San Francisco as a Travel Destination

Farah Darwich

February 21, 2019

1. Introduction

1.1 Background

When it comes to picking a city as a travel destination, whether it is to visit it yourself, to add it to the list pf cities your travel agency covers, or to talk about it in your tourism blog, there are many aspects to consider. It is not enough to rely on a city's reputation or the information you have about from the media alone to make such an important decision.

As the cultural, commercial, and financial center of Northern California, the city of San Francisco falls into the category of 'most popular cities in the U.S'. However, before it can be chosen as a travel destination for either business or personal purposes, it is important to carefully evaluate it on many aspects to determine whether or not making such a decision would serve the desired purpose.

1.2 Problem

Data that might help determine whether a given city is a suitable travel destination is the availability of shopping options, tourist spots, and accommodation options, as well as their quality and affordability. This project aims to evaluate the city of San Francisco based on those metrics to determine how efficient of a travel destination it is.

1.3 Interest

This evaluation will be of interest for stakeholders in industries related to tourism, travel agencies, as well as for people who are simply trying to decide on a destination for their next vacation.

2. Data Acquisition and Cleaning

2.1 Data Sources

In order to perform the evaluation, I will need as much metadata as possible about as many venues as possible in San Francisco.

Foursquare is a social location-based service that contains all the information we need. The Foursquare API allows application developers to interact with the Foursquare platform, which is how I obtained the dataset I worked with.

In this project I relied on the Foursquare scraper algorithm built by Data Scientist Max Woolf, and was able to obtain a dataset that contains basic metadata on 44425 venues in San Francisco including their categories, ratings, coordinates, and prices.

2.2 Data Cleaning

After acquiring the data from Foursquare, I saved it into a pandas dataframe so I can further examine and analyze it. The dataframe contained information on 44425 venues in San Francisco with their names, categories, coordinate, number of check ins, number of likes, ratings, prices, number of ratings, venue URLs, and Foursquare URLs.

I didn't need many of these columns in my work, and since I was going to base my evaluation of venue categories on availability, rating, and price, I decided to split the dataframe into two new dataframes: One to be used in evaluating venues based on price, and the other to be used in evaluating venues based on rating.

As for the rating dataframe, I deleted all columns except for the ones that contain names, categories and ratings, and then dropped all rows (venues) that had a missing value in one of those three columns.

Table 1: The first five rows of the ratings dataframe

	name	categories	rating
4	Madusalon	Salon / Barbershop	7.4
10	Mr. Smith's	Bar, Lounge, Nightclub	6.7
18	Hillside Supper Club	American Restaurant	8.2
25	Hand Touch Nails	Spa	7.7
38	Southern Pacific Brewing	Brewery, American Restaurant, Burger Joint	8.7

As for the price dataframe, I deleted all columns except for the ones that contain names, categories and prices, and then dropped all rows (venues) that had a missing value in one of those three columns.

Table 2: The first five rows of the prices dataframe

	name	categories	price
8	Lady Falcon Coffee Club	Food Truck, Coffee Shop	1.0
10	Mr. Smith's	Bar, Lounge, Nightclub	2.0
11	Madame Kim's Annex	Speakeasy	3.0
18	Hillside Supper Club	American Restaurant	2.0
21	The Treehouse @ Public Works	Lounge	3.0

In these final two dataframes that I used for the analysis, I had 8167 venues with available ratings and 8365 venues with available prices.

To recap, this evaluation was be built on the following aspects:

- Shopping Options (mainly, Shopping Malls and Shopping Plazas): Availability - Rating.
- Tourist Spots (mainly, Museums, Music Venues, and Theme Parks): Availability - Rating - Affordability.
- Accomadation Options (mainly, Hotels, Hostels, and Bead&Breakfast's): Availability - Rating - Affordability.

Keep in mind that each of those categories contain its own subcategories that will automatically be included in the analysis.

The ‘Museum’ category includes 5 sub-categories, the ‘Music Venue’ category includes 3 sub-categories, the ‘Theme Park’ category includes 1 sub-categories, and the ‘Hotel’ category includes 10 sub-categories.

