Overview

In this report, we will be examining data from an e-commerce site during November 2019. This data set contains information such as the brand, the category of the product, the user session, the price, whether the user viewed, purchased, or placed the product in their cart, the user accessed the site. The main goal of utilizing this dataset is to identify price data among the different brands. That way, someone could examine the data to see which brands sell certain kinds. Furthermore, someone could use the visualizations to see which products sold the most in a specific brand. The intent was to create a dashboard that someone in a marketing team could use in order to make decisions on what items are the most profitable and to see who the competition could be.

This dataset was accessed from Kaggle, which is a site that specializes in free, downloadable datasets that are already well cleaned. In this case, the data did not need to be cleaned. However, there were other challenges with this specific dataset. For example, the dataset was around 9 gigabytes, which meant it was extremely. This meant it was difficult to load the data. In order to load the data, I had to use Amazon Gluemaker. Glue is a product provided by Amazon Web Services (AWS) that allows for datasets to be extracted, loaded, and transferred to a new location for the data, which was Amazon Athena. Athena is an AWS database that specializes in relational databases. That way, it would be easier to perform SQL queries to extract. To create the visualization, I used Amazon’s visualization tool, called Quicksight. In order to create the visualizations, I queried the data in SQL using Athena to create subsets of the data. Then, I imported the tables into Quicksight so that I could create interactive visualizations based on those new tables.

Results

Based on the results of the visualizations, I found that more expensive products (electronic, furniture) had the most drastic drop in average price from view to purchase. However, products such as tennis rackets stayed relatively stable. In short, people were more likely to buy something such as a toy or sports equipment compared to furniture or electronics., Electronics had the highest average prices across the board with viewing, adding to cart, and purchasing. This goes without saying because electronics are usually more expensive items. Finally, one trend that can be seen is that products in both the view and purchased section are about the same average price. This means that products that are already in view are likely to be purchased as well.

Link to the GitHub repository: <https://github.com/zara-sarkar/MA346-final>

Login Details for Quicksight Dashboard (Note: this dashboard is published, it’s just that Quicksight has a lot of security stuff):

Link: <https://us-east-1.quicksight.aws.amazon.com/sn/dashboards/dcdd44c0-a4c0-4ab1-93a8-13c47e7e8ca0/views/99a148cc-7a06-4132-b0e0-34b6cf7cce30>

Account name: karsarka

Email: [zara@sarkarfamily.org](mailto:zara@sarkarfamily.org)

Password: Dataanalytics1