Capstone Project: The Battle of Neighbourhoods

By Harry Ho

1. Introduction

In the US, there are different cities and some types of businesses are the majority of each city. When the entrepreneur would like to set up his/her own business in a city, they need to consider the number of competitors in it and what type of business is the majority of the city. Therefore, they can make a better plan for choosing the city to set up their business. In the report, I will analyze the common venues in different cities by clustering different cities in US on overall types of business in neighbourhoods.

1. Data Analysis

This report will use the data from Wikipedia and Foursquare. The analysis approach is using K-means to cluster the data and display different groups on map.

Dataset

In this dataset, I will examine data from the cities of United Stated which was provided by Wikipedia.

Below is part of this dataset, which includes the city name, latitude and longitude.



Then, I combined the data from Wikipedia and Foursquare to form below table,

一張含有 螢幕擷取畫面 的圖片

自動產生的描述

After that, I find the top ten Common Venue for each cities by analysis.

一張含有 螢幕擷取畫面 的圖片

自動產生的描述

1. Results

I cluster the data to 10 cluster labels by using K-means and plot them on the map.

一張含有 螢幕擷取畫面 的圖片

自動產生的描述

一張含有 文字, 地圖 的圖片

自動產生的描述

1. Conclusion

There is just a limited range of data available on those websites, so that I cannot make an accurate conclusion for the entrepreneur. However, the entrepreneur can use the above map as a reference to choose the cities to set up their own business.