

INTRODUCTION TO DATA SCIENCE

EXPLORATORY ANALYSIS ON DRUG RELATED DEATHS

C.Diya (PES1201700246)

Namrata R (PES1201700921)

Semester-3, Section-B

Reason for choosing dataset:

Deaths related to drugs is one of the most common issues being addressed all over due to which we decided to analyse a data set based on this topic.

The data needed to be reformatted as it appeared to have missing values and did not have a standard format of entries. After restructuring this data, we wanted to provide an appropriate visual representation of the various factors responsible.

Aim and Question to be asked:

- **Aim:** To predict the different factors that are most likely to cause drug induced deaths and hence use this to reduce the latter in the future.
- **Question to be asked:** Can we isolate a target group from test samples by analysing the chosen data frame based on various parameters? And if so, what are the different factors affecting it?

Data Cleaning

Formatting the 'Date' column.

Ensuring no drug is repeated in the 'Other' column.

BEFORE

	CaseNumber	Date
0	13-16336	2013-11-09
1	12-18447	2012-12-29

AFTER

	CaseNumber	Date
0	13-16336	11-09-2013
1	12-18447	12/29/2012

```
df3['Other'].value_counts()
Out[104]: - 3726
MORPHINE 54
PCP 41
HYDROMORPH 28
BUPRENORPHONE 24
OPIATE 15
BUPREN 11
MORPH 11
10
MORPHINE RX 10
BUPRENOR 9
OPIATES 8
BUPRENO 7
U-47700 6
DUSTER 6
CODEINE 5
OTHERS 5
OPIATE SCREEN 5
HYDROMORPHONE 5
MDMA 5
KETAMINE 5
TAPENTADOL 4
HYDR-MOR 4
COD 3
CARFENTANIL 3
HYDROMORP 3
BUPRE 3
DIFLURO 2
H-MORPH 2
```

Filling in missing values in 'Age' column.

BEFORE

```
path[path.Age.isnull()]
```

	CaseNumber	Date	Sex	Race	Age	Residence City	Residence State	Residence County	Death City	Death State	...	Benzodiazepine	Methadone	Amphet	Tramad	Morphine (not heroin)	O
779	14-9876	2014-06-28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	YES	NO	NO	NO	NO	
1891	15-16348	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NO	NO	NO	NO	NO	

2 rows × 32 columns

AFTER

```
path[path.Age.isnull()]
```

	CaseNumber	Date	Sex	Race	Age	Residence City	Residence State	Residence County	Death City	Death State	...	Benzodiazepine	Methadone	Amphet	Tramad	Morphine (not heroin)	Other
--	------------	------	-----	------	-----	-------------------	--------------------	---------------------	---------------	----------------	-----	----------------	-----------	--------	--------	-----------------------------	-------

0 rows × 32 columns

Dealing with missing values of the qualitative data.

BEFORE

Residence City	Residence State	Residence County	Death City	Death State
GROTON	NaN	NEW LONDON	GROTON	NaN
WOLCOTT	NaN	NEW HAVEN	WATERBURY	NaN
ENFIELD	NaN	NaN	ENFIELD	NaN
WALLINGFORD	NaN	NaN	WALLINGFORD	NaN

AFTER

Residence City	Residence State	Residence County	Death City	Death State	...
GROTON	Data Unavailable	NEW LONDON	GROTON	Data Unavailable	...
WOLCOTT	Data Unavailable	NEW HAVEN	WATERBURY	Data Unavailable	...
ENFIELD	Data Unavailable	Data Unavailable	ENFIELD	Data Unavailable	...
WALLINGFORD	Data Unavailable	Data Unavailable	WALLINGFORD	Data Unavailable	...

Standardizing the format and cleaning rows for each drug.

