**Investigation of Spanish restaurant scene**

**Comparison of New York & Toronto**

Marcel Gerhards

Octubre 6, 2019

**Content**

[**1. Introduction and Business Problem** 2](#_Toc21294007)

[**1.1 Background** 2](#_Toc21294008)

[**1.2 Problem** 2](#_Toc21294009)

[**1.3 Interest** 2](#_Toc21294010)

[**2. Data section** 3](#_Toc21294011)

[**2.1 Data sources** 3](#_Toc21294012)

[**2.1.1 New York and Manhattan Neighborhood dataframe** 3](#_Toc21294013)

[**2.1.2 Toronto Neighborhood dataframe** 3](#_Toc21294014)

[**2.1.3 Data on restaurants in New York and Toronto** 3](#_Toc21294015)

[**2.2 Data cleaning** 4](#_Toc21294016)

[**2.2.1 New York Neighborhoods dataframe** 4](#_Toc21294017)

[**2.2.2 Toronto Neighborhoods dataframe** 4](#_Toc21294018)

[**2.3 Feature selection** 4](#_Toc21294019)

[**2.4 Datasets** 5](#_Toc21294020)

[**2.4.1 Spanish Restaurants in New York** 5](#_Toc21294021)

[**2.4.2 Tapas Restaurants in New York** 6](#_Toc21294022)

[**2.4.3 Spanish Restaurants in Manhattan** 6](#_Toc21294023)

[**2.4.4 Tapas Restaurants in Manhattan** 7](#_Toc21294024)

[**2.4.5 Tapas & Spanish Restaurants in Toronto** 7](#_Toc21294025)

# **1. Introduction and Business Problem**

## **1.1 Background**

Spanish food has become a new trend all over Europe and an increasing number of restaurants serving “tapas”, the typical small Spanish dishes are becoming extremely popular all over Europe. Major restaurant chains are opening all over Europe, following the franchise system. As these companies are having a great success, they are thinking to expand to the US to expand their market. As all these restaurant companies are Europe based, and have little experience in expanding to North America and additionally knowledge of the North American food industry is lacking a deep analysis is urgently needed. As Toronto and New York are trending cities with pioneering food industries, these two cities have been elected to perform an analysis of the current situation.

## **1.2 Problem**

Data that might contribute to the analysis of the problem are: Are there any restaurants that are already serving Spanish food? If yes, in which neighborhoods are they located? How many restaurants are there

## **1.3 Interest**

Obviously, the current Spanish restaurant industry would be very interested to know about a potential market and business opportunities overseas. Additionally this might even be a business opportunity for Spanish people who want to move abroad and open a small Spanish restaurant by themselves.

# **2. Data section**

## **2.1 Data sources**

### **2.1.1 New York and Manhattan Neighborhood dataframe**

The dataset about the New York Neighborhoods existed free on the web and was downloaded from the following link:

<https://geo.nyu.edu/catalog/nyu_2451_34572>

### **2.1.2 Toronto Neighborhood dataframe**

The dataset about the Toronto Neighborhoods was obtained from the following Wikipedia page:

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

To obtain the dataset the information was obtained by scraping the website using the BeautifulSoup package.

The geographical coordinates of the neighborhoods was obtained using the Geocoder package and downloaded from the following link:

<https://cocl.us/Geospatial_data>

### **2.1.3 Data on restaurants in New York and Toronto**

Data about the restaurants where received by calling the Foursquare API to obtain venues in the concerned areas.

## **2.2 Data cleaning**

### **2.2.1 New York Neighborhoods dataframe**

From the New York dataset only the columns of interest were selected: 'Borough', 'Neighborhood', 'Latitude', 'Longitude'. No further cleaning of the data was necessary.

### **2.2.2 Toronto Neighborhoods dataframe**

From the Toronto dataset the rows with “Not assigned” value in the Borough column were dropped. Rows with the same Postal Code and Borough were joined. Not assigned Neighborhoods were replaced by the concerning Boroughs. Then the dataset was sorted by Postal Code.

## **2.3 Feature selection**

After obtaining the venues through Foursquare API, the dataset was categorized with onehot encoding to create a new dataset with all categories and the applying category with a value as 1. Therefore, it was easy to select the rows and corresponding venues that applied to a specific category.

Two new datasets were created with Tapas restaurant and Spanish restaurants, listing all restaurants of this type in the different cities.

