Project 2 - Crime in Austin

Blake Barber and Samuel McMillan

Introduction

In the realm of criminal statistics, understanding the dynamics and distribution of crime over time is crucial for effective law enforcement and public safety planning. Our dataset encompasses a comprehensive compilation of crime reports from Austin, Texas, throughout the year 2015, detailing incidents across various council districts, along with socio-economic indicators. By employing a multifaceted analysis technique that includes temporal trend analysis via line plots and spatial distribution through heatmaps, we have unearthed patterns and insights that were previously latent. This analysis serves a pivotal role in enabling law enforcement agencies to allocate resources efficiently, optimize patrol routes, and implement targeted crime prevention strategies. Our findings reveal not only the most common crimes but also pinpoint temporal hotspots, highlighting days and months with heightened criminal activity, as well as other socio-economic trends.

This analysis is instrumental as it provides a granular view of crime trends, facilitating data-driven decision-making that can lead to safer communities. It underscores the importance of leveraging data analytics in the sphere of public safety, guiding stakeholders in crafting policies that are both proactive and reactive to the patterns of urban crime. For an in-depth exploration of our methodology and findings, please refer to our presentation slides (CS5830_Project 2) and delve into our project on github (Project 2 Github). These resources offer a deep dive into the crime in Austin and our approach to unraveling it, emphasizing the power of data in building a safer society.

Dataset

Our dataset is of crime data recorded in the bustling urban landscape of Austin, Texas, in 2015. It is a detailed ledger that includes key attributes such as the type of offense, the district in which the incident occurred, the date of the report, along with geographic coordinates for precise location mapping. Accompanying these are socio-economic markers such as poverty levels, median household income, and housing affordability, which provide a multi-dimensional view of the context in which these crimes were committed. This integration of crime data with socio-economic factors makes the dataset particularly conducive to analysis, allowing for an exploration of not just the 'where' and 'when' of crime incidents but also the 'why' that supports different trends and patterns. The granularity of the data, with daily crime occurrences over the course of a year, offers a temporal resolution that is pivotal for detecting patterns and anomalies. Each record acts as a puzzle piece, and when assembled through analytical techniques, creates an extensive picture of the crimes within the city, making it a great dataset for dissecting the crime in Austin.

Analysis Technique

We used a bar chart to display the different types of crime because bar charts are good at visually showing which types of crime were most often performed. Bar charts were also used for the crime count by council district because it effectively conveyed the different levels of crime in each district visually. A scatterplot was used to show if rent cost affects the amount of crime in the area because it showed how the variables interacted with one another. A t-test was used to see if rent price and crime count were correlated because the t-test returns values that tell the mathematical similarities between the variables. A scatterplot was used to see if 2000-2012 house value changes and amount of crime were correlated because it showed visually if they were or not. The average of home value percent change was used to see if 2000-2012 house value changes and crime were correlated because if the average was greater than 58% it would show if home value has gone up in price faster than the norm, crime rates are correlated to go up with it. Line Plots were used to show the number of crimes in 2015 because a scatter plot is effective at showing data over a period of time.Finally, a heatplot was used to show the number of crimes in 2015 because it was good to see the correlation between number of crimes and the time of year, time of month, and day of month; it effectively conveys the clusters of when crime is most likely to occur and maybe show some patterns from that.

Results

Some of the most interesting findings come from the scatterplots. Generally, it seems like people think living in a more expensive home increases your likelihood for theft to happen to you. However, the findings in the study proved that wrong. In fact, there wasn't a big enough correlation between crime count and median rent to say anything about that see figure 1 below.

