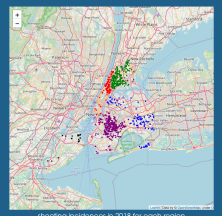
new york city shooting incidents

a wicked challenge

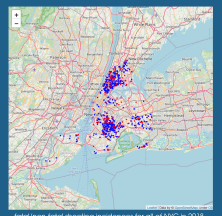
stefanie bienert [1], christopher patzanovsky [2]

[1] hasselt university, belgium
[2] hasselt university, belgium

background



shooting incidences in 2018 for each region

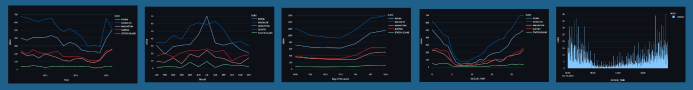


fatal/non-fatal shooting incidences for all of NYC in 2018

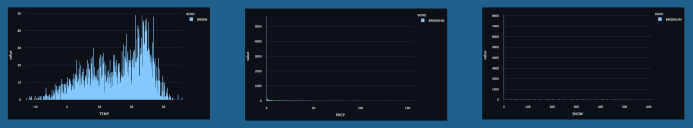
- Five regions of New York City:
  - Manhattan (1,694,263)
  - Bronx (1,472,654)
  - Brooklyn (2,736,074)
  - Queens (2,405,464)
  - Staten Island (495,747)
- Gun violence a clear issue in nyc/usa
- Citizens and tourists are afraid of being involved in shootings
- Governments looking for solutions to reduce shooting incidence frequency
- Democratisation of information on shootings, their location and associated information can inform stakeholders, policies and decision making from stakeholder

methodology

shooting incidences throughout the years, months, days of the week, hour of the day and individual minutes



weather related data



- time is clearly an important factor, from time of the year, time of the week to time of the day
- weather appears to have an impact, while exact correlations would have to be further explored
- this data needs to be taken into consideration for law makers as well as law enforcers:
  - smarter usage of police: higher precense during the night/on weekends/in the summer
  - potential reduction of police presence during worse weather
  - this leads to better budget managment
  - harder policies in hotspot areas (lockdowns during the night, higher gun control, etc.)

nyc during covid-19

- lockdowns not only fought the pandemic
- "breakout moment" after first lockdown
- zig-zags very typical for all curves, as police presence was most likely increased in those areas

Link to the Streamlit dashboard: https://bienerts-nyc-shootings-nyc-shooting-streamlit-l8l4ml.streamlit.app/

machine learning part 1: predictions of regions, zip code and race of the shooter/victim

things to consider:

- no covid data (2020 & 2021) would be included in the ML-algorithm, as the effects of the lock-downs and the following "breakouts" are too high
- one shooting may appear multiple times with the same unique incident key (multiple shooters or multiple vicitms)
- those would only be counted as one shooting here
- area codes would be extracted with the help of *geopy* for possible more precise predictions
- some data points had no information on the shooter(s), while the information on the victims was nearly entirely complete

split in two datasets

full dataset for:	partial dataset for:
- region prediction - area code prediction - victim race prediction	- shooter race prediction

decision tree & knn

- decision tree and knn used as machine learning algorithms
- only marginal differences within each method
- max depth of decision tree was usually found to be between 3-5
- knn would see a flattening of the curve after increasing the neighbors to around 20
- decision tree prediction of borough came out to be around 47.33%
- Zipcode < 5%, even with borough included only reached 11%
- race of the shooter was predictable with 76.87% accuracy with the decsion tree, 75.8% accuracy with knn
- race of the victim only slightly less at 73% and 71.4% for the decision tree and knn respectively
- A full dataset only provided a small increase in accuracy for the prediction of the borough at 47.51%
- Prediction of the race of the victim resulted in 73.91% accuracy
- Numbers for the race show a clear indication for ethnicity-related issues within new york city (potentially even the united states of america)
- more in depth analysis of the ethnic distribution will have to be conducted

machine learning part 2: predicting the probability of a shooting for a certain time/location in the form of a mobile appliaction

make new york city a safer place

- prototype of an app that takes user's location, time and date to predict probability of shooting in that area
- **problem**: dataset is filled with only shooting incidence data and not with non-shooting incidences
- create dataframe with every instance for every region/area code
- either every minute for the five regions, every hour for the five regions or every hour for the area codes
- **success!**

results

- Accuracy for area codes at 99.97%, not ideal due to lack of shootings per area
- Accuracy for region with every hour at 97.34%
- Example: probability for a shooting at 3 am on a Monday in the Bronx yields a 7.88% chance of a shooting happening within the hour
- Prediction for region on days of the week only yields a less accurate score at 69.86%
- But far more interpretable data with a shooting in Staten Island happening on any given Sunday being 13.18%

discussion

- clear correlations between time related data and shooting incidences
- covid-19 consequences had a massive impact on new yorkers
- race can be predicted with alarming accuracy
- predictive algorithm for location highly promising
- incorporation of hotels into application for user to seek shelter
- closer anaysis and incorporation of weather into predictive algorithm
- closer analysis of hotspots
- ethical challenges considering analysis of race (within regions)

references

Bienert, S., & Patzanovsky, C. (2022, January 5). New York City Shooting Incidents. GitHub. Retrieved January 6, 2023, from https://github.com/bienerts/nyc_shootings

Global trade, air homicide. (2019, 07 01). UNODC. Retrieved January 9, 2023, from https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html

Lanjewar, C. S. (2022, 08 17). Meteostat Python (Computer software). Meteostat Python Package 1.6.5. https://github.com/meteostat/meteostat-python

Mit. (2022, 11 13). geopy/geopy: Geocoding library for Python. GitHub. Retrieved January 9, 2023, from https://github.com/geopy/geopy

NYC Open Data. (2022, June 9). NYPD Shooting Incident Data (Historic). | NYC Open Data. NYC Open Data. Retrieved January 6, 2023, from https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8

python-visualization/folium. Python Data Science Meets Leaflet.js. (2022, 12 01). GitHub. Retrieved January 7, 2023, from https://github.com/python-visualization/folium

Streamlit. (2020, 1 5). streamlit/streamlit. Streamlit — The fastest way to build data apps in Python. GitHub. Retrieved January 9, 2023, from https://github.com/streamlit/streamlit

United Nations. (2022, May 9). The Sustainable Development Goals Report 2022. UN Statistics Division. Retrieved January 6, 2023, from https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf