Analysis of San Francisco Neighborhoods to Open an Indian Restaurant

Siddharth Banerjee

INTRODUCTION

- California has the highest Indian populations among US States
- Within California, San Francisco has the highest population of Indian-Americans living
- Crime rate also high in SF city
- This analysis uses Crimes dataset combined with Foursquare API to locate the perfect place to open an Indian restaurant

DATA

- a. Data Collection
  - Crimes dataset obtained from earlier Coursera course on data visualization using Python
  - Foursquare API used to obtain locations of Indian restaurants using client key and client secret
- b. Data Cleaning
  - Crimes dataset has 150500 initial rows with 13 features
  - Although the dataset belongs to 2016, I have assumed a constant rate change in crimes in all police districts, still maintaining the validity of my analysis
  - Only one row with ‘nan’ for police district – this row removed from analysis

METHODOLOGY

```
In [8]:  1  df_incidents.groupby(['Resolution']).count()
```

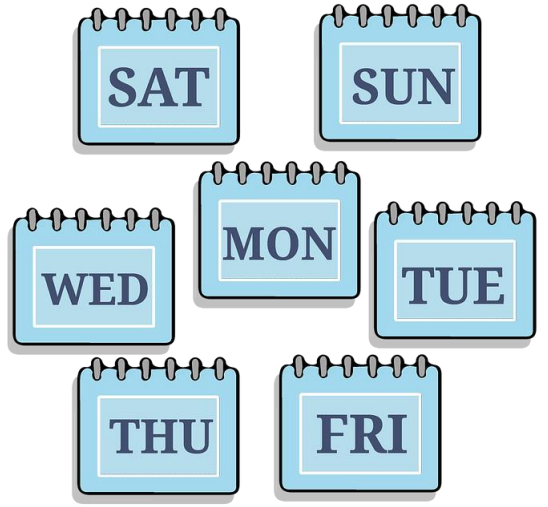
Out[8]:

Resolution	IncidntNum	Category	Descript	DayOfWeek	Date	Time
ARREST, BOOKED	39416	39416	39416	39416	39416	39416
ARREST, CITED	144	144	144	144	144	144
CLEARED-CONTACT JUVENILE FOR MORE INFO	58	58	58	58	58	58
COMPLAINANT REFUSES TO PROSECUTE	2	2	2	2	2	2
EXCEPTIONAL CLEARANCE	371	371	371	371	371	371
JUVENILE BOOKED	1056	1056	1056	1056	1056	1056
JUVENILE CITED	3	3	3	3	3	3
JUVENILE DIVERTED	2	2	2	2	2	2
LOCATED	20	20	20	20	20	20
NONE	107780	107780	107780	107780	107780	107780
NOT PROSECUTED	22	22	22	22	22	22
PROSECUTED BY OUTSIDE AGENCY	1	1	1	1	1	1
PSYCHOPATHIC CASE	17	17	17	17	17	17
UNFOUNDED	1608	1608	1608	1608	1608	1608

- a. Feature Selection
  - Under ‘Resolution’ column, 107700 rows have ‘NONE’ meaning no action taken – removed from analysis
  - Only ‘ARREST, BOOKED’ rows considered for analysis
    - Increases reliability of data used
    - Reduces computational resources

METHODOLOGY (contd.)

- b. Crime Data Variables
  - ‘DayOfWeek’: Is there any observable trend in the day of week crimes are committed? Mostly, restaurant patrons prefer weekends to visit
  - ‘PdDistrict’: Are there any visible trends in the crimes committed based on police districts? This can be a huge factor in deciding location of the Indian restaurant



