

[Open in app](#)

Parth Shukla

5 Followers [About](#)

Building solutions for housing and business problems using Data Science

 Parth Shukla · Just now · 4 min read

Note: This is a part of the capstone project for the IBM Data Science Professional Certificate.

Introduction

Mumbai (also known as Bombay, the official name until 1995) is the capital city of the Indian state of Maharashtra. According to the United Nations, as of 2018, Mumbai is the second-most populous city in the country after Delhi and the seventh-most populous city in the world with a population of roughly 20 million.

Space is limited in Mumbai and it is one of the densest cities in the world. It is because of this lack of space that we need to find a solution utilizing the power of Data Science to find the right housing conditions according to personal requirements.

Mumbai is India's largest city (by population) and is the financial and commercial capital of the country as it generates 6.16% of the total GDP. Mumbai is a city that never sleeps and its economy is always booming. Competition is high but the opportunities are endless. We need to find a way to find a solution for maximizing opportunity utilizing Data Science.

Data Description

I am using the data available by 99acres.com, and I am referring to https://github.com/dibyendutapadar/mumbai_city_analysis who has done an

excellent job of collecting and cleaning the data.

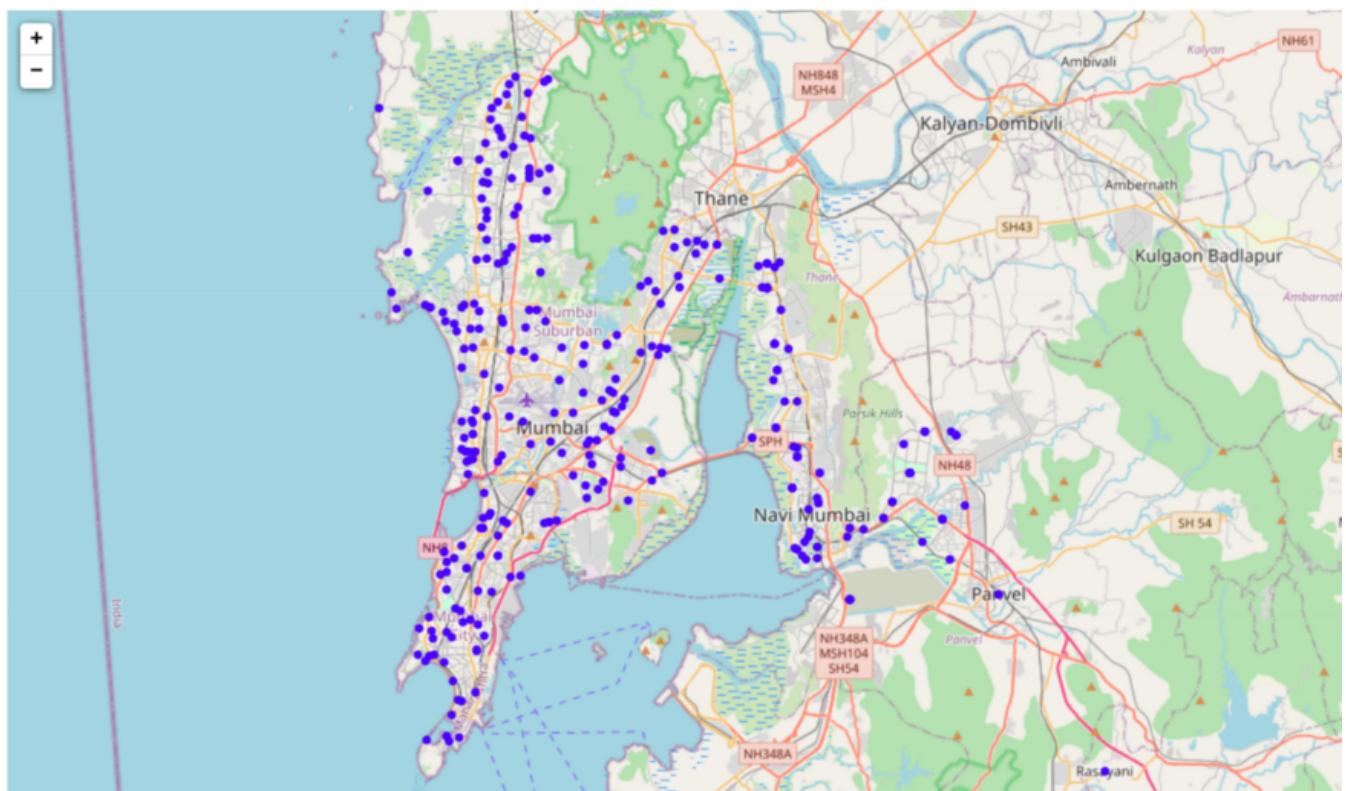
- Property rates are scrapped from 99acres.com. 99Acres is an Indian real estate database website founded in 2005.
- Nominatim API to get coordinates of neighbourhoods under analysis.
- Four Square API to get the details and types of venues in the vicinity of a neighbourhood.