Overall, 27 different categories of venues were included in this evaluation.

3. Exploratory Data Analysis (Methodology)

3.1 Shopping Options

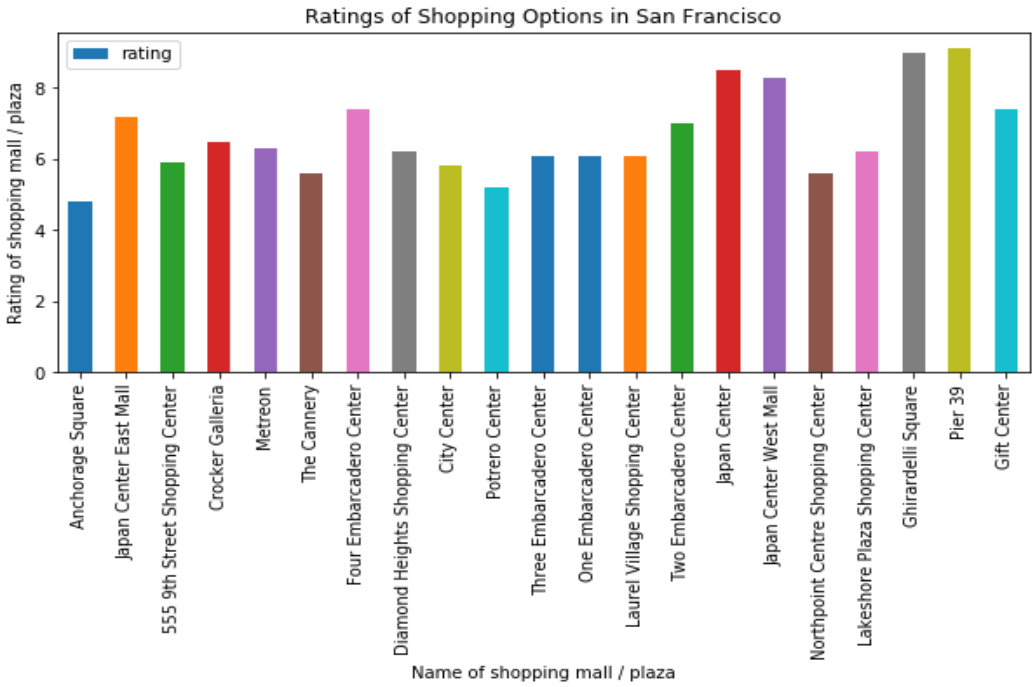
3.1.1 Shopping Options Availability

In order to determine the availability of shopping options in the City of San Francisco, I first calculated the number of shopping malls by counting the number of rows that contain the word ‘Shopping Mall’ in the categories column, and the result was 34. I did the same for shopping plazas and the result was 2. Overall, the total number of shopping options in San Francisco turned out to be 36.

3.1.2 Shopping Options Ratings

In order to evaluate the ratings for shopping options in the city of San Francisco, I first determined the number of shopping options with available ratings information from the ratings dataframe, and then calculated their overall average rating which turned out to be 7.74.

To get more in depth with the results, I generated a bar chart that contained the individual rating for each of the shopping options, and was able to use this later to give a judgement the quality of shopping options in San Francisco.

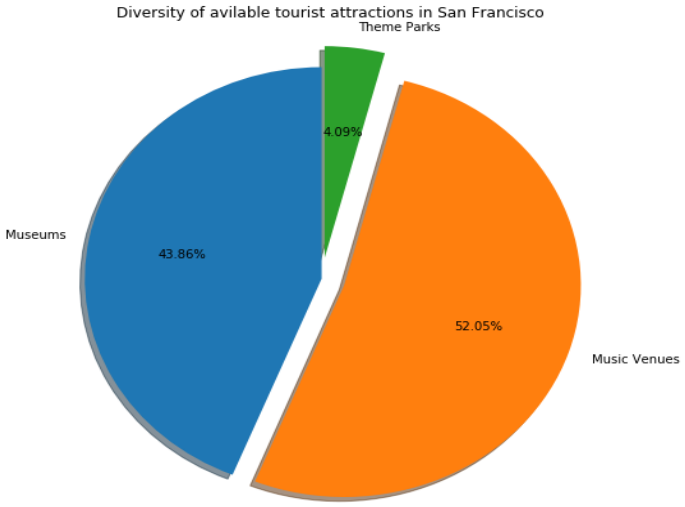


3.2 Tourist Spots

3.2.1 Tourist Spots Availability

In order to determine the availability of tourist spots in the City of San Francisco, I first calculated the number of museums by counting the number of rows that contain the word ‘Museum’ in the categories column, and the result was 150. I did the same for music venues and the result was 178. And then again for theme parks and the result was 14. Overall, the total number of tourist spots in San Francisco turned out to be 342.

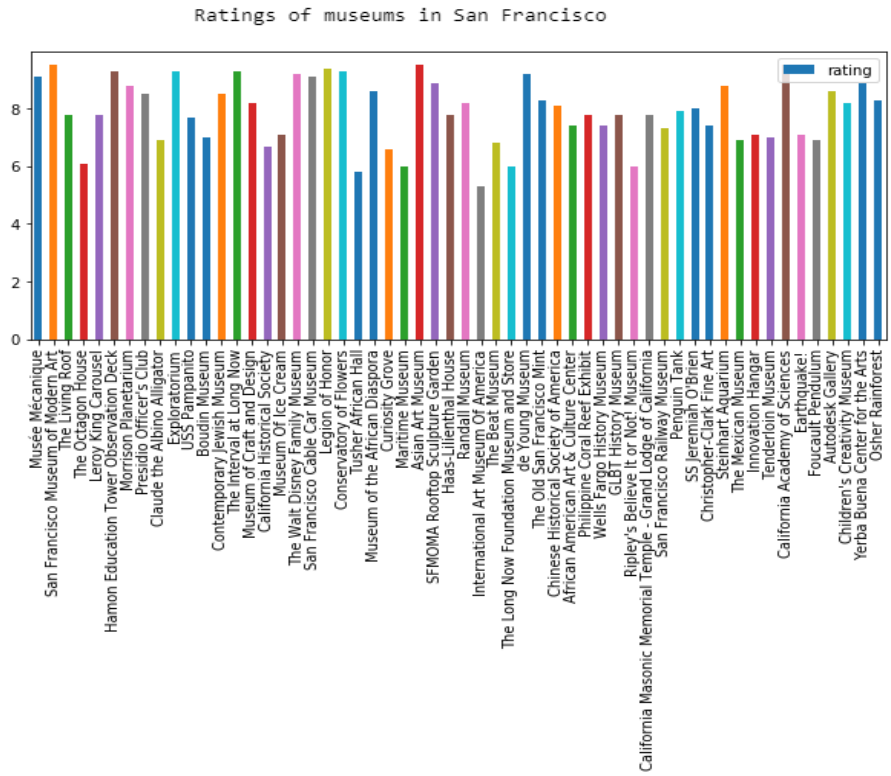
To affirm the diversity of available tourist spots in the city, I generated a pie charts of the previous results and was able to use this later to give a judgement the diversity of available tourist spots in San Francisco.

