



# Coursera Capstone

## San Diego High School Recommendation

Sunil Yadavalli  
February 2020

# Business Problem

- Assume that someone got a job opportunity in the beautiful city of San Diego, California. Where you would recommend the locality to stay so that there are enough high schools around with a minimal commute?
- Also, in the available options, which school would be the best option to opt for if there is a chance of admission?

# Data

To address the above problem, we would need the below:

- List of the neighborhoods in San Diego
- The latitude and longitude of those neighborhoods
- Venue data, particularly the high schools.
- We will leverage the Foursquare API to collect the data, apply the machine learning techniques on this data and utilize map visualization to arrive at a solution.

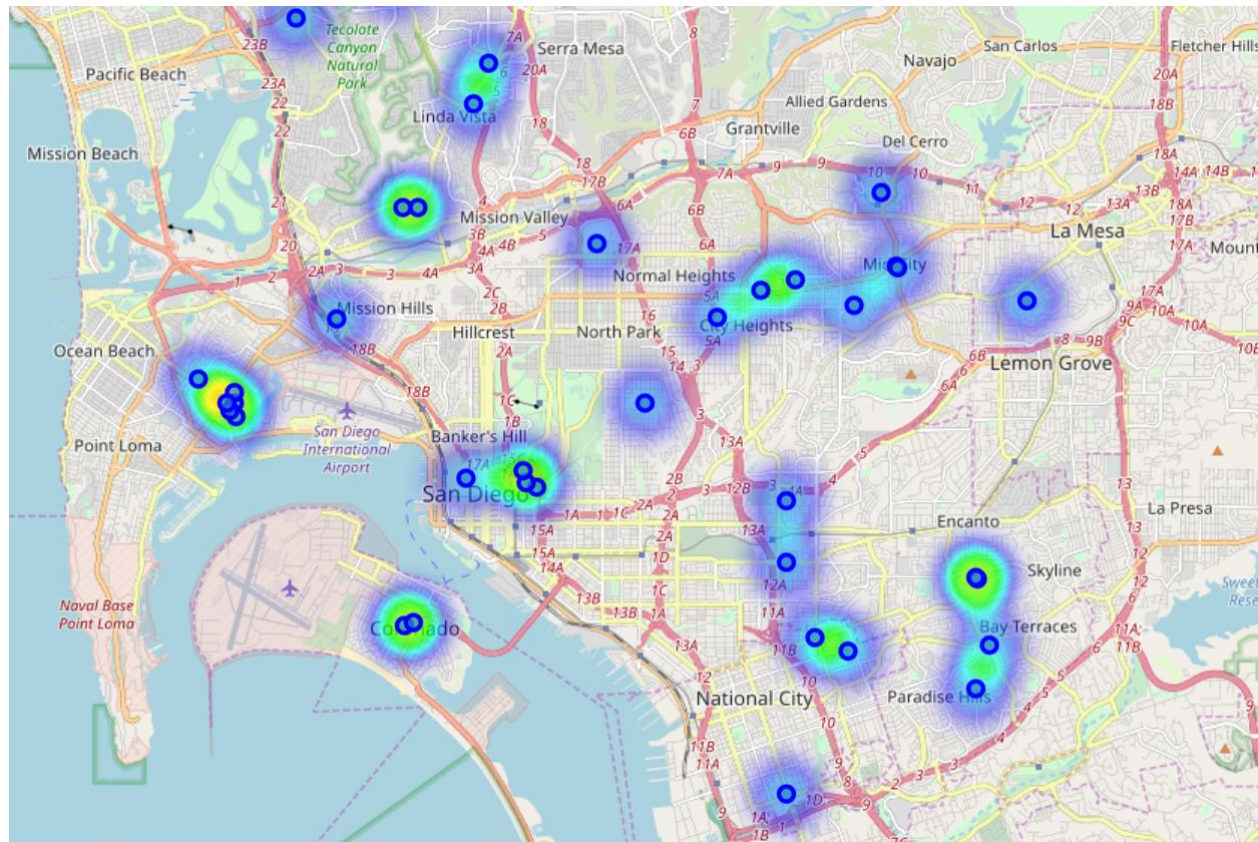
Below is the sample data

Identifier	SchoolName	ShortName	Address	PostalCode	Latitude	Longitude
4bbe5d61b083a5933ee5a1e9	San Diego SCPA	High School	2425 Dusk Dr	92139	32.679879	-117.048609
4b548e8ff964a520e5bf27e3	Helix Charter High School	High School	7323 University Ave	91942	32.754317	-117.037150
4bba201d7421a593db6bc340	Garfield High School	High School	1255 16th St	92101	32.718490	-117.149067
4dd293b61838a751965e6ca4	SDEMC	High School	1425 Russ Blvd Ste T112D	92101	32.719405	-117.151360
4bc8950d6501c9b6aa664029	San Diego High School	High School	1405 Park Blvd	92101	32.721871	-117.152180