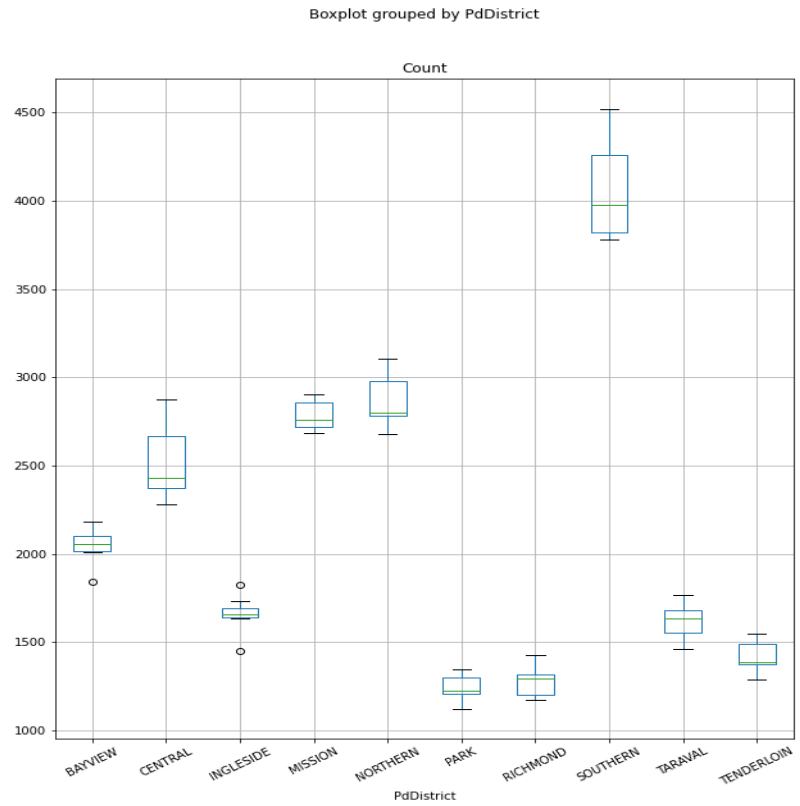
METHODOLOGY (contd.)

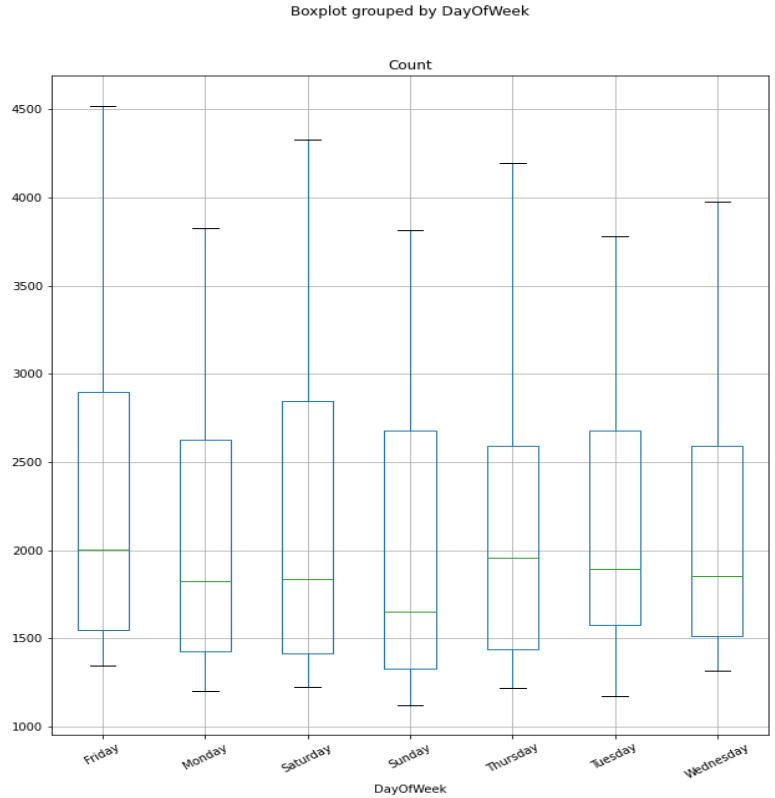
- Foursquare API
- Client id and client secret used in a query to find out Indian restaurants in San Francisco city
- Code used as shown in the side box
- Initial .json file converted to Pandas dataframe
- Data extracted for Indian restaurants

```
url = 'https://api.foursquare.com/v2/venues/search?&client_id={}&client_secret={}&v={}&query={}&ll={},{}&radius={}&limit={}'.format(
CLIENT_ID,
CLIENT_SECRET,
VERSION,
search_query #Indian,
latitude,
longitude,
radius #10000,
LIMIT #100)
```

