Customer Behavior Prediction Model Using RFM and Logistic Regression

Project MVP

Latifah Mohammed Alrasheedi

7-12-2021

Project Goal

The project's main goal is to build a model that can help identify high-value customers and predict those customers' next date of purchase.

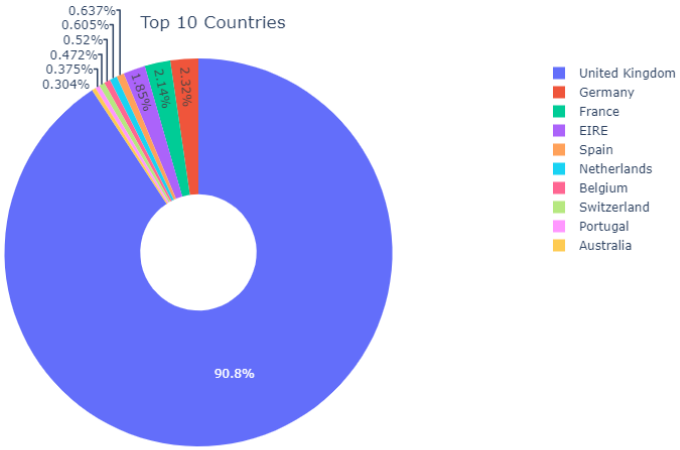
What been accomplished so far

This project will be done in three stages, these stages and their status are described in the table below:

	Description	Status	Notes
Stage 1	Data cleaning and data preprocessing using Pandas, Numpy.	**Done**	Things have been done in this stage: - Change data types. - Change columns' names to lowercase. - Drop null and duplicated values. - Drop canceled orders.
Stage 2	Conducting EDA on the data using Seaborn, Matplotlib, Plotly.	**Done**	Built some charts that helped answer my questions, for example: - Which countries have the highest number of customers? - What is the relationship between the price of a product and the number of purchases? - Which customers have the highest number of orders?
Stage 3	Building the model using Scikit-learn, Logistic Regression model, RFM Model for customer segmentation, K-Means Clustering.	**In progress**	For this stage, I did feature engineering to create two new features: - **total_amount:** which contains the amount of money spent by each customer. - **NextPurchaseDate:** which contains the number of days past after customer last purchase, and that will be my target. RFM model and logistic model are still in progress.

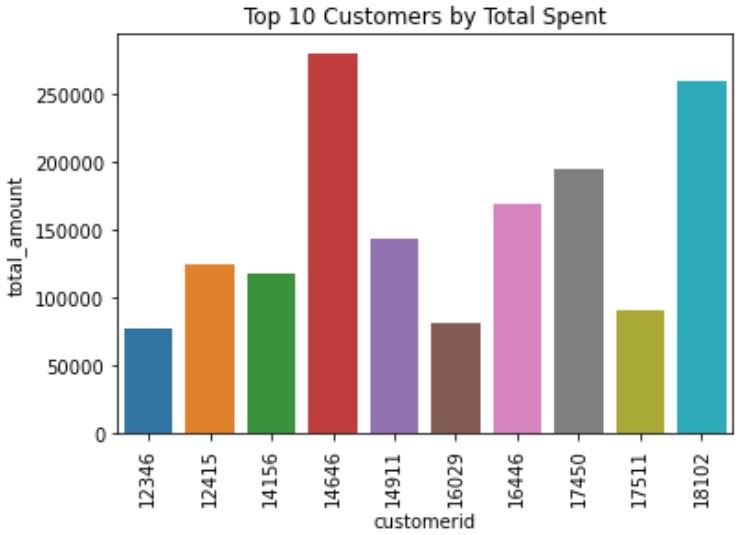
Some visualizations and their observations:

1- Which countries have the highest number of customers?



Most of the customers are from the United Kingdom, which can indicate that the highest revenues of the company come from their customers from the United Kingdom as well.

2- What is the top 10 customers who spent the most money?



We can see that the customer with id's: 14646, spent the highest amount of money. We can also dive into the top spenders and analyze them, for example, see from which country and how much their income.

The next steps

- Building the model using logistic regression and RFM for customer segmentation.
- Communicate the final results.