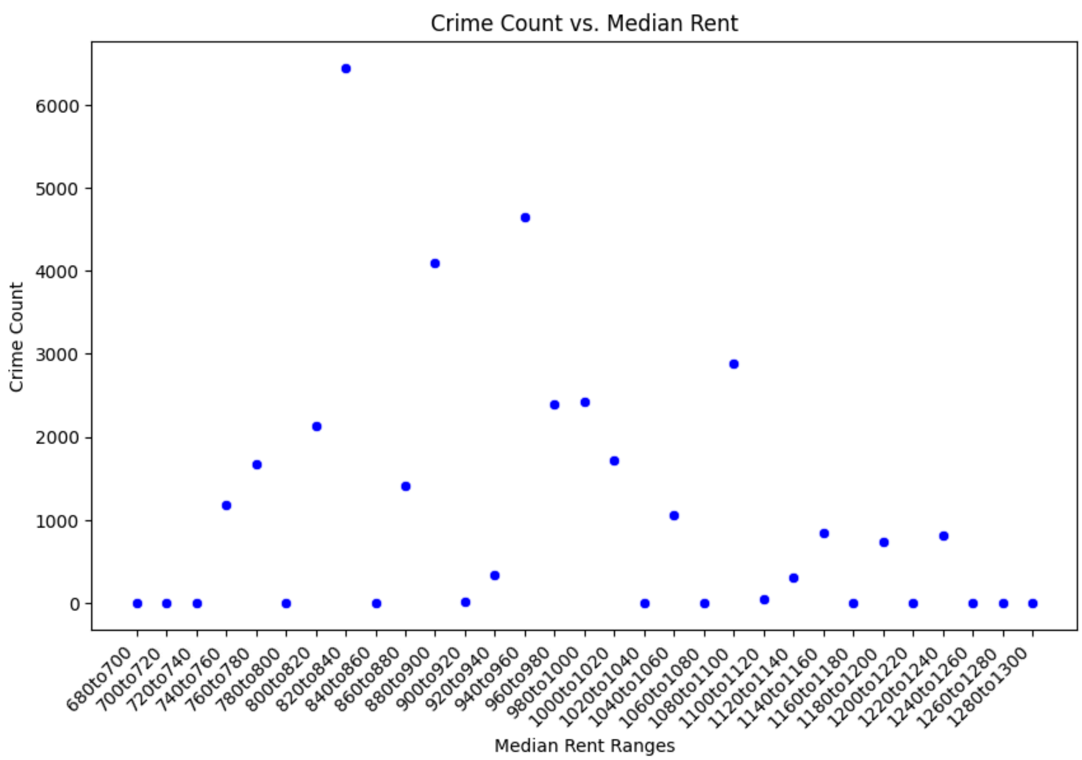

Figure 1. Crime Count vs. Median Rent in Austin Texas 2015

```
Ttest_indResult(statistic=0.5804519114950707, pvalue=0.5637836781382588)
```

While crime count vs. median rent did not yield much results, when it comes to moving into an area that just got a whole lot more expensive compared to the surrounding area, you would be hard pressed to be in a more crime free zone according to the data. It seems that it might matter to you to see if an area just got way more expensive, if you want to avoid being around more crime see figure 2 below.

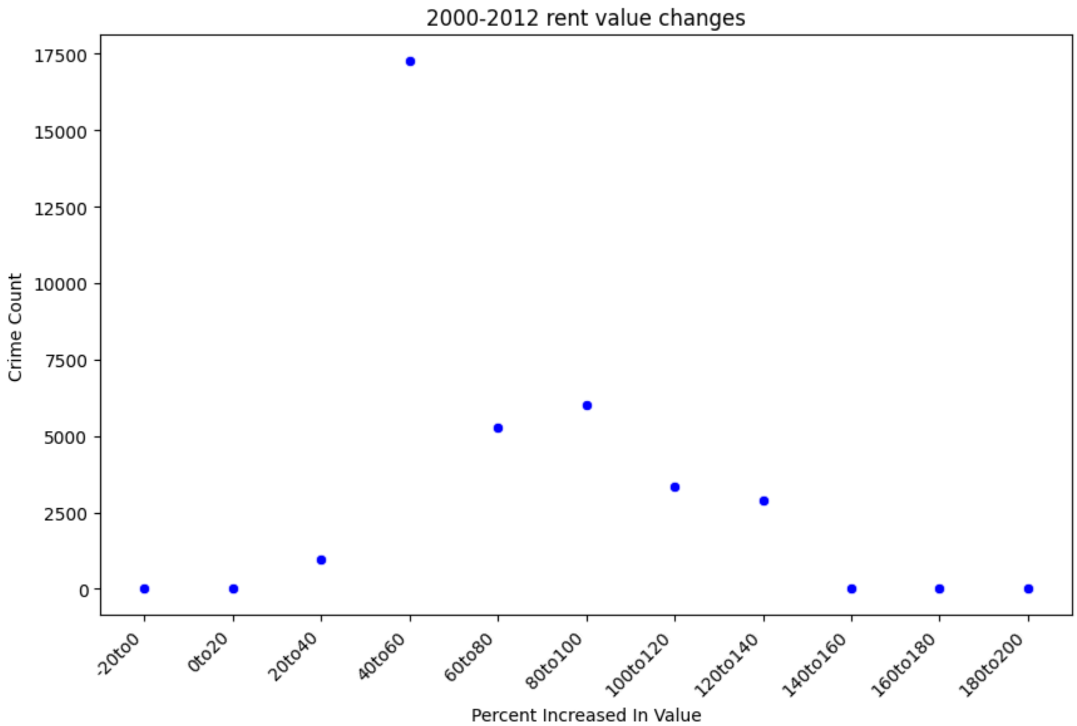

Figure 2. Rent Value Changes vs Crime Count

The average speed increase of house value in Austin, TX is 58% and after taking the mean it was seen that with greater speed of increase of house amount, is greater for most crimes as shown below making crimes and median home value strongly correlated.

```
crimeDf.Changeinmedianhomevalue20002012.mean()
```

```
75.22072962290353
```

Figure 2. Continued. This average speed of housing value increase is 75% in crime areas compared to Austin, TX average of 58%

The results for the crime count by crime type really help a person know what the main type of crime to prepare for is. For example, if robbery is the biggest threat, maybe one should lock their door more often in Austin, TX. Another interesting find from this chart is that vehicle burglary is quite high. This implies that the city may need to run a campaign that encourages people to lock their car doors and ensure things are hidden out of sight

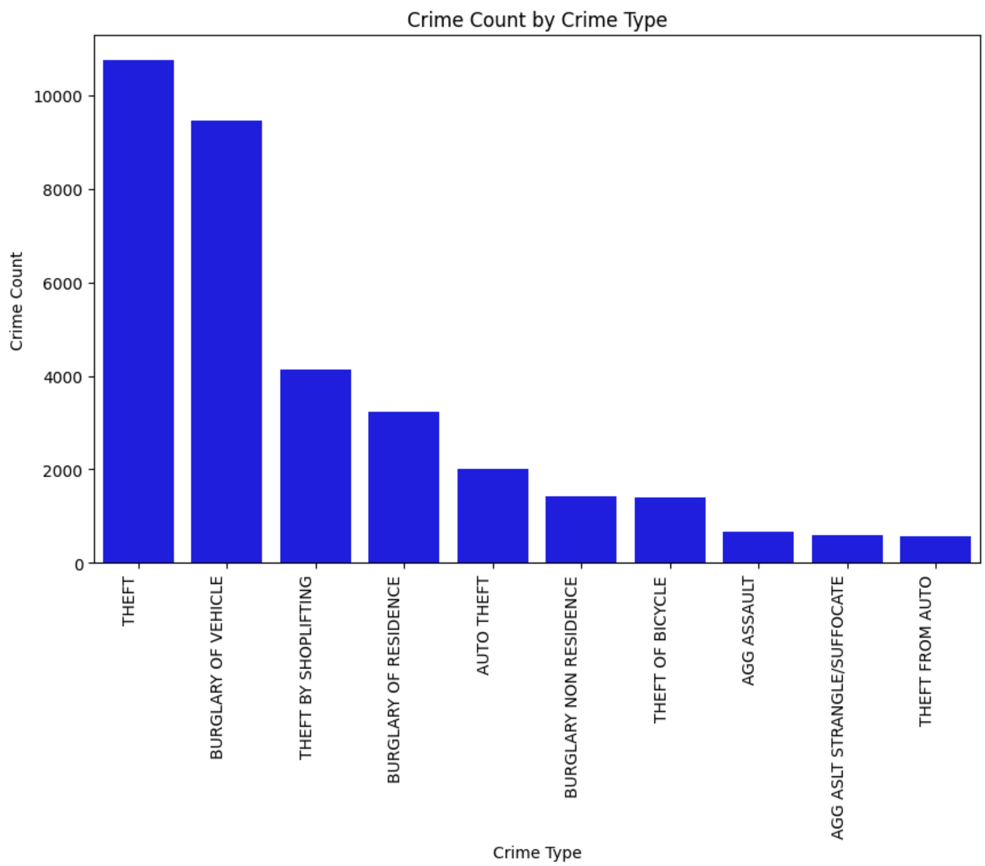

Figure 3. Crime Count by Crime Type

Crime count per council could inform a person what council in the city is the safest place to live, and where they may want to avoid. The results for the amount of crime per month and day according to the heat map shows that crime is spread out pretty uniformly over the course of the year. However, there is some evidence that crimes are more likely to happen in the summer and just before christmas in december. This could inform individuals to keep an eye out for crime during these times of the year.

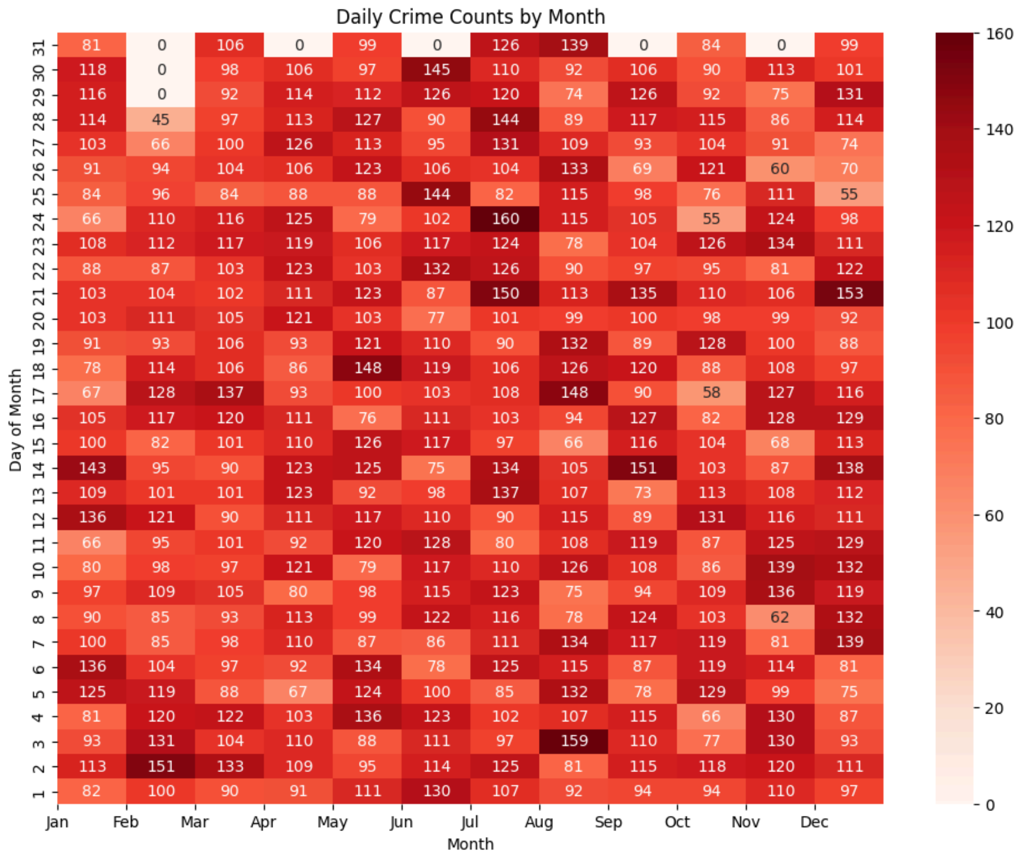

Figure 4. Heatmap of Daily Crime Counts throughout 2015

Technical

Initially for calculating the median rent compared to crime rates, we had to get median rent from a string to a float type. From there, we chose to use a plot to visually show the spread of data. A t-test was used next as a way to show if crime rates and median rent were correlated. Initially, it was noticed that there was a piece of data that was very off to the right away from the main spread of data. We noticed that the crime rates were extremely high for this one particular zip code. While diving in deeper we found out there were many empty places away from the public eye where crimes could be committed easily. We had to take that zip code out of the dataset to properly do the t-test.We also needed to allow for median rent to have smaller groupings than the first go around to be able to more easily see the trend lines of crime rates and median rent. The next analysis to compare 2000-2012 home value change with crime rates, the "-" symbol needed to be removed in the column with 2000-2012 data in it to be able to access the data without python thinking something was being subtracting. Next we needed to convert the percent of home value from a string to a float to be able to graph the data. We graphed it to visually show if there was a correlation between the speed of home value increase and crime rates. Initially, we thought a Pearson correlation would work. However, we didn't see any clear trend lines. We did notice the graph showed that a 40-60% increase in value had a much larger amount of crime. We originally thought this was an anomaly, but after some research we found out that the average home in Austin during that time frame averaged 58% increase. Therefore, we took The average of the speed of home value increases from 2000 to 2012 because if the average was greater than 58% it could be said that the speed of home value increase was correlated to crime. From the data, there does seem to be a strong correlation.

To prepare the dataset for additional analysis, we had to clean the data a bit. This involved converting the 'Report_Date' from a string format to a datetime object to facilitate time-based analysis. Furthermore, we addressed missing values and anomalies, such as zero values on dates that do not exist in certain months. The dataset was enriched by extracting additional features like the day of the week and month from the 'Report_Date', which were instrumental in the analyses.

The choice of analysis techniques was directly aligned with the complex nature of the dataset and the objectives of the project. Historical trends were visualized using line plots, which are well-suited for illustrating time-series data, allowing stakeholders to observe patterns over time. Heatmaps were employed to illuminate the spatial distribution of crime incidents, capitalizing on their ability to showcase densities and gradients effectively. These techniques were chosen for their robustness in handling large datasets and their utility in revealing underlying patterns that are not immediately apparent from raw data.

Throughout the analysis process, there were initial challenges that needed to be worked through. We experimented with different color scales to enhance the readability and interpretability of the heatmaps, eventually settling on a red color scale that appropriately indicated the intensity of crime occurrences. Additionally, the y-axis of the heatmap was inverted to align with the conventional calendar format, improving the intuitiveness of the visualization.

An alternative approach that could have been considered is the use of machine learning algorithms, such as clustering, to identify crime hotspots. However, given the exploratory nature of our project and the need for clear, immediate visualization, the chosen techniques were deemed most appropriate. Future analyses could incorporate predictive modeling to forecast crime trends and assist in proactive resource allocation.