Marijuana Related Crimes in Denver

Team Members: Farouk Ghandour, Antonio Martinez, Melissa Olivera, Omar Jamil

Github repository

Link to dataset

Link to presentation slides

Introduction/Purpose:

The rise of marijuana across the country has left people with many questions, and a city in Colorado might have the answers we seek. There has always been a relationship between marijuana and crime, but the environment has changed. Many cities are coming to grips with the reality that Marijuana legalization is coming nationwide soon. Now is the time to look at Denver, one of the first major cities in the nation to see legalization take hold. After acquiring years of crime data from all over Denver, we can paint a bigger picture and glean important insights into the criminal world and how various factors affect crime, business, and families. The dataset that we will look at is the Denver Police department database on marijuana-related crime list that started recording on January 3, 2012, and this particular subset ended on October 29, 2016. This gives a large set of over 1200 incidents that range from robbery and theft to crimes of unlawful discharge of a weapon. The purpose of this research is to explore the crimes in a large city and characterize them so that others can better combat marijuana crimes in their cities now and in the future.

We will initially look at where the crime happens and determine if variables make those neighborhoods more susceptible to crime. Following that, we can once again dig deeper into the afflicted neighborhoods and find out what time of crime happens there. Finally, we ultimately observe the rate of marijuana crimes over time after legalization.

Research Questions:

New cities looking ahead to legalization would want to know the answer to various questions. This includes questions below that we will work to answer through exploratory data analysis: We will begin with a high-level analysis to understand the high-level trends in our data and then dig deeper into some of the notable trends we identify.

1. Which neighborhoods have the highest concentrations of crimes?
    a. We will begin by analyzing overall neighborhood crime and then determine how crime by neighborhood has changed over time after the legalization of marijuana.
    b. This would allow city officials to predict where in their neighborhoods they might see an uptick in crime when new marijuana laws take effect.
2. Which Marijuana crime is most common, and is it correlated with location?
    a. To analyze marijuana crime, we will create a high-level ranking of all crime incidents. We will then identify if the type of crime correlates to the incident's location.
    b. Digging deeper into the locations of interest, there could be additional correlations with specific crimes being committed in those areas.
3. Are industry or non-industry-related crimes higher than one another? Which type of crime is most common in each Marijuana relation type category?

a. Initially, we will explore overall industry vs. non-industry related crime trends. We will then dig deeper into this by determining if there is a correlation between the type of crime and the marijuana relation type category.
b. By splitting the crime incidents into industry vs. non-industry, we can filter out crimes committed to either legal marijuana businesses or other entities.
c. Finding commonalities in the crimes committed by specific industries can better prepare businesses in cities for the type of crime they will face.

4. Is there a correlation between a neighborhood's median household income and the number of Marijuana-related crimes in the area?
    a. Using the median household income ACS Census data we have acquired, we will create visualizations to help us determine if there is a relationship between the neighborhood's household income and crime rates.
    b. Together with question 1, we can pinpoint possible contributing factors for the crime propensity in some neighborhoods versus others.
5. Did Marijuana-related crimes decline over time?
    a. By utilizing time-series analysis, we can analyze the historical trends of crime rates in Denver and determine if there has been a significant change over time.
    b. This would suggest that the legalization of marijuana since 2012 has, for a city like Denver, lessened the amount of marijuana crime.

Context and Datasets:

Main Dataset:
Data is of crimes reported to the Denver Police Department, which, upon review, were determined to have a clear connection or relation to marijuana. These data do not include police reports for violations of restricting the possession, sale, and or cultivation of marijuana. This dataset is based on the National Incident-Based Reporting System (NIBRS), which includes all victims of personal crimes and all crimes within an incident.

There are 1255 crimes listed, and there are 14 variables included:

INCIDENT_ID: Unique identifier of the police report which documents the crime
FIRST_OCCURENCE: The first possible date/time when the crime could have occurred.
LAST_OCCURENCE: The last possible date/time when the crime could have occurred.
REPORTDATE: The date when the report was taken.
INCIDENT_ADDRESS: The location of the reported crime
GEO_X: Geographical location of reported crime.
GEO_Y: Geographical location of reported crime.
DISTRICT_ID: The police district where the reported crime occurred
PRECINCT_ID: The police precinct where the reported crime occurred.
OFFENSE_CODE: The Uniform Crime Reporting (UCR) code
OFFENSE_TYPE_ID: Text description of UCR code
OFFENSE_CATEGORY_ID: Aggregate category used for similar grouping crimes
MJ_RELATION_TYPE: The type of relation to marijuana in the reported crime after review.

- Industry-related crimes involved marijuana and licensed marijuana facilities. These reported crimes are committed against the licensed industry or by the industry itself.
- Non-Industry - crimes reported where marijuana is the primary target in the commission of these crimes, but marijuana has no apparent tie to a licensed operation.

NEIGHBORHOOD_ID: Aggregate category to identify a location within the city.

