Let’s Make a Student’s Life Easier

Data Science Capstone - IBM Data Science Professional Certificate on Coursera

SAPNA TARE

Introduction

Student move to Metropolitan cities for higher educations. Moving to metro cities is extremely exciting for students, but it can also be a little overwhelming. Since these cities very crowded. Students face the problem of commute to reach their destinations. Students having part time job/night time classes struggle to reach destination from universities/colleges on time.

Solution: -

For the problem, Let’s say XYZ company decided to construct off-campus /student - housing community (campus for students) /dorms to nearby universities/colleges in different states across the USA.

Unless students who live on campus work away from the school, there is little need for transportation and commute is biggest problem. Off Campus housing should be close to universities/collages. It should be within walking distance to classrooms, libraries, university center. It should be closer to other places such as book store, a variety of food locations, and campus recreational facilities. It will save quality time of student to avoid commute

This report can help to companies those are planning to start building an off campus / Dorms for student in a metropolitan city.

XYZ company has to figure out all the neighborhood locations where top universities are exists and start their business around.

The problem we aim to analyze the Universities locations across the USA and find the best place for our students to stay and make their life easier.

Data

For this project we need the following Data

Foursquare

https://api.foursquare.com/v2/venues/explore

Top 5 states in USA for Higher educations: -

1. New York
2. Boston, MA
3. Chicago, IL
4. Jersey City, NJ
5. San Francisco, CA

Methodologies: -

Expletory Analysis

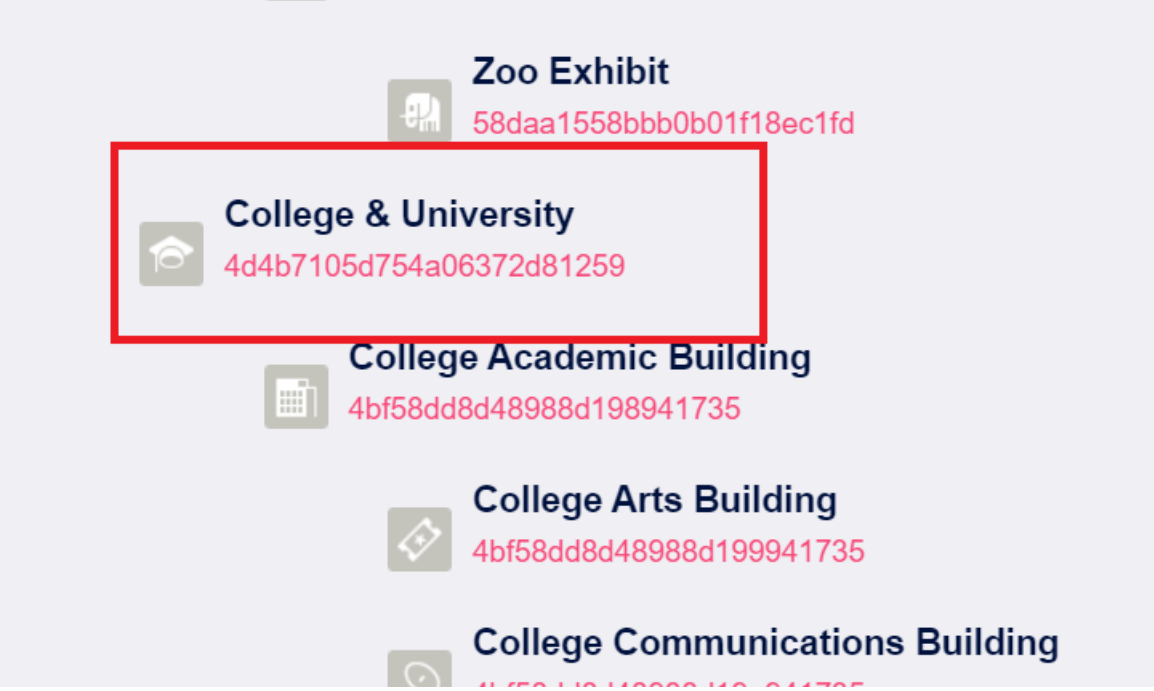
As a data scientist what would be my approach to solve this business Problem.

Main target here is to asses which City would have the highest universities locations density. I used the Four-Square API through the venues channel. I used the near query to get venues in the cities. Also, I use the CategoryID to set it to show only targeted places.

