212141

Erika Therese G. Diño

Ad Hoc Analysis Approach

I first focused on creating a table for the Customer Profile. I created a dataframe for the json file, allowing me to sort the values by the transaction date format. I dropped the unnecessary columns and decided to create separate data frames per month. With this, I dropped duplicate names in each month, giving me the unique values. By using np.count_nonzero and .isin, I determined which customers per month purchased in the current and the previous month, which was the requirement for "Repeaters".

The "Inactive" people, on the other hand, did not purchase in the current month but had purchase history in the previous months before it. Given this, I created new dataframes sorting the values by month/s, such as a Jan to Feb dataframe to compare with the March data or Jan to March dataframe to compare with the April data and so on. I did similar actions with "Repeaters", but after determining the number of similar customers in the data month and the specific dataframe to compare, I subtracted it to the number of customers in the data frame to get those who had purchase history but were inactive in the current month.

Lastly, to get the "Engaged" per month, I concatenated the created data frames from "Repeaters" wherein the duplicated names were dropped. Thus, depending on the month, I can get how many similar names were found. For example, for February, Jan_drop and Feb_drop were linked and I determined how many names had 2 duplicates. For March, I concatenated Jan_drop until Mar_drop and determined how many had 3 duplicates, and so on. For January, since there were no previous months, I placed the number of customers for that month.

I, then, created a table for the data gathered and created a multiple-bar graph as visuals. A brief analysis and description was made in the Jupyter notebook file "Explain_Customer_Profile" while the coded version was made in the Jupyter notebook file "Customer_Profile".

For the breakdown items, I used the same dataframe at the beginning, but this time, I split the items into one item per row, and separated the quantity in another column. I, then, created separate dataframes again per month. Next, I grouped it by the "items", which allowed me to get the sum per item for each month. With this, I made 2 different kinds of multiple-bar graphs to compare the data in different points of views. I also made another dataframe and graph which only showed the quantity of the 3 types of items: Candy City, Exotic Extras, and HealthyKid 3+, for further analysis.

To get the breakdown of sales per month, I looked for the selling price of each item instead of just the transaction value which may contain multiple items. For each item, I figured out which ones only had 1 quantity, and got the minimum value of all of the data to get the selling price. I then multiplied the amount to each quantity of the specific item from the breakdown of items per month allowing me to get the sales of each item per month. Furthermore, I got the total sales per month, total sales per item in all months, and the total sales for everything in 6 months. Afterwards, I created another multiple-bar graph for visuals.

A brief analysis and description for both the breakdown of items and sales are found in the Jupyter notebook file “Explain_Items_Breakdown”, while the coded version is found in the Jupyter notebook file “Items_Breakdown”.