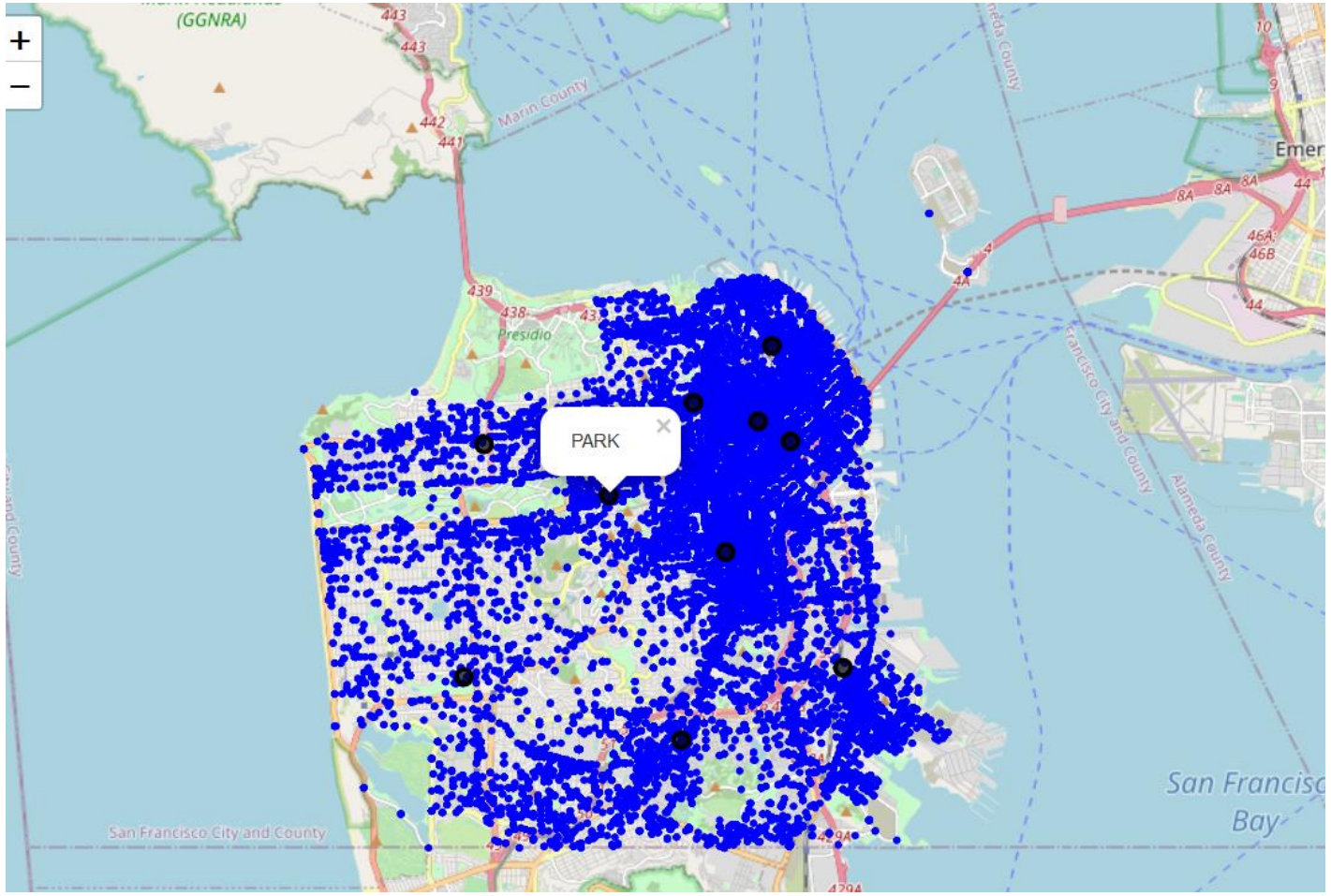
RESULTS

- Boxplot of DayOfWeek and PdDistrict as shown beside
- Thursdays and Fridays were found to have the maximum crimes reported – incidentally, it also happens to be the start to the weekend
- Park and Richmond police districts have the lowest crimes reported – so these could be possible locations for the restaurant considering lower crime rates



