

Dataset A (Crime Rates) + SVM

CISC 873 Hannah LeBlanc

Crime Rates

Robbery, Property, Burglary, Larceny Theft, Motor Vehicle Theft

All rates in 100 000 of population / year

Property crime is a category of crime that includes, among other crimes, burglary, larceny, theft, motor vehicle theft, arson, shoplifting, and vandalism. Property crime is a crime to obtain money, property, or some other benefit. This may involve force, or the threat of force, in cases like robbery or extortion. Since these crimes are committed in order to enrich the perpetrator they are considered property crimes.

Missing Values

Train size: (504, 120)

Test size: (88, 120)

Train size with dropna(any): (185, 120)

Train size with dropna(all): (504, 120)

Percent RACE - Native Hawaiian and Other Pacific Islander	248
Percent; GROSS RENT - \$1,000 to \$1,499	136
Percent; GROSS RENT - Less than \$200	136
Percent; GROSS RENT - \$200 to \$299	136
Percent; GROSS RENT - \$300 to \$499	136
Percent; GROSS RENT - \$500 to \$749	136
Percent; GROSS RENT - \$750 to \$999	136
Percent; GROSS RENT - \$1,500 or more	136
Separated; Percent of Population 15 years and over	41
Never married; Percent of Population 15 years and over	41
Widowed; Percent of Population 15 years and over	41
Divorced; Percent of Population 15 years and over	41
Now married (except separated); Percent of Population 15 years and over	41
Percent RACE - White	18
Percent RACE - Some other race	18
Percent RACE - Asian	18
Percent RACE - American Indian and Alaska Native	18
Percent RACE - Black or African American	18
Unemployment rate in Population 16 years and over	15
Percent; EMPLOYMENT STATUS - In labor force	15
Estimate; HOUSING OCCUPANCY - Rental vacancy rate	14
Total civilians employee	10
Percent; OCCUPATION - Production, transportation, and material moving occupations	10
Percent; OCCUPATION - Sales and office occupations	10
Percent; OCCUPATION - Service occupations	10
Percent; OCCUPATION - Management, business, science, and arts occupations	10
Total law_enforcement_employees	10
Law_enforcement_Total officers employee	10
Civilian employed population 16 years and over - Self-employed workers in own not incorporated business	5
Civilian employed population 16 years and over - Government workers	5

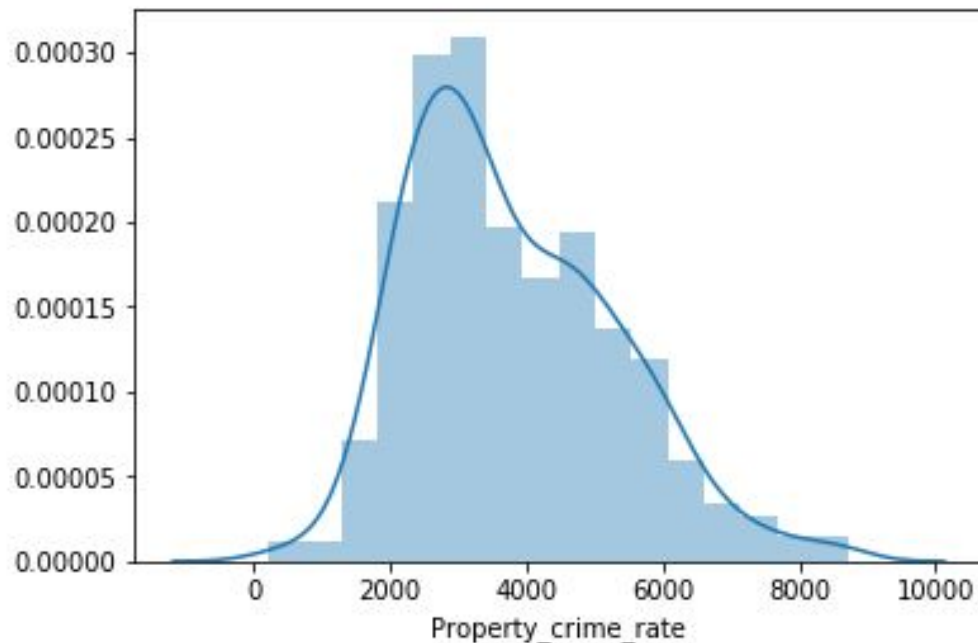
Missing Values

City	Count
City1	2
City10	2
City100	2
City101	2
City102	2
City103	2
City104	2
City105	1
City106	2
City107	2
City108	2
City109	2

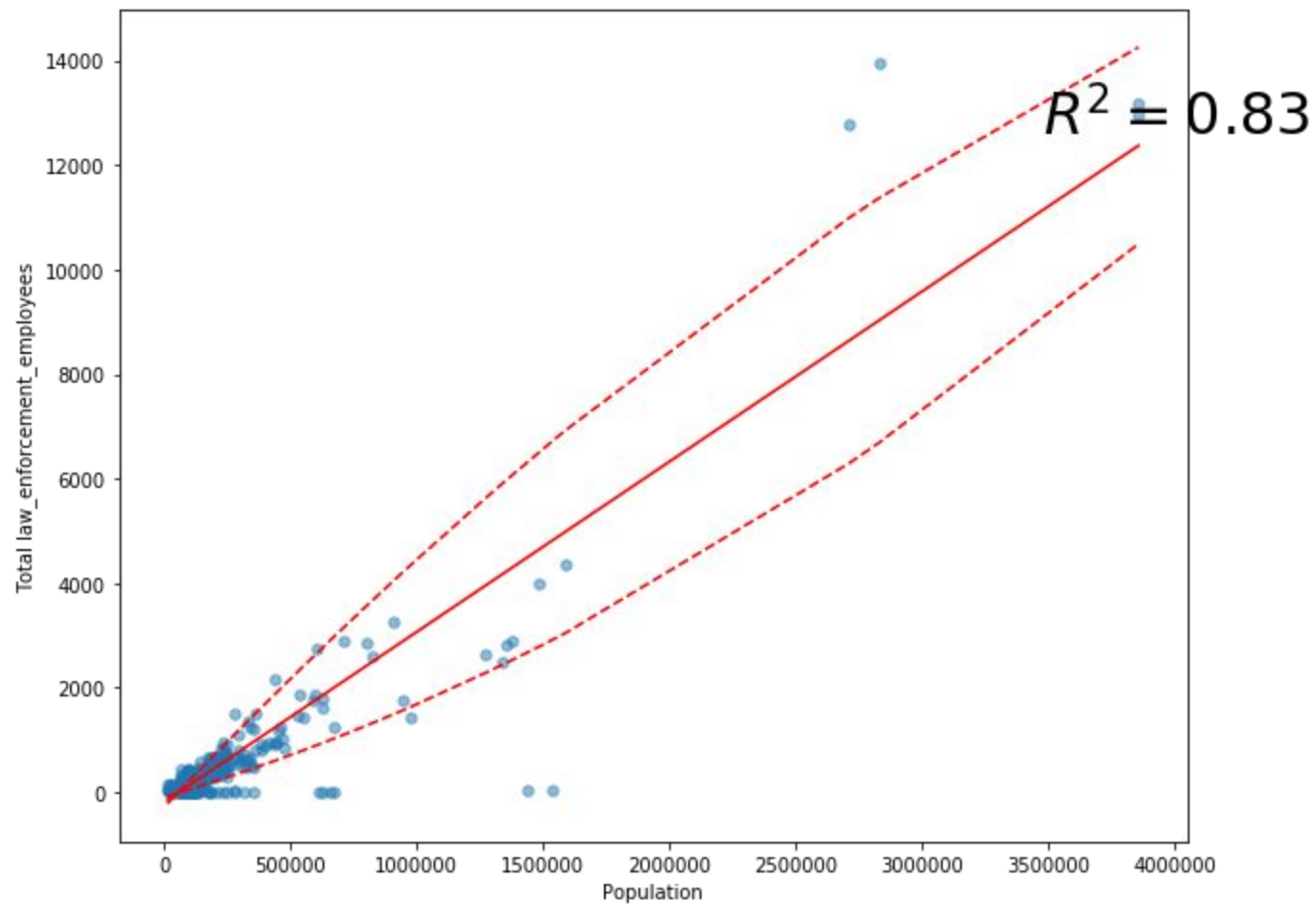
```
def impute_from_city(df, city_df):  
    #for each city  
    for i, row in df.iterrows():  
        #getting number of rows for City  
        other_rows = city_df.loc[row["City"]]["Count"]  
        if other_rows > 1:  
            # list of rows for this city  
            roi = [x for x in list(df.index[df.City == row.City]) if x != i]  
            # get row in df with same city but not same index  
            other_row = df.iloc[roi[0]]  
            df.iloc[i] = row.fillna(other_row)  
    return df
```

