**Predicting the Improvement of NBA players**

Parth Joshi

April 20, 2020

**1. Introduction**

**1.1 Scenario and Background**

I am a tech enthusiast / student currently residing in Chennai.I currently live within walking distance to CMRL metro station therefore I have access to good public transportation to work. Likewise, I enjoy many ammenities in the neighborhood , such as international cousine restaurants, cafes, food shops and entertainment. I have been offered a internship at california. Although, I am very excited about it, I am a bit stress toward the process to secure a comparable place to live in California. Therefore, I decided to apply the learned skills during the Coursera course to explore ways to make sure my decision is factual and rewarding. Of course, there are alternatives to achieve the answer using available Google and Social media tools, but it rewarding doing it myself with learned tools.

**1.2 Problem to be resolved:**

The challenge to resolve is being able to find a rental apartment unit in California that offers similar characteristics and benefits to my current situation. Therefore, in order to set a basis for comparison, I want to find a renta unit subject to the following conditions:

- Apartment with min 2 bedrooms with monthly rent not to exceed US$5000/month

- Unit located within walking distance (<=1.0 mile, 1.6 km) from a subway metro station in California.

- Area with ammenities and venues similar to the ones described for current.

**1.3 Interested Audience**

I believe this is a relevant project for a person or entity considering moving to a major city in Europe, US or Asia, since the approach and methodologies used here are applicable in all cases. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case toward the development of Data Science skills.

**2. Data Section**

**2.1 Data of Current Situation**

I Currently reside in the neighborhood of Annanagar in Chennai. I use Foursquare to identify the venues around the area of residence which are then shown in the Chennai map shown in methodology and execution in section 3.0 . It serves as a reference for comparison with the desired future location in California.

**2.2 Data Required to resolve the problem**

In order to make a good choice of a similar apartment in California, the following data is required:

* List/Information on neighborhoods form california with their Geodata ( latitud and longitude ).
* List/Information about the subway metro stations in California with geodata.
* Listed apartments for rent in California area with descriptions ( how many beds, price, location, address)
* Venues and ammenities in the Calfornia neighborhoods

**2.3 sources and manipulation**

The list of california neighborhoods is worked out during LAb exercise during the course. A csv file was created which will be read in order to create a dataframe and its mapping. The csv file 'cal\_neigh\_data.csv' has the following below data structure. The file will be directly read to the Jupiter Notebook for convenience and space savings. The clustering of neighborhoods and mapping will be shown however. An algorithm was used to determine the geodata from Nominatim . The actual algorithm coding may be shown in 'markdown' mode becasues it takes time to run.

Borough Neighborhood Latitude Longitude

35 Calfornia Turtle Bay 40.752042 -73.967708

36 Calfornia Tudor City 40.746917 -73.971219

37 Calfornia Stuyvesant Town 40.731000 -73.974052

38 Calfornia Flatiron 40.739673 -73.990947

39 Calfornia Hudson Yards 40.756658 -74.000111

A list of calfornia subway metro stops was complied in Numbers (Apple excel) and it was complemented with wikipedia data (https://en.wikipedia.org/wiki/List\_of\_Los\_Angeles\_Metro\_Rail\_stations) and information from LAX Transit authority and Google maps (https://www.google.com/maps/place/LAX/@33.9429151,118.4116813,17z/data=!3m1!4b1!4m5!3m4!1s0x80c2b11ca96d81fb:0x80d7cbf96e7a68e7!8m2!3d33.9429151!4d-118.4094926) for a final consolidated list of subway stops names and their address. The geolocation was obtained via an algorithm using Nominatim. Details will be shown in the execution of methodolody in section 3.0.

A list of places for rent was collected by web-browsing real estate companies in california :

<http://www.rentcalifornia.com/index.cfm?page=search&state=results>

<https://www.nestpick.com/search?city=new->york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAgEiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv\_MvKcrIhRWRzNkc\_r-fGi0lxoCA7cQAvD\_BwE&type=apartment&display=list

https://www.realtor.com/apartments/California\_CA

A csv file was compiled with the rental place that indicated: areas of california, address, number of beds, area and monthly rental price. The csv file "cal.csv" had the following below structure. An algorithm was used to create all the geodata using Nominatim, as shown in section 3.0. The actual algorithm coding may be shown in 'markdown' mode becasues it takes time to run. With the use of geolocator = Nominatim() , it was possible to determine the latitude and longiude for the subway metro locations as well as for the geodata for each rental place listed. The loop algorithms used are shown in the execution of data in section 3.0

"Great\_circle" function from geolocator was used to calculate distances between two points , as in the case to calculate average rent price for units around each subway station and at 1.6 km radius.