Methodology

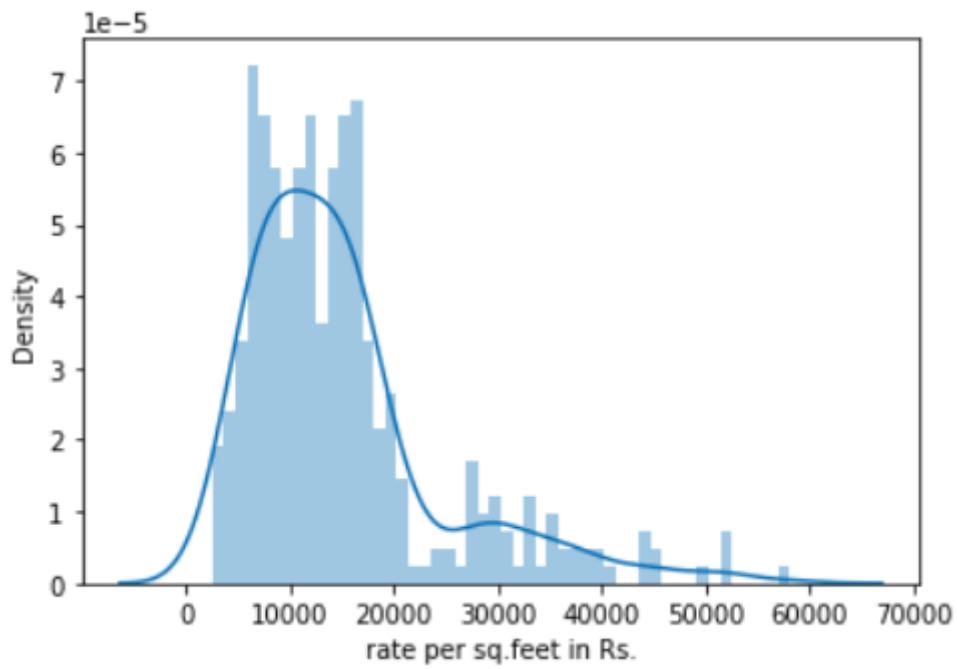
The dataset after compiling looks something like this.

Locality Name	buy_rate	buy_rate_trim	buy_rate_lower	buy_rate_higher	buy_rate_avg	buy Rates Q/Q	Rent_1B	Rent_1B_trim	...	Rent_2B_avg	Rent_2B_persqf	Rent_3B	Rent_3B_t
Worli	Rs.32,852 - 38,462/sq. ft.	32,852 - 38,462	32852.0	38462.0	35657.0	2.48%	Rs.26,392 - 34,808	26,392 - 34,808	...	81664.0	229%	Rs.1,54,460 - 1,95,769	1,54,460 - 1,95,769
Y K Nagar	Rs.5,100 - 5,398/sq. ft.	5,100 - 5,398	5100.0	5398.0	5249.0	0%	NaN	NaN	...	NaN	NaN	NaN	NaN
Yari Road	Rs.18,530 - 20,825/sq. ft.	18,530 - 20,825	18530.0	20825.0	19677.5	-1.69%	Rs.26,775 - 31,025	26,775 - 31,025	...	49929.0	254%	NaN	NaN
Yogi Nagar	Rs.16,448 - 17,808/sq. ft.	16,448 - 17,808	16448.0	17808.0	17128.0	-0.48%	NaN	NaN	...	NaN	NaN	NaN	NaN
Yogidham	Rs.5,525 - 6,418/sq. ft.	5,525 - 6,418	5525.0	6418.0	5971.5	7.81%	Rs.6,818 - 8,391	6,818 - 8,391	...	NaN	NaN	NaN	NaN

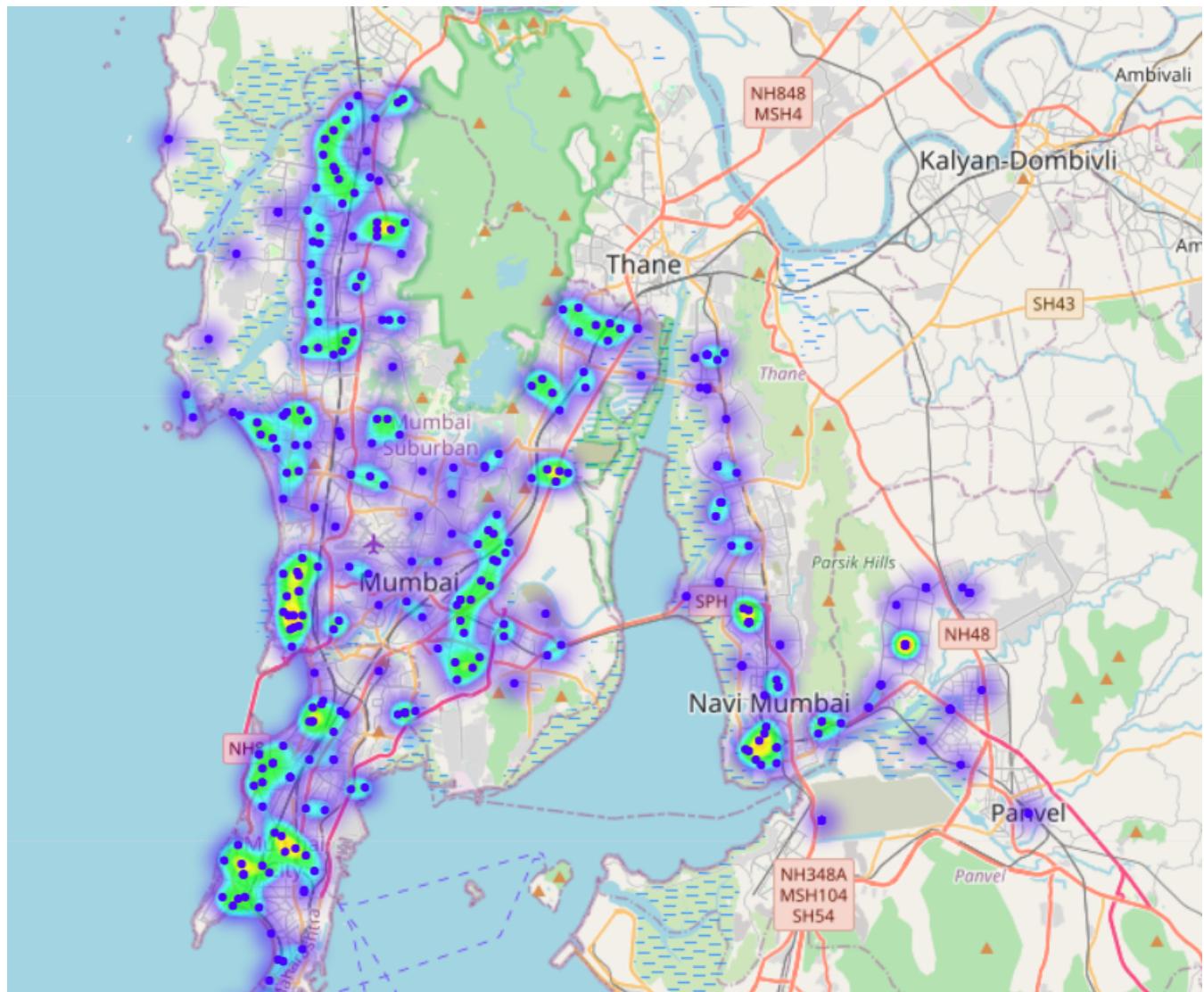
Next, we use folium to visualize all the locations we have.



