My Capstone project introduction: Business analysis to open a shop close to metro stations in Shanghai, China

Introduction of city Shanghai

Shanghai is the economics capital of China. In terms of GDP volume , consumption capacity as an individual city, it has been longly the top one among all cities in China; quoted from Wikipedia " *With a population of 24.28 million of 2019, it is the most populous urban area in China and the second most populous city proper in the world. Shanghai is a global center for finance, innovation and transportation*" The city embraces rich lifestyles, the business facilities to support such variety is substantial. So how to explore the business opportunities in such a giant city is interesting and exciting, especially if we look at the chance to brought with the convenient metro transportation system in Shanghai;

Introduction of Shanghai's metro system

The Shanghai metro system is the world's second largest metro system by route length, reaching total 676Kilometers (420mi), it is also the second largest by the number of metro stations, meanwhile it ranks second in the world by annual ridership;

Business chance exploration and problems

Such gigantic metro transportation not only provides much convenience to the daily life for people living in this city, but also renders big chance to setup commercial facilities as business opportunities, when the metro lines cover most of the city's major residence/commodity areas, more than half of the movements within the city is by metros, people gets food, drink, shopping during their movement along the metro stations, these station spots, on another hands, metro stations provide fast reachability

so people are willing to go there when shopping is considered; Hence usually the café, coffee shop, drink selling, convenient store, shopping malls, restaurants of all flavors do have good flow of people to support their business;

When at the same time, the real estate prices are already significant in such large cities, utilizing the location advantage also brings solid pressure of the housing rental cost for which, when you think of starting a new business around the metro station you also need to check what you have to pay for that commercial estate, these have to be well balanced to allow your investment gets profitable return;

Purpose of the project

In this project, we will go through all the metro stations in Shanghai, by their location geo data of latitude/longitude, we explore each of their nearby venues, finding their current business patterns, classify these patterns into different patterns, then with the information of the commercial estate price level and the opportunities analyzed from the clustering pattern, giving out the recommendation what and where to start your business closing to Shanghai metro stations;

Knowledge/Skills used

- Web scraping: This is to get the raw data of list of Shanghai metro stations, their names and belonging district in Shanghai from public Wikipedia website; Then following each station's link in the list pointing to the introduction of these stations, obtaining each of their geo data (latitude and longitude)
- Foursquare API to get the venues of each station: For each station, with their geodata (latitude, longtitude), using foursquare free account (client username/passwd) to get all venue list of each station, get all venues for each station and put into the pandas dataframe for venue information extract (venue name, category, etc) and processing
- Visualize the map with folium:
  Get the Geo data format of "Shanghai district " from website, converting the Arcgis data format to the GeoJson data format by online tool:
- Data analysis & visualization
  With the venue data obtained from the foursquare API, get information of each metro station's nearby shops, their categories and find the top categories of each station; Using machine learning K-means clustering to cluster the metro stations by their top categories of commercial shops; Analyze the clustering of the metro stations, along with the property price level information of the district the station belongs to, to provide the reference information for business chance exploration and determination (for specific metro stations, what's the current characteristic of the business shops and the chance to open new stores, considering the property price level nearby the metro station)