**Executive Summary**

**The Problem**

It’s no secret that rent prices have been skyrocketing as of lately, and many may pose the question as to “why”? We chose to investigate a specific factor that may contribute to the increased cost of living, crime rates. This report seeks to address the impact of crimes per capita on the average rental prices across 51 major US cities over a time span of 11 years, from 2005 to 2015.

We hypothesize two possible outcomes:

: Crimes per capita do NOT have an impact on average rent prices

: Crimes per capita do have an impact on average rent prices

A data-driven approach can be useful to obtain insights to this question in that our finding might be counterintuitive. There are two possible outcomes associated with the question. The second one (the alternative hypothesis) seems very obvious to a lot of people: our intuition is that higher crime rates signal unpleasant living conditions of the area, and the demand for the apartments in the area should be lower, driving rental prices lower. However, there exists another possible outcome (the null hypothesis), which could be more complicated than our reasoning above, since crime rates can be correlated with other confounding factors, which might IN REALITY have the predominant influence on the average rent prices. For example, high crime rates can be correlated with high population density; places with higher population densities host more social and economic activities and are more likely to witness violent criminal activities.

The takeaways from this data-driven study could also serve as a beneficial resource to a variety of interested parties. This can include real estate developers seeking to look for new areas to develop or help to select an area for redevelopment. Further, real estate investors may use falling crime rates to predict areas where rental prices may increase over time and use this to guide investment decisions. Law enforcement or other government agencies may also choose to use rental prices as a proxy for crimes per capita to determine if more police intervention is required, and to check if their crime rates are accurate.

**The Ideal Experiment**

Ideally, we want to have access to two groups of identical cities and compare the changes in their rent prices, where the cities in the treatment group only differ from the controlled group in the fact that the former experienced changes in crime rates in 2005-2015, and the latter did not. However, there do not exist identical cities in this world, and the best way we could thought of to mimic the ideal experiments was trying to splitting the cities into the following, where we would compare cities having crime rates changed from low->low with cities having crime rates changed from low->high and see how the average rent prices changed; similarly, we would compare cities with high->high crime rates and high->low crime rates.

Controlled groups:

* Low Crime Rates → Low Crime Rates
* High Crime Rates → High Crime Rates

Treatment groups:

* Low Crime Rates → High Crime Rates
* High Crime Rates → Low Crime Rates

With these in mind, we identified 4 cities that satisfied our requirements: New Orleans, Detroit, El Paso, and Boston. Using simple line charts, we could observe that a decrease in crime rates roughly corresponded to an increase in rent prices. However, since we only had limited data points (11 years of data) for every city and there are too many unobservable factors (not controlled) in a line chart, an “ideal experiment” like this is far from ideal. Thus, we needed more data and a more comprehensive analysis.

**New Orleans** (Low Crime Rates → High Crime Rates)

**Detroit** (High Crime Rates → Low Crime Rates)

**El Paso** (Low Crime Rates → Low Crime Rates)

**Boston** (High Crime Rates → High Crime Rates)

**The Data**

To further examine the issue of how crimes per capita may affect the rent prices in certain cities, we gathered a US crime dataset from Kaggle (1970-2015) and rent estimate datasets (2005-2015) from the Office of Policy Development and Research (PDR). Due to the differences in our datasets, we decided to study the years 2005 to 2015, and we merged the two datasets based on city, state, and year after data cleaning.

We also included and merged the following 2 datasets with our existing one because unemployment rate and median household income might be confounding variables in our experiment:

* Unemployment data - US Bureau of Labor Statistics (2005-2015)
* Median Household Income Data - US Census Dataset (2005-2015)

**The Approach & Findings**

Summary statistics roughly suggested that crimes per capita moved in the opposite direction as the average rent prices from 2005-2015, while the median household moved in the same direction. Correlation coefficients coincided with our simple observations from the summary statistics. One thing worth noticing was that the 2008 financial crisis disrupted the unemployment rate data and rendered the correlation coefficient between unemployment rate and average rent prices less meaningful.

Based on our observations above, we started our study with a simple linear regression model, where the average rent prices was the dependent variable and crimes per capita was the independent variable. The univariate model indicates that crimes per capita have a negative impact on the average rent at a 1% significance level. According to the model, a unit change in crimes per capita leads to -0.1424 unit change in the average rent (in USD).

In order to isolate the effect of crimes per capita on average rent prices from potential confounding variables, we then conducted a multivariate linear regression analysis to integrate unemployment rate and median household income into our existing model. We found that using the MLR model, crimes per capita still have a negative impact on average rent prices at a 1% significance level, and a unit change in crimes per capita leads to -0.0.0945 unit change in the average rent (in USD).

Furthermore, we investigated a subtopic about whether crimes per capita affect the average rent prices of different apartment types (0-bedroom, 1-bedroom, 2-bedroom, 3-bedroom, 4-bedroom apartments) differently. Slicing our datasets into 5 different subsets and using a similar multivariate linear regression analysis, we found out that crimes per capita do NOT have an impact on the average rent prices of 0 and 1-bedroom apartments at a 5% significance level, while they DO have a negative impact on the average rent prices of 2, 3, and 4-bedroom apartments at a 1% significance level. One thing one might notice is that we used logged variables for this subtopic study, which is because we wanted to interpret the results using percentage change instead of unit change, but this did not yield statistically significantly different results from not using logged variables.

**The Evaluation**

Pointing the lens on the number of bedrooms showed that statistically crime has no effect on rent prices for studio (0-bedroom) and 1-bedroom apartments. Inadvertently, the regression analysis showed, controlling for median income and unemployment, that crime does have an effect on 2, 3, and 4 bedroom apartment units.

Our interpretation for these differences is that the average rent prices of larger two, three and four bedroom homes might be more sensitive to changes in crimes per capita because families (instead of individuals) typically stay in larger homes, and these groups of people place a greater emphasis on safety. Therefore, would be willing to pay a higher rental price to ensure the safety of their loved ones.

**The Conclusion**

In conclusion, when we look at average rent across different apartment types in the 51 major cities from 2005-2015, we reject the null hypothesis and conclude that crimes per capita do have an effect on average rent prices. Moreover, crimes per capita have a negative effect on average rent prices.

**Limitations**

Apparently, there are caveats and limitations to our study. To begin with, when studying our sub-topic regarding how crimes per capita affect the average rent prices of different apartment types differently, we could add interaction terms in our MLR model to clearly identify the differences. Additionally, we could glean more census data and focus on crime hot spots (in specific neighborhoods) instead of aggregating our data on a city level. Last but not least, we could study specific categories of crimes and their correlation with rent prices *(e.g. property crimes may be correlated with rent more than other kind of crimes).*