

Analytical Plan and Recommendations for Concept Startup's Retail Location Launch

Annika Hettmann

22 April 2022

Analytical Plan:

The primary objective for our startup is to narrow down the top 10 market locations to launch our retail-lite store against competitors such as Wal-Mart and Target. In order to obtain these ideal market locations, we will clean up and fix missing information in the datasets that will be imported into code. Once that is completed, we will find the top 10 markets by conducting two separate calculations: overall spending, using the sums of medians of DMA spending from our competitors; and spending per capita, using the sums of medians of DMA spending divided by the number of consumer units. The analysis will be wrapped up with K-Means clustering. From there, an even narrower selection of prime locations will be determined based on accumulated data.

Technical Process:

I executed this plan through Python code on the Jupyter Notebook environment, first by loading the datasets on sales, stores with DMA, and features on variables including CPI and unemployment rate (the Features dataset). Before diving into the actual analysis, I grabbed summary statistics from the three datasets to determine whether or not missing information or outliers were affecting the data.

Some data on CPI and Unemployment Rate were missing, so I filled in missing values by extracting datasets from Kaggle, building dictionaries for CPI and Unemployment Rates with datasets consisting of the additional data, and writing a function that finalized the merge. Another dictionary was created to assign numbers to month names in this process. Any data that was not present in the Kaggle datasets was stripped from The Bureau of Labor Statistics. The updated and cleaned dataset was then saved as a copy into a separate folder.

To use data from the sheets on consumer information in the 4 specified regions (Northeast, South, West, Midwest), that information had to be imported differently due to the sheets' structure not matching that of the previous datasets with which I worked. Tableau was used to visualize the 2 bar charts of the top 10 market territories.

Finally, after obtaining the resulting top 10 markets for opening a new retail location – listed in the section below – I conducted K-Means clustering with the Features dataset. 3 datasets resulted from this unsupervised machine learning, where the number of clusters ranged between 2 and 4 per dataset. Scatterplots of the K-Means clustering were produced with particular focus on the CPI and Fuel Price variables. Overall, the most cohesive results came from the data with only 2 clusters assigned.

Results:

The top 10 markets that emerged from this analysis were as follows: Los Angeles, Cleveland-Akron, Chicago, Philadelphia, Houston, Tampa-St. Pete, San Diego, Atlanta, Denver, and Dallas-Ft. Worth. Despite differing quantitative results in the graphs below, all values were present in both graphs.

Figure 1 displays the top 10 markets based on overall spending rankings, sorted in ascending order. These values were the sums of the medians of stores. Figure 2, on the other hand, presents the top 10 markets based on spending per capita rankings, also sorted in ascending order.

These values were calculated by dividing the sums of the medians of stores by the number of consumer units.