Data Reconciliation and Sanitization:

Before analyzing the data, we first performed an initial data exploration and sanitization effort using python. Our initial assumptions prior to this exercise, consisted of the following:

1. All 14 columns were usable
2. Last_occurence_date was equal to First_occurence_date for incidents that consisted of only one offense
3. Incident_ID is the unique primary key
4. Offense_Code has a 1-1 relationship with offense_type_id

To start, we imported the file into pandas and confirmed that all 1,254 rows were imported correctly. We then checked for any duplicates, for which we found none. The type was initially set as a string for the three date columns: First_Occurence_Date, Last_Ocurrence_Date, and ReportDate. Each column was then transformed into DateTime type, and we ensured that they all had consistent “dd-mm-yy” formatting to be easily compared. Additionally, to understand the nature of the Last_Ocurrance_Date column, we created a calculated column consisting of the difference between the first and last occurrence dates:

```
0 days      809
1 days      151
2 days       13
3 days        6
6 days        2
11 days       2
42 days       2
4 days        2
8 days        2
9 days        2
51 days       1
137 days      1
187 days      1
76 days       1
185 days      1
124 days      1
149 days      1
16 days       1
37 days       1
21 days       1
dtype: int64
```

Figure 1: Difference between First_Occurence_Date and Last_Occurence_Date grouped by count

Based on these findings, we decided it was too hard to discern the significance and relationship between these two columns without proper documentation; we could not find any specific patterns. Due to this – and that Last_Ocurrence_Date had 253 null values - we opted to assume that ReportDate corresponded to the date when the incident was reported and ignore both occurrence date columns.

We then performed a quick check on the offense-related columns. First, we set out to confirm that each offense code pertained to one offense_type. However, we found that when we grouped the values of these two columns by their specific count, they had similar values but did not match exactly:

```
OFFENSE_TYPE_ID
BURGLARY - BUSINESS BY FORCE     587
ROBBERY - STREET                  80
CRIMINAL MISCHIEF - OTHER         78
THEFT - OTHER                     73
BURGLARY - RESIDENCE BY FORCE     70
BURGLARY - RESIDENCE NO FORCE     50
THEFT - SHOPLIFT                  29
ROBBERY - RESIDENCE               28
THEFT - ITEMS FROM VEHICLE        28
BURGLARY - BUSINESS NO FORCE      26
AGGRAVATED ASSAULT                25
THEFT - FROM BLDG                 21
ROBBERY - BUSINESS                19
DRUG - MARIJUANA SELL             18
CRIMINAL TRESPASSING              17
THREATS TO INJURE                 15
ASSAULT - SIMPLE                  15
```

```
OFFENSE_CODE
2203    590
2999     82
1205     80
2399     73
2202     71
2204     52
1315     33
2303     29
1208     28
2305     28
2205     26
2308     21
1202     19
3560     18
1313     18
5707     17
1316     15
```

This implies that the relationship is not 1-1, and based on the potential unreliability of the offense_code column, we decided to ignore the offense code grouping completely.

Throughout our initial exploration, we also found that the Incident_ID column was not a unique identifier:

```
2013471031    5
2014677313    4
2013282971    3
2014454249    3
2015288724    3
             ..
2012293739    1
2012170316    1
2012124878    1
2012129224    1
2016692084    1
Name: INCIDENT_ID, Length: 1196, dtype: int64
```

However, by double-clicking on one particular key, we found that a particular incident can be associated with multiple offenses and updated our assumptions accordingly. We could confirm this because all the attributes were identical except those of the offense type.

Finally, to complete the sanitization of the data, we removed all '/r' occurrences in the MF_Relation_Type column and omitted the one row where GEO_X and GEO_Y had null values. The data was then exported to a CSV file.

We also had an initial intention to utilize the Geo-x and Geo-y provided to us in the dataset to generate a map-based visualization. However, we quickly realized that these columns referenced polar coordinates and needed to be more functionally useful to produce an effective map visualization. Our goal was thus to create a set of longitude and latitude coordinates based on the address column provided. Using the Geopy library, we created our longitude and latitude columns to produce our map and dropped the Geo-X and Geo-Y columns as they were no longer helpful. After getting the longitude and longitude coordinates, we had to set the dimensions for the graph plot to plot the coordinates. After we had the maximum and minimum longitude and latitude, we exported a map with these dimensions and plotted each incident using a scatter plot.

