Adesewa Adesida

Info 371

6/15/20

CURBING CRIME IN DETROIT

<https://colab.research.google.com/drive/1nNuCoXDd5u09Fcswbhb_pgs5C0AqE7iB?usp=sharing>

Issues

Weka was not loading a lot of my datasets, only one really. So, I did most of my analysis in google colab and as much as I could with what weka could load.

**ABSTRACT**

Comparing crime data gathered from Washington D.C and Detroit, Michigan, I discovered crime trends that occur in DC and Detroit and highlighted similarities. According to The Washington Post, three decades ago DC was named the “murder capital” of The United States. Being the political hub of the US, the capital suffered from terrorist attacks, homicides, bombing, even crime ridden streets. In 2019, DC was recognized as one of the safest largest cities in the world, scoring an 87.6 out of 100. According to Forbes, Detroit, Michigan is the current crime capital of the US with 2,137 criminals per 100,000 residents while in DC all crimes went down by 1,628 citywide during the past 2 years. Using the data sets from Kaggle and the Detroit portal, I was able to find patterns in the crime data in visualization bar plots. Hence, I have inferred from my study, that the most committed crime in the US is theft. Also, crime in the US is most frequent in the evenings.

**Introduction**

According to The Washington Post, three decades ago DC was named the “murder capital” of The United States. Being the political hub of the US, the capital suffered from terrorist attacks, homicides, bombing, even crime ridden streets. In 2019, DC was recognized as one of the safest largest cities in the world, scoring an 87.6 out of 100. According to Forbes, Detroit, Michigan is the current crime capital of the US with 2,137 criminals per 100,000 residents while in DC all crimes went down by 1,628 citywide during the past 2 years. Using the datasets from Kaggle and the Detroit portal, I was able to find patterns in the crime data in visualization bar plots. Hence, I have inferred from my study, that the most committed crime in the US is theft. Also, crime in the US is most frequent in the evenings.

**Objectives**

* Are the timings of crimes similar; for instance, if assaults are committed more in the morning in DC is that the same in Detroit.
* What are the peak crime hours?
* What are the more rampant crimes in Detroit?

**Techniques**

* Data sorting and organizing
* Data cleaning
* WEKA, cross validation
* Bar plots I used these to monitor the frequency of crime trends in the data, focusing on the 2008 to 2017 timeline.

**Results**

The DC Dataset

* peak crime time in DC from 2008 to 2017 is the Evening, then day, least is midnight.
* Theft/Other is committed the most while Arson is committed the least in DC, 2008- 2017.
* 2017 has the lowest recorded crime rate in the dataset The Detroit Dataset
* Crime in Detroit grew in 2016, but spiked greatly in 2017, before that crime rate was very low. -Crime in Detroit has been growing at an exponential rate since 2016
* Larceny is the most committed crime in Detroit, 2008 - 2017
* Detroit is most dangerous in the evenings Cross referencing both Datasets
* Theft is the major problem in both places.
* The evenings are dangerous in both places.

### Problem Definition

Crime is a major issue all over the world, likewise in the United States. Even with the police working hard, there are still many people that need help, especially in Detroit. By finding similar crime trends in DC (a safer large city) and Detroit (a crime capital), the police force in Detroit can learn which crimes applying DC crime fighting tactics would provide beneficial results. This project is intended to help curb crime in Detroit, Michigan by reducing the number of resources put into less rampant crimes or at the wrong times in the US. More so, it will be beneficial to all residents of Detroit including the police.

**Data Sources**

1) <https://www.kaggle.com/vinchinzu/dc-metro-crime-data?select=dc_crime_add_vars.csv> To access the dataset, use the link above. Dataset of all the crimes in the DC metro police system ranging from Theft, Arson, Assault, Homicide, Sex Abuse, Robbery, and Burglary. Data can be easily geocoded and mapped, trends can be extracted, and predictions can be made. The original data as well as the data set is in rows and columns. The dataset also contains a csv file. Raw Data can be accessed from: <http://crimemap.dc.gov/CrimeMapSearch.aspx#tabs-GeoOther> The data is most easily accessed by downloading 1 ward at a time for the specific data range.

2) <https://data.detroitmi.gov/datasets/rms-crime-incidents?geometry=-83.581%2C42.264%2C-82.617%2C42.442&selectedAttribute=council_district> To access the dataset, use the link above. This data reflects reported criminal offenses that have occurred in the City of Detroit. Offense data was extracted from the Detroit Police Department's records management system. This data set contains the most recent data available and is updated anytime DPD sends official crime records contributing to the Michigan Incident Crime Reporting (MICR) or the National Incident Based Reporting systems (reflected by the IBR Date field). It should be noted that some incidents involve the commission of multiple offenses, such as a domestic assault where property was also vandalized. Accordingly, the data describe all offenses associated with all reported incidents. The data set contains JSON files and charts, also the data is organized into rows and columns. The original file is a Query URL, [https://gis.detroitmi.gov/arcgis/rest/services/OpenData/RMSCrimeIncidents/FeatureServer/0/query?where=1%3D1&outFields=\*&outSR=4326&f=json](https://gis.detroitmi.gov/arcgis/rest/services/OpenData/RMSCrimeIncidents/FeatureServer/0/query?where=1%3D1&outFields=*&outSR=4326&f=json)