We look at the distribution of housing prices using distplot from the seaborn library.

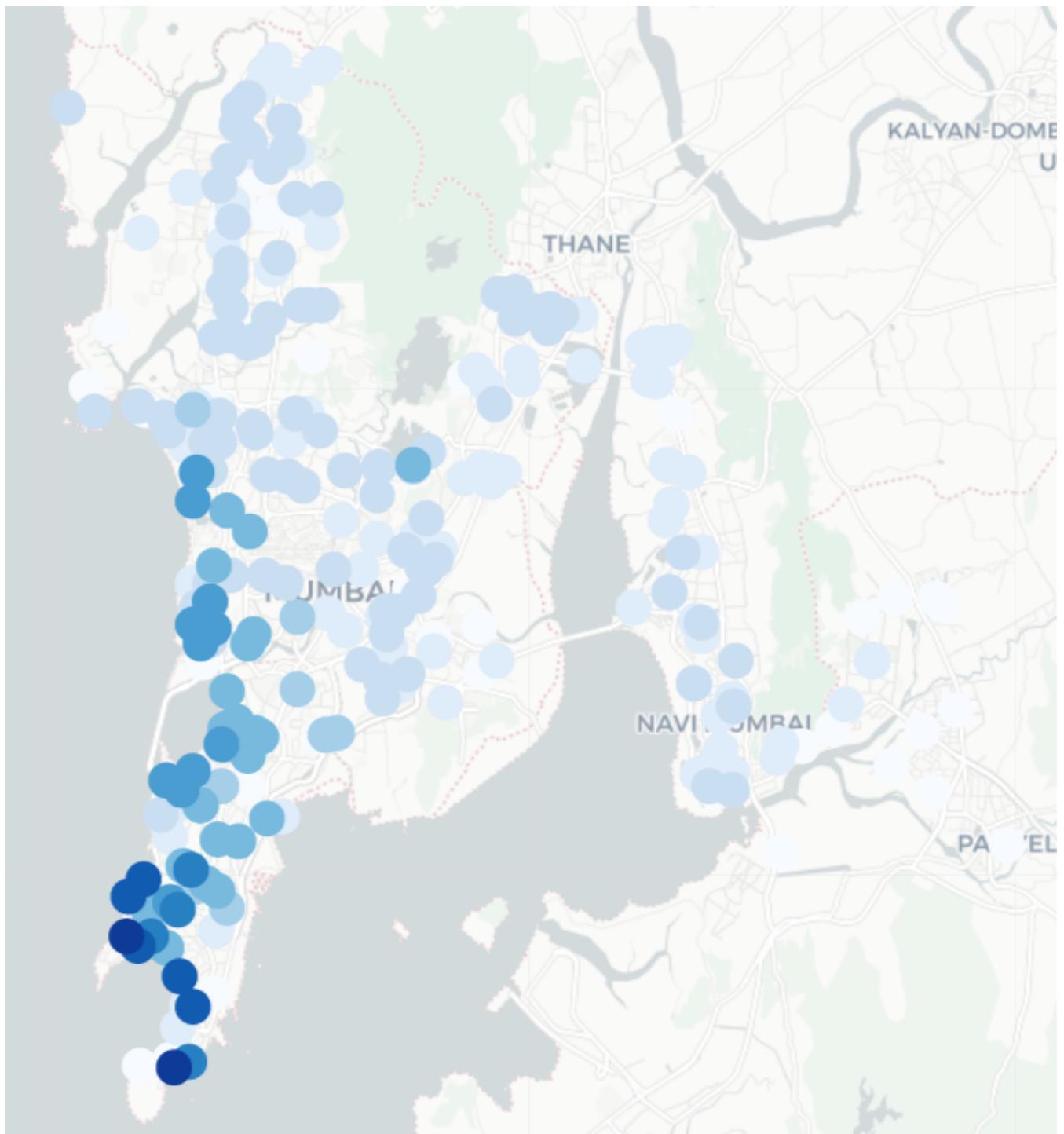


We use a heatmap to visualize the distribution of prices of houses in Mumbai.





Alternate form to visualize the prices.



Solution

For housing solution, I highly recommend you to go and check out [this repository](#).

Business solution

For business solutions, we first consider a business that you want to open up. For example, let's say you want to open a Nightclub that targets high paying customers. For this, the club should be in an area where the rent is more which indicates a high earning-high paying customer base. At the same time, you would want to consider areas where competition is less. So, ideally, you should target areas that do not contain any premium nightclubs similar to the one you want to open.