BEFORE

```
In [238]: df['Heroin'].value_counts()
Out[238]: Y      2122
          y       22
          3
          Name: Heroin, dtype: int64
```

```
In [88]: df['Fentanyl'].value_counts()
Out[88]: Y      1451
          y        9
          Y-A       2
          Y POPS     1
          Y (PTCH)   1
          Name: Fentanyl, dtype: int64
```

	Benzodiazepine	Methadone	Amphet	Tramad	Morphine (not heroin)	Other	Any Opioid
	Y	NaN	NaN	NaN	NaN	NaN	NaN
	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Y	NaN	NaN	NaN	NaN	NaN	NaN
	NaN	NaN	NaN	NaN	NaN	NaN	NaN

AFTER

	Benzodiazepine	Methadone	Amphet	Tramad	Morphine (not heroin)	Other	Any Opioid
	YES	NO	NO	NO	NO	-	NO
	NO	NO	NO	NO	NO	-	NO
	YES	NO	NO	NO	NO	-	NO
	NO	NO	NO	NO	NO	-	NO
	NO	NO	NO	NO	NO	-	NO

```
path["Oxycodone"].value_counts()
NO      3551
YES      530
Name: Oxycodone, dtype: int64
```

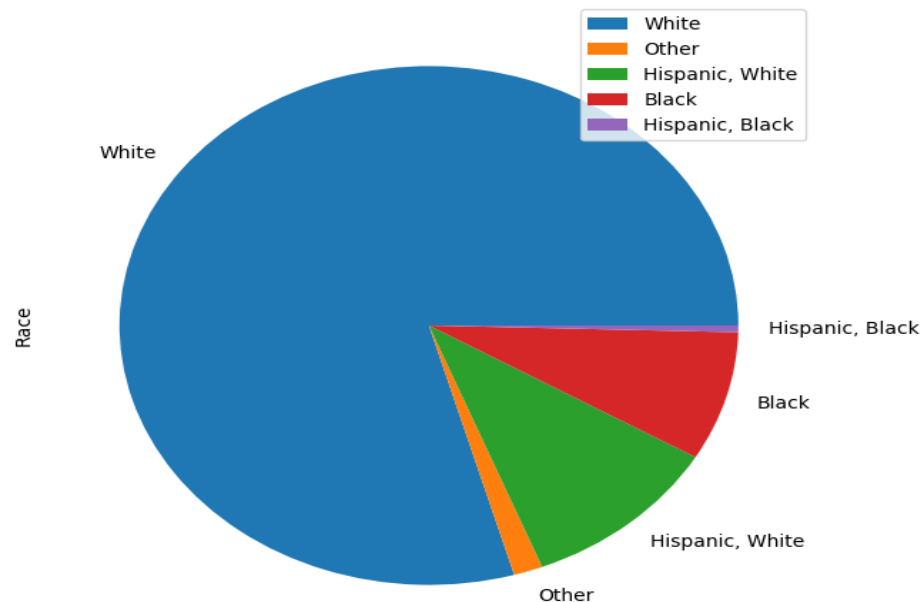
```
path["Oxymorphone"].value_counts()
NO      3985
YES       96
Name: Oxymorphone, dtype: int64
```

```
path["Hydrocodone"].value_counts()
NO      3977
YES      104
Name: Hydrocodone, dtype: int64
```

