**Group C PART2 (Chicago Crime 2023 – dataset)**

We began our search for a dataset, large and diverse enough to satisfy our curiosity, on Kaggle, since we had heard of that website from university. On Kaggle we scrolled and scrolled eventually coming across a fateful link. Said link led us to the City of Chicago’s crime data for the years of 2001-2024. This dataset was far too large for our meagre skillsets and, admittedly, too ambitious for our humble endeavor. After managing to pull out a couple solid handfuls of hair, we unanimously decided to limit the dataset to only the year of 2023 A.D.

The reason for our intrigue into Chicago and its lesser-than-ideal crime statistics, was due to the shear shock we felt when realizing that even though us and Chicago were separated by kilometers upon kilometers of land and sea, we were still just a mere click away from seeing the somewhat accurate location of a specific crime and ALL its details! This got us to ponder about the true meaning of privacy in our current digital age, and how close big data and the all-seeing eye truly is. After recuperating from our abhorrent realization, we got to analyzing the beast that is Chicago’s crime data.

To begin we devised a mighty strategic plan to hack and slice away at the data to uncover the hidden and despicable truths that lay within. We divided this mischievous plan into five main operative level categories that we further reduced to tactical level battleplans capable of ducking and weaving through the worst that Chicago could throw at us, or at least so we thought. The following are the operative and tactical categories that were arduously sifted out and selected (remember that these are national secrets, that few are privy to): (1) Geospatial Analysis equipped with Crime Hotspots and Neighborhood Comparisons to boot; truly a remarkable combination. (2) Crime Type and Frequency and the legendary duo of Crime Categories and Relationships Between Crime Type; something even Genghis Khan would’ve struggled against! (3) Arrest and Law Enforcement Analysis, with the remarkable Arrest and Law Enforcement Analysis, that even Napoleon couldn’t best! The last two are so legendary, I dare only mention them shortly, as not to cause unnecessary commotion; (4) Time-Series Forecasting with Seasonal Variations and Anomalies, and (5) Weather and Environmental Impact Analysis with Weather Correlation. Huh, I can already feel the hairs on the back of my neck rising from the mere thought of these monstrous and certainly effective strategies and tactics!

Some additional information about those-that-shall-not-be-named:

**Geospatial Analysis**

(Crime Hotspots and Neighborhood Comparisons): We brainstormed this in the early stages of planning, since we wanted to focus on the most relevant and interesting things we could figure out about this data, so we decided to immediately check if the downtown areas really were the crime dens of Chicago. – Long story short; they were.

We used latitude and longitude to create a heatmap, showing areas with the highest crime rate. With these hotspots we analyzed and compared crime rates across different neighborhoods.

This analysis we were able to accomplish with PowerBI’s built in map system, and that, without the need for a lot of input, was able to display what we wanted it to display. The only noticeable issue that did end up taking a while to fix was that there were some missing locations within the dataset, and Power Query wasn’t dealing with them properly, for whatever reason. The map did end up a bit ugly, but its beauty wasn’t exactly the “numero uno” trait we were looking for from it. We also had another map for the different neighborhoods in Chicago, and we were trying to connect them with the crime data we had, but we couldn’t get it to display properly, so we had to ditch that part and focus on other aspects of our report. This neighborhood specific crime data could’ve been very intriguing because lots of conclusions and opinions could’ve been drawn from it, though that also makes it one of the most common and attractive ways to display crime data, so we didn’t really mind too much about losing out on it.

**Crime Type and Frequency**

(Crime Categories and Relationship Between Crime Types): Since crimes like theft, battery, and criminal damages, the specific crimes we focused on in this part, and the three most popular ones to report, are crimes that can be done anywhere, we were interested in seeing if there was a clear “go-to” location for would-be criminals to do their deeds.

We investigated the distribution of different crime types, such as theft, assault, burglary, etc. This could highlight which crimes are most frequent and possibly most affected by policy or socioeconomic factors, however we didn’t include in our research any external factors such as policies and the aforementioned factors, since we were focused on analyzing the data in front of us. Additionally, we wanted to investigate if there’s a relationship between different types of crimes, such as whether certain crime types tend to increase or decrease together. We tried to accomplish this with the CORRAL function, but it doesn’t have native support in PowerBI and is only contained within Excel. We did end up creating a small program (PowerBI measure) to do this covariance analysis for us, but we couldn’t get it to work in time; there was something wrong with the dataset we were working with, and we were in the process of fixing it.

The analysis was fairly easy as the database was well constructed, thus filtering and putting all required information into charts was manageable. We have managed to put the most common crime types based on the arrest rates, as well as the percentage of reports that lead to arrests based on crime types. Thanks to that, we were also able to conduct detailed analysis based on descriptions of crimes.

**Arrest and Law Enforcement Analysis**

(Arrest Rates by Crime Type): When dealing with reports, you must remember that you are dealing with just that; reports, so obviously we wanted to know the true number of arrests that police officers must carry out, within Chicago. The arrest rate for crimes also can be a very interesting topic, since the factors that go into it are truly broad, as is the case for most data fields.

We analyzed the arrest rates for different crime types to see which types were the most and least likely to lead to arrests.

We conducted the analysis by filtering the cases that lead to arrests using filters. In our database, cases that led to arrests were marked with “true”, and those that didn’t were marked with “false”. This demographic varied a lot based on crime types.

**Time-Series Forecasting**

(Seasonal Variations and Anomalies): When dealing with crime statistics, overarching themes like policies and laws are often discussed, but immutable facts, such as the seasons and, weather, are seldom mentioned, since they aren’t as hot topics and go away by themselves. We wanted to see if the time of the year truly had a visible effect on crime rates. Though due to our analysis only being on the year 2023, we couldn’t get any conclusive evidence, however, we did get some interesting observations.

We examined seasonal spikes, like potential increases in certain types of crimes during summer months, and we investigated the time of day most of the reports were flooding in.

To do this, we had to download and include the database from [this website](https://www.visualcrossing.com/weather-data?gad_source=1&gclid=CjwKCAiAjKu6BhAMEiwAx4UsAkA0j8nidAqL1-Q-tvwgq9M42oCj5LYvb87ed0lFXOOZXqH9ayEuCBoCdjwQAvD_BwE.). After downloading the database in CSV format, it was necessary to connect two files for them to have a connection in visualization. We have included data like weather conditions and temperature on which the crime was committed. We have also included crimes committed by hour. We did it by putting both the time and count of reports in the chart, thus making it possible to see that most of the crimes were reported in the nighttime.

**Weather and Environmental Impact Analysis**

(Weather Correlation): Our logic behind choosing to analyze this was almost identical to the previous part – Time-Series Forecasting – as it stemmed from the same root, curiosity about nature and its effects on crime. Interestingly enough, it did seem to have a very strong correlation, but due to it, again, being only on the year 2023, we couldn’t get any conclusive evidence. For me (Liam) this part of the analysis was probably the most interesting to observe, purely due to how large of a correlation the dataset provided.

By integrating historical weather data (e.g., temperature, precipitation) with the crime data, we analyzed how weather conditions affected crime rates.

To do so, we have implemented the same steps as in the last part, getting data from the same website and importing it to our PowerBI file. That way, we were able to analyze and present how weather conditions and temperatures impact the crime rate.