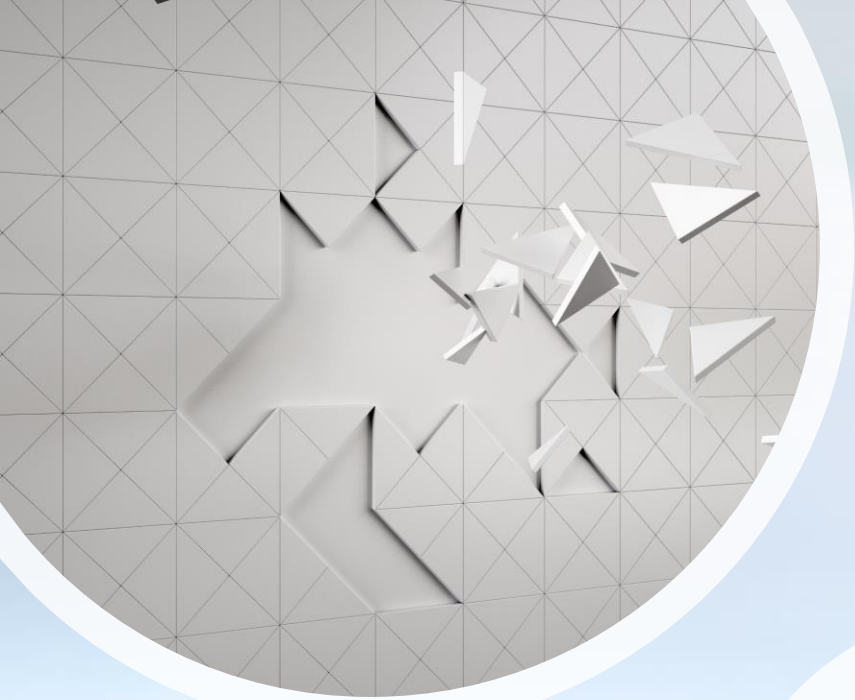
Exploring San Francisco Neighborhoods

Through the lens of police incident reports, citizen reported 311 cases, and housing prices from 2018 to 2020

San Francisco: Fast Facts

- 16th most populous city in the United States
- 4th most populous city in California
- 2019: 7th highest income county in the United States[1]
- 2019: GDP of $203.5 Billion[1]
- Expensive housing prices

[1] Source: Bureau of Economic Analysis (BEA) 2019: https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas



