CrimeRateAnalysisCode

December 11, 2022

[1]:
```
# Import packages

import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import normalize
from matplotlib.pyplot import figure
# from linearmodels import FamaMacBeth
from sklearn.preprocessing import MinMaxScaler
import statsmodels.formula.api as smf
from scipy import stats
import statsmodels.api as sm
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from numpy import mean, std, absolute, sqrt
```

[2]:
```
# Change this to run in your environment

absolute_file_path = '~/Downloads/DMProjectRichaRishabh'
```

[3]:
```
# check our accuracy for each degree, the lower the error the better!

def fetch_best_degree_regression_model(city_data, normalize_original_data,␣
↪for_our_model=True):
    number_degrees = [1,2,3,4,5,6]
    plt_mean_squared_error = []

    if normalize_original_data:
        y_values = city_data['Scaled_CrimeCount']
        if for_our_model:
            x_values = city_data.loc[:,['Gender_ratio', 'Scaled_Unemployment',␣
↪'Scaled_MedianIncomeRate', 'WorkforceCount', 'PercentNegativeUsers',␣
↪'FavorOfDemocrats']]
        else:
```

```
            x_values = city_data.loc[:,['Gender_ratio', 'Scaled_Unemployment',␣
↪'Scaled_MedianIncomeRate']]
    else:
        y_values = city_data['Cul. Crime count']
        if for_our_model:
            x_values = city_data.loc[:,['Gender_ratio', 'UnemploymentRate',␣
↪'MedianIncomeRate', 'WorkforceCount', 'PercentNegativeUsers',␣
↪'FavorOfDemocrats']]
        else:
            x_values = city_data.loc[:,['Gender_ratio', 'UnemploymentRate',␣
↪'MedianIncomeRate']]

    for degree in number_degrees:

        poly_model = PolynomialFeatures(degree=degree)

        poly_x_values = poly_model.fit_transform(x_values)
        poly_model.fit(poly_x_values, y_values)

        regression_model = LinearRegression()
        regression_model.fit(poly_x_values, y_values)
        y_pred = regression_model.predict(poly_x_values)

        plt_mean_squared_error.append(mean_squared_error(y_values, y_pred,␣
↪squared=False))

    plt.scatter(number_degrees,plt_mean_squared_error, color="green")
    plt.plot(number_degrees,plt_mean_squared_error, color="red")
    plt.title("Polynomial Regression MSE at different degrees")
    plt.show()

#     return number_degrees[plt_mean_squared_error.
↪index(min(plt_mean_squared_error))]
    return 2
```

[4]:
```
# Plot normalized histogram for model coefficients. coeff_params: dict of␣
↪coefficients with values

def plot_hist_for_model_coefficients(coeff_params, p_vals,␣
↪min_max_normalization=False, save_fig=False, city_name="", time_period=""):
    i = 0
    coeff_params_significant = {}
    for k, v in coeff_params.items():
        if p_vals[i] <= 0.05:
            coeff_params_significant[k] = v
        i += 1
```

```
    coeff_params_significant = dict(sorted(coeff_params_significant.items(),␣
↪key=lambda item: item[1], reverse=False))
    if min_max_normalization:
        scaler = MinMaxScaler()
        norm_coeff_values = scaler.fit_transform(pd.
↪DataFrame(coeff_params_significant, index=[0]))
        norm_coeff_values = [i[0] for i in norm_coeff_values]
    else:
        norm_coeff_values = list(dict(coeff_params_significant).values())
    coeff_vars = list(dict(coeff_params_significant).keys())
#     figure(figsize=(10, 8), dpi=80)
    plt.barh(coeff_vars, norm_coeff_values)
    plt.title("Attributes with significant p-values contributing towards crime␣
↪rate")
    plt.xlabel("Coefficient Values")
    plt.ylabel("Attributes")
    if save_fig:
        plt.savefig('{}_P_VAL_{}.png'.format(city_name, time_period),␣
↪transparent=True, bbox_inches='tight')
    plt.show()
```

[5]:
```
def remove_interaction_terms(items_list, indices):
    res = []
    for idx, ele in enumerate(items_list):
        if idx not in indices:
            res.append(ele)
    return res
```

[6]:
```
# Removing interaction terms with less p-val in LA
# Parameters: x=transformed x, z=interaction variable names
# Removing following variables from Polynomial Equation
# 3: WorkforceCount
# 7: Gender_ratio^2
# 10: Gender_ratio WorkforceCount
# 11: Gender_ratio PercentNegativeUsers
# 12: Gender_ratio FavorOfDemocrats
# 13: Gender_ratio SearchCountForDepression
# 14: UnemploymentRate^2
# 15: UnemploymentRate MedianIncomeRate
# 17: UnemploymentRate PercentNegativeUsers
# 18: UnemploymentRate FavorOfDemocrats
# 19: UnemploymentRate SearchCountForDepression
# 20: MedianIncomeRate^2
# 22: MedianIncomeRate PercentNegativeUsers
# 23: MedianIncomeRate FavorOfDemocrats
# 24: MedianIncomeRate SearchCountForDepression
# 25: WorkforceCount^2
```

```
# 28: WorkforceCount SearchCountForDepression
# 30: PercentNegativeUsers FavorOfDemocrats
# 31: PercentNegativeUsers SearchCountForDepression

def remove_interaction_terms_for_LA(x, z):
    new_x_ = []
    interaction_terms_indices = [3, 7, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20,␣
 ↪22, 23, 24, 25, 28, 30, 31]
    for item in x:
        temp = remove_interaction_terms(item, interaction_terms_indices)
        new_x_.append(temp)

    new_x_ = pd.DataFrame(np.array(new_x_), columns=remove_interaction_terms(z,␣
 ↪interaction_terms_indices))
    return new_x_
```

[7]:
```
# Removing interaction terms with less p-val in NYC
# Parameters: x=transformed x, z=interaction variable names
# Removing following variables from Polynomial Equation
# 1: UnemploymentRate
# 6: SearchCountForDepression
# 11: Gender_ratio PercentNegativeUsers
# 13: Gender_ratio SearchCountForDepression
# 14: UnemploymentRate^2
# 17: UnemploymentRate PercentNegativeUsers
# 19: UnemploymentRate SearchCountForDepression
# 21: MedianIncomeRate WorkforceCount
# 22: MedianIncomeRate PercentNegativeUsers
# 24: MedianIncomeRate SearchCountForDepression
# 26: WorkforceCount PercentNegativeUsers
# 28: WorkforceCount SearchCountForDepression
# 30: PercentNegativeUsers FavorOfDemocrats
# 31: PercentNegativeUsers SearchCountForDepression
# 33: FavorOfDemocrats SearchCountForDepression
# 34: SearchCountForDepression^2

def remove_interaction_terms_for_NYC(x, z):
    new_x_ = []
    interaction_terms_indices = [1, 6, 11, 13, 14, 17, 19, 21, 22, 24, 26, 28,␣
 ↪30, 31, 33, 34]
    for item in x:
        temp = remove_interaction_terms(item, interaction_terms_indices)
        new_x_.append(temp)

    new_x_ = pd.DataFrame(np.array(new_x_), columns=remove_interaction_terms(z,␣
 ↪interaction_terms_indices))
    return new_x_
```

[8]:
```python
# Removing interaction terms with less p-val in CHICAGO
# Parameters: x=transformed x, z=interaction variable names

def remove_interaction_terms_for_CHICAGO(x, z):
    new_x_ = []
    interaction_terms_indices = [1, 8, 14, 17, 19, 22, 24, 26, 28, 29, 30, 31,␣
↪33]
    for item in x:
        temp = remove_interaction_terms(item, interaction_terms_indices)
        new_x_.append(temp)

    new_x_ = pd.DataFrame(np.array(new_x_), columns=remove_interaction_terms(z,␣
↪interaction_terms_indices))
    return new_x_
```

[9]:
```python
# Performs 10-cross validation on our dataset to find RMSE

def perform_k_cross_validation(X, y):
    #define cross-validation method to use
    cv = KFold(n_splits=10, random_state=1, shuffle=True)

    #build multiple linear regression model
    model = LinearRegression()

    #use k-fold CV to evaluate model
    scores = cross_val_score(model, X, y, scoring='neg_mean_squared_error',
                             cv=cv, n_jobs=-1)

    #return MSE, std
    return mean(absolute(scores)), std(absolute(scores))
```

[10]:
```python
# Ground Truth

mse_ground_truth_before_covid, std_ground_truth_before_covid = {}, {}
mse_ground_truth, std_ground_truth = {}, {}

def regress_ground_truth(city_name, city_data, best_degree,␣
↪normalize_original_data, poly_model, before_covid=False):
    lin_model_p_vals, poly_model_p_vals = [], []
    if normalize_original_data:
        y = city_data['Scaled_CrimeCount']
        x = city_data.loc[:,['Gender_ratio', 'Scaled_Unemployment',␣
↪'Scaled_MedianIncomeRate']]
    else:
        y = city_data['Cul. Crime count']
        x = city_data.loc[:,['Gender_ratio', 'UnemploymentRate',␣
↪'MedianIncomeRate']]
```

```
    X = sm.add_constant(x)
    lin_model = sm.OLS(y, X).fit()
    for attributeIndex in range (0, len(X.columns)):
        lin_model_p_vals.append(lin_model.pvalues[attributeIndex])
    lin_model_summary = lin_model.summary()
    if before_covid:
        mse_ground_truth_before_covid[city_name], std_ground_truth_before_covid[city_name] = perform_k_cross_validation(X, y)
    else:
        mse_ground_truth[city_name], std_ground_truth[city_name] = perform_k_cross_validation(X, y)
#     linear_model = LinearRegression()
#     linear_model.fit(x, y)
#     r_sq_lm = linear_model.score(x, y)
#     features = ['Intercept'] + ['Gender_ratio', 'Scaled_Unemployment', 'Scaled_MedianIncomeRate']
#     p_val_lm = get_p_val(linear_model, x, y, features)
    if poly_model:
        transformer = PolynomialFeatures(degree=best_degree, include_bias=False)
        x_ = transformer.fit_transform(x)
        x_ = pd.DataFrame(np.array(x_), columns=transformer.get_feature_names_out())
        X = sm.add_constant(x_)
        poly_model = sm.OLS(list(y), X).fit()
        for attributeIndex in range (0, len(X.columns)):
            poly_model_p_vals.append(poly_model.pvalues[attributeIndex])
        poly_model_summary = poly_model.summary()
#     poly_model = LinearRegression().fit(x_, y)
#     r_sq_pm = poly_model.score(x_, y)
#     features = ['Intercept'] + list(transformer.get_feature_names_out())
#     p_val_pm = get_p_val(poly_model, x_, y, features)

    print("Ground Truth:\nLinear Model summary:\n {}".format(lin_model_summary))
#     print("Model Coefficients: \nLinear: {}\n".format(list(linear_model.coef_)))
    print("\n\nPlot for Linear Model Coefficients for Ground Truth:\n")
    plot_hist_for_model_coefficients(lin_model.params, lin_model_p_vals)
#     print("Polynomial: {}\n".format(list(poly_model.coef_)))
    if poly_model:
        print("\n\n\nPolynomial Model (degree={}) summary:\n {}\n".format(best_degree, poly_model_summary))
        print("\n\nPlot for Polynormial Model Coefficients:\n")
        plot_hist_for_model_coefficients(poly_model.params, poly_model_p_vals)
```

[11]:

```python
# Our Model

mse_our_model_before_covid, std_our_model_before_covid = {}, {}
mse_our_model, std_our_model = {}, {}

def regress_our_model(city_name, city_data, best_degree,␣
 ↪normalize_original_data, include_mental_health_data, poly_model,␣
 ↪covid=False, before_covid=False, save_fig=False, time_period=""):
    lin_model_p_vals, poly_model_p_vals = [], []
    if normalize_original_data:
        y = city_data['Scaled_CrimeCount']
        if include_mental_health_data:
            ori_features = ['Gender_ratio', 'Scaled_Unemployment',␣
 ↪'Scaled_MedianIncomeRate', 'WorkforceCount', 'PercentNegativeUsers',␣
 ↪'FavorOfDemocrats', 'Scaled_SearchCountForDepression']
        else:
            ori_features = ['Gender_ratio', 'Scaled_Unemployment',␣
 ↪'Scaled_MedianIncomeRate', 'WorkforceCount', 'PercentNegativeUsers',␣
 ↪'FavorOfDemocrats']
    elif include_mental_health_data:
        y = city_data['Cul. Crime count']
        ori_features = ['Gender_ratio', 'UnemploymentRate', 'MedianIncomeRate',␣
 ↪'WorkforceCount', 'PercentNegativeUsers', 'FavorOfDemocrats',␣
 ↪'SearchCountForDepression']
    else:
        y = city_data['Cul. Crime count']
        ori_features = ['Gender_ratio', 'UnemploymentRate', 'MedianIncomeRate',␣
 ↪'WorkforceCount', 'PercentNegativeUsers', 'FavorOfDemocrats']
    if covid:
        ori_features += ['CovidPercPositive']

    x = city_data.loc[:,ori_features]

    X = sm.add_constant(x)
    lin_model = sm.OLS(y, X).fit()
    for attributeIndex in range (0, len(X.columns)):
        lin_model_p_vals.append(lin_model.pvalues[attributeIndex])
    lin_model_summary = lin_model.summary()
    if not covid:
        if before_covid:
            mse_our_model_before_covid[city_name],␣
 ↪std_our_model_before_covid[city_name] = perform_k_cross_validation(X, y)
        else:
            mse_our_model[city_name], std_our_model[city_name] =␣
 ↪perform_k_cross_validation(X, y)
#     linear_model = LinearRegression()
```

```
#     linear_model.fit(x, y)
#     r_sq_lm = linear_model.score(x, y)
#     features = ['Intercept'] + ori_features
#     p_val_lm = get_p_val(linear_model, x, y, features)

    # poly = PolynomialFeatures(degree = 2)
    # X_poly = poly.fit_transform(x)
    # poly.fit(X_poly, y)
    # poly_model = LinearRegression()
    # poly_model.fit(X_poly, y)
    # r_sq_pm = poly_model.score(X_poly, y)
    if poly_model:
        transformer = PolynomialFeatures(degree=best_degree, include_bias=False)
        x_ = transformer.fit_transform(x)
        z = transformer.get_feature_names_out()
        if city_name == "LA" and not covid:
            x_ = remove_interaction_terms_for_LA(x_, z)
        elif city_name == "NYC" and not covid:
            x_ = remove_interaction_terms_for_NYC(x_, z)
        elif city_name == "CHICAGO" and not covid:
            x_ = remove_interaction_terms_for_CHICAGO(x_, z)
        else:
            x_ = pd.DataFrame(np.array(x_), columns=z)
        X = sm.add_constant(x_)
        poly_model = sm.OLS(list(y), X).fit()
        for attributeIndex in range (0, len(X.columns)):
            poly_model_p_vals.append(poly_model.pvalues[attributeIndex])
        poly_model_summary = poly_model.summary()
#     poly_model = LinearRegression().fit(x_, y)
#     r_sq_pm = poly_model.score(x_, y)
#     features = ['Intercept'] + list(transformer.get_feature_names_out())
#     p_val_pm = get_p_val(poly_model, x_, y, features)

    print("Our Model:\nLinear Model summary:\n {}".format(lin_model_summary))
#     print("Model Coefficients: \nLinear: {}\n".format(list(linear_model.
 ↪coef_)))
    print("\n\nPlot for Linear Model Coefficients for Our Model:\n")
    plot_hist_for_model_coefficients(lin_model.params, lin_model_p_vals)
#     print("Polynomial: {}\n".format(list(poly_model.coef_)))
    if poly_model:
        print("\n\n\nPolynomial Model (degree={}) summary:\n {}\n".
 ↪format(best_degree, poly_model_summary))
        print("\n\nPlot for Polynormial Model Coefficients:\n")
        plot_hist_for_model_coefficients(poly_model.params, poly_model_p_vals)
```

[12]:
```
# Our Model without ground truth columns
```

```
# mse_our_model_before_covid, std_our_model_before_covid = {}, {}
# mse_our_model, std_our_model = {}, {}

def regress_our_model_without_ground_truth(city_name, city_data,␣
↪normalize_original_data, include_mental_health_data, poly_model,␣
↪covid=False, before_covid=False, save_fig=False, time_period=""):
    lin_model_p_vals, poly_model_p_vals = [], []
    if normalize_original_data:
        y = city_data['Scaled_CrimeCount']
        if include_mental_health_data:
            ori_features = ['WorkforceCount', 'PercentNegativeUsers',␣
↪'FavorOfDemocrats', 'Scaled_SearchCountForDepression']
        else:
            ori_features = ['WorkforceCount', 'PercentNegativeUsers',␣
↪'FavorOfDemocrats']
    elif include_mental_health_data:
        y = city_data['Cul. Crime count']
        ori_features = ['WorkforceCount', 'PercentNegativeUsers',␣
↪'FavorOfDemocrats', 'SearchCountForDepression']
    else:
        y = city_data['Cul. Crime count']
        ori_features = ['WorkforceCount', 'PercentNegativeUsers',␣
↪'FavorOfDemocrats']
    if covid:
        ori_features += ['CovidPercPositive']

    x = city_data.loc[:,ori_features]

    X = sm.add_constant(x)
    lin_model = sm.OLS(y, X).fit()
    for attributeIndex in range (0, len(X.columns)):
        lin_model_p_vals.append(lin_model.pvalues[attributeIndex])
    lin_model_summary = lin_model.summary()
#     if not covid:
#         if before_covid:
#             mse_our_model_before_covid[city_name],␣
↪std_our_model_before_covid[city_name] = perform_k_cross_validation(X, y)
#         else:
#             mse_our_model[city_name], std_our_model[city_name] =␣
↪perform_k_cross_validation(X, y)
#     linear_model = LinearRegression()
#     linear_model.fit(x, y)
#     r_sq_lm = linear_model.score(x, y)
#     features = ['Intercept'] + ori_features
#     p_val_lm = get_p_val(linear_model, x, y, features)
```

```
    # poly = PolynomialFeatures(degree = 2)
    # X_poly = poly.fit_transform(x)
    # poly.fit(X_poly, y)
    # poly_model = LinearRegression()
    # poly_model.fit(X_poly, y)
    # r_sq_pm = poly_model.score(X_poly, y)
    if poly_model:
        transformer = PolynomialFeatures(degree=best_degree, include_bias=False)
        x_ = transformer.fit_transform(x)
        z = transformer.get_feature_names_out()
        if city_name == "LA" and not covid:
            x_ = remove_interaction_terms_for_LA(x_, z)
        elif city_name == "NYC" and not covid:
            x_ = remove_interaction_terms_for_NYC(x_, z)
        elif city_name == "CHICAGO" and not covid:
            x_ = remove_interaction_terms_for_CHICAGO(x_, z)
        else:
            x_ = pd.DataFrame(np.array(x_), columns=z)
        X = sm.add_constant(x_)
        poly_model = sm.OLS(list(y), X).fit()
        for attributeIndex in range (0, len(X.columns)):
            poly_model_p_vals.append(poly_model.pvalues[attributeIndex])
        poly_model_summary = poly_model.summary()
#     poly_model = LinearRegression().fit(x_, y)
#     r_sq_pm = poly_model.score(x_, y)
#     features = ['Intercept'] + list(transformer.get_feature_names_out())
#     p_val_pm = get_p_val(poly_model, x_, y, features)

    print("Our Model without ground truth:\nLinear Model summary:\n {}".
 ↪format(lin_model_summary))
#     print("Model Coefficients: \nLinear: {}\n".format(list(linear_model.
 ↪coef_)))
    print("\n\nPlot for Linear Model Coefficients for Our Model without ground␣
 ↪truth:\n")
    plot_hist_for_model_coefficients(lin_model.params, lin_model_p_vals)
#     print("Polynomial: {}\n".format(list(poly_model.coef_)))
    if poly_model:
        print("\n\n\nPolynomial Model (degree={}) summary:\n {}\n".
 ↪format(best_degree, poly_model_summary))
        print("\n\nPlot for Polynormial Model Coefficients:\n")
        plot_hist_for_model_coefficients(poly_model.params, poly_model_p_vals)
```

[13]:
```
# Plot Correlation graphs

def plot_correlation_graphs(city_data, normalize_original_data):
    print("\nCorrelation Graphs:\n")
```

```
    if normalize_original_data:
        unemployment_col = city_data['Scaled_Unemployment']
        median_income_col = city_data['Scaled_MedianIncomeRate']
        y = city_data['Scaled_CrimeCount']
    else:
        unemployment_col = city_data['UnemploymentRate']
        median_income_col = city_data['MedianIncomeRate']
        y = city_data['Cul. Crime count']

    figure(figsize=(15, 10), dpi=80)

    plt.subplot(2, 3, 1)
    plt.scatter(city_data['Gender_ratio'], y)
    plt.title("Gender Ratio v/s Crime Count")

    plt.subplot(2, 3, 2)
    plt.scatter(unemployment_col, y)
    plt.title("Scaled_Unemployment v/s Crime Count")

    plt.subplot(2, 3, 3)
    plt.scatter(median_income_col, y)
    plt.title("Scaled_MedianIncomeRate v/s Crime Count")

    plt.subplot(2, 3, 4)
    plt.scatter(city_data['WorkforceCount'], y)
    plt.title("WorkforceCount v/s Crime Count")

    plt.subplot(2, 3, 5)
    plt.scatter(city_data['PercentNegativeUsers'], y)
    plt.title("PercentNegativeUsers v/s Crime Count")

    plt.subplot(2, 3, 6)
    plt.scatter(city_data['FavorOfDemocrats'], y)
    plt.title("FavorOfDemocrats v/s Crime Count")

    plt.show()
```

[14]:
```
# Append multiple city dataframes into single list

# def build_single_dataset(city_name, city_data, all_cities_df):
#     city_data['City'] = city_name
#     all_cities_df.append(city_data)
```

[15]:
```
# Min-Max Normalization on full data

def normalize_dataset(city_data, include_mental_health_data):
    scaler = MinMaxScaler()
```

```
    city_data[["Scaled_CrimeCount"]] = scaler.fit_transform(city_data[['Cul.␣
↪Crime count']])
    city_data[["Scaled_Unemployment"]] = scaler.
↪fit_transform(city_data[['UnemploymentRate']])
    city_data[["Scaled_MedianIncomeRate"]] = scaler.
↪fit_transform(city_data[['MedianIncomeRate']])
    if include_mental_health_data:
        city_data[["Scaled_SearchCountForDepression"]] = scaler.
↪fit_transform(city_data[['SearchCountForDepression']])
```

[16]:
```
# Covid File Name Structure - <CityNameInCaps>_Covid.csv, No Header, First␣
↪Column as Percent and Second Column as MonthYear

def infer_covid_comparison(city_name, city_data, absolute_file_path,␣
↪best_degree, normalize_original_data, include_mental_health_data,␣
↪poly_model):
    city_covid = pd.read_csv("{}/{}_Covid.csv".format(absolute_file_path,␣
↪city_name), header=None)
    city_covid.columns = ['Perc', 'MonthYear']
    city_covid['MonthYear'] = pd.to_datetime(city_covid['MonthYear'],␣
↪format="%b-%Y")
    city_covid = city_covid.sort_values(by=['MonthYear'])

    city_data['MonthYear'] = pd.to_datetime(city_data['MonthYear'],␣
↪format="%b-%y")

    city_combined_dataset_covid = city_data[city_data['MonthYear'] >=␣
↪'2020-03-01']

    perc = []
    for index, row in city_combined_dataset_covid.iterrows():
        month_year = row['MonthYear']
        perc_pos_covid_cases = city_covid[city_covid['MonthYear'] ==␣
↪month_year]['Perc']
        perc.append(float(perc_pos_covid_cases))

    city_combined_dataset_covid['CovidPercPositive'] = perc

    print("\nCovid Data Comparison:\n\nRegression from Mar-2020 to Dec-2021␣
↪along with Covid Positive Cases Per Month:\n")
    regress_our_model(city_name, city_combined_dataset_covid, best_degree,␣
↪normalize_original_data, include_mental_health_data, poly_model, covid=True)
    print("\nRegression for the same time frame without considering Covid Cases:
↪\n")
    regress_our_model(city_name, city_combined_dataset_covid, best_degree,␣
↪normalize_original_data, include_mental_health_data, poly_model)
```

```
[17]: # File Name Structure - <CityNameInCaps>_GoogleTrends.csv, with original file␣
      ↪format as downloaded from google trends

      def add_mental_health_data(city_name, city_data, absolute_file_path):
          mh_data = pd.read_csv("{}/{}_GoogleTrends.csv".format(absolute_file_path,␣
      ↪city_name), header=1)
          mh_data['Month'] = pd.to_datetime(mh_data['Month'], format="%Y-%m")
          mh_data = mh_data.sort_values(by=['Month'])
          mh_data.columns = ['Month', 'Count']

          city_data['MonthYear'] = pd.to_datetime(city_data['MonthYear'],␣
      ↪format="%b-%y")
          city_data = city_data.sort_values(by=['MonthYear'])
          city_data['SearchCountForDepression'] = list(mh_data['Count'])
      #     city_data.to_csv("{}-withMentalData.csv".format(city_name))
          return city_data
```

```
[18]: # For few cities, MonthYear Column was missing, so created on the fly

      def create_month_year_column(city_data):
          city_data['YearMonth'] = pd.to_datetime(city_data['YearMonth'],␣
      ↪format="%y-%b")
          city_data["MonthYear"] = city_data['YearMonth'].dt.strftime("%b-%y")
          return city_data
```

```
[19]: def fetch_city_data_before_covid(city_data):
          city_data['Time'] = pd.to_datetime(city_data['MonthYear'], format="%b-%y")
          return city_data[city_data['Time'] < '2020-03-01']
```

```
[20]: def plot_time_series_graph(city_name, city_data, save_fig=False):
          plt.figure(figsize=(16, 8), dpi=150)
      #     city_data['WorkforceCount'].plot(label='Workforce%', color='orange')
      #     city_data['Scaled_CrimeCount'].plot(label='Crime%')
      #     plt.title('{}-Workforce% v/s Crime%'.format(city_name))
      #     plt.xlabel('Time Period (Months) - 2011-2021')
      #     plt.legend()
          fig, ax1 = plt.subplots()

          ax2 = ax1.twinx()
          ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',␣
      ↪color="#6495ED")
          ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',␣
      ↪color="#CC4F1B")

          ax1.set_xlabel('Time Period (Months) - 2011-2021')
          ax1.set_ylabel('Unemployment%', color='#6495ED')
```

```
    ax1.set_title('{}-Unemployment% v/s Crime%'.format(city_name))
    ax2.set_ylabel('Crime%', color='#CC4F1B')

#     plt.savefig('{}_TIME_SERIES_Unemployment.png'.format(city_name),␣
↪transparent=True, bbox_inches='tight')
    plt.show()
```

[21]:

```
# Driver Function to read and process data for all cities

def infer_data(city_name, city_file, all_cities_df, absolute_file_path,␣
↪normalize_original_data=True, include_mental_health_data=True,␣
↪poly_model=True):

    city_data = pd.read_csv("{}/{}".format(absolute_file_path, city_file),␣
↪header=0)

    if city_name in ["SEATTLE", "DENVER", "DALLAS", "NEW_ORLEANS", "NYC"]:
        city_data = create_month_year_column(city_data)

    if include_mental_health_data:
        city_data = add_mental_health_data(city_name, city_data,␣
↪absolute_file_path)
    if normalize_original_data:
        normalize_dataset(city_data, include_mental_health_data)

#     build_single_dataset(city_name, city_data, all_cities_df)

#     print("\nMSE for different degree of Polynomials for Ground Truth:\n")
#     best_degree_for_ground_truth =␣
↪fetch_best_degree_regression_model(city_data, normalize_original_data,␣
↪for_our_model=False)
#     print("\nMSE for different degree of Polynomials for Our Model:\n")
#     best_degree_for_our_model = fetch_best_degree_regression_model(city_data,␣
↪normalize_original_data)
    best_degree_for_ground_truth, best_degree_for_our_model = 0, 0

    print("\nBefore Covid (Mar-2020)\n\n")
    city_data_subset = fetch_city_data_before_covid(city_data)
    regress_ground_truth(city_name, city_data_subset,␣
↪best_degree_for_ground_truth, normalize_original_data, poly_model,␣
↪before_covid=True)
    regress_our_model(city_name, city_data_subset, best_degree_for_our_model,␣
↪normalize_original_data, include_mental_health_data, poly_model,␣
↪before_covid=True)
```

```
    regress_our_model_without_ground_truth(city_name, city_data,␣
↪normalize_original_data, include_mental_health_data, poly_model,␣
↪before_covid=True, save_fig=True, time_period="WithoutGroundTruth")

    print("\nTill Dec-2021\n\n")
    regress_ground_truth(city_name, city_data, best_degree_for_ground_truth,␣
↪normalize_original_data, poly_model)
    regress_our_model(city_name, city_data, best_degree_for_our_model,␣
↪normalize_original_data, include_mental_health_data, poly_model)
    regress_our_model_without_ground_truth(city_name, city_data,␣
↪normalize_original_data, include_mental_health_data, poly_model,␣
↪before_covid=True, save_fig=True, time_period="WithoutGroundTruth")

    if city_name in ["LA", "CHICAGO", "NYC"]:
        infer_covid_comparison(city_name, city_data, absolute_file_path,␣
↪best_degree_for_our_model, normalize_original_data,␣
↪include_mental_health_data, poly_model)

    plot_correlation_graphs(city_data, normalize_original_data)

    plot_time_series_graph(city_name, city_data, save_fig=True)
```

[22]:
```
# Create Final Dataset including all cities and run regression

# def infer_final_dataset_all_cities(all_cities_df):

#     all_cities_final_dataset = pd.concat(all_cities_df)
#     all_cities_final_dataset['MonthYear'] = pd.
↪to_datetime(all_cities_final_dataset['MonthYear'], format="%b-%y")
#     all_cities_final_dataset['CrimeCount'] = all_cities_final_dataset['Cul.␣
↪Crime count']
#     all_cities_final_dataset = all_cities_final_dataset.loc[:,['MonthYear',␣
↪'City', 'CrimeCount', 'Gender_ratio', 'UnemploymentRate',␣
↪'MedianIncomeRate', 'WorkforceCount', 'PercentNegativeUsers',␣
↪'FavorOfDemocrats']]
#     all_cities_final_dataset = all_cities_final_dataset.
↪sort_values(by=['MonthYear', 'City'])

#     all_cities_final_dataset.to_csv('final_dataset.csv', index=False,␣
↪header=True)

#     def ols_coef(x,formula):
#         return smf.ols(formula,data=x).fit().params
```

```
#     gamma = (all_cities_final_dataset.groupby('MonthYear').
↪apply(ols_coef,'CrimeCount ~ 1 + Gender_ratio + UnemploymentRate +␣
↪MedianIncomeRate + WorkforceCount + PercentNegativeUsers +␣
↪FavorOfDemocrats'))
#     def fm_summary(p):
#         s = p.describe().T
#         s['std_error'] = s['std']/np.sqrt(s['count'])
#         s['tstat'] = s['mean']/s['std_error']
#         return s[['mean','std_error','tstat']]
#     print(fm_summary(gamma))

#     all_cities_final_dataset = all_cities_final_dataset.set_index(['City',␣
↪'MonthYear'])
#     print(all_cities_final_dataset)

#     model = FamaMacBeth.from_formula('CrimeCount ~ 1 + Gender_ratio +␣
↪UnemploymentRate + MedianIncomeRate + WorkforceCount + PercentNegativeUsers␣
↪+ FavorOfDemocrats', data=all_cities_final_dataset)
#     result = model.fit(cov_type= 'kernel',debiased = False, bandwidth = 3)

#     print(result.summary)
```

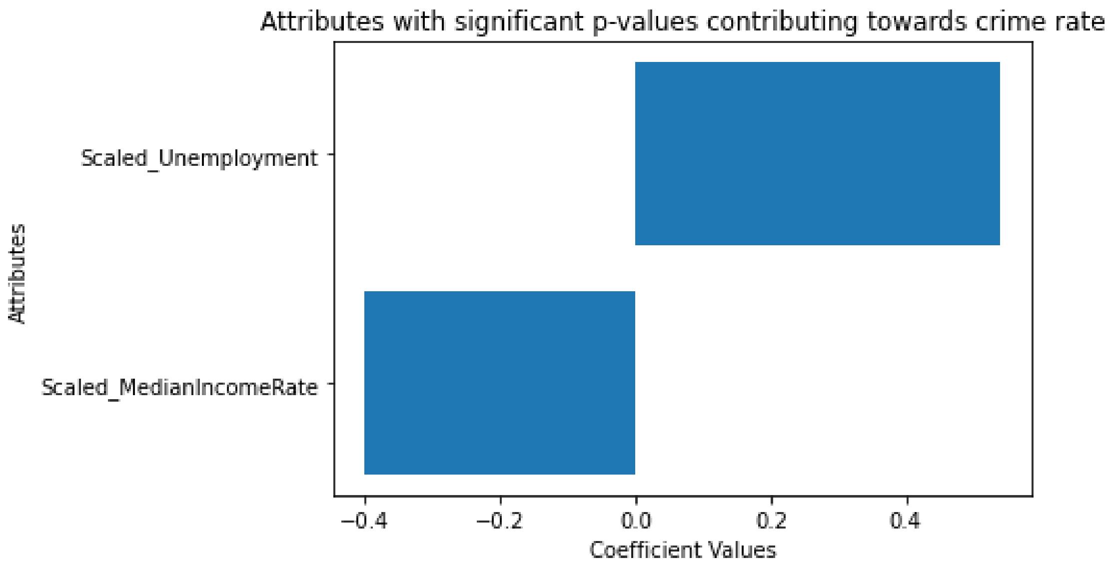
[23]:
```
# Plot Side-by-side bar graph for MSE in both the models

def plot_rmse(mse_our_model, std_our_model, mse_ground_truth, std_ground_truth,␣
↪before_covid=False):
    mse_our_model_values, std_our_model_values, mse_ground_truth_values,␣
↪std_ground_truth_values = [], [], [], []
    for city in cities:
        mse_our_model_values.append(mse_our_model[city])
        std_our_model_values.append(std_our_model[city])
        mse_ground_truth_values.append(mse_ground_truth[city])
        std_ground_truth_values.append(std_ground_truth[city])

    x_indices = np.arange(len(cities))  # the x locations for the groups
    width = 0.35       # the width of the bars

    fig = plt.figure()
    ax = fig.add_subplot(111)
    rects1 = ax.bar(x_indices, mse_ground_truth_values, width,␣
↪color='lightblue', yerr=std_ground_truth_values)
    rects2 = ax.bar(x_indices+width, mse_our_model_values, width,␣
↪color='royalblue', yerr=std_our_model_values)

    # add some
    ax.set_ylabel('Mean Square Error')
```

```
    ax.set_title('MSE comparison of both the models')
    ax.set_xlabel('Cities')
    ax.set_xticks(x_indices + width / 2)
    ax.set_xticklabels(cities, rotation=90)
    ax.legend( (rects1[0], rects2[0]), ('Reference Paper', 'Our Model') )

#     if before_covid:
#         plt.savefig('RMSE_before_covid.png', transparent=True,␣
 ↪bbox_inches='tight')
#     else:
#         plt.savefig('RMSE.png', transparent=True, bbox_inches='tight')

    plt.show()
```

[24]:
```
# Main Initiating Code Block

# cities_files = {'LA': 'CSV_LA_city_final_dataset_v2.csv','CHICAGO':␣
 ↪'chicago_city_final_dataset_v2.csv','NYC': 'NYC_final_dataset_v2.csv'}
cities = ["NYC", "LA", "CHICAGO", "NASHVILLE", "BOSTON", "SEATTLE", "DENVER",␣
 ↪"DALLAS", "NEW_ORLEANS", "INDIANAPOLIS"]
# cities = ["LA"]
all_cities_df = []

for city_name in cities:
    print("\n\n{}:\n".format(city_name))
    city_file = "{}_final_dataset2.csv".format(city_name)
    infer_data(city_name, city_file, all_cities_df, absolute_file_path,␣
 ↪normalize_original_data=True, poly_model=False)

print("\nBefore Covid (Mar-2020)\n\n")
plot_rmse(mse_our_model_before_covid, std_our_model_before_covid,␣
 ↪mse_ground_truth_before_covid, std_ground_truth_before_covid,␣
 ↪before_covid=True)
print("\nTill Dec-2021\n\n")
plot_rmse(mse_our_model, std_our_model, mse_ground_truth, std_ground_truth)

# infer_final_dataset_all_cities(all_cities_df)
```

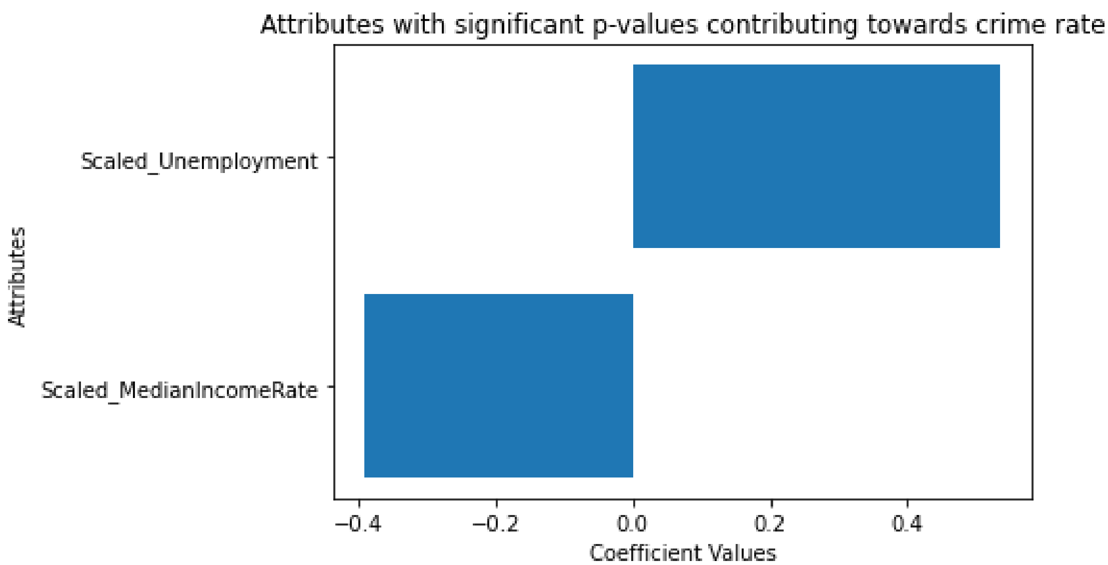
```
NYC:


Before Covid (Mar-2020)


Ground Truth:
```

```
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.834
Model:                            OLS   Adj. R-squared:                  0.829
Method:                 Least Squares   F-statistic:                     177.7
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.32e-41
Time:                        16:45:20   Log-Likelihood:                 119.19
No. Observations:                 110   AIC:                            -230.4
Df Residuals:                     106   BIC:                            -219.6
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                             coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                     27.2471     16.375      1.664      0.099      -5.218
59.712
Gender_ratio             -50.7675     31.349     -1.619      0.108    -112.919
11.384
Scaled_Unemployment        0.5384      0.173      3.109      0.002       0.195
0.882
Scaled_MedianIncomeRate   -0.3993      0.061     -6.541      0.000      -0.520
-0.278
==============================================================================
Omnibus:                       35.741   Durbin-Watson:                   1.219
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               90.769
Skew:                          -1.198   Prob(JB):                     1.95e-20
Kurtosis:                       6.750   Cond. No.                     5.39e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 5.39e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Ground Truth:
```



