Rental Housing in San Francisco¶

Table of Contents

1. Problem and Discussion

Often relocation is a time consuming process. Finding affordable housing with safe neighborhoods and prefered venues is a big challenge. Data science can save time for meeting such criteria by providing interactive visual tools through Jupyter notebook. A Goal of this project is to provide such a sample project for finding a new housing in San Francisco utilizing **Folium** library to make visual segmentation and clustering data in a map. This notebook allows users to tweek few parameters and shows desired housings on top of crime rate in neighborhoods, rental price ranges and venues on interactive maps.

This project can help rentees considering moving to San Francisco or renters deciding reasonable rents since interactive visual aids can quickly allow users to see intuitive and interactive visual information. The use of FourSquare data and mapping techniques combined with data analysis will help providing clustered venues along with rents and crime rate in a single map. This project can serve a good practical case toward the development of Data Science skills.

2. Description of Data and How it will be Used

In this jupyter notebook, main focal area is set to San Francisco. This notebook will use geojson data from DataSF (https://data.sfgov.org/api/geospatial/pty2-tcw4?method=export&format=GeoJSON) for geographical information and police department incident report from DataSF (https://data.sfgov.org/api/views/wg3w-h783/rows.csv?accessType=DOWNLOAD) for stata analysis and finally use python-craigslist to retrieve a set of most recent posts on interactive maps. The raw data from craigslist is programatically scraped and the data will generate statistics and interactive visual aids for users.

Use Foursquare and geopy data to map top 10 venues for all San Francisco neighborhoods and clustered in groups (as per Course LAB). Use foursquare and geopy data to map the location of available rental housings and crime rates separately and on top of the above clustered map in order to identify the venues. The markers of rental housing display the rental price and URL to the posts in the popups. Alternatively Boxplot and Choropleth Maps shows rents statistics and mean rental values respectively in the neighborhoods.

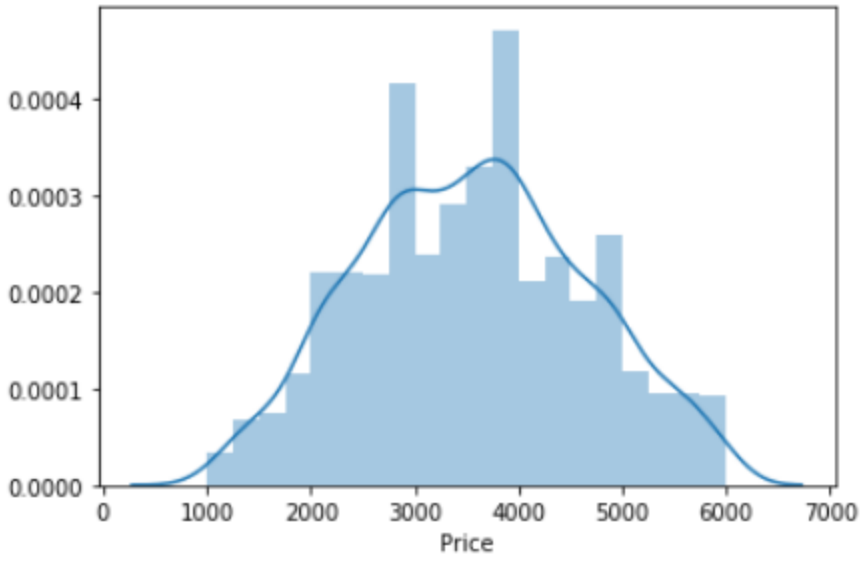
3 Methodology Section

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicate the execution of steps.

