Analysis of Venus from Districts of Shanghai and Beijing and A Tentative Clustering Scheme of Those Districts

Yuhui Cao

May 28, 2020

1. Introduction

1.1 Background

During the past 3 decades, China has been the fastest developing country in terms of culture, economy and education. Shanghai, the center of Chinese economy, and Beijing, the capital city of China, are certainly the two main metropolises that catch eyes of people from all over the world. During the past 3 decades, urban expansion, new immigrants and population influx have been the main issues of both cities. With such issues, there are other new topics presenting to not only government, city planners and entrepreneurs from both internal and external companies, but also original residents of the cities and incoming immigrants. In order to have a more appropriate resource allocation, we need to dig deep into the venues data and give some insights to the publics.

1.2 Problem

In order to have a more appropriate resource allocation, and due to the lack of information of each district and the lack of clustering scheme of all the districts from the two cities, this project, aims here to analyzing venues of each districts and try to cluster those districts based on the similarity of common venues. The venues information will be acquired from Foursquare website, by providing location information of a specific district.

1.3 Interest

Obviously, government, city planners, entrepreneurs, original residents and even new immigrants are interested in the developmental information of each district and comparison between two cities. We can give out some insight information about those topics by using one of the most commonly used unsupervised machine learning techniques, K-Means clustering, on venues information of each district from the two main metropolises of China.