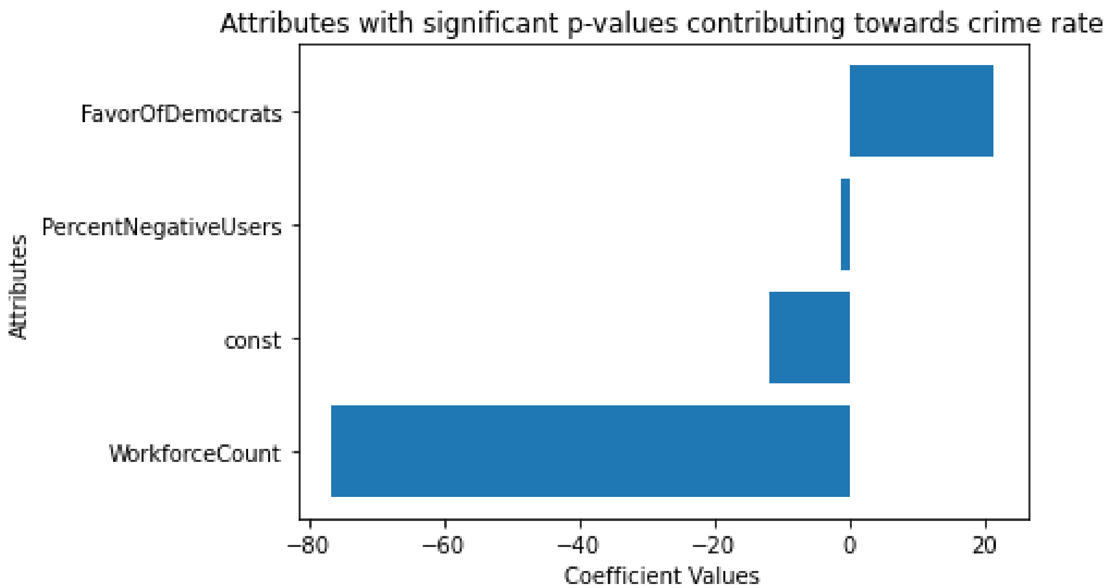

```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.842
Model:                            OLS   Adj. R-squared:                  0.831
Method:                 Least Squares   F-statistic:                     77.80
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.98e-38
Time:                        16:45:22   Log-Likelihood:                 121.94
No. Observations:                 110   AIC:                            -227.9
Df Residuals:                     102   BIC:                            -206.3
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              17.2737     44.739      0.386      0.700
-71.466     106.013
Gender_ratio                      -33.3607     88.219     -0.378      0.706
-208.343     141.621
Scaled_Unemployment                 0.5347      0.217      2.465      0.015
0.104       0.965
Scaled_MedianIncomeRate            -0.3891      0.063     -6.224      0.000
-0.513      -0.265
WorkforceCount                     -7.6674     14.337     -0.535      0.594
```

```
-36.105      20.770
PercentNegativeUsers                   -0.1104      0.135     -0.817      0.416
-0.378       0.158
FavorOfDemocrats                        1.5674      2.028      0.773      0.441
-2.454       5.589
Scaled_SearchCountForDepression         0.0834      0.047      1.778      0.078
-0.010       0.176
==============================================================================
Omnibus:                       32.899   Durbin-Watson:                   1.254
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               75.456
Skew:                          -1.142   Prob(JB):                     4.12e-17
Kurtosis:                       6.353   Cond. No.                     1.94e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.94e+04. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```



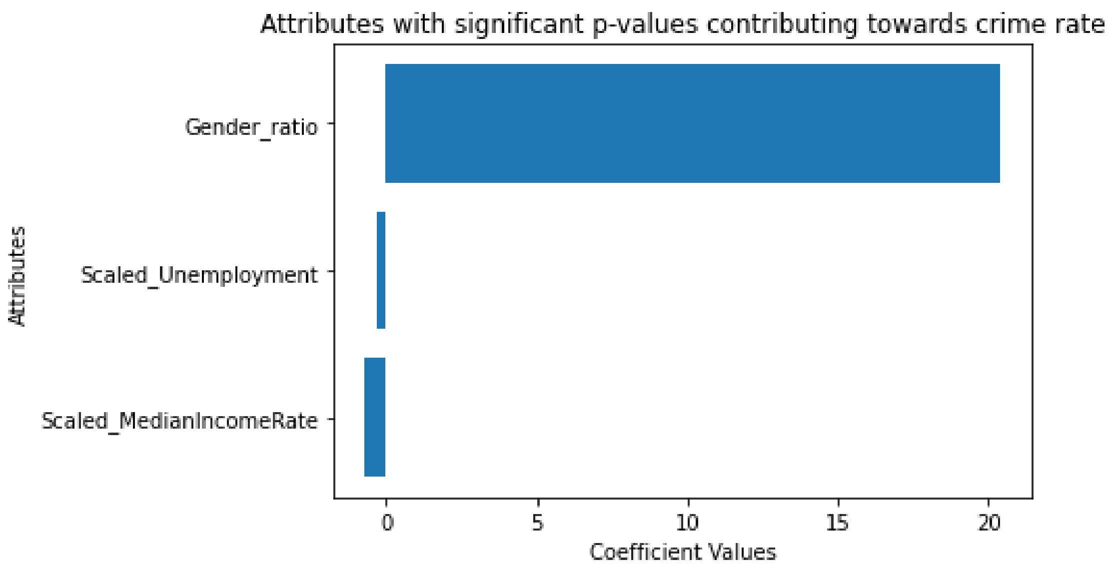

```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.438
```

```
Model:                            OLS   Adj. R-squared:                  0.420
Method:                 Least Squares   F-statistic:                     24.72
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.80e-15
Time:                        16:45:22   Log-Likelihood:                 29.632
No. Observations:                 132   AIC:                            -49.26
Df Residuals:                     127   BIC:                            -34.85
Df Model:                           4
Covariance Type:            nonrobust
==================================================================================
==================
                                      coef    std err          t      P>|t|
[0.025      0.975]
----------------------------------------------------------------------------------
------------------
const                             -11.9249      2.214     -5.386      0.000
-16.306      -7.543
WorkforceCount                    -76.7194      9.522     -8.057      0.000
-95.561     -57.878
PercentNegativeUsers               -1.2691      0.296     -4.293      0.000
-1.854      -0.684
FavorOfDemocrats                   21.5093      3.085      6.972      0.000
15.404      27.614
Scaled_SearchCountForDepression    -0.1113      0.089     -1.251      0.213
-0.287       0.065
==============================================================================
Omnibus:                       18.806   Durbin-Watson:                   0.460
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               25.765
Skew:                           0.770   Prob(JB):                     2.54e-06
Kurtosis:                       4.521   Cond. No.                         764.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Our Model without ground truth:
```



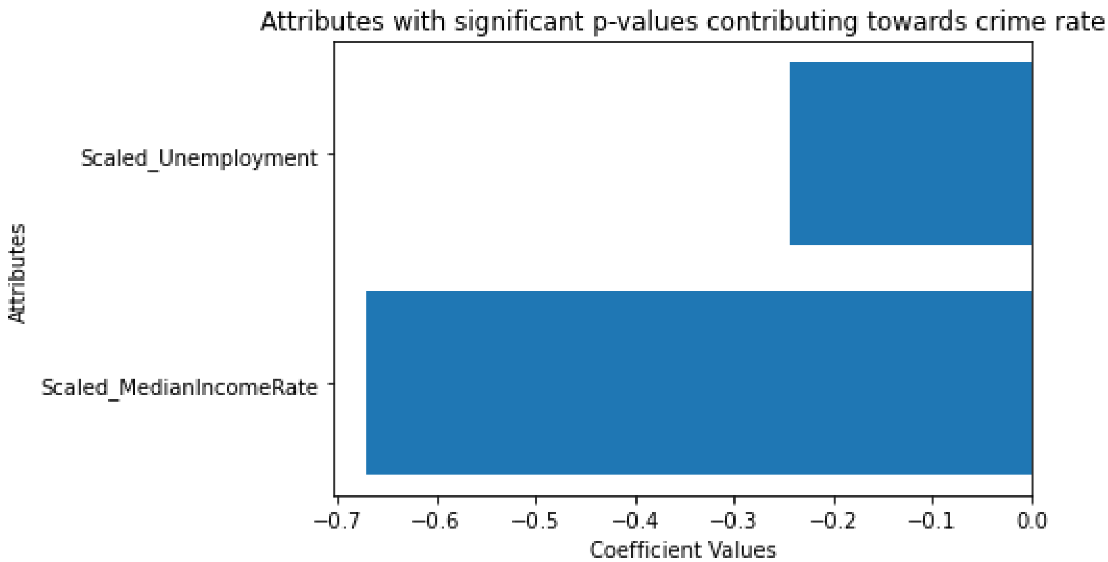

```
Till Dec-2021


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.885
Model:                            OLS   Adj. R-squared:                  0.882
Method:                 Least Squares   F-statistic:                     327.8
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           7.19e-60
Time:                        16:45:22   Log-Likelihood:                 134.28
No. Observations:                 132   AIC:                            -260.6
Df Residuals:                     128   BIC:                            -249.0
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                             coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                     -9.7313      5.039     -1.931      0.056     -19.702
0.240
Gender_ratio              20.3820      9.620      2.119      0.036       1.347
39.417
Scaled_Unemployment       -0.2677      0.043     -6.162      0.000      -0.354
```

```
-0.182
Scaled_MedianIncomeRate    -0.6854      0.028    -24.075      0.000      -0.742
-0.629
==============================================================================
Omnibus:                       31.872   Durbin-Watson:                   1.179
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               74.247
Skew:                          -0.964   Prob(JB):                     7.54e-17
Kurtosis:                       6.128   Cond. No.                     1.76e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.76e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.890
Model:                            OLS   Adj. R-squared:                  0.884
Method:                 Least Squares   F-statistic:                     143.6
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.41e-56
Time:                        16:45:22   Log-Likelihood:                 137.43
```

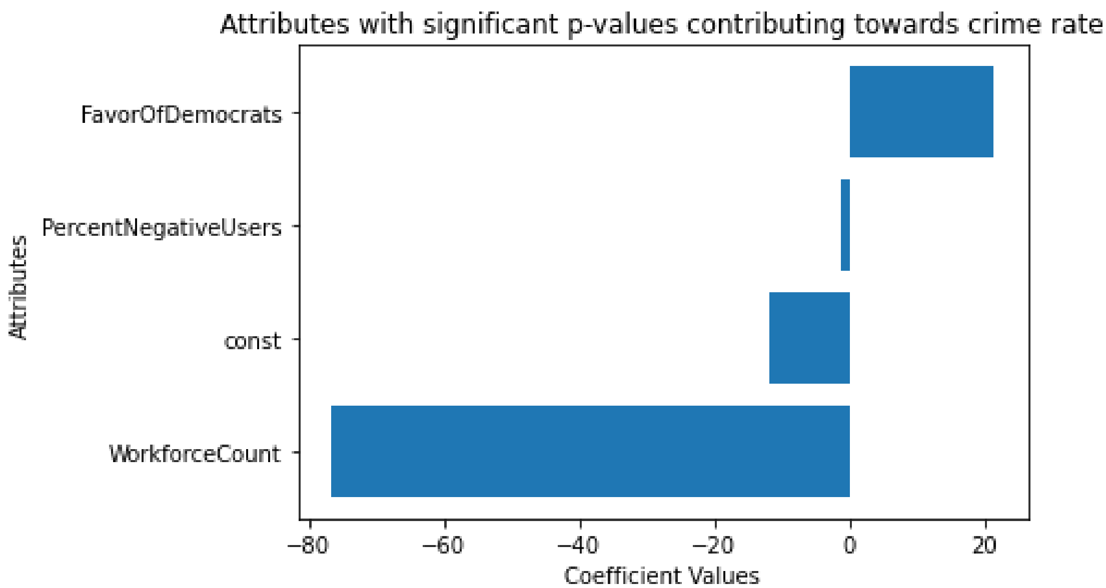
```
No. Observations:                 132   AIC:                            -258.9
Df Residuals:                     124   BIC:                            -235.8
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                   coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                           -5.9746      5.525     -1.081      0.282
-16.911       4.961
Gender_ratio                    11.5131     10.661      1.080      0.282
-9.589      32.615
Scaled_Unemployment             -0.2438      0.048     -5.101      0.000
-0.338      -0.149
Scaled_MedianIncomeRate         -0.6702      0.033    -20.346      0.000
-0.735      -0.605
WorkforceCount                  -7.3509      5.573     -1.319      0.190
-18.381       3.679
PercentNegativeUsers            -0.2464      0.140     -1.760      0.081
-0.524       0.031
FavorOfDemocrats                 1.6608      1.709      0.972      0.333
-1.722       5.044
Scaled_SearchCountForDepression  0.0531      0.042      1.258      0.211
-0.031       0.137
==============================================================================
Omnibus:                       31.465   Durbin-Watson:                   1.200
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               75.290
Skew:                          -0.940   Prob(JB):                     4.48e-17
Kurtosis:                       6.187   Cond. No.                     2.47e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.47e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```




```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.438
Model:                            OLS   Adj. R-squared:                  0.420
Method:                 Least Squares   F-statistic:                     24.72
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.80e-15
Time:                        16:45:22   Log-Likelihood:                 29.632
No. Observations:                 132   AIC:                            -49.26
Df Residuals:                     127   BIC:                            -34.85
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                     coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                            -11.9249      2.214     -5.386      0.000
-16.306      -7.543
WorkforceCount                   -76.7194      9.522     -8.057      0.000
-95.561     -57.878
PercentNegativeUsers              -1.2691      0.296     -4.293      0.000
-1.854      -0.684
FavorOfDemocrats                  21.5093      3.085      6.972      0.000
15.404      27.614
Scaled_SearchCountForDepression   -0.1113      0.089     -1.251      0.213
```

```
-0.287      0.065
==============================================================================
Omnibus:                       18.806   Durbin-Watson:                   0.460
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               25.765
Skew:                           0.770   Prob(JB):                     2.54e-06
Kurtosis:                       4.521   Cond. No.                         764.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:



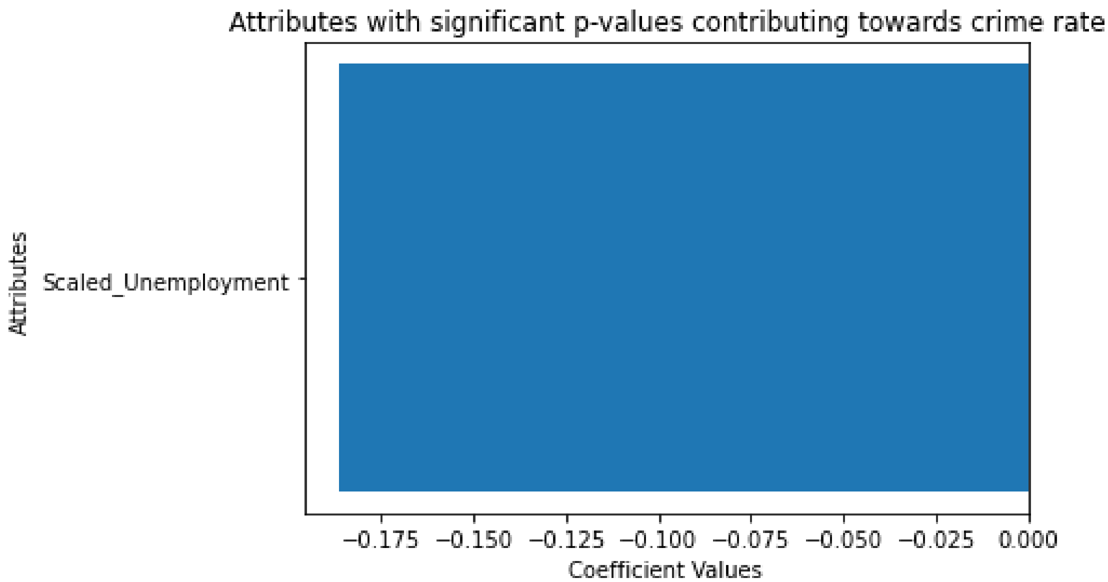
Covid Data Comparison:

Regression from Mar-2020 to Dec-2021 along with Covid Positive Cases Per Month:

```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.568
Model:                            OLS   Adj. R-squared:                  0.432
Method:                 Least Squares   F-statistic:                     4.200
```

```
Date:                Sun, 11 Dec 2022   Prob (F-statistic):            0.0125
Time:                        16:45:23   Log-Likelihood:                39.654
No. Observations:                  22   AIC:                           -67.31
Df Residuals:                      16   BIC:                           -60.76
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
===================
                                    coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
Gender_ratio                      1.0775      0.748      1.440      0.169
-0.508       2.663
Scaled_Unemployment              -0.1864      0.063     -2.954      0.009
-0.320      -0.053
Scaled_MedianIncomeRate          -0.7378      0.819     -0.901      0.381
-2.475       0.999
WorkforceCount                   -0.0758      0.071     -1.067      0.302
-0.226       0.075
PercentNegativeUsers             -1.5430      0.834     -1.850      0.083
-3.311       0.225
FavorOfDemocrats                  1.3225      0.895      1.478      0.159
-0.575       3.220
Scaled_SearchCountForDepression   0.1518      0.084      1.802      0.090
-0.027       0.330
CovidPercPositive                -0.2757      0.148     -1.859      0.082
-0.590       0.039
==============================================================================
Omnibus:                        0.969   Durbin-Watson:                   2.645
Prob(Omnibus):                  0.616   Jarque-Bera (JB):                0.409
Skew:                           0.334   Prob(JB):                        0.815
Kurtosis:                       3.027   Cond. No.                     5.73e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 1.41e-34. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.


Plot for Linear Model Coefficients for Our Model:


/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1807615713.py:1
9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
```

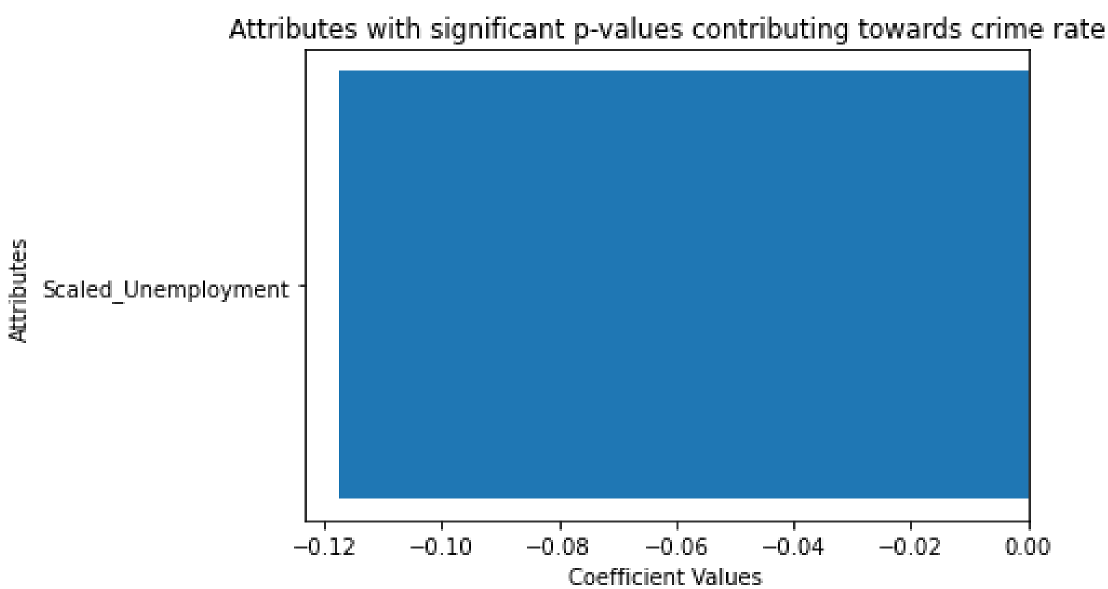
```
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  city_combined_dataset_covid['CovidPercPositive'] = perc
```



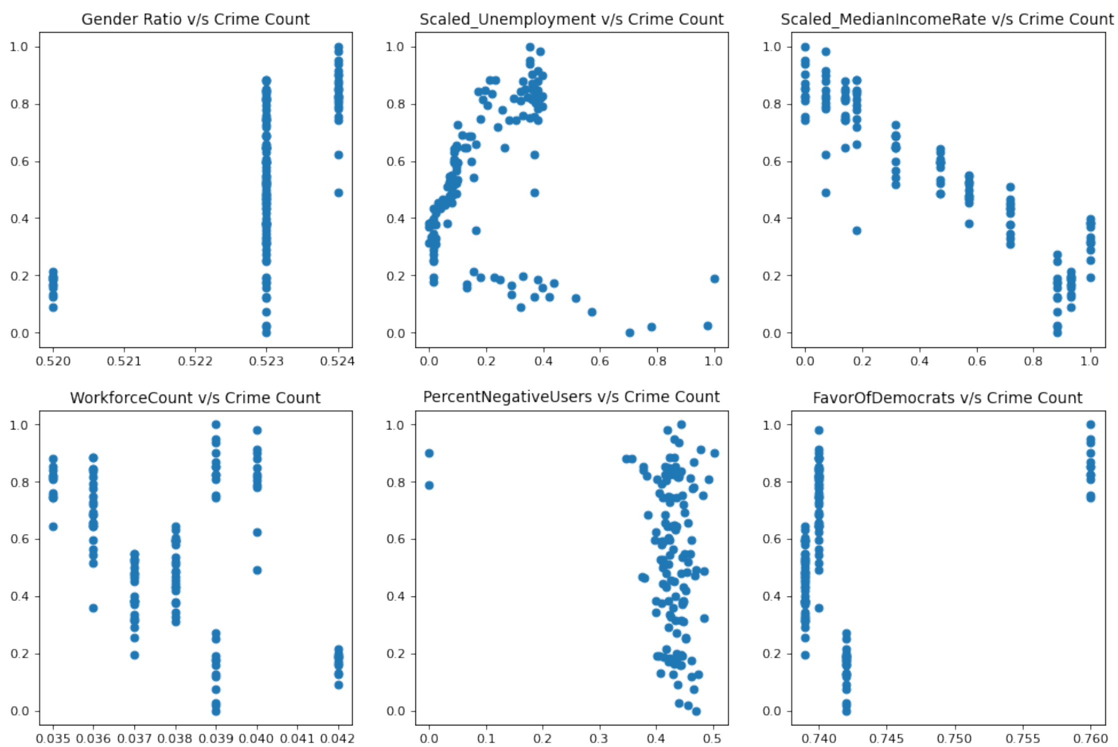

```
Regression for the same time frame without considering Covid Cases:

Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.474
Model:                            OLS   Adj. R-squared:                  0.350
Method:                 Least Squares   F-statistic:                     3.833
Date:                Sun, 11 Dec 2022   Prob (F-statistic):             0.0213
Time:                        16:45:23   Log-Likelihood:                 37.503
No. Observations:                  22   AIC:                            -65.01
Df Residuals:                      17   BIC:                            -59.55
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
```

```
Gender_ratio                        0.5494      0.740      0.742      0.468
-1.012       2.111
Scaled_Unemployment                -0.1175      0.055     -2.151      0.046
-0.233      -0.002
Scaled_MedianIncomeRate            -0.1462      0.808     -0.181      0.858
-1.850       1.558
WorkforceCount                     -0.0246      0.070     -0.351      0.730
-0.172       0.123
PercentNegativeUsers               -1.1112      0.857     -1.297      0.212
-2.919       0.696
FavorOfDemocrats                    0.6929      0.886      0.782      0.445
-1.177       2.563
Scaled_SearchCountForDepression     0.0247      0.053      0.469      0.645
-0.086       0.136
==============================================================================
Omnibus:                        2.053   Durbin-Watson:                   2.341
Prob(Omnibus):                  0.358   Jarque-Bera (JB):                0.899
Skew:                           0.465   Prob(JB):                        0.638
Kurtosis:                       3.342   Cond. No.                     2.97e+18
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 5.23e-36. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Plot for Linear Model Coefficients for Our Model:

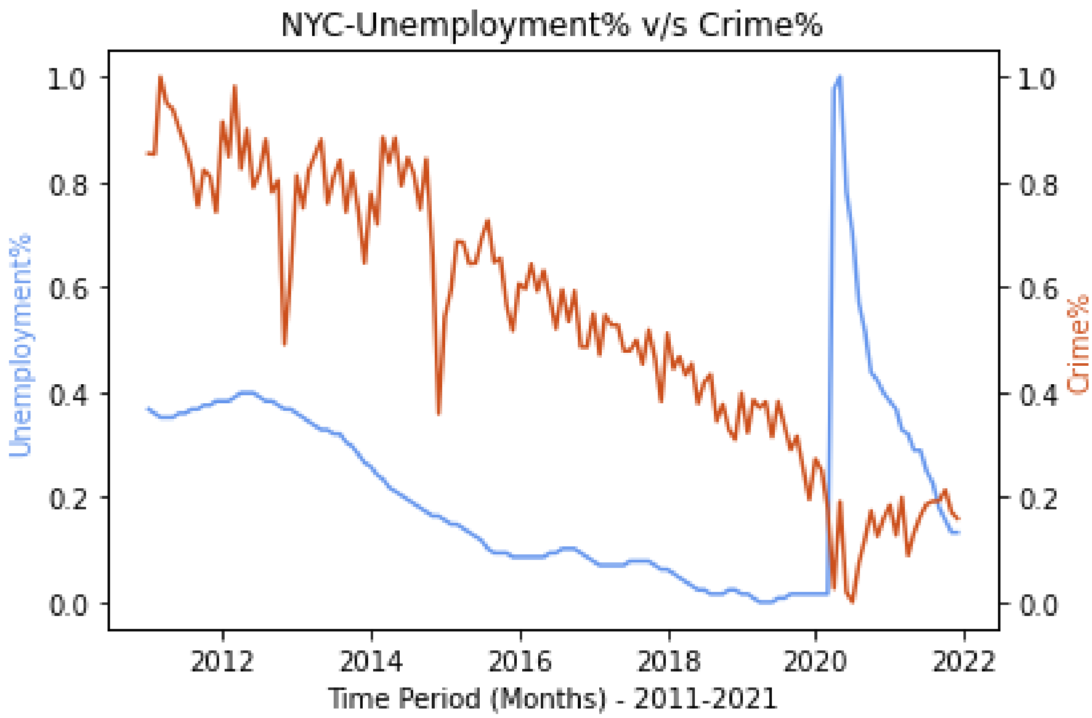


Correlation Graphs:



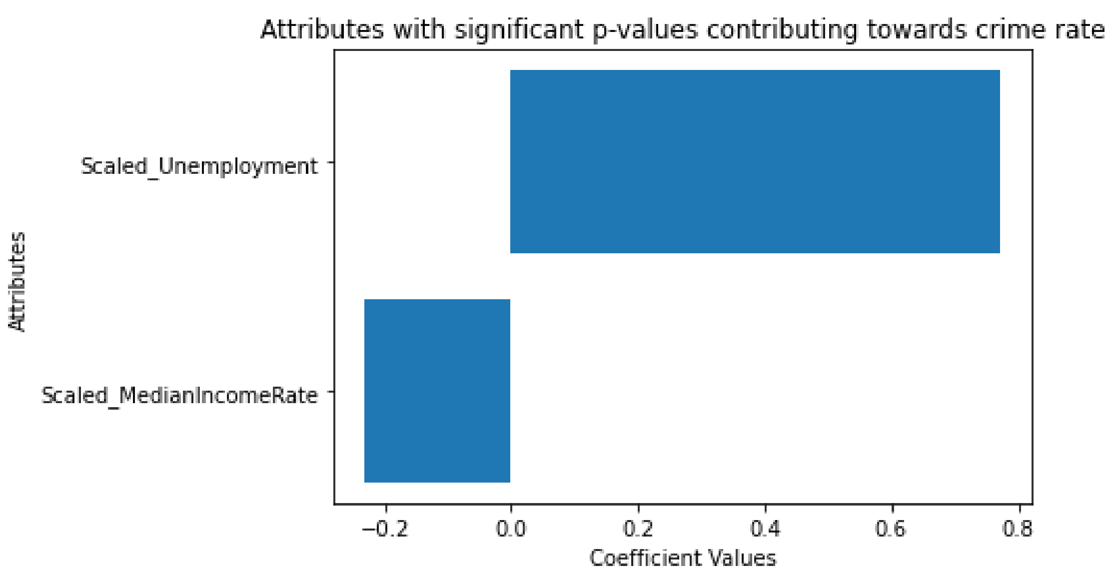

```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

```
<Figure size 2400x1200 with 0 Axes>
```



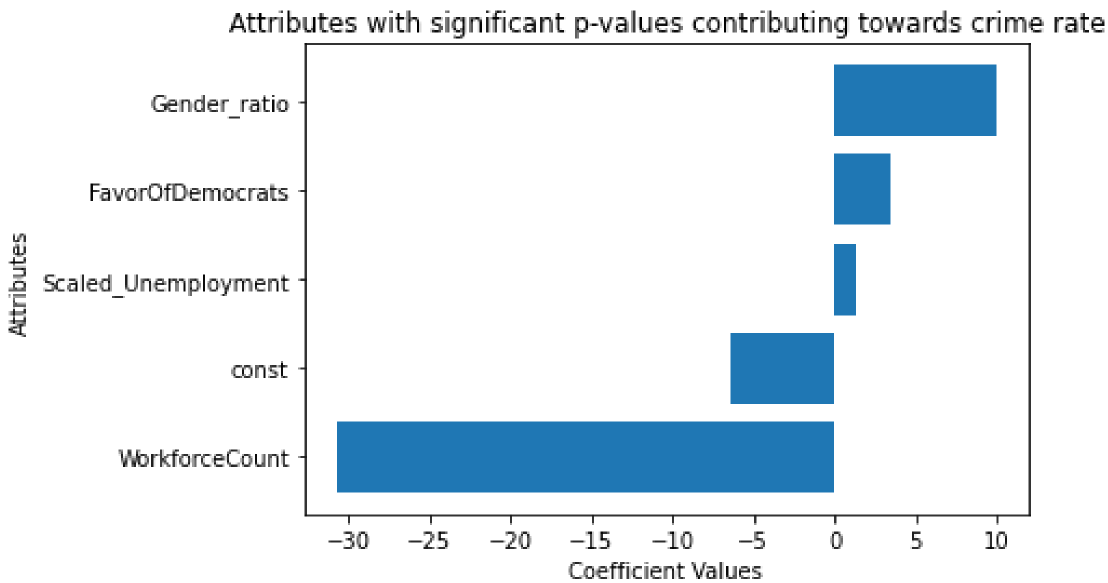

```
LA:


Before Covid (Mar-2020)


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.864
Model:                            OLS   Adj. R-squared:                  0.860
Method:                 Least Squares   F-statistic:                     224.9
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           8.48e-46
Time:                        16:45:23   Log-Likelihood:                 119.41
No. Observations:                 110   AIC:                            -230.8
Df Residuals:                     106   BIC:                            -220.0
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                             coef    std err          t      P>|t|      [0.025
```

```
0.975]
--------------------------------------------------------------------------------
-----------
const                   -0.2077      2.189     -0.095      0.925      -4.547
4.132
Gender_ratio             1.3841      4.296      0.322      0.748      -7.134
9.902
Scaled_Unemployment      0.7707      0.108      7.155      0.000       0.557
0.984
Scaled_MedianIncomeRate -0.2302      0.065     -3.562      0.001      -0.358
-0.102
==============================================================================
Omnibus:                        0.714   Durbin-Watson:                   1.177
Prob(Omnibus):                  0.700   Jarque-Bera (JB):                0.312
Skew:                           0.049   Prob(JB):                        0.856
Kurtosis:                       3.242   Cond. No.                         725.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
                            OLS Regression Results
```

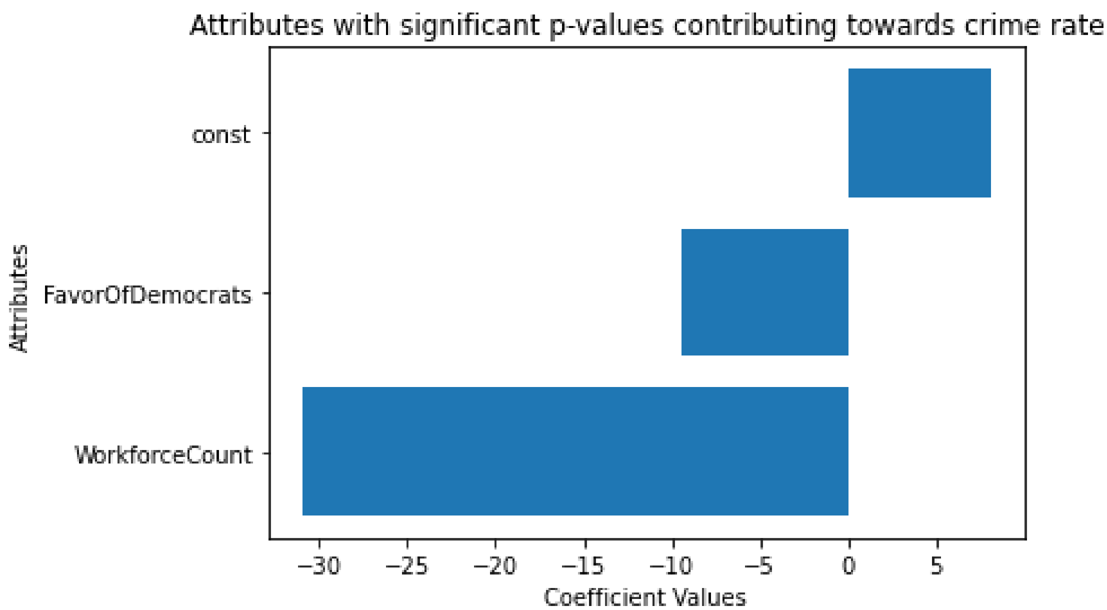
```
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.890
Model:                            OLS   Adj. R-squared:                  0.883
Method:                 Least Squares   F-statistic:                     118.5
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.75e-46
Time:                        16:45:23   Log-Likelihood:                 131.23
No. Observations:                 110   AIC:                            -246.5
Df Residuals:                     102   BIC:                            -224.8
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              -6.4065      2.537     -2.525      0.013
-11.438      -1.375
Gender_ratio                        9.9219      4.337      2.287      0.024
1.319      18.525
Scaled_Unemployment                 1.2847      0.145      8.869      0.000
0.997       1.572
Scaled_MedianIncomeRate            -0.1025      0.073     -1.409      0.162
-0.247       0.042
WorkforceCount                    -30.6813      6.485     -4.731      0.000
-43.544     -17.818
PercentNegativeUsers                0.0555      0.063      0.880      0.381
-0.070       0.181
FavorOfDemocrats                    3.3964      1.121      3.029      0.003
1.173       5.620
Scaled_SearchCountForDepression     0.0823      0.045      1.817      0.072
-0.008       0.172
==============================================================================
Omnibus:                        3.685   Durbin-Watson:                   1.202
Prob(Omnibus):                  0.158   Jarque-Bera (JB):                3.051
Skew:                          -0.349   Prob(JB):                        0.217
Kurtosis:                       3.422   Cond. No.                     1.48e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.48e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```




```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.575
Model:                            OLS   Adj. R-squared:                  0.562
Method:                 Least Squares   F-statistic:                     43.04
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           8.83e-23
Time:                        16:45:24   Log-Likelihood:                 41.755
No. Observations:                 132   AIC:                            -73.51
Df Residuals:                     127   BIC:                            -59.10
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               8.0421      0.954      8.429      0.000
6.154       9.930
WorkforceCount                    -30.9226      9.846     -3.140      0.002
-50.407     -11.438
PercentNegativeUsers               -0.2529      0.145     -1.750      0.083
-0.539       0.033
FavorOfDemocrats                   -9.4381      1.518     -6.216      0.000
-12.443      -6.433
Scaled_SearchCountForDepression    -0.1676      0.086     -1.939      0.055
```

```
-0.339       0.003
==============================================================================
Omnibus:                        5.613   Durbin-Watson:                   0.417
Prob(Omnibus):                  0.060   Jarque-Bera (JB):                2.849
Skew:                          -0.048   Prob(JB):                        0.241
Kurtosis:                       2.287   Cond. No.                         873.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:

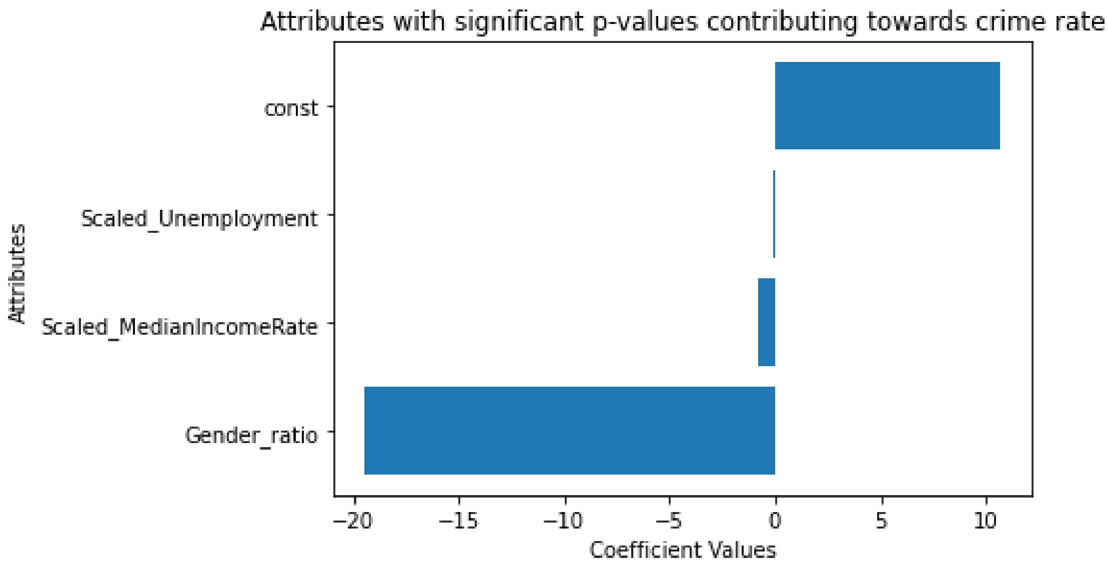


Till Dec-2021