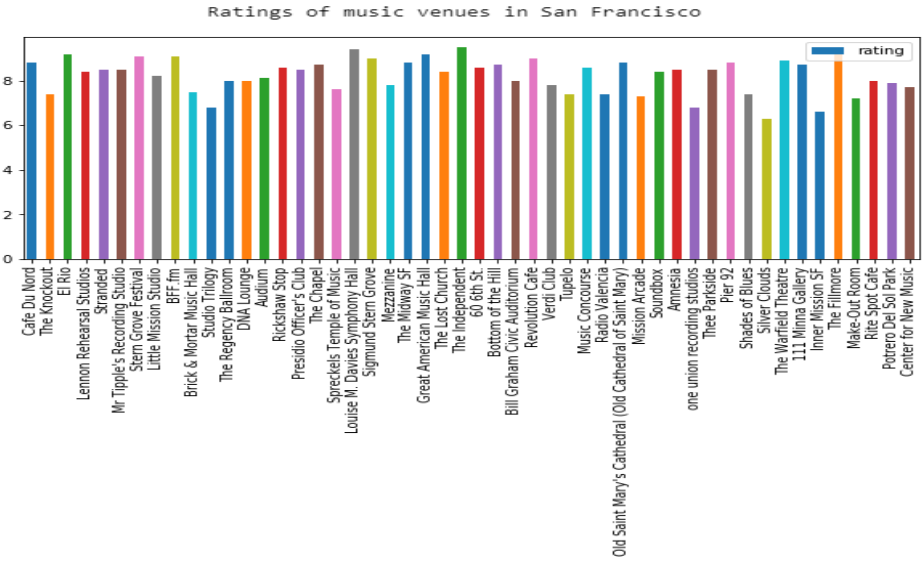


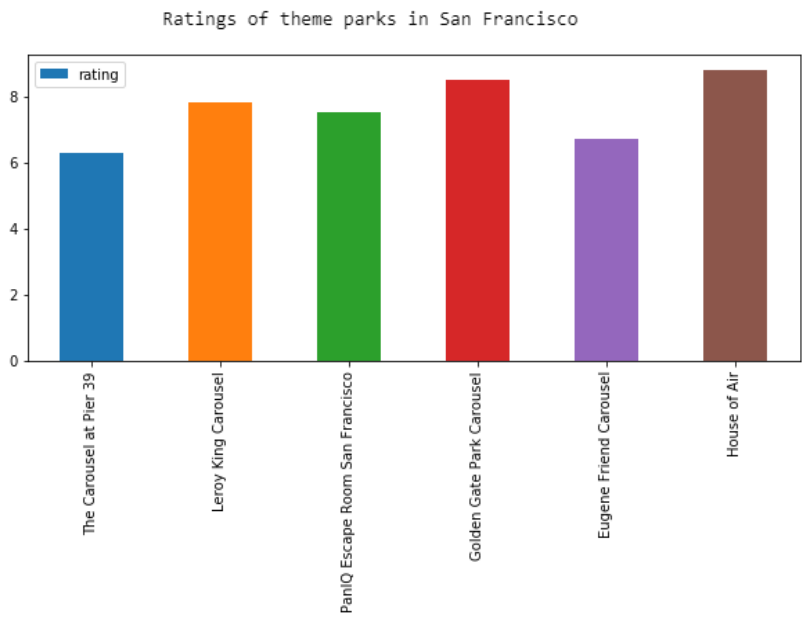
3.2.2 Tourist Spots Ratings

In order to evaluate the ratings for tourist spots in the city of San Francisco, I first determined the number of tourist spots with available ratings information from the ratings dataframe, and then calculated their overall average rating which turned out to be 7.91.

To get more in depth with the results, I generated three bar charts that contained the individual rating for each of the museums, music venues, and theme parks, and was able to use this later to give a judgement the quality of shopping options in San Francisco.







