Capstone Project Submission

Team Member's Name, Email and Contribution:

Individual-Kamalnayan kumar

Email-nayankumar7099@gmail.com

1 :- kamalnayan kumar

- **Data Cleaning**
  a) **Dropping duplicates**
  b) **Handling null and missing values**
  c) **Handling Outliers**

- **EDA (Univariate and Bivariate Analysis)**
  a) **Most frequently ordered products**
  b) **Products with the most quantity ordered**
  c) **Products that make the most money**
  d) **Top 10 countries who's buying maximum products**
  e) **Customers who have placed the most orders**
  f) **Countries from which most orders have been made**
  g) **Hours for which most orders are made**
  h) **Day of the month on which most orders are made**
  i) **Day of the week on which most orders are made**
  j) **Most orders made according to the month**

- **plotting the correlation matrix**
- **Pre-processing the data**
- **check the distribution of the numerical features**
- **Forming the segmentation criteria**

- **Machine Learning Clustering algorithms :**
  a) **K-Means**
  b) **K-Means with Elbow method**
  c) **Hierarchical clustering**

Github Link:-

https://github.com/nayankr77/Online_Retail_Customer_Segmentation

Business all over the world is growing everyday. With the help of technology, they have access to a wider market and hence, a large customer base.

Customer segmentation refers to categorizing into different groups with similar characteristics. Customer segmentation can help businesses focus on each customer group in a different way, in order to maximize benefits for customers as well as the business.

This project mainly deals in segmenting customers of an online business storfe in the UK.

We have done in this project:

1. **Initial preparations.**
2. **Data Cleaning.**
3. **Feature Engineering.**
4. **EDA.**
5. **Forming the segmentation criteria.**
6. **Pre-processing the data.**
7. **Model Implementation.**
8. **Conclusion.**

Problem Statement :

In this project, your task is to identify major customer segments on a transactional data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.

Conclusion

1. The product 'White hanging heart t-light holder' is the most frequently ordered product, around 1700 times. 'Jumbo bag red retrospot' is the second most ordered product, around 1300 times.

2. The product 'Pack of 72 retrospot cake cases' has the most quantity ordered, around 15,000 units. 'Assorted colour bird ornament' is second with around 13,000 units ordered.

3. The product "Product Bunting" has made the most money, around 35,000 sterling. "White Hanging heart T-light holder" being the second, which has made around 32,000 sterling.

4. The customer with the ID: 17841 has the highest number of orders and the customer with the ID: 18118 has the lowest number of orders.

5. The United Kingdom has the most orders placed, with around 3 lakh orders. Germany being second, but way less than the United Kingdom.

6. Most orders are made in the 12th hour, i.e 12pm to 1pm, and the least orders are made in the 6th hour, i.e 6am to 7am.

7. The 6th day of the month has the highest number of orders and the 31st day has the lowest.

8. Most of the orders are made on Thursday,around 66 thousand, and the least number of orders are made on Friday, around 46 thousand.

9. The most number of orders are made in the 11th month, i.e December, and the least in the 2nd month, i.e February.

- **Conclusion from model implementation:**

K means with elbow method is the best model and simple K means is the worst performing model

Actions to take for each cluster:

1. Perform targeted analysis and targeted advertisement for each cluster.

2. Advertise products that can be presented with a discount to the customers in the lesser important clusters, which could convert the customers in these less important groups to customers of more important clusters.