```
Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.857
Model:                            OLS   Adj. R-squared:                  0.854
Method:                 Least Squares   F-statistic:                     256.3
```

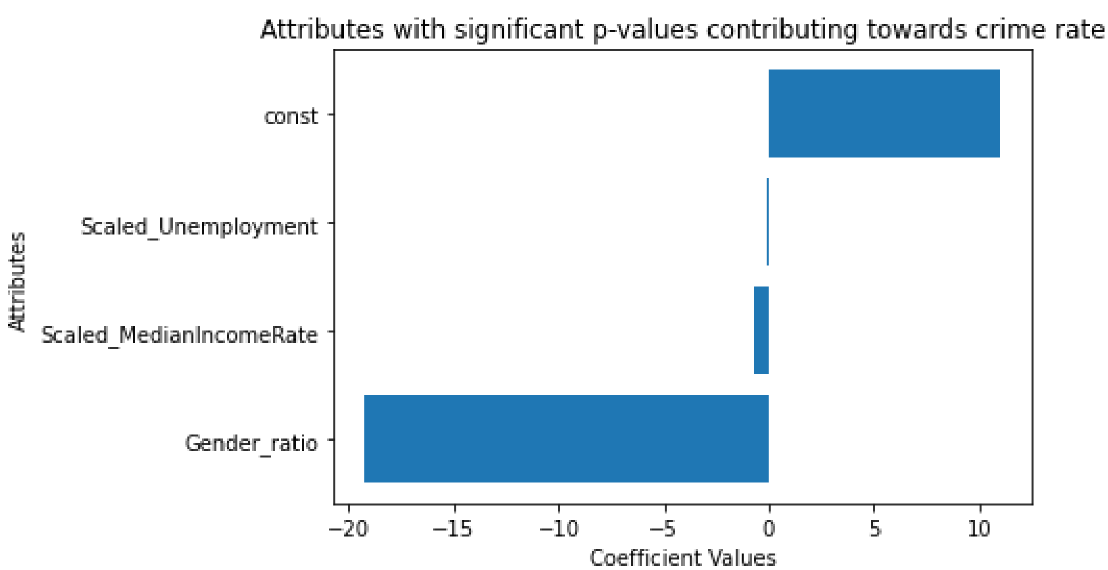
```
Date:                Sun, 11 Dec 2022   Prob (F-statistic):          6.51e-54
Time:                        16:45:24   Log-Likelihood:                113.70
No. Observations:                 132   AIC:                           -219.4
Df Residuals:                     128   BIC:                           -207.9
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                           coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                   10.6498      2.114      5.037      0.000       6.466
14.834
Gender_ratio           -19.4411      4.189     -4.641      0.000     -27.730
-11.153
Scaled_Unemployment     -0.1041      0.044     -2.376      0.019      -0.191
-0.017
Scaled_MedianIncomeRate  -0.7496     0.028    -26.773      0.000      -0.805
-0.694
==============================================================================
Omnibus:                        0.100   Durbin-Watson:                   0.969
Prob(Omnibus):                  0.951   Jarque-Bera (JB):                0.239
Skew:                           0.046   Prob(JB):                        0.887
Kurtosis:                       2.813   Cond. No.                         642.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Ground Truth:




```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.858
Model:                            OLS   Adj. R-squared:                  0.850
Method:                 Least Squares   F-statistic:                     107.1
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           1.78e-49
Time:                        16:45:24   Log-Likelihood:                 114.08
No. Observations:                 132   AIC:                            -212.2
Df Residuals:                     124   BIC:                            -189.1
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                  coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                          10.9824      2.320      4.733      0.000
6.390      15.575
Gender_ratio                  -19.1958      4.336     -4.428      0.000
-27.777     -10.614
Scaled_Unemployment            -0.1280      0.056     -2.282      0.024
-0.239      -0.017
Scaled_MedianIncomeRate        -0.7417      0.048    -15.383      0.000
-0.837      -0.646
WorkforceCount                  4.4923      6.928      0.648      0.518
```

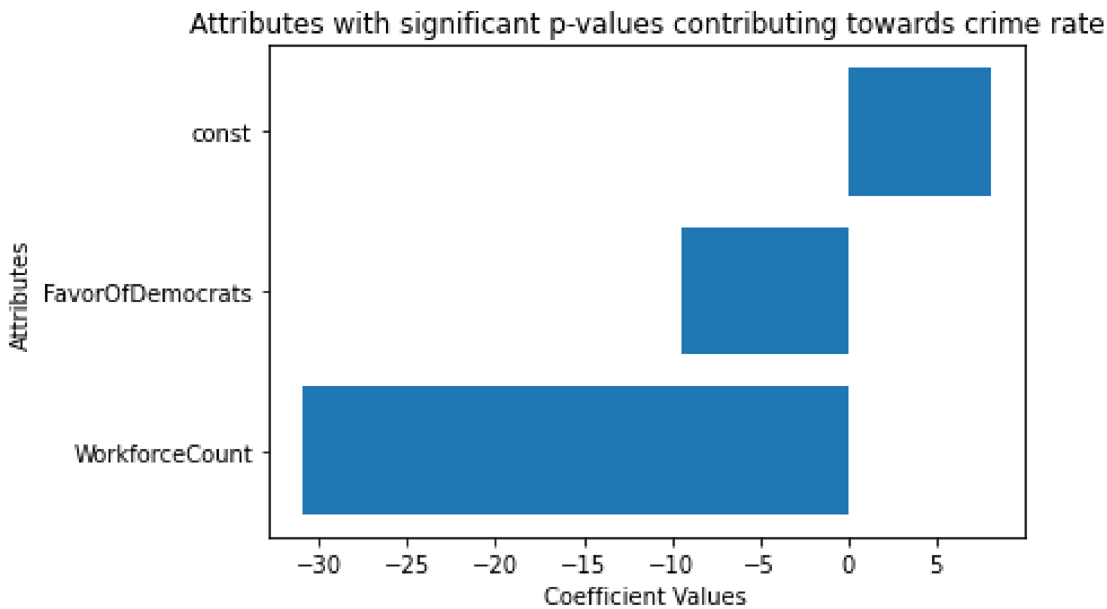
```
-9.221      18.205
PercentNegativeUsers                -0.0257      0.086     -0.299      0.765
-0.195       0.144
FavorOfDemocrats                    -0.7881      1.234     -0.638      0.524
-3.231       1.655
Scaled_SearchCountForDepression      0.0128      0.053      0.242      0.809
-0.092       0.117
==============================================================================
Omnibus:                        0.103   Durbin-Watson:                   0.980
Prob(Omnibus):                  0.950   Jarque-Bera (JB):                0.173
Skew:                           0.065   Prob(JB):                        0.917
Kurtosis:                       2.879   Cond. No.                     1.19e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.19e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```




```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.575
```

```
Model:                            OLS   Adj. R-squared:                  0.562
Method:                 Least Squares   F-statistic:                     43.04
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           8.83e-23
Time:                        16:45:24   Log-Likelihood:                 41.755
No. Observations:                 132   AIC:                            -73.51
Df Residuals:                     127   BIC:                            -59.10
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               8.0421      0.954      8.429      0.000
6.154       9.930
WorkforceCount                    -30.9226      9.846     -3.140      0.002
-50.407     -11.438
PercentNegativeUsers               -0.2529      0.145     -1.750      0.083
-0.539       0.033
FavorOfDemocrats                   -9.4381      1.518     -6.216      0.000
-12.443      -6.433
Scaled_SearchCountForDepression    -0.1676      0.086     -1.939      0.055
-0.339       0.003
==============================================================================
Omnibus:                        5.613   Durbin-Watson:                   0.417
Prob(Omnibus):                  0.060   Jarque-Bera (JB):                2.849
Skew:                          -0.048   Prob(JB):                        0.241
Kurtosis:                       2.287   Cond. No.                         873.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:



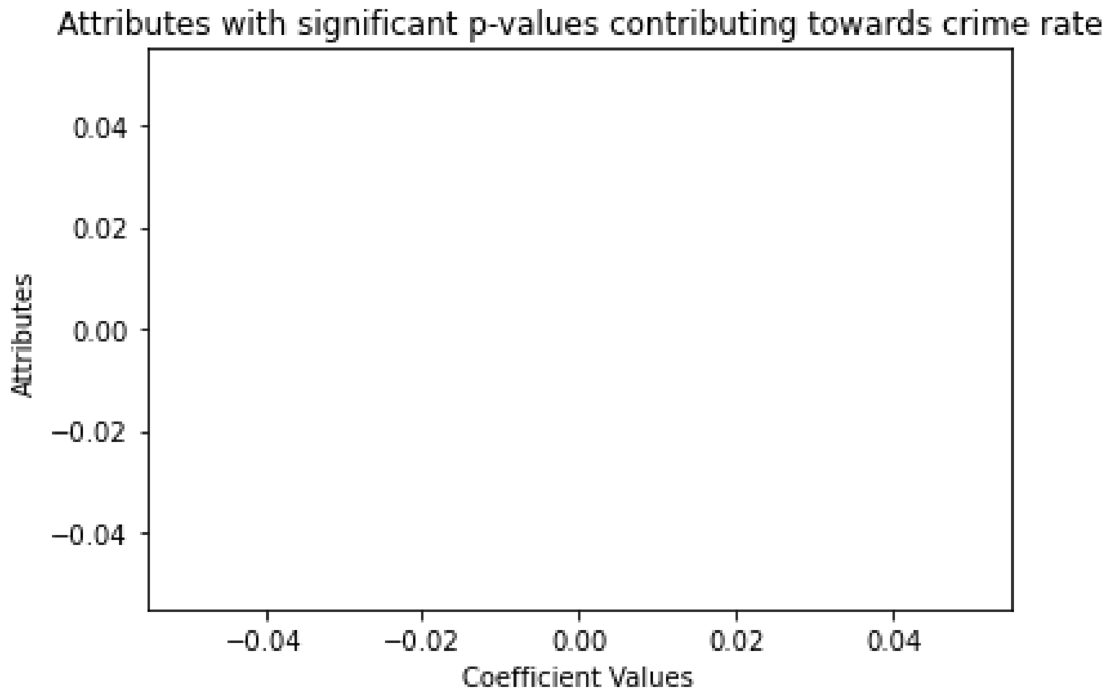
Covid Data Comparison:

Regression from Mar-2020 to Dec-2021 along with Covid Positive Cases Per Month:

Our Model:
Linear Model summary:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.256
Model:                            OLS   Adj. R-squared:                  0.024
Method:                 Least Squares   F-statistic:                     1.101
Date:                Sun, 11 Dec 2022   Prob (F-statistic):              0.398
Time:                        16:45:24   Log-Likelihood:                 33.988
No. Observations:                  22   AIC:                            -55.98
Df Residuals:                      16   BIC:                            -49.43
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
Gender_ratio                        0.0420      0.291      0.145      0.887
-0.574       0.658
Scaled_Unemployment                 0.0262      0.066      0.395      0.698
```

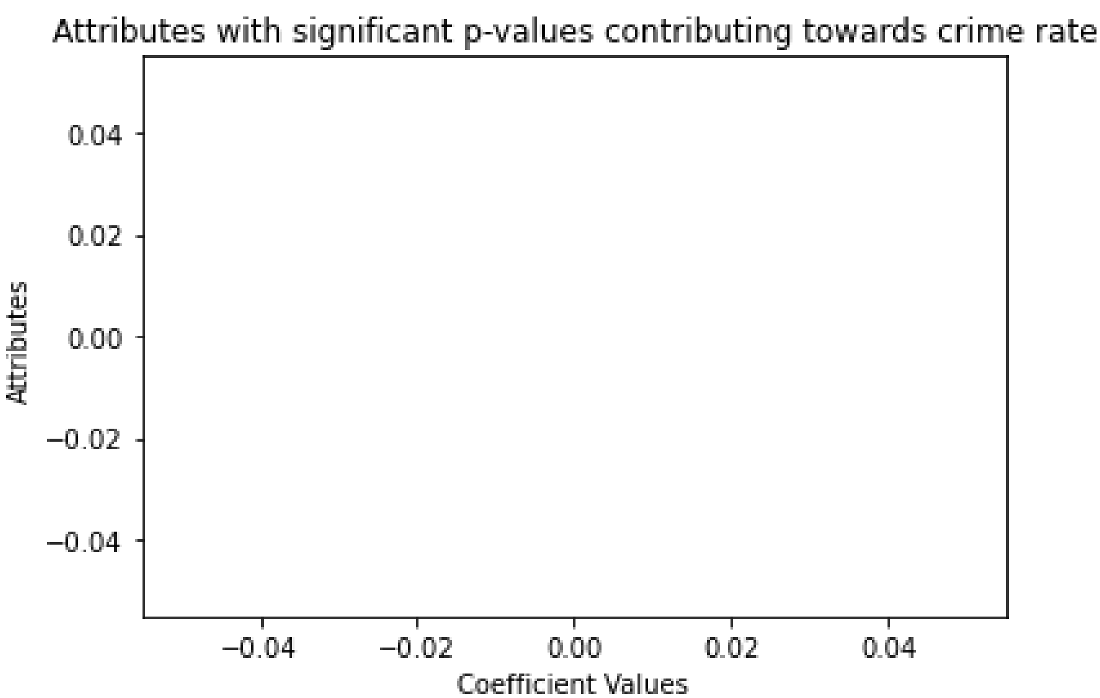
```
-0.114       0.167
Scaled_MedianIncomeRate             0.1545      0.211      0.731      0.475
-0.293       0.602
WorkforceCount                      0.0026      0.011      0.234      0.818
-0.021       0.026
PercentNegativeUsers               -0.2587      0.639     -0.405      0.691
-1.612       1.095
FavorOfDemocrats                    0.0614      0.393      0.156      0.878
-0.771       0.894
Scaled_SearchCountForDepression     0.0706      0.086      0.817      0.426
-0.113       0.254
CovidPercPositive                  -0.4031      0.351     -1.147      0.268
-1.148       0.342
==============================================================================
Omnibus:                        0.621   Durbin-Watson:                   1.761
Prob(Omnibus):                  0.733   Jarque-Bera (JB):                0.638
Skew:                          -0.093   Prob(JB):                        0.727
Kurtosis:                       2.187   Cond. No.                     2.90e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 6.03e-34. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.


Plot for Linear Model Coefficients for Our Model:


/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1807615713.py:1
9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  city_combined_dataset_covid['CovidPercPositive'] = perc
```



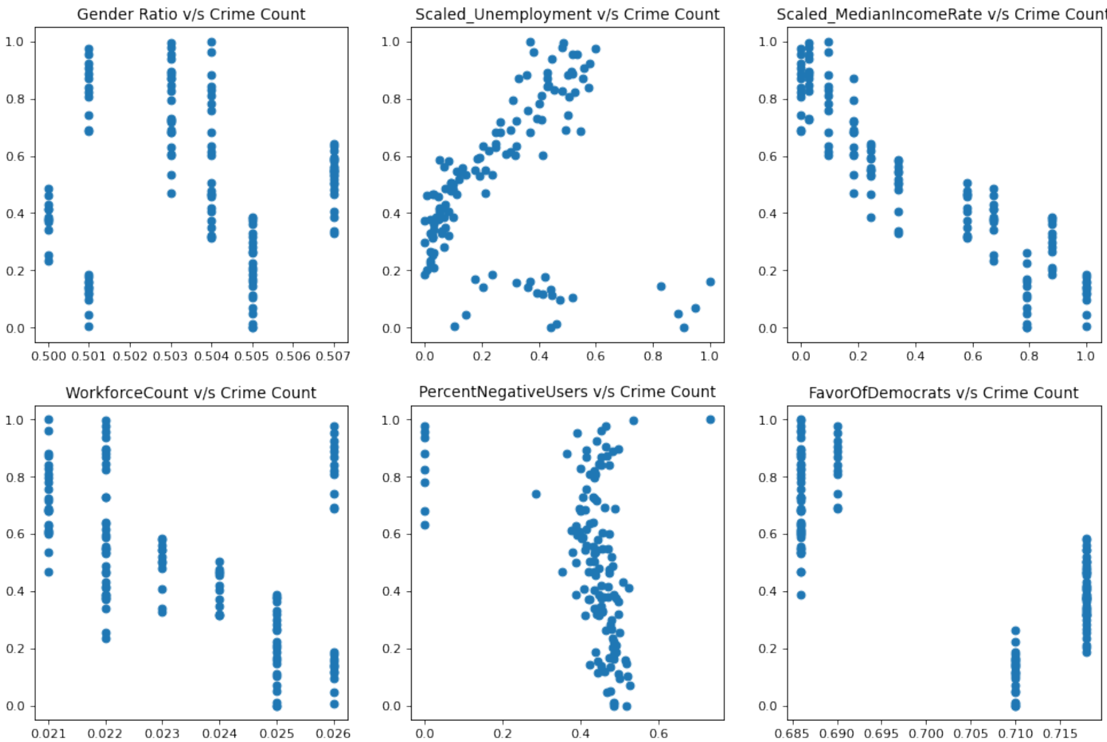

```
Regression for the same time frame without considering Covid Cases:

Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.195
Model:                            OLS   Adj. R-squared:                  0.005
Method:                 Least Squares   F-statistic:                     1.028
Date:                Sun, 11 Dec 2022   Prob (F-statistic):              0.421
Time:                        16:45:24   Log-Likelihood:                 33.118
No. Observations:                  22   AIC:                            -56.24
Df Residuals:                      17   BIC:                            -50.78
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
Gender_ratio                       -0.0043      0.291     -0.015      0.988
-0.617       0.609
Scaled_Unemployment                 0.0309      0.067      0.462      0.650
```

```
-0.110       0.172
Scaled_MedianIncomeRate            0.2549      0.194      1.314      0.206
-0.154       0.664
WorkforceCount                     0.0012      0.011      0.112      0.912
-0.022       0.024
PercentNegativeUsers              -0.3802      0.636     -0.598      0.558
-1.721       0.961
FavorOfDemocrats                   0.0008      0.393      0.002      0.998
-0.828       0.829
Scaled_SearchCountForDepression    0.0879      0.086      1.023      0.320
-0.093       0.269
==============================================================================
Omnibus:                        2.374   Durbin-Watson:                   1.756
Prob(Omnibus):                  0.305   Jarque-Bera (JB):                1.193
Skew:                          -0.151   Prob(JB):                        0.551
Kurtosis:                       1.900   Cond. No.                     3.81e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 3.48e-34. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Plot for Linear Model Coefficients for Our Model:

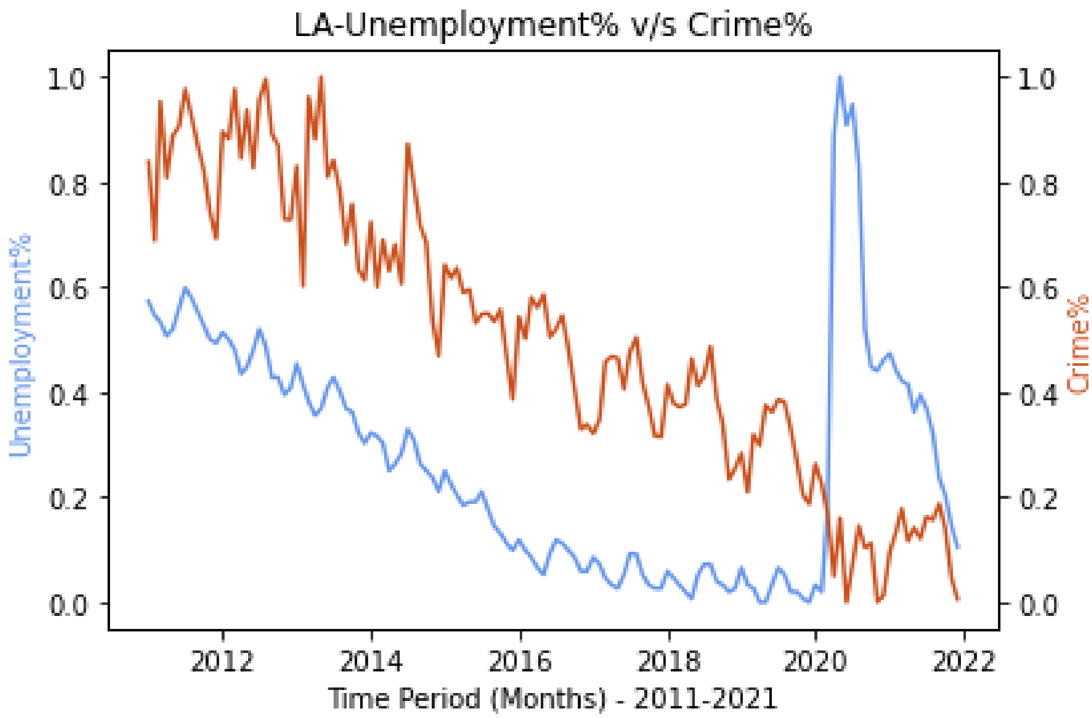


Correlation Graphs:




```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

```
<Figure size 2400x1200 with 0 Axes>
```



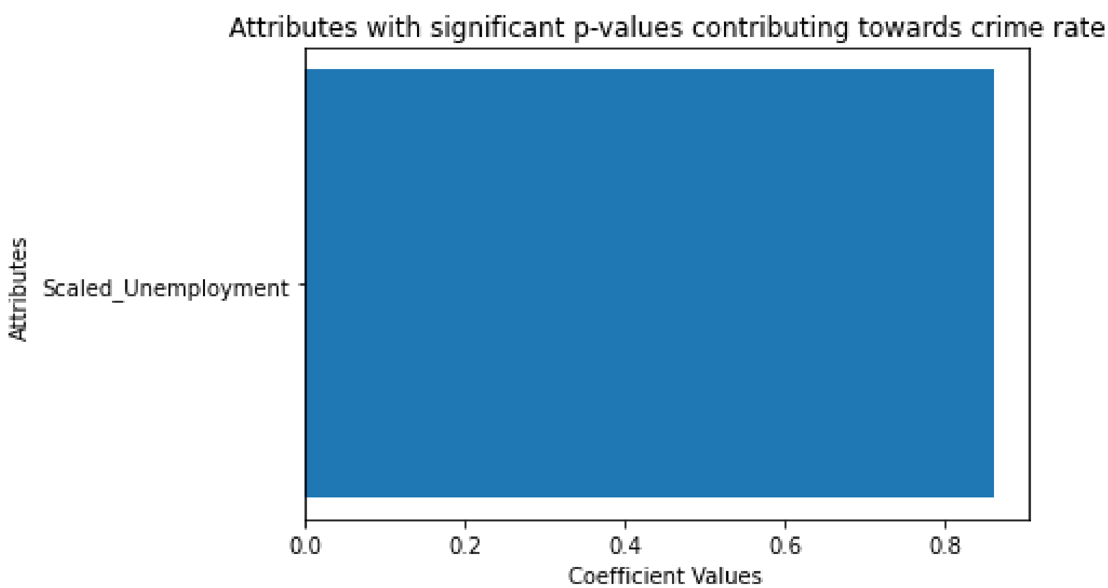
CHICAGO:

Before Covid (Mar-2020)

Ground Truth:
Linear Model summary:

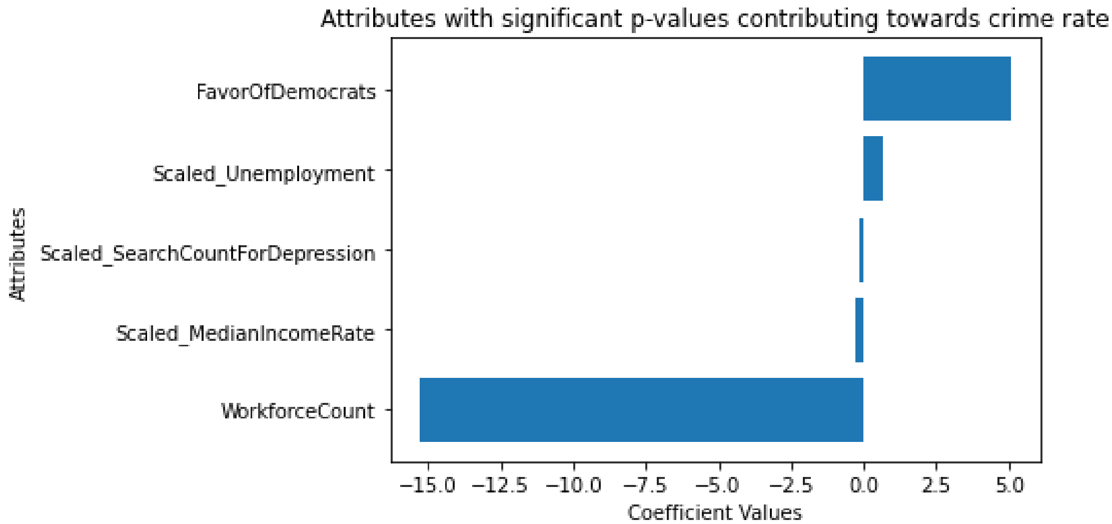
```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.777
Model:                            OLS   Adj. R-squared:                  0.770
Method:                 Least Squares   F-statistic:                     122.8
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.33e-34
Time:                        16:45:25   Log-Likelihood:                 105.41
No. Observations:                 110   AIC:                            -202.8
Df Residuals:                     106   BIC:                            -192.0
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                              coef    std err          t      P>|t|      [0.025
```

```
0.975]
-------------------------------------------------------------------------------
-----------
const                      3.8112      3.316      1.149      0.253      -2.764
10.386
Gender_ratio              -6.5741      6.382     -1.030      0.305     -19.227
6.079
Scaled_Unemployment        0.8608      0.173      4.986      0.000       0.519
1.203
Scaled_MedianIncomeRate   -0.1702      0.108     -1.573      0.119      -0.385
0.044
==============================================================================
Omnibus:                        5.500   Durbin-Watson:                   0.573
Prob(Omnibus):                  0.064   Jarque-Bera (JB):                3.443
Skew:                          -0.252   Prob(JB):                        0.179
Kurtosis:                       2.295   Cond. No.                         960.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
```

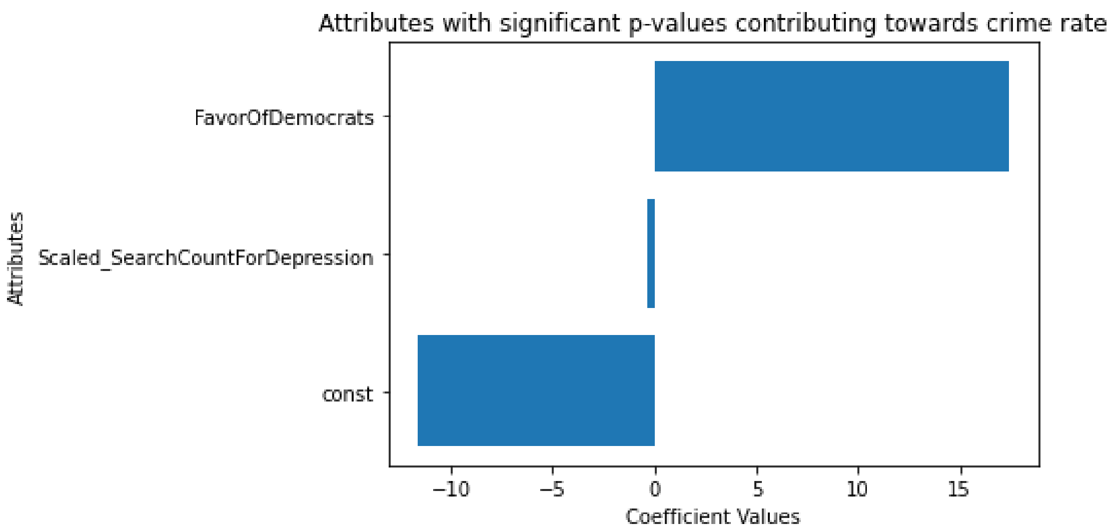
```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.805
Model:                            OLS   Adj. R-squared:                  0.791
Method:                 Least Squares   F-statistic:                     60.05
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.37e-33
Time:                        16:45:25   Log-Likelihood:                 112.82
No. Observations:                 110   AIC:                            -209.6
Df Residuals:                     102   BIC:                            -188.0
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               1.8651      4.017      0.464      0.643
-6.103       9.833
Gender_ratio                       -8.7363      6.720     -1.300      0.197
-22.065       4.593
Scaled_Unemployment                 0.6259      0.177      3.532      0.001
0.274       0.977
Scaled_MedianIncomeRate            -0.2615      0.122     -2.149      0.034
-0.503      -0.020
WorkforceCount                    -15.2992      7.122     -2.148      0.034
-29.425      -1.173
PercentNegativeUsers                0.1180      0.107      1.104      0.272
-0.094       0.330
FavorOfDemocrats                    5.0899      2.230      2.283      0.025
0.667       9.513
Scaled_SearchCountForDepression    -0.1202      0.049     -2.477      0.015
-0.216      -0.024
==============================================================================
Omnibus:                        4.584   Durbin-Watson:                   0.612
Prob(Omnibus):                  0.101   Jarque-Bera (JB):                2.888
Skew:                          -0.205   Prob(JB):                        0.236
Kurtosis:                       2.320   Cond. No.                     1.47e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.47e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```




```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.235
Model:                            OLS   Adj. R-squared:                  0.211
Method:                 Least Squares   F-statistic:                     9.769
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           6.37e-07
Time:                        16:45:25   Log-Likelihood:                 12.958
No. Observations:                 132   AIC:                            -15.92
Df Residuals:                     127   BIC:                            -1.503
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                             -11.6414      3.309     -3.518      0.001
-18.189      -5.094
WorkforceCount                    -18.9933     14.044     -1.352      0.179
-46.785       8.798
PercentNegativeUsers                0.3077      0.251      1.228      0.222
-0.188       0.803
FavorOfDemocrats                   17.4337      5.001      3.486      0.001
7.538       27.329
Scaled_SearchCountForDepression    -0.3310      0.092     -3.608      0.000
```

```
-0.513      -0.149
==============================================================================
Omnibus:                        5.838   Durbin-Watson:                   0.144
Prob(Omnibus):                  0.054   Jarque-Bera (JB):                6.010
Skew:                          -0.510   Prob(JB):                       0.0495
Kurtosis:                       2.774   Cond. No.                     1.05e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.05e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Plot for Linear Model Coefficients for Our Model without ground truth:

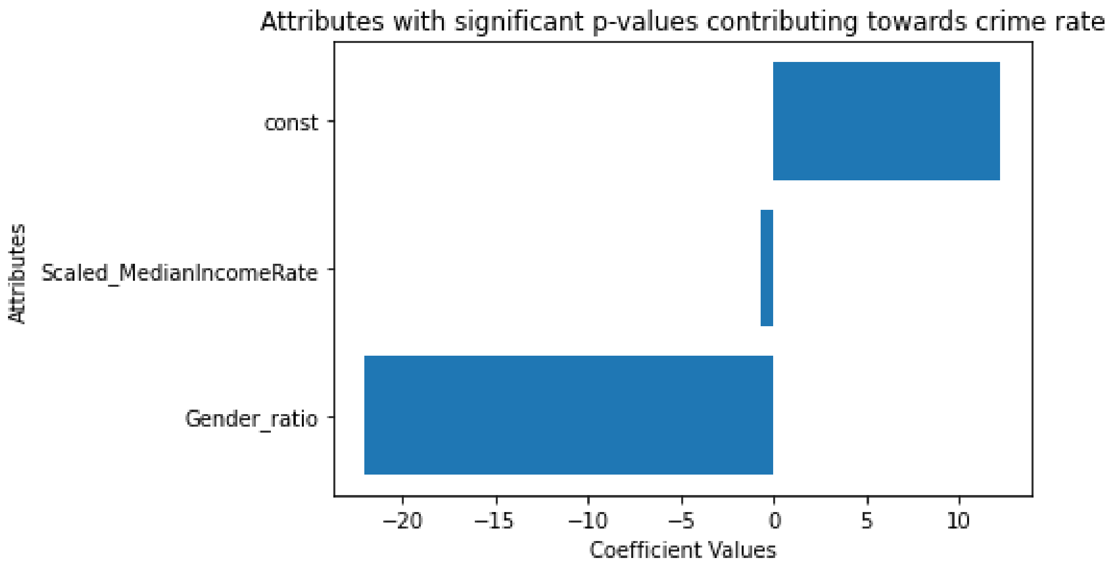


Till Dec-2021

Ground Truth:
Linear Model summary:

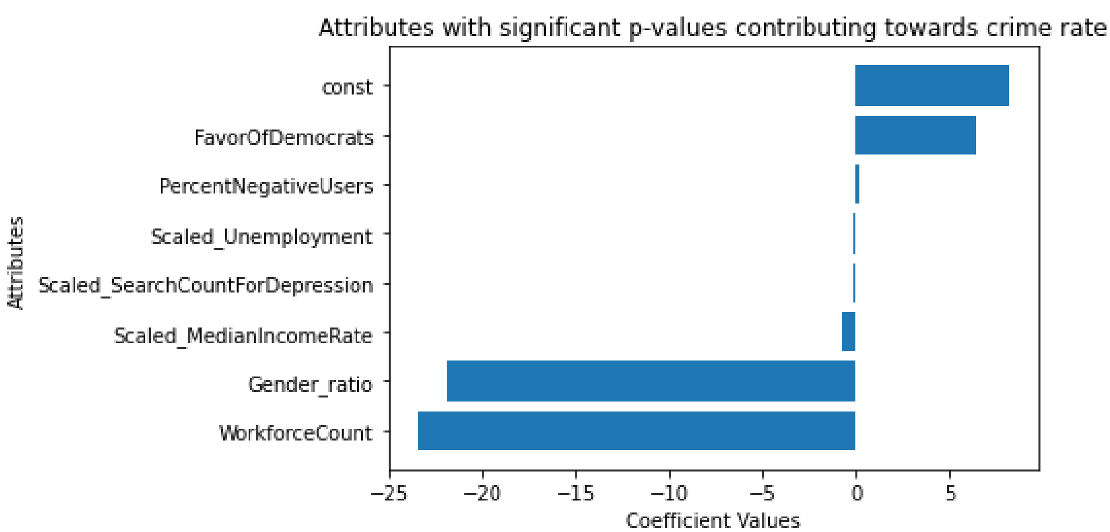
```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.842
Model:                            OLS   Adj. R-squared:                  0.839
Method:                 Least Squares   F-statistic:                     227.8
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.87e-51
```

```
Time:                        16:45:25   Log-Likelihood:                 117.15
No. Observations:                 132   AIC:                            -226.3
Df Residuals:                     128   BIC:                            -214.8
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                          coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                  12.1945      3.184      3.830      0.000       5.894
18.495
Gender_ratio          -22.0308      6.187     -3.561      0.001     -34.273
-9.788
Scaled_Unemployment    -0.1000      0.057     -1.766      0.080      -0.212
0.012
Scaled_MedianIncomeRate    -0.7663      0.032    -24.223      0.000      -0.829
-0.704
==============================================================================
Omnibus:                        0.946   Durbin-Watson:                   0.554
Prob(Omnibus):                  0.623   Jarque-Bera (JB):                0.959
Skew:                           0.060   Prob(JB):                        0.619
Kurtosis:                       2.600   Cond. No.                         980.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.872
Model:                            OLS   Adj. R-squared:                  0.865
Method:                 Least Squares   F-statistic:                     120.5
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.33e-52
Time:                        16:45:25   Log-Likelihood:                 130.85
No. Observations:                 132   AIC:                            -245.7
Df Residuals:                     124   BIC:                            -222.6
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                  coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                           8.2253      3.782      2.175      0.032
0.740      15.711
Gender_ratio                  -21.8722      5.990     -3.652      0.000
-33.727     -10.017
Scaled_Unemployment            -0.1198      0.058     -2.053      0.042
-0.235      -0.004
Scaled_MedianIncomeRate        -0.7440      0.030    -24.400      0.000
-0.804      -0.684
WorkforceCount                -23.4736      5.852     -4.011      0.000
```

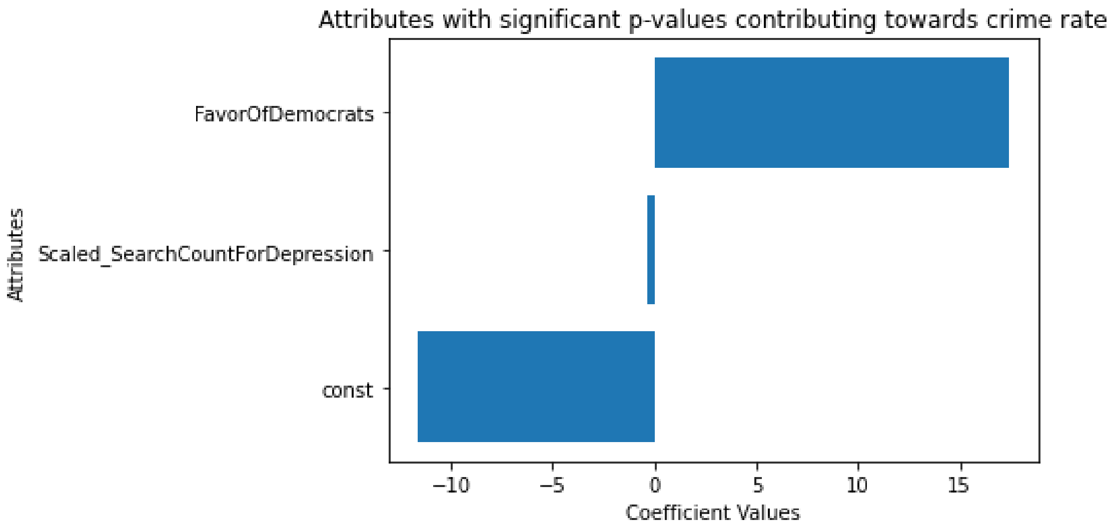
```
-35.056     -11.891
PercentNegativeUsers                  0.2090      0.105      1.988      0.049
0.001       0.417
FavorOfDemocrats                      6.4892      2.224      2.918      0.004
2.087      10.891
Scaled_SearchCountForDepression     -0.1277      0.041     -3.150      0.002
-0.208      -0.047
==============================================================================
Omnibus:                        0.423   Durbin-Watson:                   0.626
Prob(Omnibus):                  0.809   Jarque-Bera (JB):                0.572
Skew:                          -0.017   Prob(JB):                        0.751
Kurtosis:                       2.679   Cond. No.                     1.39e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.39e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```




```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.235
Model:                            OLS   Adj. R-squared:                  0.211
```

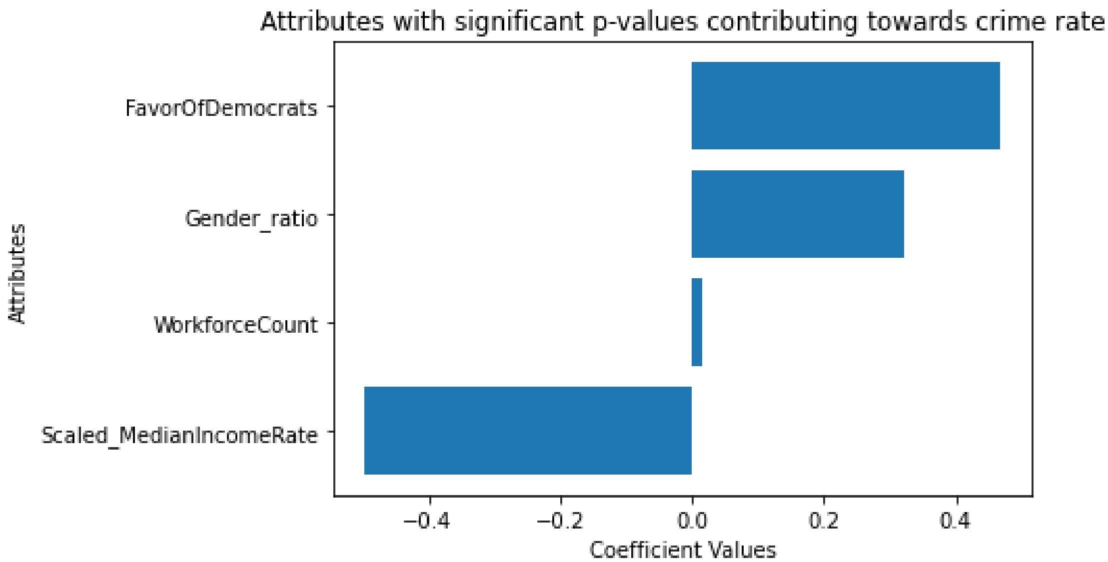
```
Method:                 Least Squares   F-statistic:                     9.769
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           6.37e-07
Time:                        16:45:25   Log-Likelihood:                 12.958
No. Observations:                 132   AIC:                            -15.92
Df Residuals:                     127   BIC:                            -1.503
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                     coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                            -11.6414      3.309     -3.518      0.001
-18.189      -5.094
WorkforceCount                   -18.9933     14.044     -1.352      0.179
-46.785       8.798
PercentNegativeUsers               0.3077      0.251      1.228      0.222
-0.188       0.803
FavorOfDemocrats                  17.4337      5.001      3.486      0.001
7.538      27.329
Scaled_SearchCountForDepression   -0.3310      0.092     -3.608      0.000
-0.513      -0.149
==============================================================================
Omnibus:                        5.838   Durbin-Watson:                   0.144
Prob(Omnibus):                  0.054   Jarque-Bera (JB):                6.010
Skew:                          -0.510   Prob(JB):                       0.0495
Kurtosis:                       2.774   Cond. No.                     1.05e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.05e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Plot for Linear Model Coefficients for Our Model without ground truth:




```
Covid Data Comparison:

Regression from Mar-2020 to Dec-2021 along with Covid Positive Cases Per Month:

Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.491
Model:                            OLS   Adj. R-squared:                  0.331
Method:                 Least Squares   F-statistic:                     3.082
Date:                Sun, 11 Dec 2022   Prob (F-statistic):             0.0389
Time:                        16:45:25   Log-Likelihood:                 41.161
No. Observations:                  22   AIC:                            -70.32
Df Residuals:                      16   BIC:                            -63.78
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
Gender_ratio                        0.3230      0.120      2.689      0.016
0.068       0.578
Scaled_Unemployment                -0.1164      0.059     -1.961      0.067
-0.242       0.009
Scaled_MedianIncomeRate            -0.4951      0.139     -3.574      0.003
```

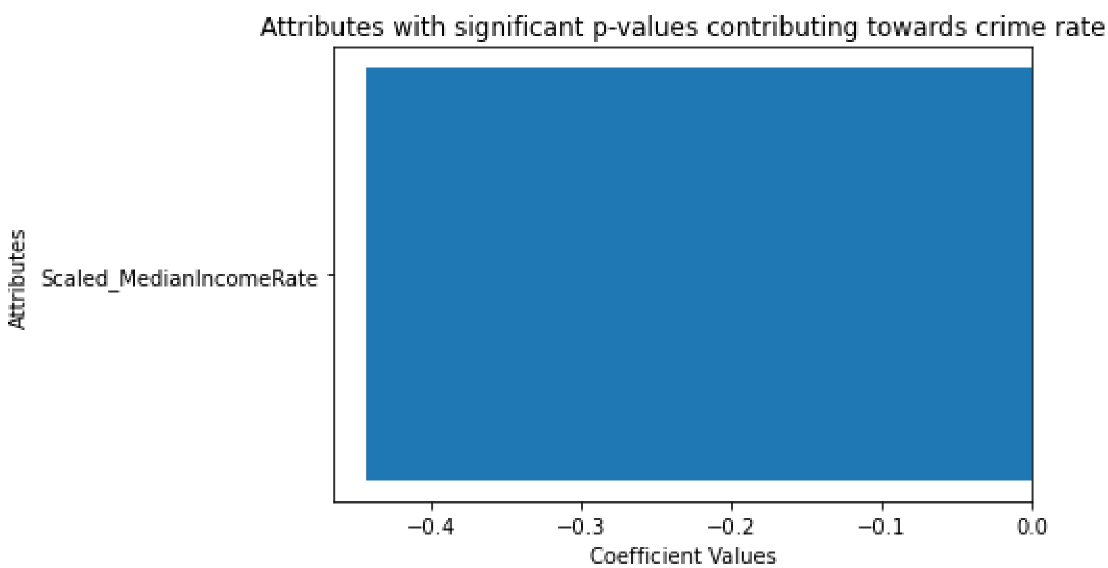
```
-0.789      -0.201
WorkforceCount                        0.0155      0.007      2.280      0.037
0.001       0.030
PercentNegativeUsers                  0.1907      0.326      0.585      0.567
-0.501       0.882
FavorOfDemocrats                      0.4663      0.173      2.689      0.016
0.099       0.834
Scaled_SearchCountForDepression       0.1280      0.086      1.488      0.156
-0.054       0.310
CovidPercPositive                    -0.4447      0.238     -1.870      0.080
-0.949       0.059
==============================================================================
Omnibus:                        3.575   Durbin-Watson:                   2.116
Prob(Omnibus):                  0.167   Jarque-Bera (JB):                1.457
Skew:                           0.182   Prob(JB):                        0.483
Kurtosis:                       1.793   Cond. No.                     3.38e+18
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 4.04e-36. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.


