

Applied Data Science Capstone Project

Part-1

Exploring New York City's Diverse Food Cultures for Tourism Purposes

Shabbir Ahmad

Contents

Introduction and Problem Description	2
Data Description	2
Requirements.....	2
Data Sources.....	2
Audience/Stakeholders	3
References	3

Introduction and Problem Description

New York City is the most populous city in the United States and the largest metropolitan area in the world by urban landmass (World Urban Areas, 2018). The city has remained a major entry point for immigrants over the years. According to recent statistics, about 36 percent of the city's population is foreign-born (NYC Mayor's Office, 2018). The ethnic diversity of New York City is reflected in its diverse cuisine. Almost all the ethnic cuisines are available in the city. John Corry, the New York Times reporter rightly pointed out that "One of the glories of New York is its ethnic food" (Food Reference, 2020).

Tourism industry in the city is in boom for over a decade. The city welcomed over 65 million tourists in 2019 which included over 13 million foreign tourists (NYC and Company, 2020). While visiting different sites, tourists often find it difficult to find the restaurants and eateries of their choice in those areas. For instance, in some neighbourhoods, there are numerous outlets of Italian and Caribbean food but there might be only one or two Mexican or Indian restaurants. Moreover, sometimes the tourists are also overwhelmed by the food choices and the decision-making process may be time consuming.

In this project, I aim to explore all the boroughs of New York City and locate all the possible food options in specific areas and neighbourhoods of the city. Eateries for different food categories and cuisine belonging to different ethnic groups and countries will be located in the city.

Data Description

Requirements

For this project, data related to boroughs and neighbourhood of New York City will be required. To extract data, I will also use the Foursquare API. Latitude and longitude of the locations will be needed mainly. Moreover, for analysis purpose. To create data frames and analyse data, I will use different libraries like pandas, numpy, matplotlib and folium (for mapping). K-mean clustering algorithm will be used for data analysis.

Data Sources

The data for this project will be collected from:

- New York City dataset: This dataset will provide the addresses of neighbourhoods of the city. A json file will be extracted from here.

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

- Foursquare API: It is a location data provider and will be used to make API calls to extract data regarding different venues in New York City's neighbourhoods. To access the API, 'CLIENT_ID', 'CLIENT_SECRET' and 'VERSION' will be defined.

<https://developer.foursquare.com/docs/>

After processing and cleaning the data, it will be analysed through k-means clustering. The data will be segmented into clusters. The clusters will be analysed to determine the most common cuisine(s) in different neighbourhoods of New York City.

Audience/Stakeholders

- The audience of this report are mainly local and foreign tourists in New York City who would like to find the eatery of their choice conveniently or explore the diverse food choices available at a particular locality.
- Findings of this report can also be used by the tourism department of the city as well as by tour operators to guide the tourists regarding the locations of the diverse food options.

References

Food Reference (2020). Food Quotes. John Corry. NY Times Reporter.

NYC and Company (2020). Travel and Tourism trend report. Retrieved from <https://indd.adobe.com/view/e91e777a-c68b-4db1-a609-58664a52cffd>

NYC Mayor Office (2018). State of our immigrant city. Annual Report-2018. Retrieved from
https://www1.nyc.gov/assets/immigrants/downloads/pdf/moia_annual_report_2018_final.pdf

World Urban Areas (2018). Demographia World Urban Areas. Retrieved from
<http://demographia.com/db-worldua.pdf>