Visual Representation

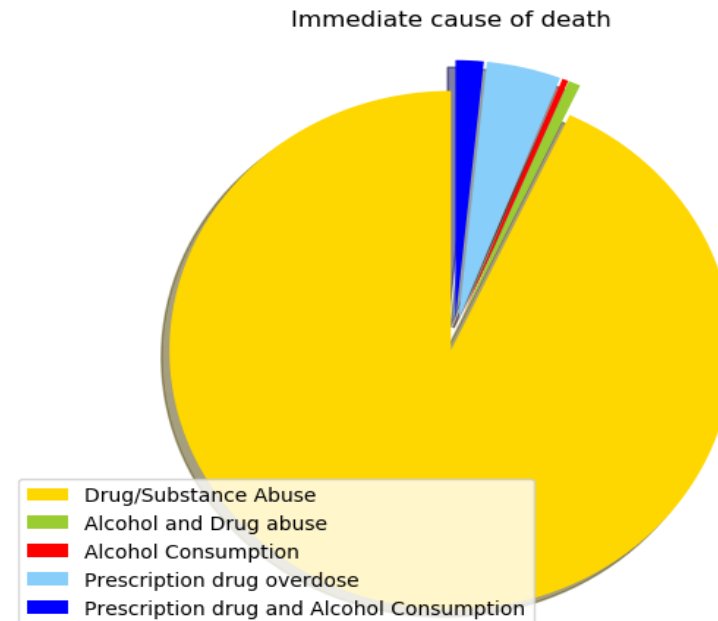
Distribution of number of deaths based on 'Race'.

- Percentage distribution of each race in that region is clearly observed.
- Most affected race in that region is the 'White' race.



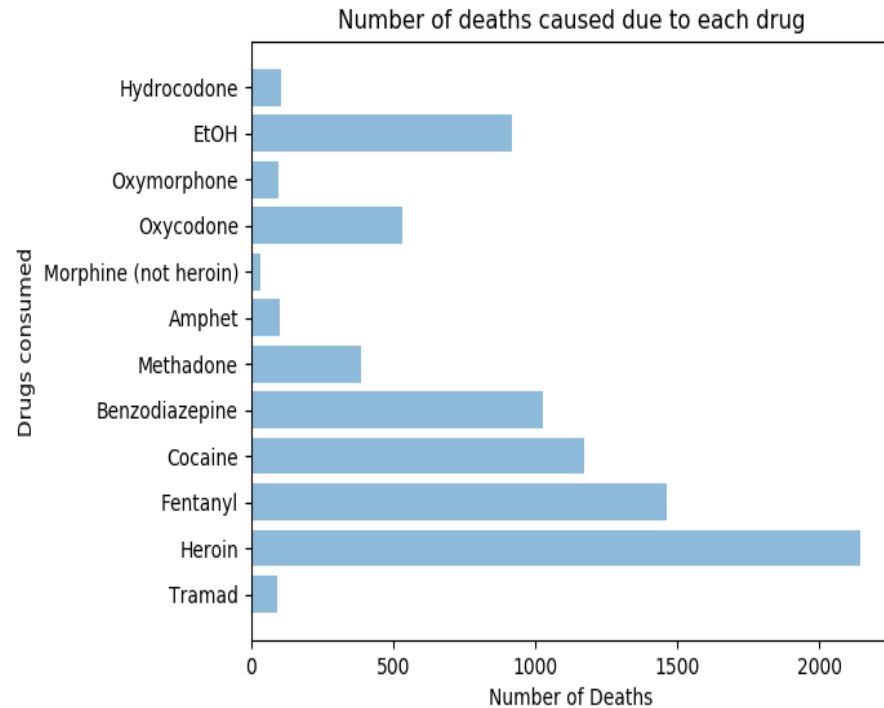
Distribution of the "Immediate cause of death".

- Drug/Substance abuse is responsible for the greatest number of deaths
- However, a notable number of deaths are caused due to prescribed medications.