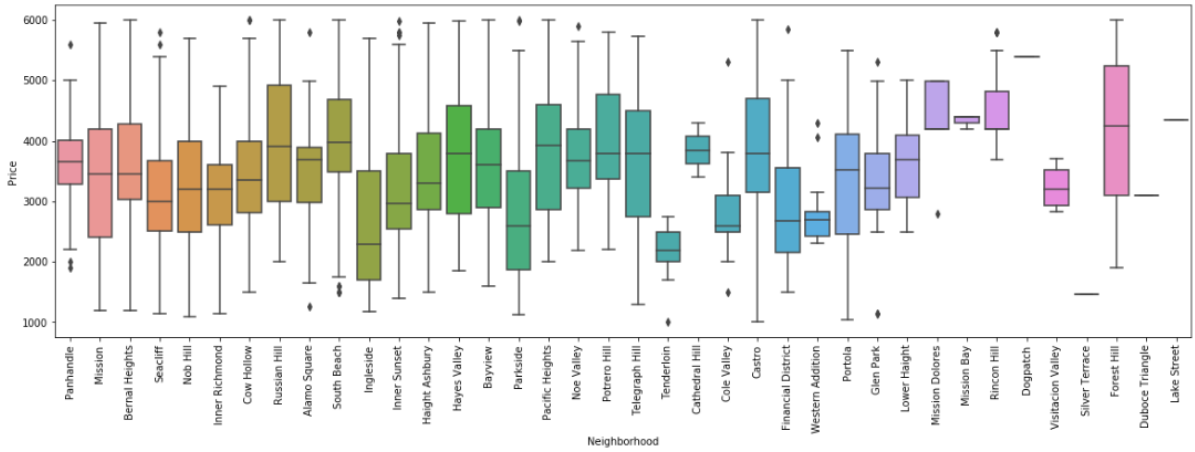
There are four key data components, crime incidents, mean rental price, clustered venues and desired rental properties for completing combined maps. All components need to be grouped and mapped geographic polygons to appear correctly in choropleth map. First crime incidents and mean rental values are grouped and loaded into dataframe respectively. Clustered venues are constructed from the centroids of the neighborhoods. Finally desired rental properties and clustered venues of neighborhoods are superimposed on top of either crime rate or mean rental value in each choropleth map.

