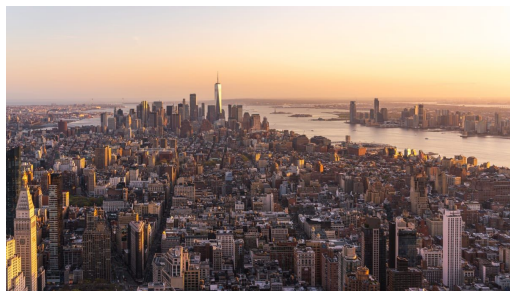# What do taxi trips and crime levels have in common?

Maria Kesa

February, 2023

Problem: Citizens need to be aware of criminal activity patterns in their environment to be on alert and safe

We would like to build a solution that alerts the taxi driver when he enters an area where there is a statistically higher chance of criminal activity during that day of the week.

Knowing risks associated with an area could lead to increased purchase of specialized insurance policies.

We use data from **Uber in New York** (taxi drop-offs) and **NYPD** (crime statistics).

NB! *Data from Uber is from 2014 and New York is*

since 2022, so this is a proof-of-concept data analysis!

Take the results with a grain of salt, imagine a hypothetical

world where data was collected at the same time with more rigour.

Average crime occurrences and taxi trips from a location exhibit non-linearity and are uncorrelated across clusters learned from latitude and longitude of events

Crime Count vs Taxi Count Per Day

Crime Locations by Cluster
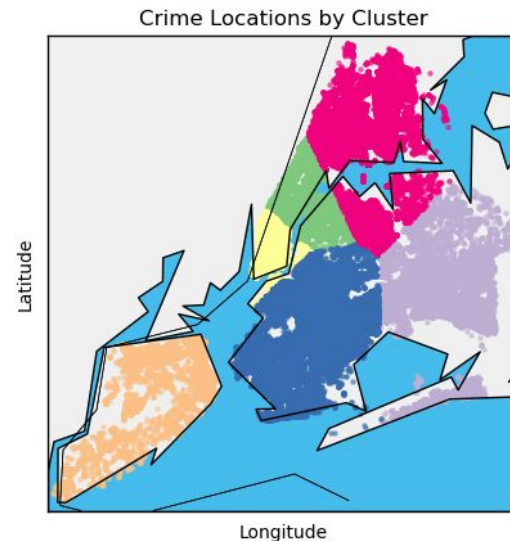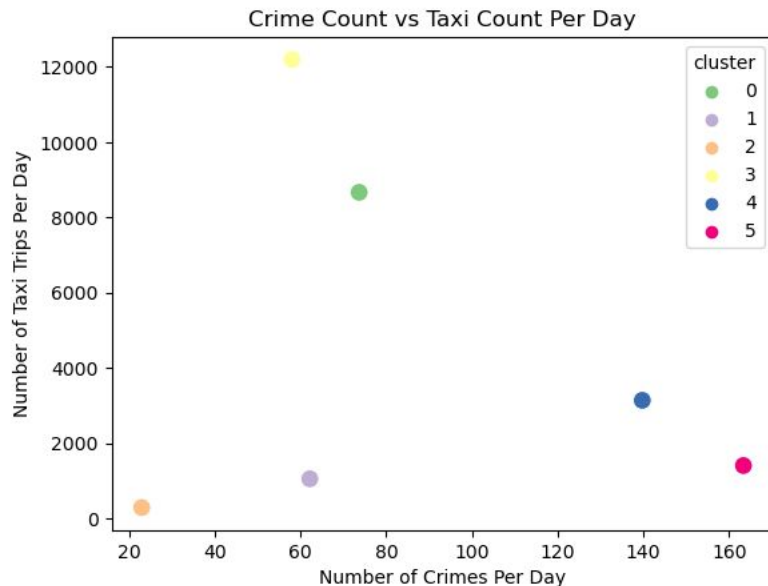
We clustered events in the taxi and the crime dataset jointly according to their geographical latitude and longitude. We then counted the number of occurrences in each cluster and normalized the crimes activity and number of taxi trips by the number of days over which each data set was collected.

This analysis revealed the presence of two groups–
 (1)   A group with intermediate crime levels and low taxi trips (Queens)
 (2)   A group with low number of taxi trips and number of crimes per day (Mid Island)
 (3)   and (0)   Intermediate crime levels, but high taxi trips (near Statue of Liberty)
 (4)   and (5)   Low taxi trips, but high crime levels (Brooklyn, Manhattan)

I made these plots myself:-)
Special thanks to ChatGPT!

Proposed Analyses

1.) In the previous slide we saw the aggregation of data across the entire measurement period, but we would like to achieve more granularity by jointly analyzing taxi trip data and crime statistics across weekdays.

2.) Another analysis that would further prime citizens to dangerous stimuli would be to find out if particular types of crime (robbery, murder, drunk driving) have any geographical clustering.

3.) Are there any seasonality patterns in crime aggregated across geographical areas (background danger level) and within geographical areas?

Value proposal

Intelligent behavior is impossible without awareness. Knowledge of risk permits safeguarding and risk pooling (insurance).

Organized and unorganized crime present significant risk to citizens.

It is important to stay safe and be aware of the risks that are associated with traveling and tourism.

Therefore, we propose an initial data analysis to explore the potential of open public data to help citizens make informed mobility decisions.

This analysis could be the start of a data app that displays risk of crime with time of week granularity and perhaps in the future even hourly risk (more difficult to collect and analyze). The Uber data that I propose to study contains taxi cab pick-up data, but it is not difficult to acquire drop-off data (Freedom of Information Act) which would further help citizens to stay safe.

Concluding remarks

→ Crime is a shifting target. Criminals have always found new ways to get around advances on the side of society and police.

→ Unintended consequences: Is it possible we will see a negative correlation between crime and taxi trips in geographic locations in the future? Would that be a good outcome?