

CHAMPO-CARPET CASE STUDY



Done by:

The Night's Watch



SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Team Members:

Priyanshu Kundu (Team Lead)

Jahnvi Darbhamulla

Kruthi M

Brijesh Kumar

Jerin Joe

PROJECT SUMMARY

We are proposing a recommendation system that can provide the best path towards increasing the conversion rate. Using the Unsupervised Machine Learning model we are able to categorize groups based on country, preferences and which items lead to higher purchases. On understanding these clusters we can identify groups of interest which can increase the number of purchases from customers who sample certain types of products. Using the Champo-Carpet dataset, we analyze clusters using k-means clustering.

We have confirmed 6 different clusters which are based on different countries and their customer needs based on item characteristics.

Therefore can confirm how to improve conversion rates and overall increase profit margin. We will use the random forest model because it has the highest accuracy. To increase its reach and obtain a good understanding of client needs from various geographic locations, we will launch it in an online market and promote third party signup bypasses.



Pc # : M-61586
Quality : Hand Knotted
Design : Ikat
Color : Black
Content : Jute

TABLE OF CONTENTS

| | |
|---|-----------|
| Introduction and Problem Definition..... | 2 |
| Objective of the project..... | 4 |
| Methodology..... | 4 |
| Main results..... | 9 |
| Conclusions | 10 |
| Appendices..... | 11 |
| References..... | 12 |

CHAPTERS

Problem Statement:

There is a need to increase revenue by having a higher conversion rate from samples to purchased goods. Based on the customer demand and other features we need to create a recommendation system which makes the order conversion rate better by using supervised learning and unsupervised learning models.

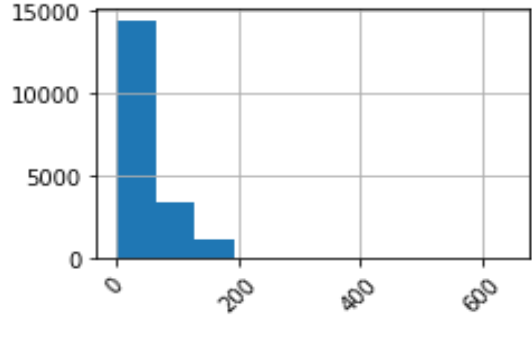
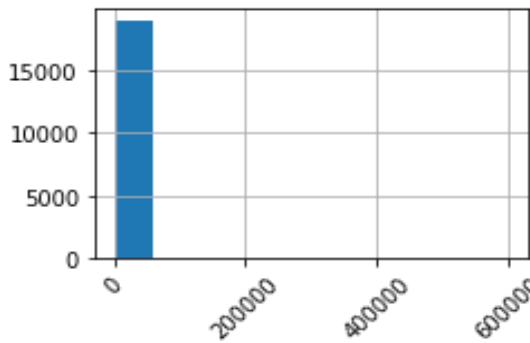
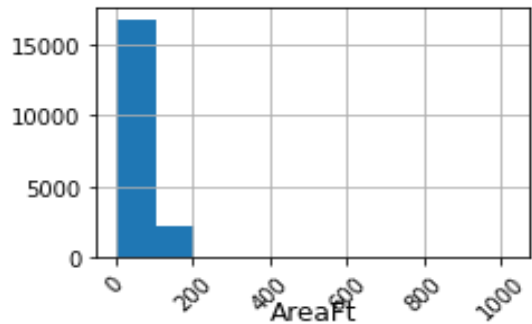
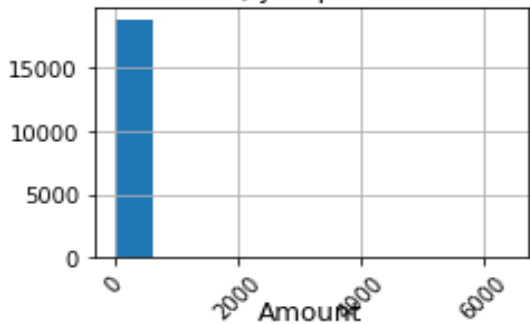
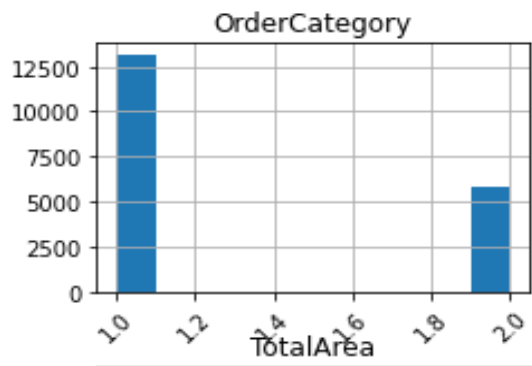
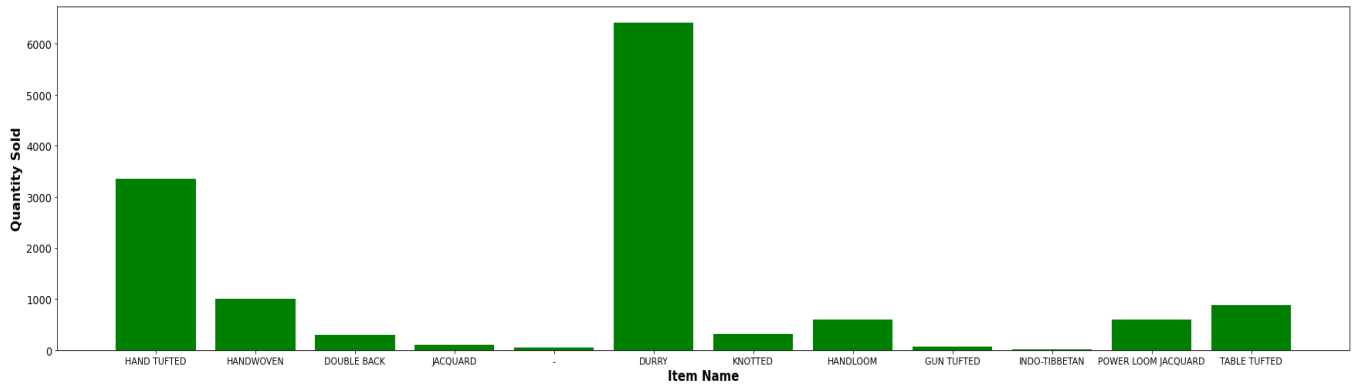
Approach:

We provide a key insight using EDA(Exploratory Data Analysis) with the help of data visualization. For the Exploratory Data analysis we have found that there are no outliers in OrderType, OrderCategory, CustomerCode, CountryName, Custorderdate, UnitName whereas there are outliers in CustomerOrderNo, DataSample.

* **We have to convert :**

1. Converting values of categorical with character into factor format
2. Converting values of categorical with posix format into date format

* There are a lot of NA Values and since '0' has occurred most no of times we can consider replacing NA values with it



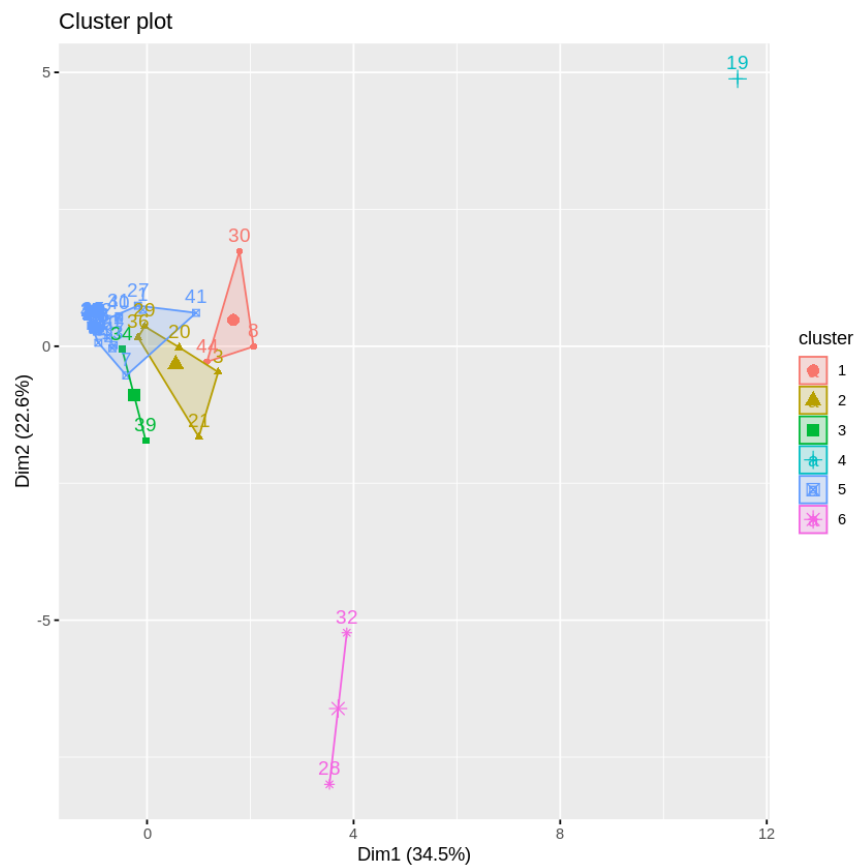
We also develop ML models to help identify features that contribute towards the conversion of samples sent to the customer.

Analysis Attributes:

- CustomerCode
- QtyRequired
- itemName
- ShapeName
- AreaFt
- OrderConversion

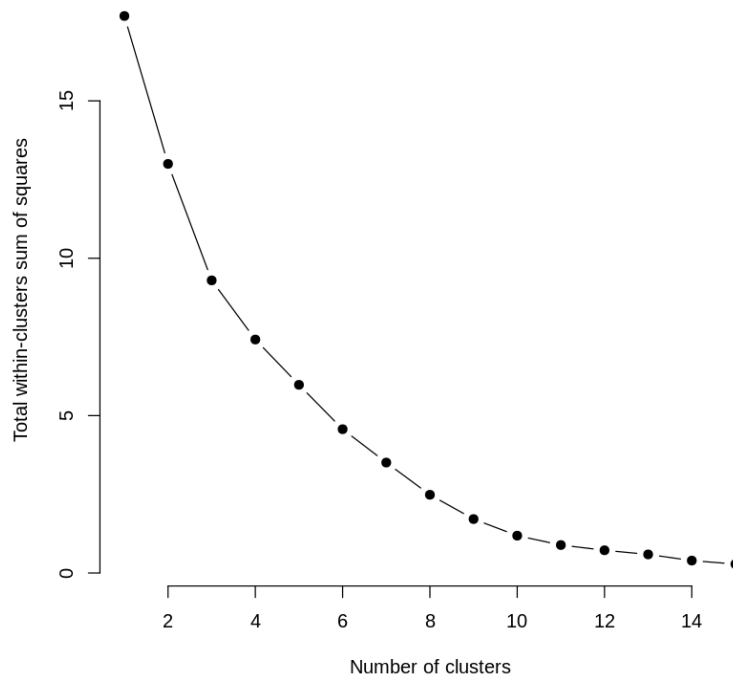
Random forest approach gives us an accuracy of 91.4% with a split dataset of 80:20 train_test split.

We also try the unsupervised clustering model (K-Means) and have segmented the Champto-Carpets customer into 6 categories.



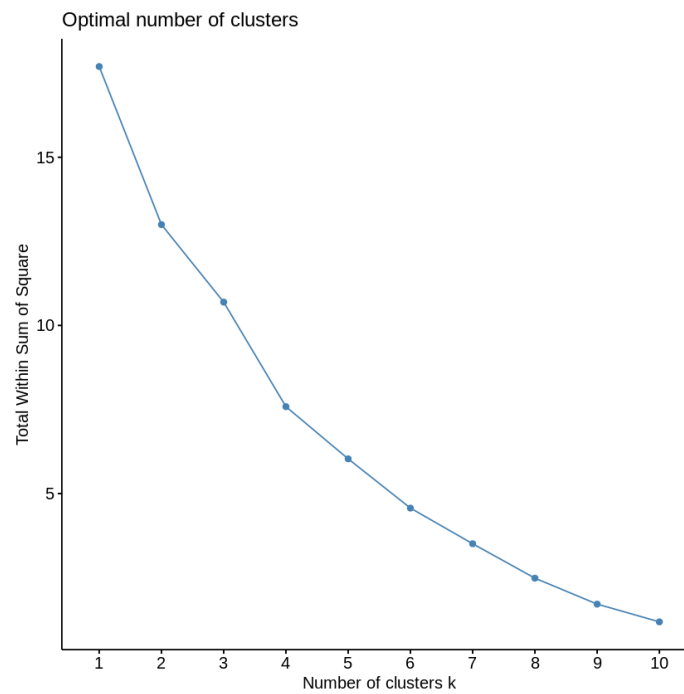
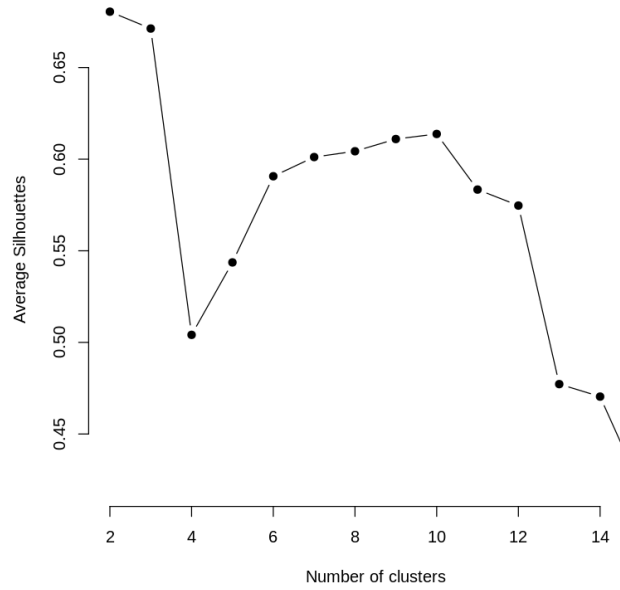
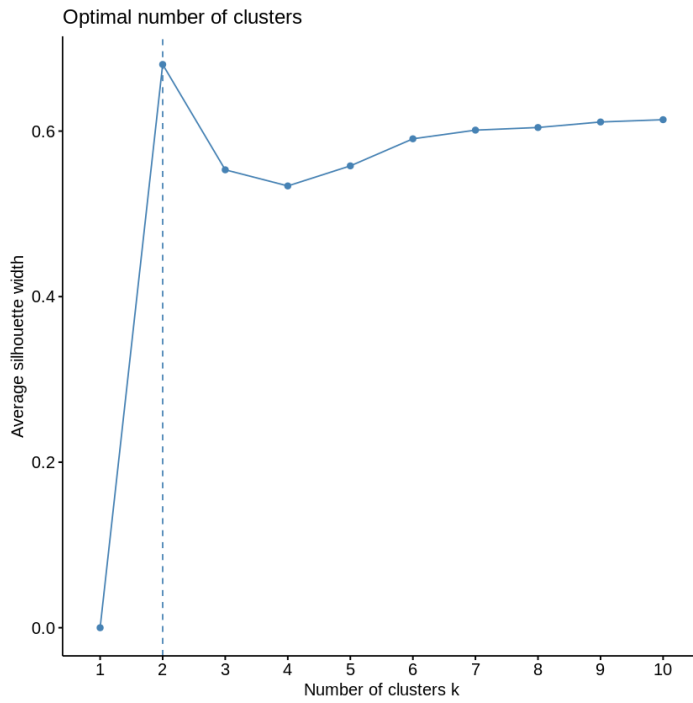
Identifying similar customers and sending samples accordingly, in order to ensure order conversion after sample testing.

| Cluster | Country | Item | Characteristics |
|-----------|-------------------|--------------------------|---|
| Cluster 1 | USA, Israel | Hand tufted, Durry | Darker shades |
| Cluster 2 | Brazil | Double back, Jacquard | Lighter shades of Homer and tikki designs |
| Cluster 3 | USA | Hand tufted, Double back | Darker shades of Double back |
| Cluster 4 | Australia, Brazil | Double back, Knotted | Multicolored shades |
| Cluster 5 | Canada | Double back, Hand tufted | Rectangular shapes |
| Cluster 6 | Australia | Hand tufted | Bombay Print designs |



Based on the results from K-means clustering optimal number of clusters is six.

Graphs - Data Visualisation



Cluster Characteristics:

Cluster 1:

- The highest revenue generating Customers from this cluster belong to USA,UK preferring Hand tufted , Durry, Hand Woven , Double back items.
- Durry items are popular in the USA and Israel with higher preference in embroidered designs.

Cluster 2:

- The highest revenue generating Customers from this cluster belong to USA,Brazil preferring Durry, Double back ,Jacquard, Knotted items.
- Double Back items are popular in USA

Cluster 3:

- Small quantity orders are in majority in the USA.
- Double back and Hand tufted are popular

Cluster 4:

- Australia and Brazil customers are the highest revenue generator with the preference of hand tufted,double backed , knotted items from the cluster.

Cluster 5:

- Majority of customers are from Canada and place orders in small quantities.
- Popular items here are the Double back, Hand tufted, hand woven strictly in rectangular shapes are preferred.

Cluster 6:

- Highest revenue generating customers are from Australia with higher preference of hand tufted items of Bombay print design from this cluster.

Conclusion:

Each model serves a different purpose and is important for further decision making. Using the K-means clustering model, companies can gain knowledge about the distribution of their customers and it will help companies to form better strategies targeting different segments. Here, 6 clusters are formed on the basis of attributes like- country name, item name, color, shape, etc.

The rules generated using the recommendation system can help companies to gain new customers. Using the above cluster content we are able to make a recommendation system that can state which products are the highest sellers and provide a path to increase conversion rate. Some strategies to be applied are:

- Customers frequently buying can recommend samples based on existing orders.
- Recommendation can change according to the countries as customers from different countries show different preferences in carpet type and design.

APPENDICES (SOFTWARE OUTPUT)

The data strategy to build recommender system for this problem statement would be based on the customer segmentation done using clustering above. Some of the possible strategies to be considered based on the 4 clusters are: -

- Customers frequently buying can be recommended samples based on their existing order details (Knowledge based recommender system).
 - Customers with similar buying choices but not buying frequently can be recommended based on the order details of similar customers with high sales record (Collaborative filtering recommendation system).
 - We can use the data from “Raw Data – Order and Sample” in order to group them based on “Customer Code” and segment them accordingly
-
- The decision tree has an accuracy of 91.15% and Kappa value of 0.7082 with the split condition being 80% for train and 20% for test.
 - The prop.table output shows that only 21% of the orders will be converted whereas majority (79%) of the orders won't lead to sale.

```
> table(ConvertsTS$OrderConversion)
```

```
  No  Yes  
930 234
```

```
> prop.table(table(ConvertsTS$OrderConversion))
```

```
      No      Yes  
0.7989691 0.2010309
```

REFERENCES

1. Research Paper :

<https://fliphtml5.com/kzrhj/jhnq/basic>

2. Exploratory Data Analysis + Model Approach:

https://colab.research.google.com/drive/1Jvh8pmfRrjSayPO_JXl4AwE9e71NGSfc?usp=sharing

3. Github link:

https://github.com/Kruthim1304/Datathon_IIM_Bangalore/blob/main/Datathon.ipynb