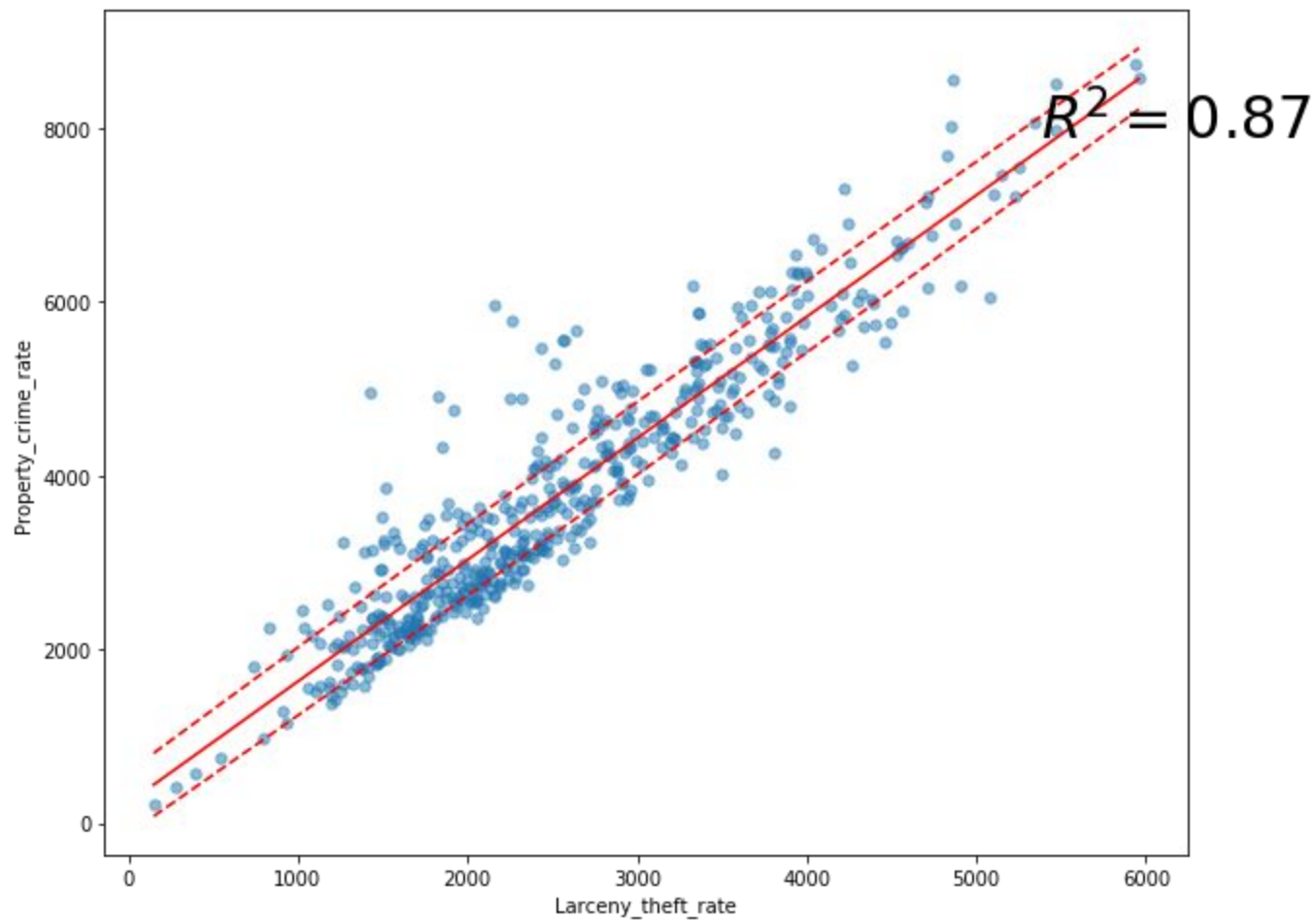
Percent; GROSS RENT - \$1,500 or more	115
Percent; GROSS RENT - Less than \$200	115
Percent; GROSS RENT - \$200 to \$299	115
Percent; GROSS RENT - \$300 to \$499	115
Percent; GROSS RENT - \$500 to \$749	115
Percent; GROSS RENT - \$750 to \$999	115
Percent; GROSS RENT - \$1,000 to \$1,499	115
Now married (except separated); Percent of Population 15 years and over	29
Widowed; Percent of Population 15 years and over	29
Divorced; Percent of Population 15 years and over	29
Separated; Percent of Population 15 years and over	29
Never married; Percent of Population 15 years and over	29
Percent RACE - Native Hawaiian and Other Pacific Islander	16
Percent RACE - American Indian and Alaska Native	9
Percent RACE - Black or African American	9
Percent RACE - White	9
Percent RACE - Asian	9
Percent RACE - Some other race	9
Percent; EMPLOYMENT STATUS - In labor force	6
Unemployment rate in Population 16 years and over	6
Percent; OCCUPATION - Management, business, science, and arts occupations	5
Percent; OCCUPATION - Production, transportation, and material moving occupations	5
Percent; OCCUPATION - Sales and office occupations	5
Percent; OCCUPATION - Service occupations	5
Civilian employed population 16 years and over - Private wage and salary workers	4
Civilian employed population 16 years and over - Unpaid family workers	4
Civilian employed population 16 years and over - Self-employed workers in own not incorporated business	4
Civilian employed population 16 years and over - Government workers	4
Estimate; HOUSING OCCUPANCY - Rental vacancy rate	4
Total law_enforcement_employees	3

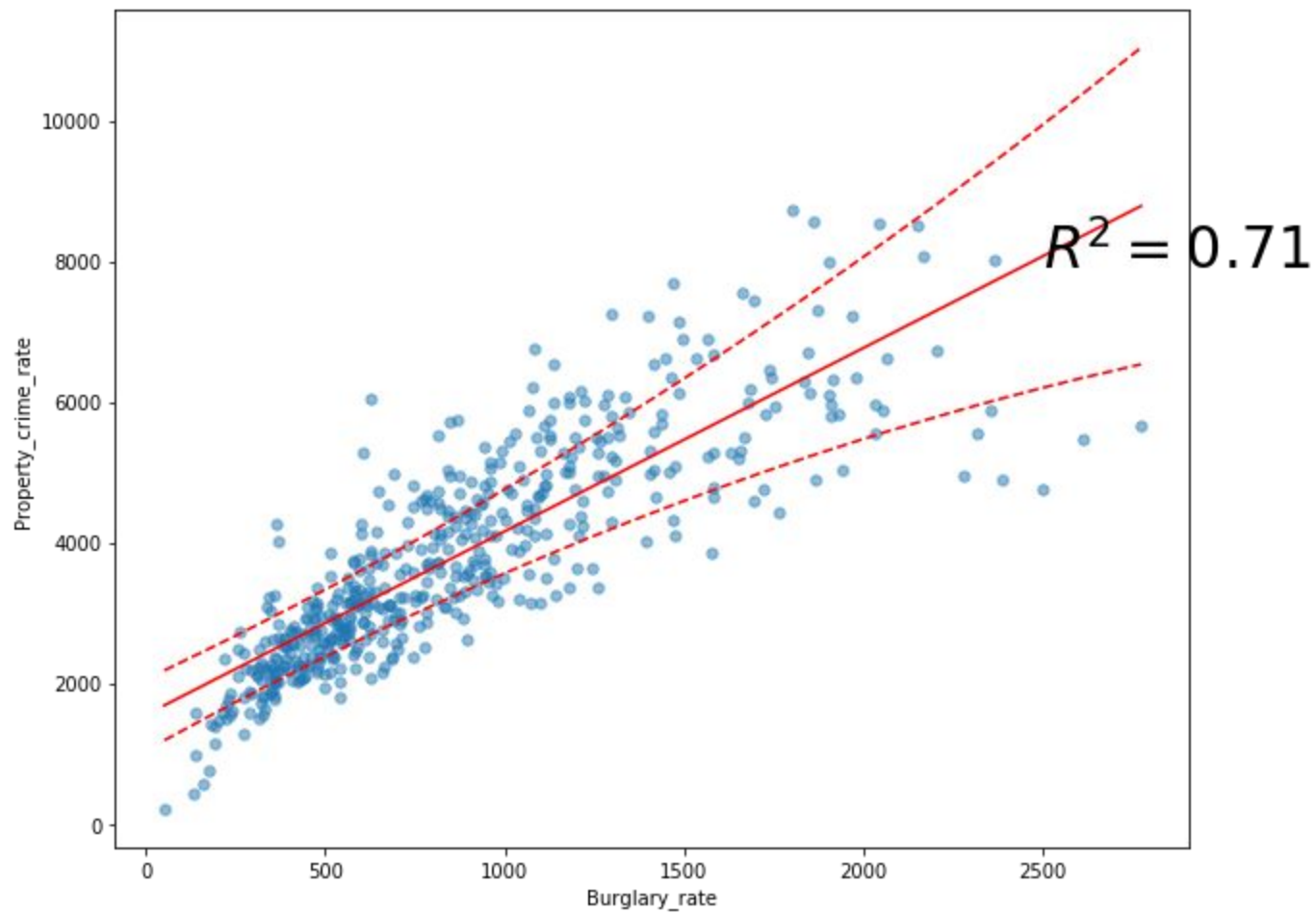
Property Crime Rate

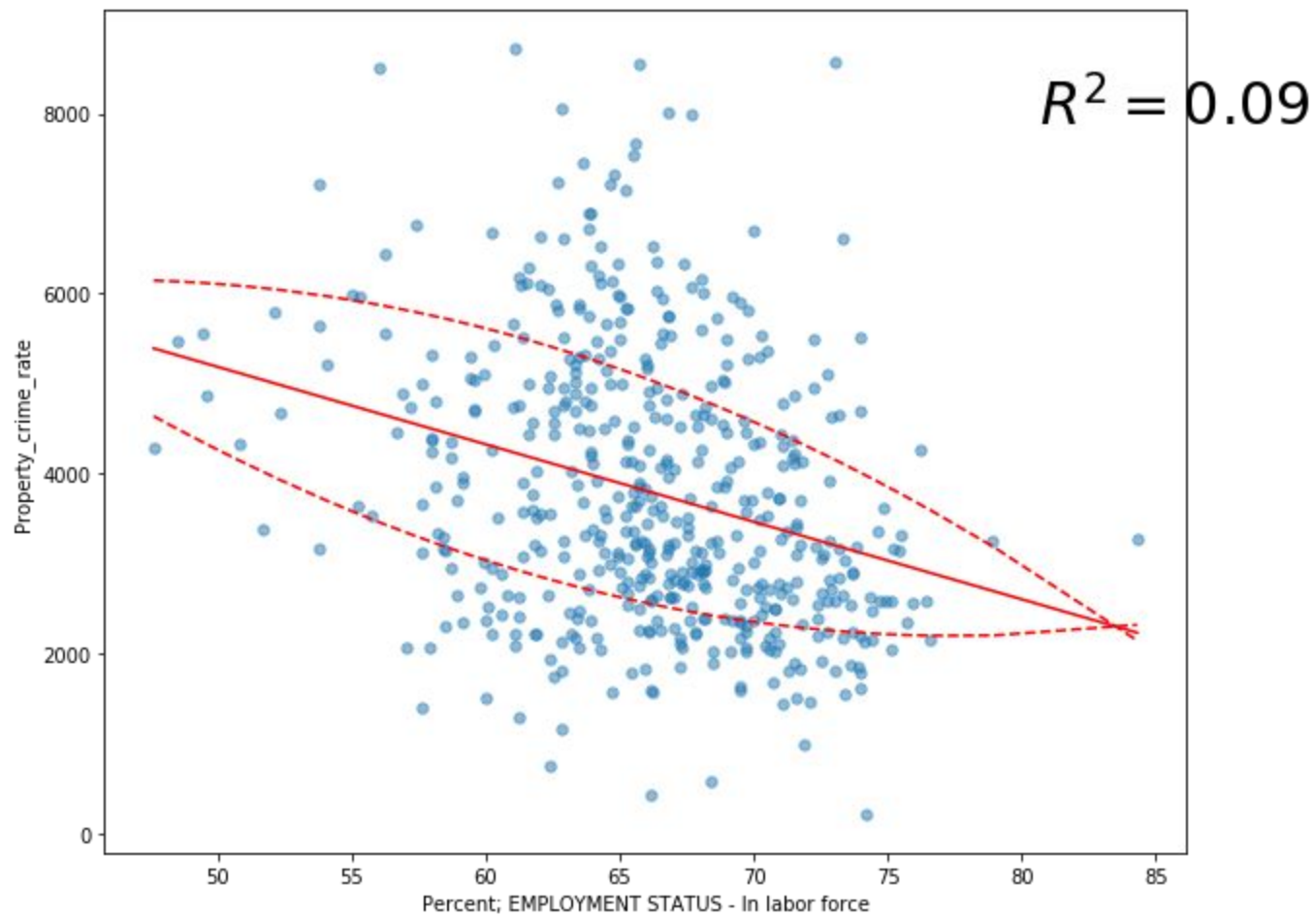


```
count      504.000000  
mean      3790.993837  
std       1519.426682  
min        210.700000  
25%       2623.800000  
50%       3509.250000  
75%       4809.450000  
max       8730.100000  
Name: Property_crime_rate,
```







SVR

```
model2 = SVR(kernel='rbf')
```

```
X = df_train.drop(columns=['Robbery_rate', 'Property_crime_rate', 'Burglary_rate',  
                           | 'Larceny_theft_rate', 'Motor_vehicle_theft_rate', 'City']).values  
y = df_train['Property_crime_rate'].values
```

Train

R²: -0.0288

MSE: 2618091.8138

Test

R²: -0.032

MSE: 2284217.113

StandardScaler

```
model3 = SVR(kernel='rbf',C=1000)
```

Before

Train

R²: 0.7382

MSE: 666260.0587

Test

R²: 0.4861

MSE: 1137586.0277

After

Train

R²: 0.6481

MSE: 895453.6127

Test

R²: -0.0225

MSE: 2263290.6344

5 Fold GridSearchCV

```
# Tuning hyper-parameters for r2
```

```
Best parameters set found on development set:
```

```
{'C': 10, 'kernel': 'linear'}
```

Train

R²: 0.6277

MSE: 947301.3764

Test

R²: 0.383

MSE: 1365801.2983

```
# Tuning hyper-parameters for neg_mean_squared_error
```

```
Best parameters set found on development set:
```

```
{'C': 10, 'kernel': 'linear'}
```