

Coursera Capstone

IBM Applied Data Science Capstone

Opening an Indian Restaurant in New York



Introduction

New York, often called **New York City** to distinguish from New York State, or **NYC** for short, is the most populous city in the United States. With an estimated 2020 population of 8,253,213 distributed over about 302.6 square miles (784 km²), New York City is also the most densely populated major city in the United States. Located at the southern tip of the State of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban area. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports, and is the most photographed city in the world. Home to the headquarters of the United Nations, New York is an important center for international diplomacy, and has sometimes been called the capital of the world.

This final project explores the best locations for Indian restaurants throughout the city of New York. Potentially the owner of the new Indian 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. So our project will attempt to answer the questions "Where the Indian restaurant should be open by the investor ?" and "Where Should I eat to have great Indian food ?"

Business Problem

The objective of this Capstone project is to analyze and select the best locations in the city of New York to open a new Indian restaurant. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question: Where the Indian restaurant should be open by the investor, in the city of New York ?

Target Audience of this project and some demographic facts

This project is particularly useful to developers and investors looking to open or invest in an Indian restaurant in the city of New York. Overall, New York is a great place to open a restaurant with a multicultural food. As New York is the most diverse city in the world (800 languages are spoken in New York). With its diverse culture, comes diversity in the food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Japanese, French, etc. Why did we decide to focus on Indian cuisine in our project ? Now when the idea of a healthy and spicy lifestyle conquered the minds of people all over the country, Indian restaurants became extremely popular, as they offer a healthy and spicy alternative to regular American eating habits.

Data

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

- New York City data containing the neighborhoods and boroughs.
- 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 this data to perform further analysis of the neighborhoods.

Data Source and methods to extract them

New York City data containing the neighborhoods and boroughs will be obtained from the open data source: https://cocl.us/new_york_dataset. 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 (open source dataset), working with API (Foursquare), data cleaning, data wrangling, to 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 techniques were used.