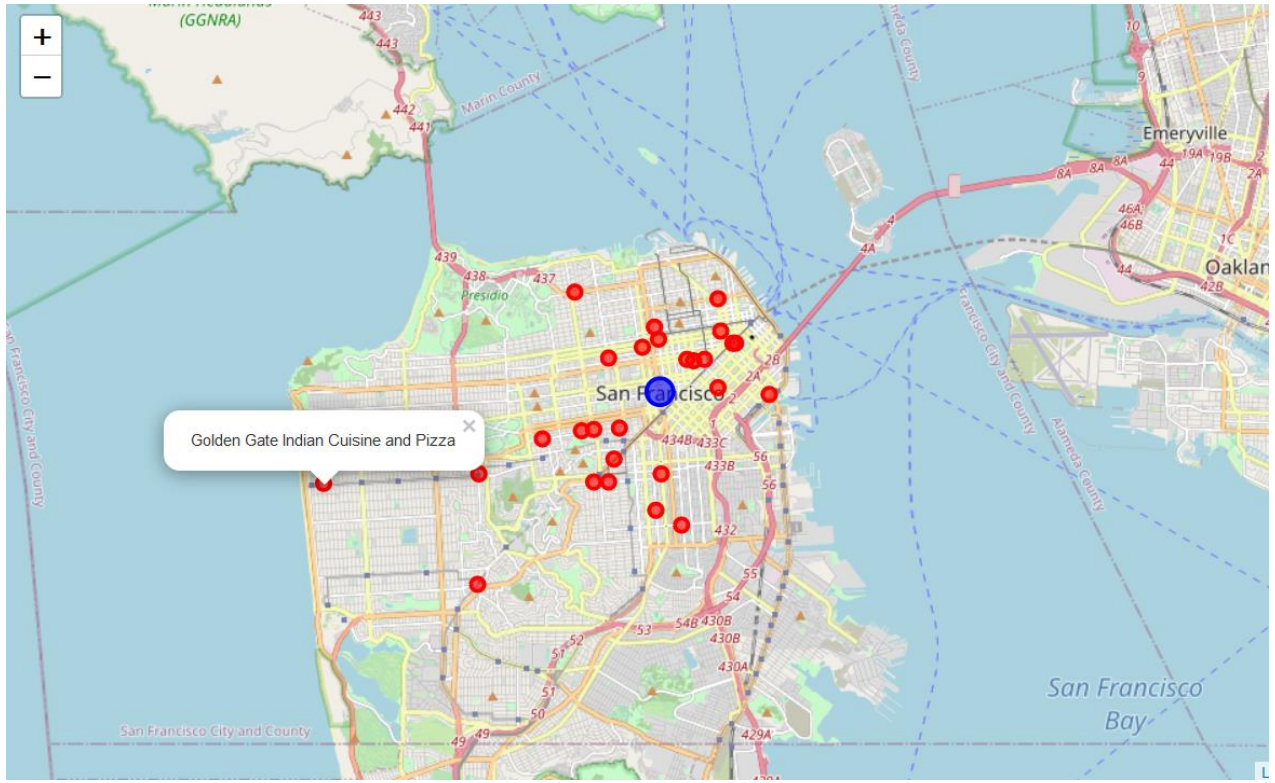


RESULTS (contd.)

