

Introduction

Members of Generation Z have gradually become the main force of the consumption market. They grew up in the era of technology and information explosion, and also witnessed the development of internet giants, like Google, Amazon, Facebook, etc. Obviously, their consumption preferences are different from former generations. They love cool things and love to share that in the internet. Therefore, for those companies which provide cool products and services, having the Generation Z consumers is of great significance to achieve their long-term development and increase market shares.

However, to win the Generation Z consumers is not an easy task. According to consumption characteristics, different companies should improve their products, services, marketing etc. While among all of those, location is a top priority, because a bad location will always discount consumption experiences and satisfaction. This is especially vital for a new opening service.

So, my target audience are those companies who want to win more market shares and have the plan to open several new service sites in cities. To help them make a wise decision on location, I will take advantage of the Foursquare location data and provide valuable suggestions to them.

In this capstone project, we choose San Francisco as our target city. San Francisco is a great city, diverse, dynamic and entrepreneurial. With its economy humming, the city attracts many young people to live and work. After analyzing the location data of San Francisco, we can have a visual sight of the distribution of location, and then recommend some locations to some business.

Data

To successfully finish this capstone project, we need the following kinds of data:

- a) The neighborhoods in San Francisco, and their latitude and longitude coordinates.
- b) The venue data of those neighborhoods.

At first, we can get the list of neighborhoods from Wikipedia via the link: https://en.wikipedia.org/wiki/List_of_neighborhoods_in_San_Francisco using BeautifulSoup library. Then we can use geopy library to get the latitude and longitude values of those neighborhoods. With these data, we can create a map with neighborhoods superimposed on top.

Next, we can apply Foursquare API to get venue data of the above neighborhoods, which are the most important data for our analysis. After analyzing neighborhoods data, we use K-means method to cluster them. Based on this part's work, we present the findings and make the recommendation for some business.