**Acquiring Data**

In my search for datasets to compare, it was a hassle finding data with similar criteria. The dc dataset was the easier of the two. From the onset of this project, I wanted to focus on the timings and characteristics of the crimes in the respective locations. I found the DC dataset on Kaggle with a summary of its contents and examples of its incorporation into various projects. This being my first encounter with Kaggle, I hoped to find data with similar constraints on Kaggle for Detroit. Sadly enough, there was none. Nonetheless, I continued my search on google and came across the Detroit datasets portal. After comparing some of the datasets I found with the DC dataset I had, I found that the RMS incident dataset had the most similar constraints. Thus, my datasets became the Kaggle DC dataset and the RMS Detroit dataset.

**Loading Data**

My code below explains how I loaded the dataset. Refer to google colab link above and PowerPoint for pictures.

**Exploring Data**

My Jupyter notebooks is malfunctioning so I have attached a link to the google colab that shows my code interacting with the data. <https://colab.research.google.com/drive/1nNuCoXDd5u09Fcswbhb_pgs5C0AqE7iB?usp=sharing> Both Datasets are very large, The DC dataset contains 342867 rows and 32 columns of data, likewise the Detroit dataset contains 287371 rows and 24 columns. The Dc dataset shows the date, year, month, district, offense, neighborhood, crime type of the crimes and more in DC from 2008 to 2017 while the Detroit Dataset shows the crime id, report number, address, offense description, offense category and more in Detroit from 1915 to 2020. The Datasets contain a lot of information, however, not all are relevant to my project. Moreover, I explored the timeframe and the categorization of crimes in both datasets. I found that they are grouped differently, but not so differently that it is impossible to regroup.

**Visualizing Data**

My code below shows how I visualized the data and explored it. I plot my graphs in the analysis section after I have cleaned my data.

**Analyzing Data**

For my analysis, I decided to do series bar plots of the crime growth in relation to years, time of day and a bar plot to visualize the crime rates of the different categories of crime. I did this because I wanted to answer my objective questions for my analysis. Starting with the DC data, I found that crime in DC is more rampant in the evening, theft was the most common crime reported in DC from 2008- 2017 and 2017 crime in DC significantly reduced. From my analysis of the Detroit Dataset, I found that crime in DC grew in 2016 but largely spiked in 2017. Larceny is defined as theft of personal property, that is the most committed crime in the Detroit Dataset and Detroit is most dangerous in the evenings.

**Results:**

min 1915.000000

25% 2017.000000

50% 2018.000000

75% 2019.000000

max 2020.000000

This proves my inference that crime in Detroit majorly spiked in 2017 and has been growing ever since 2016.

**The DC Dataset**

* peak crime time in DC from 2008 to 2017 is the Evening, then day, least is midnight.
* Theft/Other is committed the most while Arson is committed the least in DC, 2008- 2017.
* 2017 has the lowest recorded crime rate in the dataset

**The Detroit Dataset**

* Crime in Detroit grew in 2016, but spiked greatly in 2017, before that crime rate was very low.
* Crime in Detroit has been growing at an exponential rate since 2016
* Larceny is the most committed crime in Detroit, 2008 - 2017
* Detroit is most dangerous in the evenings

**Cross referencing both Datasets**

* Theft is the major problem in both places.
* The evenings are dangerous in both places.

WEKA

To choose the best class type to represent my data, I used a 10-fold cross validation of different class types. Using the Crime type as the class yielded the best results.

=== Run information ===

Scheme: weka.classifiers.rules.ZeroR

Relation: dc\_crime\_add\_vars

Instances: 342867

Attributes: 32

X

REPORT\_DAT

SHIFT

OFFENSE

METHOD

BLOCK

DISTRICT

PSA

WARD

ANC

NEIGHBORHOOD\_CLUSTER

BLOCK\_GROUP

CENSUS\_TRACT

VOTING\_PRECINCT

CCN

START\_DATE

END\_DATE

XBLOCK

YBLOCK

optional

date

year

month

day

hour

minute

second

EW

NS

quad

crimetype

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

ZeroR predicts class value: Non-Violent

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 280528 81.8183 %

Incorrectly Classified Instances 62339 18.1817 %

Kappa statistic 0

Mean absolute error 0.2975

Root mean squared error 0.3857

Relative absolute error 100 %

Root relative squared error 100 %

Total Number of Instances 342867

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

1.000 1.000 0.818 1.000 0.900 ? 0.500 0.818 Non-Violent

0.000 0.000 ? 0.000 ? ? 0.500 0.182 Violent

Weighted Avg. 0.818 0.818 ? 0.818 ? ? 0.500 0.702

=== Confusion Matrix ===

a b <-- classified as

280528 0 | a = Non-Violent

62339 0 | b = Violent