

Opening a Hungarian Fine Dining Restaurant in the United States

Coursera: Applied Data Science Capstone Project

Kristof Domokos Szabo, Budapest, 2020.04.28

Contents

1. Introduction	1
2. Problem Statement.....	1
3. Data.....	2
4. Methodology.....	2
5. Results.....	2
6. Conclusion.....	8

1. Introduction

In 2019, the United States restaurant market has generated 839 billion USD, an increase of 3.6 percent from the previous year.¹ According to Business Insider, the average American family spends around 4400 annually eating out, with an average of 5.9 times a week.² The restaurant market is ever changing, and these days fine dining seems to be taking the spotlight. From 2017 to 2018 alone, the fine dining industry saw a market increase of 20.5 %.³ This report aims to support a hypothetical group of investors in opening up a Hungarian Fine Dining restaurant in the United States.

2. Problem Statement

The author would like to highlight that as Hungary is a small nation of 10 million people and the European fine dining industry being highly competitive, the group of investors would only consider opening the restaurant in one of the major cities in the United States. Furthermore it is important to note that due to the Hungarian Fine Dining brand being relatively unknown, the project will also consider the size of Hungarian-American diaspora in these cities.

¹ <https://www.nrn.com/sales-trends/us-restaurant-sales-reach-record-863b-2019-nra-says>

² <https://www.businessinsider.com/what-people-spend-on-dining-out-2019-8?r=US&IR=T>

³ <https://www.boutiquehotelier.com/top-trends-whats-hot-in-fine-dining-for-2019/>

Considering the fact that Hungarian-American diaspora could serve as a strong customer upon which the business can be expanded.

The following project aims to pinpoint locations, where exploration into opening such a restaurant is underpinned by a strong business case.

3. Data

The data used in this project has been retrieved from multiple sources and have been processed with careful consideration regarding its accuracy. The author would like to highlight that sources such as Wikipedia would usually not be considered for doing studies and is only considered a valid source in the scope of this hypothetical study.

Several articles have been scraped and used for analysis from Wikipedia, primarily to draw up conclusions around the Hungarian-American diaspora and exploring the possible economic strength of their respective states. Python package BeautifulSoup has been used to scrape information regarding the size of [Hungarian-American diaspora in each state](#). This information served as the foundation for further analysis. In order to gauge economic strength of these states, [statewide economic strength](#), [size of population of these states and their biggest cities](#) and [economic power of its biggest cities](#) have also been included in the analysis.

In order to better understand the specifics of the possible location Foursquare API has been used for two main purposes. Firstly, to better understand the presence of Hungarian diaspora in the primary location. Secondly, to have a clear understanding around competition in the city and recommend a location accordingly.

Geo data has also been used to better visualize results of the analysis. The author principally relied on Geocoding data provided for New York City by Coursera and a GeoJSON file to show results of Pandas dataframes on a Neighborhood level.

4. Methodology

According to the author, the research and analysis leading up to the conclusion can be separated into 3 distinct phases. During the first phase, the author conducted web research in order to establish some key information related to the size of the Hungarian-American diaspora in the United States of America. In order to do this, Python package BeautifulSoup has been used. BeautifulSoup is aimed at enabling Python users to scrape certain parts of HTML code with the possibility of systematically analyze them with Python library Pandas.

In order to further support findings around the city of choice, Pandas dataframes enhanced from several Wikipedia sources have been further strengthened by Foursquare API (venue) data. Venue data has been used to support claims around possible locations for the restaurant in the city of choice as well as a foundation for analyzing the competition in the city. Geocoding data has also been received from Coursera in order to support easy visualization of findings.

Last but not least, findings synthesized in Pandas dataframes have been visually enhanced to strengthen key findings with the use of Folium library and corresponding GeoJSON data.

5. Results

In the first phase of the research, the author extracted the states in the US with a considerable size of Hungarian-American population. 7 states have been identified on the corresponding Wikipedia website. The results have been scraped and put into a Pandas Dataframe.

	State	Population
1	Ohio	203,417
2	New York	157,863
3	California	133,988
4	Pennsylvania	132,184
5	New Jersey	115,615
6	Michigan	98,036
7	Florida	96,885

This dataframe has been paired with median income for each one of the states, as economic strength is a good indicator for wealth spent on leisure activity.

	State	Population	Median Income
4	New Jersey	115,615	\$81,740
2	California	133,988	\$75,277
1	New York	157,863	\$67,844
3	Pennsylvania	132,184	\$60,905
5	Michigan	98,036	\$56,697
0	Ohio	203,417	\$56,111
6	Florida	96,885	\$55,462

Once the relative economic strength of each state has been established, the need for establishing a restaurant in a major city is also a decisive factor. In order to understand the possible cities for the restaurant, the Pandas dataframe has been further enhanced with scraped data. Each state's largest city and its population size has been added to the view.