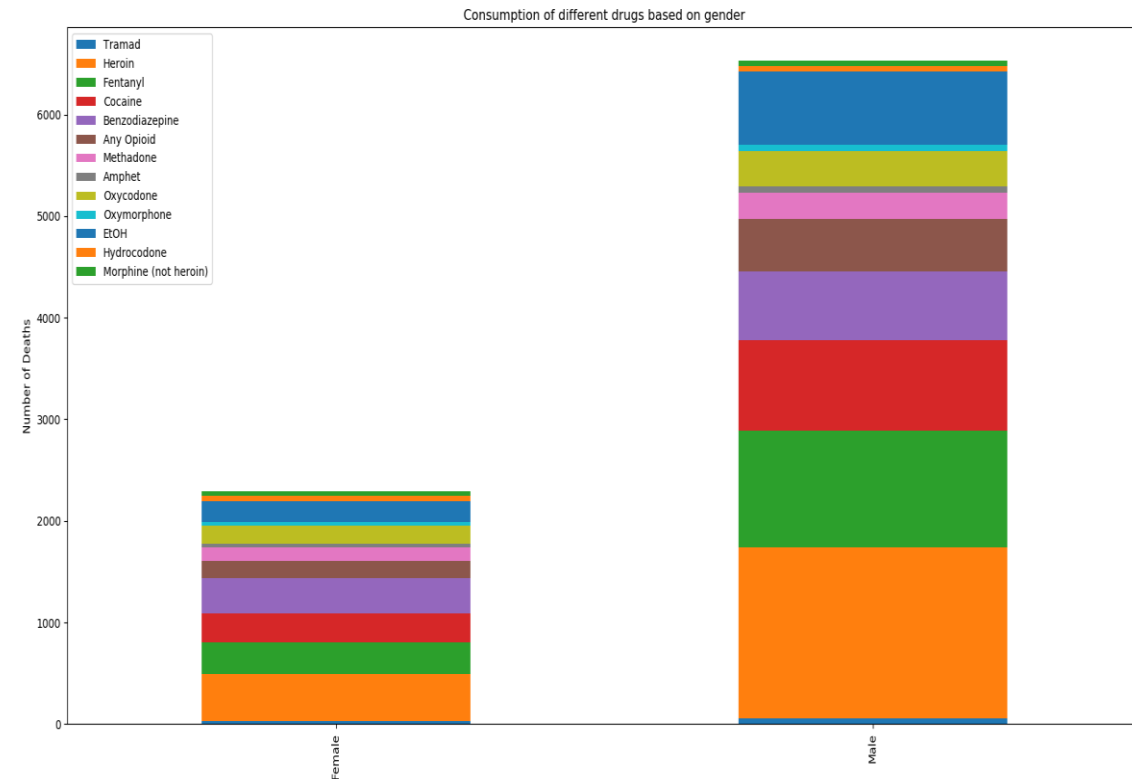
Number of Deaths vs Drugs Consumed.

- Comparative proportion of deaths caused by each drug can be observed.
- Heroin caused the greatest number of deaths in this region.



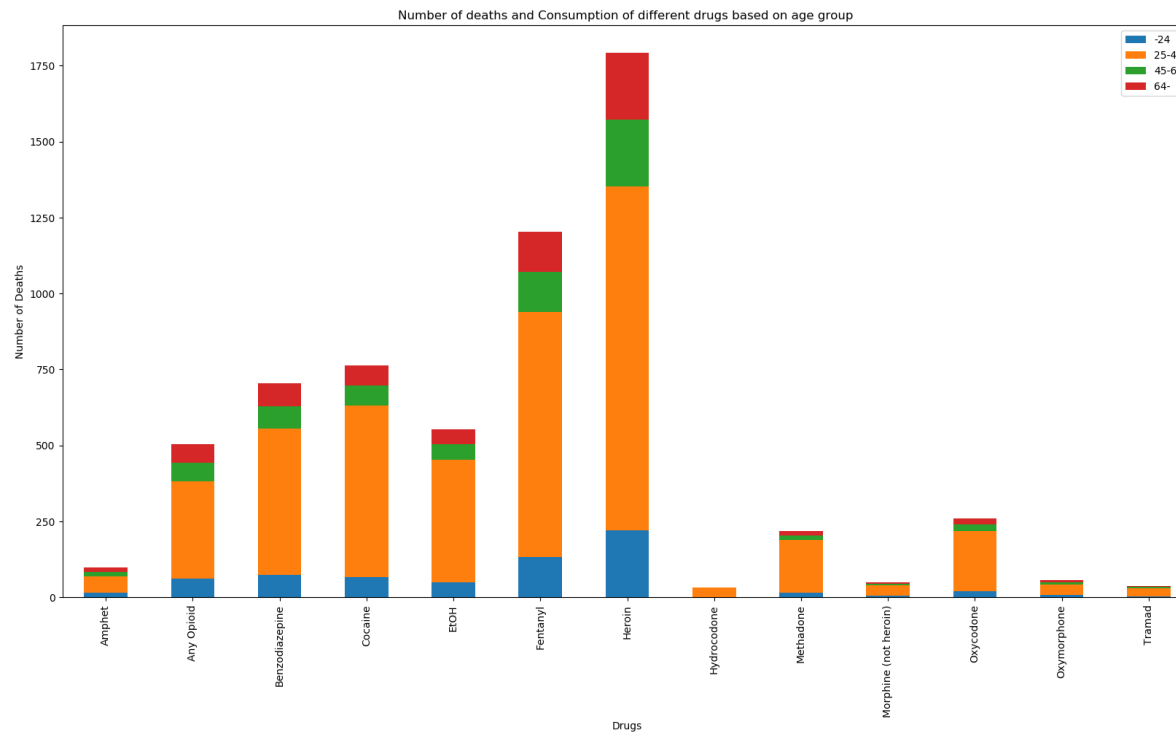
Number of deaths due to each drug for Male and Female.

