

Coursera Capstone

San Francisco New High School Analysis

By Henry Bonnie



Introduction:

San Francisco is one of the most lively cities in the country. It has a population of 800,000 people and is a great city for families. Despite this, there are a limited number of high schools available in the Bay Area. In addition to this, the secondary school selection process is incredibly competitive and there are a sparse amount to choose from. There is a large youth population fighting over a select number of schools so it just makes sense that a new grades 9-12 high school is in order.

Business Problem:

The problem I am addressing in this report is finding out the optimal location to put a new secondary school within San Francisco. To solve this issue, I am going to need to use data and map school locations.

Data

The data I will need to find the optimal location is:

- Schools and their locations in San Francisco
- Youth population per district
- Schools per neighborhood
- School proximity

Methodology

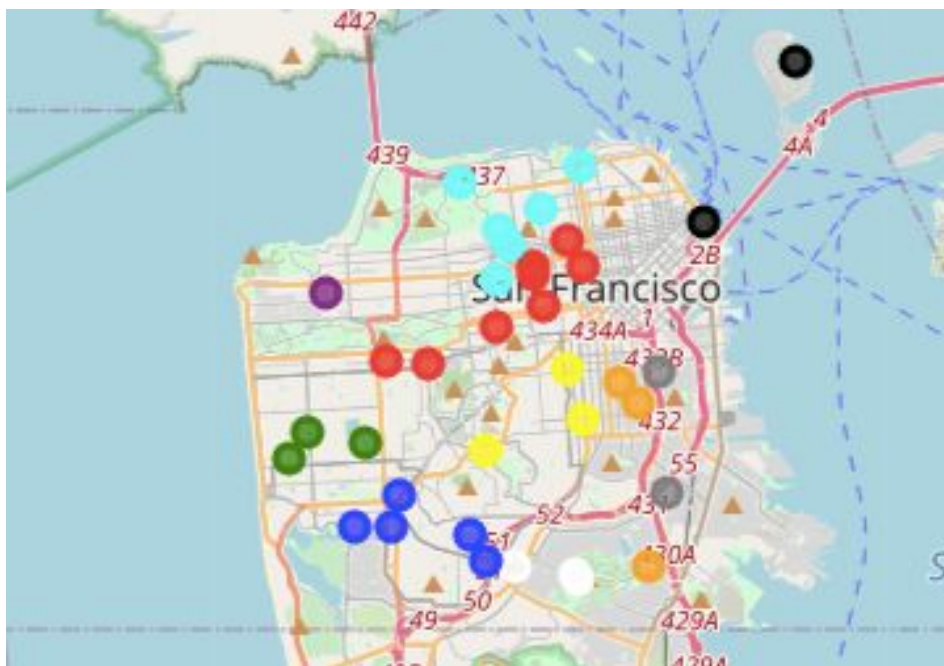
To solve this problem, the first thing I needed to find was a dataset that appropriately fit my project. Using sfgov.org, I found a .csv file on schools in their data records. I downloaded the data and read it into a dataset where I extracted name, address, country name, grade range, district, and coordinates. This preliminary dataset still needed to be cleaned. I selected schools within only the 'San Francisco' county as well as selected schools with a grade range of 9-12 as I was attempting to look at only secondary schools. Through this, I came up with the dataset displayed below:

	Name	Address	County Name	Grade Range	District	Latitude	Longitude
0	Marshall, Thurgood Marshall High School	45 CONKLING ST, San Francisco, CA 94124	SAN FRANCISCO	9-12	10	37.736309	-122.401649
1	San Francisco University High School	3065 JACKSON ST, SAN FRANCISCO, CA 94115	SAN FRANCISCO	9-12	2	37.790657	-122.445351
2	Saint Ignatius College Preparatory	2001 37TH AVE, SAN FRANCISCO, CA 94116	SAN FRANCISCO	9-12	4	37.748863	-122.495819
3	Hearst, Phoebe Apperson Hearst Home	3045 SANTIAGO ST, SAN FRANCISCO 94116	SAN FRANCISCO	9-12	4	37.743630	-122.500053
4	Burton, Phillip And Sala Burton High School	400 MANSELL ST, San Francisco, CA 94134	SAN FRANCISCO	9-12	9	37.721546	-122.406555
5	Washington, George Washington High School	600 32ND AVE, San Francisco, CA 94121	SAN FRANCISCO	9-12	1	37.777905	-122.491013
6	San Francisco Waldorf High School	470 WEST PORTAL AVE, SAN FRANCISCO, CA 94127	SAN FRANCISCO	9-12	7	37.736015	-122.471199
7	Lincoln, Abraham Lincoln High School	2162 24TH AVE, San Francisco, CA 94116	SAN FRANCISCO	9-12	4	37.746594	-122.480240
8	Life Learning Academy Charter School	651 8TH TI ST, SAN FRANCISCO, CA 94130	SAN FRANCISCO	9-12	6	37.825512	-122.367996
9	Gateway High School / Kipp Sf Bay Academy	1430 SCOTT ST, San Francisco, CA 94115	SAN FRANCISCO	9-12	5	37.783264	-122.436691
10	Galileo High School	1150 FRANCISCO ST, San Francisco, CA 94109	SAN FRANCISCO	9-12	2	37.803791	-122.424149

I took this dataframe and grouped it by district and used the count function to see how many schools are within each district. I would use this to do calculations in further sections. The data is displayed below:

Number of Schools	
District	
1	1
2	6
4	3
5	8
6	2
7	5
8	3
9	3
10	2
11	3

Once I had the dataset, I decided to display the locations on the city so I could visualize where a majority of the schools are located. Using the folium libraries, I created a map with markers over the locations of all the schools. I categorized them by district using different colors. The resulting map is displayed below:



After this, I was still curious about populations and how the amount of schools in a district compares to the adolescent population. Using San Francisco's city planning website, I found a pdf file containing demographic data for each district. I manually scraped the website for the population per district and percent of population ages 5-17. Once I had that data, I used it to calculate the amount of adolescents per school in each district. I dropped the third district from this data as an outlier as it is the downtown area of San Francisco whose real estate prices are too high to house a school. The resulting dataset is displayed below:

	Population	Percent Ages 5-17	Adolescent Population	Kids Per School
District				
1	78060	9	7025.4	7025.400000
2	67190	7	4703.3	783.883333
4	75120	11	8263.2	2754.400000
5	80760	6	4845.6	605.700000
6	66430	5	3321.5	1660.750000
7	72820	11	8010.2	1602.040000
8	71040	6	4262.4	1420.800000
9	83680	11	9204.8	3068.266667
10	73150	15	10972.5	5486.250000
11	85090	12	10210.8	3403.600000

Results

From the data I collected and analyzed, it seemed clear that the optimal location to place a school would be within district 1. Both of my main indicators, the map and kids per school, pointed to district 1. Visually, it is clear to see that district 1 doesn't have a great selection of schools to choose from. On top of this, they have the largest number of kids per school. Adding another school within this district would be beneficial to reduce this number as well as provide a more accessible option for children living within this district.

Discussion

From my analysis, I would recommend district 1 as the optimal location to place a new school. But there are more factors I would like to incorporate if I were to recreate this experiment. I would like to look further into real estate prices per district so I can see where it would be easiest to undertake a city project like this. In addition to this, I was curious about looking at traffic and amount of driving per district so I could conclude from which district it is most difficult to travel to schools.

Conclusion

In doing this capstone, the scarcity of 9-12 schools in San Francisco definitely shined in my eyes. I had experienced this problem first hand as a student applying for schools, but it was interesting looking at this issue from a strictly data-focused perspective. Through this, I learned which districts and locations are in need of a new school. In my opinion, I think it would be incredibly beneficial for district 1 containing the Richmond district and sea cliff neighborhood to have a new school accessible to their children.

Data Used

- <https://data.sfgov.org/Economy-and-Community/Schools/tpp3-epx2/data>
- https://default.sfplanning.org/publications_reports/SF_NGBD_SocioEconomic_Profiles/2010-2014_ACS_Profile_SupeDistricts_v3AH.pdf