**Recommender System on Contractor Business**

**Richer Ge**

October 21, 2020

1. **Introduction**

**1.1 Background**

In a city of your choice, if a contractor is trying to start their own business in Jinan, where would you recommend that they setup their office? This is a typical clustering problem with dataset labeled.

**1.2 Business and Data Understanding**

Obviously, it should be close to provider and Customer markets. And data needed should be geological location information about specific borough and the neighborhoods in that borough.

We assume it is in Jinan, China. And in China borough and the neighborhoods are not quiet distracted.

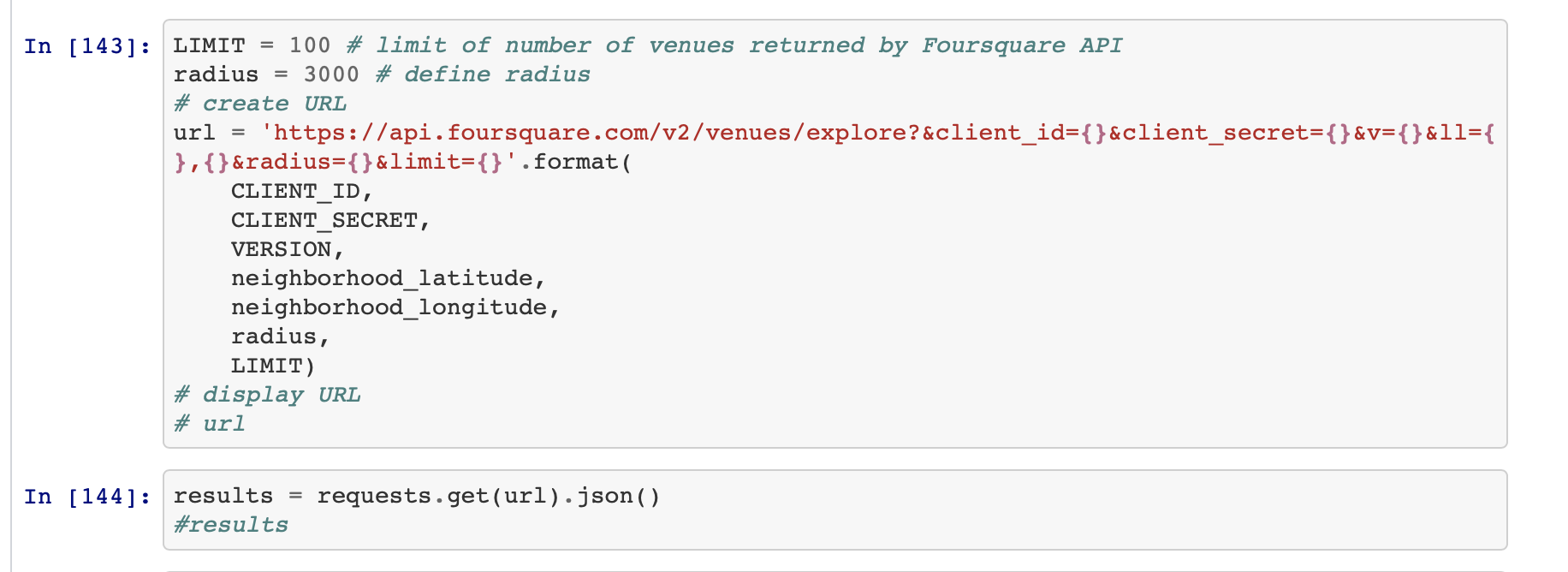
1. **Data acquisition and cleaning**

**2.1 Data Sources**

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. A typical request from Foursquare will be in this format:

1. CLIENT INFORMATION

2. GEO PARAMETERS

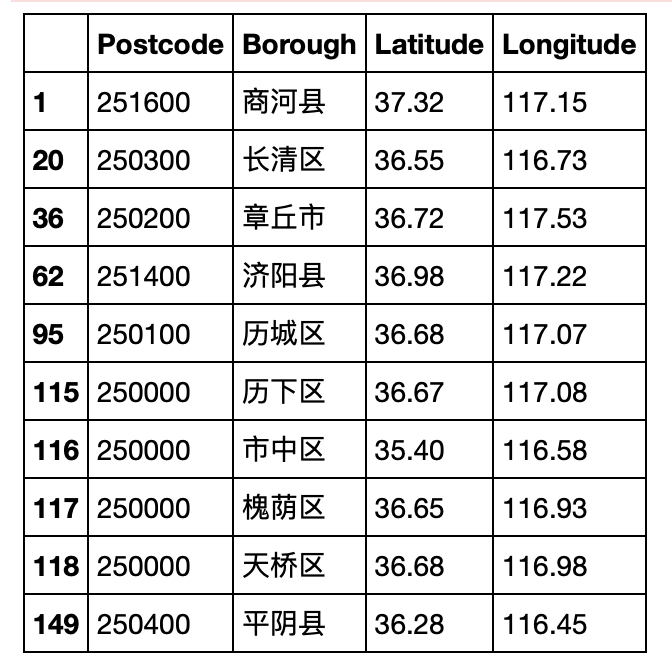
****

What’s more, by requesting from the website about all postcodes in China, I draw the paticular table on 'http://tools.2345.com/yb.htm'.