- Male has higher number of deaths than female.
- Heroin caused the greatest number of deaths for both male and female.



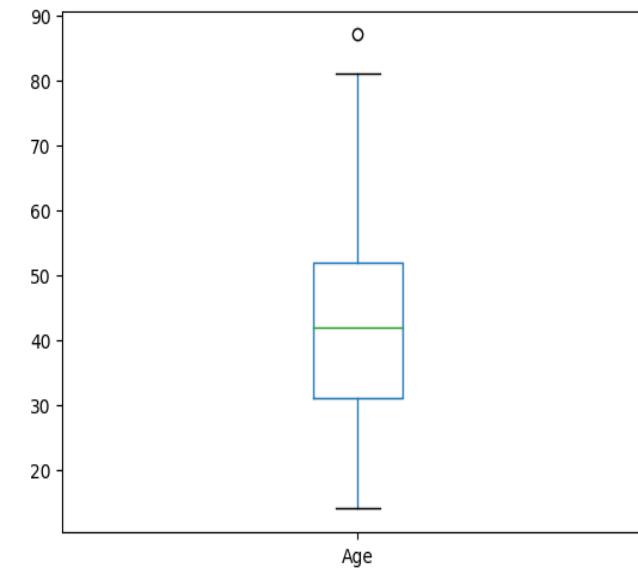
Comparative Distribution of the age groups vs Number of deaths based on the different drugs

- The most number of deaths(most affected) are in the age group is from 25-44 years.
- Comparison between the age group, type of drug and number of deaths can be clearly observed.



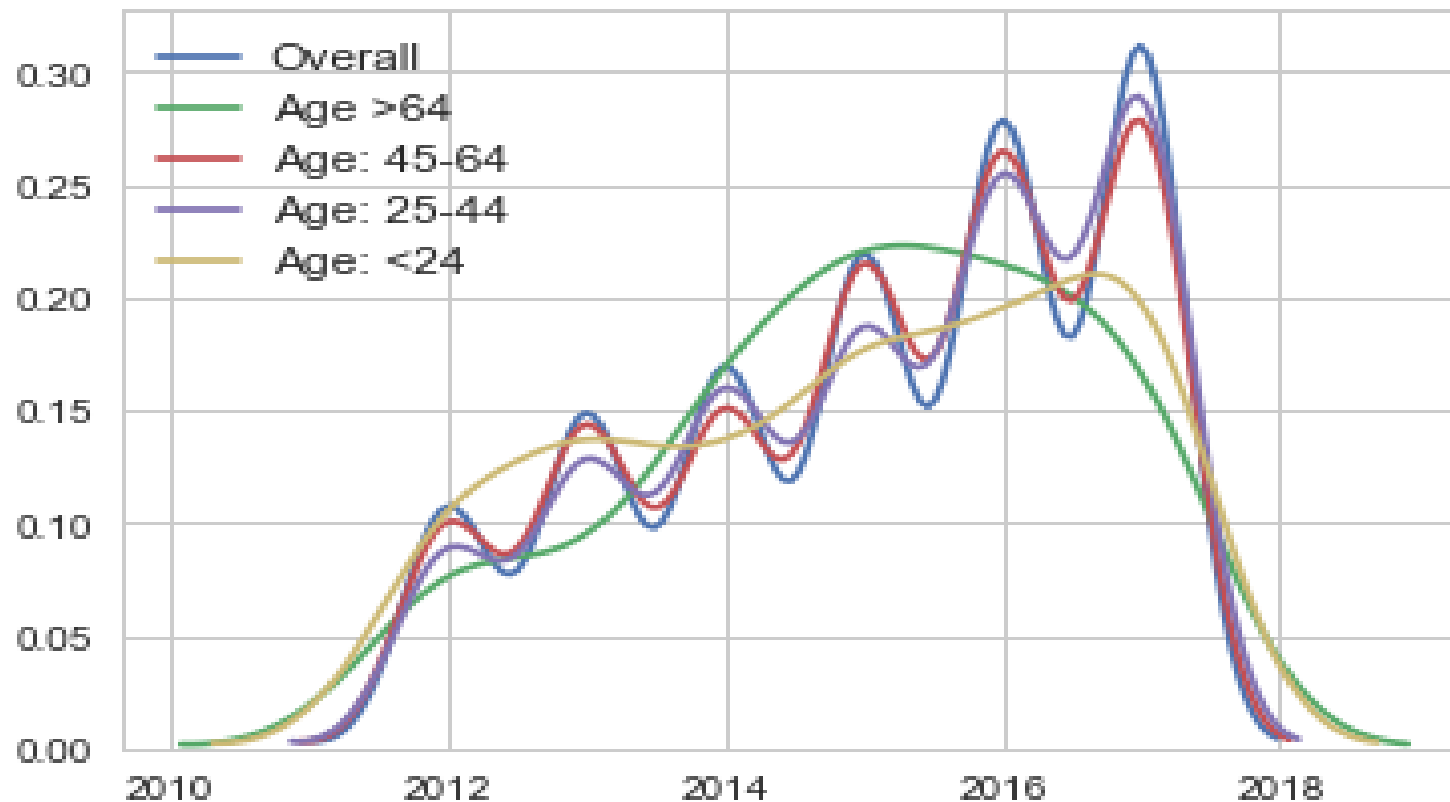
Detecting outliers in 'Age' entries.

- There is one evident outlier(anomaly) i.e., at the age 87.
- The median age is 42.0 years.



Variation of number of deaths caused in each age group over a span of five years.

- 25-44 and 45-64 years is the most affected age group between the years 2016 to 2018.
- For >64 years, the number of deaths peaks in 2015 but has decreased in the recent years.
- For <24 years, rise in number of deaths is not very notable.



Conclusion:

- For the dataset chosen by us, we have observed that the drug that is responsible for the most number of deaths is Heroin.
- Compared to other age groups, the ratio of then number of deaths caused by drugs is significantly higher for age group 25-44 years.
- Although, the greatest number of deaths were caused due to recreational drugs, we observed a notable number of deaths caused due to prescribed medications. From this, we can deduce that a lot of people misuse prescription drugs.
- Unfortunately, over the years the number of drug induced deaths have increased considerably.
- Thus, it is important to work with the target group in order to reduce the impact of drug usage .