Opening a New Sushi Restaurant in San Francisco, California

By Rey Andrew Doctora

Business Problem Introduction

For the final capstone project, we received a request from our Japanese client who was interested in opening a new sushi chain in the city of San Francisco, California, USA.The key stakeholder was the Director of Operations who specified two criteria for site selection:

- A location not crowded with competitor restaurant cuisines
- Low theft rates since the chain maintains expensive equipment and ingredients what would be cost-prohibitive to replace from a supply-chain perspective

Data Used for Analysis

Based on the requirements defined in the problem, factors that influenced the stakeholder's decision included:

- Number of competitor restaurants in a location
- Crime data for a location

For the first criteria, we used data from **Foursquare** to generate a list of restaurants in San Francisco by location.

For the second criteria, we used public **San Francisco crimes data** to investigate theft and robberies by location to present a proposal for site selection for the year 2016. This data set was acceptable to the Director as a preliminary indicator of trends.

We used district parameters as defined by the San Francisco Police Department to define San Francisco locations. Data sources extracted for analysis included:

- Theft events by location over-layed onto a San Francisco Folium map
- Number of restaurants by San Francisco location will be obtained using Foursquare API

Methodology

Our methodology employed a 10-step process to import the data, wrangle the data, and provide outputs to the data that resulted in our recommendations. The steps listed include:

- Import all data libraries into Python notebook
- Extract 2016 San Francisco Crimes data using the PdDistrict as basis for San Francisco Districts
- Data wrangle police incidents file to select only Burglary, Larceny/Theft, and Robberies
- Use an agglomerative approach to aggregate theft incidents by District
- Using the Districts, determine latitude and longitude geolocator data for each district
- Use Foursquare API to get all venues in defined Districts
- Data wrangle venues data to select only restaurants in each District
- Aggregate restaurants by District
- Create a map to overlay theft incidents data as well
- Merge theft incidents data with restaurants data and sort by theft incidents to analyze the results

Results and Discussion of Findings

PdDistrict	Total Theft
TENDERLOIN	2338
INGLESIDE	2628
PARK	2758
BAYVIEW	3192
TARAVAL	3480
RICHMOND	3629
MISSION	4661
CENTRAL	7780
NORTHERN	8410
SOUTHERN	10633

Neighborhood	Restaurant Venues
TENDERLOIN	4
SOUTHERN	7
NORTHERN	8
PARK	10
CENTRAL	11
BAYVIEW	12
INGLESIDE	13
RICHMOND	13
TARAVAL	15
MISSION	18

After wrangling the data, we rank sorted total theft incidents by District as noted on the left table.

Similarly, we ranked sorted total restaurant venues as returned by our Foursquare API as noted on the right table.

We then merged the two tables together to discern any patterns capturing the attributes of theft and total restaurants in vicinity.

PdDistrict	Total Theft	Restaurant Venues
TENDERLOIN	2338	4
INGLESIDE	2628	13
PARK	2758	10
BAYVIEW	3192	12
TARAVAL	3480	15
RICHMOND	3629	13
MISSION	4661	18
CENTRAL	7780	11
NORTHERN	8410	8
SOUTHERN	10633	7

Our analysis indicated that the **Tenderloin** district appears to exhibit the attributes of comparatively minimized theft incidents and fewest restaurant venues compared with other districts in San Francisco. Additional districts the client may wish to consider include the **Ingleside** and **Park** districts which seem to have the second and third lowest theft incidents, respectively, throughout the city. While the **Southern** and **Northern** districts appears to have the second and third number of restaurant venues in the city, they have nearly four to five times as many theft incidents compared with the lowest ranking districts

Only two factors were considered in this project's scope, namely **theft incidents** and **restaurant venues,** which may present limitations to the results. We therefore further recommend that additional iterations of data analysis be completed to fine-tune the requirements of the client and better serve their business needs. Other parameter considerations for future iterations include:

- For theft data, while the Tenderloin may have the lowest theft rates, other crime types, homelessness, or other socioeconomic data may be found to be higher than other districts to the dismay of the client.
- For restaurant venues, the data only filtered on restaurant types. Other venues, such as bake shops or bars may have an impact on customer needs as well as adjusting parameters on the Foursquare API.

Conclusion

This project was initiated on behalf of a client whose business problem required selecting a location in San Francisco, CA to open a new sushi restaurant based on the criteria of fewest theft incidents and minimum adjacent restaurants.

Using Foursquare and theft incident data made publicly available by the San Francisco Police Department, we analyzed numerous data sets to conclude that opening a location in the **Tenderloin, Ingleside, or Park Districts** are optimal locations based on the criteria supplied.

As data analysis is an iterative process, we further suggest to run additional analyses to determine if other attributes may present opportunities to fine-tune site selection.