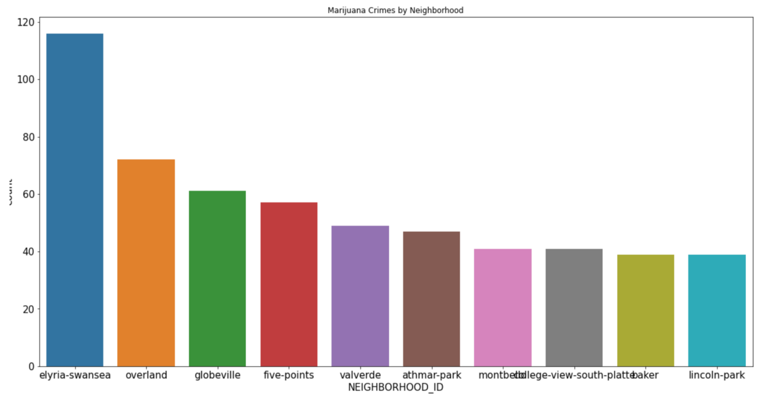
Additionally, we have consolidated data from ACS Census data (medhhincome.pdf (denvergov.org)) consisting of Denver's median household income categorized by neighborhood. The data has been processed into an excel table. The neighborhood data has been formatted consistently to match the original data set. Once joined to the primary dataset, we can visualize the number of incidents by neighborhood and compare them against the corresponding median household income. This will help us analyze the correlation between these two variables.

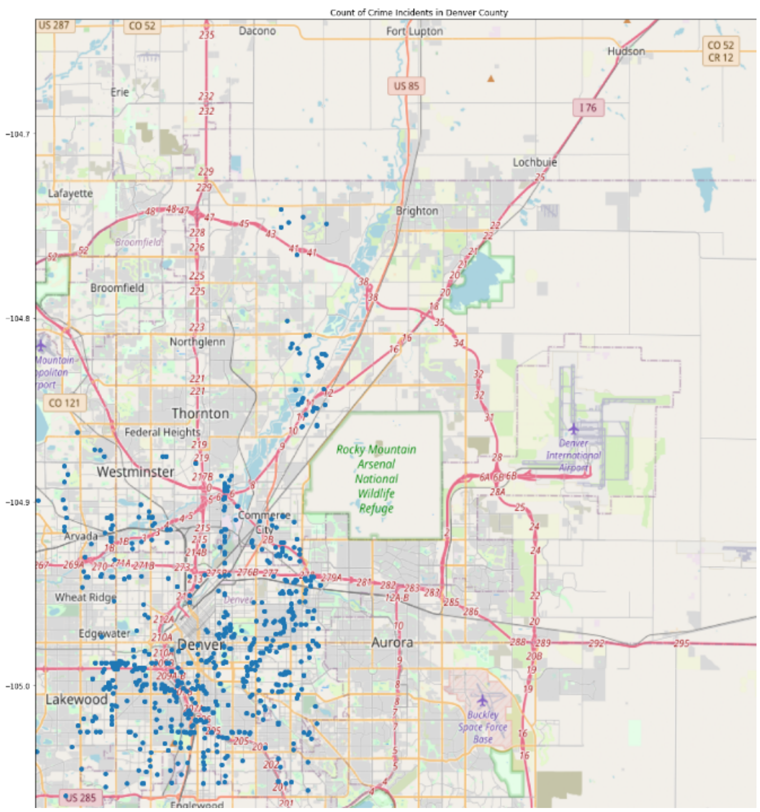
After completing the data preparation, we began our data exploration to identify and summarize high-level trends.

Plots & Visuals

Crimes by Neighborhood

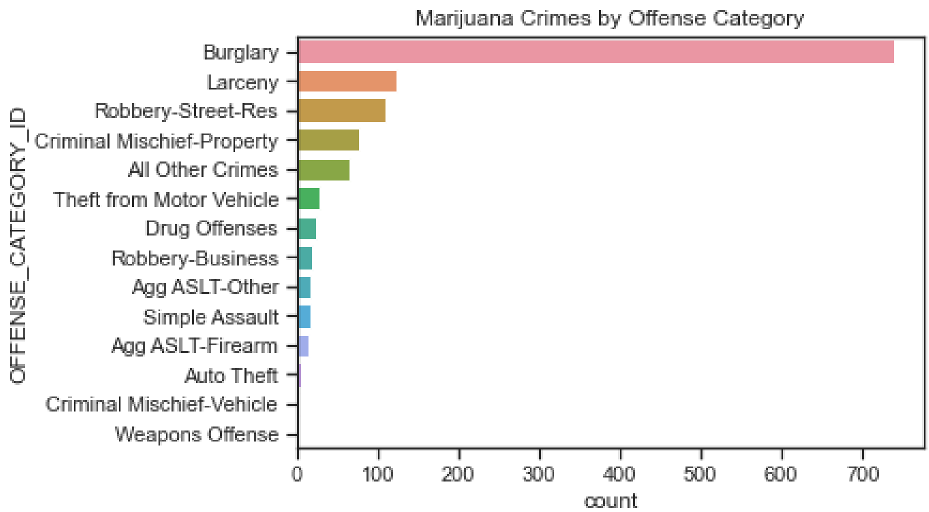
Regarding location, the majority of the crimes occurred in the Elyria-Swansea neighborhood. In order to get a better geographical sense of the data, our team plotted the coordinates of the crimes using geopy on a map. As a result, we identified that most crimes occurred just east of Denver, in neighborhoods such as Elyria-Swansea, Overland, and Globeville.



