Analysing eCommerce Customer Behaviour

Overview

The project aims to analyse customer behaviour in a eCommerce setting of multicategory store and a cosmetic store. Two different datasets from multicategory store is to be used with date difference of 6 months. One dataset from cosmetic shop will be used to analyse any correlation between other two datasets. The eCommerce Behaviour Data from multicategory store contains behaviour data for October 2019 and April 2020. The eCommerce Behaviour Data from cosmetic store contains behaviour data for November 2019.

Problem Statement

The dataset can be used to provide answers to customer behaviour questions such as: How many sessions does a customer create on average for multicategory and cosmetic store each? How likely is a customer purchase an item for multicategory and cosmetic store each? What is the difference in customer behaviour from October 2019 to April 2019 in multicategory store?

Scope

The scope of the project would be to (1) Acquire the data from the data source and store it into data lake (2) Perform ETL process and design a data warehouse (3) Automate the process using automated data pipeline in a cloud-based environment (4) Analyse the data to find solutions to the problems

Data Source(s)

Data collected by Open CDP project and is directly sourced from https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store?select=2019-Nov.csv for multicategory store and from https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-cosmetics-shop?select=2019-Nov.csv for Cosmetic store.

Architecture of the Solution

1. The data would be acquired and stored in a data lake
2. The data from data lake would be then put through ETL process to form a data warehouse. Cloud solutions will be used to scale the database
3. The data will be analysed from the data warehouse to get solutions to the problem statement

Technology

The choice of technology would depend on the stage of project. The initial choice of technologies is:

1) Data Acquisition: Python, API modules

2) Data Exploration: Python, Pandas, SQL
3) ETL: Python, SQL