Regression Analysis

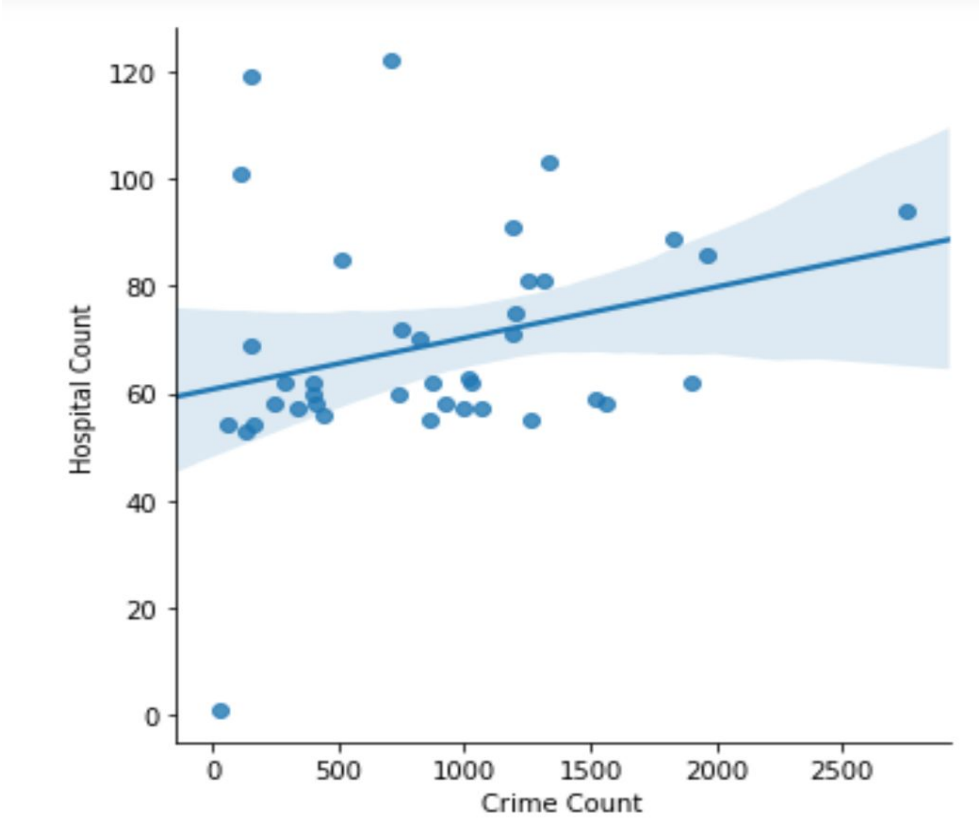
Goal: After creating a zip code ranking interface, we decided to analyze a few data sets. The datasets we focused on were based upon some hypotheses that we came up with. The datasets we worked with were crime, hospital, and housing sales. Instructions for loading the analysis will be below.

Instruction:

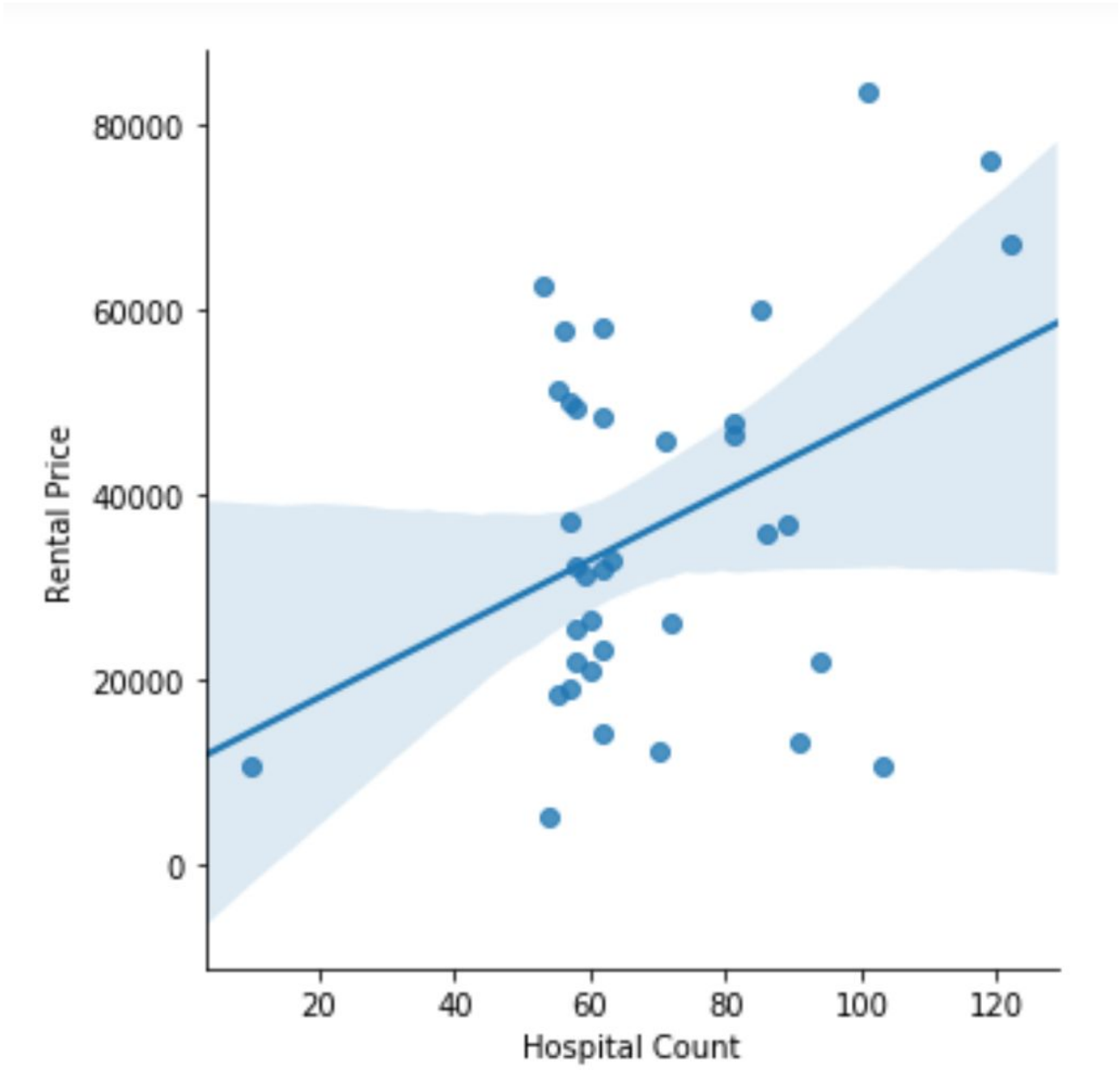
1. Dowload CrimeAnalysis.py, CrimeAnalysis.csv, HospitalAnalysis.csv, and HousingSalesAnalysis.csv.
2. Run Regression_Analysis.py
3. You should see three graphs and their respective correlation coefficients, the same as the screen-shots shown below

We first converted the median housing purchase price to a yearly rental rate, using a conservative value of 6%. Then we analyzed further by taking the rental rate with the following analysis.

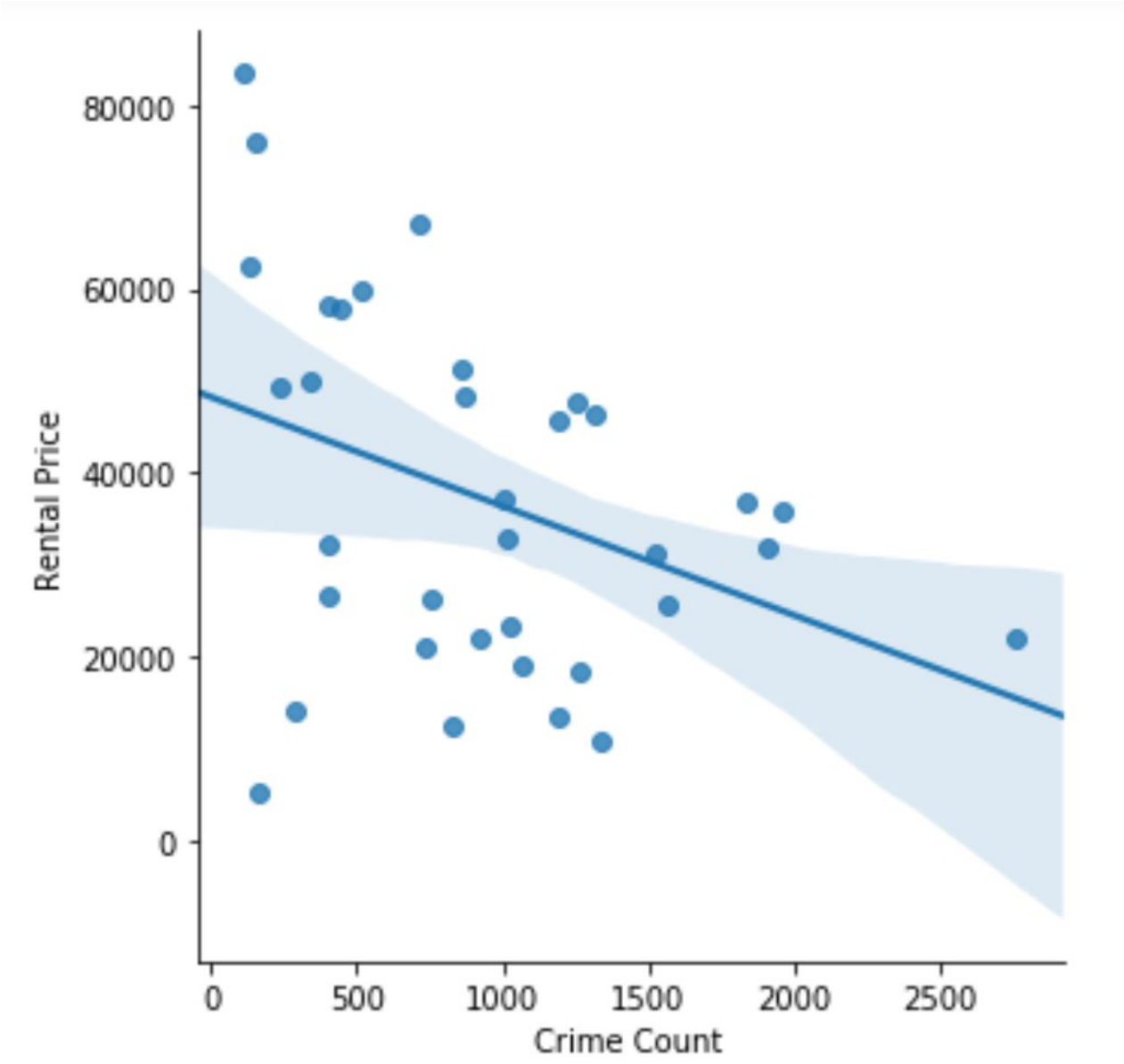
The first hypothesis we came up with was that there should be a relationship between crime and hospitals because high crime areas warrant more hospitals to treat the victims. Although the regression may show some regression between the two variables, the correlation coefficient between crime count and hospital count low (0.28).



The second hypothesis we came up with was that there should be a relationship between hospitals and rent because richer neighborhoods can afford to go to hospitals more frequently. The regression shows less dependence between the two variables than the correlation coefficient. The regression shows a concentration of 60 hospitals among various rental prices, while the correlation coefficient shows a value of .4 which was better than the relationship between crime and hospital. This shows that it is important to not only look at the graph but also at the computed correlation coefficient when you analysis the relationship between two values.



The third hypothesis was as crime increases, the rental rate will decrease. The graph below shows that between rental and crime there is some correlation between these two values. As crime increases, rental prices start to drop. This rental price was the median rental price of each zip code. This shows that if you want to live cheap you may have to pay with your safety (Correlation Coefficient = ~-0.37).




```
Hospital Count vs Crime Count- 0.2801974340553681
Hospital Count vs Rental Price- 0.3993229388702484
Crime Count vs Rental Price- -0.3720979110362589
```