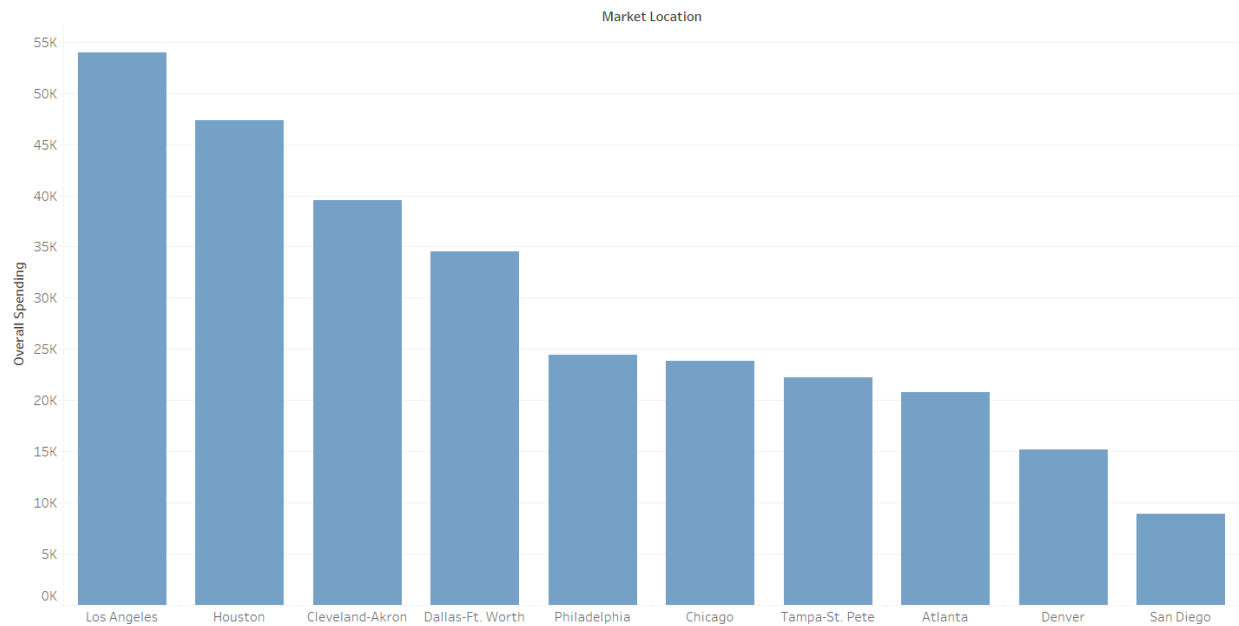


Figure 1: Bar chart of Overall Spending (Sum of Store Medians) in Top 10 Markets

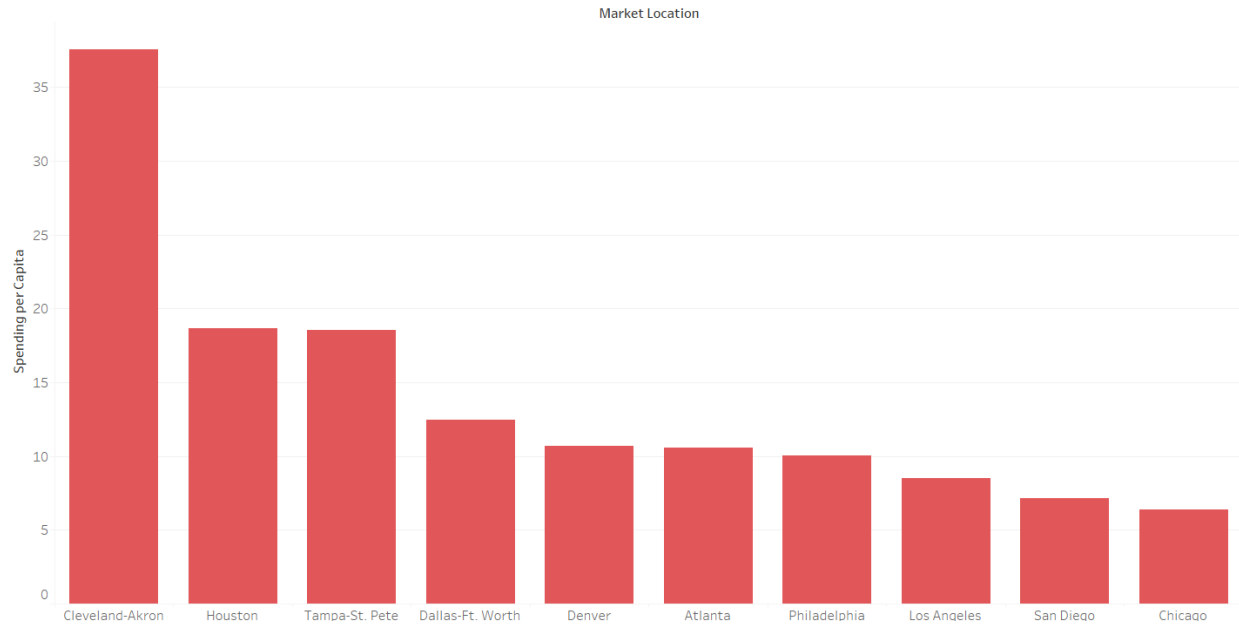


Figure 2: Bar chart of Spending Per Capita (Sums of Store Medians/Number of Consumer Units) in Top 10 Markets

Shifting towards the results of the K-Means clustering, the next pair of visualizations demonstrate the clarity with which only 2 clusters were needed to differentiate data. The dark-colored cluster consisted of data points that had high CPI values and lower fuel prices (more desirable market locations), while the opposite was true for the lighter cluster. The x-axis represents the CPI and fuel price, respectively. The y-axis is the store number.

In this first graph visualizing CPI values (Figure 3), nearly all of the store numbers with high CPIs correspond to the DMAs that were listed on the 2 bar graphs. An interesting observation to note is that Los Angeles has multiple stores, yet some of them are grouped in the yellow cluster. This occurs with several other locations with 2 or more stores, such as Chicago and Austin.