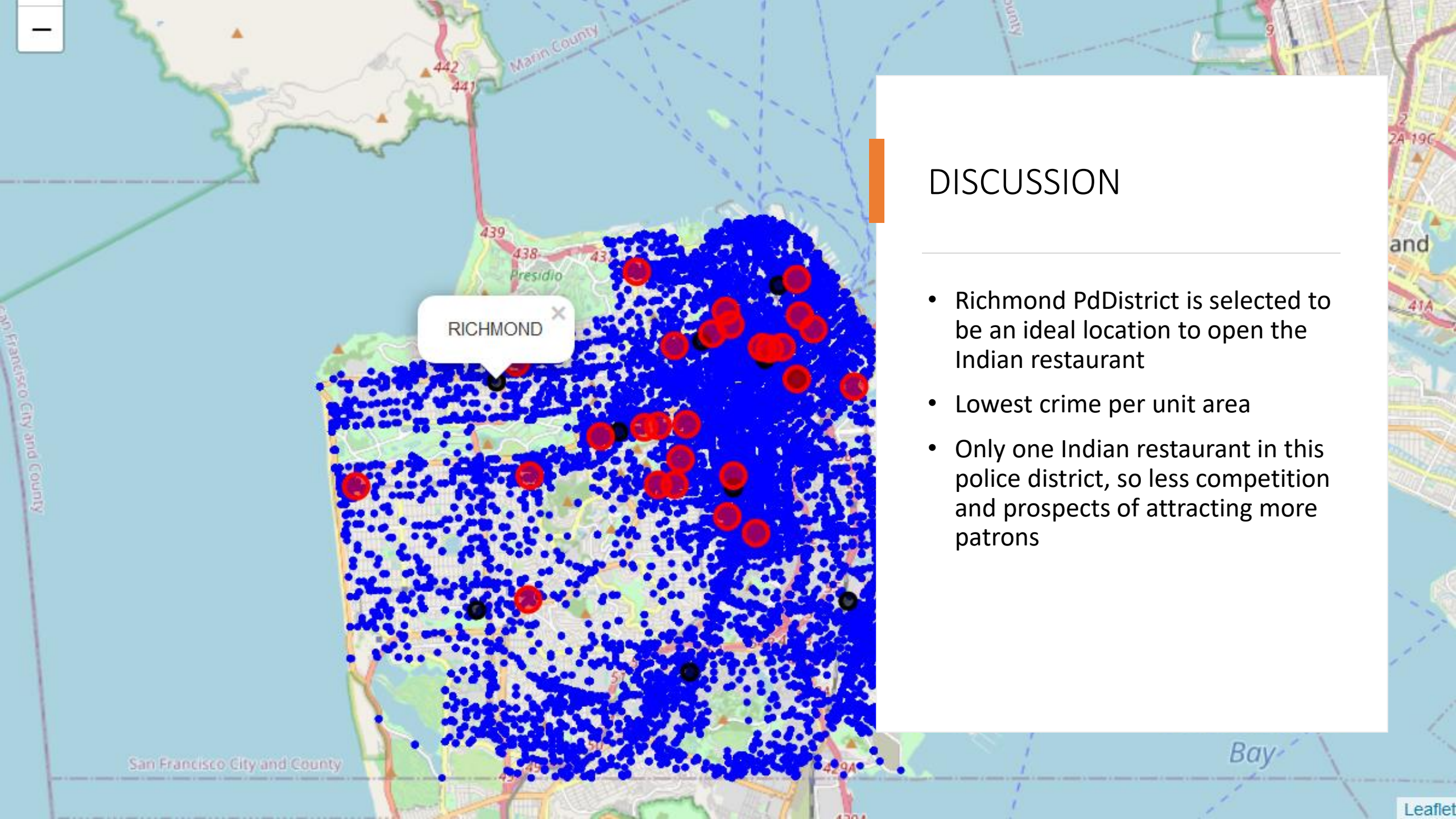


- Crimes plotted using Folium – these are shown using blue dots
- Police districts location obtained as mean of lat, long values grouped by PdDistrict column
- These are shown using black dots – in the figure, one can see Park PdDistrict label

RESULTS (contd.)



