Best Restaurant Investment?

IBM Applied Data Science Capstone Course Project – Battle of Neighborhoods

~ Parth Parikh

Table of Contents

Introduction:

Background:

- After Covid-19, restaurants are again in business, more than ever.
- Mass vaccination drives in countries like United States and Toronto.
- New York City and Toronto are huge financial capitals and vast multicultural cities

Introduction:

Problem:

- Restaurant investment is unpredictable currently.
- What restaurant type to invest in and in what neighborhood of the chosen cities?

Introduction:

Solution:

- Data science has a data driven answer for this.
- Exploratory data analysis can be used to find out what restaurant most people in city would go to.
- Making the correct decision through data.

Data Acquisition and Processing

- IBM cloud storage data was used to obtain neighborhood data of New York City and coordinate data for Toronto.
- Web-scraping was used to get Toronto postal codes data from Wikipedia.
- geopy, folium, libraries used to get coordinates of a location and plot them.
- Foursquare API used to get restaurant details.

Data Acquisition and Processing

- Data cleaning was performed to end up with Neighborhood names, coordinates, 10 most visited restaurant of each neighborhood, their coordinates, and categories.
- All data was loaded in various pandas dataframe.
- One Hot Encoding was used to arrange frequently visited restaurant types in each neighborhood by their frequency.

Data Analysis

- All data was processed by total number of neighborhoods where each restaurant type was in top-10.
- Restaurant types were sorted by the above said value.
- Data from both cities had to be normalized for apt comparison.

Data Analysis

- All values were normalized by min-max method of statistics.
- Normalized data was then merged for both cities by restaurant type.
- Normalized values were added from both cities.
- Restaurant type with highest sum was chosen as optimum.

Data Analysis

- Neighborhoods were sorted according to the optimum restaurant type.
- For each city, the neighborhood where the optimum restaurant type was most optimum was/were chosen as optimal locations for this investment.

Results

- Deli/Bodega type restaurants were found out to be consistently one of the most frequently visited in maximum neighborhoods across both New York City as well as Toronto.
- Following neighborhoods were chosen in each city:
  - New York City – Arden Heights
  - Toronto – North Park, Maple Leaf Park, or Upwood Park

Future Directions:

- These results are, although strongly data driven, preliminary as several factors were excluded from the analysis.
- These factors include:
  - Competition from other restaurants
  - Initial cost and yearly maintenance cost
  - Pricing and profitability
  - Labor availability

Future Directions:

- In future, when more data pertaining to these other important factors will be available, even strongly data driven answer can be found.
- This model, although not tested for accuracy, provides a fairly accurate base for future developments on this.

Thank you!