

Capstone Project

Toronto Restaurant Analysis

Ruan H.

Business Problem

The client would like to open a new restaurant.

This restaurant would be situated in the financial district of Toronto.

They are specifically targeting employees' lunches near the "First Canadian Place" building, which is located in the heart of the financial district. They believe there is room for another restaurant / take-away.

They would like an analysis of what is already available. Once the type of food is selected then they would like a recommendation on the price, should it be low-end or high-end?

The Data: Sources

- Foursquare API
- Self generated excel sheet with restaurant prices

The data analysis is heavily reliant on the values returned from the foursquare api.

The location of first Canadian place is sent, and then a radius of 1.5km is selected. The requests are sent through several times to create a collection of data on different restaurant types.

The fields mainly used in the final cluster analysis involves restaurant type (steak, thai, sushi, etc), distance from the building and average price of restaurant.

Prices for each restaurant could only be obtained through extremely difficult, thus an average value was used based on the average price of each mean for that specific restaurant.

The data was uploaded for the notebook to be able to access it.

Methodology

The methodology involved used the following steps:

1. Collect all cuisines data
2. Collect average price for data
3. Combine the data
4. Show range in prices
5. Show frequency of each type
6. Show a cluster of each restaurant type
7. Predict possible prices for cuisines to estimate business possibilities

Exploratory Data Analysis

➤ **Data Transformation**

I started off by collecting all the relevant cuisine information to get an idea of the data.

The following cuisines were queried.

- Thai
- Steak
- Italian
- Sandwich
- Sushi
- French
- Mexican

This resulted in 159 rows.

Going through the data it was noted that some venues found were not related.

Example of this is that “Italian” returned the “Italian consulate” as a venue.

These rows had to be remove.

The 11 rows of data consisted of the following erroneous data and were removed

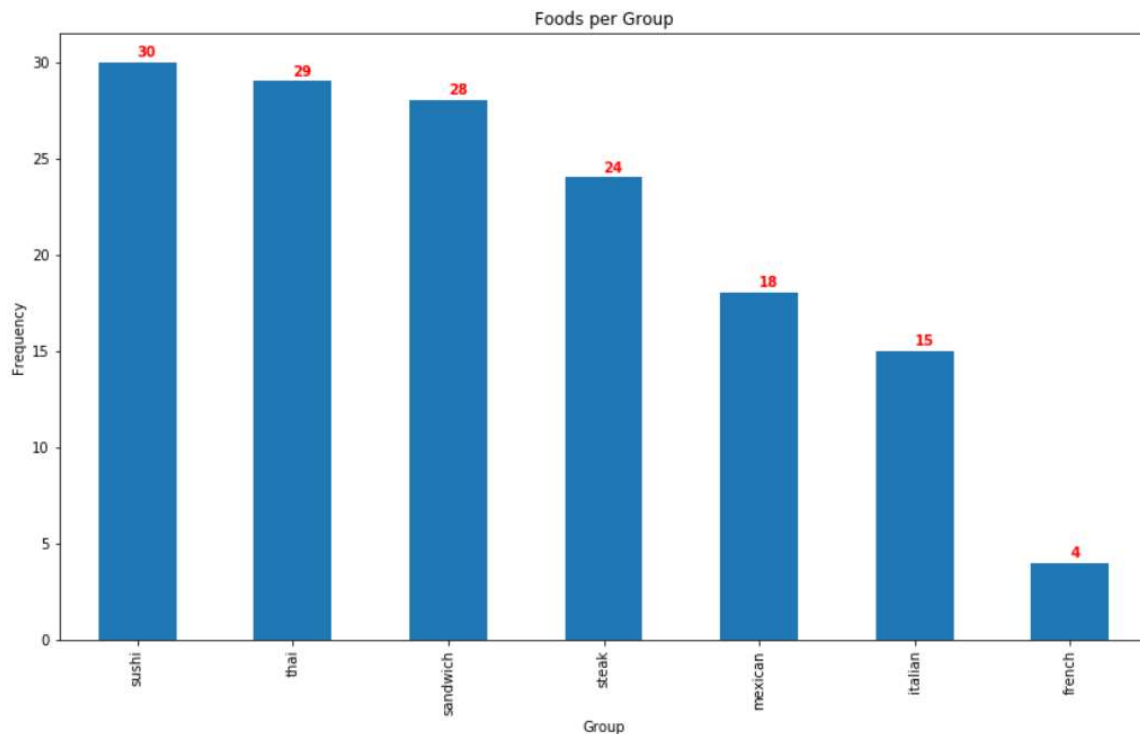
- ☐ Embassy / Consulate (e.g. Italian Consulate)
- ☐ Residential Building (Apartment / Condo)
- ☐ Neighborhood
- ☐ Gym / Fitness Center
- ☐ Clothing Store
- ☐ Government Building

➤ Overview of the Data

Sushi was found to be most popular (by hitting the limit of 30 restaurants) within the 1.5km radius.