**2.2 Data cleaning**

Here is the data cleaning part in beautifulsoup to clean the nonetype and make everty elements unique.

****

****

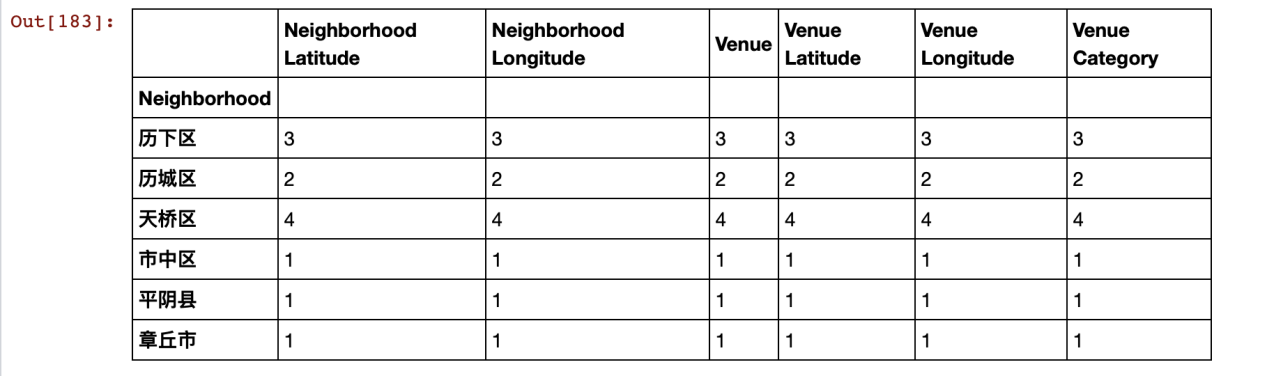
1. **Exploratory Data Analysis**

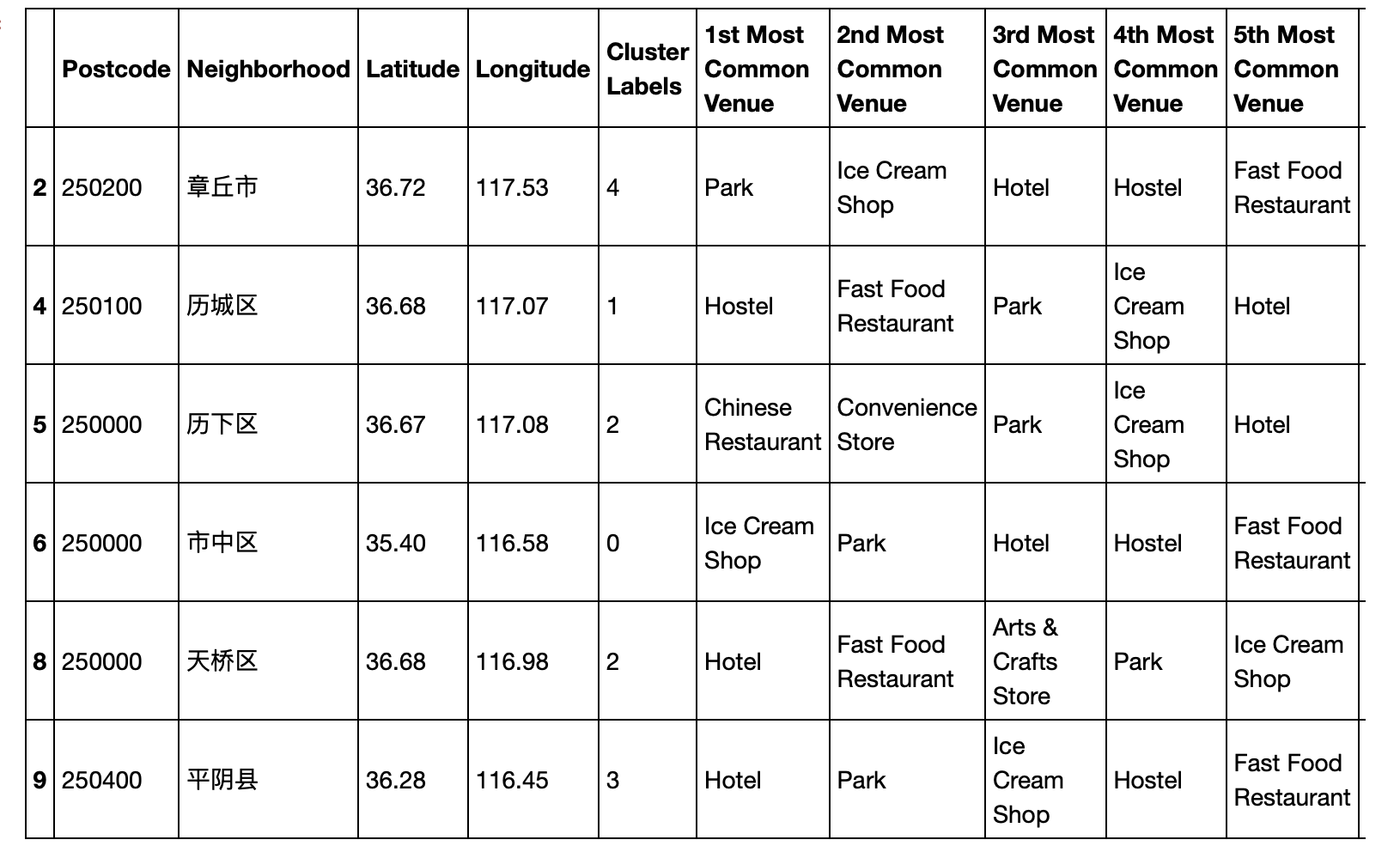
**3.1 Find top 100 in Lixia**

Identifying Postal Codes (and then Neighborhoods) in Jinan.

Connecting to Foursquare and Retrieving Locational Data for Each Venue in Every Neighborhood.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 3000 meter. The two pictures below distributively showing the clusters from the standpoint of geo location and common-sharing.



****

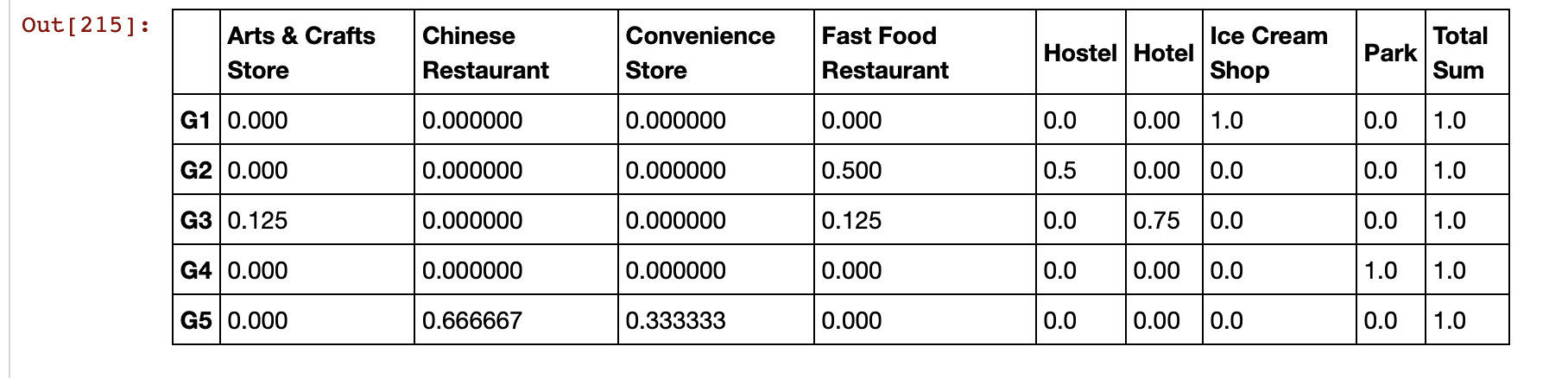
1. **Predictive Modeling**

Processing the Retrieved Data and Creating a DataFrome for All the Venues inside Jinan.

Then One-hot Encoding !

Applying one of Machine Learning Techniques (K-Means Clustering)

Now, we focus on the centers of clusters and compare them for their “Total Types”. The group which its center has the highest "Total Sum" will be our best recommendation to the contractor. {Note: Total Sum = Sum of Total Types}. This algorithm although is pretty straightforward yet is strongly powerful. We can see the below picture.

****

**Here we can easily see G1 has the greastest feature of glocery business.**

1. **Conclusions**

Although some techniques are quite straightforward but powerful. And Although this is quite simple, further more I will use more structural datasets to evaluate NBA games to improve my data analysis ability.

Finally, I have one word for everybody sharing the course: Hard work pays off.