

# ЗАДАНИЕ 13

на практические занятия по дисциплине  
«Прикладная теория вероятностей и статистика»

## «Регрессионный анализ»

### Задание

Реализовать python программу для вычисления линейной зависимости между двумя параметрами выборки. Программа должна:

- 1) с помощью метода наименьших квадратов найти коэффициенты линейной регрессии;
- 2) построить график, на котором в виде точек отобразить исходные данные (plt.scatter) и аппроксимирующую прямую;
- 3) вычислить коэффициент корреляции.

В качестве исходные данных взять информацию о стоимости домов (kc\_house\_data.csv) и рассчитать корреляцию между следующими столбцами:

- 1) price и sqft\_living;
- 2) price и sqft\_lot;
- 3) sqft\_living и sqft\_lot.

### Листинг:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import pearsonr

# Чтение данных
df = pd.read_csv('kc_house_data.csv')

def regression_and_correlation(x, y, xlabel, ylabel):
    # Подготовка данных
    X = np.array(df[x])
    Y = np.array(df[y])
```

```

# Вычисление коэффициентов линейной регрессии
A = np.vstack([X, np.ones(len(X))]).T
m, c = np.linalg.lstsq(A, Y, rcond=None)[0]

# Построение графика
plt.figure(figsize=(10, 6))
plt.scatter(X, Y)
plt.plot(X, m * X + c, color='red')
plt.xlabel(xlabel)
plt.ylabel(ylabel)
plt.title(f'{xlabel} vs {ylabel}')
plt.show()

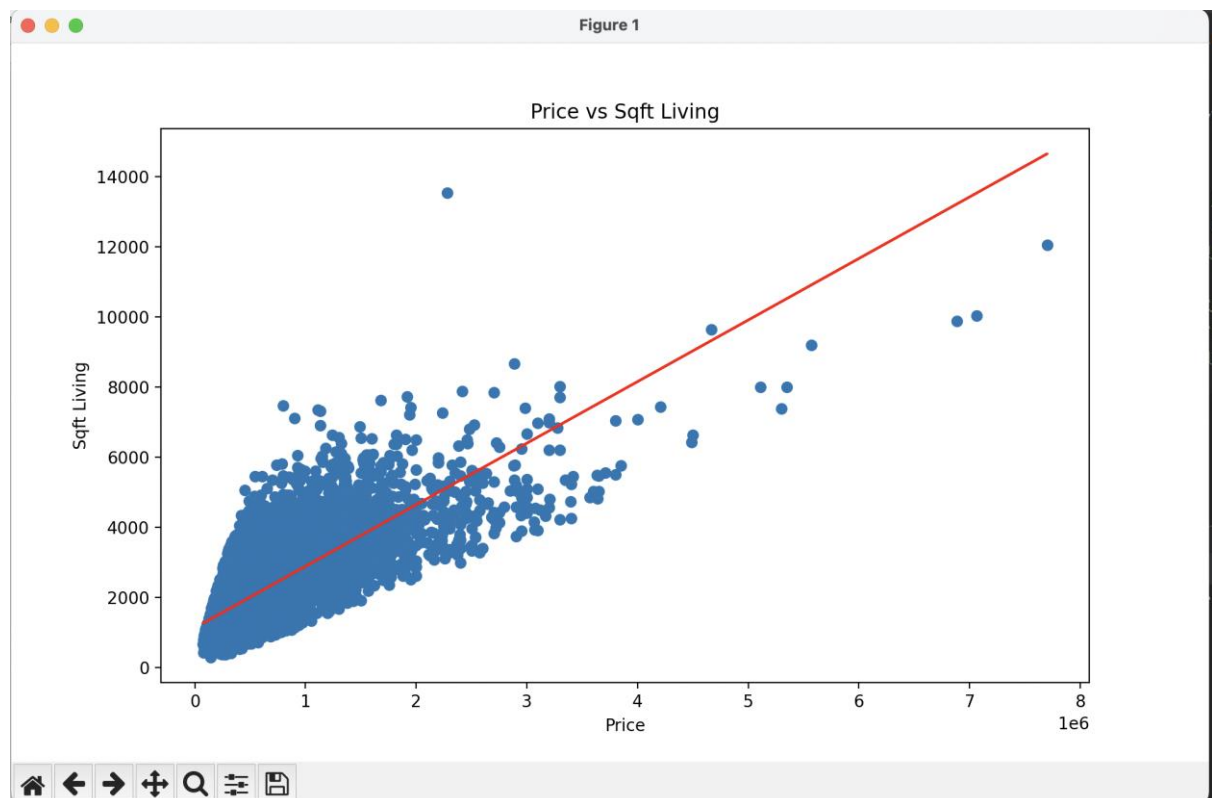
# Вычисление коэффициента корреляции
corr, _ = pearsonr(df[x], df[y])
print(f'Коэффициент корреляции между {xlabel} и {ylabel}: {corr:.3f}\n')