Thai the 2nd most popular with 29 venues and

French was the least popular with 4 venues only

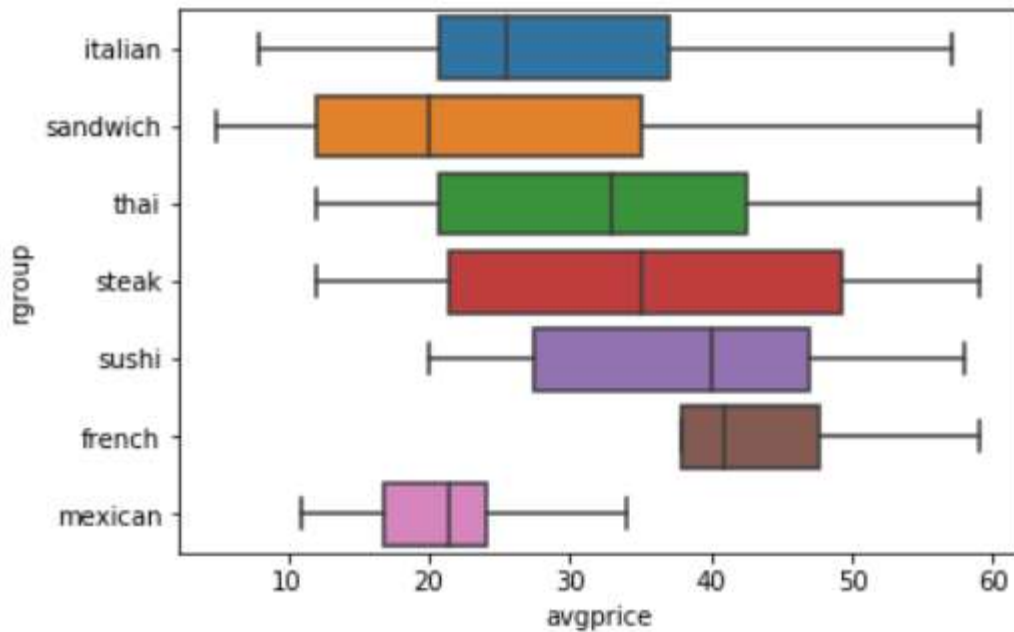


After the data of the average price of each venue was added onto the data frame one could analyze the cuisine with its average price.

A box plot was used to see the high, low and median points of each cuisine.

Some items noted:

- ▶ Italian had the widest range with very cheap and very expensive
- ▶ Sandwiches had the bulk in the cheap area
- ▶ French ranged only in expensive ranges

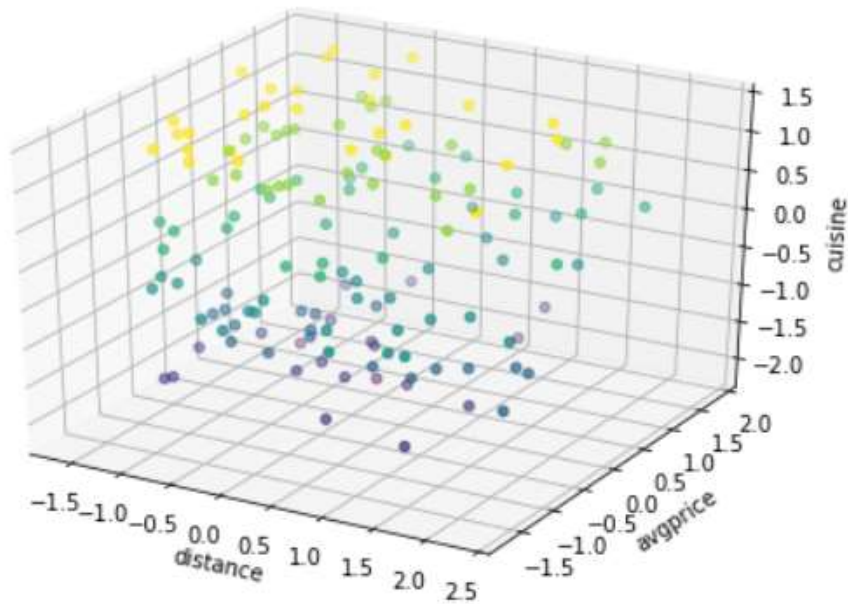


➤ Predictive Analytics

It was finally decided to use a K-Means clustering to determine the average price of a restaurant. Using this model business will be able to supply the distance from the target market and the cuisine type to determine the average price of a restaurant

The client can then determine the potential profitability of a new business.

After a K-Means model was built (and normalized) and the average prices were predicted



The results showed the average prices that could be used

	rgroup	avgprice
0	french	44.750000
1	italian	29.750000
2	mexican	21.500000
3	sandwich	24.862069
4	steak	35.041667
5	sushi	38.111111
6	thai	32.785714

Conclusion

Sushi is the most saturated with more than 30 restaurants therefore one can either

- 1) go with Thai (2nd highest) with an average price of \$33 or
- 2) go for the least saturated market (French) with only 4 restaurants, average price of \$ 45

French would be the final recommendation as there's less competition and a higher price allowed.

There the profit margins might be easier to navigate in a difficult industry