically generated

**Cluster Interpretation:**

Chart, box and whisker chart

Description automatically generatedChart

Description automatically generated

**Cluster 1 – Active customers**

Customers in this cluster are most active on our website with most recent purchase records. However, they have least number of transactions and thus generated least revenue for our business. Based on this kind of purchasing behavior we are segmenting these customers as active customers who are more price sensitive.

**Cluster 2 – Impulse customers**

Looking at their purchase history, number of transactions we can infer these are impulse customers who make purchases on the spot without any prior planning.

**Cluster 3 – Prime Customers**

Customers in this cluster seem to be occasional buyers making purchases based on their specific need or requirement. These are the Customers who are generating highest revenue for our business and also have highest number of transactions.

**Marketing Strategies:**

**Cluster-1,** we observe that they are active but price sensitive therefore, we can Encourage them to make more frequent purchases through our targeted promotions or we can attract them using special offers or discount.

**Cluster-2**, we observe these customers are impulse buyers therefore using social media effectively we can target these customers into buying a product of their interest. Targeted ads about products or services in which the user has interest can also prove effective with such customers. We can also create sense of urgency by offering limited time promotions or discount which will encourage impulse purchases.

**Cluster-3,** These are the customers who are generating highest revenue we should try to convert these customers into our loyal customers.we can have personalized communications to talk about their specific needs and preferences via targeted email campaigns or other form of direct marketing. We can also offer rewards to incentivize prime customers to repeat their purchase and become more loyal to us.