Growing City, Growing Problems

- Crime
- Vehicle break-ins
- Homeless Encampments

- High Housing Prices

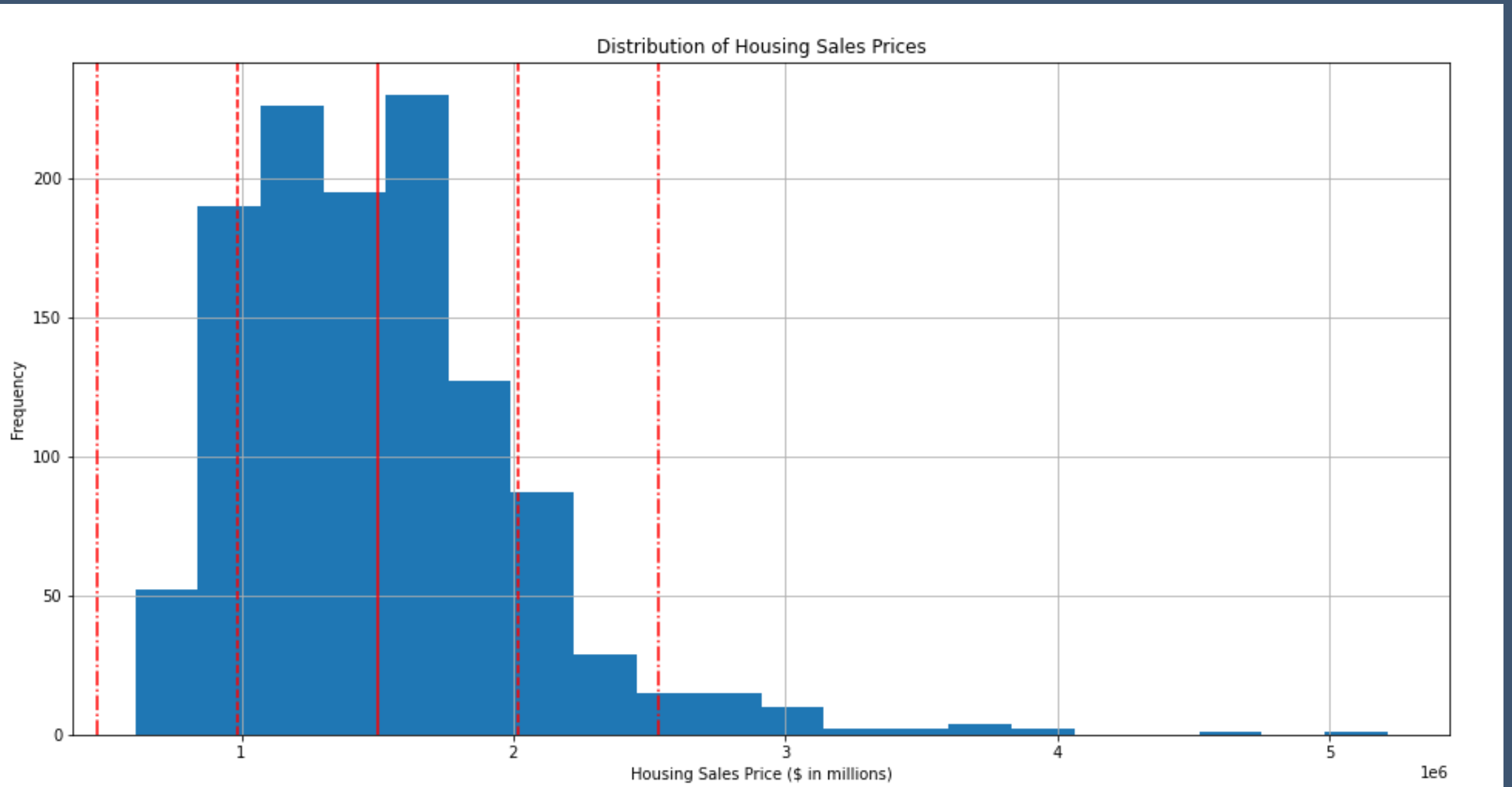
Data Sources

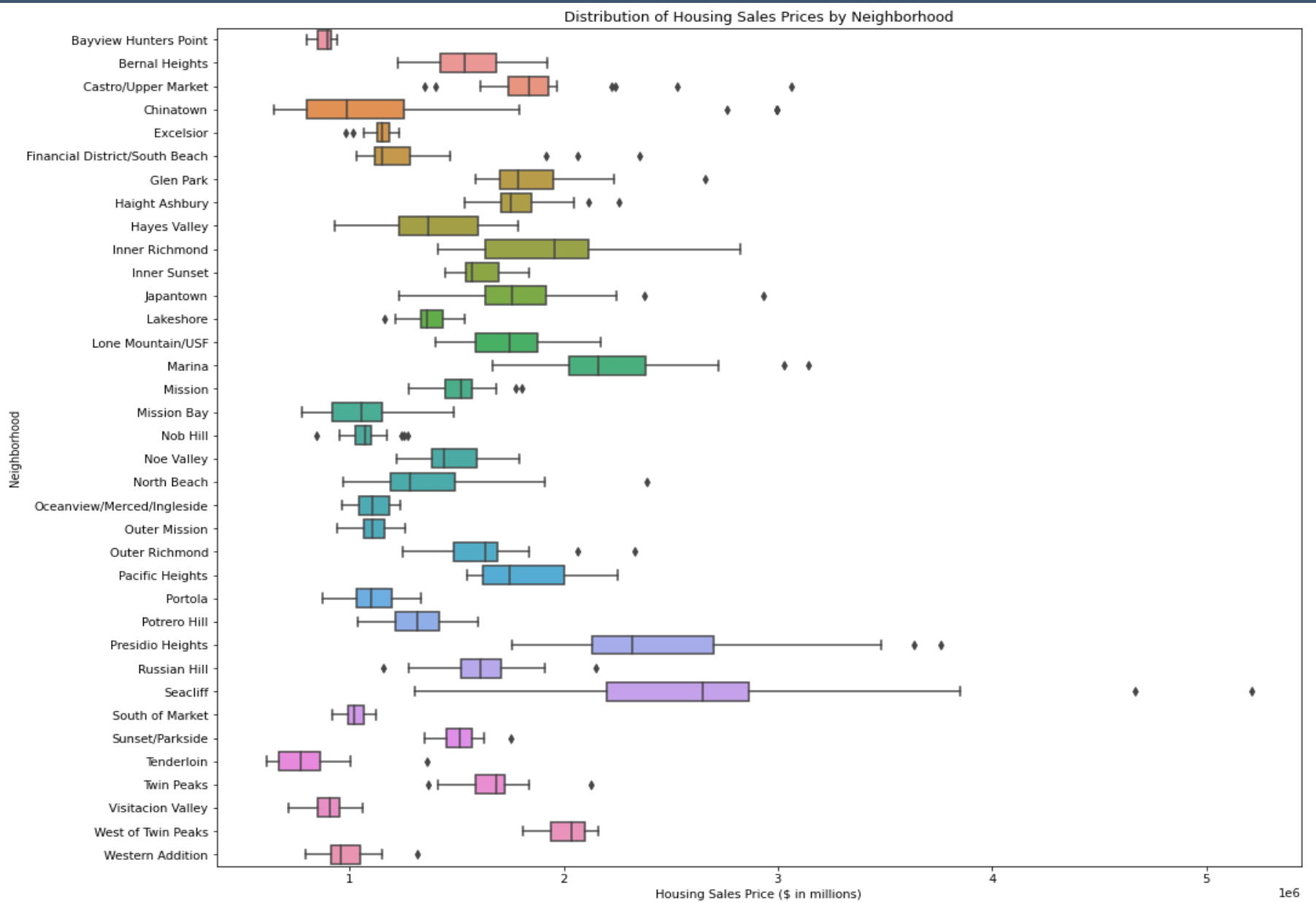
- San Francisco Police Department Incident Reports: 2018 to Present | DataSF | City and County of San Francisco
- 311 Cases | DataSF | City and County of San Francisco
- Redfin's Downloadable Housing Market Data median housing sales prices

Seeking answers to…

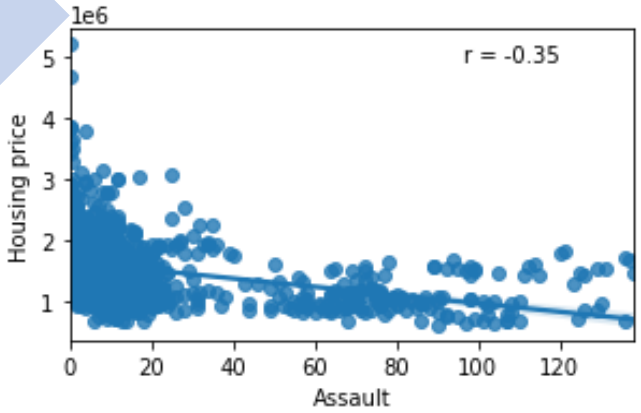
Can we predict housing prices from features derived from the police incident reports and 311 cases?

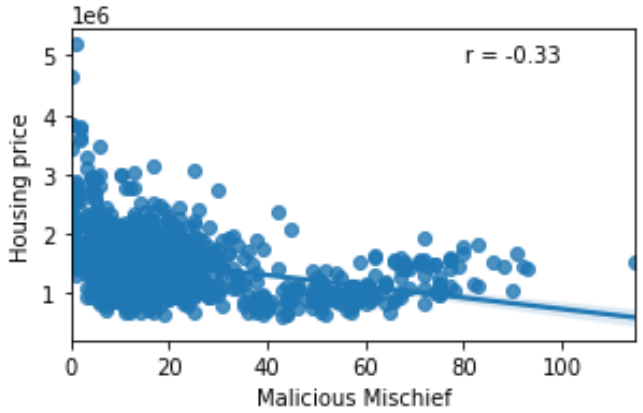
How do the neighborhoods differ from one another?