## **2.4 Datasets**

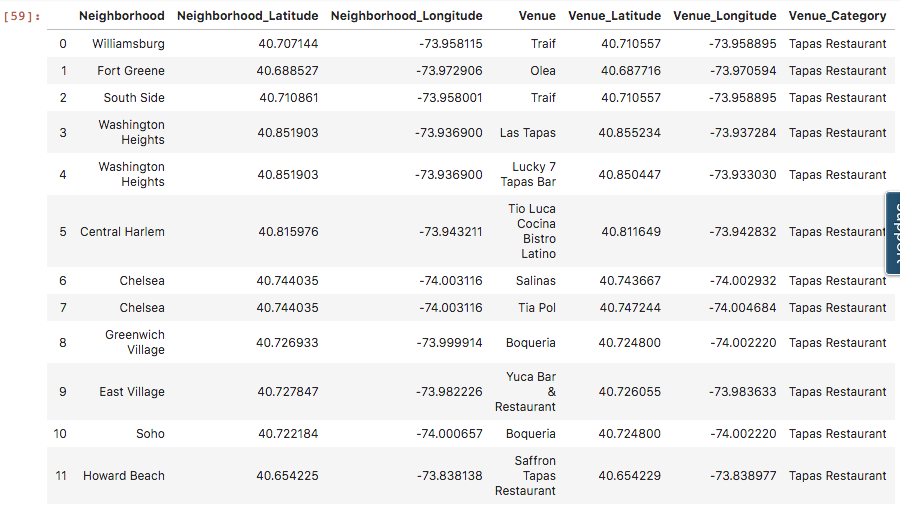
### **2.4.1 Spanish Restaurants in New York**

(only a part)

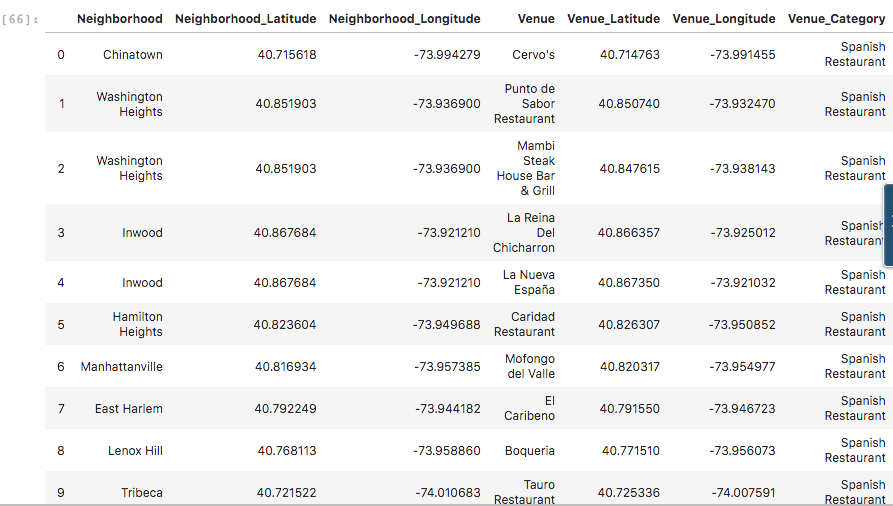




### **2.4.2 Tapas Restaurants in New York**

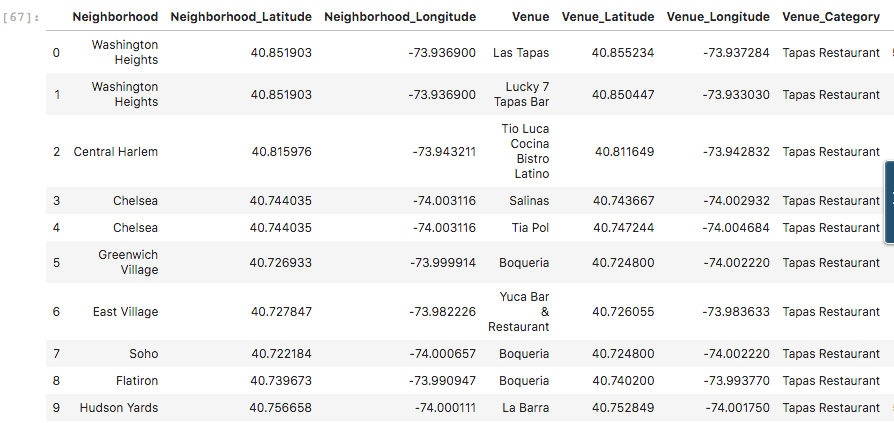


### **2.4.3 Spanish Restaurants in Manhattan**





### **2.4.4 Tapas Restaurants in Manhattan**



### **2.4.5 Tapas & Spanish Restaurants in Toronto**

There was no Spanish restaurant and only one Tapas restaurant in Toronto!!