Here's the code for that(I promise this would make more sense once you check the repo). For this example, we have considered premium areas to be the areas where the rent is higher than 17,000(which was almost the mean of rent in Mumbai). Obviously, this can be varied according to the use case.

```
filtered = onehot_venues[onehot_venues[business] == 0]

areas = onehot_venues['Neighborhood'].unique() # Areas in Mumbai

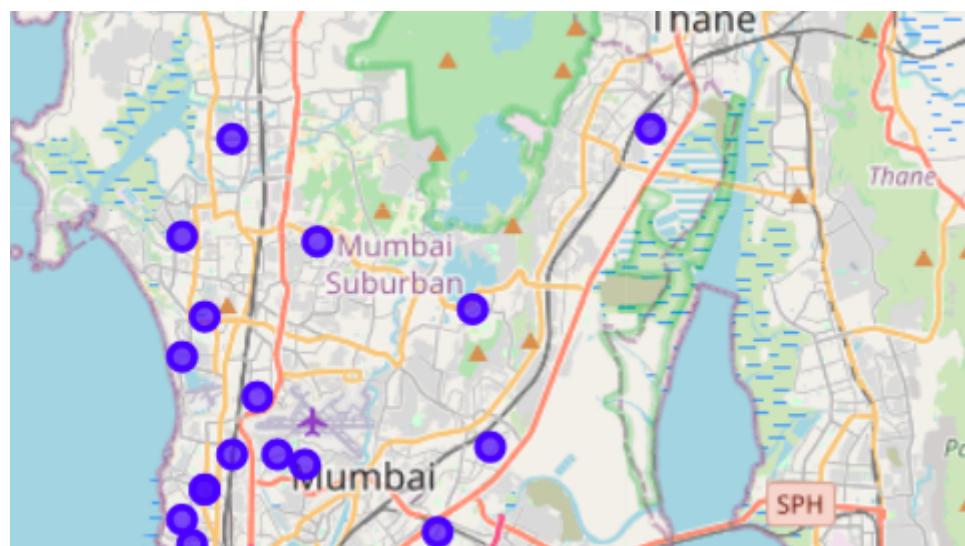
useful_areas = filtered['Neighborhood'].unique()

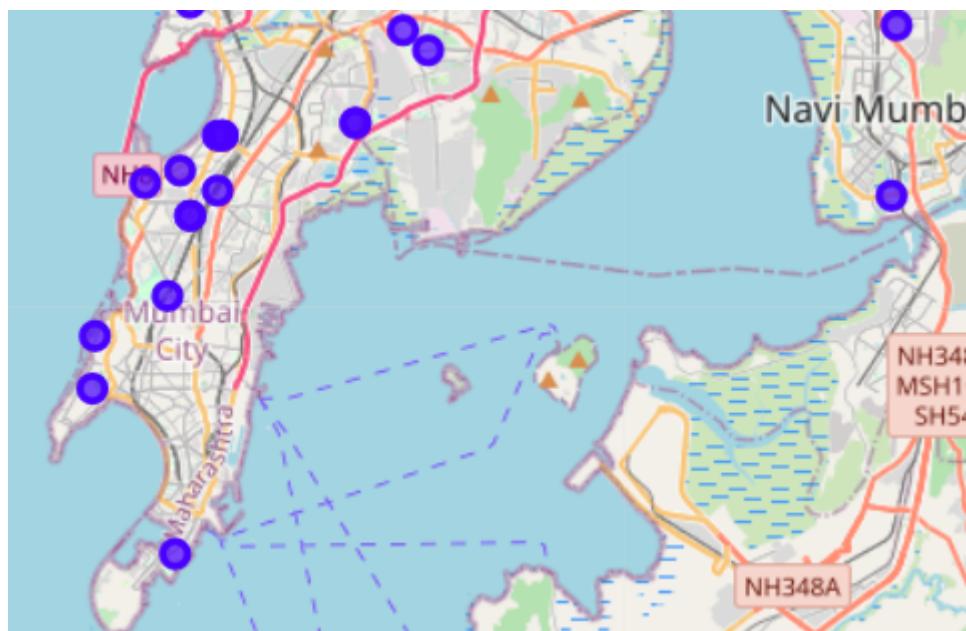
areas_df = pd.DataFrame(columns=['Area', 'Latitude', 'Longitude', 'Avg_rate'])