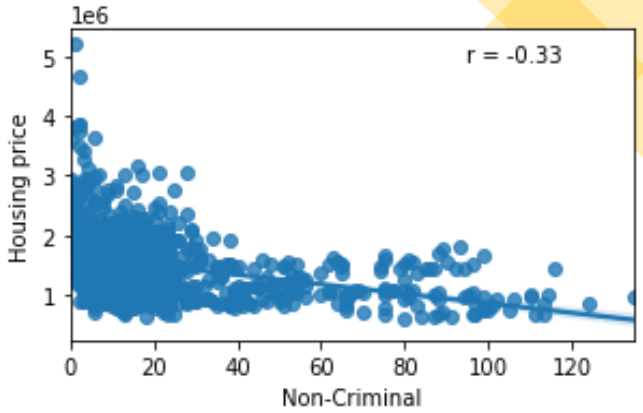

Distribution of Housing Sales Prices
Frequency
0
50
100
150
200
1
2
3
4
5
1e6
Housing Sales Price ($ in millions)

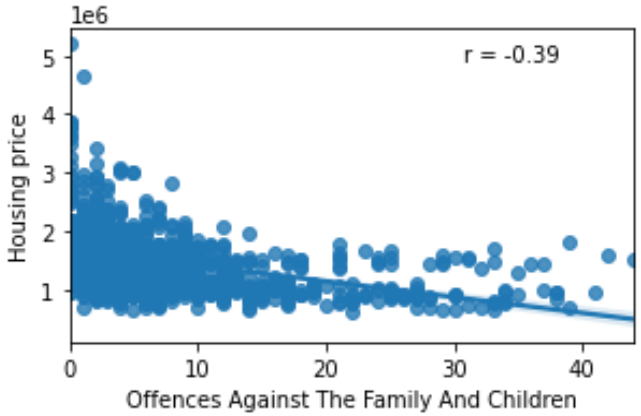

Distribution of Housing Sales Prices by Neighborhood
Bayview Hunters Point
Bernal Heights
Castro/Upper Market
Chinatown
Excelsior
Financial District/South Beach
Glen Park
Haight Ashbury
Hayes Valley
Inner Richmond
Inner Sunset
Japantown
Lakeshore
Lone Mountain/USF
Marina
Mission
Mission Bay
Nob Hill
Noe Valley
North Beach
Oceanview/Merced/Ingleside
Outer Mission
Outer Richmond
Pacific Heights
Portola
Potrero Hill
Presidio Heights
Russian Hill
Seacliff
South of Market
Sunset/Parkside
Tenderloin
Twin Peaks
Visitacion Valley
West of Twin Peaks
Western Addition
Neighborhood
1
2
3
4
5
Housing Sales Price ($ in millions)
1e6

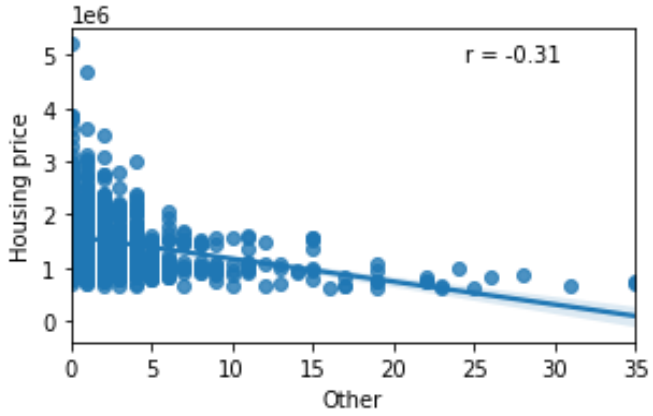
Highest Negative Correlation of Features with Housing Prices

