A picture containing computer

Description automatically generated

AN ANALYSIS OF ACCIDENTAL DRUG-RELATED DEATHS BY DRUG OVERDOSE

IN

THE STATE OF CONNECTICUT, USA

FROM 2012 - 2019

FOLASHIKEMI MARYAM ASANI OLANIYAN – P2586104

# ABSTRACT

An analysis of accidental drug-related deaths by drug overdose derived from an investigation by the Chief Medical Examiner of The State of Connecticut from 2012 to 2018. The analysis shows the demography of the number of deaths, the trend of deaths by each drug, the location of deaths and geographical map showing location of deaths.

The report shows a significant increase in drug-related death cases per year. There are key locations with significantly higher cases and the numbers appear to good yearly. From the demography analysis, there is a huge variation in drug-related death numbers by sex and race but the normal distribution by age. The geographical heat-map further emphasis our analysis giving insights on the distribution of death and residence locations.

Table of Contents

Introduction to The Data **3**

Source and License 3

Data Format 3

Data, Scope, Range and Accuracy3

Methodology4

Data Cleansing4

Preparing the Data for Analysis5

Data Analysis 6

Distribution of Deaths by Year 6

Demography of Deaths by Year 6

Sex3

Age3

Race3

Drug Analysis6

Understanding the Location, Death and Residence6

Death City Analysis3

Death Location3

Geographical Analysis3

Heroin3

Any Opioid3

Fentanyl 3

Race3

Conclusion and Avenues for Future Research3

References3

Appendix I3

Appendix II3

Sex3

Age3

Race3

Appendix III3

Race3

Race

# INTRODUCTION TO THE Data

SOURCE AND LICENSE

The State of Connecticut gave access to the data. It was derived from an investigation by the office of the Chief Medical Examiner. The data is public property as it was obtained from the general public. This report will explore, access and derive insightful knowledge which can be accessed by any citizen with interest in preventing accidental drug-related deaths.

The data set within this report were derived from data.gov (2019) along with the metadata in JSON file format. The data on data.gov are provided freely and at no cost to the public. The United States published this data freely and updates under the US’s Open Government license, permitting the copying, publishing and exploiting 0f the data for commercial and non-commercial purposes.

DATA FORMAT

The complete data from 202 – 20 8 is published as a CSV file. The CSV file contained anonymised information describing each accidental death associated with drug overdose with a “Y” value under the different substance columns that indicate the particular substance was detected. The CSV file contains several fields relating to the location and demography of Deaths; this report will primarily look at the year, age, sex, race, location, latitude, longitude and each drug. All Drug are Categorical data types: for a full explanation of each drug per year see. Appendix I

The geographical coordinates used to plot the Death City Location and Residence City Location boundaries of the Accidental Drug-Related Deaths were similarly derived from the CSV. They were split into latitude and longitude for analysis.

DATA, SCOPE, RANGE AND ACCURACY

The data within this report covers the day, month and years from January 2012 to December 2018; inclusively. However, data.gov does point out that Any Opioid is ticked if the Medical Examiner cannot conclude whether it’s RX Morphine or heroin based morphine in the toxicity results.

Similarly, while this report limits its scope to accidental drug-related deaths with The State Of Connecticut. It is important to note that some residences live outside of Connecticut and their Deaths may be recorded in multiple states.

A large part of the dataset had commas within the rows and this resulted in data inconsistency and accuracy. A find and replace comma with blank was implemented on the raw file uploading to hue. The field unused in this analysis were dropped.

# MethODology

OBTAINING THE DATA

The raw data is a CSV (Comma-Separated Values) file was which was uploaded to the HDFS via hue. The analysis was conducted using Apache Spark in Jupyter Notebook. The first set was to create a spark context (sc) and import the relevant libraries required for analysis.

The spark context was used to create an RDD (Resilient Distributed Datasets) using sc.textFile(“Data Path in Hue”). For the workflow chart: See Appendix I

# DATA CLEANSING

Headers were removed from the data by creating a new RDD called header and calling an action to output the first row. Creating a new RDD filters off the headered from the RDD named header.

A screenshot of a social media post

Description automatically generated

A screenshot of a social media post

Description automatically generated

To ensure there were no rows with length less than or equal to zero:

A screenshot of a cell phone

Description automatically generated

When analysing the column Residence City Geocode an attempt to plot the geolocations should error. To rectify the issue, the null values in the column were removed

A screenshot of a social media post

Description automatically generated

# PREPARING THE DATA FOR ANALYSIS

To make the data set easier to analyse on an ad-hoc basis, it was converted to a dataframe. The main difference between an RDD and a dataframe is the RDD is a distributed collection of data elements spread across many machines in a cluster while a dataFrame is a distributed collection of data organized into named columns(reference). The RDD data was split by a delimiter (‘,’). A schema was defined explicitly by mapping the fields with the objects data type. After creating the dataframe, DeathCityGeo and Residence CityGeo were spilt into Latitude and Longitude and their data type was changed to float.

A screenshot of a cell phone

Description automatically generated

Likewise, the Date column was converted to date type and Year, month and day columns were created.

A screenshot of a cell phone

Description automatically generated

The unrequired for field for this analysis were removed by using drop() action.



The dataframe was registered as a Table using SQL queries

A screenshot of a cell phone

Description automatically generated

At any point, the results returned by the SQL query could easily be converted to Pandas to create visuals and illustrations that help summarise our findings. We could connect to libraries such as Matplotlib, pixie dust and gmplot. Thus, the data Is left unchanged within the dataframe and the SQL Context is used to query the data for desired results.

A screenshot of a cell phone

Description automatically generated

# DATA ANALYSIS

A picture containing drawing

Description automatically generatedDISTRIBUTION OF DEATHS BY YEAR

A simple count of the total accidental drug-related deaths in The State of Connecticut reveals that there has been a continuous increase in the number deaths since 2012 and the highest being 2017.

# DEMOGRAPHY OF DEATHS BY YEAR

SEX

It was observed that there is a significant ratio of Male to Female, this is rather amazing because the percentage of women to men in Connecticut is 51.2% to 48.8%. (reference).

A screenshot of a cell phone

Description automatically generated

AGE

The Age range of Deaths suggests that the highest death case by age were of persons aged 33 year in 2017 with nearly 40 individuals involvA screenshot of a cell phone

Description automatically generateded.

RACE

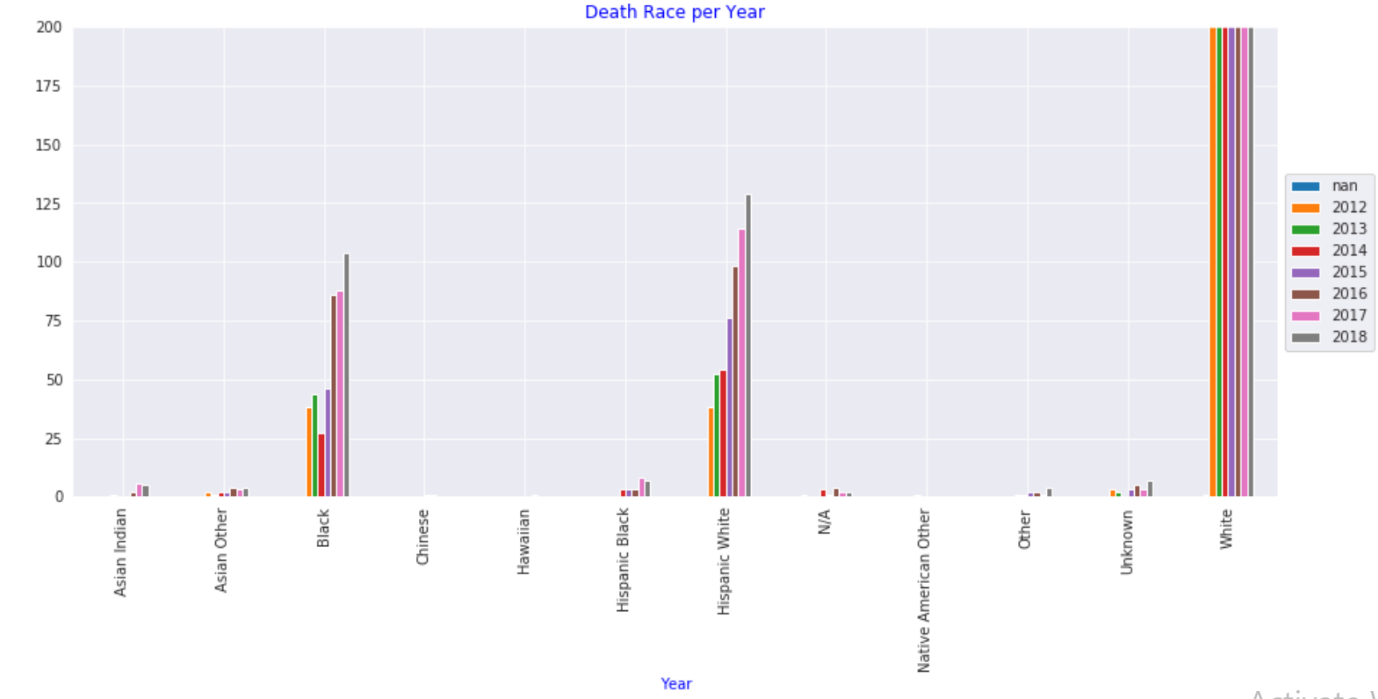
The population of Connecticut is White: 76.36%, Black or African American: 10.56%, Other race: 5.17%, Asian: 4.43%, Two or more races: 3.17%, Native American: 0.27%, Native Hawaiian or Pacific Islander: 0.03%(Reference).

The trend of the death by race shows that whites had the largest death cases with 2017 begin the highest number of cases.

A screenshot of a cell phone

Description automatically generated

A closer look at the data shows that other races also have highly significant cases when compare to their percentage in the population. There’s been a drastic increase in the number of cases for Hispanic whites and blacks.



# DRUG ANALYSIS

The data set contains 15 fields of the different Drugs taken that lead to overdose. Drugs: Any Opioid had led to the highest number of deaths. A good recommendation to the DEA (Drug Enforcement Administration) in Connecticut will be to monitor the prescription of drugs containing Opioids. A study conducted showed that Opioids are often prescribed as pained killers but have very high addiction tendencies (Reference).

The second-largest drugs cause of death is Heroin then followed by Cocaine and Tramad.

A geographical comparing the location of death and residence by these drugs will show the exact location of occurrence. This may suggest the presence of a drug-dealer in the area.

A picture containing table, sitting, colorful

Description automatically generated

# UNDERSTANDING THE LOCATION OF DEATH AND RESIDENCE

DEATH CITY ANALYSIS

Most of Connecticut’s population live in its capital Hartford, a pie chart plot shows that most of the drug-related deaths happened there. The availability of rehabilitation centres in the locations with the highest cases will help in preventing further deaths as more individuals addicted to the drug will be able to receive adequate medical intervention.

A screenshot of a cell phone

Description automatically generated

A picture containing table, different, group

Description automatically generated

DEATH LOCATION

Most of the death cases happened in the residences while others happened at the hospital.A screenshot of a cell phone

Description automatically generated

GEOGRAPHICAL ANALYSIS

A close up of a map

Description automatically generatedTo get a better view of the Death City Location, we created a plot on the map.

From the heatmap, the concentrated points are Hartford, Bridgeport, New Haven, Waterbury and New Britain. This is unsurprising as many news reports mentioned arrests of drug dealers in these areas.

# Conclusion and Avenues for Future Research

While analysing this data, we have discovered that males have larger drug-related deaths, the highest number of cases have been amongst the Whites. We also found that most of the deaths took place in residences in Hartford. We realised that the growing number of cases of ‘Any Opioid’ deaths suggest an opioid crisis that need to be curbed via medical intervention and the law enforcement agencies.

This analysis looked at 2012 -2018, updating the data in future years will help in future analysis to see if trends drastically reduce with improved drug-related death prevention methods.

# REFERENCES

# aPPENDIX