The Battle of Neighborhoods - final report

Coursera
IBM Applied Data Science Capstone

1. Introduction

Business problem: Comparing neighborhood similarity in different cities (New York and Los Angeles)

Target group and use: Individuals interested in moving to new countries to quantify how similar or dissimilar other neighborhoods or cities are to their current place of living

2. Data

Source: Foursquare

Data: Location data for the two selected cities - New York and Los Angeles. This data includes information on the venues within a given radius, including the type of category. By comparing the frequency of each category, thereby showing us which venue categories are more frequent in each city.

3. Methodology

The steps to conduct this study were the following:

1. Extract data using the Foursquare API and load it into two separate dataframes
2. Combine data into a single dataframe
3. Filter columns and extract category information
4. Plot data on map to verify correctness
5. Create pivot table to summarize data, cleaning data as necessary
6. Visualize results

4. Results

Six categories where there is a big difference in the number of venues.

➢ New York appears to have more of a sweet tooth as bakeries and ice cream shops are more frequent
➢ Los Angeles has a greater number of café's, fast food restaurants and coffee shops
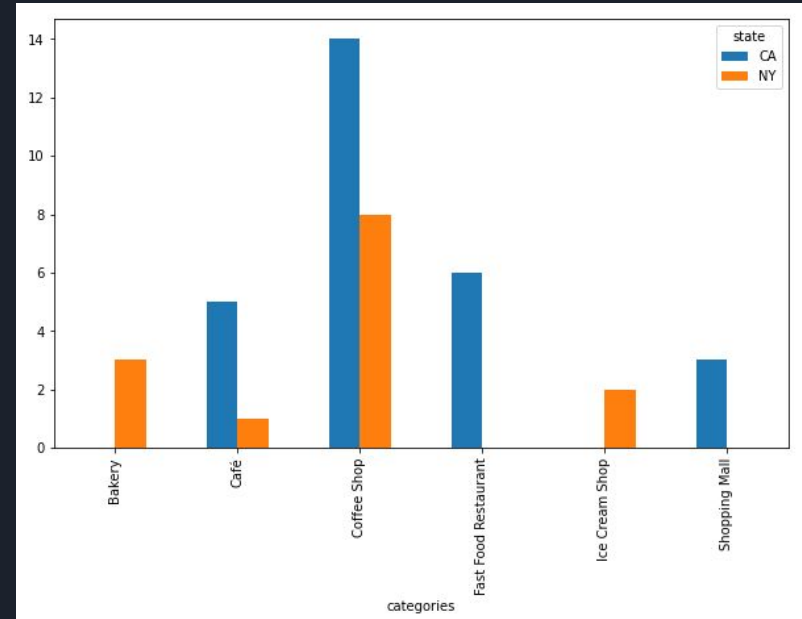
5. Discussion

More complexity than anticipated due to large number of factors to compare

➢ Need to select category of comparison (in this case restaurants)

Foursquare data has some issues (e.g. naming, categorization)

➢ Would need to be further investigated if building a full tool

6. Conclusion

Good progress on building a comparability index

Code easy to build on and to modify to desired category or other cities