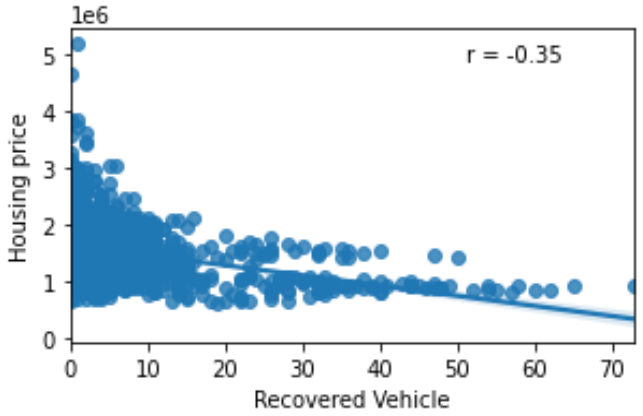


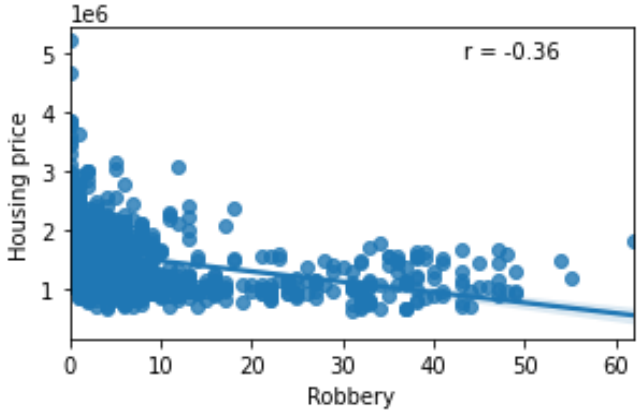


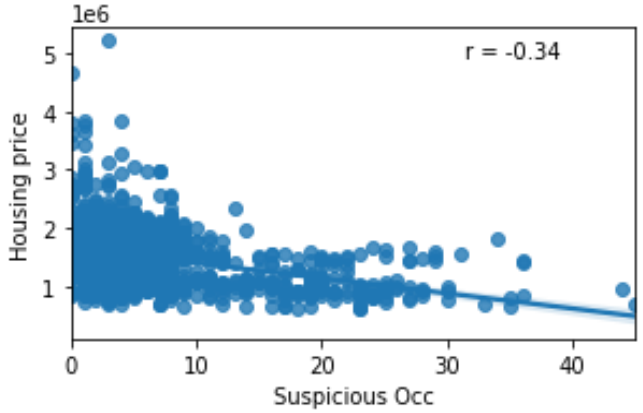


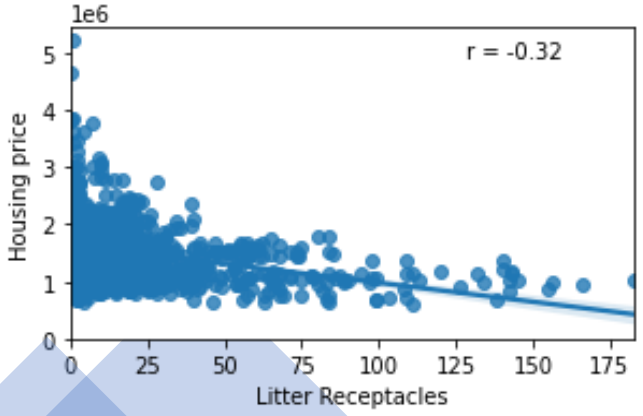