Example: -

Find Category ID

https://developer.foursquare.com/docs/resources/categories



Find Venue using Category ID
https://api.foursquare.com/v2/venues/explore?&client_id=&client_secret=&v=20180605&New York, NY&limit=100&categoryId= **4d4b7105d754a06372d81259**

Using Category Id for collage /universities from Foursquare API we fetch the venues which limits us to maximum of 100 venues per query.

Moreover, I repeated this request for the 5 major targeted cities and got their top 100 venues. I saved the name and coordinate data only from the result and plotted them on the map for visual inspection.

Next, to get an indicator of the density of colleges/universities Places where XYZ companies targeted to start construction for off campus for students, I calculated a center coordinate of the venues to get the mean longitude and latitude values. Then I calculated the mean of the Euclidean distance from each venue to the mean coordinates. That was my indicator; mean distance to the mean coordinate

1. Foursquare API
2. IBM Watson cloud -python
   Libraries needed
   - Numpy and pandas: - Data handing
   - geoPy: - to get coordinates latitude and longitude information
   - folium: - to visualize result on map

Results: -

1. **Total number of Universities in targeted locations.** 1

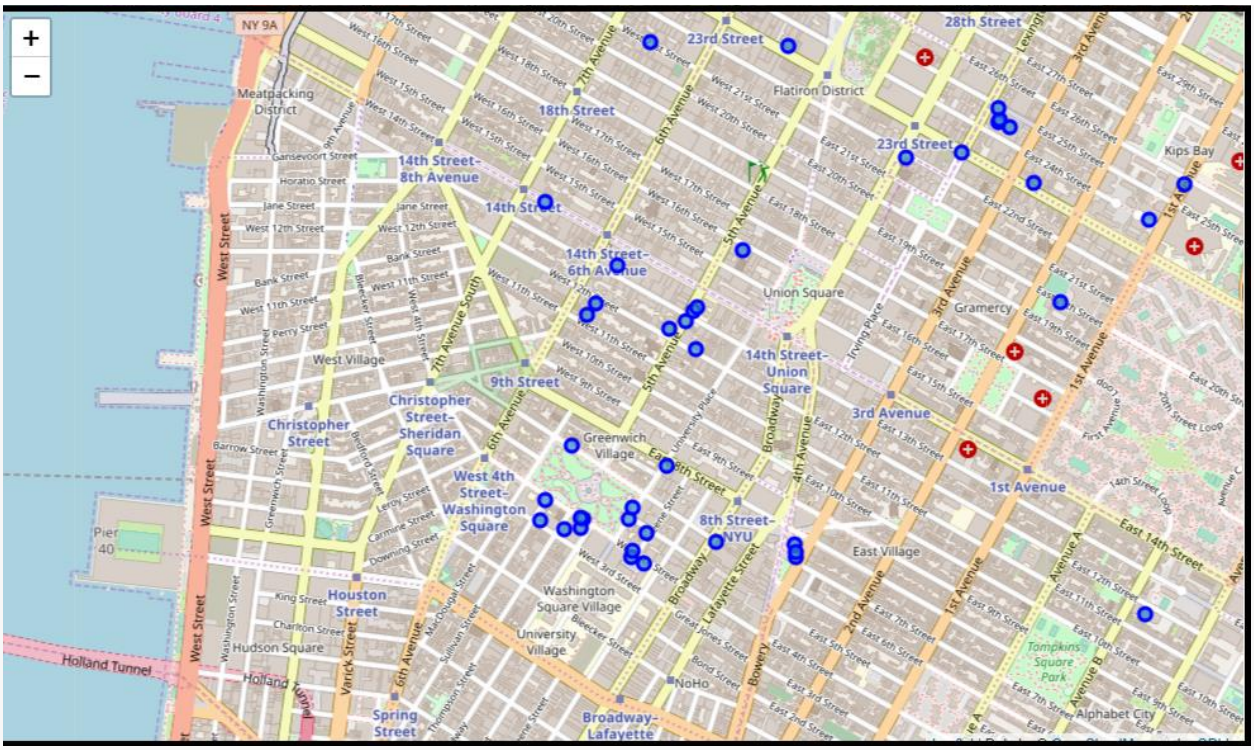
```
Total number of Collages and universities places in New York, NY =  244
Showing Top 100
Total number of Collages and universities places in Chicago, IL =  240
Showing Top 100
Total number of Collages and universities places in San Francisco, CA =  225
Showing Top 100
Total number of Collages and universities places in Jersey City, NJ =  132
Showing Top 100
Total number of Collages and universities places in Boston, MA =  244
Showing Top 100
```

2. **Mean distance from Mean coordinates**

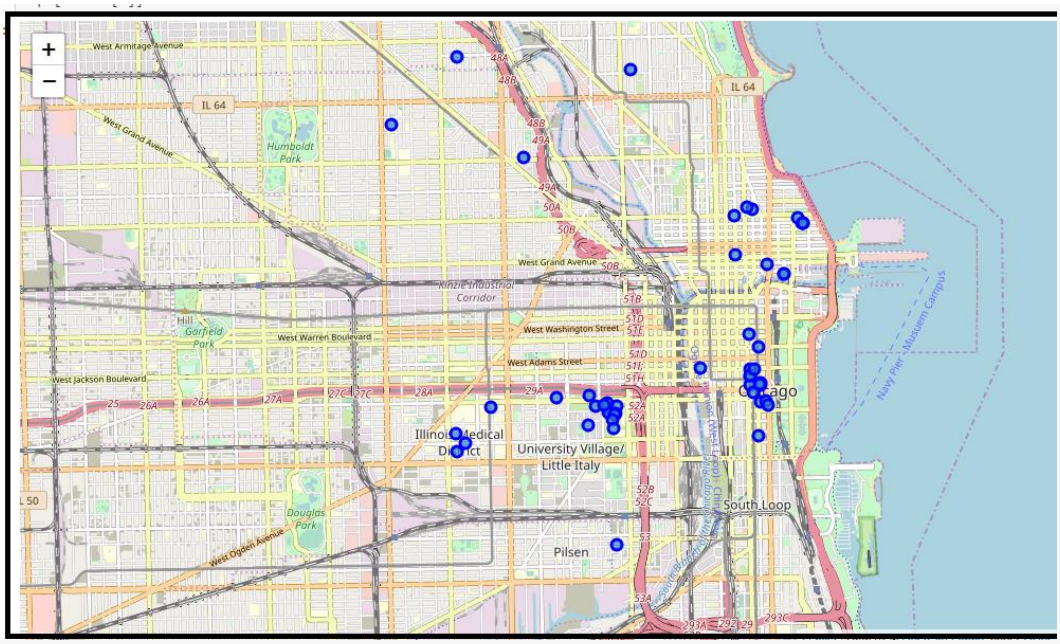
```
New York, NY
Mean Distance from Mean coordinates
0.026473758343131547
Chicago, IL
Mean Distance from Mean coordinates
0.06117101621818162
San Francisco, CA
Mean Distance from Mean coordinates
0.03598265681770688
Jersey City, NJ
Mean Distance from Mean coordinates
0.013631532866919085
Boston, MA
Mean Distance from Mean coordinates
0.02325207138763866
```

Visualize Result

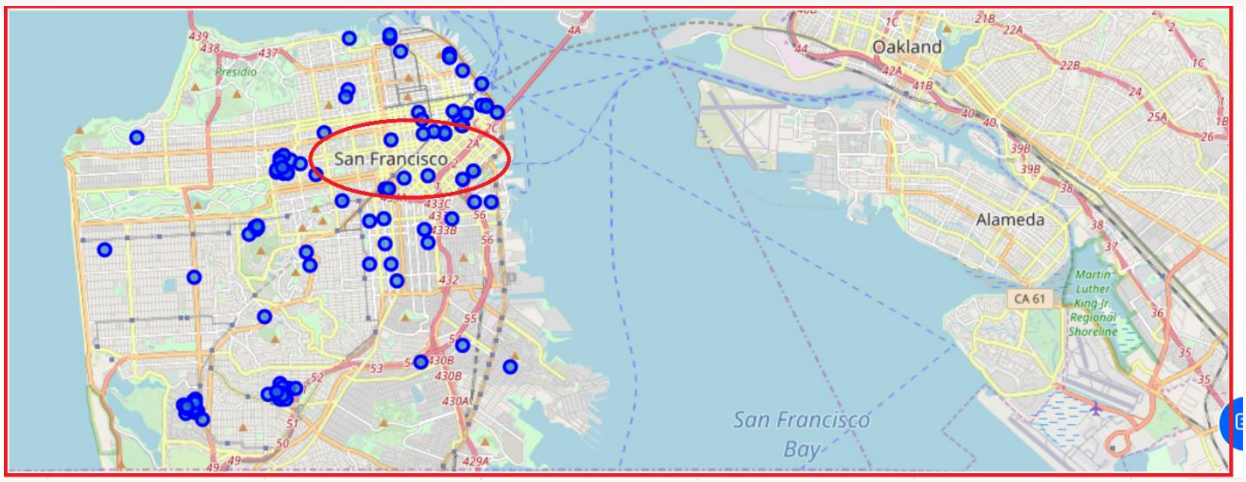
New York:


