Summary:

I approached this problem with the intent of using a linear model to find the effect and significance of independent variables to give insight to my fictional client. (In this case, the city of San Francisco)

What Worked:

I like to follow the Cross-Industry Standard Process for Data Mining (CRISP-DM) when I am working on a new project. That gave me direction in my analysis.

Also, my EDA let me clean up outliers and visually check my data.

What didn’t work:

I had planned to use the StatsModels.api library as it very neatly provides a summary of statistical significance and linear coefficients, but I ran into an issue with fitting my model and was running out of time so I switched to scikit-learn to at least see my coefficients and draw a conclusion.

Also, due to time constraints I was unable to write my usual README.md that would have summarized my work and provided a collection of graphs and my stats models significance chart.