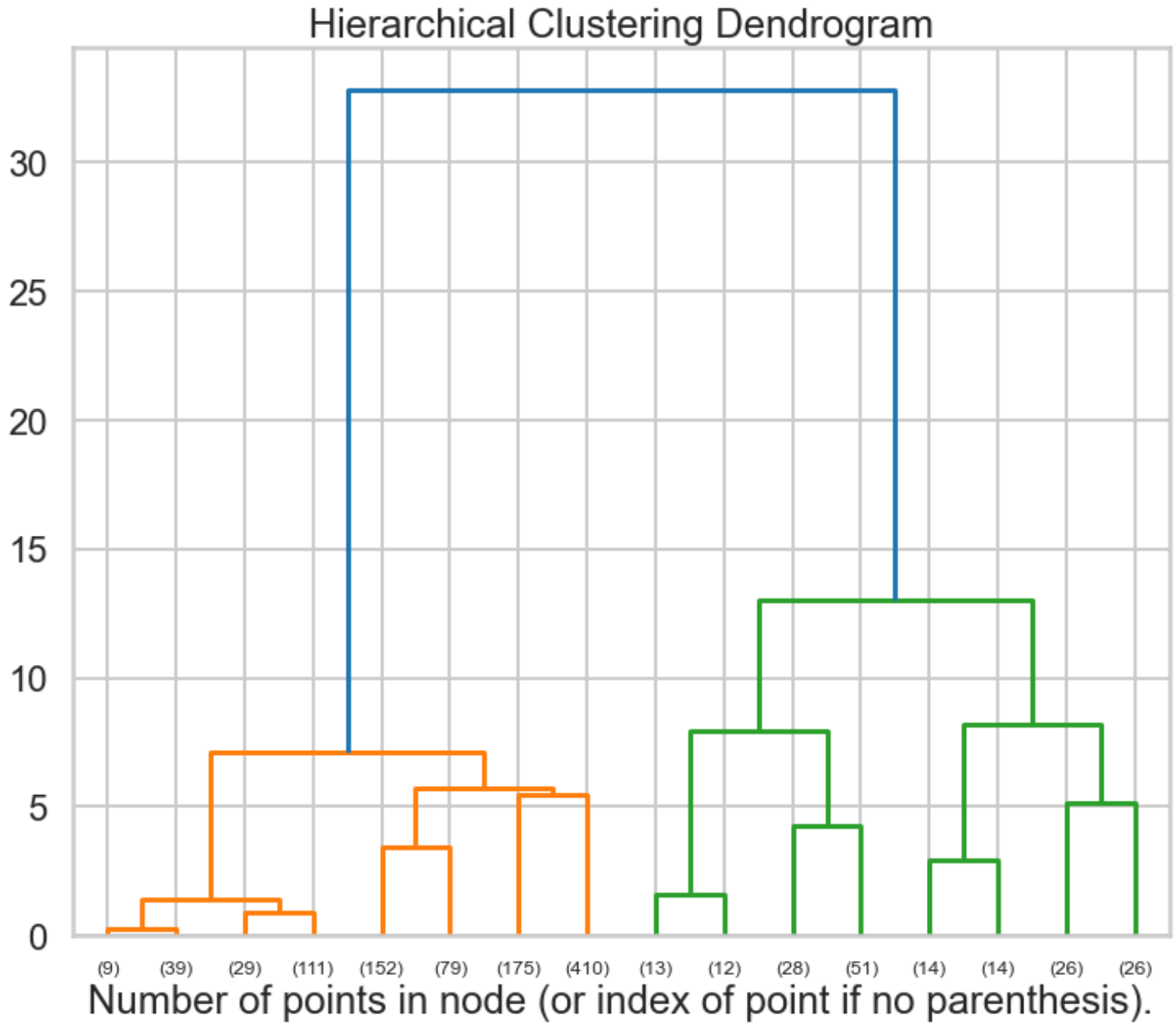




Modeling – Predicting Housing Prices

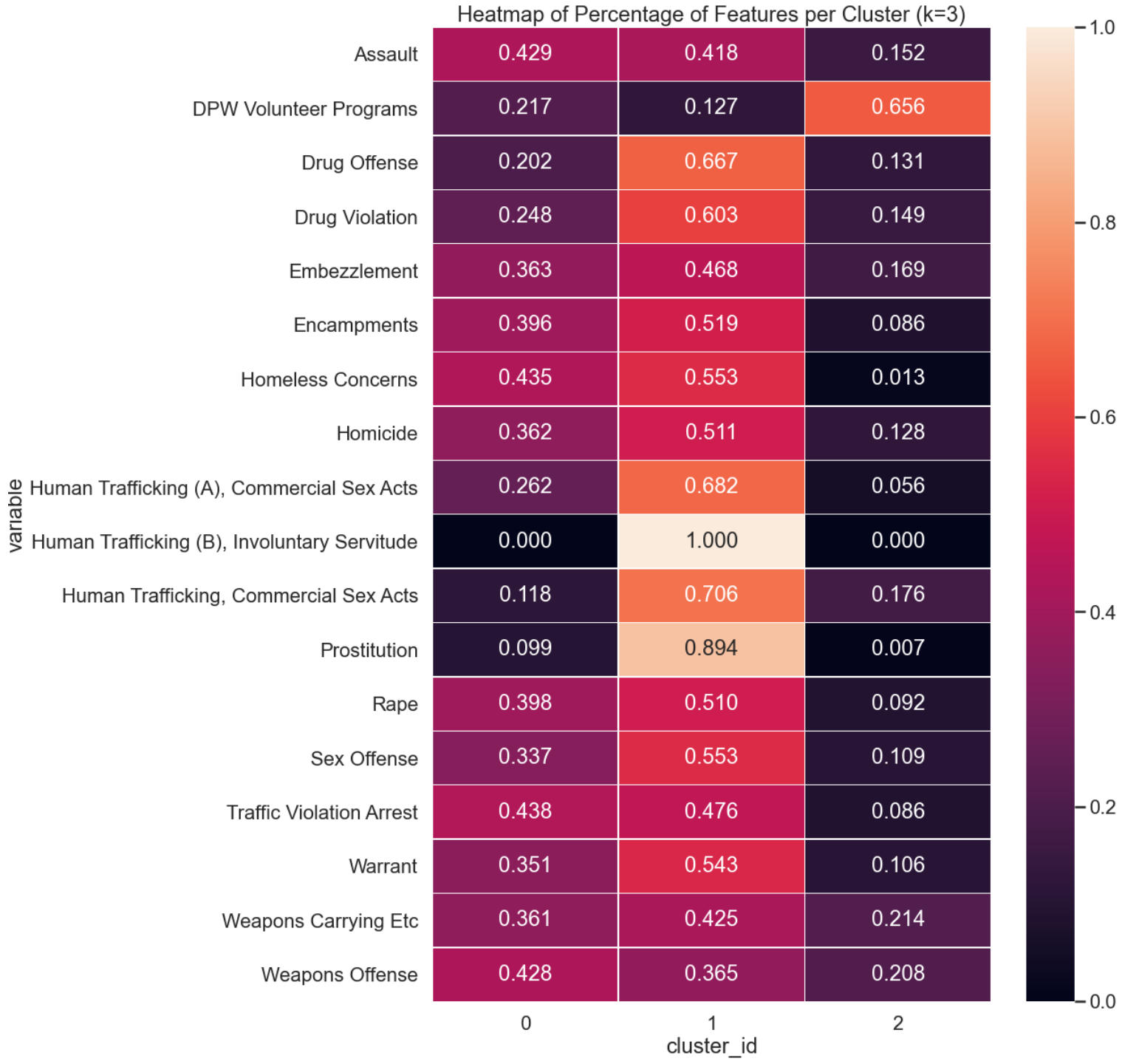
- Best Model after using GridSearchCV to tune hyperparameters:
  - Min Max Scaler
  - PCA, n_components = 7
  - Random Forest Regressor, max_depth = 10, n_estimator = 31
- R-squared = 0.349708
- Mean Absolute Error (MAE): $247,694.77
- Sqrt(Mean Squared Error): $365,848.02