Foursquare is used to find the avenues at california neighborhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

**2.4 How the data will be used to solve the problem**

The data will be used as follows:

Use Foursquare and geopy data to map top 10 venues for all california neighborhoods and clustered in groups ( as per Course LAB)

Use foursquare and geopy data to map the location of subway metro stations , separately and on top of the above clustered map in order to be able to identify the venues and ammenities near each metro station, or explore each subway location separately

Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations.

create a map that depicts, for instance, the average rental price per square ft, around a radious of 1.0 mile (1.6 km) around each subway station - or a similar metrics. I will be able to quickly point to the popups to know the relative price per subway area.

Addresses from rental locations will be converted to geodata( lat, long) using Geopy-distance and Nominatim.

Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies such as Metro, etc.

**2.5 Mapping of Data**

The following maps were created to facilitate the analysis and the choice of the palace to live.

California map of Neighborhoods

California subway metro locations

California map of places for rent

California map of clustered venues and neighborhoods

Combined maps of California rent places with subway locations

Combined maps of California rent places with subway locations and venues clusters.

**3. Methodology section:**

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

**The analysis and the stragegy:**

The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed : location near a subway, rental price and similar venues to Chennai. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.

The procesing of these DATA and its mapping will allow to answer the key questions to make a decision:

- what is the cost of available rental places that meet the demands?

- what is the cost of rent around a mile radius from each subway metro station?

- what is the area of California with best rental pricing that meets criteria established?

- What is the distance from work place and the tentative future rental home?

- What are the venues of the two best places to live? How the prices compare?

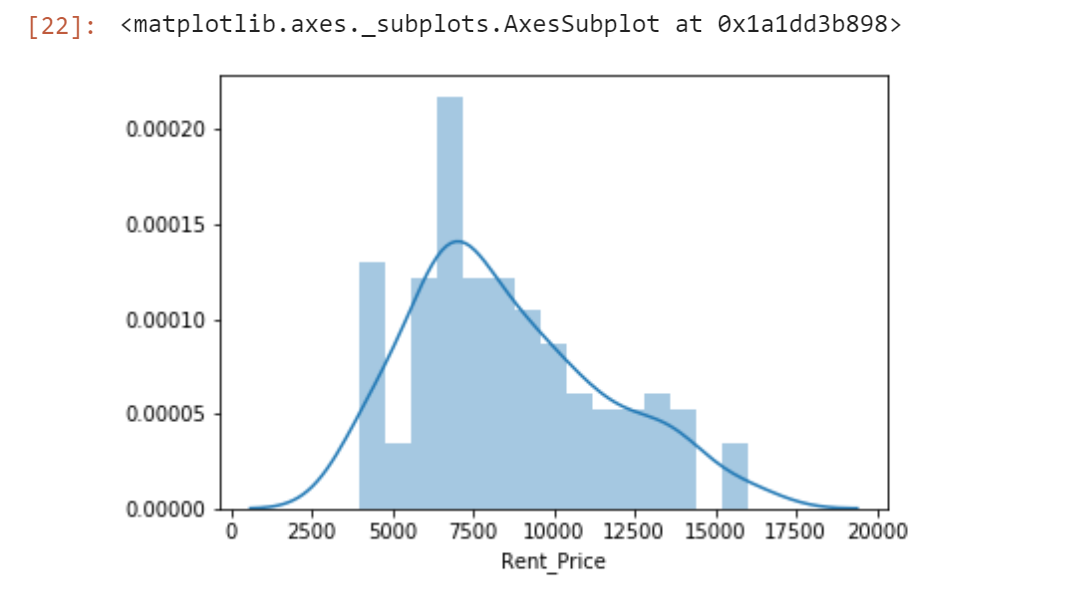
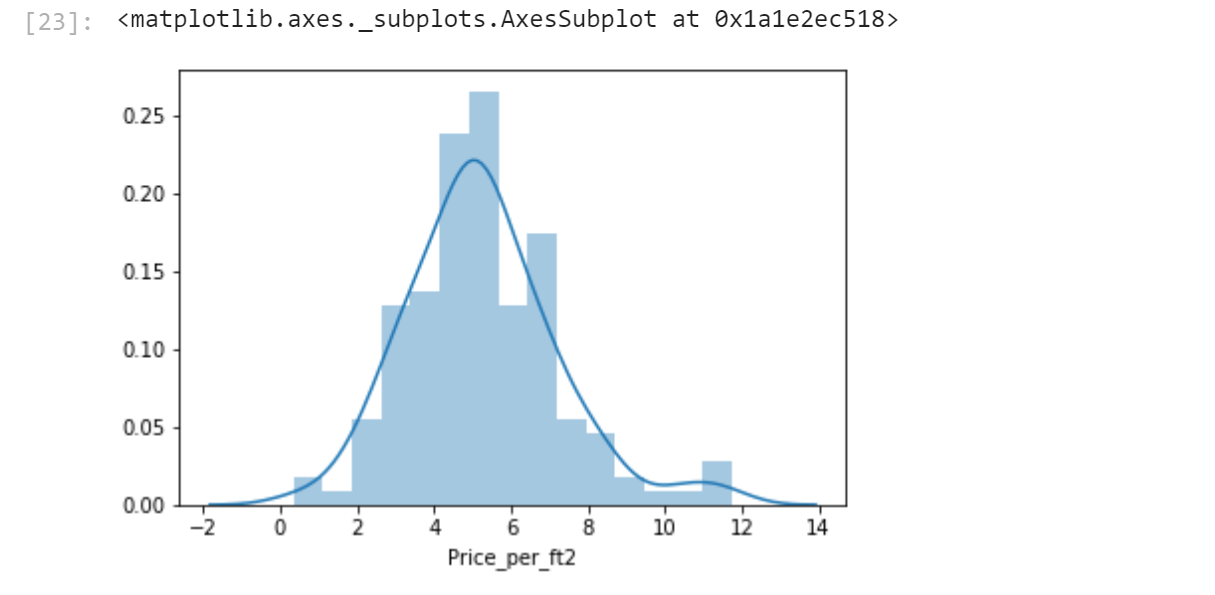
- How venues distribute among California neighborhoods and around metro stations?

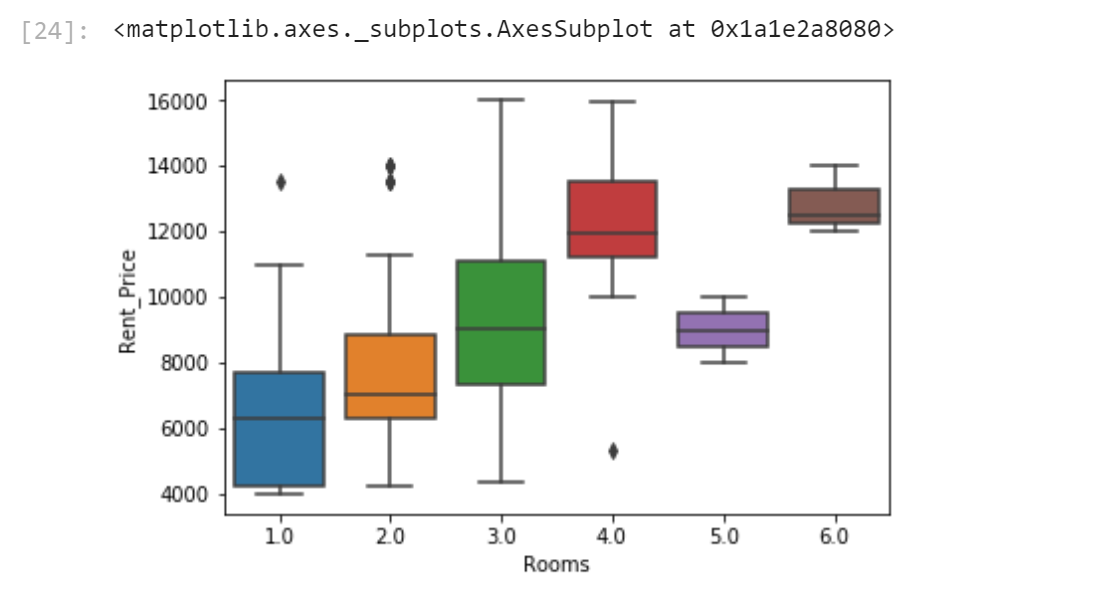
- Are there tradeoffs between size and price and location?

- Any other interesting statistical data findings of the real estate and overall data.

**4.0 Results**

**ONE CONSOLIDATE MAP**

Let's consolidate all the required inforamtion to make the apartment selection in one map. 



Apartment 1 rent cost is US5500 slightly above the US5000 budget. Apt 1 is located 400 meters from subway station at 59th Street and work place ( Park Ave and 53rd) is another 600 meters way. I can walk to work place and use subway for other places aroung. Venues for this apt are as of Cluster 2 and it is located in a fine district in the East side of California.

Apartment 2 rent cost is US6935, just under the US5000 budget. Apt 2 is located 60 meters from subway station at Fulton Street, but I will have to ride the subway daily to work , possibly 40-60 min ride. Venues for this apt are as of Cluster 3.¶

Based on current Chennai venues, I feel that Cluster 2 type of venues is a closer resemblance to my current place. That means that Apartment 1 is a better choice since the extra monthly rent is worth the conveniences it provides.

**5.0 DISCUSSION**

I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned. I have created a good project that I can present as an example to show my potential. I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases.

**6.0 CONCLUSION**

I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation.This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations.One must keep abreast of new tools for DS that continue to appear for application in several business fields.