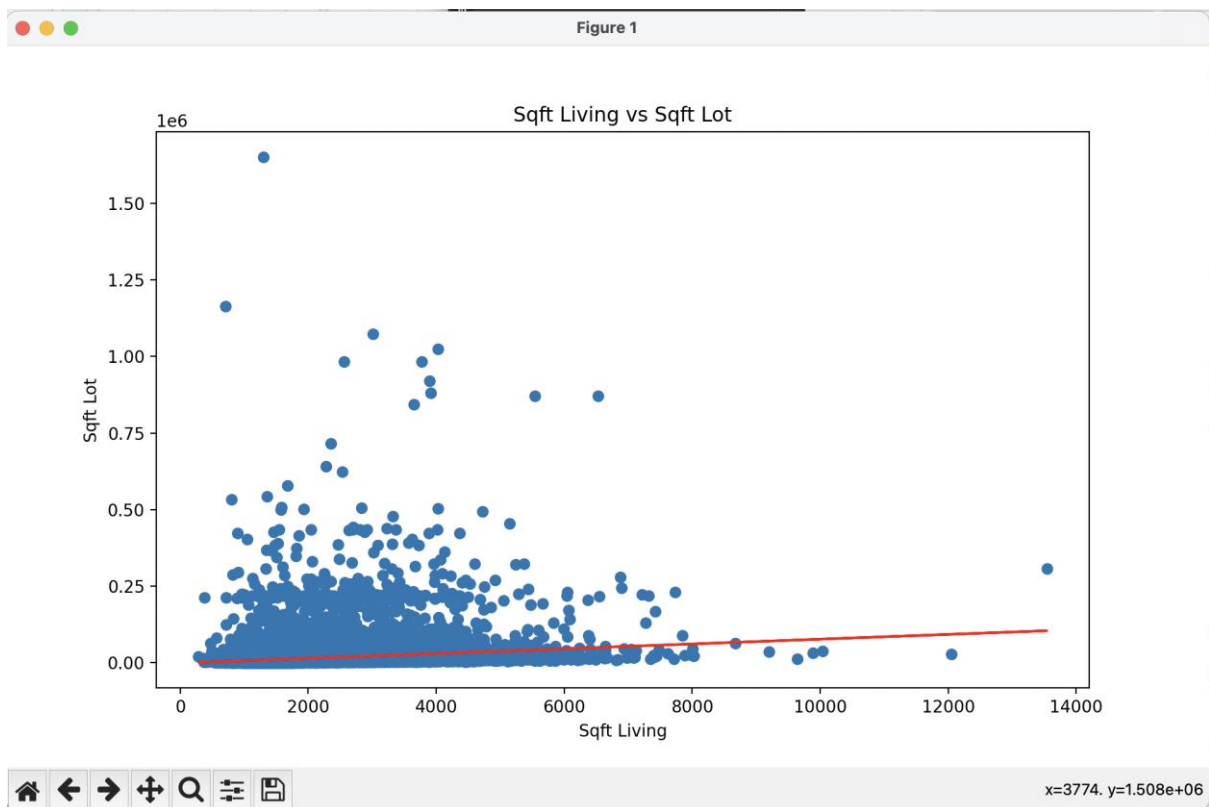
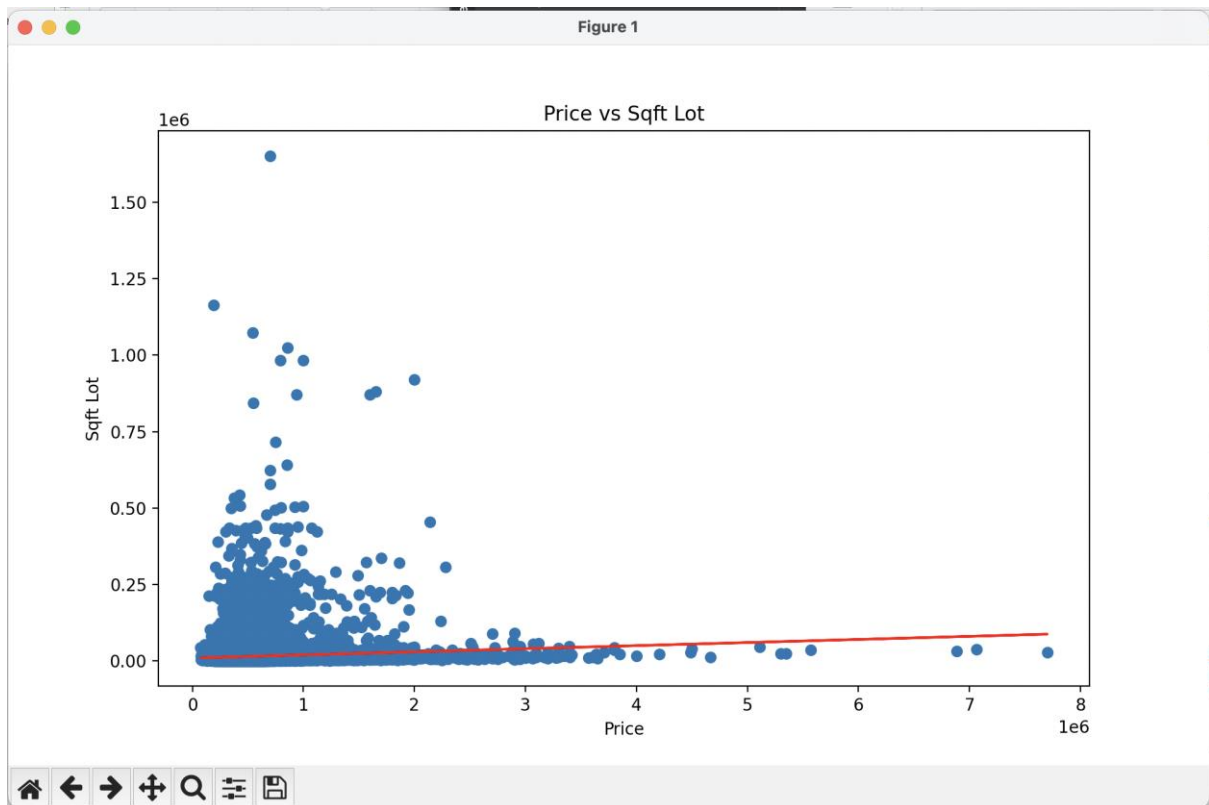
regression_and_correlation('price', 'sqft_living', 'Price', 'Sqft Living')
regression_and_correlation('price', 'sqft_lot', 'Price', 'Sqft Lot')
regression_and_correlation('sqft_living', 'sqft_lot', 'Sqft Living', 'Sqft Lot')

```

Вывод:

Скриншот:





Вывод полностью:

/Users/andrey/Documents/PyCharm/pythonProject/bin/python

/Users/andrey/Documents/PyCharm/pythonProject/main.py

Коэффициент корреляции между Price и Sqft Living: 0.702

Коэффициент корреляции между Price и Sqft Lot: 0.090

Коэффициент корреляции между Sqft Living и Sqft Lot: 0.173

Process finished with exit code 0