BOSTON CRIMES

Sofya Denoun – Félix Poirier – Bhumana Prannoy – Ismail Zizi – Michaël Ullah

Context

Crime incident reports are provided by the Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This dataset contains records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident (I, II and II, corresponding respectively to crimes, offences and infractions) as well as when and where it occurred. Each row represents a crime report, including the type of crime, date and time, location and other details. Records begin on June 14th, 2015 and end on September 3rd, 2018.

Inspiration

We first wanted to get a good feeling for the crime situation in Boston, answering some key questions: What types of crimes are most common? What different types of crimes are most likely to occur? Does the frequency of crimes change over the day? Week? Year? We used extensive visualizations, including geographical ones. An analytical work is also performed in the aggregation of the crimes per period of time/district.
We then will forecast crimes. For this, we will use Python's library Facebook Prophet. We aggregated the forecasts per district and gravity of crime, and displayed the results on interactive maps for several forecast horizons.

Story Telling

To guide the presentation, we considered that our work will be useful to a superhero – Batman – to plan his interventions.

EDA Conclusions

1. The most frequent "crimes" in Boston are Motor Vehicle Accidents while the second most are Larcenies.
2. Serious crimes are most likely to occur on Friday and least likely to occur on Sunday.
3. Serious crimes are most likely to occur in the afternoon and evening.
4. Guns are mostly used in Aggravated Assault – followed by Homicides.
5. Holidays and temperature have a big impact on the number of crimes committed.
6. Crimes keeps on increasing every year by 1%.
7. Serious crimes are most common in the city center, especially districts A1 and D4.
8. Dorchester, Roxbury, and South End districts have the highest cumulative incidents across the dataset.
9. B2 is the most affected district.
10. Criminal Districts distributions are homogeneous among all types (gravity) of crimes.

Model and Predictions of crimes

A short-term time-series-analysis is used to predict the time and location of the crime. The prediction length is 1 month. We used standard linear regression to predict crime in each district and for each UCR type (Uniform crime reports types). We also computed and plotted Autocorrelation and partial autocorrelation for top 12 districts along with the UCR part.
We then used a powerful time-series library to perform the forecast: Facebook Prophet. We used short-term predictions as an insight for crime prevention. We will subsequently include considerations about confidence intervals.