Violent Crimes Vs Rent prices

**Question 1:**

The problem we’re seeking answers to is the impact of crimes per capita on average rent prices in 51 major US cities.

**Question 2:**

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 first one seems very obvious to a lot of people: our intuition is that higher violent 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 a second possible outcome, which could be more complicated than our reasoning above, since violent crime rates can be correlated with other confounding factors. 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.

Given the second scenario mentioned above, we deem it necessary to investigate our question, because our findings can be useful to real-estate developers who seek to build new apartment buildings. It might seem natural for them to avoid building apartment buildings in areas with higher crime rates because they would assume that such places would have lower rental rates; however, if we found out that higher crime rates were not reflective of poorer living conditions and didn’t necessarily push rental rates lower, we would offer a new angle for developers when they choose new address to build. Moreover, if we found out that higher crime rates did result in lower rental rates (i.e. first possible outcome), such a fact might be useful to municipal law enforcement agencies as they try to monitor illegal activities, in that a decrease in rental rates might signal an increase in criminal activities.

**Question 3:**

To examine the issue of how violent crime rates may affect the housing market in certain cities, we will combine and analyze a US crime dataset from Kaggle and rent estimate datasets from the Office of Policy Development and Research. Due to the differences in our datasets, we will be combining the county data within the PD&R dataset to match with the city level data from Kaggle. Our independent variables will be the violent crime rates and the corresponding year, while our dependent variable will be the rent prices. We will compare cities that have the corresponding changes in crime rates in the 5 years between 2010-2015:

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

With these changes in mind, we will be analyzing the changes within apartment rent prices throughout the same time period. Cities that have drastic changes in crime rates will be used to identify if there is a relationship between the two variables and the strength of that relationship. Cities that have no significant changes in crime rates will be used to isolate the two variables from any systematic changes that occur due to changing situations in the housing market. An A/B test will be used to carry out our experiment, the earlier versions of these cities will be used as the A test, and the later years will be used as the B portion of the test. Furthermore, through the use of descriptive machine learning models and logistic regressions we can determine a population average for rent prices (Null Hypothesis). For the models that compare significant changes in crime rates, we will find the sample average for rent prices, which we can use to find the sample average for rent prices (Alternative Hypothesis).

**Question 4: What type of data will you be using/collecting and why is this data appropriate for this problem?**

To adequately explore our hypothesis, we will be using two different datasets from two different sources. For violent crime, we will be using the kaggle dataset “Analysis of US Crime data”. We will specifically be examining the number of violent crimes in **49** major U.S. cities. For rent prices, we will be using a dataset provided by the Office of Policy Development and Research. We will be examining the 50th percentile rent estimates, aggregated at a county level.

Since the violent crime data is aggregated on a city level, and the rent estimates are aggregated on a county level, we will be combining the data within the “Analysis of US Crime data” to re-aggregate at a county level. We will then be comparing the data between these two sources in a time horizon spanning 11 years - from 2005 to 2015.

This data is well suited to test our hypothesis as it will allow for a direct comparison between violent crimes and rent estimates at a county level across the United States. Further, our time span of 15 years is long enough to draw conclusions across different economic periods.

<https://www.kaggle.com/code/ayush1498/analysis-of-us-crime-data/data>

<https://www.huduser.gov/portal/datasets/50per.html#2012>