NY 9A
Meatpacking District
West Street
14th Street–8th Avenue
14th Street
14th Street–6th Avenue
West Village
Christopher Street
Christopher Street–Sheridan Square
9th Street
West 4th Street–Washington Square
Houston Street
Hudson Square
Holland Tunnel
Pier 40
Spring Street
Broadway–Lafayette
NoHo
University Village
Washington Square Village
Greenwich Village
8th Street–NYU
East Village
Union Square
14th Street–Union Square
Flatiron District
23rd Street
18th Street
28th Street
Gramercy
3rd Avenue
1st Avenue
Kips Bay
Tompkins Square Park
Alphabet City
East 14th Street

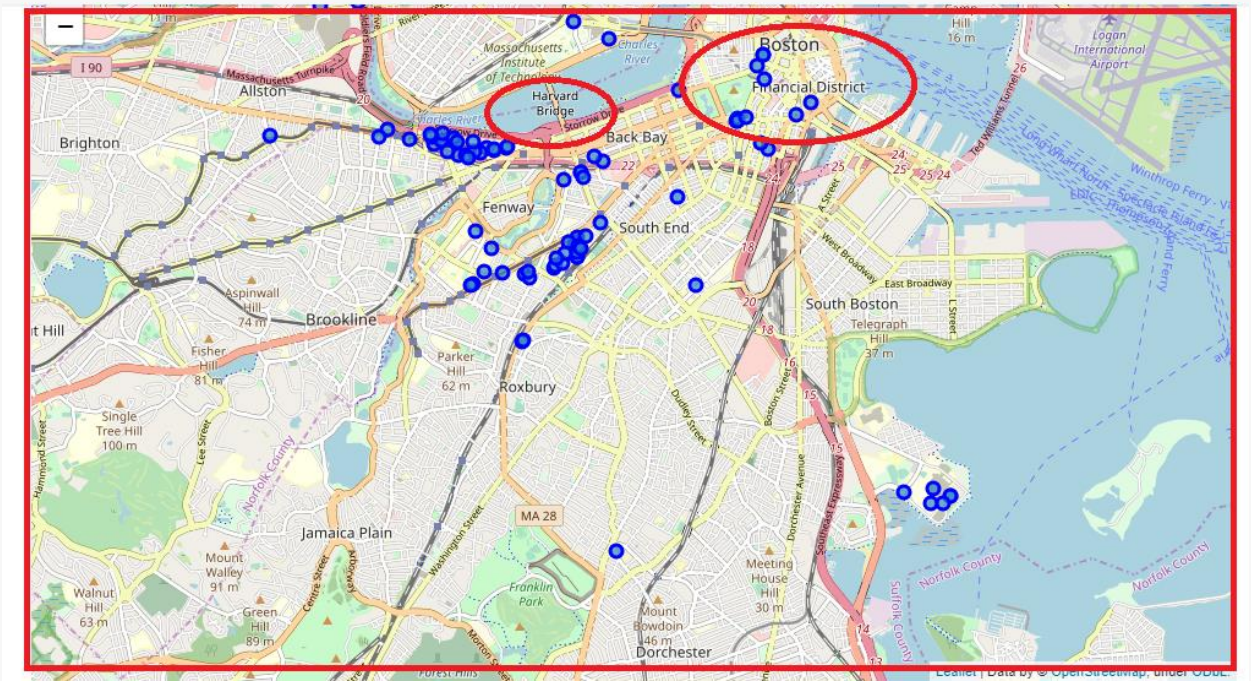
Chicago, IL



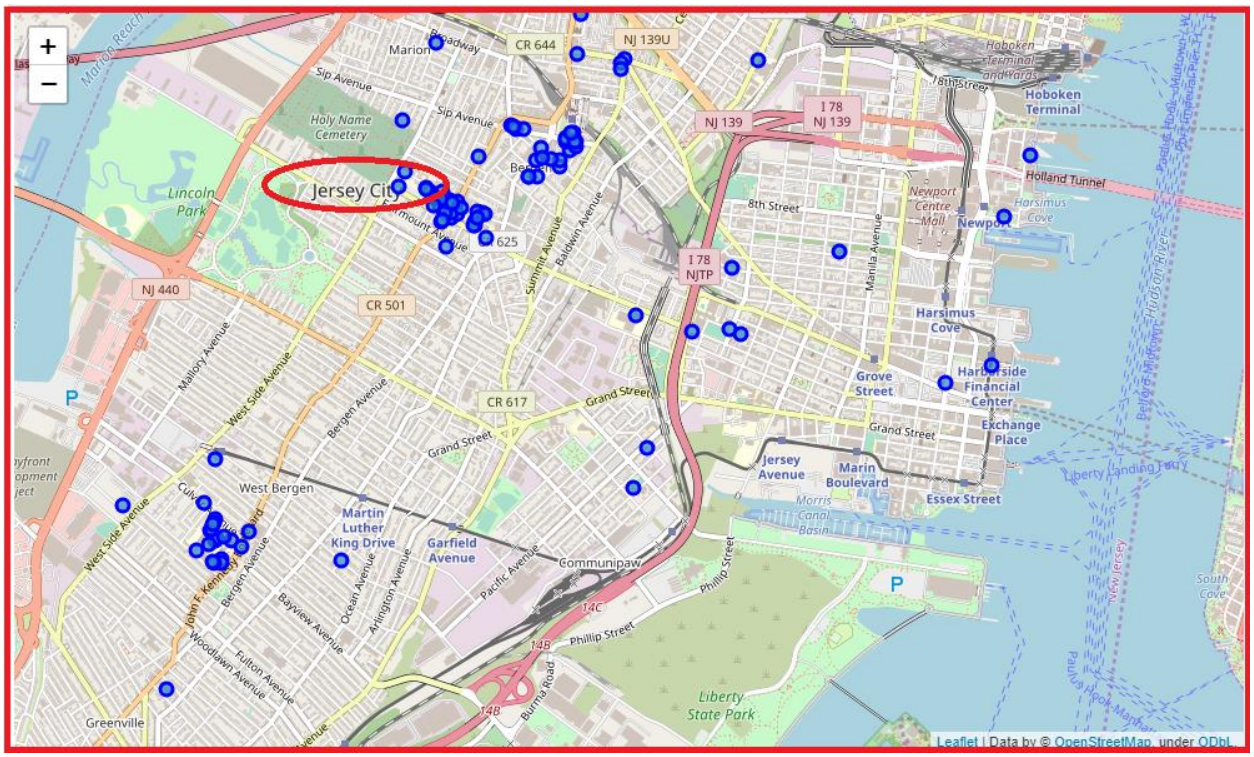
San Francisco, CA



Boston, MA



Jersey City, NJ



Therefore, our results are:

1. New York (244)
2. Boston, MA (244)
3. San Francisco (225)
4. Chicago (240)
5. Jersey City (132)

Discussion:

Observation 1: - Based on the above research for XYZ company. I conclude that company should target first New York and Boston have more universities and colleges than other cities. In such city's student life would be easier in terms of commute/transport. That is good idea to construct dorms and off campus in such locations. and We get more clarity using Data Visualization.

Observation 2: - I noticed in the figure is that there is a really far away universities in San Francisco that is probably giving it a higher Mean. So, I checked what if I removed it,

it would not harm anyone. The new MDMC was: 0.0519953, putting it one place up on the list replacing Chicago, IL. One consideration to do further work on is to move the location of the Foursquare API query until we get all the universities places in each city and do the calculations again.

Conclusion:

Based on the analysis conclude that company should target first New York and Boston because of more universities and colleges than other cities. In such city's student life would be easier in terms of commute/transport. That is good idea to construct dorms and off campus. Student prefer off campus and housing which are closer to collages and other facilities this decision is profitable to company because more student.

I have learned lot of techniques and modeling in this full course. This is great level of learning and lot of hands on labs will helpful to build tiny concepts from scratch (ML models recommender system, Python, IBM Watson, IBM SkillNet, folium). With this course we got better understanding of neighborhood with respect to most common venues using Foursquare API.

In future we can explore more analysis with respect to cost and other factors to run and start any business using ML techniques.