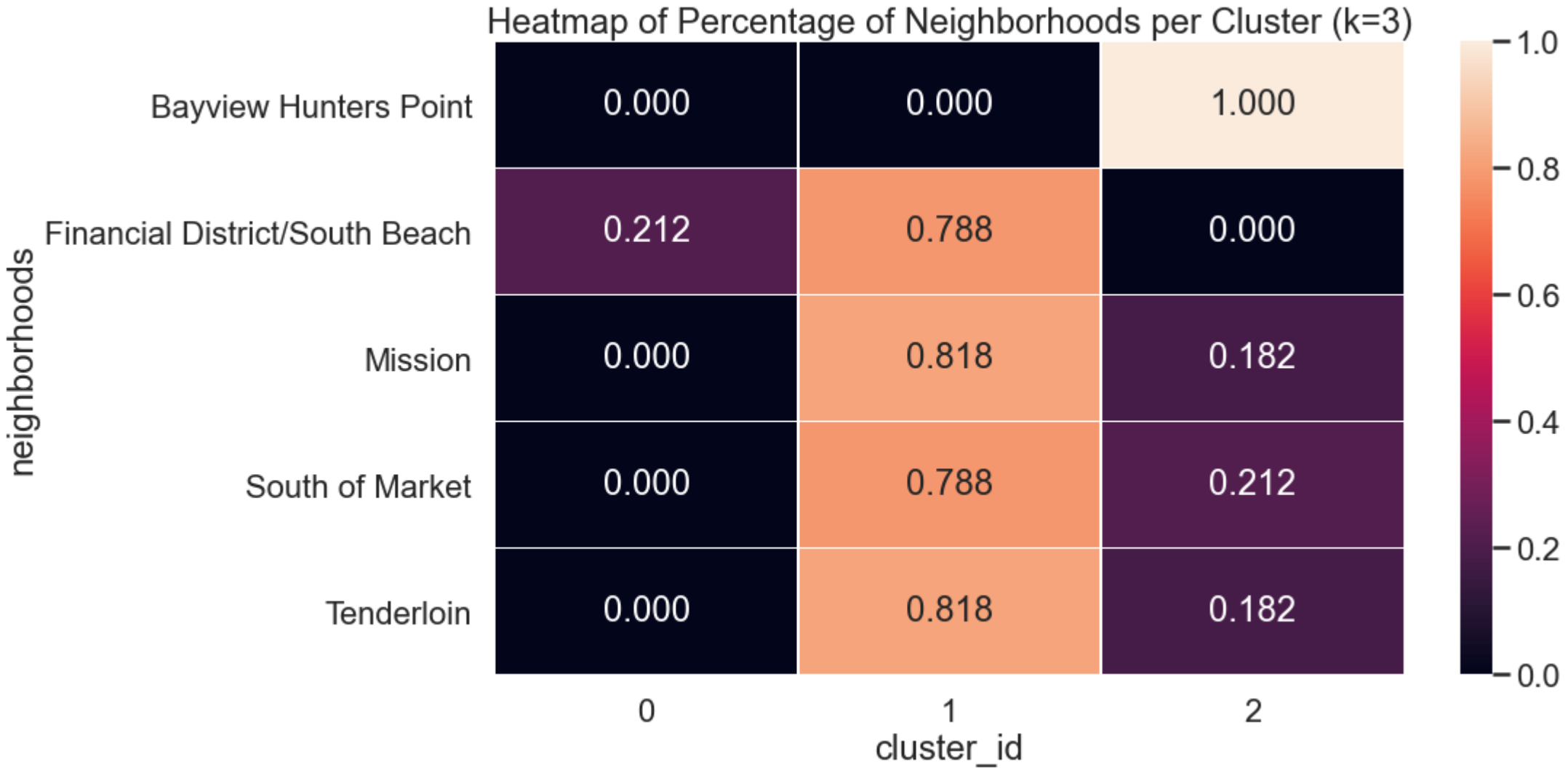
Modeling – Clustering Neighborhoods



Features of neighborhoods in clusters

- K-Means with n_clusters = 3
- Cluster_id = 0 : default catch-all
- Cluster_id = 1 : violent crimes
- Cluster_id = 2 : DPW Volunteer Programs and nearly no Encampments or Homeless Concerns




Heatmap of Percentage of Neighborhoods per Cluster (k=3)
neighborhoods
Bayview Hunters Point
Financial District/South Beach
Mission
South of Market
Tenderloin
0.000
0.000
1.000
0.212
0.788
0.000
0.000
0.818
0.182
0.000
0.788
0.212
0.000
0.818
0.182
0
1
2
cluster_id
1.0
0.8
0.6
0.4
0.2
0.0

Further Research

1. Increase time frame of data
2. Improve model – clustering : increase granularity of police incident and 311 case data
3. Improve model – predicting housing prices : add property data