Xinyu Chen

MGS3101

Nov 14, 2024

Raza Rafique

Project Data Acquisition

I bought this dataset from Chinese database platform. The dataset contains 3,900 records involving multiple variables, including numeric and typed data about online shopping. The numerical variables are Customer ID, Age, Purchase Amount (USD), Review Rating, and Previous Purchases. The Type variables include Gender, Promo Code Used, Payment Method and Frequency of Purchases and so on. These variables cover the customer's basic information, shopping behavior and purchasing preferences. I am going to use these variables to build a relationship between extrinsic variables and online consumer behavior. So that I can know what factors will affect them, which can help companies predict the customers’ behavior so that they can conduct efficient business activities.

The mean Age was 44 and ranged from 18 to 70. The average Purchase Amount (USD) is about $59.76, with fluctuations ranging from $20 to $100. The average value of Previous Purchases is 25, with a maximum value of 50. There are four categories for each Category, and six categories for each Shipping Type. I will provide a more detailed analysis in a future report. There are no missing values in the dataset, demonstrating the diversity of customers' shopping habits, preferences, and their associated attributes. So, it is very suitable for my research on what factors can affect customer behavior on online shopping.