3.2.3 Tourist Spots Affordability

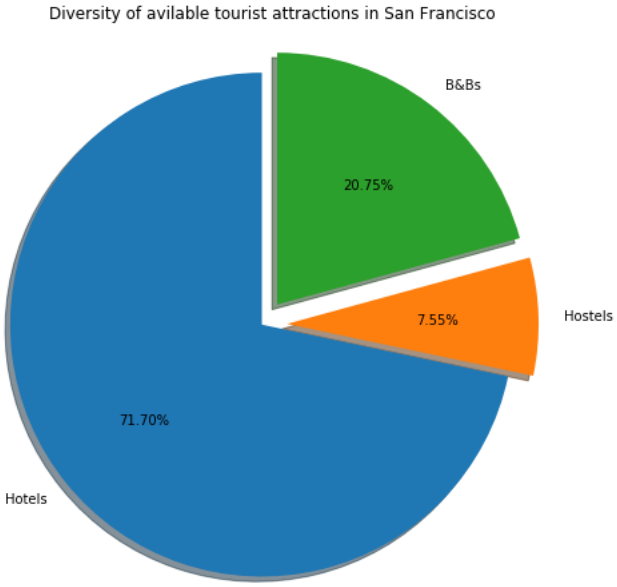
In order to evaluate the affordability of tourist spots in the city of San Francisco, I retrieved the minimum and maximum values for the price column for each of the three categories (museums, music venues, theme parks). Foursquare provides pricing information as a score from 1 to 4, with 4 being the most pricy. The results for tourist Spots showed that pricing for museums has a score of 2, while pricing for music venues has a score that ranges from 1 to 2. There were no information available on theme park prices.

3.3 Accommodation Options

3.3.1 Accommodation Options Availability

In order to determine the availability of accommodation options in the City of San Francisco, I first calculated the number of hotels by counting the number of rows that contain the word ‘Hotel’ in the categories column, and the result was 418. I did the same for hostels and the result was 44. And then again for bed and breakfasts and the result was 121. Overall, the total number of accommodation options in San Francisco turned out to be 583.

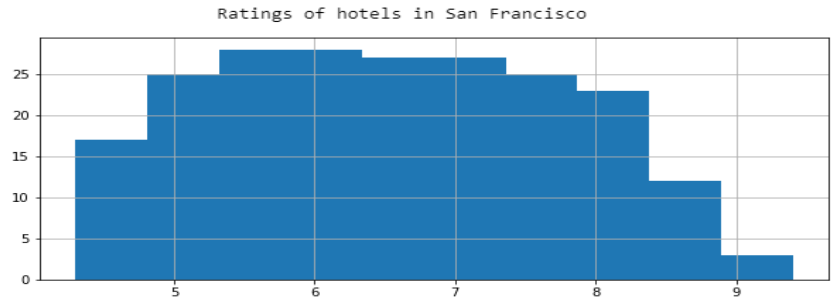
To affirm the diversity of available accommodation options in the city, I generated a pie charts of the previous results and was able to use this later to give a judgement the diversity of available accommodation options in San Francisco.

