

CIE 427 Project Proposal

Prepared by: Rawan Eldalil and Ghada Ghanem

Multi-category Store Dataset

The multi-category dataset contains behavior data for 7 months from a large multi-category online store collected by Open CDP project. Each row in the file represents an event. All events are related to products and users. There are different types of events : Purchase, view, remove, cart. In this project, we'll do the analytics for the 5 GB of data collected in one month, October 2019 [\[1\]](#).

Project Requirements

- Top 10 most purchased and viewed brands. Done
- For each of the top 10 brands, what items that is mostly purchased (e.g. samsung best sells smartphones). Done
- Top 20 most purchasing users. Done
- For the most-purchasing users, what categories they mostly purchase. Done
- Brands that is mostly added to and removed from cart. Done
- Categories that are mostly viewed.

Machine Learning Requirement

An XGBoost model will be trained to predict if a user will purchase a given added-to-cart product or not.