Coursera Capstone Project

IBM Data Science Professional

Searching for a good area to build a new Hotel in Shanghai, China

Author: Lucas Kanz

10.07.2020

Introduction

Vacation is for many cultures and regions in the World a very important thing. One part of a successful vacation is to be able to book a hotel in your favorite destination. As Shanghai is one of the fastest growing economies and also a venue that's getting more and more important every year for people searching for a interesting and beautiful vacation destination, as well as for business people searching for a place to stay.

On the other hand for business developers it's hard to find a good place to open a new venue as the city is getting more and more crowded, and you don't want to make the mistake of building a new, expensive hotel complex into an area that Is already filled with such. Therefore there is a need for a clear and detailed analysis were the best location for such a new opening should be, as it is crucial for the success of the upcoming business.

Business Problem

The goal of this part of the Capstone Project is to find the perfect spot were such a hotel could be placed to be not in the middle of an area already saturated with such venues. It will enable a minimized business risk, as dependency on competitors should be minimized too for a successful start of the new Hotel business the possible customer wants to open.

As building a Hotel complex is very cost intensive usually costing between 30 and 200 million depending on the type of Hotel, the success of it needs to be guaranteed to not lead to a financial disaster for the business developer. Gaining insights through valuable data is a important key to lead the path for a successful future.

Audience

This analysis is especially important **for venue developers, hoteliers as well as large business groups looking for additional ways of generating income through a customer serving business like a Hotel or Hotel Chain**. As more regions in the world get access to more and more wealth and start traveling the world the need for more and well designed Hotels is increasing in the last couple of years together with an increase in competition especially in regions of extraordinary growth.

Data

- Scraping Wikipedia sited to get neighborhoods in Shanghai (*)
- add geo data to these scraped neighborhoods via geocoder (add longitudinal and latitudinal coordinates that are needed for Fourspace API)
- clean the dataset and group by neighborhoods to prepare for clustering

```
: # group
venues_df.groupby(["Neighborhood"]).count()
```

- add venue data by using Fourspace
- take the mean of occurrence of venues
- exclude all other venues except hotels in the dataset
- Cluster neighborhoods using k-means clustering
- Analyse which Cluster Area would be best to open a Hotel to avoid too much competition by creating table and visual data via Folium

We will use web scraping of a Wikipedia Website that's showing neighborhoods in Shanghai:

(*) https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Shanghai

After that Fourspace will help us to find venues nearby and cluster areas with a high amount of Hotels.

One example what could be gathered by the Foursquared API is different kind of venues around a specific area, like in our case 100 venues in a radius of 2500 meters:

[17]:

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Anting	31.2989	121.1576	Alibaba	31.297209	121.162602	German Restaurant
1	Anting	31.2989	121.1576	Wirtshaus	31.291667	121.154532	Bar
2	Anting	31.2989	121.1576	Life Hub (嘉亭荟城市生活广场)	31.289792	121.157673	Shopping Mall
3	Anting	31.2989	121.1576	Starbucks (星巴克)	31.291264	121.142850	Coffee Shop
4	Anting	31.2989	121.1576	KFC (肯德基)	31.297443	121.158709	Fast Food Restaurant

These can include categories of venues just as German Restaurants, Bars, Shopping Malls, Coffee Shops, Night Clubs, etc. In addition Foursquare is also providing the Venue Latitude and Longitude data for a short comparison where within the Neighborhood the venue is located.