With Figure 4 showing fuel prices, the distinction between clusters is not as strict but still evident. As I discussed earlier, the same stores with low CPI values have generally higher fuel prices in the market area. If it costs customers more money to fuel their vehicles and they are able to reduce how frequently they drive, the market's CPI is negatively affected by this reduction in foot traffic.

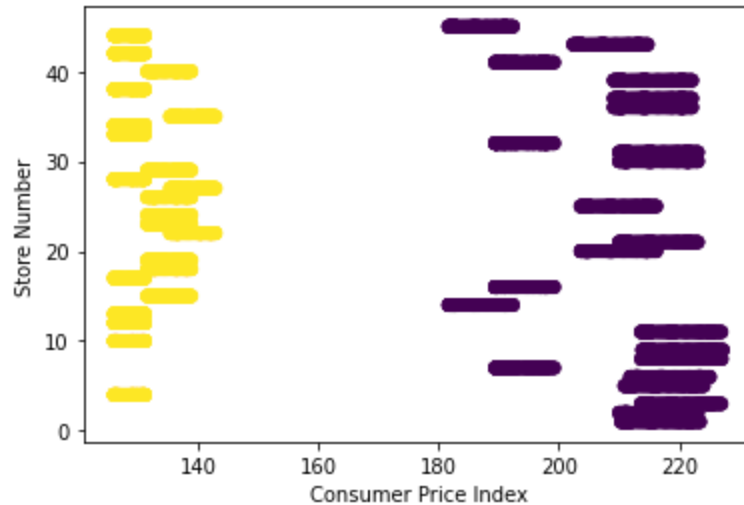


Figure 3: Scatterplot of K-Means Clustering of Store Number by Consumer Price Index

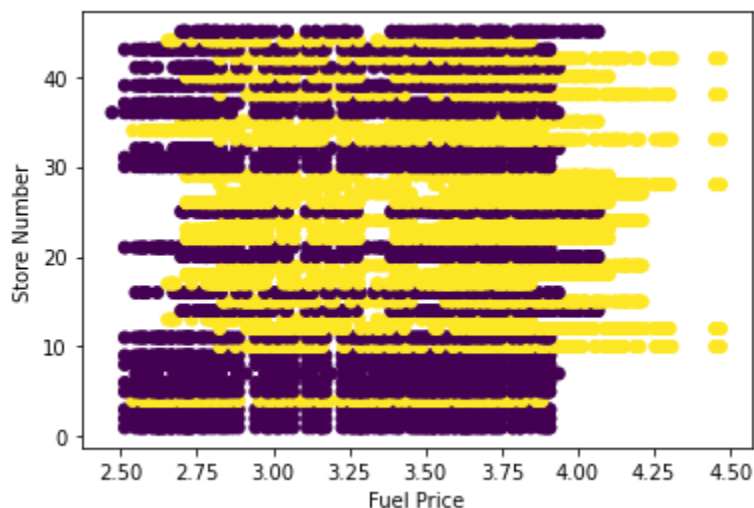


Figure 4: Scatterplot of K-Means Clustering of Store Number by Fuel Price

Market Recommendations for Retail Location Launch:

Houston maintained a 2nd-place position in both graphs. Based on the data obtained on overall spending and spending per capita, it would be ideal to focus on that city for launching our new retail-lite location. Cleveland-Akron and Dallas-Ft. Worth also consistently ranked highly, so they are also locations worth further consideration for launch if Houston becomes an undesirable location (or if we are able to open an additional location). Los Angeles has potential to be successful based on its high overall spending, but its lower spending per capita implies a risky market territory. If we are able to open up multiple locations in the long term and are financially sound, it is certainly an area to consider. A similar

situation applies to the other market areas listed on the above graphs, though their differences between overall spending and spending per capita are not as stark.

Another factor to consider is household vehicle ownership, the data of which is provided in its own Excel file. Of the cities graphed that overlap with the cities listed in the results section, Los Angeles had the largest number of households with at least 1 car. However, considering the findings from this analysis, this does not make as strong of a case to launch the retail-lite location in Los Angeles as opposed to cities like Houston.

Links to External Data Used:

[Kaggle Dataset on Consumer Price Index](#)

[Kaggle Dataset on Unemployment Rates](#)

[Bureau of Labor Statistics Data Viewer](#)