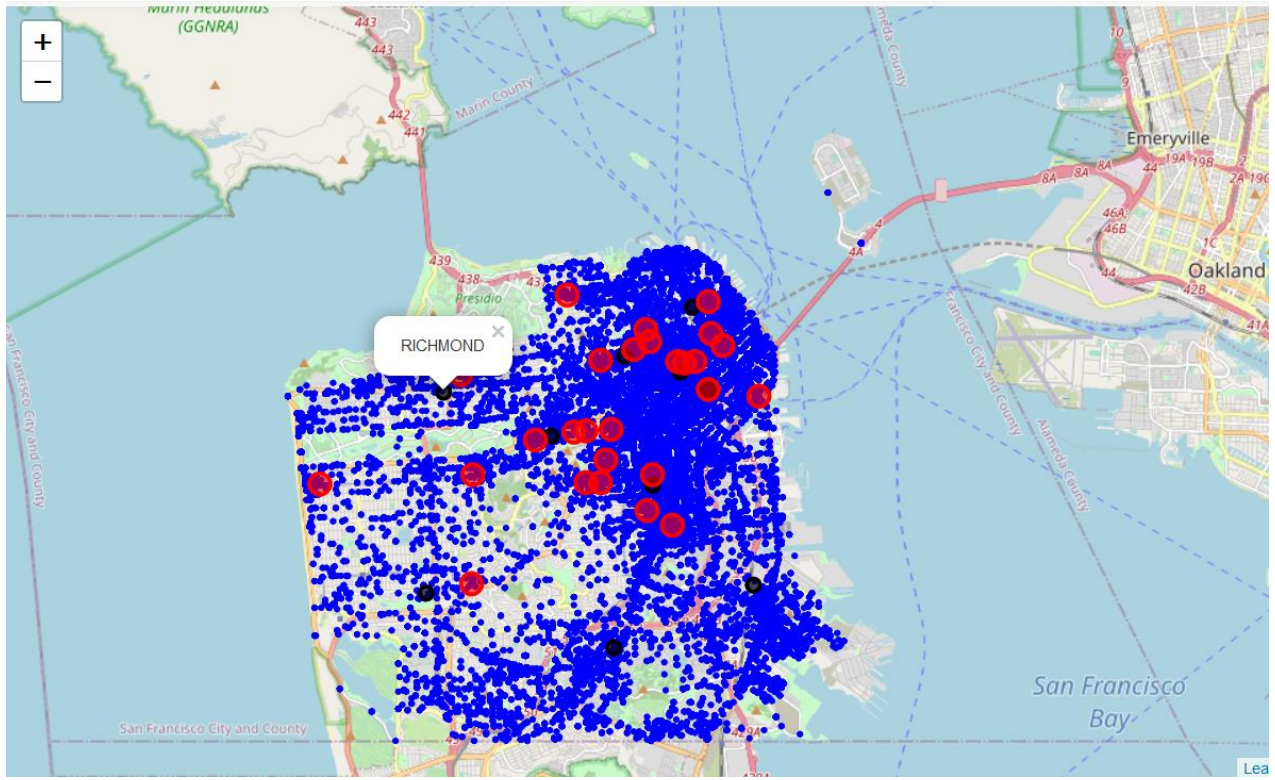
- Foursquare API used to locate Indian restaurants in SF city
- .json file converted into Pandas dataframe
- Restaurant data extracted from this dataframe and plotted using Folium as shown using red dots



DISCUSSION

- Richmond PdDistrict is selected to be an ideal location to open the Indian restaurant
- Lowest crime per unit area
- Only one Indian restaurant in this police district, so less competition and prospects of attracting more patrons

CONCLUSION



- This analysis used Python pandas to bring extracted data into appropriate dataframes
- Foursquare API used for obtaining locations of Indian restaurants in SF city
- Folium used for visualizing data
- Richmond police district best suitable location for opening Indian restaurant considering just the crime data
- Future studies can look into other factors such as commercial rental space rates, availability of labor, etc.