Customer Behavior Prediction Model Using RFM and Logistic Regression

Tuwaiq Boot Camp Project 2021

Latifah Mohammed Alrasheedi

11-29-2021

Purpose of this project

The study of customer behavior in online shopping usually deals with customers identification and their buying behavior patterns. In this project, we will see if we can build a model that can help us identify high valued customers and predict the next date that a customer will buy from the site again. By identifying the loyalty levels of a customer, a company can focus its marketing campaigns based on the loyalty of a customer. Moreover, by predicting the next date of purchase, the marketing team can target the specific customers based on the date or the type of product they will buy. That will help increase the customers' loyalty and decrease time and effort put into a marketing campaign.

Project Goals:

- Improve customers loyalty.
- Increase time efficiency.
- Design promotions packages that fit the target customers.

Data Description

The data is a transnational data set that contains all the transactions occurring between 01/12/2010 and 09/12/2011 for UK-based online retail found on Kaggle.

Number of features: 8
Number of records: 541908

- Data sample

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
475939	576927	22750	FELTCRAFT PRINCESS LOLA DOLL	4	11/17/2011 11:02	3.75	12567.0	France
135859	547914	POST	POSTAGE	1	3/28/2011 12:35	18.00	12594.0	Italy
151455	549556	23256	CHILDRENS CUTLERY SPACEBOY	2	4/10/2011 12:53	4.15	17372.0	United Kingdom
209613	555257	21294	ETCHED GLASS COASTER	2	6/1/2011 15:28	0.39	17596.0	United Kingdom
46496	540353	22198	LARGE POPCORN HOLDER	4	1/6/2011 14:53	1.65	13764.0	United Kingdom

- Data link
  https://www.kaggle.com/carrie1/ecommerce-data

Tools and algorithms

- Language: Python.
- Environment: Jupyter Notebook.
- For data processing: panda, NumPy
- For data visualization: Seaborn, Matplotlib, Plotly
- Algorithms and modeling: Scikit-learn, Logistic Regression model, RFM Model for customer segmentation, K-Means Clustering.

Future Work

For the next step, I will start the data preprocessing and data cleaning. Then I will conduct EDA on the data to discover the data' patterns which will help in understanding the data better. After the EDA, I will start building the model that will help me study the customer behavior and predict when a customer will purchase from the site again. Last but not least, I will share the results and present them as presentation form.