COURSERA CAPSTONE

# IBM APPLIED DATA SCIENCE CAPSTONE

## CHINESE RESTAURANT

By Salima Tankibayeva

# Problem

This Capstone Project aims to find out  the best areas to open Chinese restaurants through data analysis and visualization. Location is an important factor that defines viability of the business.

# Data Section

1. Location data of neighborhoods and boroughs

   url: 'https://cocl.us/new_york_dataset'

2. Geocoder package to define coordinates for data visualization
3. Foursquare API, a location data provider,  for venue details

# Methodology

1. First step is extract and manage data in neat data frame. Find out the largest cluster of venues by boroughs and neighborhoods.
2. Filter restaurants by ratings, likes and tips.
3. Visualize data through bar charts and maps.

# Results

1. Top three boroughs by the prevalence of Chinese restaurants:  Manhattan, Brooklyn, and Queens.
2. Top five neighborhoods by the prevalence of Chinese restaurants: Chinatown, Little Neck, Homecrest, Little Italy, and Washington Heights.
3. Top three boroughs with the highest average rating:  Manhattan, Queens, and Brooklyn.
4. Top three neighborhoods with the highest average rating: West Village, Chelsea, and Upper West Side.

# Discussion

We see that many Chinese restaurants are located in Brooklyn, Manhattan, or Queens. Referring to these places as popular locations for Chinese restaurants and using information that there is a huge gap between average ratings we can conclude that a good place to establish a Chinese restaurant is either Brooklyn or Queens, where are weaker opponents.

# Conclusion

Big cities provide enough data to manipulate and analyze which approach can provide more benefits. The pattern shows that best Chinese restaurants are located in Manhattan, however, the amount of restaurants is slightly different from Queens and Brooklyn. Further research needed due to some limitations.