Plot for Linear Model Coefficients for Our Model:


/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1807615713.py:1
9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  city_combined_dataset_covid['CovidPercPositive'] = perc
```



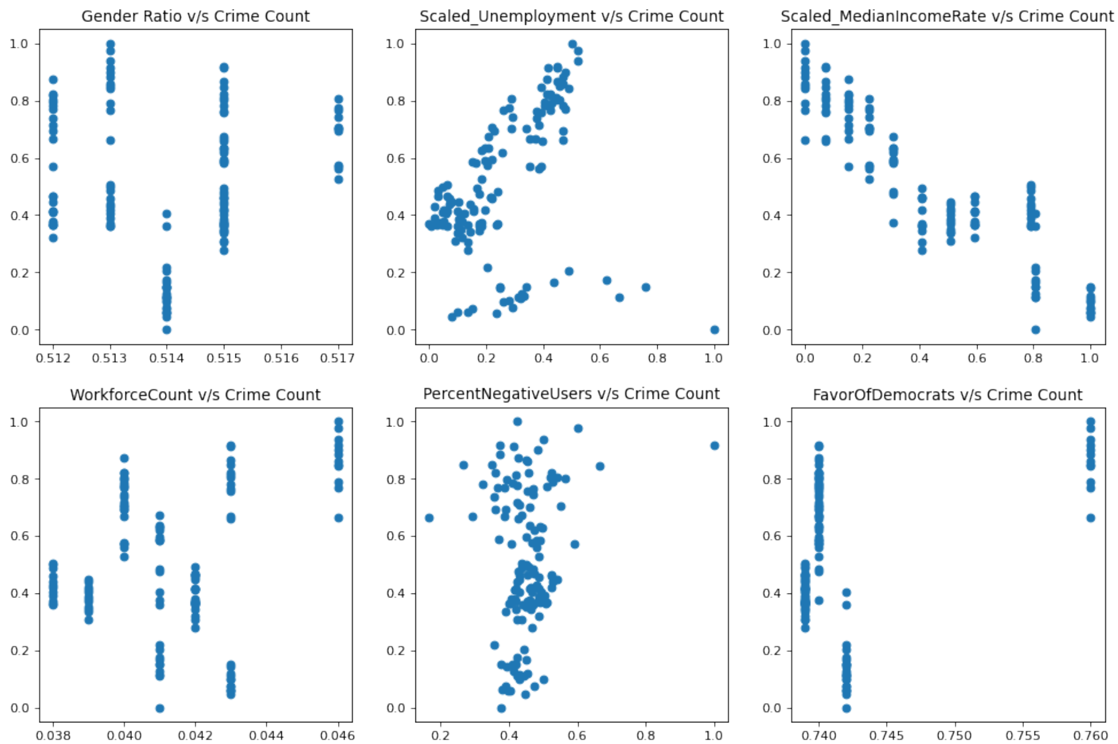

```
Regression for the same time frame without considering Covid Cases:

Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.379
Model:                            OLS   Adj. R-squared:                  0.233
Method:                 Least Squares   F-statistic:                     2.596
Date:                Sun, 11 Dec 2022   Prob (F-statistic):             0.0734
Time:                        16:45:25   Log-Likelihood:                 38.986
No. Observations:                  22   AIC:                            -67.97
Df Residuals:                      17   BIC:                            -62.52
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
Gender_ratio                        0.2576      0.123      2.093      0.052
-0.002       0.517
Scaled_Unemployment               -0.1220      0.063     -1.922      0.072
-0.256       0.012
Scaled_MedianIncomeRate           -0.4436      0.145     -3.051      0.007
-0.750      -0.137
```

```
WorkforceCount                     0.0119      0.007      1.701      0.107
-0.003      0.027
PercentNegativeUsers               0.3433      0.338      1.015      0.324
-0.370      1.057
FavorOfDemocrats                   0.3719      0.178      2.093      0.052
-0.003      0.747
Scaled_SearchCountForDepression    0.0235      0.070      0.336      0.741
-0.124      0.171
==============================================================================
Omnibus:                      1.146   Durbin-Watson:                   1.791
Prob(Omnibus):                0.564   Jarque-Bera (JB):                0.834
Skew:                        -0.092   Prob(JB):                        0.659
Kurtosis:                     2.064   Cond. No.                     5.82e+18
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 1.36e-36. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Plot for Linear Model Coefficients for Our Model:

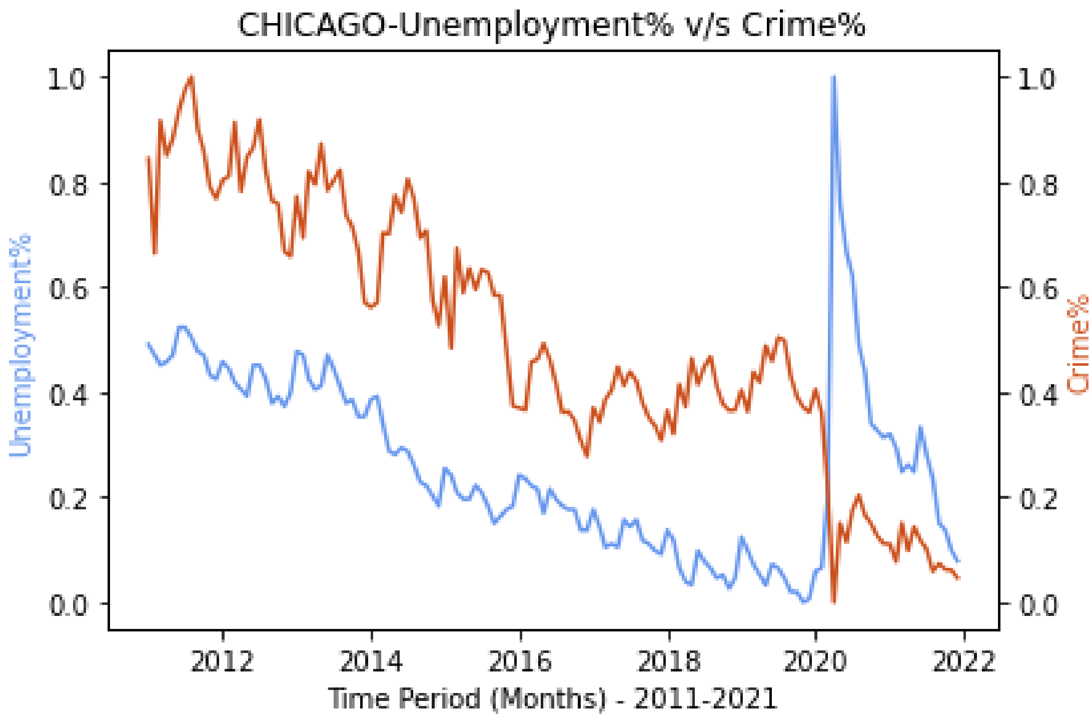


Correlation Graphs:



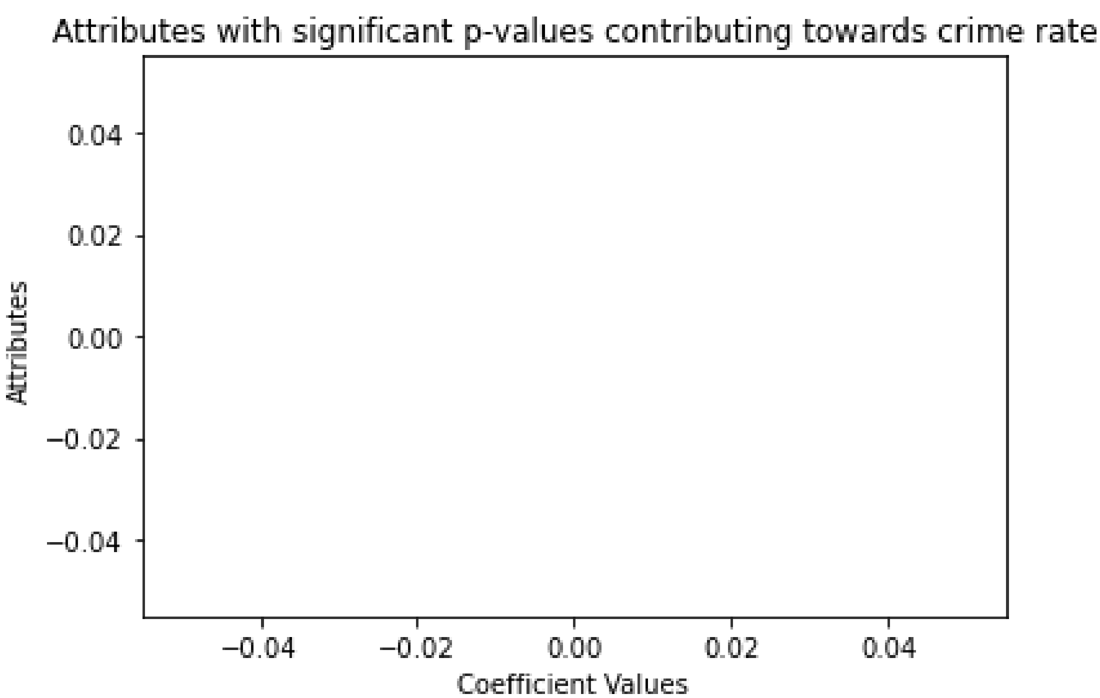

```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

```
<Figure size 2400x1200 with 0 Axes>
```




```
NASHVILLE:


Before Covid (Mar-2020)


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.058
Model:                            OLS   Adj. R-squared:                  0.009
Method:                 Least Squares   F-statistic:                     1.190
Date:                Sun, 11 Dec 2022   Prob (F-statistic):              0.322
Time:                        16:45:26   Log-Likelihood:                 16.556
No. Observations:                  62   AIC:                            -25.11
Df Residuals:                      58   BIC:                            -16.60
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                              coef    std err          t      P>|t|      [0.025
```

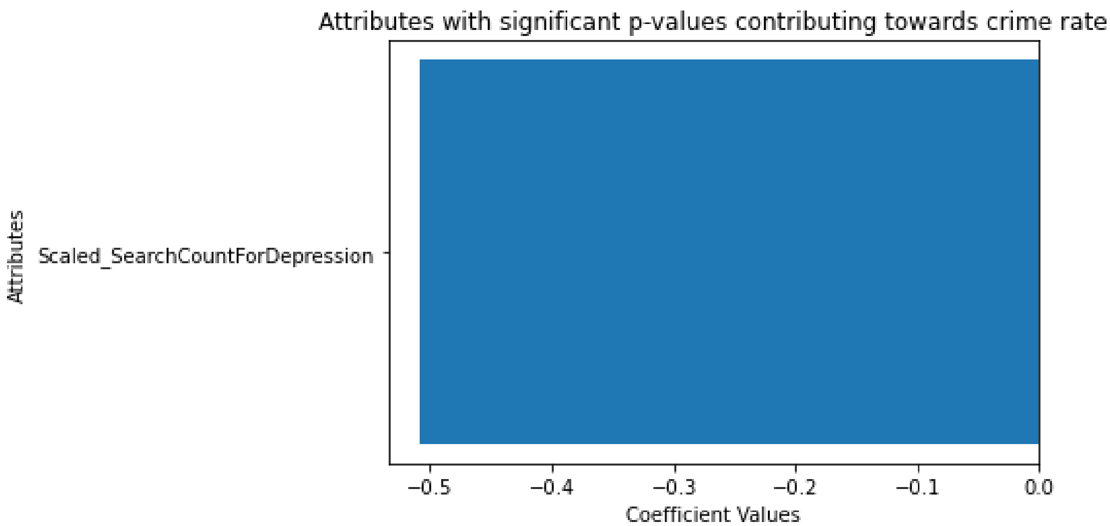
```
0.975]
--------------------------------------------------------------------------------
-----------
const                       25.2928     21.064      1.201      0.235     -16.871
67.456
Gender_ratio               -47.2977     40.717     -1.162      0.250    -128.802
34.207
Scaled_Unemployment         -0.2374      1.015     -0.234      0.816      -2.269
1.795
Scaled_MedianIncomeRate     -0.2129      0.152     -1.404      0.166      -0.516
0.091
==============================================================================
Omnibus:                       13.969   Durbin-Watson:                   1.196
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               16.279
Skew:                          -0.966   Prob(JB):                     0.000292
Kurtosis:                       4.604   Cond. No.                     2.31e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.31e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Ground Truth:
```



```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.346
Model:                            OLS   Adj. R-squared:                  0.261
Method:                 Least Squares   F-statistic:                     4.081
Date:                Sun, 11 Dec 2022   Prob (F-statistic):            0.00117
Time:                        16:45:26   Log-Likelihood:                 27.867
No. Observations:                  62   AIC:                            -39.73
Df Residuals:                      54   BIC:                            -22.72
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              29.6462     25.376      1.168      0.248
-21.229      80.521
Gender_ratio                      -51.5976     49.685     -1.038      0.304
-151.210      48.015
Scaled_Unemployment                -1.1478      0.913     -1.257      0.214
-2.978       0.683
Scaled_MedianIncomeRate            -0.1912      0.135     -1.414      0.163
-0.462       0.080
WorkforceCount                      6.1483      9.365      0.657      0.514
-12.627      24.923
PercentNegativeUsers                0.0516      0.611      0.084      0.933
-1.174       1.277
FavorOfDemocrats                   -3.4526      2.498     -1.382      0.173
-8.460       1.555
Scaled_SearchCountForDepression    -0.5094      0.107     -4.753      0.000
-0.724      -0.294
==============================================================================
Omnibus:                        8.402   Durbin-Watson:                   1.828
Prob(Omnibus):                  0.015   Jarque-Bera (JB):                7.749
Skew:                          -0.760   Prob(JB):                       0.0208
Kurtosis:                       3.828   Cond. No.                     3.97e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

```
[2] The condition number is large, 3.97e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

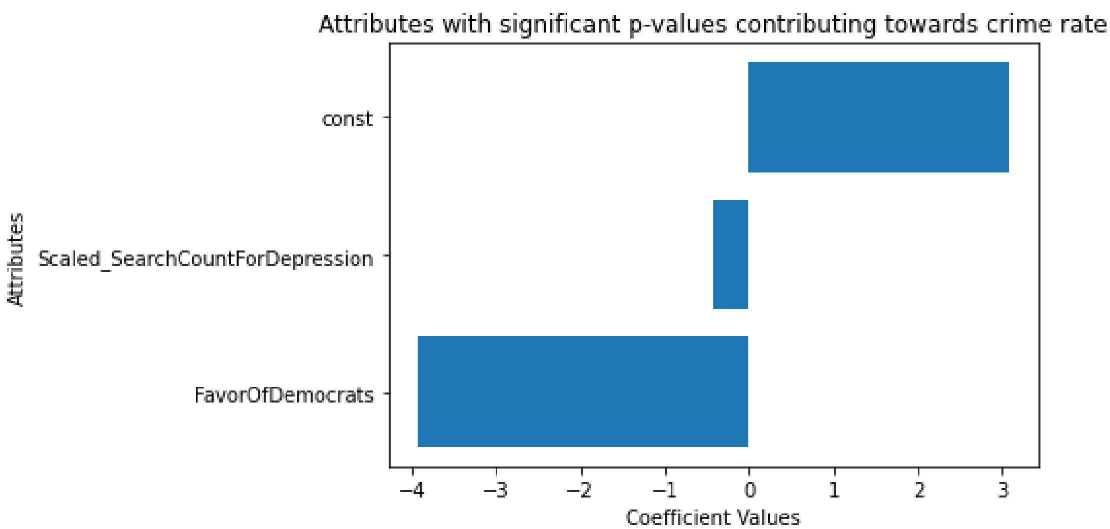
```
Plot for Linear Model Coefficients for Our Model:
```




```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.285
Model:                            OLS   Adj. R-squared:                  0.249
Method:                 Least Squares   F-statistic:                     7.863
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.18e-05
Time:                        16:45:26   Log-Likelihood:                 25.564
No. Observations:                  84   AIC:                            -41.13
Df Residuals:                      79   BIC:                            -28.97
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               3.0887      0.679      4.550      0.000
1.737       4.440
WorkforceCount                     -2.9960      6.354     -0.472      0.639
-15.643       9.651
```

```
PercentNegativeUsers                 0.8915      0.560      1.591      0.116
-0.224      2.007
FavorOfDemocrats                    -3.9353      0.920     -4.279      0.000
-5.766     -2.105
Scaled_SearchCountForDepression    -0.4172      0.100     -4.188      0.000
-0.615     -0.219
==============================================================================
Omnibus:                        5.727   Durbin-Watson:                   1.328
Prob(Omnibus):                  0.057   Jarque-Bera (JB):                4.994
Skew:                          -0.551   Prob(JB):                       0.0823
Kurtosis:                       3.462   Cond. No.                         412.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:

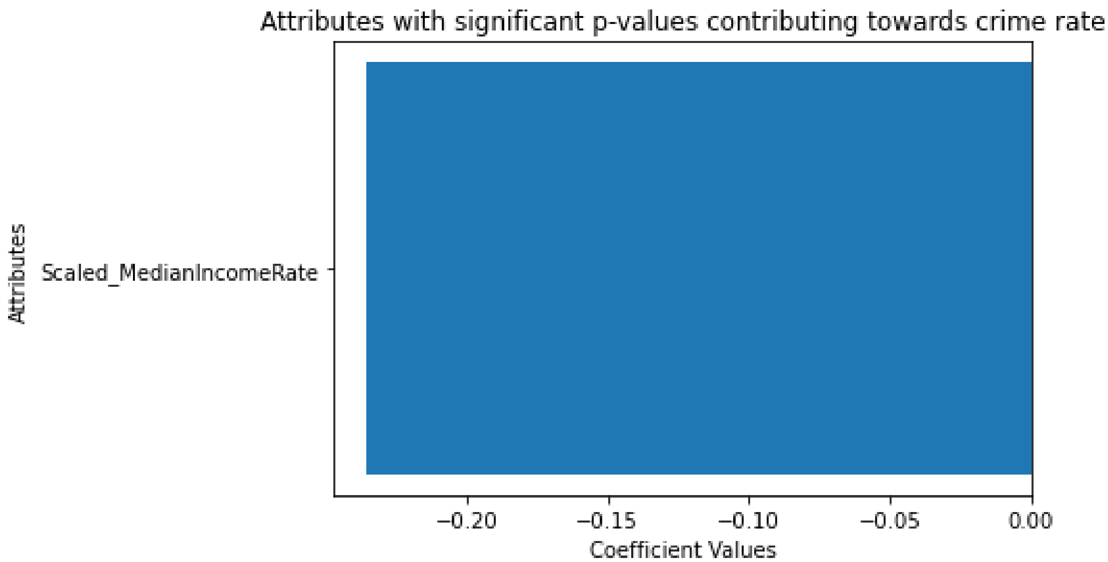


Till Dec-2021

Ground Truth:
Linear Model summary:

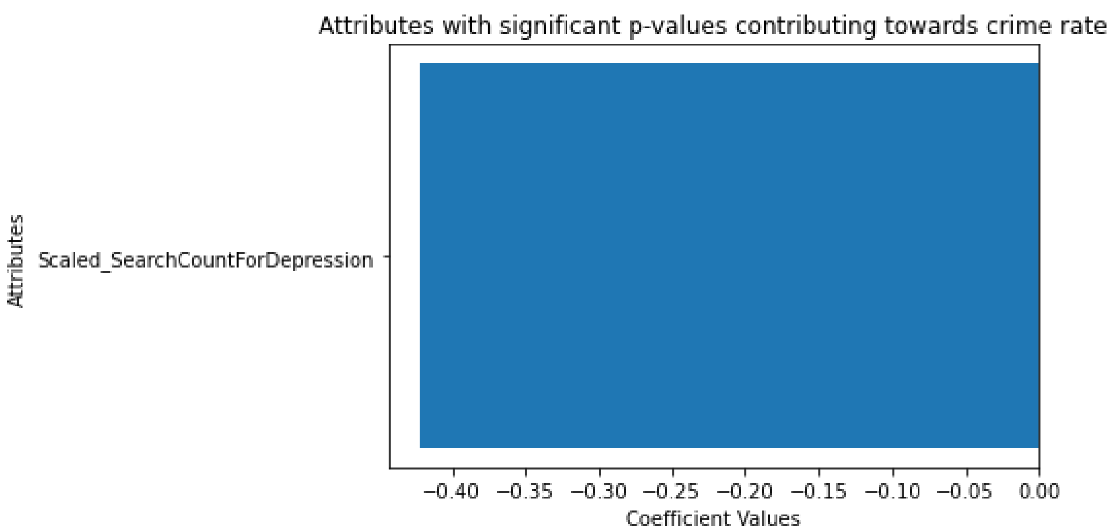
```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.129
```

```
Model:                            OLS   Adj. R-squared:                  0.097
Method:                 Least Squares   F-statistic:                     3.963
Date:                Sun, 11 Dec 2022   Prob (F-statistic):             0.0109
Time:                        16:45:26   Log-Likelihood:                 17.307
No. Observations:                  84   AIC:                            -26.61
Df Residuals:                      80   BIC:                            -16.89
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                           coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                   30.5696     18.361      1.665      0.100      -5.971
67.110
Gender_ratio           -57.4840     35.408     -1.623      0.108    -127.948
12.980
Scaled_Unemployment     -0.1250      0.128     -0.973      0.333      -0.381
0.131
Scaled_MedianIncomeRate    -0.2351      0.078     -3.025      0.003      -0.390
-0.080
==============================================================================
Omnibus:                       10.873   Durbin-Watson:                   1.043
Prob(Omnibus):                  0.004   Jarque-Bera (JB):               11.215
Skew:                          -0.756   Prob(JB):                      0.00367
Kurtosis:                       3.958   Cond. No.                     2.32e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.32e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Ground Truth:
```



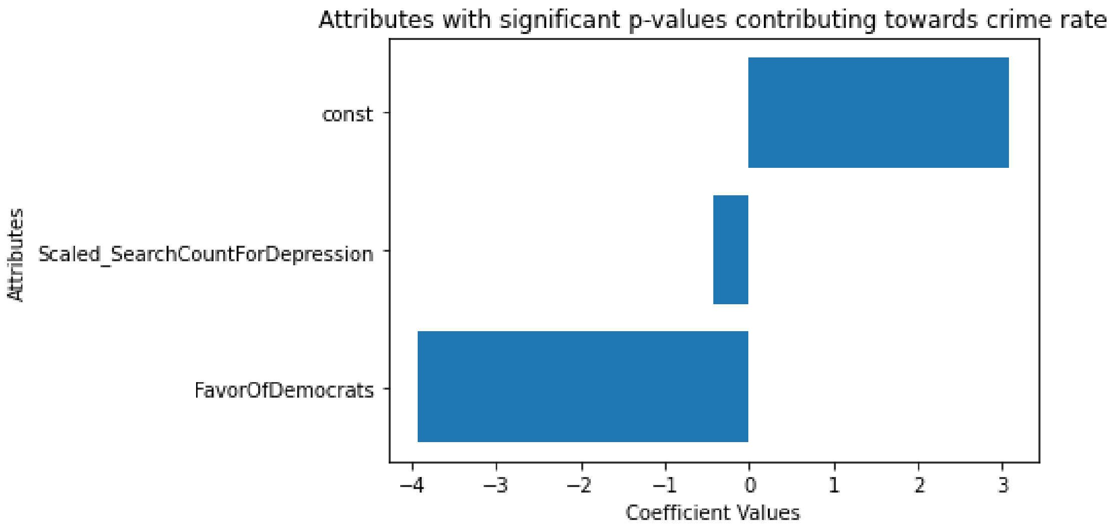

```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.319
Model:                            OLS   Adj. R-squared:                  0.257
Method:                 Least Squares   F-statistic:                     5.097
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           8.82e-05
Time:                        16:45:26   Log-Likelihood:                 27.655
No. Observations:                  84   AIC:                            -39.31
Df Residuals:                      76   BIC:                            -19.86
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              47.7534     24.250      1.969      0.053
-0.544      96.051
Gender_ratio                      -88.9294     47.761     -1.862      0.066
-184.053       6.194
Scaled_Unemployment                -0.0661      0.143     -0.461      0.646
-0.352       0.219
Scaled_MedianIncomeRate            -0.1551      0.112     -1.390      0.169
-0.378       0.067
WorkforceCount                      8.6963      9.930      0.876      0.384
```

```
-11.081      28.473
PercentNegativeUsers                 0.8941      0.568      1.574      0.120
-0.237       2.025
FavorOfDemocrats                    -2.1912      1.689     -1.297      0.198
-5.555       1.173
Scaled_SearchCountForDepression    -0.4226      0.102     -4.150      0.000
-0.625      -0.220
==============================================================================
Omnibus:                        3.943   Durbin-Watson:                   1.365
Prob(Omnibus):                  0.139   Jarque-Bera (JB):                3.213
Skew:                          -0.446   Prob(JB):                        0.201
Kurtosis:                       3.351   Cond. No.                     4.13e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 4.13e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```

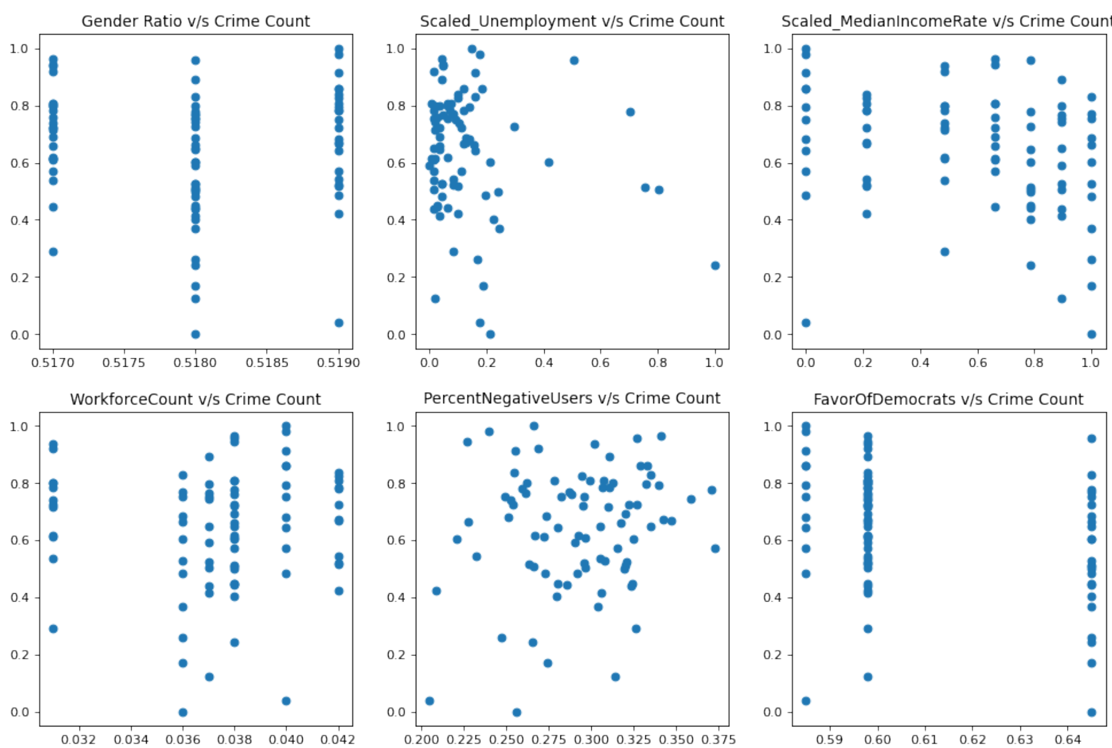



```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.285
Model:                            OLS   Adj. R-squared:                  0.249
```

```
Method:                 Least Squares   F-statistic:                     7.863
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.18e-05
Time:                        16:45:26   Log-Likelihood:                 25.564
No. Observations:                  84   AIC:                            -41.13
Df Residuals:                      79   BIC:                            -28.97
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               3.0887      0.679      4.550      0.000
1.737       4.440
WorkforceCount                     -2.9960      6.354     -0.472      0.639
-15.643       9.651
PercentNegativeUsers                0.8915      0.560      1.591      0.116
-0.224       2.007
FavorOfDemocrats                   -3.9353      0.920     -4.279      0.000
-5.766      -2.105
Scaled_SearchCountForDepression    -0.4172      0.100     -4.188      0.000
-0.615      -0.219
==============================================================================
Omnibus:                        5.727   Durbin-Watson:                   1.328
Prob(Omnibus):                  0.057   Jarque-Bera (JB):                4.994
Skew:                          -0.551   Prob(JB):                       0.0823
Kurtosis:                       3.462   Cond. No.                         412.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Our Model without ground truth:
```



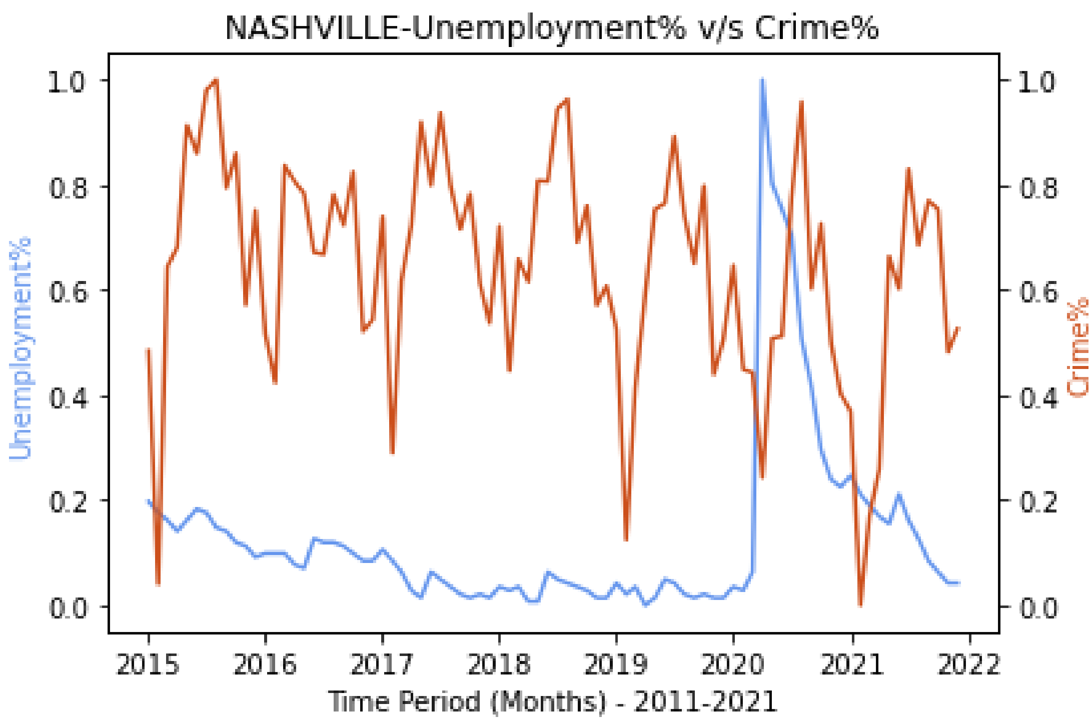
Correlation Graphs:




```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
```

```
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

<Figure size 2400x1200 with 0 Axes>



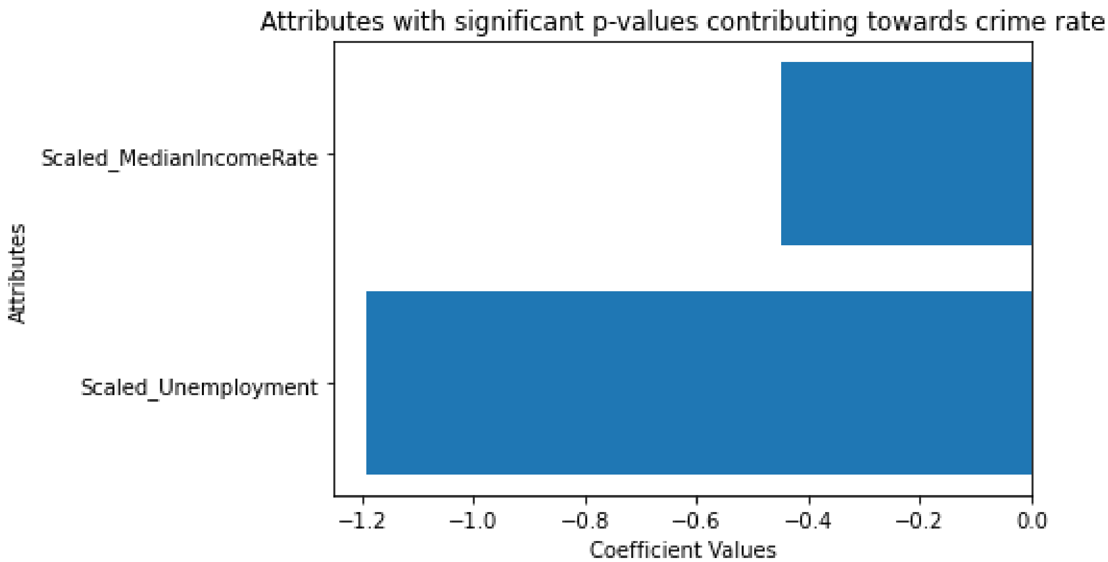
BOSTON:

Before Covid (Mar-2020)

Ground Truth:
Linear Model summary:

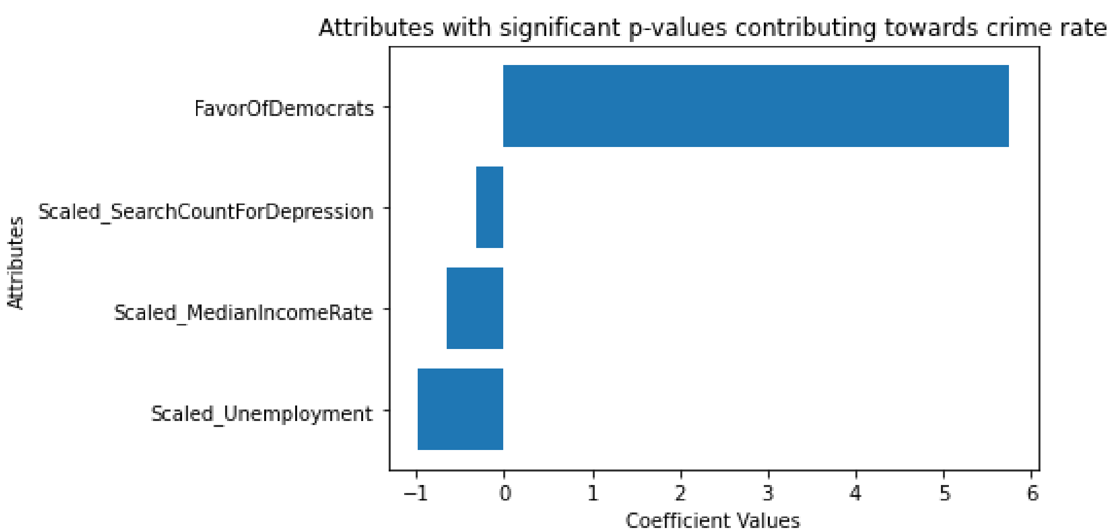
```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.155
Model:                            OLS   Adj. R-squared:                  0.126
```

```
Method:                 Least Squares   F-statistic:                     5.365
Date:                Sun, 11 Dec 2022   Prob (F-statistic):            0.00194
Time:                        16:45:27   Log-Likelihood:                 36.943
No. Observations:                  92   AIC:                            -65.89
Df Residuals:                      88   BIC:                            -55.80
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                            coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                     6.7389      5.868      1.148      0.254      -4.922
18.400
Gender_ratio            -11.0116     11.369     -0.969      0.335     -33.605
11.582
Scaled_Unemployment      -1.1918      0.396     -3.007      0.003      -1.979
-0.404
Scaled_MedianIncomeRate  -0.4468      0.147     -3.037      0.003      -0.739
-0.154
==============================================================================
Omnibus:                       27.005   Durbin-Watson:                   1.020
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               55.350
Skew:                          -1.095   Prob(JB):                     9.57e-13
Kurtosis:                       6.105   Cond. No.                         877.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```



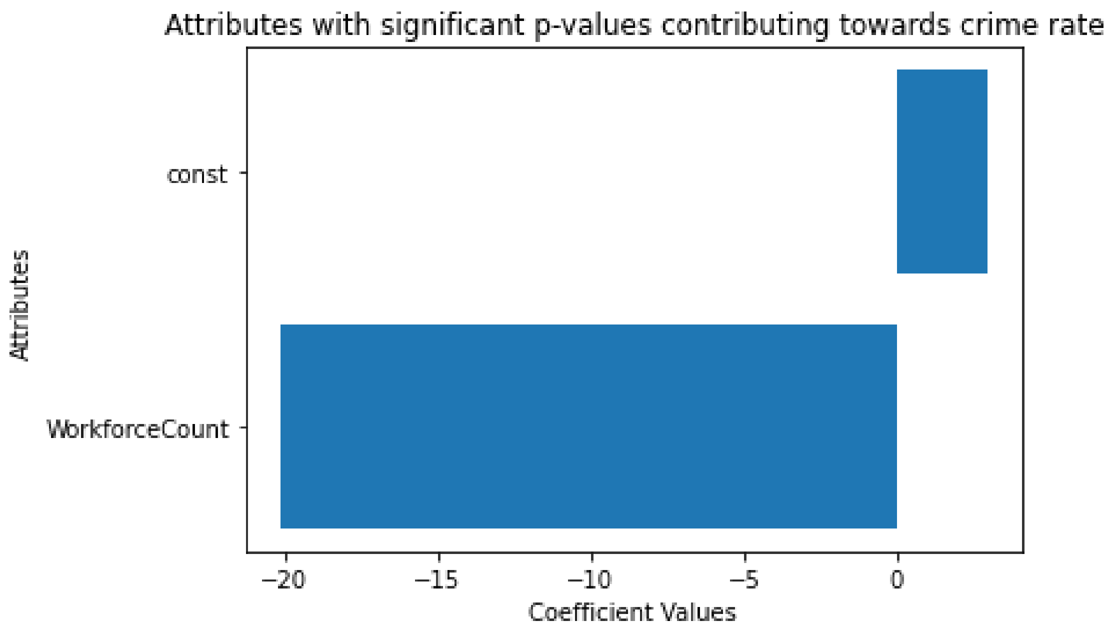

```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.456
Model:                            OLS   Adj. R-squared:                  0.411
Method:                 Least Squares   F-statistic:                     10.07
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.42e-09
Time:                        16:45:27   Log-Likelihood:                 57.244
No. Observations:                  92   AIC:                            -98.49
Df Residuals:                      84   BIC:                            -78.31
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               3.6810      5.353      0.688      0.494
-6.965      14.327
Gender_ratio                      -12.7624      9.638     -1.324      0.189
-31.929       6.404
Scaled_Unemployment                -0.9871      0.370     -2.670      0.009
-1.722      -0.252
Scaled_MedianIncomeRate            -0.6493      0.136     -4.763      0.000
-0.920      -0.378
WorkforceCount                     -7.3452      5.313     -1.383      0.170
```

```
-17.910       3.219
PercentNegativeUsers                  -0.4246      0.251     -1.694      0.094
-0.923       0.074
FavorOfDemocrats                       5.7494      1.455      3.951      0.000
2.856       8.643
Scaled_SearchCountForDepression       -0.3163      0.071     -4.437      0.000
-0.458      -0.175
==============================================================================
Omnibus:                       62.383   Durbin-Watson:                   1.804
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              369.691
Skew:                          -2.081   Prob(JB):                     5.28e-81
Kurtosis:                      11.895   Cond. No.                     1.22e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.22e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```



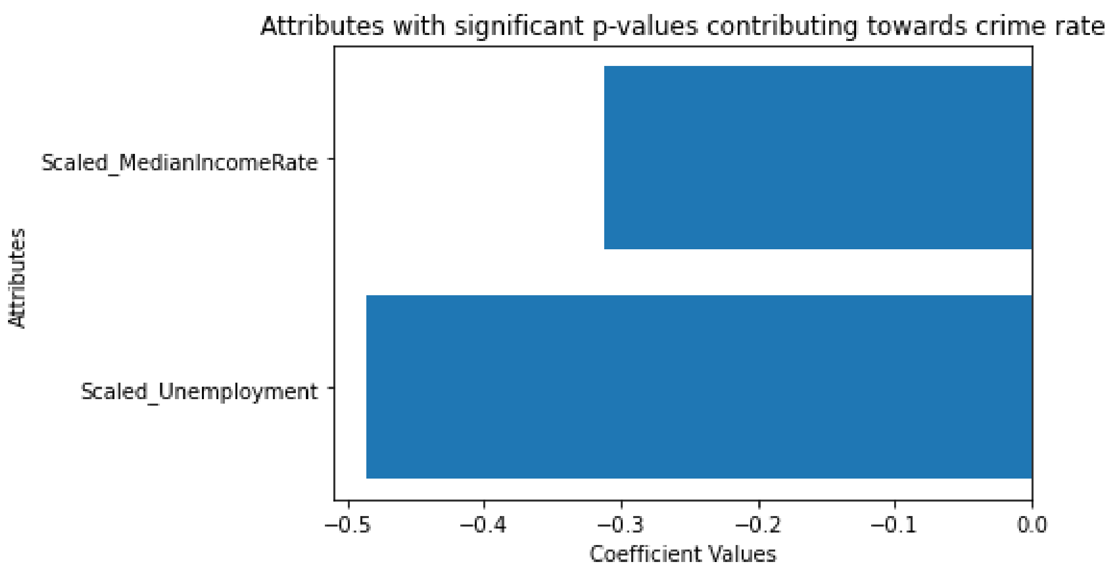

```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.147
Model:                            OLS   Adj. R-squared:                  0.116
```

```
Method:                 Least Squares   F-statistic:                     4.704
Date:                Sun, 11 Dec 2022   Prob (F-statistic):            0.00153
Time:                        16:45:27   Log-Likelihood:                 20.606
No. Observations:                 114   AIC:                            -31.21
Df Residuals:                     109   BIC:                            -17.53
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                    coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                             2.9348      0.776      3.780      0.000
1.396       4.473
WorkforceCount                  -20.1339      6.048     -3.329      0.001
-32.121      -8.147
PercentNegativeUsers             -0.5454      0.363     -1.501      0.136
-1.265       0.175
FavorOfDemocrats                 -1.5521      0.945     -1.642      0.103
-3.425       0.321
Scaled_SearchCountForDepression  -0.1523      0.087     -1.753      0.082
-0.324       0.020
==============================================================================
Omnibus:                       13.750   Durbin-Watson:                   0.626
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               14.911
Skew:                          -0.790   Prob(JB):                     0.000578
Kurtosis:                       3.803   Cond. No.                         445.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Our Model without ground truth:
```



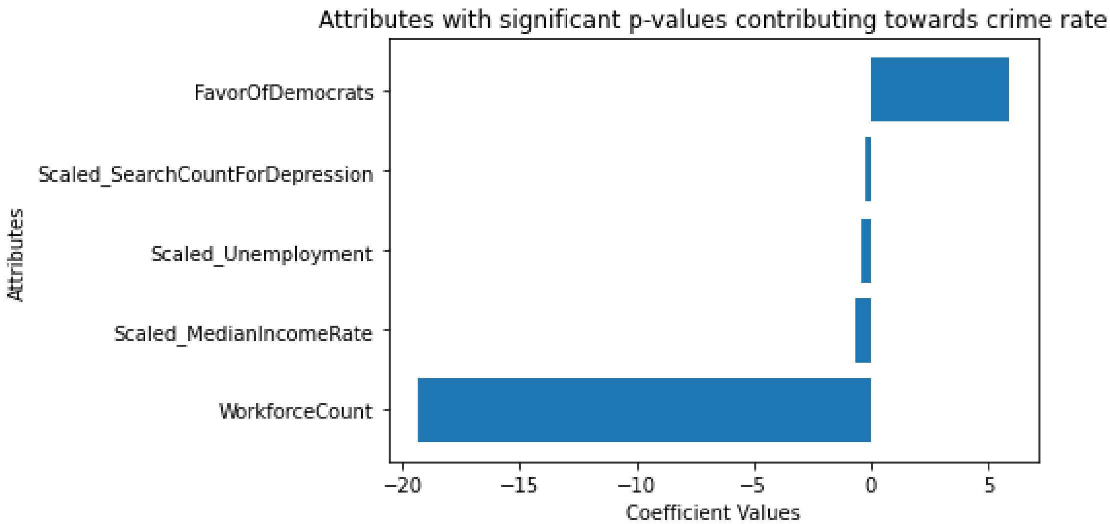

```
Till Dec-2021


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.406
Model:                            OLS   Adj. R-squared:                  0.389
Method:                 Least Squares   F-statistic:                     25.02
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.05e-12
Time:                        16:45:27   Log-Likelihood:                 41.175
No. Observations:                 114   AIC:                            -74.35
Df Residuals:                     110   BIC:                            -63.41
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                              coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                       3.0473      5.229      0.583      0.561      -7.315
13.410
Gender_ratio               -4.2061     10.068     -0.418      0.677     -24.159
15.747
```

```
Scaled_Unemployment        -0.4855      0.095     -5.132      0.000      -0.673
-0.298
Scaled_MedianIncomeRate    -0.3119      0.053     -5.913      0.000      -0.416
-0.207
==============================================================================
Omnibus:                       16.919   Durbin-Watson:                   0.804
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               22.552
Skew:                          -0.773   Prob(JB):                     1.27e-05
Kurtosis:                       4.534   Cond. No.                         865.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.600
Model:                            OLS   Adj. R-squared:                  0.573
Method:                 Least Squares   F-statistic:                     22.70
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           1.58e-18
Time:                        16:45:27   Log-Likelihood:                 63.738
No. Observations:                 114   AIC:                            -111.5
```

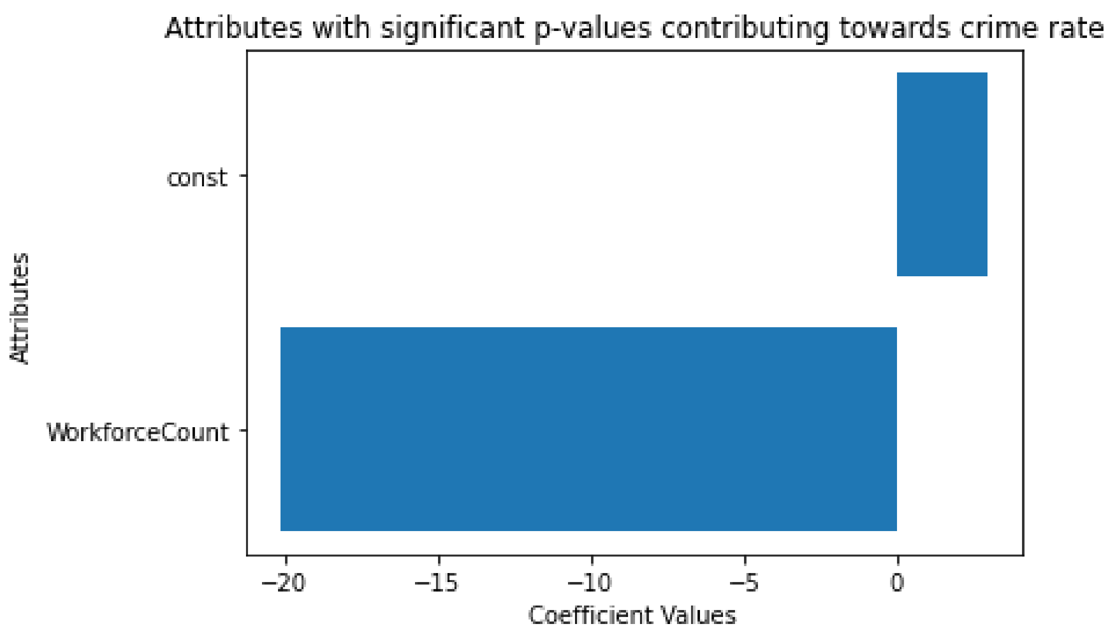
```
Df Residuals:                     106   BIC:                            -89.59
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                   coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                            1.4250      5.359      0.266      0.791
-9.200      12.050
Gender_ratio                    -8.1040      9.358     -0.866      0.388
-26.656      10.448
Scaled_Unemployment             -0.3978      0.092     -4.335      0.000
-0.580      -0.216
Scaled_MedianIncomeRate         -0.6313      0.100     -6.305      0.000
-0.830      -0.433
WorkforceCount                 -19.3595      4.582     -4.225      0.000
-28.443     -10.276
PercentNegativeUsers            -0.3063      0.256     -1.197      0.234
-0.813       0.201
FavorOfDemocrats                 5.9098      1.381      4.278      0.000
3.171       8.649
Scaled_SearchCountForDepression    -0.2234      0.063     -3.547      0.001
-0.348      -0.099
==============================================================================
Omnibus:                       43.500   Durbin-Watson:                   1.314
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              133.275
Skew:                          -1.356   Prob(JB):                     1.15e-29
Kurtosis:                       7.550   Cond. No.                     1.29e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.29e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```



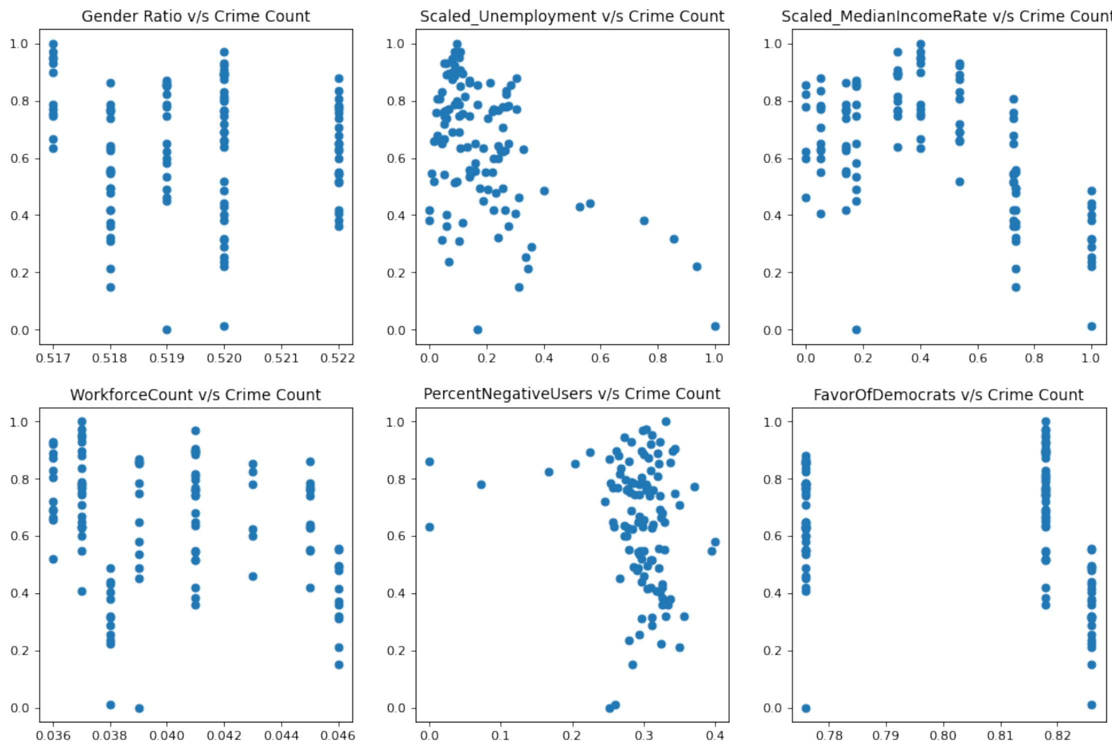

```
Our Model without ground truth:
Linear Model summary:
                             OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.147
Model:                            OLS   Adj. R-squared:                  0.116
Method:                 Least Squares   F-statistic:                     4.704
Date:                Sun, 11 Dec 2022   Prob (F-statistic):            0.00153
Time:                        16:45:27   Log-Likelihood:                 20.606
No. Observations:                 114   AIC:                            -31.21
Df Residuals:                     109   BIC:                            -17.53
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               2.9348      0.776      3.780      0.000
1.396       4.473
WorkforceCount                    -20.1339      6.048     -3.329      0.001
-32.121      -8.147
PercentNegativeUsers               -0.5454      0.363     -1.501      0.136
-1.265       0.175
FavorOfDemocrats                   -1.5521      0.945     -1.642      0.103
-3.425       0.321
Scaled_SearchCountForDepression    -0.1523      0.087     -1.753      0.082
-0.324       0.020
```

```
==============================================================================
Omnibus:                       13.750   Durbin-Watson:                   0.626
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               14.911
Skew:                          -0.790   Prob(JB):                     0.000578
Kurtosis:                       3.803   Cond. No.                         445.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:

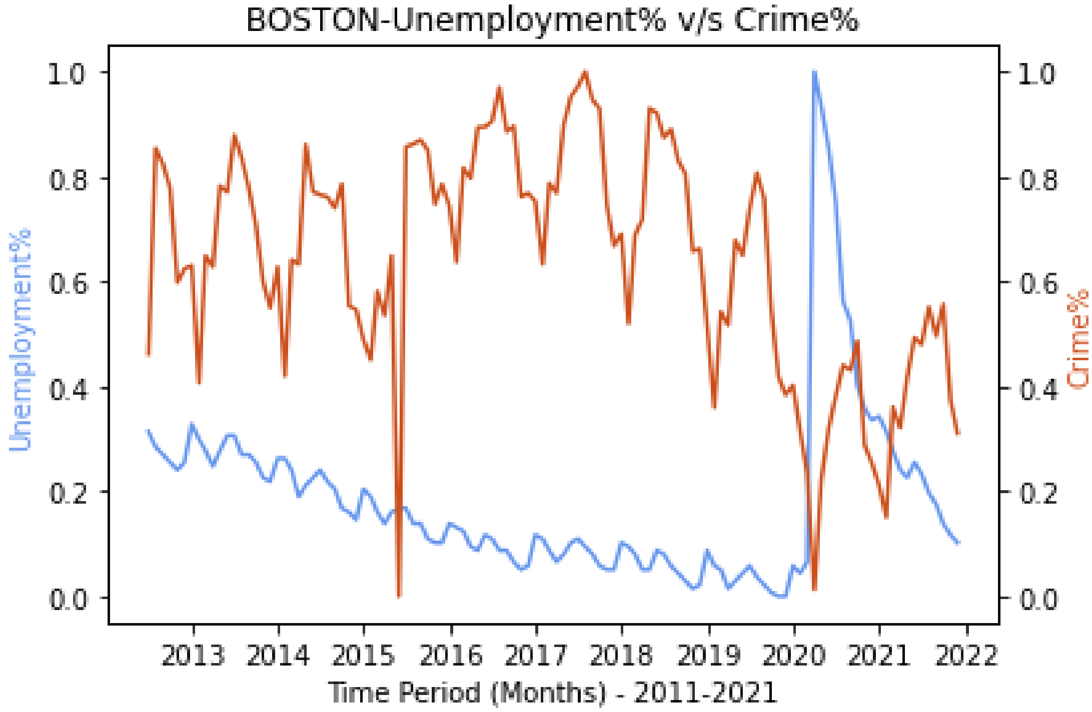


Correlation Graphs:




```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

```
<Figure size 2400x1200 with 0 Axes>
```



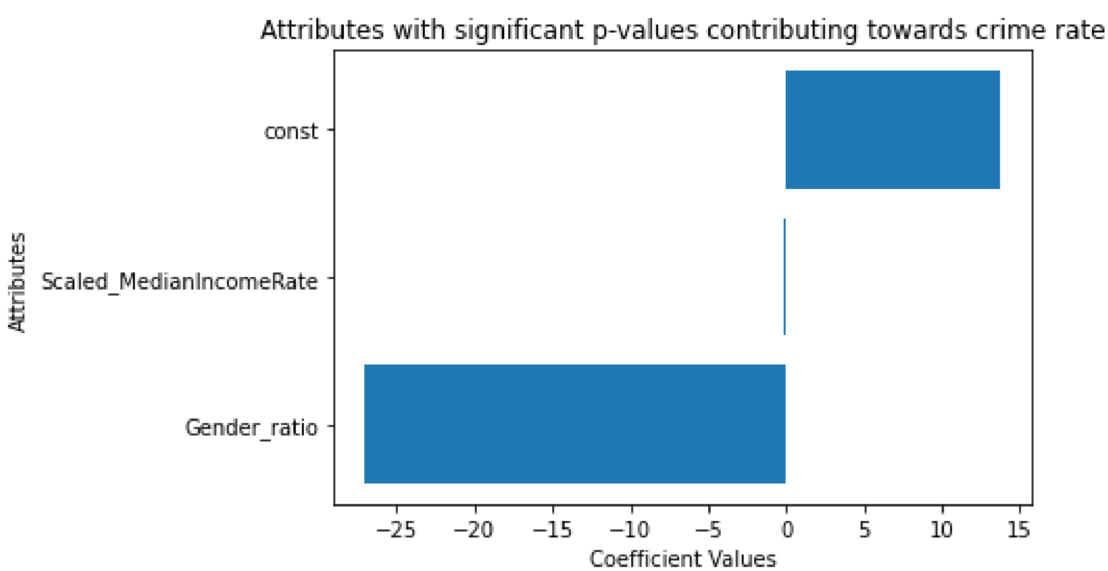
SEATTLE:

Before Covid (Mar-2020)

Ground Truth:
Linear Model summary:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.179
Model:                            OLS   Adj. R-squared:                  0.156
Method:                 Least Squares   F-statistic:                     7.701
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           0.000105
Time:                        16:45:28   Log-Likelihood:                 186.98
No. Observations:                 110   AIC:                            -366.0
Df Residuals:                     106   BIC:                            -355.2
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                              coef    std err          t      P>|t|      [0.025
```

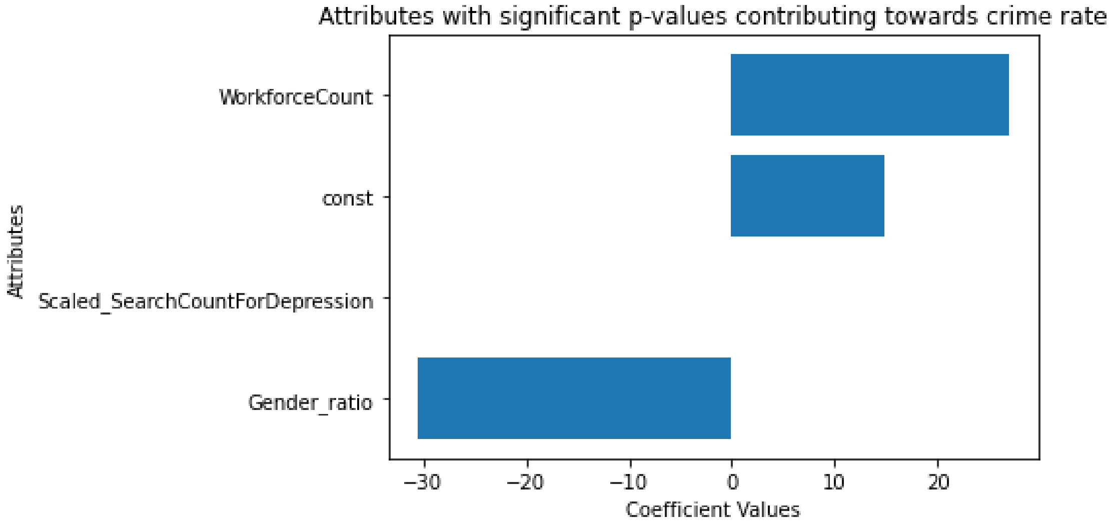
```
0.975]
------------------------------------------------------------------------------
-----------
const                     13.7158      3.497      3.923      0.000      6.783
20.648
Gender_ratio             -27.0115      6.995     -3.862      0.000    -40.879
-13.144
Scaled_Unemployment       -0.0643      0.065     -0.993      0.323     -0.193
0.064
Scaled_MedianIncomeRate   -0.1855      0.041     -4.526      0.000     -0.267
-0.104
==============================================================================
Omnibus:                       15.992   Durbin-Watson:                   1.581
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               20.839
Skew:                          -0.761   Prob(JB):                     2.98e-05
Kurtosis:                       4.494   Cond. No.                     2.14e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.14e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
```

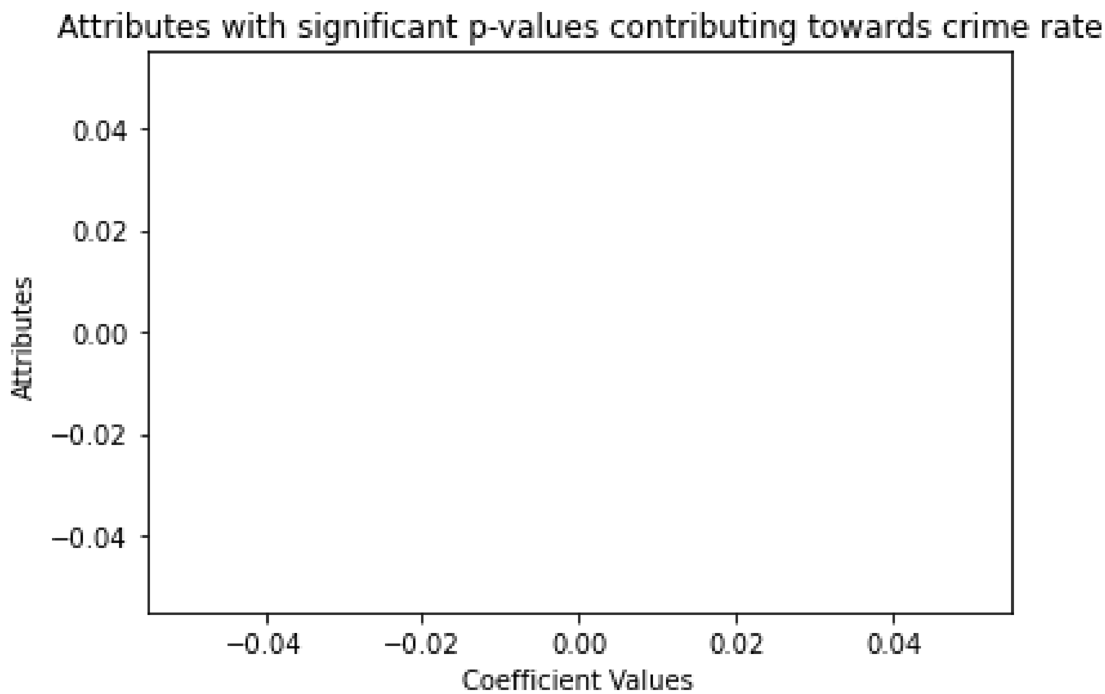
```
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.269
Model:                            OLS   Adj. R-squared:                  0.219
Method:                 Least Squares   F-statistic:                     5.368
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.95e-05
Time:                        16:45:28   Log-Likelihood:                 193.39
No. Observations:                 110   AIC:                            -370.8
Df Residuals:                     102   BIC:                            -349.2
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              14.7662      3.506      4.212      0.000
7.812      21.720
Gender_ratio                      -30.5726      6.887     -4.439      0.000
-44.232     -16.913
Scaled_Unemployment                -0.0263      0.070     -0.375      0.709
-0.165       0.113
Scaled_MedianIncomeRate             0.0180      0.086      0.209      0.835
-0.153       0.189
WorkforceCount                     26.9918     10.923      2.471      0.015
5.326      48.657
PercentNegativeUsers                0.0696      0.131      0.532      0.596
-0.190       0.329
FavorOfDemocrats                   -0.3303      0.580     -0.569      0.571
-1.482       0.821
Scaled_SearchCountForDepression    -0.0544      0.022     -2.430      0.017
-0.099      -0.010
==============================================================================
Omnibus:                       27.653   Durbin-Watson:                   1.926
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               48.804
Skew:                          -1.077   Prob(JB):                     2.53e-11
Kurtosis:                       5.451   Cond. No.                     3.93e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 3.93e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Plot for Linear Model Coefficients for Our Model:




```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.008
Model:                            OLS   Adj. R-squared:                 -0.024
Method:                 Least Squares   F-statistic:                    0.2467
Date:                Sun, 11 Dec 2022   Prob (F-statistic):              0.911
Time:                        16:45:28   Log-Likelihood:                 133.24
No. Observations:                 132   AIC:                            -256.5
Df Residuals:                     127   BIC:                            -242.1
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                     coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                             -0.1487      0.592     -0.251      0.802
-1.319       1.022
WorkforceCount                     1.2868      5.360      0.240      0.811
-9.320      11.894
PercentNegativeUsers              -0.1390      0.231     -0.602      0.548
-0.596       0.318
FavorOfDemocrats                   0.4476      0.625      0.716      0.475
-0.789       1.684
```

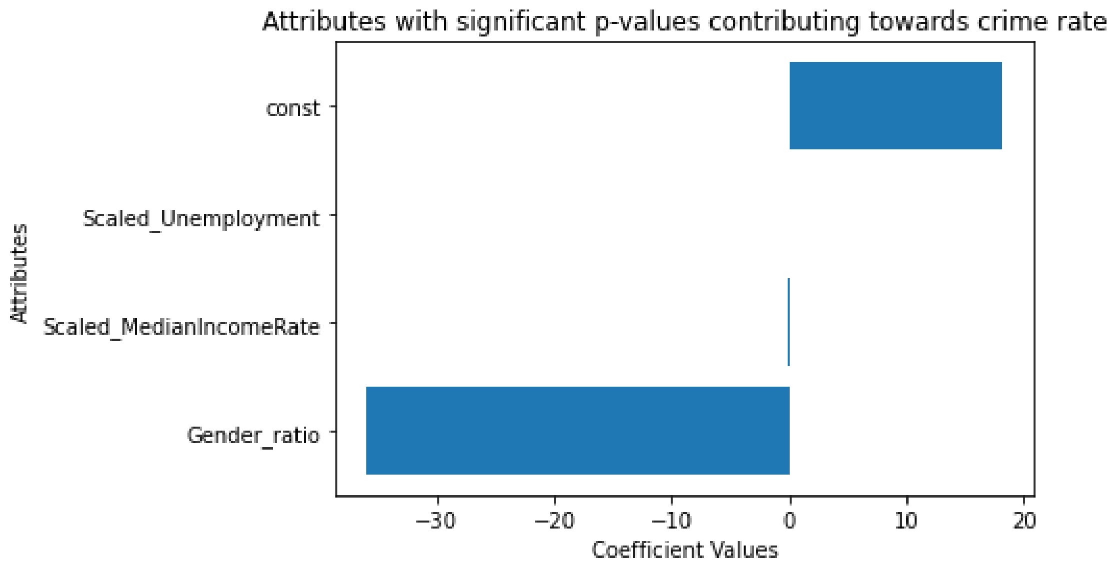
```
Scaled_SearchCountForDepression     0.0149      0.039      0.380      0.704
-0.062      0.092
==============================================================================
Omnibus:                      193.546   Durbin-Watson:                   1.969
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            14902.790
Skew:                           5.667   Prob(JB):                         0.00
Kurtosis:                      53.805   Cond. No.                         935.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Our Model without ground truth:
```




```
Till Dec-2021


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
```

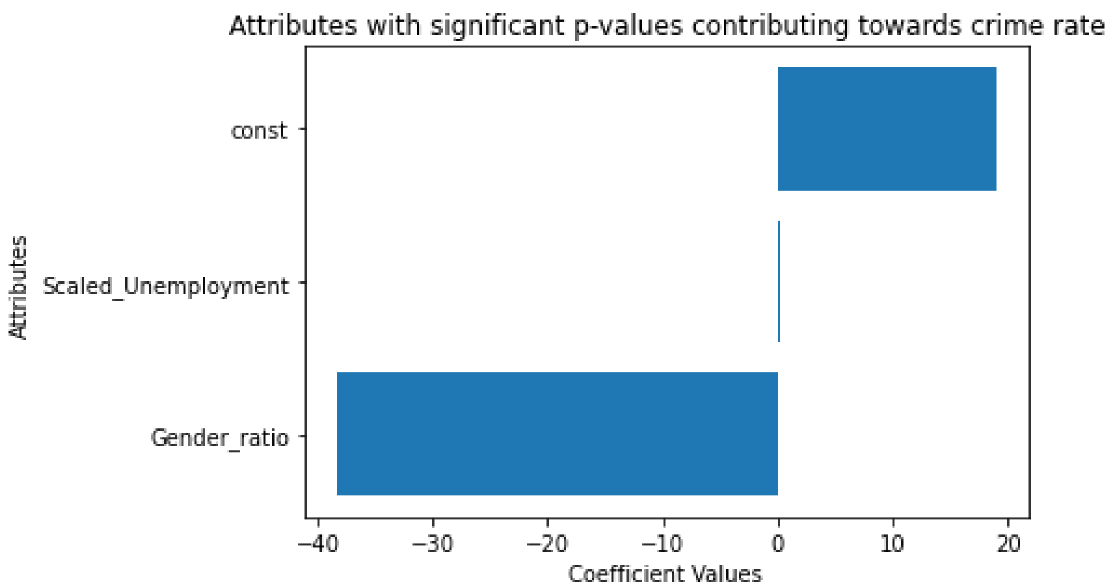
```
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.093
Model:                            OLS   Adj. R-squared:                  0.072
Method:                 Least Squares   F-statistic:                     4.372
Date:                Sun, 11 Dec 2022   Prob (F-statistic):            0.00576
Time:                        16:45:28   Log-Likelihood:                 139.17
No. Observations:                 132   AIC:                            -270.3
Df Residuals:                     128   BIC:                            -258.8
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                              coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                      18.1398      6.188      2.931      0.004       5.895
30.385
Gender_ratio              -35.9424     12.359     -2.908      0.004     -60.397
-11.487
Scaled_Unemployment         0.1281      0.049      2.618      0.010       0.031
0.225
Scaled_MedianIncomeRate    -0.1940      0.076     -2.566      0.011      -0.344
-0.044
==============================================================================
Omnibus:                      173.240   Durbin-Watson:                   2.156
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            10191.230
Skew:                           4.738   Prob(JB):                         0.00
Kurtosis:                      44.990   Cond. No.                     2.26e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.26e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Plot for Linear Model Coefficients for Ground Truth:




```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.119
Model:                            OLS   Adj. R-squared:                  0.070
Method:                 Least Squares   F-statistic:                     2.401
Date:                Sun, 11 Dec 2022   Prob (F-statistic):             0.0244
Time:                        16:45:28   Log-Likelihood:                 141.12
No. Observations:                 132   AIC:                            -266.2
Df Residuals:                     124   BIC:                            -243.2
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              18.9070      6.342      2.981      0.003
6.354      31.460
Gender_ratio                      -38.2771     12.504     -3.061      0.003
-63.026     -13.528
Scaled_Unemployment                 0.1792      0.062      2.883      0.005
0.056       0.302
Scaled_MedianIncomeRate             0.0345      0.169      0.205      0.838
-0.299       0.368
WorkforceCount                     27.5330     21.478      1.282      0.202
```

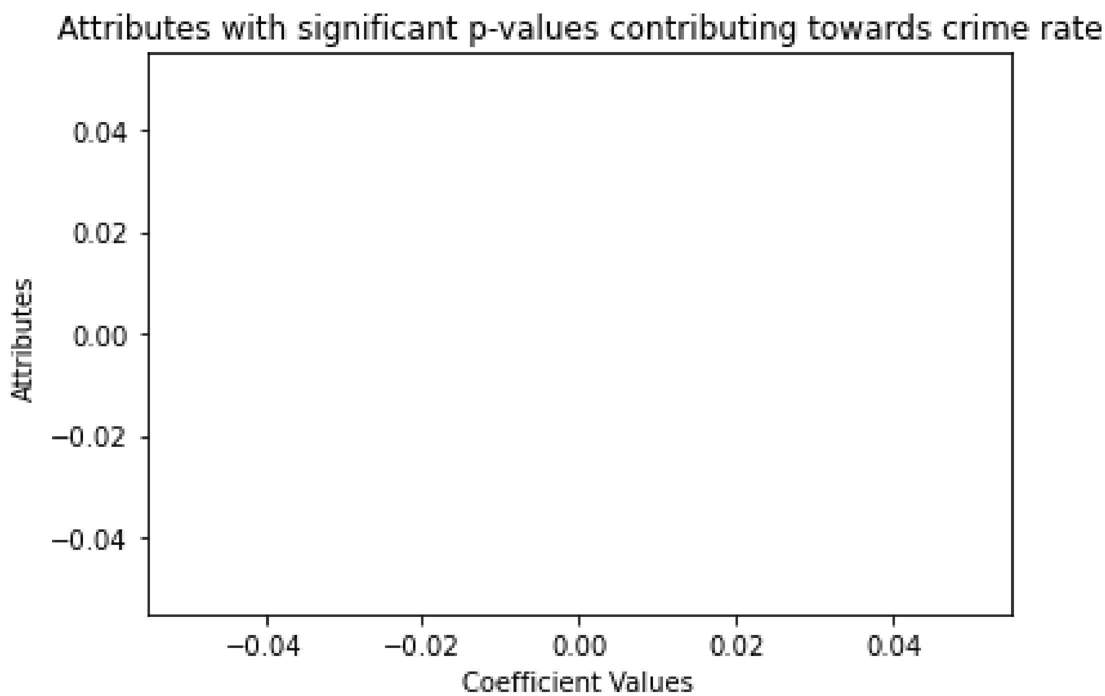
```
-14.978      70.044
PercentNegativeUsers                 -0.0811      0.225     -0.360      0.719
-0.527       0.364
FavorOfDemocrats                     -0.8022      0.750     -1.069      0.287
-2.287       0.683
Scaled_SearchCountForDepression      -0.0156      0.038     -0.410      0.682
-0.091       0.060
==============================================================================
Omnibus:                      170.283   Durbin-Watson:                   2.266
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             9831.650
Skew:                           4.600   Prob(JB):                         0.00
Kurtosis:                      44.266   Cond. No.                     4.38e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 4.38e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```



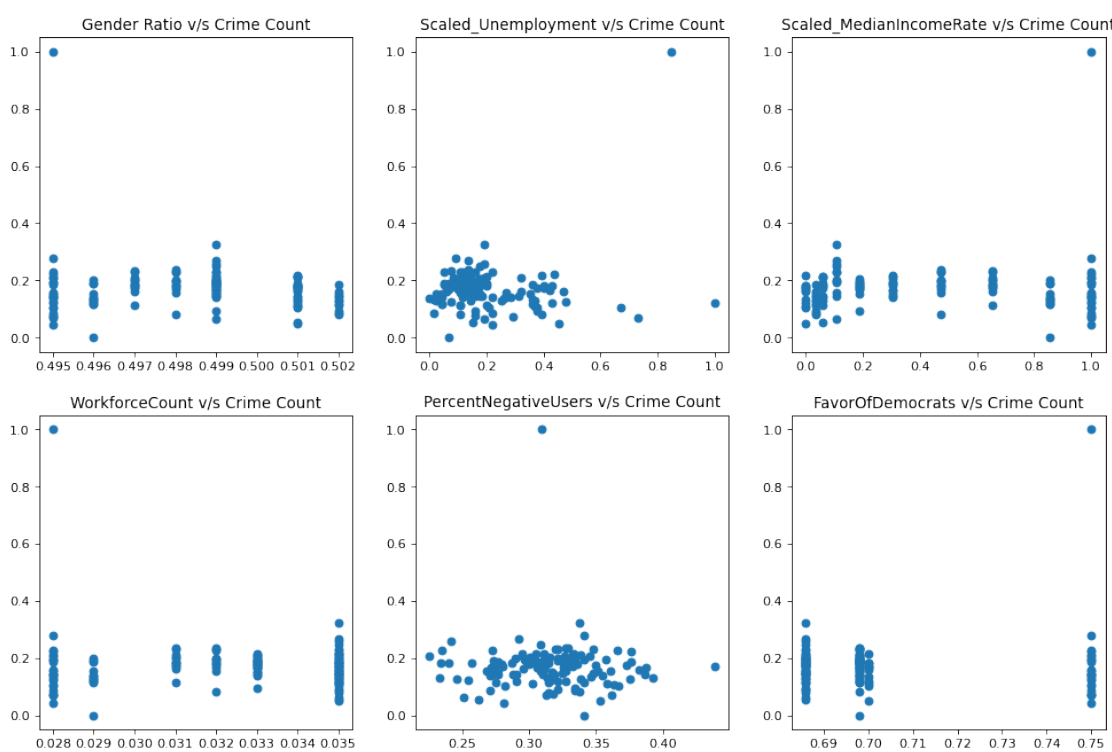

```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
```

```
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.008
Model:                            OLS   Adj. R-squared:                 -0.024
Method:                 Least Squares   F-statistic:                    0.2467
Date:                Sun, 11 Dec 2022   Prob (F-statistic):              0.911
Time:                        16:45:28   Log-Likelihood:                 133.24
No. Observations:                 132   AIC:                            -256.5
Df Residuals:                     127   BIC:                            -242.1
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              -0.1487      0.592     -0.251      0.802
-1.319       1.022
WorkforceCount                      1.2868      5.360      0.240      0.811
-9.320      11.894
PercentNegativeUsers               -0.1390      0.231     -0.602      0.548
-0.596       0.318
FavorOfDemocrats                    0.4476      0.625      0.716      0.475
-0.789       1.684
Scaled_SearchCountForDepression     0.0149      0.039      0.380      0.704
-0.062       0.092
==============================================================================
Omnibus:                      193.546   Durbin-Watson:                   1.969
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            14902.790
Skew:                           5.667   Prob(JB):                         0.00
Kurtosis:                      53.805   Cond. No.                         935.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:

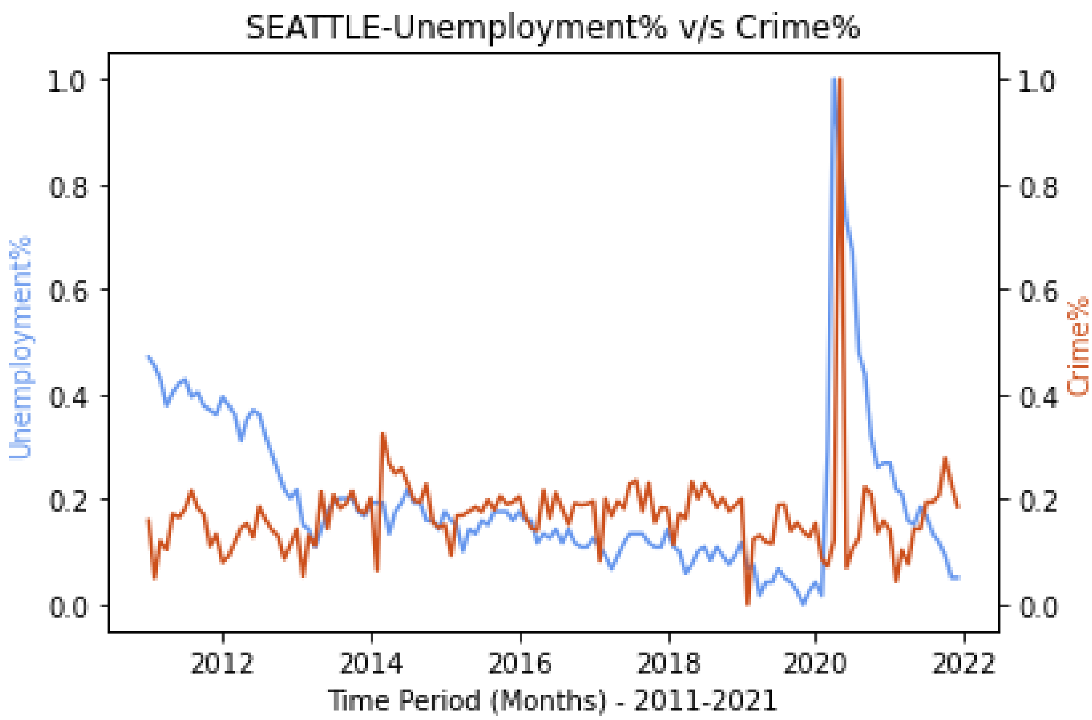


Correlation Graphs:



```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

<Figure size 2400x1200 with 0 Axes>

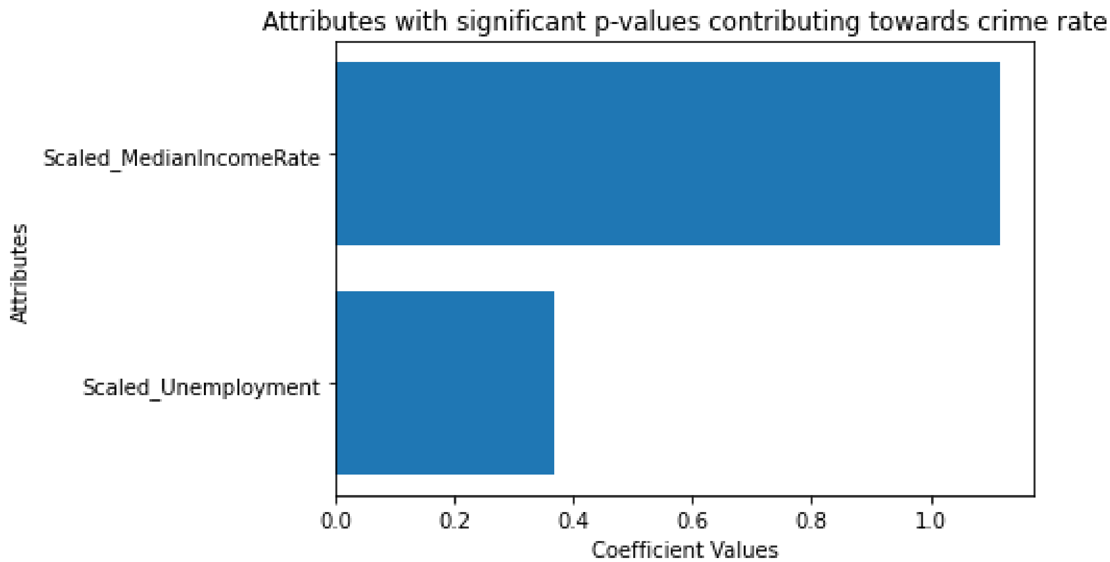


DENVER:

Before Covid (Mar-2020)

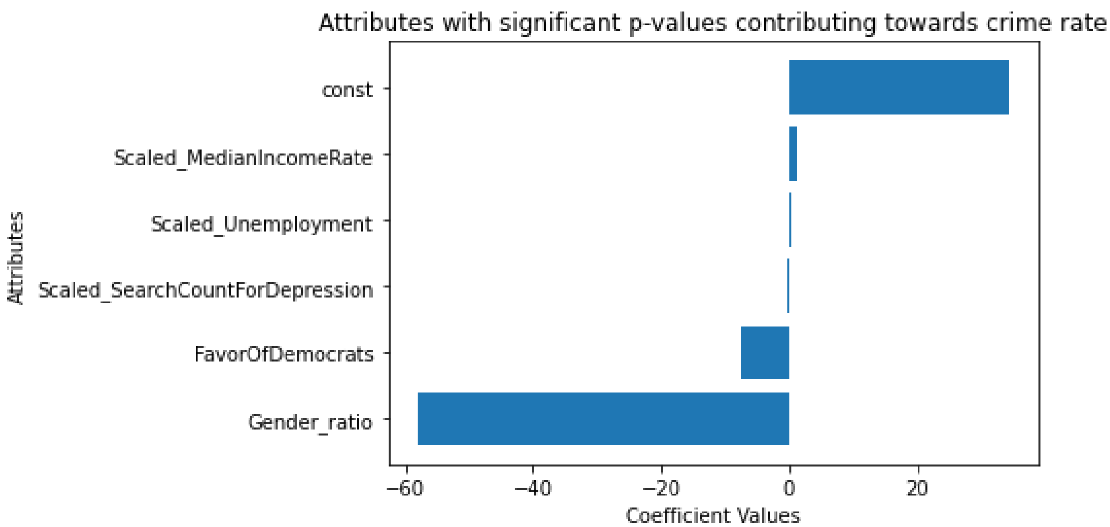
Ground Truth:

```
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.804
Model:                            OLS   Adj. R-squared:                  0.798
Method:                 Least Squares   F-statistic:                     144.8
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.41e-37
Time:                        16:45:29   Log-Likelihood:                 84.786
No. Observations:                 110   AIC:                            -161.6
Df Residuals:                     106   BIC:                            -150.8
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                            coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                    -2.9581     13.854     -0.214      0.831     -30.425
24.509
Gender_ratio              5.5403     27.692      0.200      0.842     -49.361
60.442
Scaled_Unemployment       0.3660      0.103      3.566      0.001       0.163
0.569
Scaled_MedianIncomeRate   1.1174      0.094     11.901      0.000       0.931
1.304
==============================================================================
Omnibus:                       45.501   Durbin-Watson:                   0.392
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              187.698
Skew:                          -1.327   Prob(JB):                     1.75e-41
Kurtosis:                       8.823   Cond. No.                     3.40e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 3.4e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Ground Truth:
```



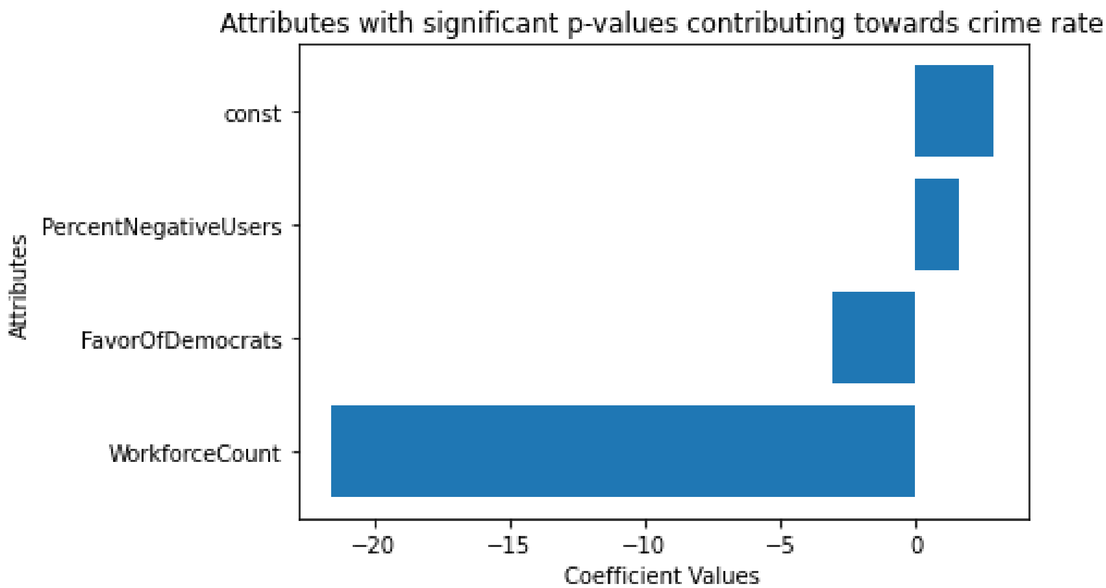

```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.891
Model:                            OLS   Adj. R-squared:                  0.884
Method:                 Least Squares   F-statistic:                     119.6
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.04e-46
Time:                        16:45:29   Log-Likelihood:                 117.32
No. Observations:                 110   AIC:                            -218.6
Df Residuals:                     102   BIC:                            -197.0
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                              coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                      34.4739     12.006      2.871      0.005
10.660      58.288
Gender_ratio              -58.1437     23.386     -2.486      0.015
-104.530     -11.758
Scaled_Unemployment         0.4250      0.088      4.825      0.000
0.250       0.600
Scaled_MedianIncomeRate     1.2283      0.074     16.706      0.000
1.082       1.374
WorkforceCount              3.6717      3.239      1.134      0.260
```

```
-2.752      10.096
PercentNegativeUsers                   -0.1995      0.247     -0.806      0.422
-0.690       0.291
FavorOfDemocrats                       -7.6505      1.039     -7.361      0.000
-9.712      -5.589
Scaled_SearchCountForDepression        -0.2229      0.046     -4.805      0.000
-0.315      -0.131
==============================================================================
Omnibus:                        3.928   Durbin-Watson:                   0.494
Prob(Omnibus):                  0.140   Jarque-Bera (JB):                2.188
Skew:                          -0.021   Prob(JB):                        0.335
Kurtosis:                       2.310   Cond. No.                     4.84e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 4.84e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```



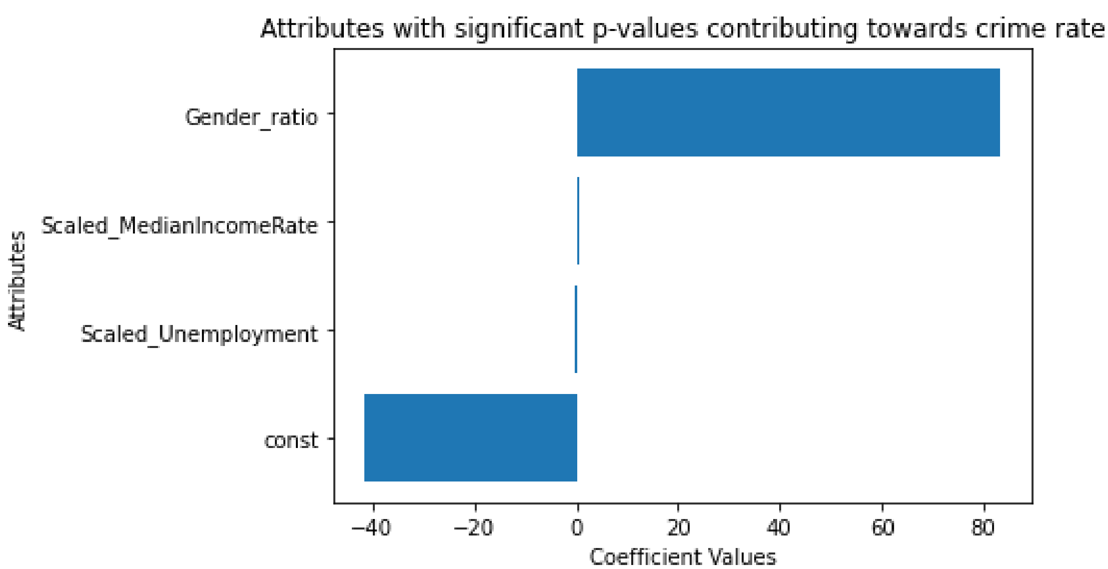

```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.277
Model:                            OLS   Adj. R-squared:                  0.254
```

```
Method:                 Least Squares   F-statistic:                     12.16
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.12e-08
Time:                        16:45:29   Log-Likelihood:                 26.188
No. Observations:                 132   AIC:                            -42.38
Df Residuals:                     127   BIC:                            -27.96
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                     coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              2.9093      0.682      4.268      0.000
1.560       4.258
WorkforceCount                   -21.5788      6.051     -3.566      0.001
-33.553      -9.605
PercentNegativeUsers               1.5894      0.446      3.564      0.001
0.707       2.472
FavorOfDemocrats                  -3.0935      0.843     -3.669      0.000
-4.762      -1.425
Scaled_SearchCountForDepression    0.0762      0.089      0.854      0.395
-0.100       0.253
==============================================================================
Omnibus:                       13.018   Durbin-Watson:                   0.272
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               13.886
Skew:                           0.771   Prob(JB):                     0.000966
Kurtosis:                       3.382   Cond. No.                         472.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Our Model without ground truth:
```



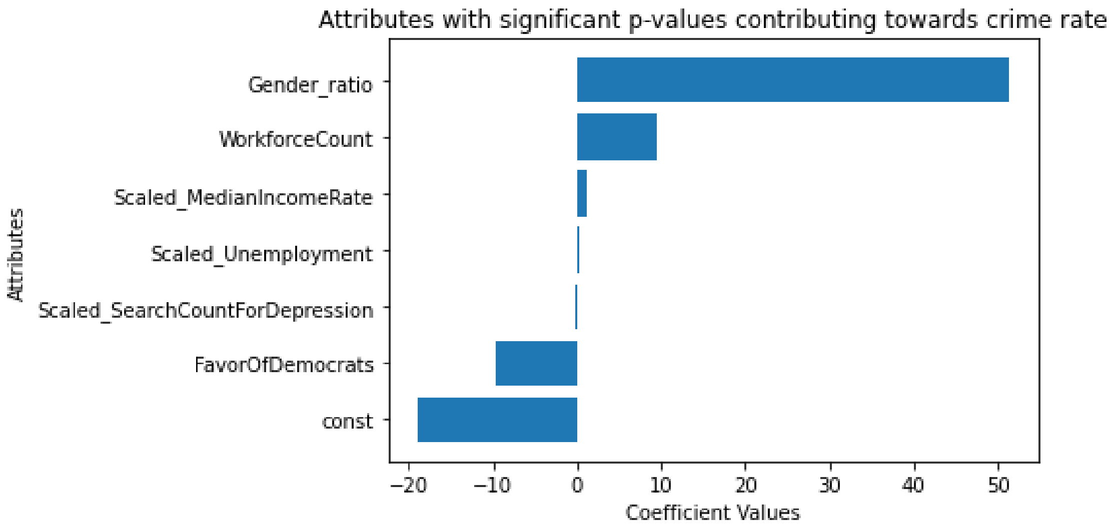

```
Till Dec-2021


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.596
Model:                            OLS   Adj. R-squared:                  0.586
Method:                 Least Squares   F-statistic:                     62.89
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.69e-25
Time:                        16:45:29   Log-Likelihood:                 64.567
No. Observations:                 132   AIC:                            -121.1
Df Residuals:                     128   BIC:                            -109.6
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                             coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                    -41.3537      9.626     -4.296      0.000     -60.401
-22.306
Gender_ratio              83.1072     19.225      4.323      0.000      45.067
121.147
Scaled_Unemployment       -0.3693      0.067     -5.484      0.000      -0.503
```

```
-0.236
Scaled_MedianIncomeRate     0.5486      0.071      7.724      0.000       0.408
0.689
==============================================================================
Omnibus:                        3.288   Durbin-Watson:                   0.248
Prob(Omnibus):                  0.193   Jarque-Bera (JB):                2.970
Skew:                           0.366   Prob(JB):                        0.227
Kurtosis:                       3.072   Cond. No.                     2.02e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.02e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.807
Model:                            OLS   Adj. R-squared:                  0.796
Method:                 Least Squares   F-statistic:                     73.99
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.08e-41
Time:                        16:45:29   Log-Likelihood:                 113.30
```

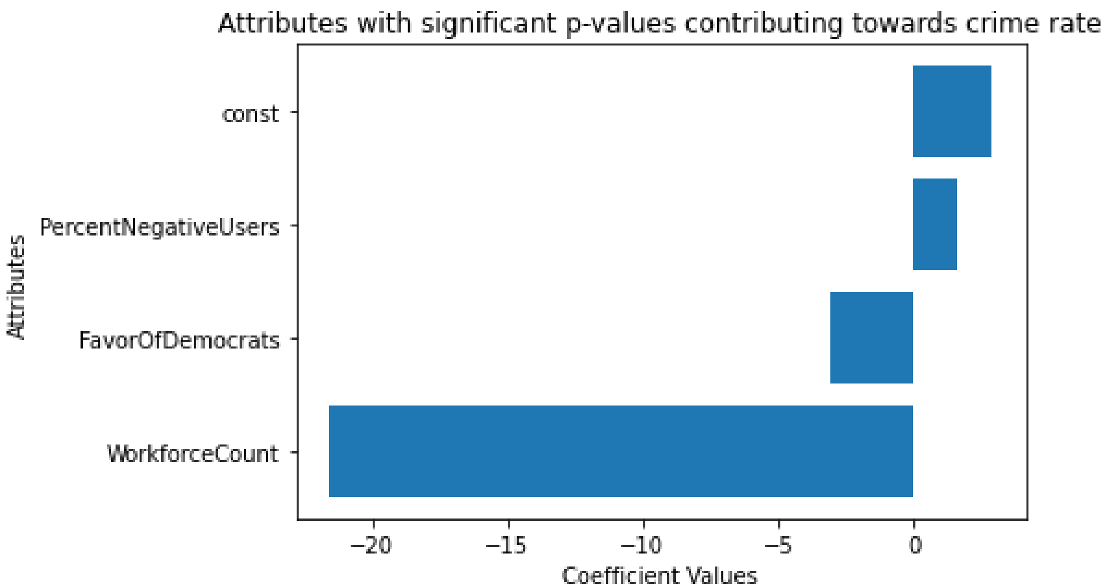
```
No. Observations:                 132   AIC:                            -210.6
Df Residuals:                     124   BIC:                            -187.5
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                    coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                           -18.9525      8.401     -2.256      0.026
-35.580      -2.325
Gender_ratio                     51.3957     16.213      3.170      0.002
19.306      83.485
Scaled_Unemployment               0.1716      0.073      2.357      0.020
0.028       0.316
Scaled_MedianIncomeRate           1.1022      0.074     14.974      0.000
0.957       1.248
WorkforceCount                    9.4968      3.816      2.489      0.014
1.944      17.049
PercentNegativeUsers             -0.1267      0.257     -0.492      0.624
-0.636       0.383
FavorOfDemocrats                 -9.6450      0.900    -10.722      0.000
-11.425      -7.865
Scaled_SearchCountForDepression  -0.1330      0.049     -2.715      0.008
-0.230      -0.036
==============================================================================
Omnibus:                       25.056   Durbin-Watson:                   0.273
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                9.581
Skew:                           0.425   Prob(JB):                      0.00831
Kurtosis:                       1.990   Cond. No.                     3.08e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 3.08e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```



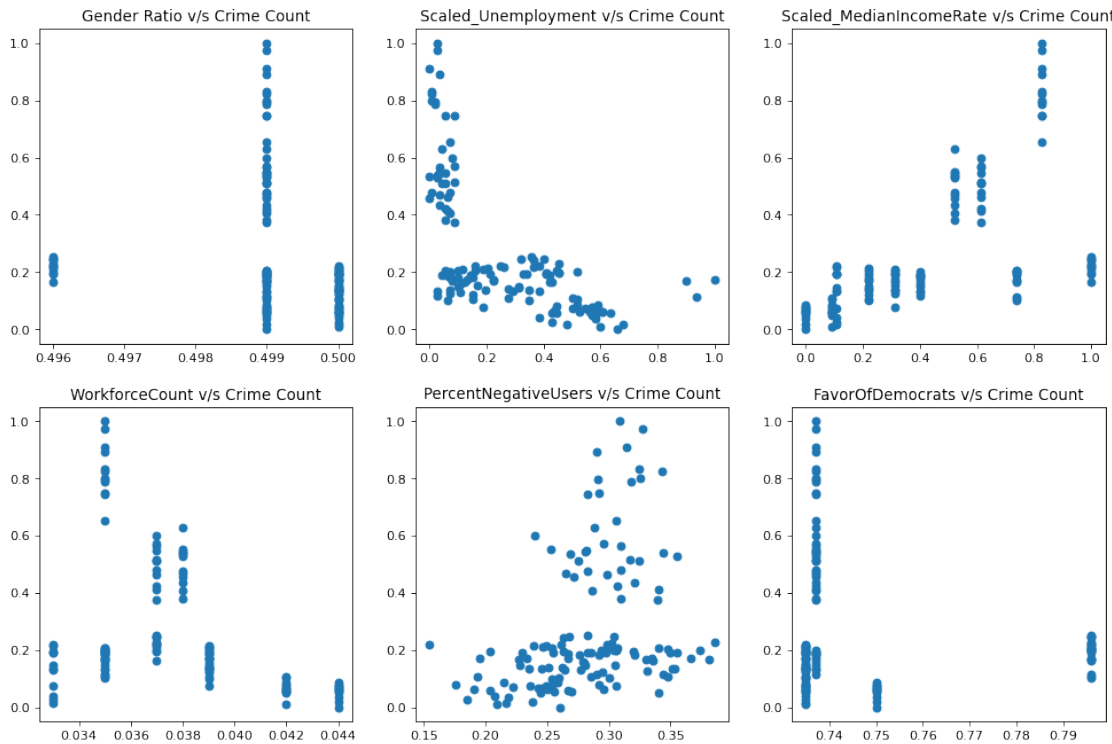

```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.277
Model:                            OLS   Adj. R-squared:                  0.254
Method:                 Least Squares   F-statistic:                     12.16
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.12e-08
Time:                        16:45:29   Log-Likelihood:                 26.188
No. Observations:                 132   AIC:                            -42.38
Df Residuals:                     127   BIC:                            -27.96
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               2.9093      0.682      4.268      0.000
1.560       4.258
WorkforceCount                    -21.5788      6.051     -3.566      0.001
-33.553      -9.605
PercentNegativeUsers                1.5894      0.446      3.564      0.001
0.707       2.472
FavorOfDemocrats                   -3.0935      0.843     -3.669      0.000
-4.762      -1.425
Scaled_SearchCountForDepression     0.0762      0.089      0.854      0.395
-0.100       0.253
```

```
==============================================================================
Omnibus:                       13.018   Durbin-Watson:                   0.272
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               13.886
Skew:                           0.771   Prob(JB):                     0.000966
Kurtosis:                       3.382   Cond. No.                         472.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:

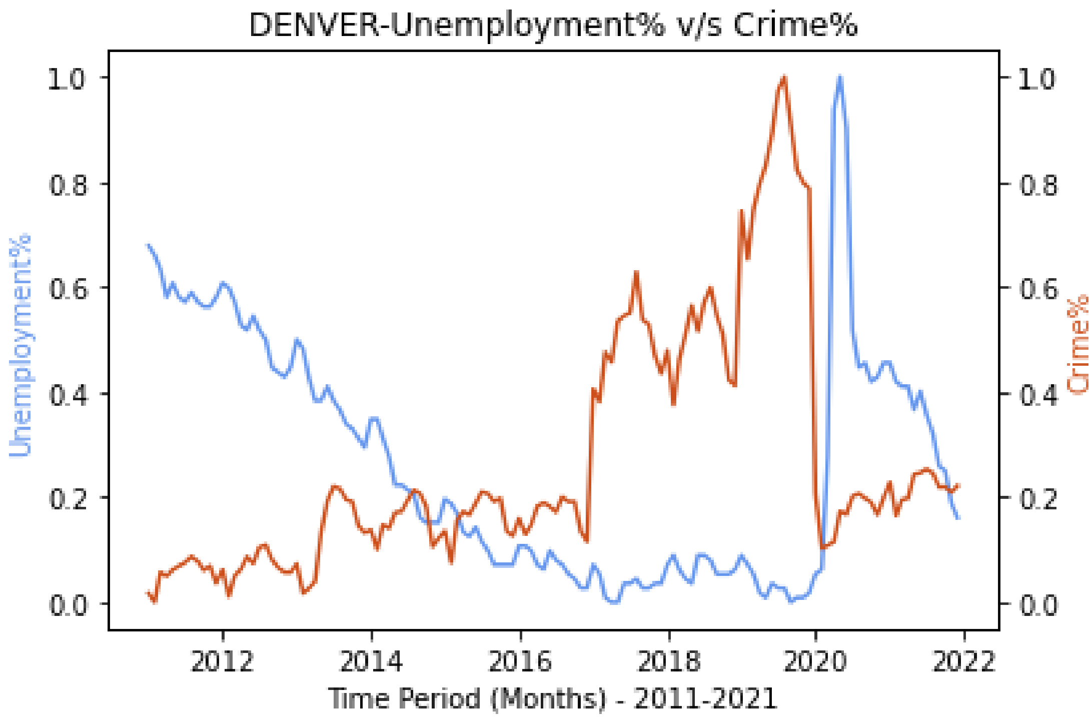


Correlation Graphs:



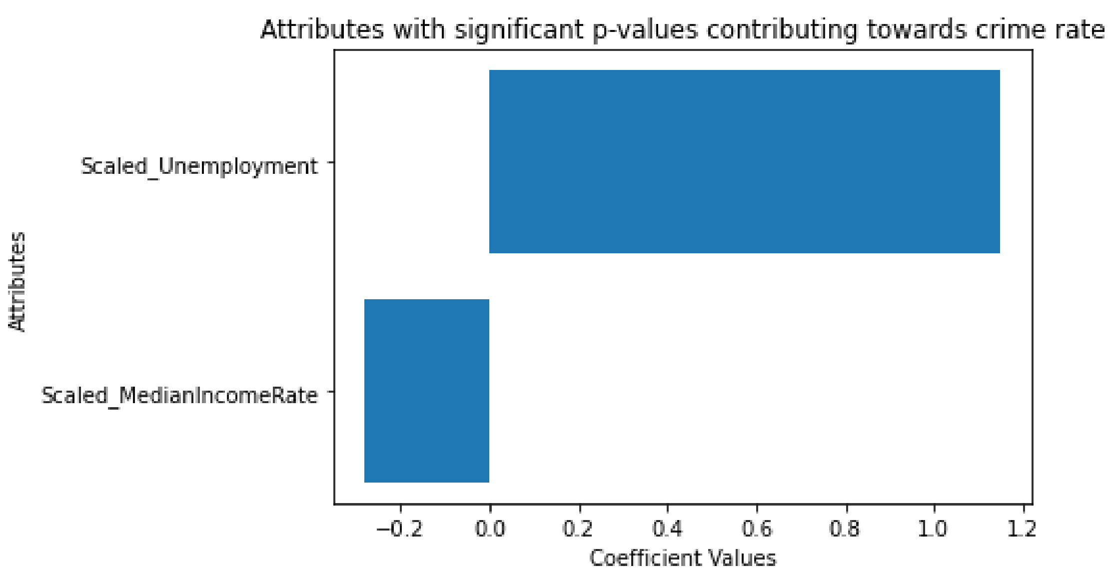

```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

```
<Figure size 2400x1200 with 0 Axes>
```




```
DALLAS:


Before Covid (Mar-2020)


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.527
Model:                            OLS   Adj. R-squared:                  0.505
Method:                 Least Squares   F-statistic:                     24.10
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           1.33e-10
Time:                        16:45:30   Log-Likelihood:                 41.919
No. Observations:                  69   AIC:                            -75.84
Df Residuals:                      65   BIC:                            -66.90
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                              coef    std err          t      P>|t|      [0.025
```

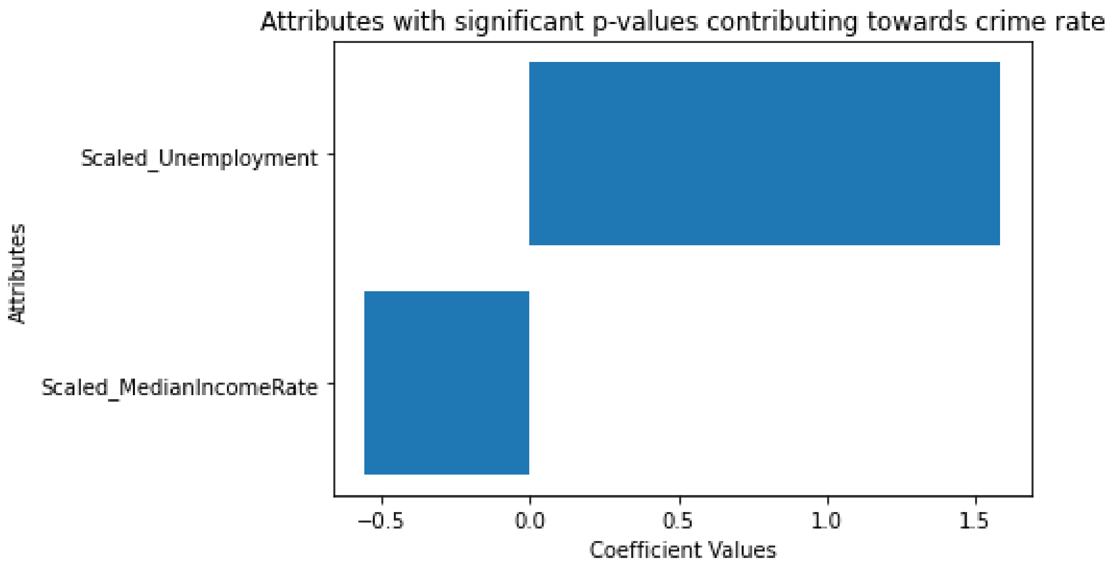
```
0.975]
------------------------------------------------------------------------------
-----------
const                     4.5994      3.188      1.443      0.154      -1.768
10.967
Gender_ratio             -8.1077      6.305     -1.286      0.203     -20.700
4.485
Scaled_Unemployment       1.1476      0.433      2.653      0.010       0.284
2.012
Scaled_MedianIncomeRate  -0.2790      0.093     -2.999      0.004      -0.465
-0.093
==============================================================================
Omnibus:                        8.945   Durbin-Watson:                   0.873
Prob(Omnibus):                  0.011   Jarque-Bera (JB):               10.962
Skew:                           0.544   Prob(JB):                      0.00416
Kurtosis:                       4.622   Cond. No.                         524.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
                            OLS Regression Results
```

```
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.563
Model:                            OLS   Adj. R-squared:                  0.513
Method:                 Least Squares   F-statistic:                     11.24
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.76e-09
Time:                        16:45:30   Log-Likelihood:                 44.706
No. Observations:                  69   AIC:                            -73.41
Df Residuals:                      61   BIC:                            -55.54
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              -5.3942      5.639     -0.957      0.343
-16.671       5.883
Gender_ratio                        6.5236      9.218      0.708      0.482
-11.909      24.956
Scaled_Unemployment                 1.5814      0.512      3.086      0.003
0.557        2.606
Scaled_MedianIncomeRate            -0.5540      0.159     -3.488      0.001
-0.871      -0.236
WorkforceCount                     30.1947     19.269      1.567      0.122
-8.335      68.725
PercentNegativeUsers                0.0567      0.268      0.212      0.833
-0.479       0.592
FavorOfDemocrats                    3.4560      1.939      1.783      0.080
-0.421       7.333
Scaled_SearchCountForDepression     0.1059      0.123      0.859      0.394
-0.140       0.352
==============================================================================
Omnibus:                        2.781   Durbin-Watson:                   0.999
Prob(Omnibus):                  0.249   Jarque-Bera (JB):                2.168
Skew:                           0.197   Prob(JB):                        0.338
Kurtosis:                       3.774   Cond. No.                     1.75e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.75e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

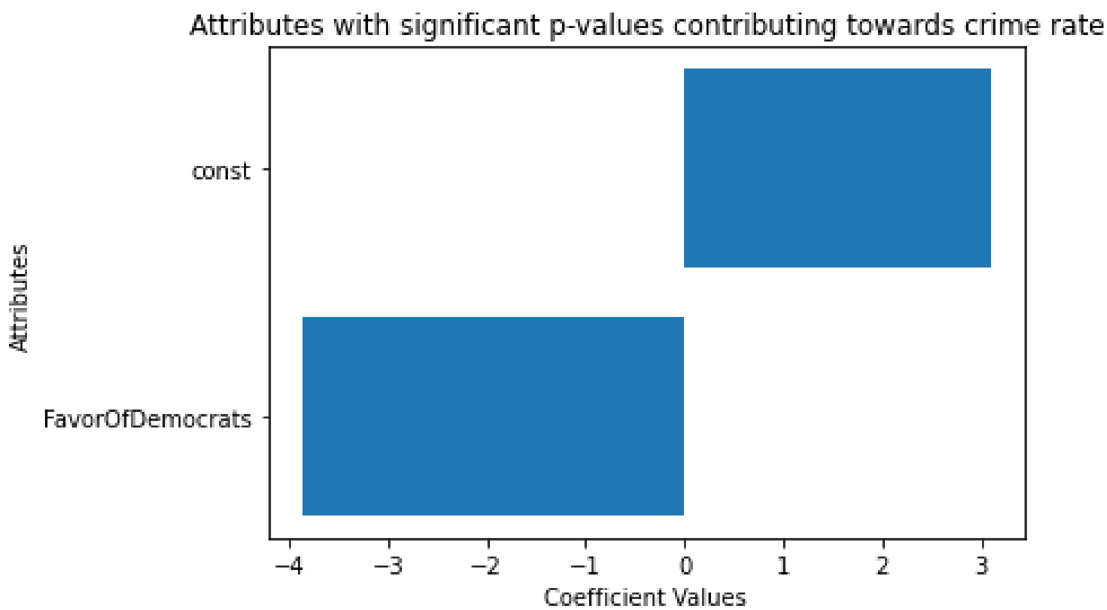
Plot for Linear Model Coefficients for Our Model:




```
Our Model without ground truth:
Linear Model summary:
                             OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.375
Model:                            OLS   Adj. R-squared:                  0.346
Method:                 Least Squares   F-statistic:                     12.91
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.81e-08
Time:                        16:45:30   Log-Likelihood:                 36.864
No. Observations:                  91   AIC:                            -63.73
Df Residuals:                      86   BIC:                            -51.17
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                  coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                           3.0907      0.427      7.240      0.000
2.242       3.939
WorkforceCount                 -3.9987     11.473     -0.349      0.728
-26.807      18.809
PercentNegativeUsers           -0.4985      0.305     -1.633      0.106
-1.105       0.108
FavorOfDemocrats               -3.8625      0.908     -4.252      0.000
-5.668      -2.057
Scaled_SearchCountForDepression   -0.1874      0.098     -1.906      0.060
```

```
-0.383      0.008
==============================================================================
Omnibus:                        9.059   Durbin-Watson:                   0.824
Prob(Omnibus):                  0.011   Jarque-Bera (JB):                9.005
Skew:                           0.638   Prob(JB):                       0.0111
Kurtosis:                       3.863   Cond. No.                         831.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:



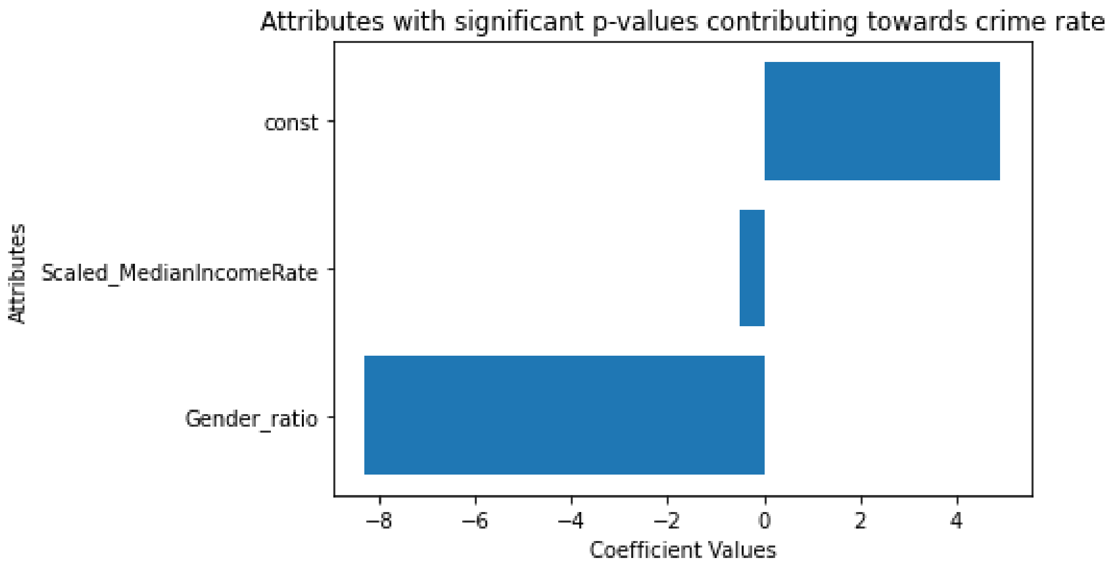
Till Dec-2021

Ground Truth:
Linear Model summary:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.519
Model:                            OLS   Adj. R-squared:                  0.503
Method:                 Least Squares   F-statistic:                     31.32
```

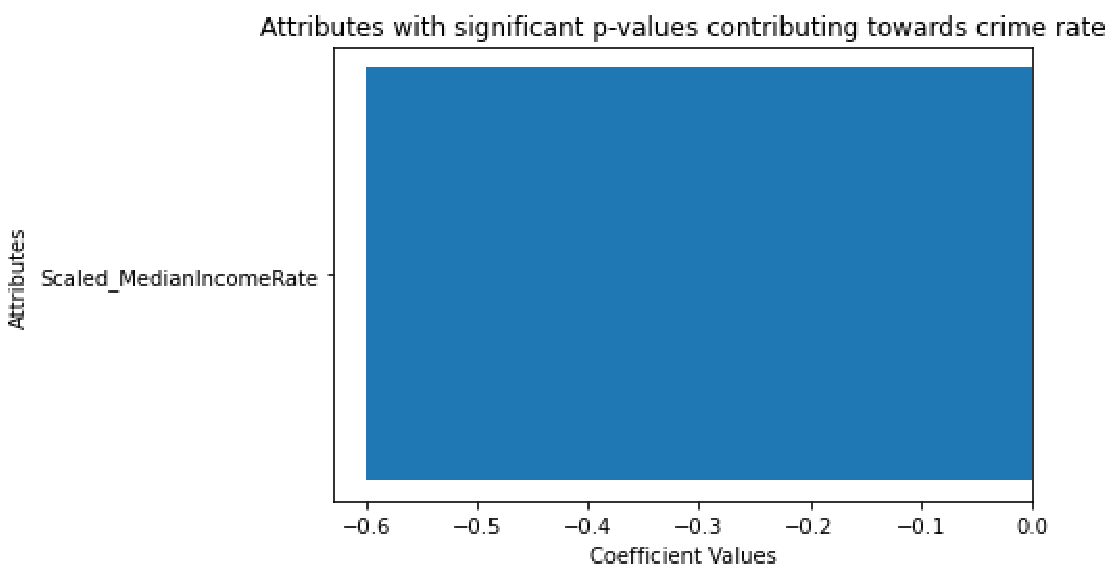
```
Date:                Sun, 11 Dec 2022   Prob (F-statistic):          7.99e-14
Time:                        16:45:30   Log-Likelihood:                48.780
No. Observations:                  91   AIC:                           -89.56
Df Residuals:                      87   BIC:                           -79.52
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                          coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                   4.9098      2.009      2.444      0.017       0.917
8.903
Gender_ratio           -8.2919      3.964     -2.092      0.039     -16.170
-0.413
Scaled_Unemployment    -0.0628      0.083     -0.758      0.451      -0.228
0.102
Scaled_MedianIncomeRate -0.4986     0.054     -9.288      0.000      -0.605
-0.392
==============================================================================
Omnibus:                        0.862   Durbin-Watson:                  0.976
Prob(Omnibus):                  0.650   Jarque-Bera (JB):               0.883
Skew:                           0.224   Prob(JB):                       0.643
Kurtosis:                       2.822   Cond. No.                        372.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Ground Truth:



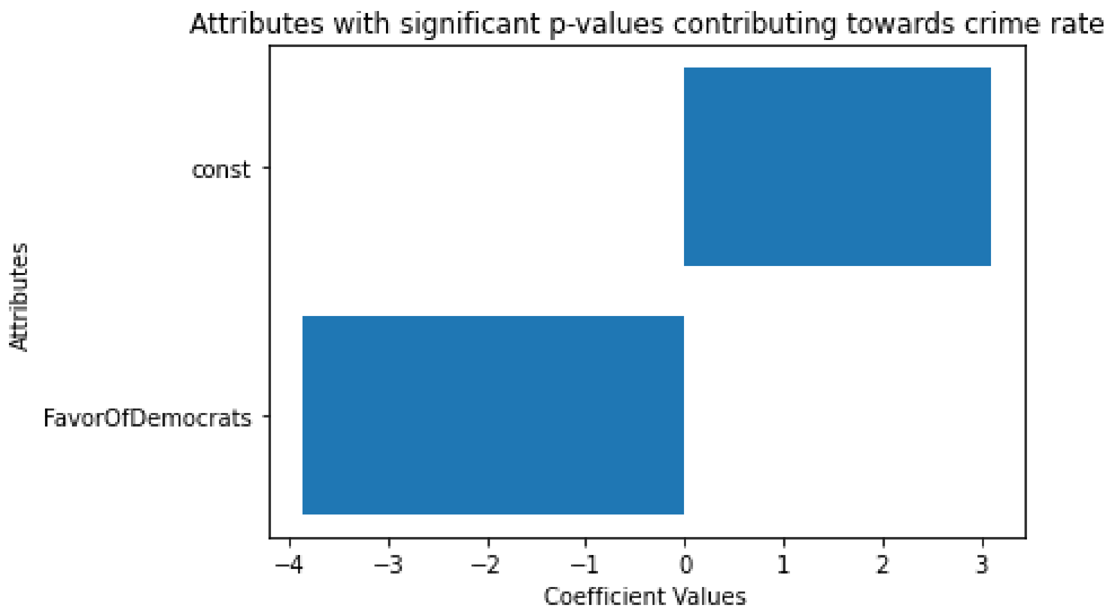

```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.526
Model:                            OLS   Adj. R-squared:                  0.486
Method:                 Least Squares   F-statistic:                     13.17
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.70e-11
Time:                        16:45:30   Log-Likelihood:                 49.448
No. Observations:                  91   AIC:                            -82.90
Df Residuals:                      83   BIC:                            -62.81
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               1.1522      4.783      0.241      0.810
-8.360      10.665
Gender_ratio                       -1.9348      7.844     -0.247      0.806
-17.536      13.667
Scaled_Unemployment                -0.0528      0.098     -0.537      0.593
-0.248       0.143
Scaled_MedianIncomeRate            -0.5992      0.162     -3.701      0.000
-0.921      -0.277
WorkforceCount                     20.0799     18.777      1.069      0.288
```

```
-17.268      57.427
PercentNegativeUsers                -0.0286      0.288     -0.099      0.921
-0.601       0.544
FavorOfDemocrats                     0.3751      1.605      0.234      0.816
-2.817       3.567
Scaled_SearchCountForDepression      0.0088      0.095      0.092      0.927
-0.181       0.198
==============================================================================
Omnibus:                        0.604   Durbin-Watson:                   0.995
Prob(Omnibus):                  0.739   Jarque-Bera (JB):                0.741
Skew:                           0.159   Prob(JB):                        0.690
Kurtosis:                       2.692   Cond. No.                     1.95e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.95e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```

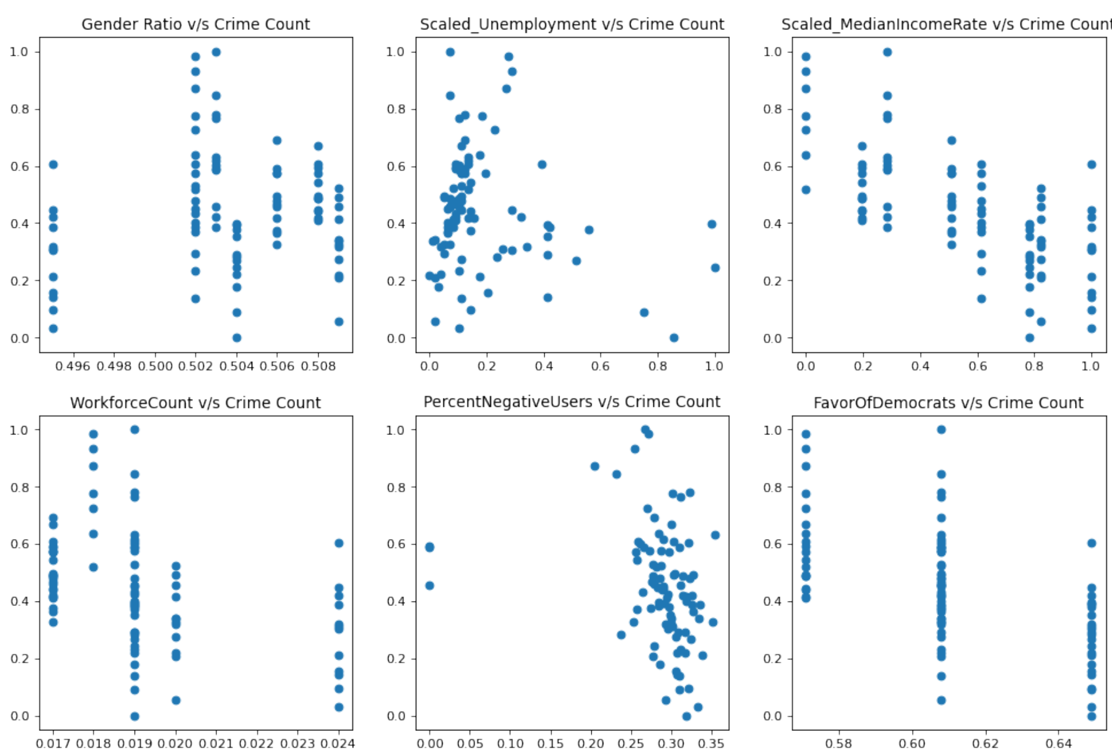



```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.375
```

```
Model:                            OLS   Adj. R-squared:                  0.346
Method:                 Least Squares   F-statistic:                     12.91
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.81e-08
Time:                        16:45:30   Log-Likelihood:                 36.864
No. Observations:                  91   AIC:                            -63.73
Df Residuals:                      86   BIC:                            -51.17
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               3.0907      0.427      7.240      0.000
2.242       3.939
WorkforceCount                     -3.9987     11.473     -0.349      0.728
-26.807      18.809
PercentNegativeUsers               -0.4985      0.305     -1.633      0.106
-1.105       0.108
FavorOfDemocrats                   -3.8625      0.908     -4.252      0.000
-5.668      -2.057
Scaled_SearchCountForDepression    -0.1874      0.098     -1.906      0.060
-0.383       0.008
==============================================================================
Omnibus:                        9.059   Durbin-Watson:                   0.824
Prob(Omnibus):                  0.011   Jarque-Bera (JB):                9.005
Skew:                           0.638   Prob(JB):                       0.0111
Kurtosis:                       3.863   Cond. No.                         831.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Our Model without ground truth:
```

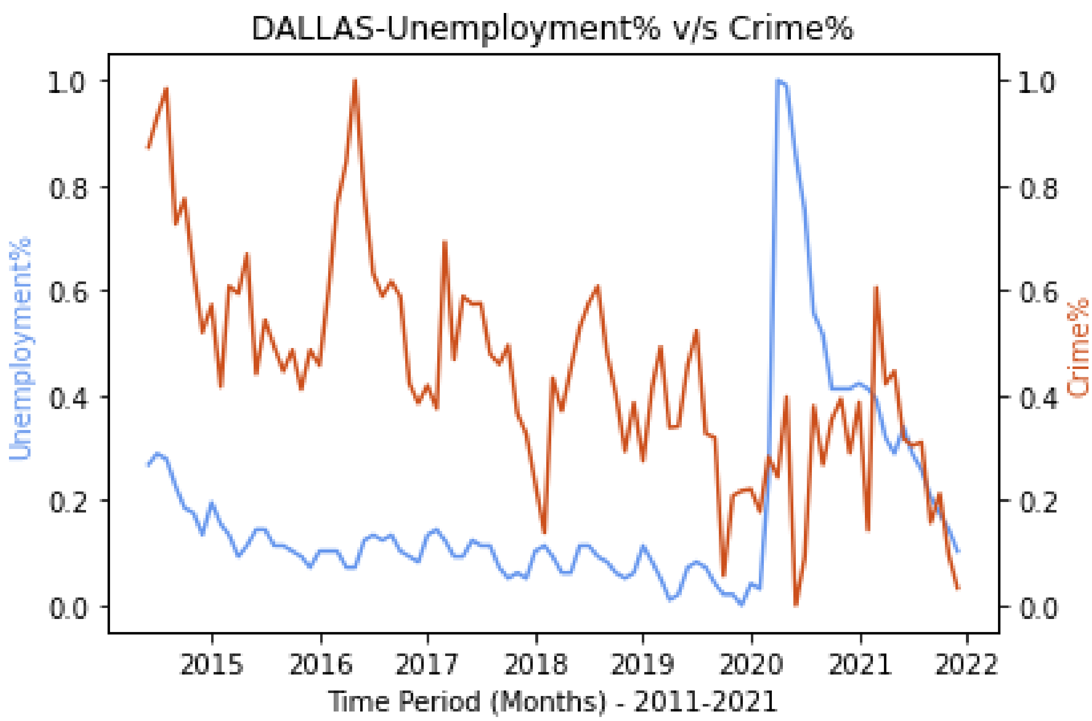


Correlation Graphs:



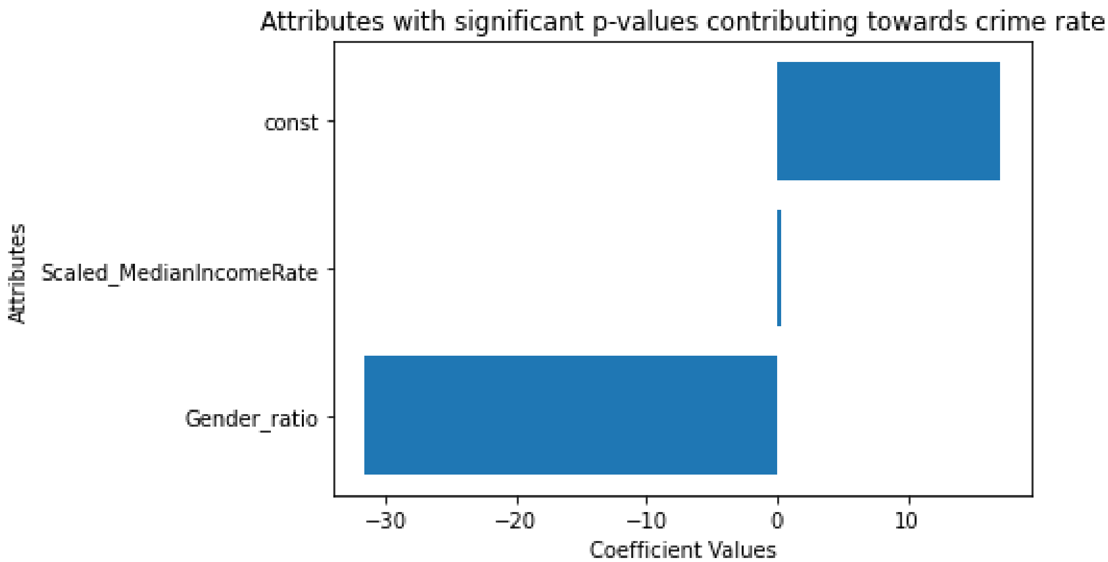
```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

```
<Figure size 2400x1200 with 0 Axes>
```



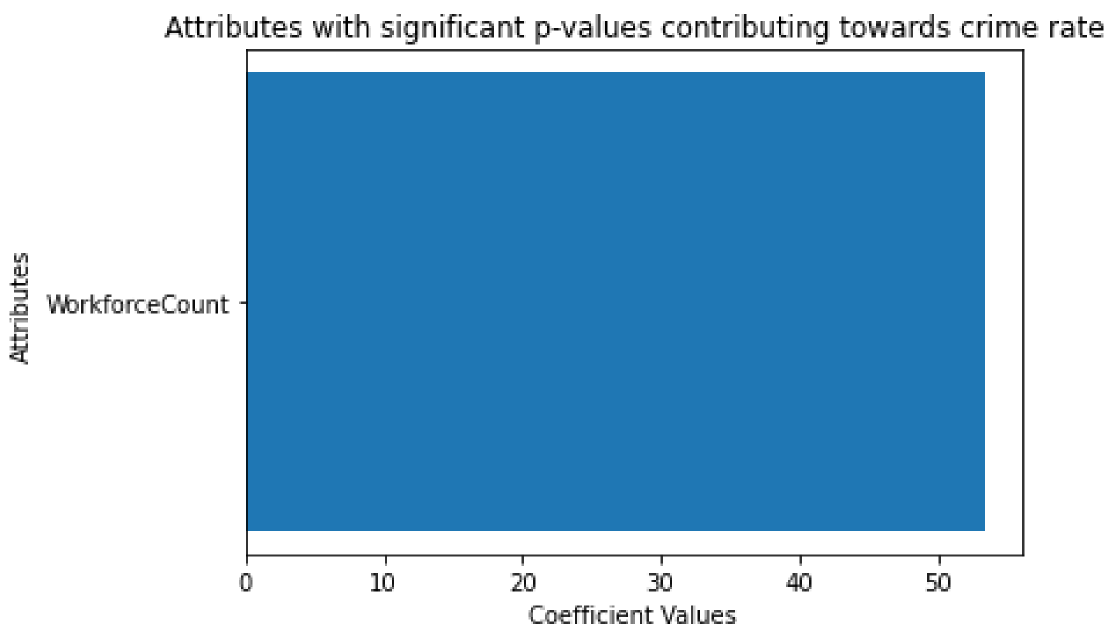

```
NEW_ORLEANS:

Before Covid (Mar-2020)

Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
```

```
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.174
Model:                            OLS   Adj. R-squared:                  0.150
Method:                 Least Squares   F-statistic:                     7.432
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           0.000145
Time:                        16:45:31   Log-Likelihood:                 58.774
No. Observations:                 110   AIC:                            -109.5
Df Residuals:                     106   BIC:                            -98.75
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                            coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                    17.0589      4.420      3.859      0.000       8.296
25.822
Gender_ratio            -31.5360      8.437     -3.738      0.000     -48.262
-14.810
Scaled_Unemployment      -0.2596      0.314     -0.826      0.411      -0.883
0.364
Scaled_MedianIncomeRate   0.3305      0.080      4.118      0.000       0.171
0.490
==============================================================================
Omnibus:                       59.259   Durbin-Watson:                   0.799
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              287.392
Skew:                          -1.756   Prob(JB):                     3.92e-63
Kurtosis:                      10.097   Cond. No.                         802.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```



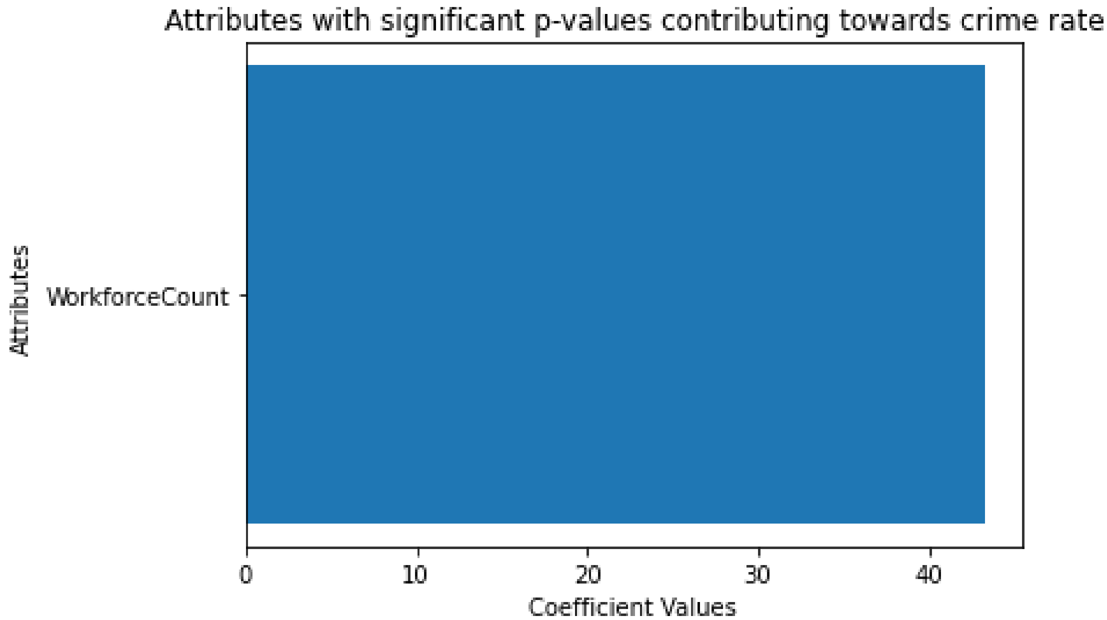

```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.309
Model:                            OLS   Adj. R-squared:                  0.262
Method:                 Least Squares   F-statistic:                     6.526
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           2.29e-06
Time:                        16:45:31   Log-Likelihood:                 68.629
No. Observations:                 110   AIC:                            -121.3
Df Residuals:                     102   BIC:                            -99.65
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                 coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                         -8.4221      7.704     -1.093      0.277
-23.703       6.859
Gender_ratio                  11.3918     12.625      0.902      0.369
-13.649      36.433
Scaled_Unemployment           -0.1678      0.307     -0.546      0.586
-0.777       0.442
Scaled_MedianIncomeRate       -0.0557      0.125     -0.444      0.658
-0.304       0.193
WorkforceCount                53.3968     13.820      3.864      0.000
```

```
25.985      80.809
PercentNegativeUsers                 -0.0741      0.124     -0.597      0.552
-0.320       0.172
FavorOfDemocrats                      0.4838      4.173      0.116      0.908
-7.793       8.761
Scaled_SearchCountForDepression      -0.0944      0.105     -0.903      0.369
-0.302       0.113
==============================================================================
Omnibus:                       63.824   Durbin-Watson:                   1.000
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              381.243
Skew:                          -1.827   Prob(JB):                     1.64e-83
Kurtosis:                      11.356   Cond. No.                     2.14e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.14e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```



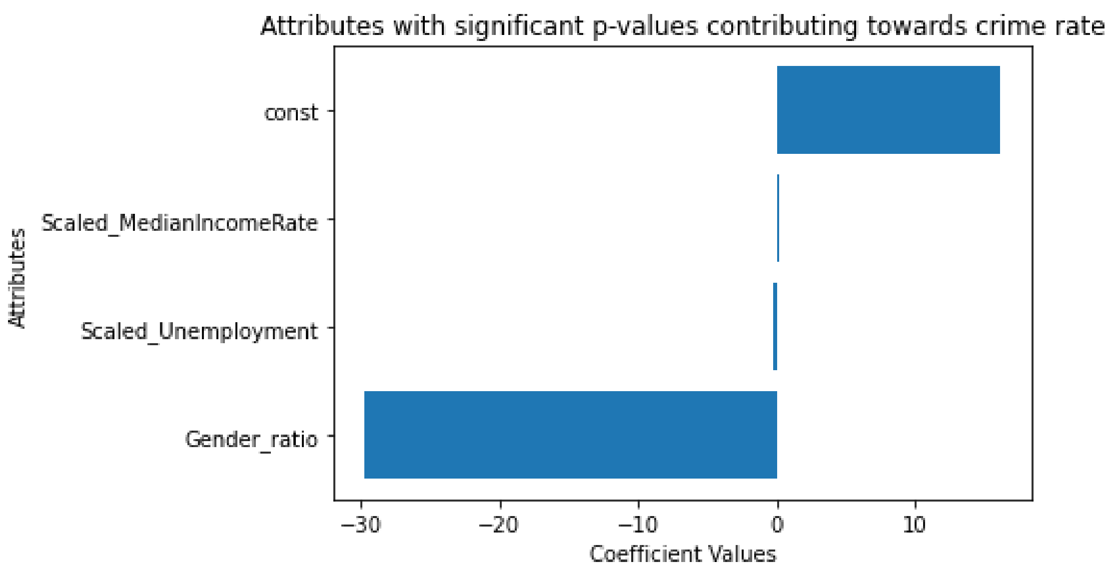

```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
```

```
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.346
Model:                            OLS   Adj. R-squared:                  0.325
Method:                 Least Squares   F-statistic:                     16.79
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.54e-11
Time:                        16:45:31   Log-Likelihood:                 82.423
No. Observations:                 132   AIC:                            -154.8
Df Residuals:                     127   BIC:                            -140.4
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                     coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                             -1.1539      1.501     -0.769      0.443
-4.123       1.816
WorkforceCount                    43.2886      7.716      5.611      0.000
28.021      58.556
PercentNegativeUsers              -0.0035      0.115     -0.030      0.976
-0.231       0.224
FavorOfDemocrats                  -0.6496      1.492     -0.435      0.664
-3.602       2.303
Scaled_SearchCountForDepression    0.0483      0.077      0.630      0.530
-0.103       0.200
==============================================================================
Omnibus:                       49.530   Durbin-Watson:                   1.008
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              190.272
Skew:                          -1.297   Prob(JB):                     4.82e-42
Kurtosis:                       8.279   Cond. No.                         944.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Our Model without ground truth:
```



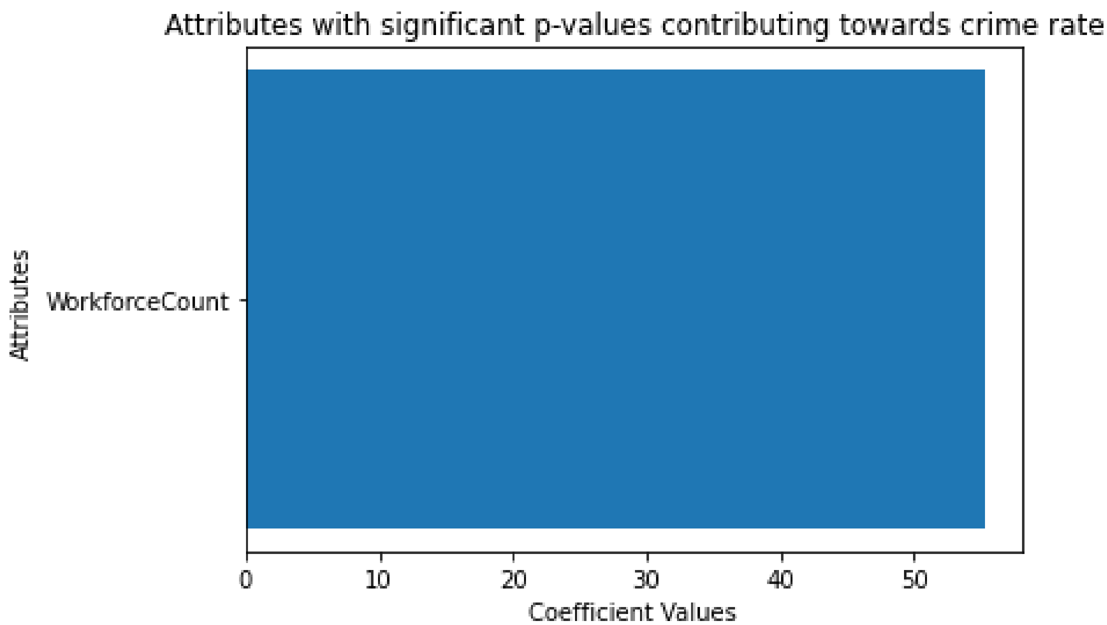

```
Till Dec-2021


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.190
Model:                            OLS   Adj. R-squared:                  0.171
Method:                 Least Squares   F-statistic:                     10.01
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           5.68e-06
Time:                        16:45:31   Log-Likelihood:                 68.319
No. Observations:                 132   AIC:                            -128.6
Df Residuals:                     128   BIC:                            -117.1
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                             coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                     16.1179      3.407      4.731      0.000       9.377
22.858
Gender_ratio             -29.6911      6.558     -4.528      0.000     -42.667
-16.716
```

```
Scaled_Unemployment        -0.2674      0.084     -3.171      0.002      -0.434
-0.101
Scaled_MedianIncomeRate     0.2052      0.069      2.953      0.004       0.068
0.343
==============================================================================
Omnibus:                       33.459   Durbin-Watson:                   0.765
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               83.602
Skew:                          -0.984   Prob(JB):                     7.02e-19
Kurtosis:                       6.366   Cond. No.                         699.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.367
Model:                            OLS   Adj. R-squared:                  0.331
Method:                 Least Squares   F-statistic:                     10.26
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.28e-10
Time:                        16:45:31   Log-Likelihood:                 84.559
No. Observations:                 132   AIC:                            -153.1
```

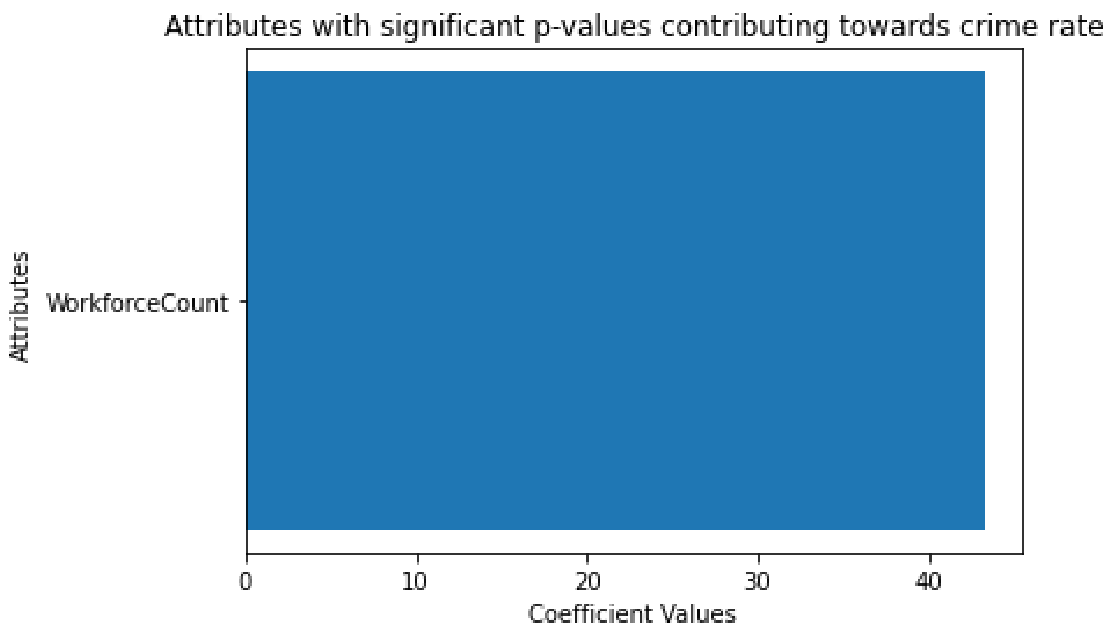
```
Df Residuals:                     124   BIC:                            -130.1
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                   coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                          -10.1459      6.825     -1.487      0.140
-23.655       3.363
Gender_ratio                    16.1323     10.553      1.529      0.129
-4.755      37.020
Scaled_Unemployment             -0.0304      0.089     -0.343      0.732
-0.206       0.145
Scaled_MedianIncomeRate         -0.0911      0.114     -0.795      0.428
-0.318       0.136
WorkforceCount                  55.2672     13.302      4.155      0.000
28.940      81.595
PercentNegativeUsers            -0.0470      0.117     -0.403      0.688
-0.278       0.184
FavorOfDemocrats                -0.6228      3.539     -0.176      0.861
-7.627       6.381
Scaled_SearchCountForDepression    -0.0381      0.090     -0.422      0.674
-0.217       0.141
==============================================================================
Omnibus:                       59.073   Durbin-Watson:                   1.035
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              291.185
Skew:                          -1.484   Prob(JB):                     5.89e-64
Kurtosis:                       9.643   Cond. No.                     2.25e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.25e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


Plot for Linear Model Coefficients for Our Model:
```



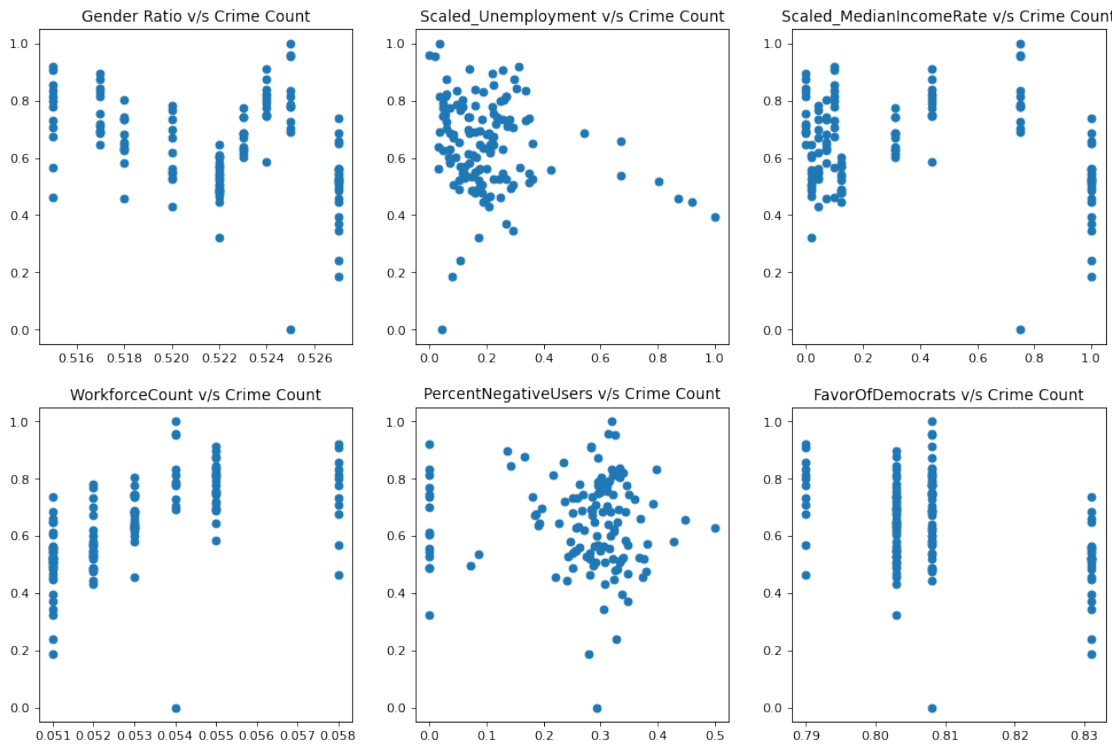

```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.346
Model:                            OLS   Adj. R-squared:                  0.325
Method:                 Least Squares   F-statistic:                     16.79
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.54e-11
Time:                        16:45:31   Log-Likelihood:                 82.423
No. Observations:                 132   AIC:                            -154.8
Df Residuals:                     127   BIC:                            -140.4
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              -1.1539      1.501     -0.769      0.443
-4.123       1.816
WorkforceCount                     43.2886      7.716      5.611      0.000
28.021      58.556
PercentNegativeUsers               -0.0035      0.115     -0.030      0.976
-0.231       0.224
FavorOfDemocrats                   -0.6496      1.492     -0.435      0.664
-3.602       2.303
```

```
Scaled_SearchCountForDepression     0.0483      0.077      0.630      0.530
-0.103       0.200
==============================================================================
Omnibus:                       49.530   Durbin-Watson:                   1.008
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              190.272
Skew:                          -1.297   Prob(JB):                     4.82e-42
Kurtosis:                       8.279   Cond. No.                         944.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:

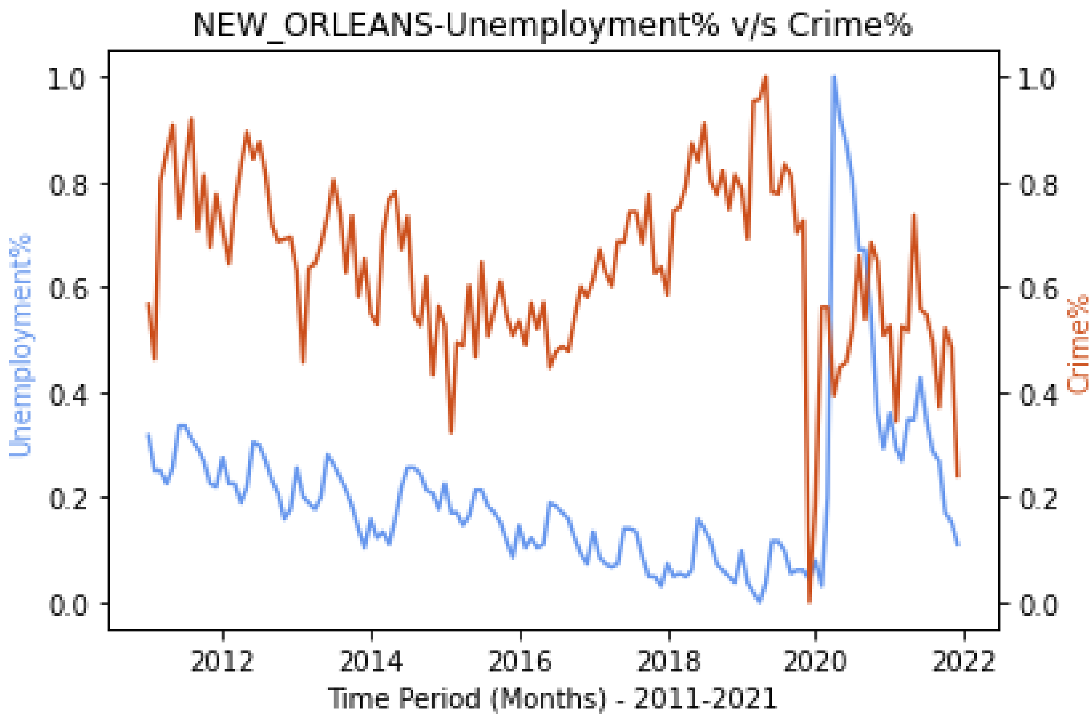


Correlation Graphs:



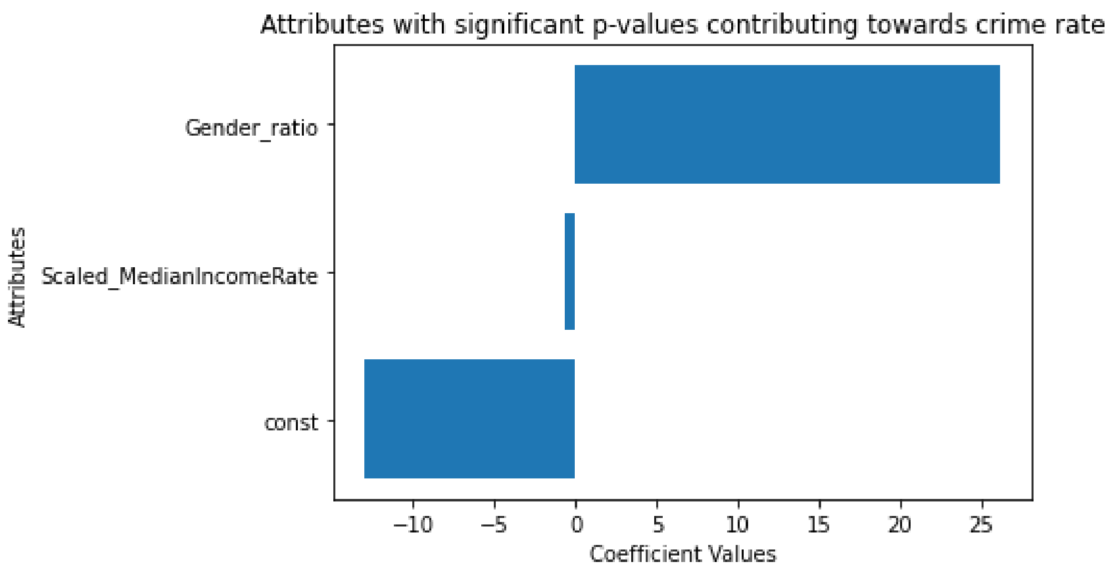

```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

```
<Figure size 2400x1200 with 0 Axes>
```



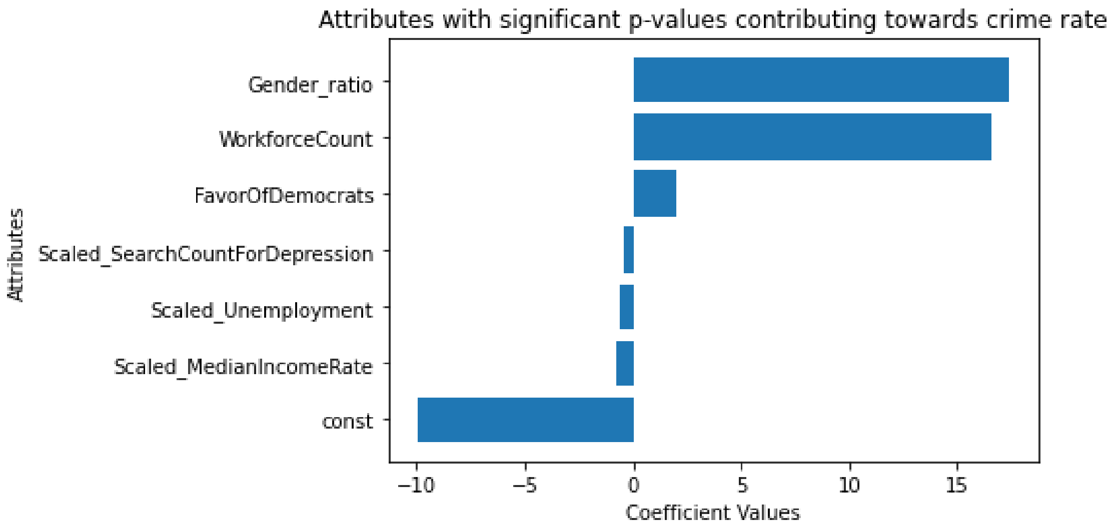

```
INDIANAPOLIS:


Before Covid (Mar-2020)


Ground Truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.566
Model:                            OLS   Adj. R-squared:                  0.554
Method:                 Least Squares   F-statistic:                     46.13
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.71e-19
Time:                        16:45:32   Log-Likelihood:                 80.659
No. Observations:                 110   AIC:                            -153.3
Df Residuals:                     106   BIC:                            -142.5
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                              coef    std err          t      P>|t|      [0.025
```

```
0.975]
--------------------------------------------------------------------------------
-----------
const                    -12.9353      2.881     -4.490      0.000     -18.647
-7.224
Gender_ratio              26.1183      5.583      4.678      0.000      15.050
37.186
Scaled_Unemployment       -0.0185      0.111     -0.166      0.868      -0.239
0.202
Scaled_MedianIncomeRate   -0.6882      0.124     -5.550      0.000      -0.934
-0.442
==============================================================================
Omnibus:                        3.462   Durbin-Watson:                   0.711
Prob(Omnibus):                  0.177   Jarque-Bera (JB):                3.221
Skew:                          -0.211   Prob(JB):                        0.200
Kurtosis:                       3.725   Cond. No.                         668.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Ground Truth:
```




```
Our Model:
Linear Model summary:
                            OLS Regression Results
```

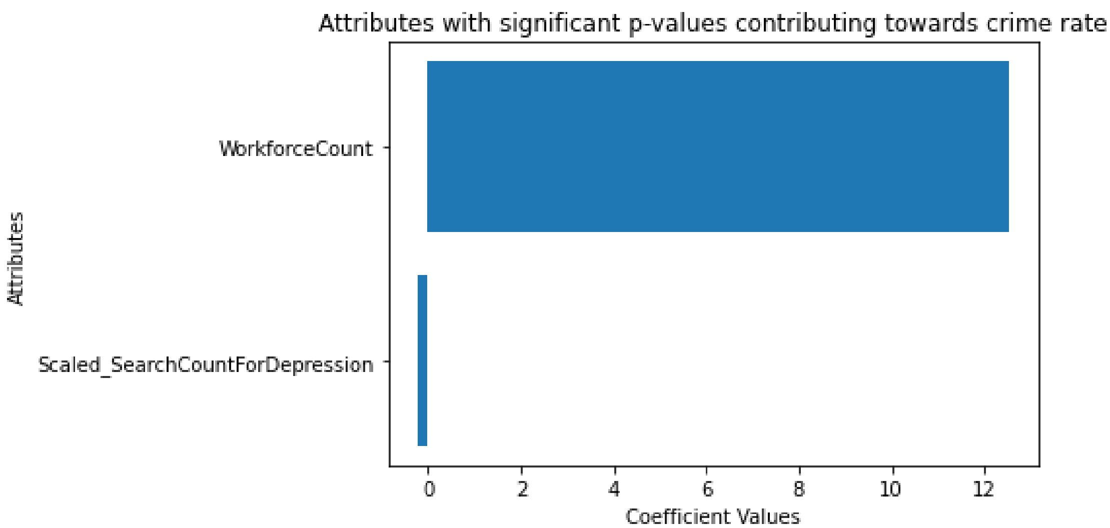
```
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.763
Model:                            OLS   Adj. R-squared:                  0.747
Method:                 Least Squares   F-statistic:                     46.87
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           4.20e-29
Time:                        16:45:32   Log-Likelihood:                 113.86
No. Observations:                 110   AIC:                            -211.7
Df Residuals:                     102   BIC:                            -190.1
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              -9.9486      2.314     -4.299      0.000
-14.538      -5.359
Gender_ratio                       17.3994      4.714      3.691      0.000
8.050      26.749
Scaled_Unemployment                -0.5442      0.119     -4.584      0.000
-0.780      -0.309
Scaled_MedianIncomeRate            -0.7472      0.101     -7.368      0.000
-0.948      -0.546
WorkforceCount                     16.5675      4.442      3.730      0.000
7.757      25.378
PercentNegativeUsers                0.1508      0.130      1.157      0.250
-0.108      0.409
FavorOfDemocrats                    2.0264      0.746      2.718      0.008
0.547      3.505
Scaled_SearchCountForDepression    -0.3757      0.048     -7.860      0.000
-0.471      -0.281
==============================================================================
Omnibus:                        2.641   Durbin-Watson:                   1.230
Prob(Omnibus):                  0.267   Jarque-Bera (JB):                2.280
Skew:                          -0.149   Prob(JB):                        0.320
Kurtosis:                       3.639   Cond. No.                         975.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


Plot for Linear Model Coefficients for Our Model:
```




```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.160
Model:                            OLS   Adj. R-squared:                  0.133
Method:                 Least Squares   F-statistic:                     6.042
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           0.000175
Time:                        16:45:32   Log-Likelihood:                 45.280
No. Observations:                 132   AIC:                            -80.56
Df Residuals:                     127   BIC:                            -66.15
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               0.7992      0.456      1.752      0.082
-0.104       1.702
WorkforceCount                     12.5342      5.315      2.358      0.020
2.017      23.051
PercentNegativeUsers               -0.4293      0.237     -1.813      0.072
-0.898       0.039
FavorOfDemocrats                   -1.1609      0.876     -1.325      0.188
-2.895       0.573
Scaled_SearchCountForDepression    -0.2289      0.087     -2.637      0.009
-0.401      -0.057
```

```
==============================================================================
Omnibus:                        5.261   Durbin-Watson:                   0.360
Prob(Omnibus):                  0.072   Jarque-Bera (JB):                4.935
Skew:                           0.358   Prob(JB):                       0.0848
Kurtosis:                       3.621   Cond. No.                         457.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:



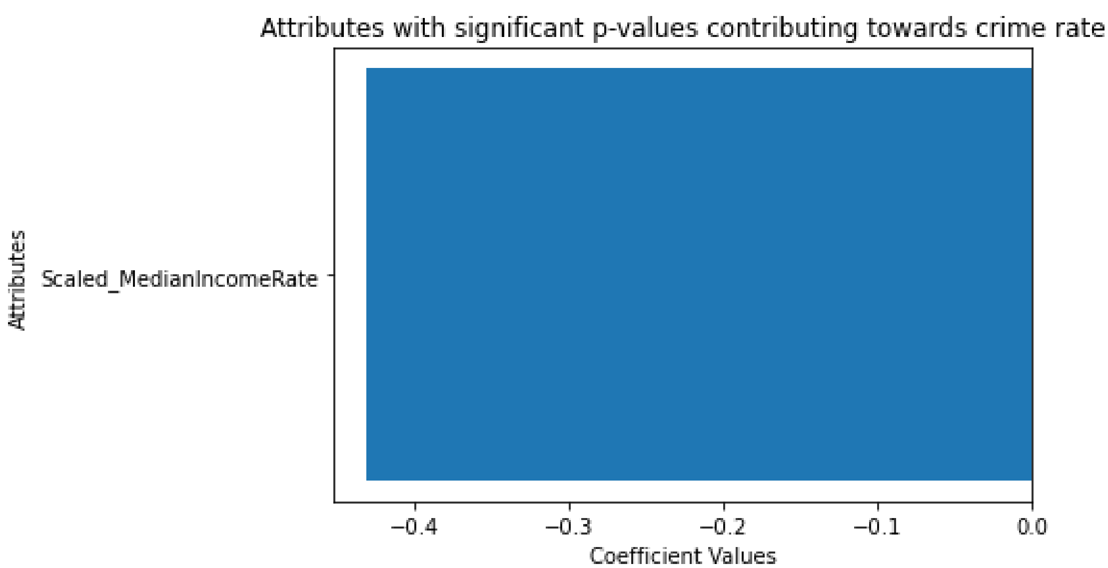
Till Dec-2021

Ground Truth:
Linear Model summary:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.552
Model:                            OLS   Adj. R-squared:                  0.542
Method:                 Least Squares   F-statistic:                     52.67
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.06e-22
Time:                        16:45:32   Log-Likelihood:                 86.845
No. Observations:                 132   AIC:                            -165.7
Df Residuals:                     128   BIC:                            -154.2
```

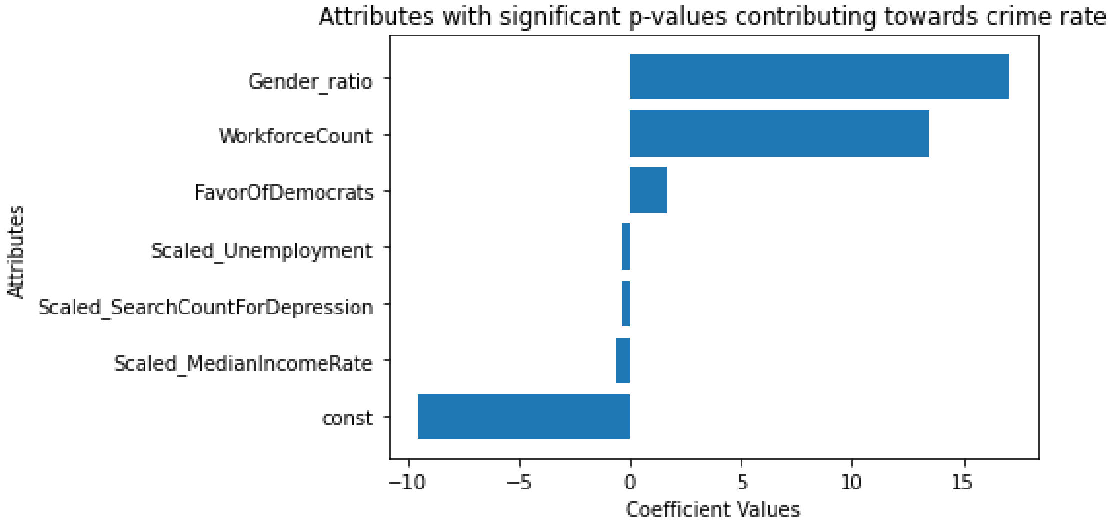
```
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
===========
                             coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----------
const                     -3.8825      2.635     -1.473      0.143      -9.097
1.332
Gender_ratio               8.4422      5.083      1.661      0.099      -1.615
18.499
Scaled_Unemployment        0.0708      0.058      1.219      0.225      -0.044
0.186
Scaled_MedianIncomeRate   -0.4305      0.049     -8.788      0.000      -0.527
-0.334
==============================================================================
Omnibus:                        6.143   Durbin-Watson:                   0.555
Prob(Omnibus):                  0.046   Jarque-Bera (JB):                7.404
Skew:                           0.274   Prob(JB):                       0.0247
Kurtosis:                       4.023   Cond. No.                         635.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Ground Truth:



```
Our Model:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.799
Model:                            OLS   Adj. R-squared:                  0.787
Method:                 Least Squares   F-statistic:                     70.26
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           3.94e-40
Time:                        16:45:32   Log-Likelihood:                 139.56
No. Observations:                 132   AIC:                            -263.1
Df Residuals:                     124   BIC:                            -240.1
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                              -9.5504      1.861     -5.133      0.000
-13.233      -5.868
Gender_ratio                       17.0663      3.548      4.811      0.000
10.045      24.088
Scaled_Unemployment                -0.3287      0.059     -5.598      0.000
-0.445      -0.212
Scaled_MedianIncomeRate            -0.6004      0.040    -15.035      0.000
-0.679      -0.521
WorkforceCount                     13.4592      2.671      5.039      0.000
8.173      18.745
PercentNegativeUsers                0.1383      0.121      1.144      0.255
-0.101       0.378
FavorOfDemocrats                    1.6635      0.538      3.091      0.002
0.598       2.729
Scaled_SearchCountForDepression    -0.3606      0.044     -8.159      0.000
-0.448      -0.273
==============================================================================
Omnibus:                        6.452   Durbin-Watson:                   1.211
Prob(Omnibus):                  0.040   Jarque-Bera (JB):                7.442
Skew:                          -0.315   Prob(JB):                       0.0242
Kurtosis:                       3.978   Cond. No.                         789.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

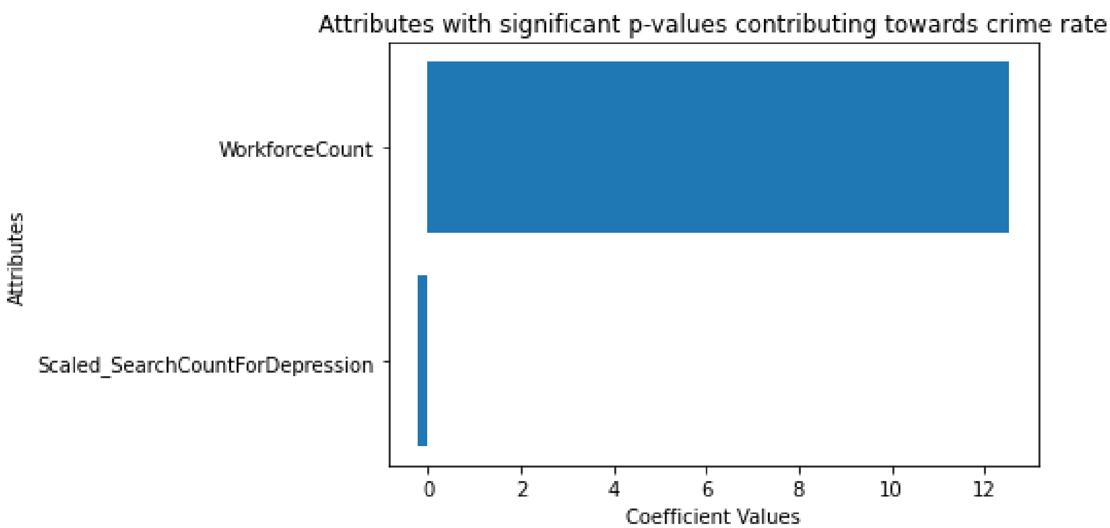
Plot for Linear Model Coefficients for Our Model:



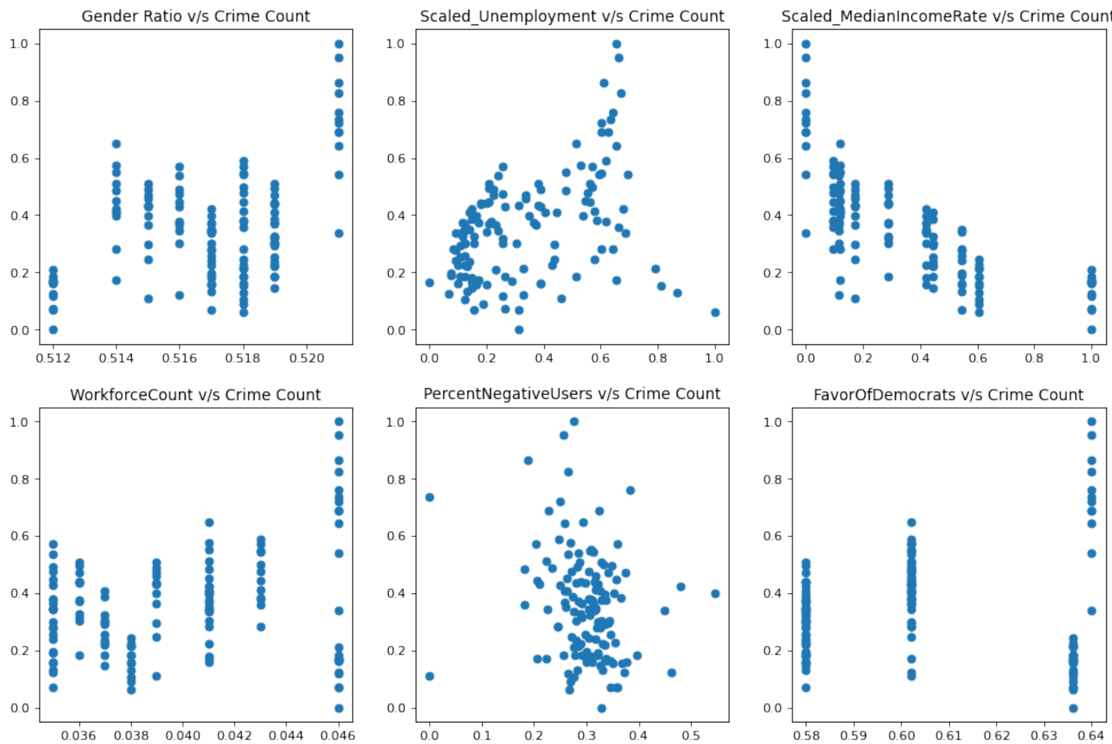

```
Our Model without ground truth:
Linear Model summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Scaled_CrimeCount   R-squared:                       0.160
Model:                            OLS   Adj. R-squared:                  0.133
Method:                 Least Squares   F-statistic:                     6.042
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           0.000175
Time:                        16:45:32   Log-Likelihood:                 45.280
No. Observations:                 132   AIC:                            -80.56
Df Residuals:                     127   BIC:                            -66.15
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
===================
                                      coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------------
const                               0.7992      0.456      1.752      0.082
-0.104       1.702
WorkforceCount                     12.5342      5.315      2.358      0.020
2.017      23.051
PercentNegativeUsers               -0.4293      0.237     -1.813      0.072
-0.898       0.039
```

```
FavorOfDemocrats                    -1.1609      0.876     -1.325      0.188
-2.895       0.573
Scaled_SearchCountForDepression    -0.2289      0.087     -2.637      0.009
-0.401      -0.057
==============================================================================
Omnibus:                        5.261   Durbin-Watson:                   0.360
Prob(Omnibus):                  0.072   Jarque-Bera (JB):                4.935
Skew:                           0.358   Prob(JB):                       0.0848
Kurtosis:                       3.621   Cond. No.                         457.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Plot for Linear Model Coefficients for Our Model without ground truth:

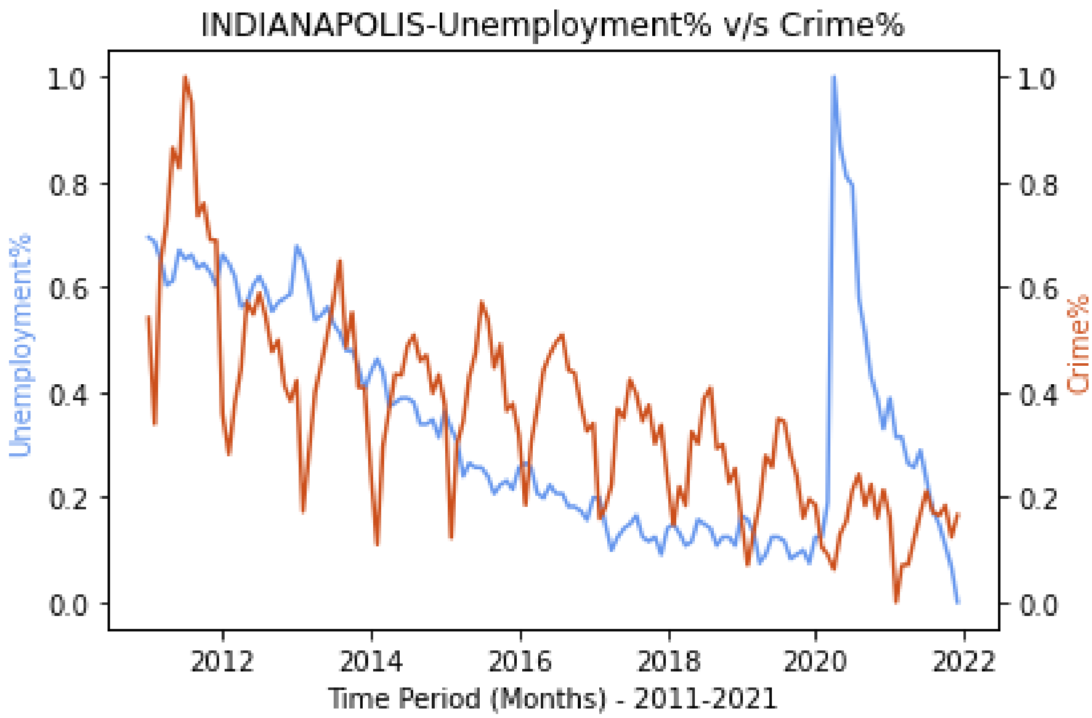


Correlation Graphs:

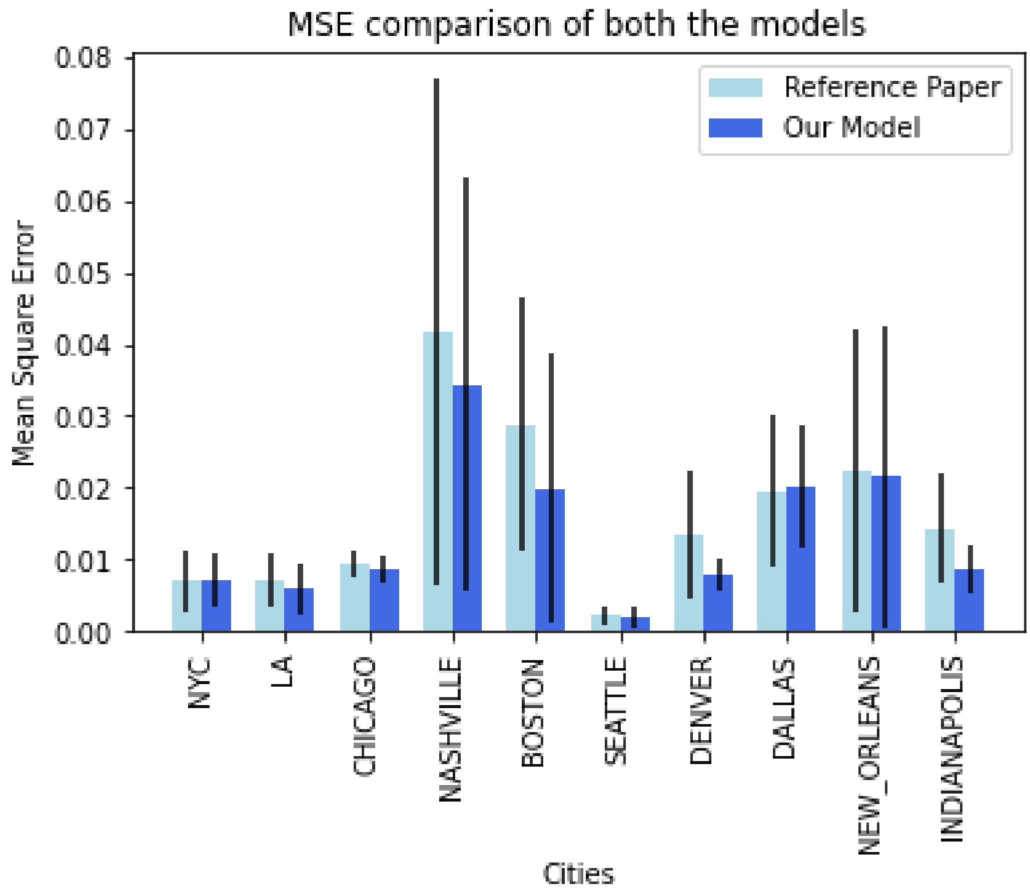



```
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
1: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "g-" (-> color='g'). The keyword argument will take precedence.
  ax1.plot(city_data["MonthYear"], city_data['Scaled_Unemployment'], 'g-',
color="#6495ED")
/var/folders/j_/kdgw7x6d25j6yr_1c3mwp9m40000gn/T/ipykernel_39508/1320670805.py:1
2: UserWarning: color is redundantly defined by the 'color' keyword argument and
the fmt string "b-" (-> color='b'). The keyword argument will take precedence.
  ax2.plot(city_data["MonthYear"], city_data['Scaled_CrimeCount'], 'b-',
color="#CC4F1B")
```

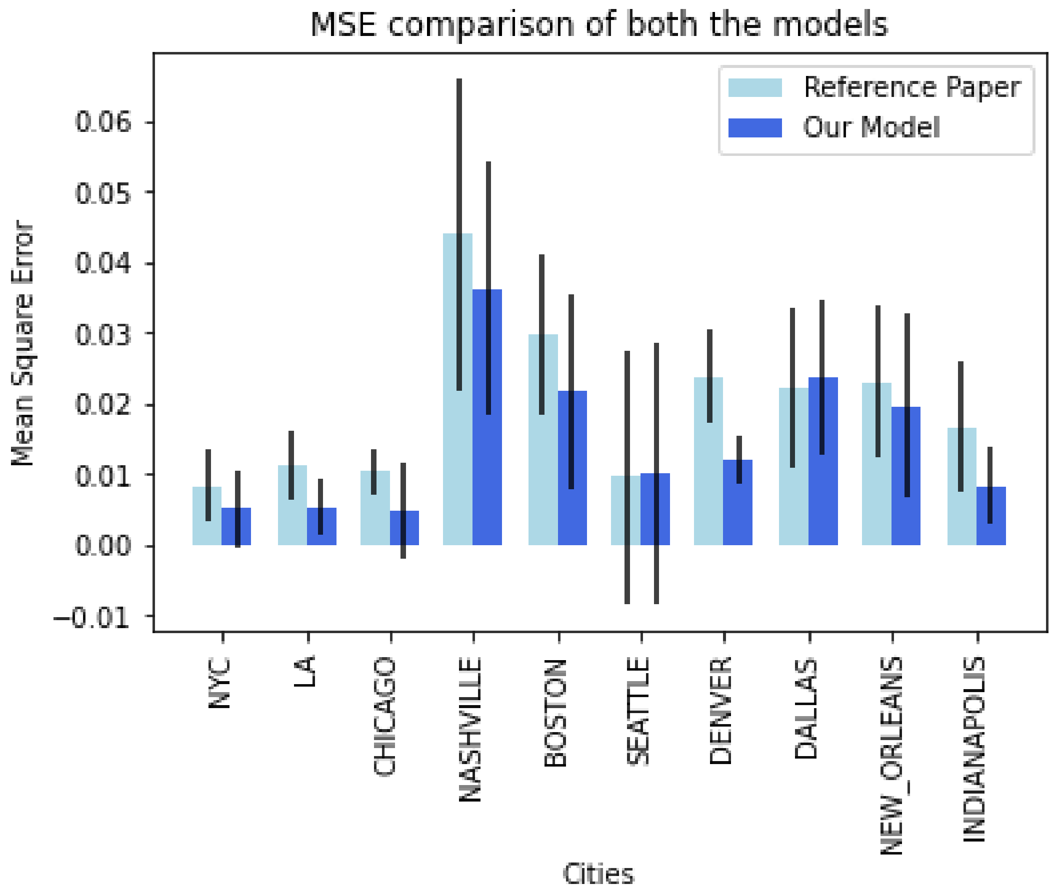
```
<Figure size 2400x1200 with 0 Axes>
```



Before Covid (Mar-2020)

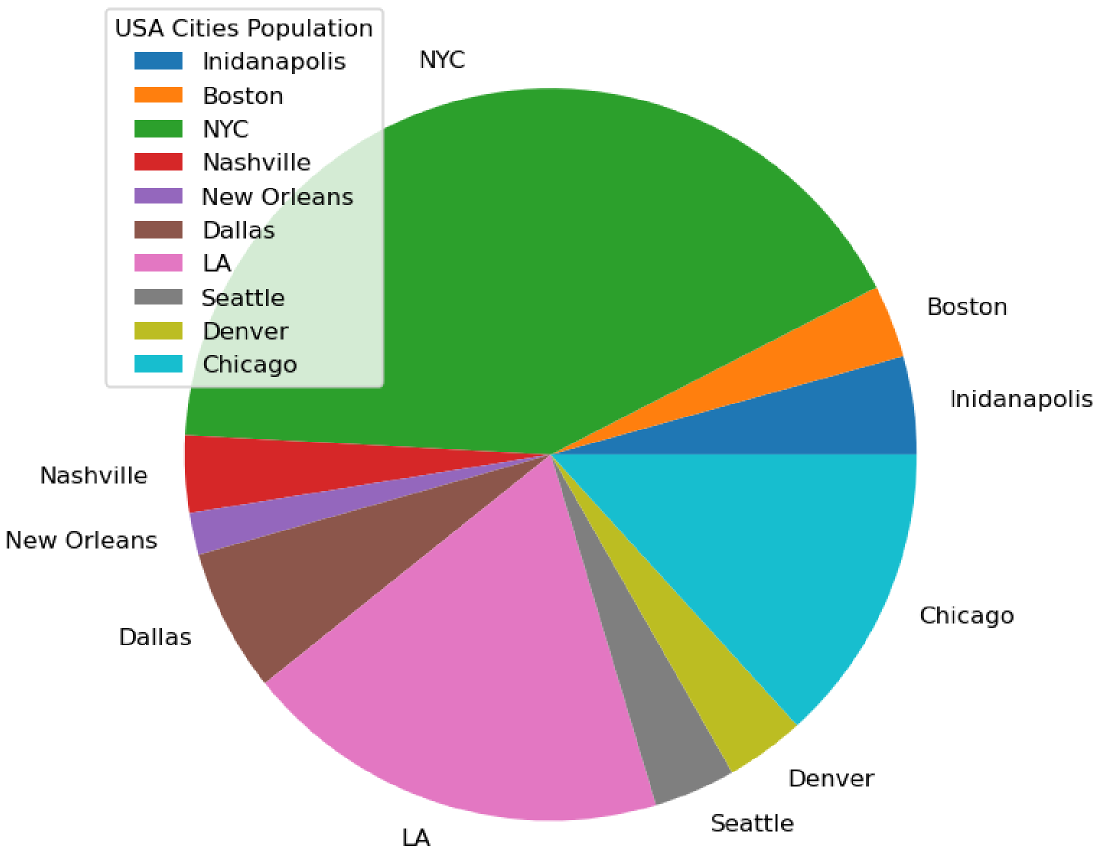


Till Dec-2021



[25]:
```
# USA Cities Population Pie Chart - Cities chosen for analysis for this project

USA = 331900000

total = 882039 + 654776 + 8468000 + 692587 + 376971 + 1288000 + 3849000 +␣
↪733919 + 711463 + 2697000

def get_pop_perc(pop):
#     return round(float(pop / USA) * 100, 2)
    return pop/total

inidanapolis = get_pop_perc(882039)
boston = get_pop_perc(654776)
nyc = get_pop_perc(8468000)
nashville = get_pop_perc(692587)
new_orleans = get_pop_perc(376971)
dallas = get_pop_perc(1288000)
la = get_pop_perc(3849000)
```

```
seattle = get_pop_perc(733919)
denver = get_pop_perc(711463)
chicago = get_pop_perc(2697000)

# others = float(100 - (inidanapolis + boston + nyc + nashville + new_orleans +␣
↪dallas + la + seattle + denver + chicago))

figure(figsize=(8, 7), dpi=120)
labels = ["Inidanapolis", "Boston", "NYC", "Nashville", "New Orleans",␣
↪"Dallas", "LA", "Seattle", "Denver", "Chicago"]
plt.pie([inidanapolis, boston, nyc, nashville, new_orleans, dallas, la,␣
↪seattle, denver, chicago], labels=labels)
plt.legend(title = "USA Cities Population")
plt.savefig('CitySelPopWise.png', transparent=True)
plt.show()
```



[26]:
```
# Playing around with FamaMacBeth Algo, with Newey-West adjustment

# import pandas as pd
# from linearmodels import FamaMacBeth # import package
# ## 'panel' is a multi-index Pandas panel dataframe (stock code - date)
# panel = pd.read_csv('~/Downloads/Sample.csv')
# panel['date'] = pd.to_datetime(panel['Time'], format="%b-%y") # date should␣
↪be `datetime` format
# panel = panel.set_index(['City', 'date']) # multi-index

# print(panel)

# mod = FamaMacBeth.from_formula('Output ~ 1 + Input1 + Input2 + Input3',␣
↪data=panel)
# ## `bandwidth` is the lagged number of Newey-West, normally, bandwidth = 4(T/
↪100) ^ (2/9)
# ## Remove all Settings in parentheses if newey-West adjustments are not␣
↪required.
# res = mod.fit(cov_type= 'kernel',debiased = False, bandwidth = 3)
# print(res.summary)
```

[27]:
```
# Barebone implementation of FamaMacBeth

# def ols_coef(x,formula):
#     return smf.ols(formula,data=x).fit().params
# panel = pd.read_csv('~/Downloads/Sample.csv')
# panel['date'] = pd.to_datetime(panel['Time'], format="%b-%y")
# gamma = (panel.groupby('date').apply(ols_coef,'Output ~ 1 + Input1 + Input2 +␣
↪Input3'))
# print(gamma.head())

# def fm_summary(p):
#     s = p.describe().T
#     s['std_error'] = s['std']/np.sqrt(s['count'])
#     s['tstat'] = s['mean']/s['std_error']
#     return s[['mean','std_error','tstat']]
# print(fm_summary(gamma))
```

[28]:
```
# Chicago Covid Comparison

# chicago_covid = pd.read_csv('~/Documents/Practice/Chicago_Covid.csv',␣
↪header=None)
# chicago_covid.columns = ['Perc', 'MonthYear']
# chicago_covid['MonthYear'] = pd.to_datetime(chicago_covid['MonthYear'],␣
↪format="%b-%Y")
# chicago_covid = chicago_covid.sort_values(by=['MonthYear'])
```

```
# # chicago_covid   # Mar2020 to Nov2022

# chicago_dataset = pd.read_csv('~/Documents/Practice/
↪chicago_city_final_dataset_v2.csv')
# chicago_dataset['MonthYear'] = pd.to_datetime(chicago_dataset['MonthYear'],␣
↪format="%b-%y")

# chicago_combined_dataset_covid = chicago_dataset[chicago_dataset['MonthYear']␣
↪>= '2020-03-01']

# perc = []
# for index, row in chicago_combined_dataset_covid.iterrows():
#     month_year = row['MonthYear']
#     perc_pos_covid_cases = chicago_covid[chicago_covid['MonthYear'] ==␣
↪month_year]['Perc']
#     perc.append(float(perc_pos_covid_cases))

# chicago_combined_dataset_covid['CovidPercPositive'] = perc

# regress_our_model_for_covid(chicago_combined_dataset_covid, 2)
# regress_our_model(chicago_combined_dataset_covid, 2)
```

[29]:
```
# Trying out OLS Regression with multivariate implementation

# import statsmodels.api as sm

# city_data = pd.read_csv('~/Documents/Practice/CSV_LA_city_final_dataset_v2.
↪csv')
# normalize_dataset(city_data)
# y = city_data['Scaled_CrimeCount']
# x = city_data.loc[:,['Gender_ratio', 'Scaled_Unemployment',␣
↪'Scaled_MedianIncomeRate']]

# X = sm.add_constant(x)
# lin_model = sm.OLS(y, X)
# regr_results = lin_model.fit()
# print(regr_results.summary())
# print(list(regr_results.params))

# transformer = PolynomialFeatures(degree=6, include_bias=False)
# x_ = transformer.fit_transform(x)
# x_ = pd.DataFrame(np.array(x_), columns=transformer.get_feature_names_out())
# X = sm.add_constant(x_)
# poly_model = sm.OLS(y, X)
# regr_results = poly_model.fit()
# print(regr_results.summary())
```

[30]:

```
# Playing around with the interaction terms

# city_data = pd.read_csv("~/Documents/Practice/chicago_city_final_dataset_v2.
↪csv".format(absolute_file_path, city_file), header=0)
# city_data = add_mental_health_data("CHICAGO", city_data, "~/Documents/
↪Practice")
# y = city_data['Cul. Crime count']
# x = city_data.loc[:,['Gender_ratio', 'UnemploymentRate', 'MedianIncomeRate',␣
↪'WorkforceCount', 'PercentNegativeUsers', 'FavorOfDemocrats',␣
↪'SearchCountForDepression']]
# transformer = PolynomialFeatures(degree=2, include_bias=False)
# x_ = transformer.fit_transform(x)
# z = transformer.get_feature_names_out()

# # for idx, ele in enumerate(z):
# #     if idx in [3, 7, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 23, 24,␣
↪25, 28, 30, 31]:
# #         print("{}: {}".format(idx, ele))
# # print(remove_interaction_terms(z, [19, 26, 33, 34]))
# # print(list(z[:10]) + list(z[11:19]) + list(z[20:31]) + list(z[32:]))

# # print(z[:7] + [z[12]] + [z[17]] + [z[21]] + [z[24]] + [z[26]])

# # Removing interaction terms
# new_x_ = []
# interaction_terms = [1, 8, 14, 17, 19, 22, 24, 26, 28, 29, 30, 31, 33]
# for item in x_:
# #     temp = list(item[:5]) + list(item[6:10]) + list(item[11:19]) +␣
↪list(item[20:22]) + list(item[23:25]) + list(item[28:31]) + list(item[32:])
#     temp = remove_interaction_terms(item, interaction_terms)
#     new_x_.append(temp)

# # # # new_x_ = pd.DataFrame(np.array(new_x_), columns=list(z[:5]) + list(z[6:
↪10]) + list(z[11:19]) + list(z[20:22]) + list(z[23:25]) + list(z[28:31]) +␣
↪list(z[32:]))
# new_x_ = pd.DataFrame(np.array(new_x_), columns=remove_interaction_terms(z,␣
↪interaction_terms))
# # new_x_ = pd.DataFrame(np.array(x_), columns=z)
# X = sm.add_constant(new_x_)
# regr_results = sm.OLS(list(y), X).fit()
# regr_results.summary()
```

[31]:

```
# 0: Gender_ratio
# 1: UnemploymentRate
# 2: MedianIncomeRate
# 3: WorkforceCount
```

```
# 4: PercentNegativeUsers
# 5: FavorOfDemocrats
# 6: SearchCountForDepression
# 7: Gender_ratio^2
# 8: Gender_ratio UnemploymentRate
# 9: Gender_ratio MedianIncomeRate
# 10: Gender_ratio WorkforceCount
# 11: Gender_ratio PercentNegativeUsers
# 12: Gender_ratio FavorOfDemocrats
# 13: Gender_ratio SearchCountForDepression
# 14: UnemploymentRate^2
# 15: UnemploymentRate MedianIncomeRate
# 16: UnemploymentRate WorkforceCount
# 17: UnemploymentRate PercentNegativeUsers
# 18: UnemploymentRate FavorOfDemocrats
# 19: UnemploymentRate SearchCountForDepression
# 20: MedianIncomeRate^2
# 21: MedianIncomeRate WorkforceCount
# 22: MedianIncomeRate PercentNegativeUsers
# 23: MedianIncomeRate FavorOfDemocrats
# 24: MedianIncomeRate SearchCountForDepression
# 25: WorkforceCount^2
# 26: WorkforceCount PercentNegativeUsers
# 27: WorkforceCount FavorOfDemocrats
# 28: WorkforceCount SearchCountForDepression
# 29: PercentNegativeUsers^2
# 30: PercentNegativeUsers FavorOfDemocrats
# 31: PercentNegativeUsers SearchCountForDepression
# 32: FavorOfDemocrats^2
# 33: FavorOfDemocrats SearchCountForDepression
# 34: SearchCountForDepression^2
```