**SECTION -1 (EDA)**

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | |  | **Kirana Club**  **PROJECT REPORT OF DATA SCIENCE INTERN**  **CHAHAT GUPTA**  **9050767636**  **Chahat.gupta1212@gmail.com** | |

**Section 1: EDA**

You are given a dataset that contains all the transactions occurring for a store.

**On this dataset do an extensive EDA and explain your findings using relevant visualizations.**

Some of the questions you can answer in the EDA:

1. Can customers be segmented into different categories? If yes then perform analysis on the same and also propose categories. If no, then explain why?
2. How would you define a loyal customer?
3. What is the most popular time of year based on this sales data?
4. Is there any seasonality in data? Explain with supportive evidence.
5. Discuss customer's lifetime with respect to the given dataset.

***Observation 1:-***

1. In this analysis, we have performed geographic segmentation by examining the distribution of transactions, average purchase value, and the number of purchases by country. We have identified that the majority of transactions in this dataset are from the United Kingdom, followed by Germany, France, and EIRE. Additionally, we have found that the average purchase value varies significantly across countries, with the highest average purchase value from the Netherlands, followed by Australia and Sweden. This information can be used to identify high-value customers and to develop targeted marketing strategies for each country.
2. The United Kingdom has the highest number of unique customers, with over 3900 unique customer IDs.
3. Other countries with a high number of unique customers include Germany, France, and Spain.
4. Some countries have relatively fewer unique customers, such as Bahrain, Saudi Arabia, and RSA.
5. The number of unique customers in a country does not necessarily correspond to the number of transactions or the total amount of revenue generated in that country. Further analysis is needed to understand customer behavior and purchase patterns in each country.
6. This information can be used to develop targeted marketing campaigns and strategies for each country, with a focus on retaining existing customers and attracting new customers in countries with a smaller customer base.

***Observation 2:***

1. for this dataset the number of loyal customers for each country in the dataset. It appears that the vast majority of loyal customers are from the United Kingdom, with 805 loyal customers identified. France and Germany have the next highest number of loyal customers, each with 23, followed by Belgium with 7, Switzerland with 4, and several other countries with 3 or fewer loyal customers identified.

***Observation 3:-***

Observations of (sheet-2)

* The data has been grouped by year and month, and the total sales for each group have been calculated.
* The year range is from 2010 to 2011, and the month range is from 1 to 12.
* The highest sales were recorded in October 2011, with a total sales value of 263434.090, while the lowest sales were recorded in February 2011, with a total sales value of 127448.770.
* There is a clear seasonality pattern in the data, with sales peaking in the last quarter of the year (October to December) and declining in the first quarter (January to March).

Observations: (Sheet-1)

* The sales for December 2009 were the highest, with a total of 198307.760.
* There is a dip in sales in February 2010, with a total of 116774.536.
* The sales increased significantly in March 2010 with a total of 268304.241.
* The sales peaked in November 2010 with a total of 301913.862.
* December 2010 had a total sales of 166198.750, which is lower than the sales in December 2009.

***Observation 4:-***

(in 2009,2010,2011)

* The monthly sales data indicates that there is a peak in sales in the month of December, which coincides with the Christmas festival. This suggests that customers tend to do more shopping during the festive season. To take advantage of this trend, businesses can create attractive offers and discounts during the holiday season to entice customers and increase sales. By doing so, businesses can capitalize on the increased demand and potentially generate higher revenue during this period. Therefore, it is important to analyze the sales data to identify such trends and take appropriate actions to ensure customer satisfaction and increase profitability.

***Observation 5:-***

* For example, customer ID 12347 has a CLV of 441.11, which means that this customer has spent 441.11 pounds more than the average customer over the assumed 3-year lifespan. On the other hand, customer ID 12346 has a negative CLV of -21.56, which means that this customer has spent 21.56 pounds less than the average customer over the assumed 3-year lifespan