San Francisco Housing Rental Price Statistics

Upper 3000 dollar per month rent is around the mean value



Display Boxplot for Each Neighborhood

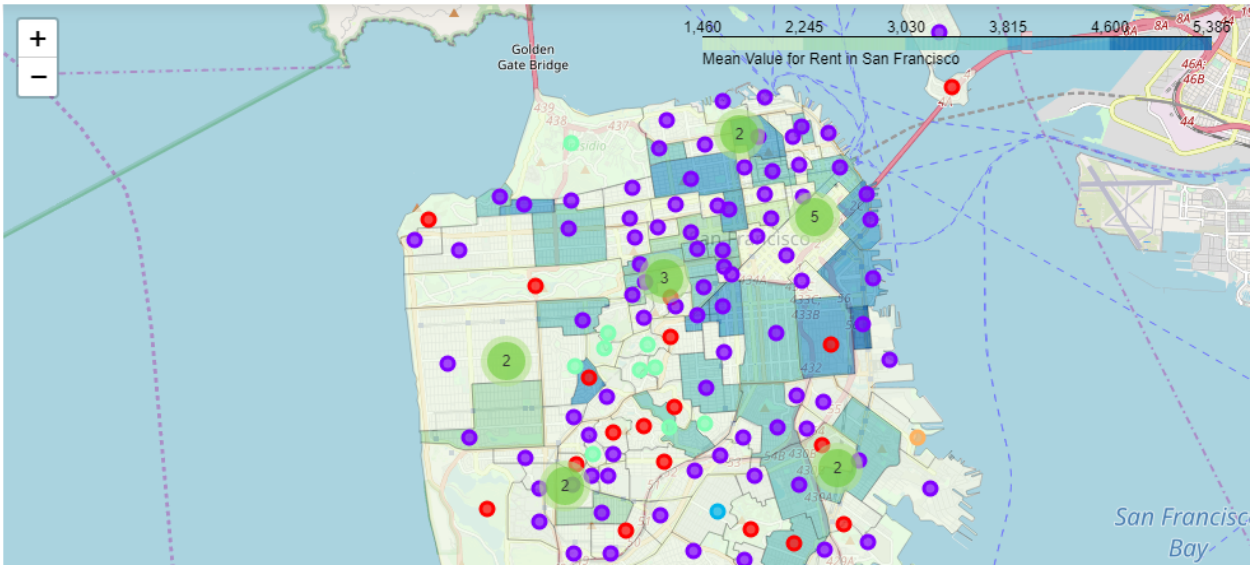


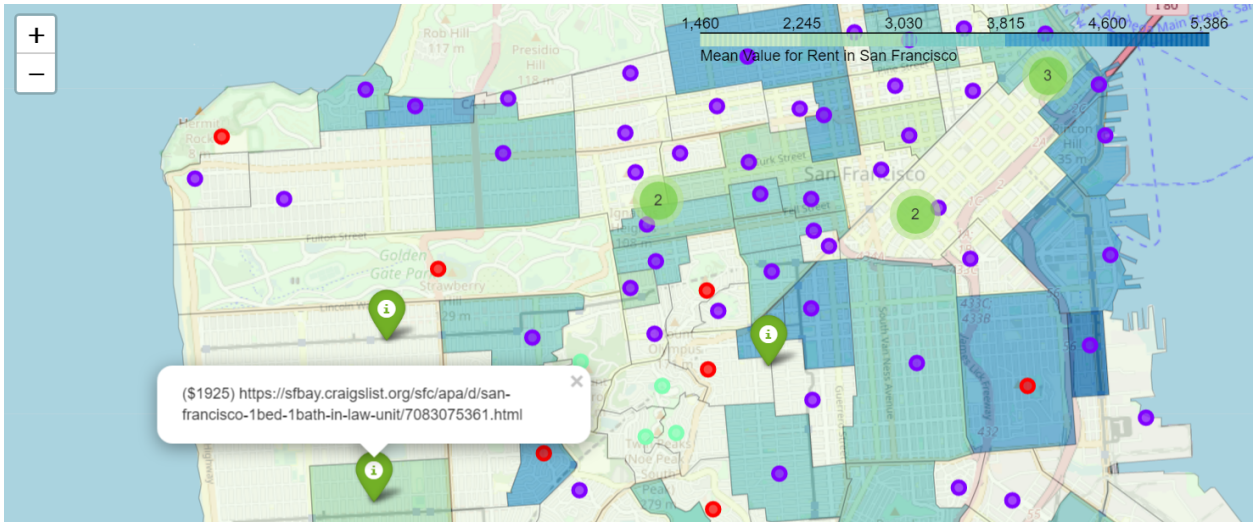
Find desired housing with new search criteria

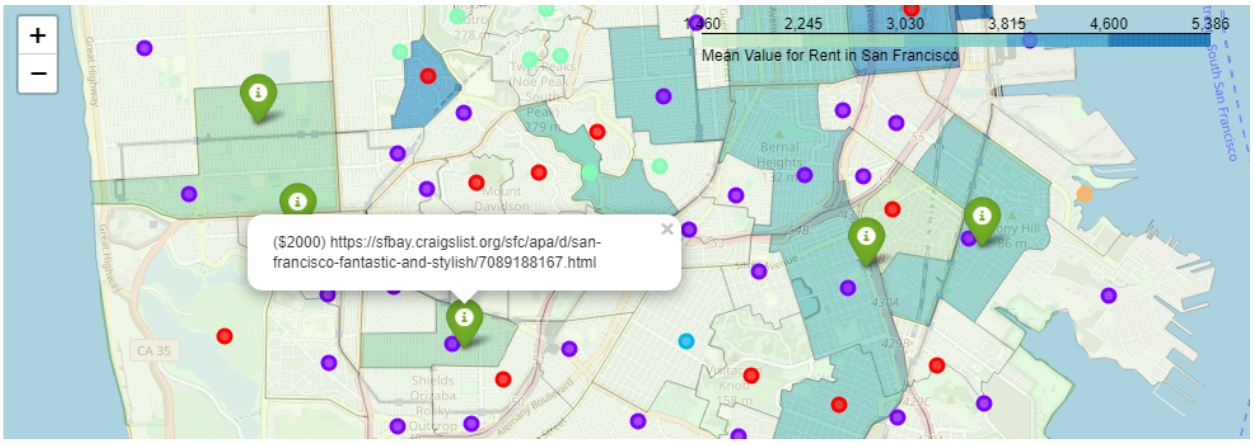
```
cl_h = CraigslistHousing(site='sfbay', area='sfc', category='apa',
                         filters={'max_price': 2200, 'min_price': 1800,
                                  'min_bedrooms':1,  'min_bathrooms':1, 'min_ft2': 650, 'private_bath': True,
                                  'private_room': True})

newResults = cl_h.get_results(sort_by='newest', geotagged=True, limit=100)

newRents = populateAptDF(newResults)
newRents.tail()
```

	Labels	Datetime	Price	Neighborhood	Latitude	Longitude
23	https://sfbay.craigslist.org/sfc/apa/d/san-fra...	2020-03-07 17:36	1800	Cole Valley	37.791500	-122.401800
24	https://sfbay.craigslist.org/sfc/apa/d/fairfie...	2020-03-07 17:36	2150	Financial District	38.242300	-122.131400
25	https://sfbay.craigslist.org/sfc/apa/d/fairfie...	2020-03-07 17:36	2150	Financial District	38.242300	-122.131400
26	https://sfbay.craigslist.org/sfc/apa/d/san-fra...	2020-03-07 17:36	2000	Panhandle	37.777433	-122.448683
27	https://sfbay.craigslist.org/sfc/apa/d/san-fra...	2020-03-07 17:36	2200	Inner Sunset	37.761370	-122.484496