areas_df['Area'] = useful_areas

for index, row in areas_df.iterrows():
    df = onehot_venues[onehot_venues['Neighborhood'] == row['Area']]
    row['Latitude'] = df['Neighborhood Latitude'].mean()
    row['Longitude'] = df['Neighborhood Longitude'].mean()
    row['Avg_rate'] = df['buy_rate_avg'].mean()

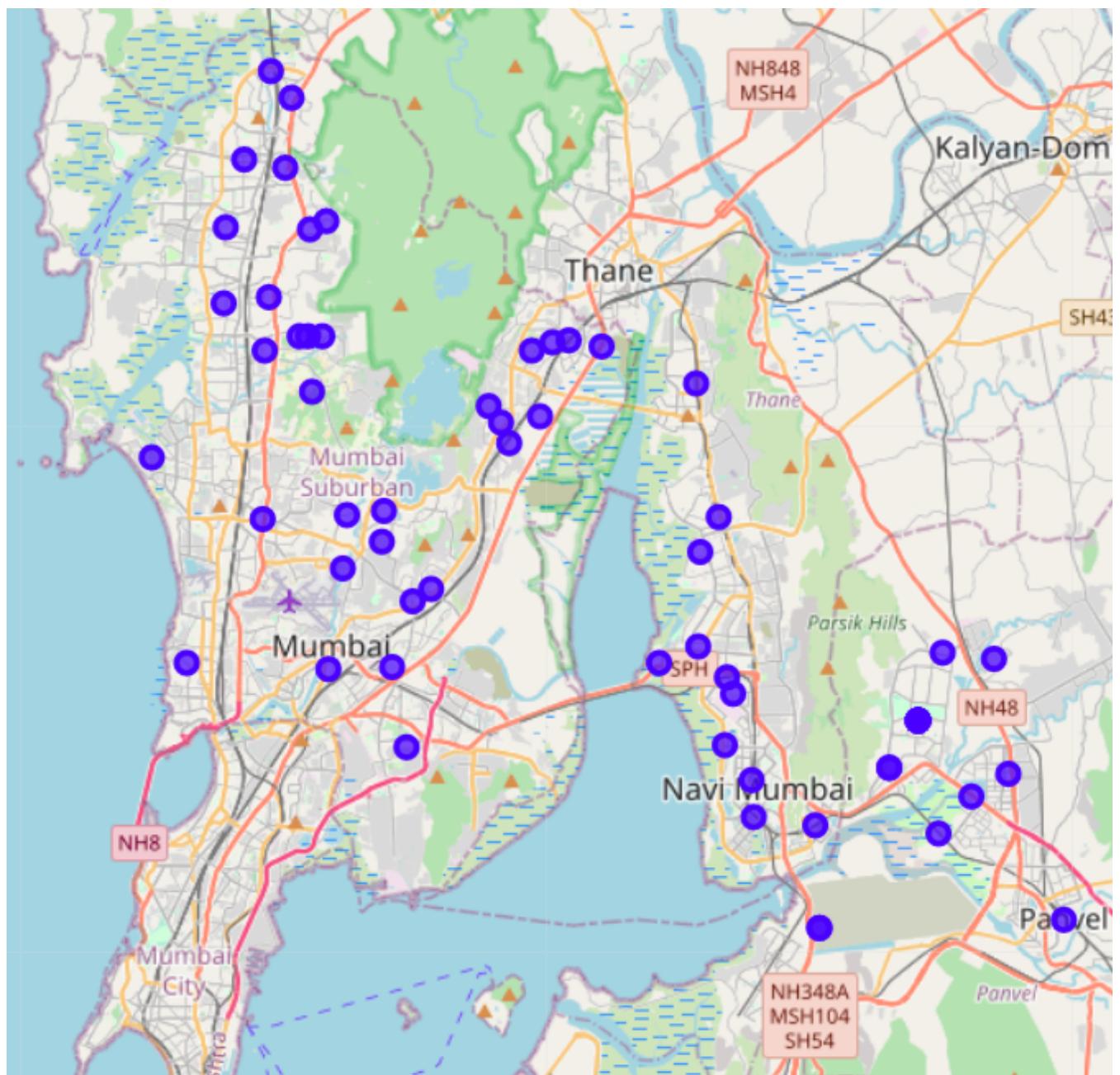
if premium:
    areas_df = areas_df[areas_df['Avg_rate'] > 17000]
else:
    areas_df = areas_df[areas_df['Avg_rate'] <= 17000]
```

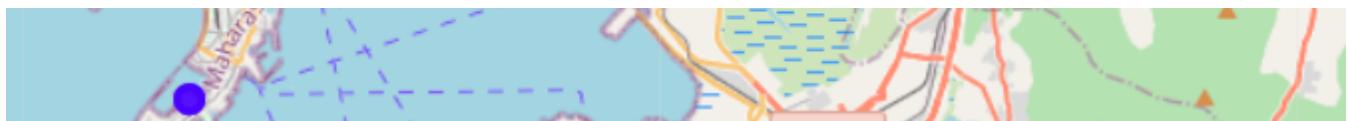
Here's the visualisation using folium





On the other hand, a map for “non-premium nightclubs” would look like





Conclusion

As you can see, we have only skimmed the surface of the vast possibilities Data Science withholds to solve your personal or business problems. As new people move into cities and competition keeps on increasing, applying data science to make your decisions can give you a huge advantage over your competition.

References

1. [My repo](#)
2. [Reference repo](#)
3. [99acres.com](#)
4. [Foursquare API](#)
5. [Nominatim API](#)

Data Science Data Visualization IBM Solutions

About Help Legal

Get the Medium app

