

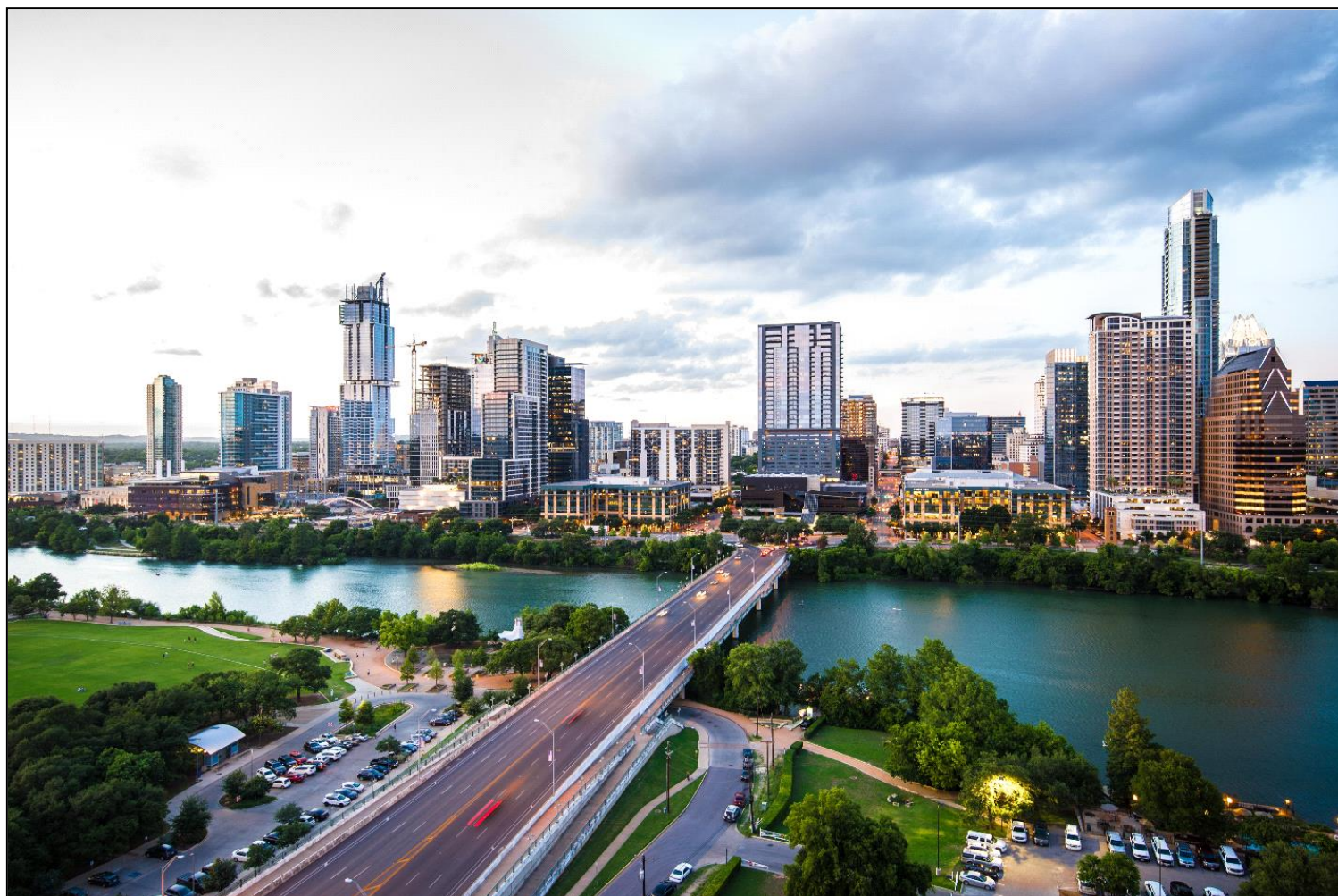
# **Coursera Capstone**

IBM Applied Data Science Capstone

## ***Opening an Ethnical East European Restaurant in Houston, Texas***

By: Denis Nemchenko

November 2019



# Introduction

For many people visiting ethnical restaurants is a great way to enjoy a new food experience during weekends and holidays. People eat every day, multiple times a day. Though during the weekdays people usually eat the food they already know at the places they know well, during weekends and holidays, many people like to get a new gastronomic experience. Besides, there is a big East European Community in Houston and there are only two or three East European restaurants/cafes in the Greater Houston area. Potentially the owner of the new East European restaurant can have great success and consistent profit. However, as with any business, opening a new restaurant requires serious considerations and is more complicated than it seems from the first glance. In particular, the location of the restaurant is one of the most important factors that will affect whether it will have success or a failure.

## Business Problem

The objective of this Capstone project is to analyze and select the best locations in the city of Houston, Texas to open a new East European restaurant. Using Data Science methodology and Machine learning instruments such as clustering, this project aims to provide solutions to answer the business question: in the city of Houston, Texas, if a developer is looking for an opportunity to open a new East European restaurant, where is the best place do it.

## Target Audience of this project and some demographic facts

This project is particularly useful to developers and investors looking to open or invest in a new East European restaurant in the city of Houston, Texas. Overall, Houston is a great place to open a restaurant with an ethnical cuisine. Houston, with 2.3 million residents, is the fourth most populous city in the United States, trailing only New York, Los Angeles and Chicago. The city is the largest in the South and the Southwest. Houston is the nation's demographic future. More than 145 different languages are spoken in Houston. That is the third largest number of languages spoken in a U.S. city behind New York (192) and LA (185). Houston is considered to have one of the best culinary scenes in the country, with cuisine from around the world. There are more than **10,000 restaurants** in the Houston area with culinary choices that represent more than 70 countries and American

regions. Houstonians dine out nearly more than residents of any other city—6.9 times per week, compared with the national average of 4.9. The average meal in Houston—\$35.57—is less than the national average of \$36.40, according to Zagat. In '16, Yelp.co. listed Houston as having more than 75 national categories of cuisine, and listed over 600 vegan friendly restaurants, more than 150 farm-to-table restaurants and more than 700 food trucks in Houston.

## Data

**To solve the problem, we will need the following data:**

- List of neighborhoods in Houston. It defines the scope of this project, confined to the city of Houston, the largest in the South and the Southwest.
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data.
- Venue data, particularly data related to restaurants. We are going to use these data to perform clustering of the neighborhoods.

### Data Source and methods to extract them

We will use the Wikipedia data to extract the list of Houston neighborhoods. To extract the needed data we will apply web scrapping techniques, as well as Python requests and beautiful soup packages. After it, we will get the geographical coordinates of the neighborhoods (latitude and longitude) using Python Geocoder package.

Finally, we will use Foursquare API to get the venue data for the neighborhoods defined at the previous step. Foursquare has one of the largest databases of 105+ million places and over 125,000 developers use this application. Foursquare API provides many categories of the venue data; we are particularly interested in the restaurant data to solve the business problem defined above. This project will require using of many data science skills, from web scrapping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next Methodology section, we will discuss and describe any exploratory data analysis that we did, any inferential statistical testing that we performed, and what machine learning technique were used.