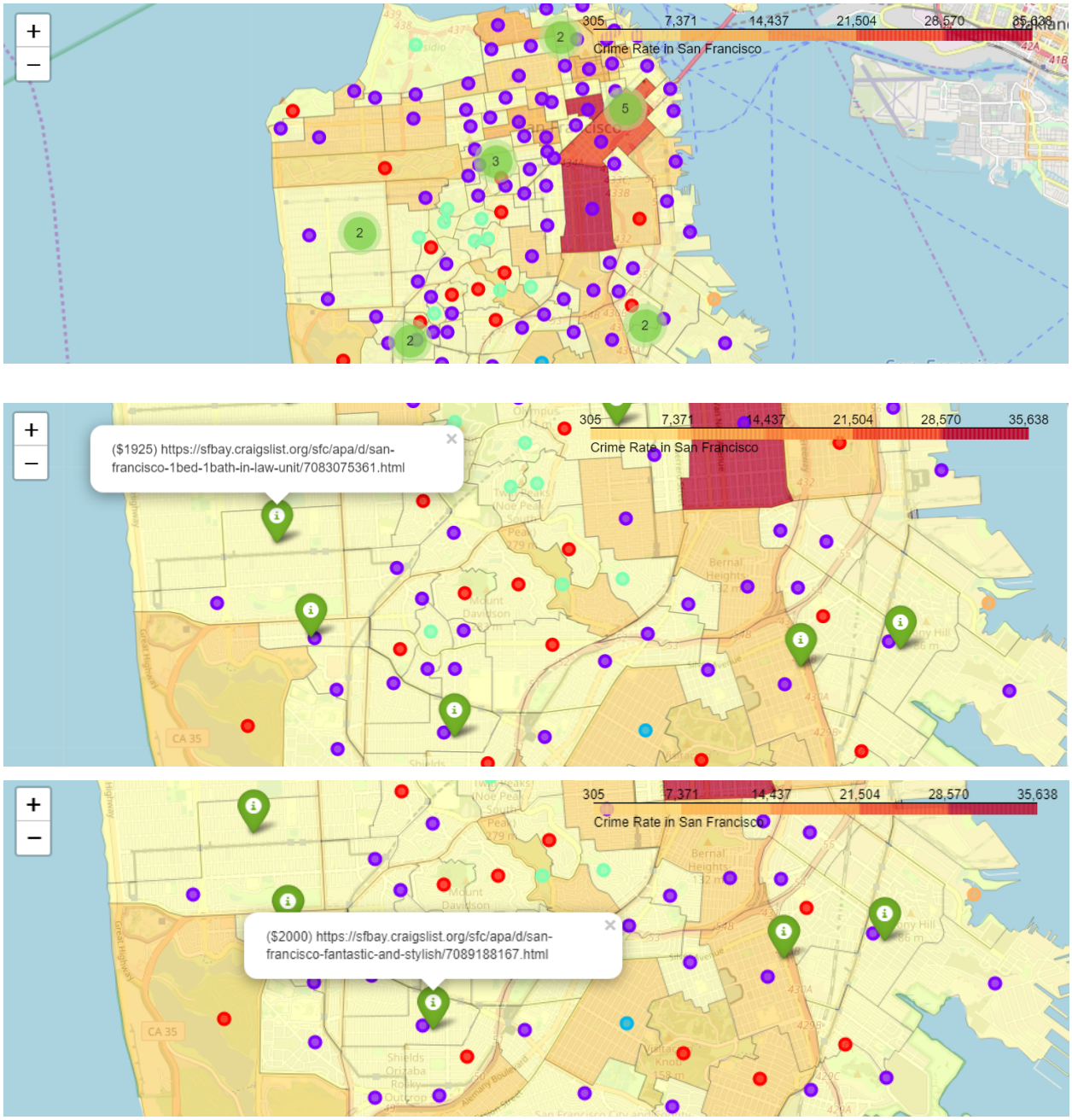
Available Housings and Clustered Venue on Top of Mean Rent Value







Available Housings and Clustered Venue on Top of Crime Rate



4 Execution and Results

Those last two combined choropleth maps provide intuitive available housings on top of mean rental price or crime rate associated with clustered venues. Listed marker shows the rental price as well as the url and trend of the neighborhood if it is safer or more affordable in the

neighborhood. It is also interactive so that a user can easily explore and find one's prefered housing on the fly.

5 Discussion and Observations

In this project, python-craigslist library limits to 3000 data and includes the false lists that do not belong to San Francisco. Often specifying more filters reduces the overall data size. Thus filter and search criteria is intentionally broadened such that the result can be used and compensated to draw enough information in the final choropleth map in one running operation.

6 Conclusion

This project has shown a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence.