

CMP7205 – Applied Statistics

REPORT TITLE

Applied Statistics Report

Zan Zver, 18133498



BIRMINGHAM CITY
University

MSc (Hons) in Big Data Analytics
Faculty of Computing, Engineering and the Built Environment
Birmingham City University

Contents

1	Section title	3
1.1	Subsection	3
2	Executive Summary	4
2.1	Subsection	4
3	Introduction	5
3.1	Subsection	5
4	Methodology	6
4.1	Subsection	6
5	Datasets	7
5.1	Subsection	7
6	Results and Discussion	15
6.1	Subsection	15
7	Conclusions	16
7.1	Subsection	16
8	References	17
8.1	Subsection	17

List of Tables

List of Figures

Abstract

Summary of the entire report

1 Section title

This is the introduction section

1.1 Subsection

2 Executive Summary

This is the Executive Summary section

2.1 Subsection

briefly introduce the problem, major questions and interesting findings of your report. Although present at the start of the report, this must be written after the rest of the report is complete and must be no longer than half a page;

3 Introduction

This is the Introduction section

3.1 Subsection

describe the purpose of the report - introduce the problem domain with suitable citations and list a set of related statistical questions;

4 Methodology

This is the Methodology section

4.1 Subsection

list and justify the number of statistical techniques selected (such as those taught) to answer the previously outlined questions (or intermediate questions within) by showing how they are apt and what information can be extracted from them. This concerns marking criterion 2 listed later.

5 Datasets

This is the Datasets section

Source: <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2>

5.1 Subsection

describe and justify the datasets used to show what information is contained in the datasets and how it aids to answer the relevant questions. If the description is very long, then summarise the key information and detail the columns in the datasets in appropriate appendices. Also, properly cite the dataset in the References section and include their exact hyperlinks;

(Data Pre-processing): discuss any data pre-processing that is required to prepare the data for subsequent analysis.

Column Name	Description	Type
ID	Unique identifier for the record.	Number
Case Number	The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.	Plain Text
Date	Date when the incident occurred. this is sometimes a best estimate.	Date and Time
Block	The partially redacted address where the incident occurred, placing it on the same block as the actual address.	Plain Text
IUCR	The Illinois Unifrom Crime Reporting code. This is directly linked to the Primary Type and Description. See the list of IUCR codes at https://data.cityofchicago.org/d/c7ck-438e .	Plain Text
Primary Type	The primary description of the IUCR code.	Plain Text
Description	The secondary description of the IUCR code, a subcategory of the primary description.	Plain Text
Location Description	Description of the location where the incident occurred.	Plain Text
Arrest	Indicates whether an arrest was made.	Checkbox

Domestic	Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.	Checkbox
Beat	Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts. See the beats at https://data.cityofchicago.org/d/aerh-rz74 .	Plain Text
District	Indicates the police district where the incident occurred. See the districts at https://data.cityofchicago.org/d/fthy-xz3r .	Plain Text
Ward	The ward (City Council district) where the incident occurred. See the wards at https://data.cityofchicago.org/d/sp34-6z76 .	Number
Community Area	Indicates the community area where the incident occurred. Chicago has 77 community areas. See the community areas at https://data.cityofchicago.org/d/cauq-8yn6 .	Plain Text
FBI Code	Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS). See the Chicago Police Department listing of these classifications at https://ucr.fbi.gov/nibrs/2011/resources/nibrs-offense-codes/view	Plain Text
X Coordinate	The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.	Number

Y Coordinate	The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.	Number
Year	Year the incident occurred.	Number
Updated On	Date and time the record was last updated.	Date and Time
Latitude	The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.	Number
Longitude	The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.	Number
Location	The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.	Location

Illinois Uniform Crime Reporting (IUCR)

<https://data.cityofchicago.org/Public-Safety/Chicago-Police-Department-Illinois-Uniform-Crime-R/c7ck-438e>

Column Name	Description	Type
IUCR	Crime ID	Plain Text
Primary description	First (main) description of the crime	Plain Text
Secondary description	Additional description of the crime	Plain Text
Index code	There are 2 codes, "I" (Index) and "N" (Non-Index). "I" code indicates crimes that are collected nation-wide (by FBI) while "N" is for other (usually smaller) crimes.	Plain Text

Active	Whether the code is active. Retired codes (No) are present in this dataset for historical reference. There is a filtered view for this dataset showing only active codes.	Boolean
--------	---	---------

Beats

<https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Beats-current-/aerh-rz74>

Column Name	Description	Type
The_geom	List of locations (as multi polygon) containing LAT and LON.	Multipolygon - location
District	ID of the police district	Number
Sector	Geographically divided area with associated ID	Number
Beat	No description.	Number
Beat_num	Indicates Beat ID from main dataset	Number

Police Districts

<https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Districts-current-/fthy-xz3r>

Column Name	Description	Type
The_geom	List of locations (as multi polygon) containing LAT and LON where districts operate.	Multipolygon - location
Dist_label	Name of the district	Plain text
Dist_num	Number of the district	Plain text

Community Areas

<https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6>

Column Name	Description	Type
The_geom	List of locations (as multi polygon) containing LAT and LON.	Multipolygon - location
Perimeter	No description.	Number

Area	No description.	Number
Comarea_	No description.	Number
Comarea_ID	No description.	Number
Area_numbe	ID of the area. This is the the same as in main dataset. Note, column name does not have r in number.	Number
Community	Name of the community / region.	Plain Text
Area_num_1	ID of the area. This is the the same as in main dataset. Note, column name does not have r in number.	Number
Shape_area	No description.	Number
Shape_len	No description.	Number

FBI codes

<https://ucr.fbi.gov/nibrs/2011/resources/nibrs-offense-codes/view>

Group A Offenses

Offense Code	Offense Description	Crime Against
Arson		
200	Arson	Property
Assault Offenses		
13A	Aggravated Assault	Person
13B	Simple Assault	Person
13C	Intimidation	Person
Bribery		
510	Bribery	Property
Burglary/Breaking & Entering		
220	Burglary/Breaking & Entering	Property
Counterfeiting/Forgery		
250	Counterfeiting/Forgery	Property
Destruction/Damage/Vandalism of Property		
290	Destruction/Damage/Vandalism of Property	Property
Drug/Narcotic Offenses		
35A	Drug/Narcotic Violations	Society
35B	Drug Equipment Violations	Society

Embezzlement		
270	Embezzlement	Property
Extortion/Blackmail		
210	Extortion/Blackmail	Property
Fraud Offenses		
26A	False Pretenses/Swindle/Confidence Game	Property
26B	Credit Card/Automated Teller Machine Fraud	Property
26C	Extortion/Blackmail	Property
26D	Welfare Fraud	Property
26E	Wire Fraud	Property
Gambling Offenses		
39A	Betting/Wagering	Property
39B	Operating/Promoting/Assisting Gambling	Property
39C	Gambling Equipment Violations	Property
39D	Sports Tampering	Property
Homicide Offenses		
09A	Murder & Nonnegligent Manslaughter	Property
09B	Negligent Manslaughter	Property
09C	Justifiable Homicide	Person/Not a Crime
Kidnapping/Abduction		
100	Kidnapping/Abduction	Person
Larceny/Theft Offenses		
23A	Pocket-picking	Person
23B	Purse-snatching	Person
23C	Shoplifting	Person
23D	Theft From Building	Person
23E	Theft From Coin-Operated Machine or Device	Person
23F	Theft From Motor Vehicle	Person
23G	Theft of Motor Vehicle Parts or Accessories	Person
23H	All Other Larceny	Person

Motor Vehicle	Theft	
240	Motor Vehicle Theft	Property
Pornography /	Obscene Material	
370	Pornography/Obscene Material	Society
Prostitution Offenses		
40A	Prostitution	Society
40B	Assisting or Promoting Prostitution	Society
Robbery		
120	Robbery	Property
Sex Offenses, Forcible		
11A	Forcible Rape	Person
11B	Forcible Sodomy	Person
11C	Sexual Assault With An Object	Person
11D	Forcible Fondling	Person
Sex Offenses, Nonforcible		
36A	Incest	Person
36B	Statutory Rape	Person
Stolen Property Offenses		
280	Stolen Property Offenses	Property
Weapon Law Vio- lations		
520	Weapon Law Violations	Society

Group B Offenses

Offense Code	Offense Description	Crime Against
90A	Bad Checks	Property
90B	Curfew/Loitering/Vagrancy Violations	Society
90C	Disorderly Conduct	Society
90D	Driving Under the Influence	Society
90E	Drunkenness	Society
90F	Family Offenses, Nonviolent	Society
90G	Liquor Law Violations	Society
90H	Peeping Tom	Society

90I	Runaway	Not a Crime
90J	Trespass of Real Property	Society
90Z	All Other Offenses	Person, Prop- erty, or Society

TO DO:

1- data import

import data to R, break data to date & time, remove X and y coordinate, remove location

DONE!!!!!!

2- remove empty data and remove outliers

3- see how data is shaped, what correlations are formed, etc...

6 Results and Discussion

This is the Results and Discussion section

6.1 Subsection

Conduct the statistical analysis and report the results. Show how the statistical assumptions required for applying a statistical test are fulfilled. Discuss and evaluate the results to effectively answer the questions outlined earlier;

- o Detail the results in tables and figures as appropriate. Each table and figure must be numbered, and must have appropriate legends (if required) and captions. The captions must be a set of grammatically complete sentences (up to 3 lines long) that concisely describe the information in the tables and the figures. No such table and figure should be included that is not discussed in the text. The numerical comparisons should be supported by appropriate tests of statistical significance.

7 Conclusions

This is the conclusions section.

7.1 Subsection

Summarize the key findings of your report so that if someone does not have the time to go through the whole report, he/she can still understand the important results.

8 References

This is the references section

8.1 Subsection