Applied Data Science Capstone Project
The Battle of the Neighborhoods

”Manhattan eats Finance”

By Fabian Schmidt

12/30/2020

1. Introduction/Business Problem

'Restaurbank' is a small New York City-based finance company with a niche business model. They are specialized in loan-funding small and medium sized restaurant businesses within Manhattan. Restaurants are considered being a riskier industry sector than others due to higher default rates and less fixed assets that could be sold to repay the loan. Despite this fact, usually 'Restaurbank's business model works quite well as they claim to perform a more rigorous due diligence before handing out the loan. Also, they argue their collaborative approach to credit lending with advice and networking among the restaurant community makes their clients more successful.

However, due to the current events around the COVID-19 pandemic new clients who would like to start a restaurant in Manhattan are rare. At the same time, a higher client default rate is putting risk management back onto the firm's schedule. So, the company's problems are both acquiring new, solid clients as well as protecting their existing business from further defaults.

Therefore, the risk management department set up a task force with a few statisticians to assess if this new hype topic called 'Data Science' might help the firm with their current challenges. They stumbled upon the geospatial data provider Foursquare and came up with an idea on how help both their department with managing risk as well as the front office staff in their client acquisition. They scheduled a meeting with the managers of the front office and the risk management department to present their results.

In the meeting they introduced their 'Manhattan eats Finance' initiative: By using restaurant ratings on Foursquare, they want to find hot- and blackspots of great restaurants. The front office can leverage these findings to guide new clients asking for a loan to areas where there are fewer great restaurants and therefore competition is less intense. Also, they can approach the owners of top-rated restaurants to restructure their debt with 'Restaurbank', providing lower interest rates than they currently pay for their loans. The risk management department could include the restaurant rating from Foursquare in their credit risk assessment and management, assuming lower rated restaurants might not be able to withstand competition and default.

2. Data

To achieve this, different types of data from Foursquare will be used. At first, a list of restaurants across Manhattan is retrieved that contains the venue ID, name, its coordinates given by longitude and latitude as well as the venue's category. To do so, Foursquare is queried with the search term restaurant. Below is a screenshot of the first few rows of the resulting Pandas DataFrame. Note that the venue category also has values like 'bakery' or 'snack place'.

Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
5894c9a15e56b417cf79e553	Xi'an Famous Foods	40.715232	-73.997263	Chinese Restaurant
3fd66200f964a520bce61ee3	La Bella Ferrara	40.717450	-73.998373	Bakery
5c965dad5455b2002c058659	Yi Ji Shi Mo Noodle Corp	40.718254	-73.995930	Chinese Restaurant
4bcf9774a8b3a5939497625f	Shanghai 21	40.714423	-73.998904	Shanghai Restaurant
4a00e0a7f964a520bc701fe3	Singapore Malaysia Beef Jerky	40.718527	-73.995824	Snack Place

In a second step the scores for these venues are queried. To retrieve these a premium call to the Foursquare API is needed querying the details of every single venue in the list. Therefore, the amount of data is limited, as only 500 such premium calls per day are allowed with the free Foursquare developer account. Below you can find a screenshot of the resulting DataFrame including the venue ID as well as the venue's score.

Venue ID	Venue Score
5894c9a15e56b417cf79e553	8.9
3fd66200f964a520bce61ee3	8.9
5c965dad5455b2002c058659	8.9
4bcf9774a8b3a5939497625f	8.9
4a00e0a7f964a520bc701fe3	8.8

In the next step both both DataFrames are joined on the venue ID index resulting the data set that will be used for the further project.

	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Score
Venue ID					
5894c9a15e56b417cf79e553	Xi'an Famous Foods	40.715232	-73.997263	Chinese Restaurant	8.9
3fd66200f964a520bce61ee3	La Bella Ferrara	40.717450	-73.998373	Bakery	8.9
5c965dad5455b2002c058659	Yi Ji Shi Mo Noodle Corp	40.718254	-73.995930	Chinese Restaurant	8.9
4bcf9774a8b3a5939497625f	Shanghai 21	40.714423	-73.998904	Shanghai Restaurant	8.9
4a00e0a7f964a520bc701fe3	Singapore Malaysia Beef Jerky	40.718527	-73.995824	Snack Place	8.8

3. Methodology

4. Results

5. Discussion

6. Conclusion