# Methodology

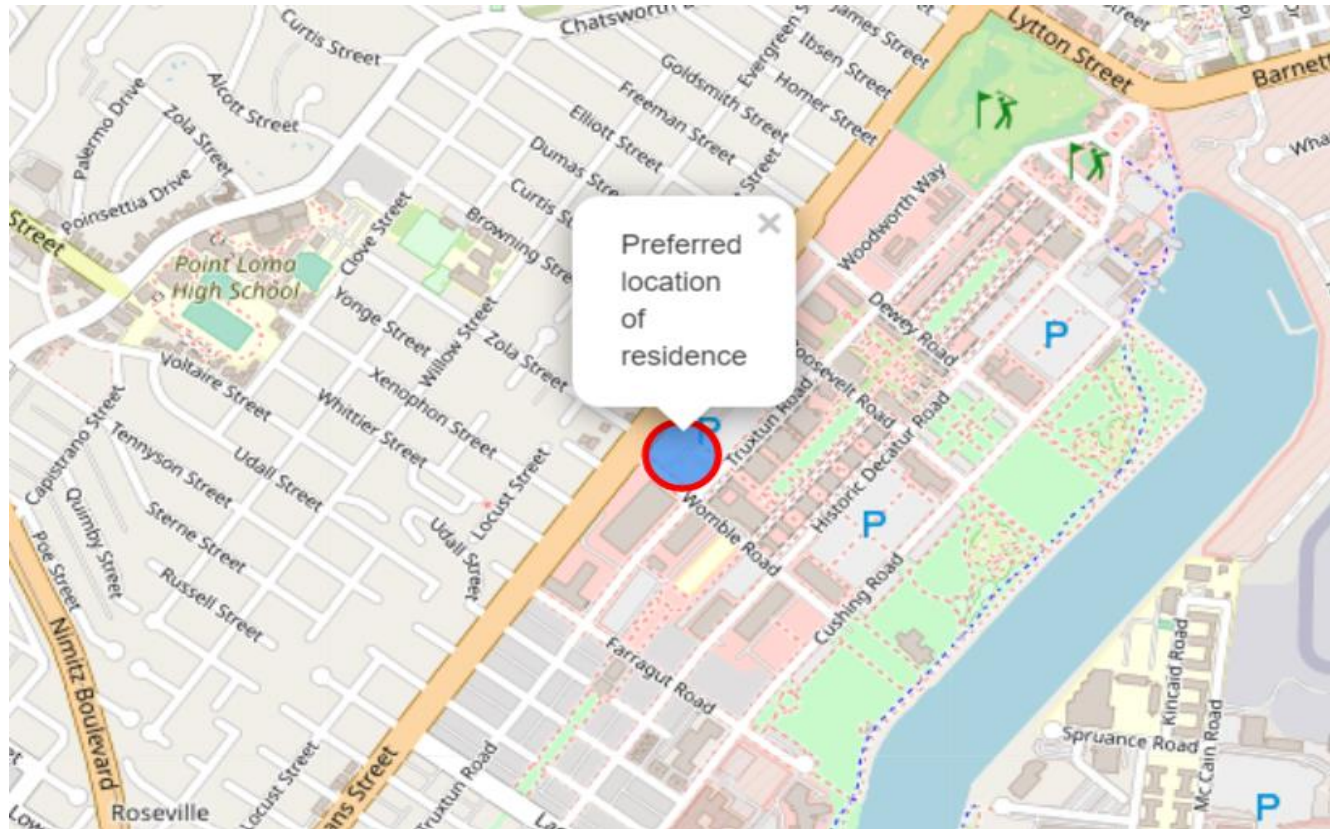
- We will use Foursquare API to get the top 50 venues that are within a radius of 10000 meters
- We use the CategoryID of the High School (which can found on the Foursquare API documentation) and pass it as a parameter to the call to get the high school information.
- After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighborhoods in a map using Folium package.

- Heatmap with the high schools visualization is shown in the below figure





- Based on the randomly picked latitude and longitude the proposed area of residence to look for is around the Womble Road



# Discussion

- We could further update our dataframe based on the ratings for each of the high school. Also, additional tips on them would help us further make a better choice. However, due to the limitation of the regular API membership we could not do that.
- This could be the potential area of interest when we deal with a real-world problem.



# Conclusion

- In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning and lastly providing recommendations to the relevant stakeholders
- To answer the business question that was raised in the business problem section, the answer proposed by this project is: The neighborhoods having the postal code 92106 are the most preferred locations to find a residence.



THANK YOU