**Exploring Chinese Restaurants in NYC**

**IBM Data Science Capstone Project**

**- Simaant Patil**

**INDEX:**

Introduction

Data

Methodology

Results

Discussion

Conclusion

**Introduction:**

The idea behind developing this project was to understand the spread of different cuisines across NYC and evaluate it to solve a business problem. The business problem in this case is figuring out the right location in NYC for a client so that he’s able to open a Chinese restaurant at a place in the city that would yield him maximum profits by attracting more customers.

To solve this problem, it is important to understand the neighbourhoods in the city and to figure out the distribution of variety of cuisines in such neighbourhoods. Analysing and understanding this will provide a lot of insight on the distribution of restaurants as per the cuisines and will thus prove to be very helpful for upcoming restaurants.

**Data:**

The data used here contains information about the neighbourhoods and boroughs present in NYC and is obtained from the following link, <https://geo.nyu.edu/catalog/nyu_2451_34572>, which is the NYU Spatial Data Repository. The data present here is in json format and it is converted into a data-frame with 5 boroughs and 306 neighbourhoods.

Apart from this, the cuisine data for NYC was obtained from Wikipedia from the following website, <https://en.wikipedia.org/wiki/Cuisine_of_New_York_City>. Here, too, the data was in json format and was converted into a data-frame. The data was further cleaned to obtain a data-frame with just names of the major towns like Manhattan, Bronx, Brooklyn and Staten Island and the cuisines in each one of them.

The cuisine data was used to visualize the distribution of cuisines with the help of a word-cloud. Later, this result is used to look at the distribution of restaurants and different public spots in one city.

**Methodology:**

There are two steps that are implemented as a part of the entire methodology for the project

**Step1: Formation of word-clouds as per the cuisines**

The initial data contained information about boroughs, street, zip codes, etc. So, a new data-frame was created with just the columns containing boroughs and the cuisine description. The data frame was then split into the cuisines in Manhattan, Brooklyn, Bronx and Staten Island.

Later, each of these locations is analysed to figure out the distribution of cuisines and then a word cloud is visualized.

**Manhattan Word cloud:**

**A screenshot of a cell phone

Description automatically generated**

**Brooklyn Word cloud:**

**A screenshot of a cell phone

Description automatically generated**

**Bronx Word cloud:**

**A screenshot of a cell phone

Description automatically generated**

**Staten Island Word cloud:**

**A screenshot of a cell phone

Description automatically generated**