Coursera Capstone Project

Introduction

It is for those who are planning to start a new restaurant. I'll provide suggestions on what all factors need to be considered while starting a new venture.

Business Problem

I will focus on problems one might face while opening a new restaurant. Well, there are many factors to be considered to open a small or big level business. For example, if Ves wants to open an Indian restaurant, the first and foremost important point to decide is the location for his new restaurant. On what basis can Ves decide his restaurant's location? Does Ves want to rent the place or can he afford to buy it? While selecting the place there are key points to consider like he needs to check out how many Indian restaurants are there in that specific location? If incase there are already two Indian restaurants which have good ratings, will it be risky to open a new one near these restaurants? What factors will help Ves to run his business above average? (Out of scope for this project: Budget for Kitchen tools, restaurant furniture, and decorations, hire new chefs and waiters, budget everybody's salary. Decide on Menu details, restaurant advertisement, publish discount coupons, restaurant website.)

Discussion

Firstly, Ves needs to choose a suitable location for his restaurant. Let's say Ves wants to open a new Indian Restaurant in San Francisco. San Francisco is a famous tourist place, and also in Silicon Valley, because of that real estate is very costly. So, assuming that Ves wants to rent a place for his new restaurant. And now he needs to figure out how many restaurants are there in say neighborhood A, B, C, etc. If there are more than two Indian restaurants in a neighborhood then that would be a great risk to open a new restaurant of the same cuisine in that neighborhood. Selecting a place where there is less or no restaurant would be of great choice, considering the rent of the neighborhood too. Why I'm emphasizing that there should be fewer restaurants is, so that Ves will face less competition with the same cuisine restaurants. He needs to look for a place where many people frequently visit so that his business is above average. Places like Downtown, Movie theatre, Malls & Gas stations would help his business running. Restaurants' ratings, check-in of customers might help in deciding the location crowd. I would also suggest that Ves should check for the opening and closing timings of these restaurants. He may try opening his place 30 mins before these restaurants open up and close 30 mins or 1 hour later than this restaurants, this might help to get more customers.

Data Description

As Ves is planning to open a restaurant in San Francisco, and assuming that he is going to rent a place. So first I took the rent dataset from (https://web.archive.org/web/20180822060556/https://www.rentcafe.com/average-rent-market-trends/us/ca/san-francisco/ , https://www.zillow.com/research/data/) according to neighborhood wise, so that it's easy for us to check the rent data neighborhood wise. In this dataset I couldn't

get all neighborhoods rent information. So, I managed to use only that information which I could get from the website. I have cleaned the dataset and I'm going to compare the rent data of year 2018. Because for this project we just need to analyze the current rent range. Since from the webpages I didn't get all neighborhood's rent data, I planned to test only for the data I have retrieved.

I'm going to use a formula to find which neighborhood is good to open a new restaurant. Before coming up with a formula, I was wondering what all attributes/factors can we consider because it's really unfair to compare data of 10 years old restaurant with 1-year old restaurant. Like for example, the check-in count of 10 years old restaurant will be more compared to a 1 year old or 6 months old restaurant. And also, after analyzing data I found that in many restaurants check-in count is zero. I thought check-in count would be really be helpful to figure out the number of crowds visiting a particular neighborhood, but because of data discrepancy I avoided it. Then I conclude that every restaurant would definitely have ratings. Even if 100 customers have visited a 1-year old restaurant, the rating will be out of 5 stars, and same goes for a 10year old restaurant.

In [1]:
```
import pandas as pd
git = 'https://raw.githubusercontent.com/Auxilin/Battle-of-Neighborhoods-for-new-Restaurant-location/master/SFRent_Dataset.csv'
sfran_Data = pd.read_csv(git)
sfran_Data
```

Out[1]:

	Neighborhood	City	State	2018-07
0	Hayes Valley	San Francisco	CA	3030
1	Van Ness - Civic Center	San Francisco	CA	2977
2	Tenderloin	San Francisco	CA	2977
3	Downtown	San Francisco	CA	3040
4	Western Addition	San Francisco	CA	2989
5	Marina	San Francisco	CA	2948
6	Russian Hill	San Francisco	CA	2954
7	Lower Pacific Heights	San Francisco	CA	2963
8	Nob Hill	San Francisco	CA	3050
9	Pacific Heights	San Francisco	CA	2973
10	Noe Valley	San Francisco	CA	2818
11	Stonestown	San Francisco	CA	2640
12	Merced Heights	San Francisco	CA	2640
13	Lakeside	San Francisco	CA	2640
14	North of Panhandle	San Francisco	CA	3048
15	Alamo Square	San Francisco	CA	3048
16	Telegraph Hill	San Francisco	CA	3055
17	North Beach	San Francisco	CA	3055
18	Anza Vista	San Francisco	CA	3099
19	Glen Park	San Francisco	CA	2967
20	Forest Knolls	San Francisco	CA	3005
21	Sunnyside	San Francisco	CA	3039
22	Financial District	San Francisco	CA	3472
23	Mission Dolores Park	San Francisco	CA	3389
24	Twin Peaks	San Francisco	CA	3154
25	South of Market	San Francisco	CA	3490
26	Potrero Hill	San Francisco	CA	3490
27	Lone Mountain	San Francisco	CA	3283
28	Golden Gate Park	San Francisco	CA	3283
29	Yerba Buena Gardens	San Francisco	CA	3563
30	Presidio Heights	San Francisco	CA	3484
31	Lake Street	San Francisco	CA	3484
32	Jordan Park	San Francisco	CA	3484
33	Richmond District	San Francisco	CA	3484
34	South Beach	San Francisco	CA	3646
35	Mission Bay	San Francisco	CA	3711

