**Capstone Project Submission**

| **Team Member’s Name, Email and Contribution:**  **Individual-Punam Nagrale**  **Email-punamnagrale456@gmail.com**   1. **Punam Nagrale : Email:punamnagrale456@gmail.com**      * + **Data Cleaning**      1. **Dropping duplicates**     2. **Handling null and missing values**     3. **Handling Outliers**      * + **EDA (Univariate and Bivariate Analysis)**   1. **Most frequently ordered products**   2. **Products with the most quantity ordered**   3. **Products that make the most money**   4. **Top 10 countries who's buying maximum products**   5. **Customers who have placed the most orders**   6. **Countries from which most orders have been made**   7. **Hours for which most orders are made**   8. **Day of the month on which most orders are made**   9. **Day of the week on which most orders are made**   10. **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 :**      1. **K-Means**     2. **K-Means with Elbow method**     3. **Hierarchical clustering** |
| --- |
| **Github Link:-**  **https://github.com/punamnagrale456/Online\_Retail\_Customer\_Segmentation\_Capstone\_Project**  **Please paste the GitHub Repo link :-** [**https://drive.google.com/drive/folders/1mMvklPUsTsiwfNI\_8wo6XdJYo6pryYG7?usp=share\_link**](https://drive.google.com/drive/folders/1mMvklPUsTsiwfNI_8wo6XdJYo6pryYG7?usp=share_link)   | **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. **1. Initial preparations.** 2. **2. Data Cleaning.** 3. **3. Feature Engineering.** 4. **4. EDA.** 5. **5. Forming the segmentation criteria.** 6. **6. Pre-processing the data.** 7. **7. Model Implementation.** 8. **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.** | | | --- | --- | |