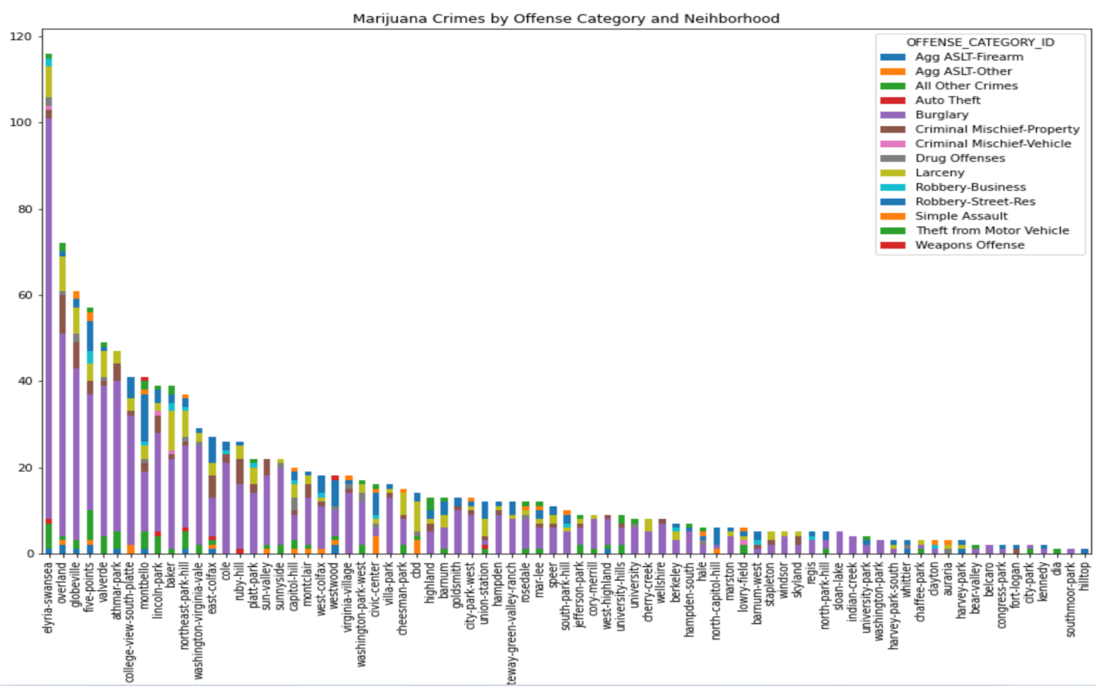


Crimes by Offense Category

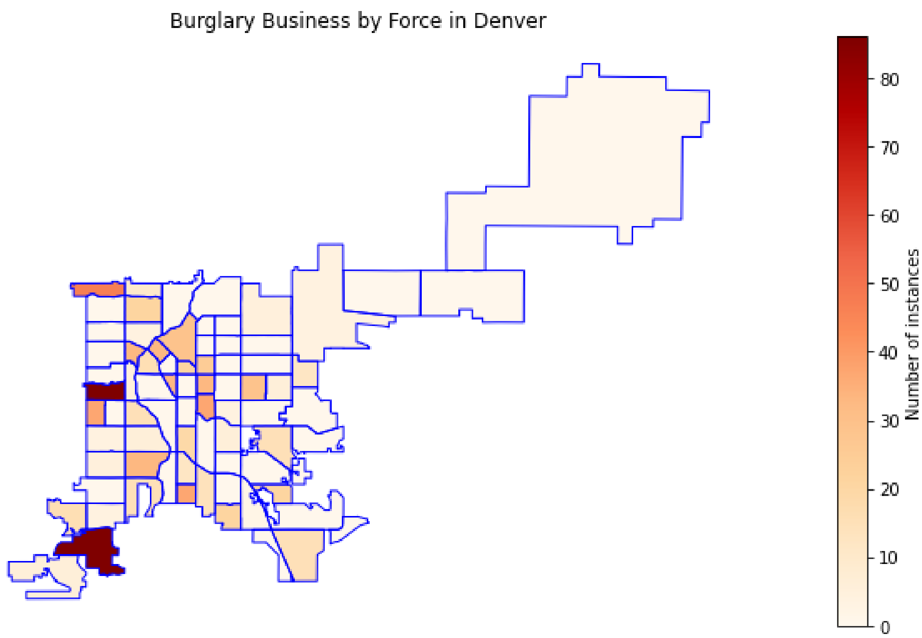
At a high level, the category with the highest amount of crimes is burglary followed by larceny and robbery. We will dive deeper into the data and assess if there is a relationship between types of crimes and neighborhoods.



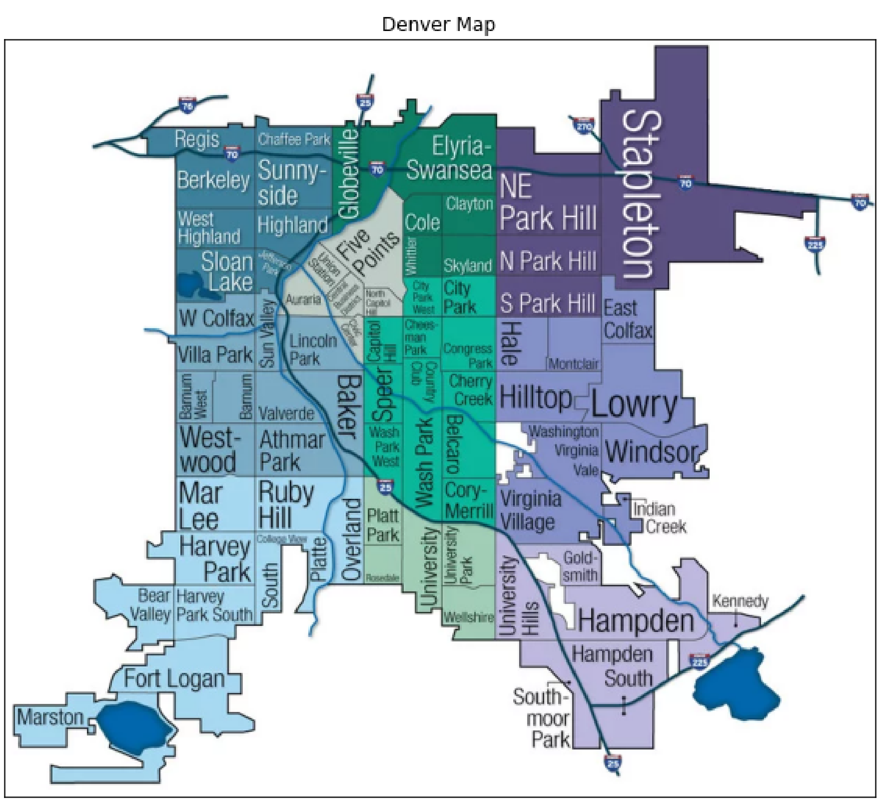
There is no correlation between the neighborhood and the type of crime. Burglary was the overwhelming type of crime per neighborhood. Larceny and theft are the second and third most common types of crimes, but they are far less common than burglary. This goes against our initial hypothesis, as the location in Denver does not seem to impact the type of crime committed.



The most common Marijuana crime was Burglary of a Business by Force

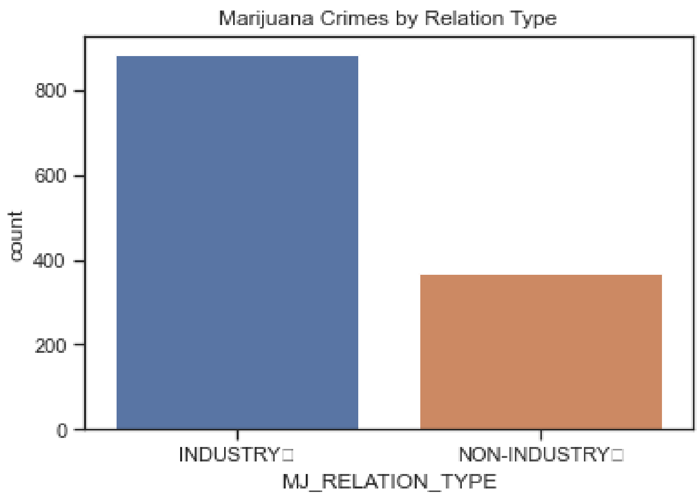


The heatmap shows that the only correlation between forceful burglary of a business and the neighborhoods is that the neighborhood near downtown Denver is more likely to be impacted. This also correlates with their location on major highways but does not show any other correlation between burglary and location. The one noticeable outlier is the neighborhood of Fort Logan, which is known to be one of the neighborhoods with the highest overall crime rate in Denver.

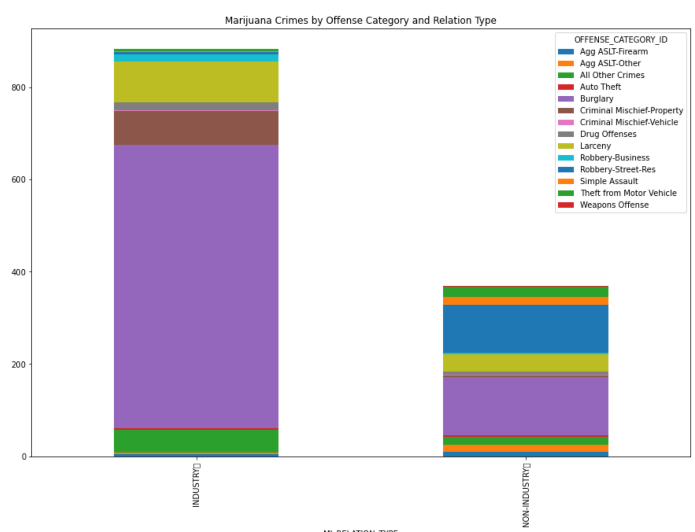


Crimes by Relation Type

Regarding relation type, we analyzed industry vs. non-industry related crime at a high level, below. Industry related crime was much more common than non-industry related crime.

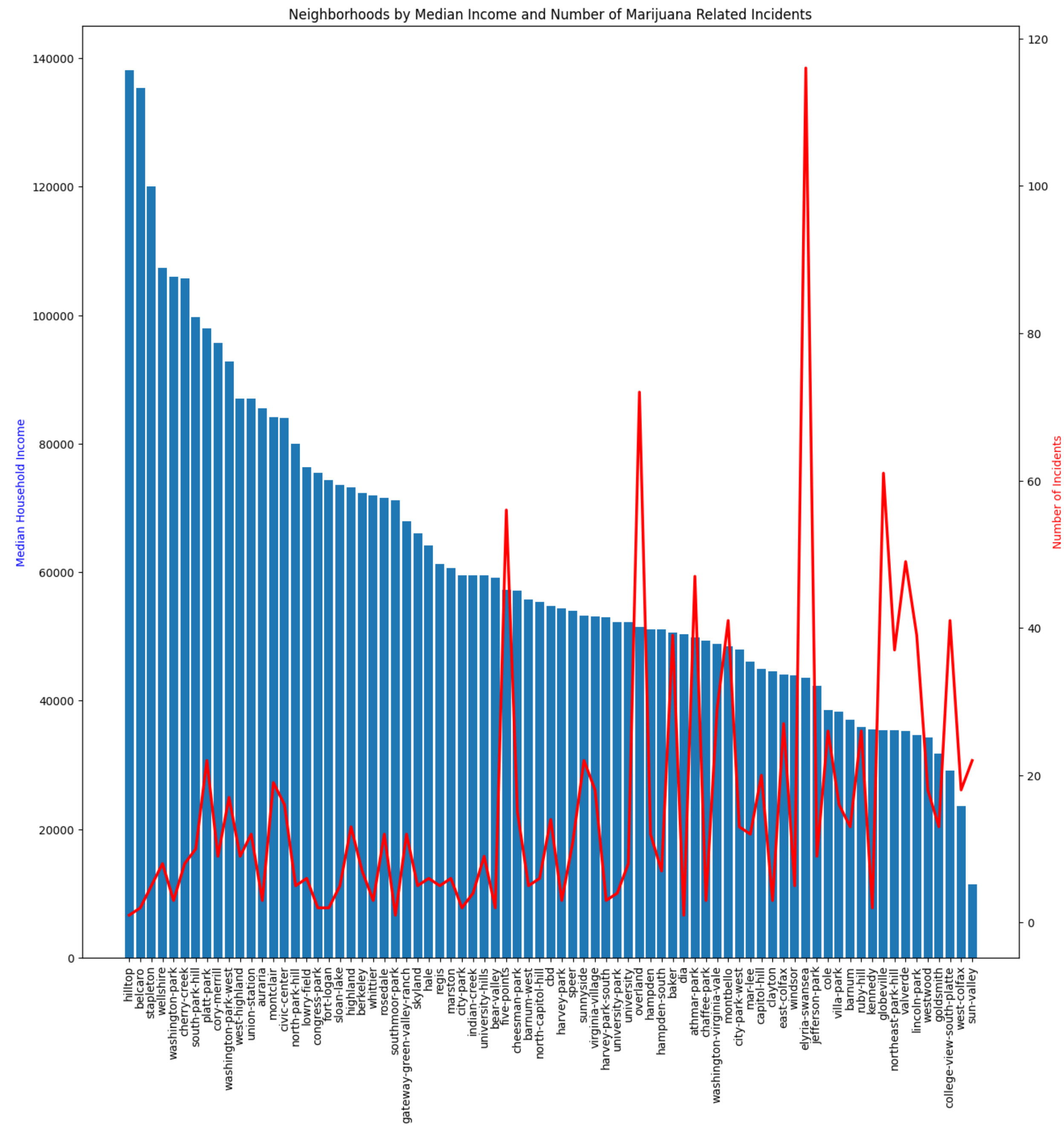


As we dive deeper into the data and see the relation type broken down by offense category, we can see that industry marijuana (such as dispensaries) is subject to many more crimes than non-industry. For example, although burglary is common in both, it seems that non-industry related crimes receive much more severe types of crimes, such as aggravated assault by firearms.

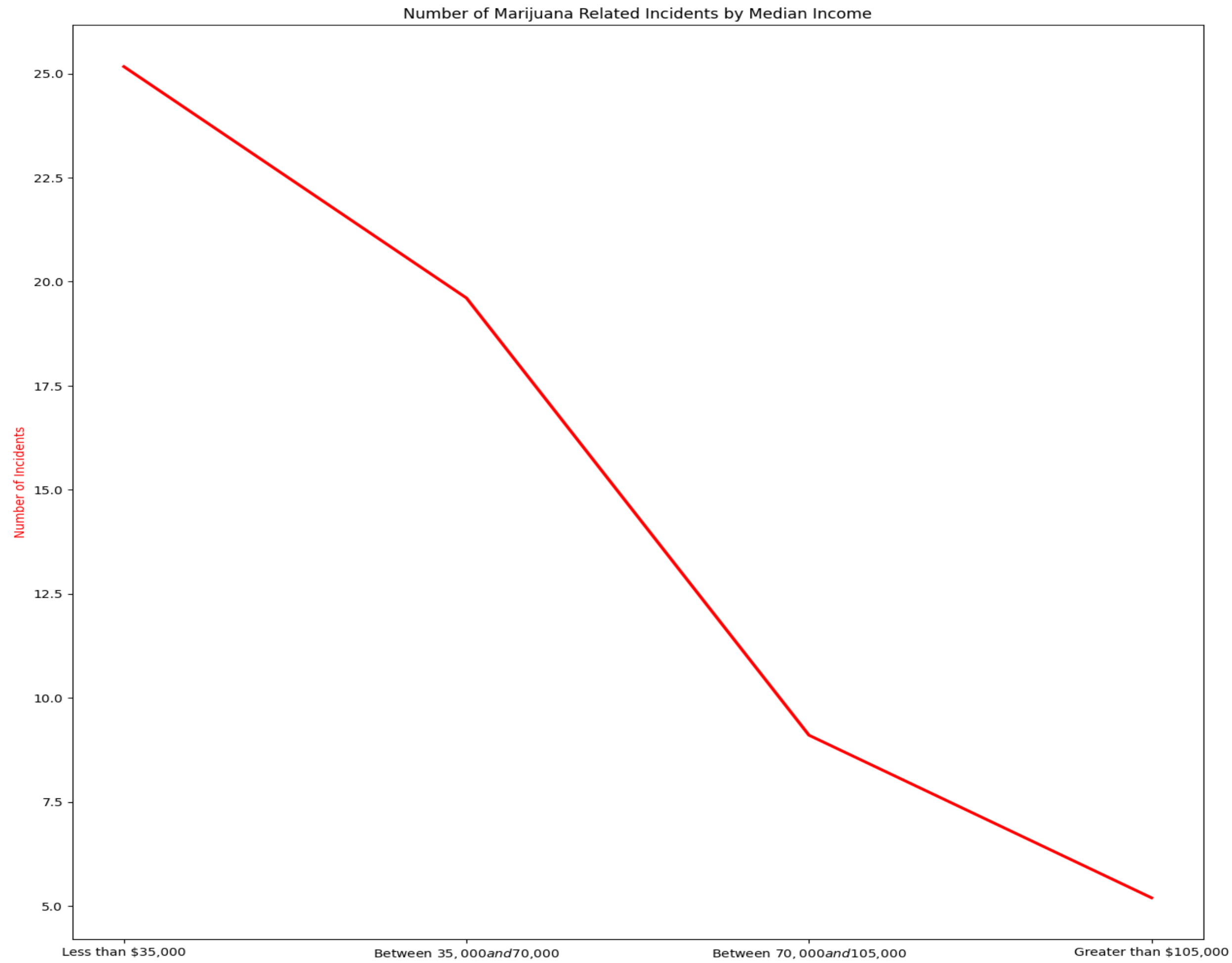


Crimes by Median Household Income

To answer question 4, we plotted the graph shown below with two y-axes: one for the median household income and another for the number of marijuana-related incidents. The shared x-axis represents the neighborhood in Denver. While the graph does seem to indicate that the number of incidents is more prevalent in neighborhoods with lower average household incomes, we decided to categorize our independent variable further and group the x-axis to assess the correlation better.

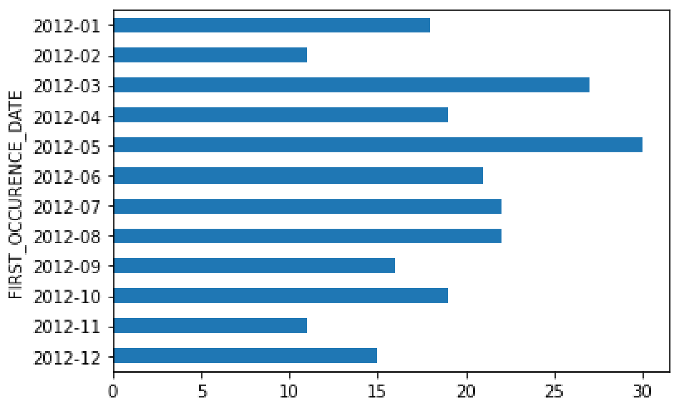


This graph (shown below) provides a much more unambiguous indication of the correlation between Median Household Income and the Total Number of Incidents by neighborhood. Neighborhoods that earn less than $35,000 per household have an average of 25 Marijuana related incidents, while those that earn more than $105,00 average only 5 Marijuana related incidents. The graph suggests a strong inverse relationship between both variables without carrying out a statistical analysis. Burglary is the more prevalent type of offense within poorer neighborhoods.

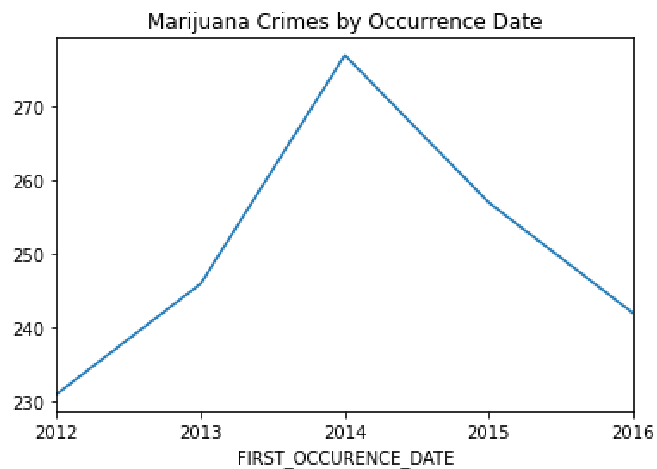


Crimes Over Time

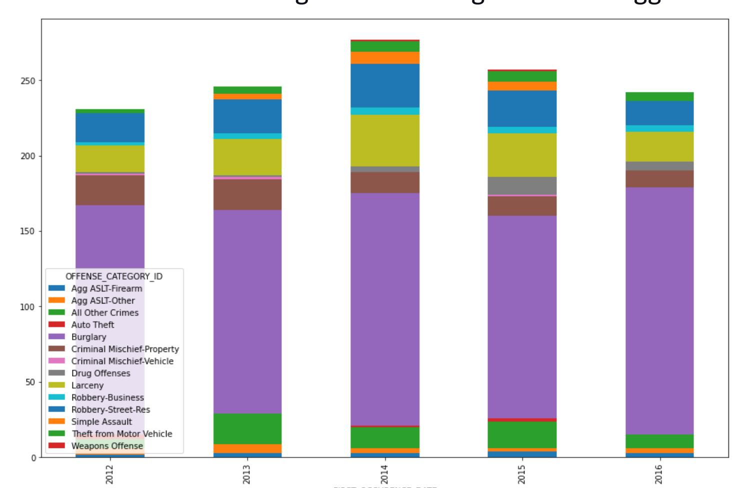
In the first year of marijuana legalization, marijuana-related crime counts changed in the months following legalization. For example, the below visualization shows that after November 2012, marijuana-related crimes dropped significantly compared to previous months. This trend, however, was not consistent past 2012.



Below we can see that from 2012-2016, overall marijuana-related crime counts increased initially and then decreased after 2014.



Diving deeper into crimes over time by offense category, we can see there was an important shift that occurred. Prior to the legalization of marijuana, the top crime categories were burglary, aggravated assaults, and larceny. In the years following legalization, these remained the top crime categories, however there was a larger shift to burglaries from aggravated assaults.



Conclusion

Our team set out to analyze marijuana-related crime in Denver to answer five main research questions. While the data offered great insight into the related crimes in this area, we were challenged by the need for proper documentation. As a result, we had to rely on assumptions to perform an adequate analysis. These assumptions were refined throughout our initial data exploration, and ambiguous attributes were ignored for the sake of reliability.

Regardless of our challenges, we leveraged various python data science libraries to transform and visualize the dataset and answer each research question. Our findings indicate that most marijuana-related crimes were concentrated in east Denver. Furthermore, these crimes - which predominantly consisted of burglary, followed by larceny and then Robbery - were shown to happen more in industry-licensed facilities than in non-licensed facilities. However, it is hard to assess this accuracy without understanding how often non-licensed related crimes go unreported.

By leveraging our subset dataset, we were also able to evaluate the possible correlation between average income and the amount of crime for each neighborhood in Denver. A strong inverse correlation exists between economic affluence and crime in a specific area. While economic instability and an overwhelming desire to meet basic needs might contribute to why we see such a correlation, it is hard to conclude this without having more information on the perpetrators themselves.

Finally, we used a time series to assess the historical crime rate trends and determine if legalization has lessened the amount of Marijuana crime. While immediately after legalization, we can observe a drop in Marijuana-related crime, the years to follow had more crimes reported than the data we had for the pre-legalization era, for which we only had ten months' worth of incidents. However, the increase in crimes that followed Marijuana legalization might be driven by increased reported incidents as victims become more comfortable with involving the police in these types of crimes.

Within the limits of our data, we have been successful in gathering insights to answer each one of our research questions. We also have identified opportunities to expand the scope of this research project further and gather additional supplemental data to provide more certainty in our findings.