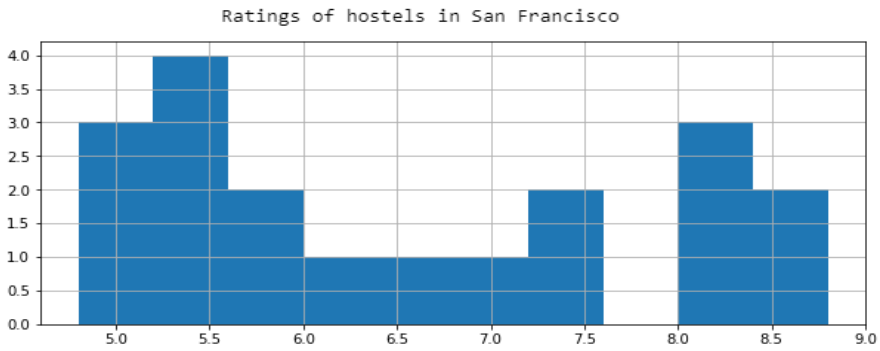


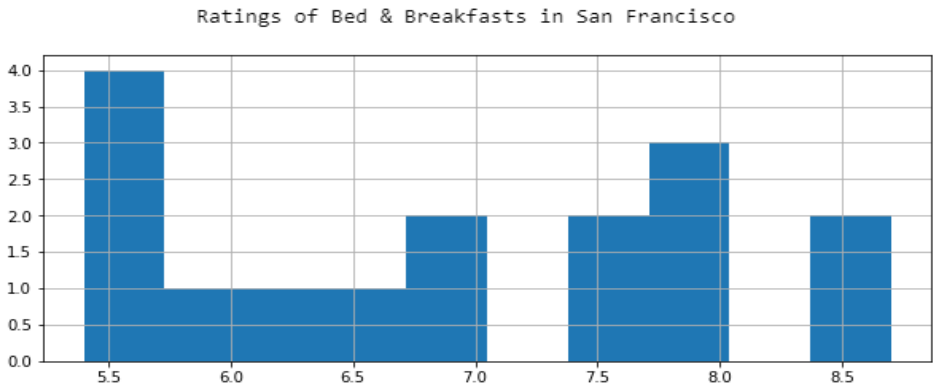
3.3.2 Accommodation Options Ratings

In order to evaluate the ratings for accommodation options in the city of San Francisco, I first determined the number of accommodation options with available ratings information from the ratings dataframe, and then calculated their overall average rating which turned out to be 6.67.

To get more in depth with the results, I generated three histograms (I didn't use bar charts here because of the large number of rows to plot) that contained the individual rating for each of the hotels, hostels, and bed and breakfasts, and was able to use this later to give a judgement the quality of accommodation options in San Francisco.







3.3.3 Accommodation Options Affordability

In order to evaluate the affordability of accommodation options in the city of San Francisco, I retrieved the minimum and maximum values for the price column for each of the three categories (hotels, hostels, bed and breakfasts). Foursquare provides pricing information as a score from 1 to 4, with 4 being the most pricy. The results for tourist Spots showed that pricing for hotels has a score that ranges from 1 to 3, while pricing for music venues has a score of 3. There were no information available on bed and breakfast prices.

4. Results and Discussion

4.1 Shopping Options

When it comes to shopping options, it seems that the city of San Francisco has a less than impressive number of shopping options (malls and plazas) available, however it makes up for it with a respectable average rating of 7.74/10. After visualizing the ratings for shopping options, I found that none of their ratings goes below 5, and most of them are between 7 and 9.

4.2 Tourist Spots

When it comes to tourist spots, San Francisco is at an abundance. There is a relatively high diversity of tourist attractions available, with a noticeably lower number of theme parks compared to museums and music venues. As for the ratings, most of the museums scored between 6 and 9 and none of them went below 5. Music venues showed better results as most of them scored between 7 and 9 and none of them went below 6.5. Same goes for theme parks which scored between 6 and 9. The tourist spots in San Francisco have an average rating of 7.91, which is very impressive. What was really note-worthy is how affordable the tourist spots turned out to be, especially compared to their relatively high ratings; both museums and music venues don’t go above 2 (on a scale of 1 to 4, 4 being the most pricey).

4.3 Accommodation Options

When it comes to accommodation options, San Francisco delivers. There is a high number of options available, with a decent diversity. However, it seems like that high number has affected the average rating of accommodation options, which stands at 6.67. This number is not necessarily bad, but it implies that a considerable number of those venues has a rating of 5 and even below. This was proven when I visualized those ratings, and it turns out that most hotel ratings are between 5 and 8 and some of them went below 5. The most common ratings for hostels were 5~6 and 8~9, with the rest in between. Same thing for B&Bs. As for affordability, hotel prices ranged between 1 and 3 (understandably, considering the wide range of ratings they have), while hostels scored an even 3. There were no information available on B&B prices.

5. Conclusion

The purpose of this project was to get a feel of how well the city of San Francisco performs as a tourist/travel destination. I evaluated three main aspects: Shopping options based on their availability and ratings, tourist spots based on their availability, ratings and prices, an accommodation options based on their availability, ratings, and prices as well. I relied on the Foursquare API to get relevant data on venues in San Francisco to perform the previous analysis. While conducting the analysis, I used data visualization to get a better understanding of my findings.

Overall, San Francisco performed relatively well on all of these aspects, as the results showed it has a respectable number of highly-rated shopping options, an abundance of highly rated and very affordable tourist attractions, and a large collection of accommodation options with a wide range of ratings and prices to suit everyone.

The final decisions will be made by stakeholders depending on their main objective, and other factors that come with it.