	State	Population	Median Income	Most populous	City population
0	New Jersey	115,615	\$81,740	Newark	277,140
1	California	133,988	\$75,277	Los Angeles	4,792,621
2	New York	157,863	\$67,844	New York City	8,175,133
3	Pennsylvania	132,184	\$60,905	Philadelphia	1,526,006
4	Michigan	98,036	\$56,697	Detroit	713,777
5	Ohio	203,417	\$56,111	Columbus	879,170
6	Florida	96,885	\$55,462	Jacksonville	880,619

Even though New Jersey has the highest median income out of all of the states, the size of the Hungarian-American diaspora is relatively small and its biggest city only has a population of 277K

citizens. The author recommends that the cities of Los Angeles, New York City and Philadelphia will be looked into for further analysis.

To gain a widened perspective on the possible locations, data related to median household income has also been scraped from Wikipedia and added for the top 3 cities in the dataframe.

	State	Hungarian Ancestry Population	Median Income	Biggest City	City population	Median Household Income
2	New York	157,863	\$67,844	New York City	8,175,133	
1	California	133,988	\$75,277	Los Angeles	4,792,621	
3	Pennsylvania	132,184	\$60,905	Philadelphia	1,526,006	

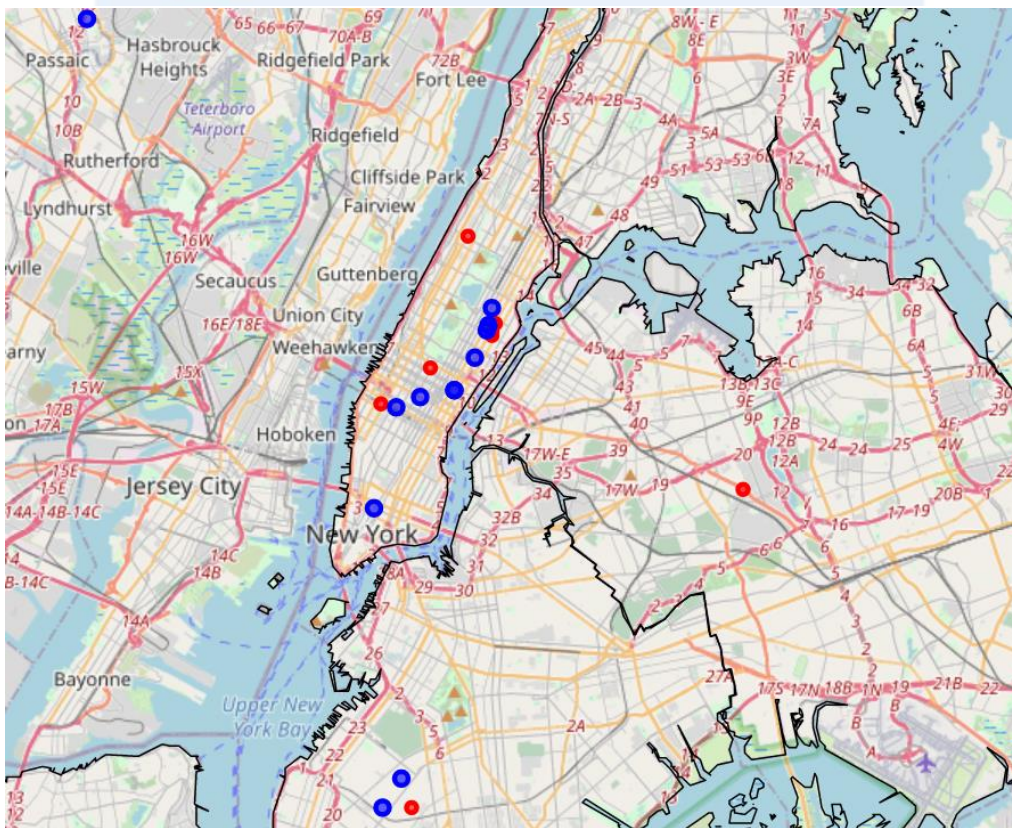
In the first phase of the analysis, it has been established that New York City has the highest median household income, by far the largest city population and Hungarian-American diaspora. It has been established the New York City will be looked into for further details.

In the second phase of the research, the author wanted to establish whether there any Hungarian restaurants in the New York City area and where are they concentrated. In order to do this, REST API calls have been to the Foursquare API and its results have been put into a Pandas Dataframe. Using Folium library, these results have been put on a New York City map.



In this section it has been established that the majority of the Hungarian restaurants are in the packed Manhattan borough. With the high number of restaurants here, it will likely results in unwanted competition for the newly opened restaurant.

However, the author would like to point out that venue data has also been scraped and visualized for places significant for the Hungarian community. The author believes that proximity to these venues for the new restaurant will be crucial, Hungarian-Americans could serve as a strong foundation for its customer base.



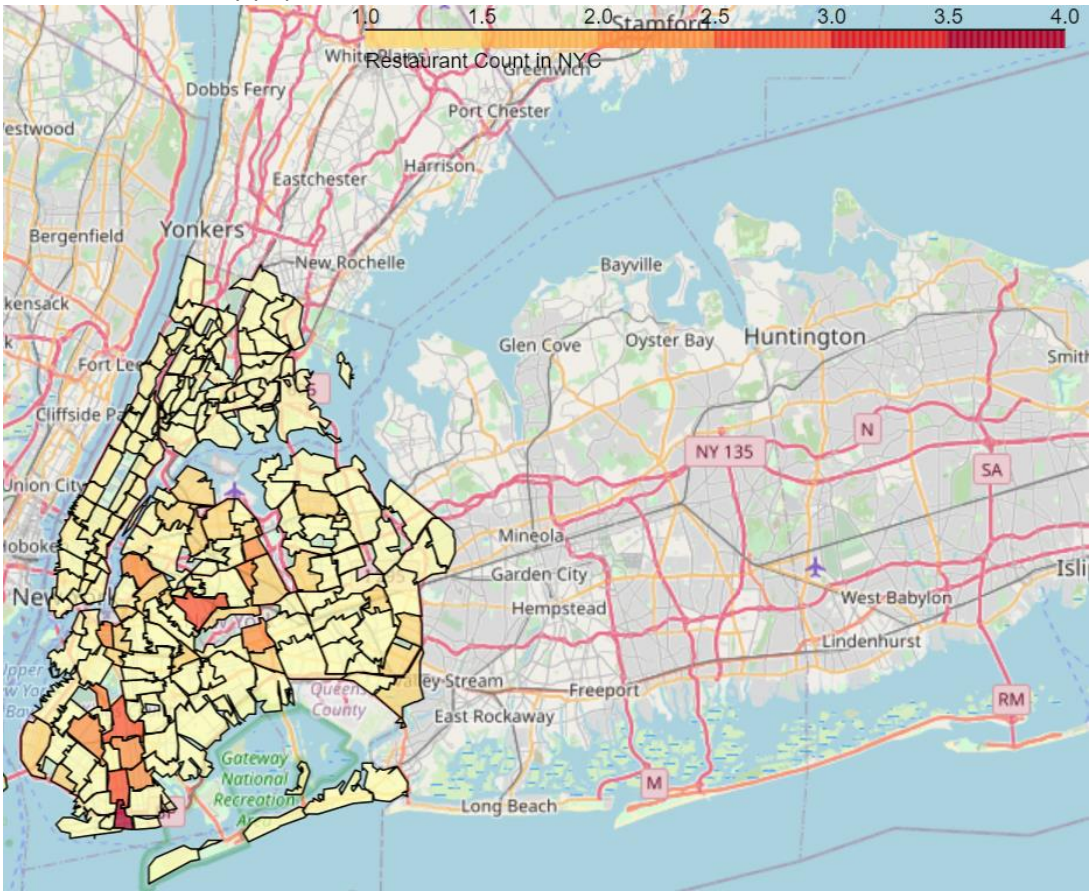
This analysis has also shown that the majority of Hungarian-American venues are centered in the Manhattan borough, while two venues were found close to a Hungarian eatery in Brooklyn. For the

scope of this project and based on this analysis the author looked into opening a restaurant in the Queens or Brooklyn boroughs, in close proximity to Manhattan.

In the third phase of this project, Foursquare venue data has been used again to analyze all possible competition for the new restaurant in Brooklyn and Queens. For this neighborhood geocoding has been used and the count of restaurants in each neighborhood have been pulled into a Pandas dataframe.

Borough count			Borough Neighborhood Latitude Longitude			
0	Brooklyn	58	46	Brooklyn	Bay Ridge	40.625801 -74.030621
47	Queens	37	47	Brooklyn	Bensonhurst	40.611009 -73.995180

When visualizing these results on a map, it seems that the majority of neighborhoods close to Manhattan are not densely populated with restaurants.



6. Conclusion

In the scope of this analysis, it has been decided that with the relative size of Hungarian diaspora, overall city population size and economic strength, New York City would be an ideal location to open a new Hungarian Fine Dining restaurant. The author first looked into direct competition by analyzing Hungarian restaurants in New York, majority of them operating in Manhattan borough. For the scope of this project it has been established that competition of this strength in the same borough is likely not beneficial for a new restaurant.

By further looking into details, it has been established the majority of Hungarian venues are also in Manhattan, hence establishing the fact the close proximity to Manhattan would be crucial to appeal to Hungarian-American customer base. The author also conducted an analysis of other restaurants in Queens and New York boroughs and concluded the fact that the majority of neighborhoods in close proximity of Manhattan have no more than 3 restaurants.

In conclusion the author recommends that for further refinement, Brooklyn and Queens neighborhoods in close proximity to Manhattan should be further explored to establish the location for the new restaurant. The author would also like to point out the fact that this study is sensitive to its limitations and only serves as an initial indicator for pinpointing a location in New York City. Further analysis would be needed to strengthen the hypothesis which is out of scope for this project.