Below is the formula for the solution.

finalScore = (rentScore)0.6+(ratingScore)0.4

Closer the finalScore value to 1, better choice of neighborhood for the client to decide. I have given more weightage to rent than rating. Let's see what is rentScore and ratingScore means.

rentScore can be calculated by (maxrentofneighborhood-currentrentofneighborhood)/(maxrentofneighborhood-minrentofneighborhood). rentScore value for each neighborhood can be obtained from above rent dataset. ratingScore can be calculated by (maxgoodrest-currentrestratingofN)/(maxgoodrest-mingoodrest). ratingScore can be calculated after retrieving information from foursquare location.

Methodology

To calculate the formula first let's calculate the rentScore. From above rentScore formula we need maximum and minimum rent of neighborhood. Maximum and minimum values can be retrieved using max and min built in functions.

maxrentofN = max(dataframename['rentcolumn'])

minrentofN = min(dataframename['rentcolumn'])

In [2]:
```
newurl = 'https://raw.githubusercontent.com/Auxilin/Battle-of-Neighborhoods-for-new-Restaurant-location/master/ResultsSheet.csv'
Results = pd.read_csv(newurl)
rent_Score = Results[['Neighborhood' , 'RentScore']]
rent_Score.head()
```

Out[2]:

	Neighborhood	RentScore
0	Hayes Valley	0.635854
1	Van Ness - Civic Center	0.685341
2	Tenderloin	0.685341
3	Downtown	0.626517
4	Western Addition	0.674136

Next in order to calculate ratingScore, we need to retrieve rating of each Indian restaurant neighborhood-wise. So, let's the analyze data. Firstly, we need the co-ordinates of the neighborhood's, co-ordinates or latitude & longitude can be obtained by passing the Neighborhood name value through geocoding. The restaurant details can be retrieved using search endpoint of foursquare location. For our project we need only Indian restaurant data, and in search endpoint there is a attribute called category id, i.e for each category (like Indian or Italian or Mexican Restaurant) foursquare has a defined category which will help us to get the desired data. In this search response, we'll retrieve the venue id of all the Indian venues. And

then pass this venue id's through venue_id endpoint to get rating of each Indian restaurant. Let's save the data in a dataframe, for further testing.

Now we need to select a neighborhood in which we have to reduce the competition for our new restaurant, hence we should test with only good rating restaurants. For this I'm going to consider restaurants which have rating greater than or equal to 7 in foursquare.

```
In [3]: Results[['Neighborhood' , 'GoodRatingRestaurant']].head()
```

Out[3]:

	Neighborhood	GoodRatingRestaurant
0	Hayes Valley	2
1	Van Ness - Civic Center	3
2	Tenderloin	8
3	Downtown	6
4	Western Addition	0

After getting counts of good rating restaurant in each neighborhood, we can calculate our rentRating now. Max and min can be calculated using respective functions.

maxgoodrest = max(dataframename['ratingcountcolumn'])

mingoodrest = min(sf_Neighborhood['ratingcountcolumn'])

Thus, now we can calculate ratingScore using the formula ratingScore = (maxgoodrest-currentrestratingofN)/(maxgoodrest-mingoodrest).

Once we have rentScore and ratingScore we can use our formula to calculate the finalScore. And below are the first 5 rows from the final results dataframe.

Result and Conclusion

```
In [4]: Results[(Results['FinalScore'] >= 0.8)].sort_values('FinalScore', ascending=False)
```

Out[4]:

	Neighborhood	2018-07	Latitude	Longitude	RentScore	GoodRatingRestaurant	RatingScore	FinalScore
11	Stonestown	2640	37.727446	-122.474895	1.000000	0	1.000	1.000000
12	Merced Heights	2640	37.717507	-122.470281	1.000000	0	1.000	1.000000
13	Lakeside	2640	37.731967	-122.474257	1.000000	0	1.000	1.000000
10	Noe Valley	2818	37.751591	-122.432081	0.833800	1	0.875	0.850280
5	Marina	2948	37.799793	-122.435205	0.712418	0	1.000	0.827450
19	Glen Park	2967	37.733108	-122.433784	0.694678	0	1.000	0.816807
9	Pacific Heights	2973	37.792717	-122.435644	0.689076	0	1.000	0.813445
4	Western Addition	2989	37.779559	-122.429810	0.674136	0	1.000	0.804482

Above is the result of my formula which I used to analyze best neighborhood's for new Indian restaurant. If you see carefully, I have listed Neighborhood's which has Finalscore greater or equal to 0.8, so that Ves has more options to choose from. There are 3 Neighborhood's in above results which are best suited for Ves as there is no competition, since there is no good rating restaurants and rent is also comparatively low according to our formula.