

Opioid Overdose Death Prediction using Machine Learning

Import Libraries

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
```

```
In [2]: df = pd.read_csv("Accidental_Drug_Related_Deaths_2012-2021.csv")
df.head()
```

Date	Date Type	Age	Sex	Race	Ethnicity	Residence City	Residence County	Residence State	Injury City	Injury County	Injury State	Injury Place	Description of Injury	Death City	Death County	Death State	Location	Location if Other	Cause of Death	Manner of Death	Other Significant Conditions	Heroin	Heroin death certificate (DC)	Control
0	06/29/2012	37.0	Male	Black	NAN	STAMFORD	FARFIELD	NAN	STAMFORD	NAN	CT	Residence	Used Cocaine	NAN	NAN	NAN	Residence	NAN	Cocaine Toxicity	Accident	NAN	NAN	NAN	
1	06/27/2012	37.0	Male	White	NAN	NORWICH	NEW LONDON	NAN	NORWICH	NAN	CT	Residence	Drug Use	NORWICH	NEW LONDON	NAN	Hospital	NAN	Heroin Toxicity	Accident	NAN	Y	NAN	NA
2	03/24/2014	28.0	Male	White	NAN	HEBRON	NAN	NAN	HEBRON	NAN	CT	Residence	Drug Use	MARLBOROUGH	NAN	NAN	Hospital	NAN	Heroin Toxicity	Accident	NAN	Y	NAN	NA
3	12/31/2014	26.0	Female	White	NAN	BALTIM	NAN	NAN	NAN	NAN	NAN	NAN	NAN	BALTIM	NEW LONDON	NAN	Residence	NAN	Acute Heroin Intoxication	Accident	NAN	Y	NAN	NA
4	01/30/2016	41.0	Male	White	NAN	SHELTON	FARFIELD	CT	SHELTON	NAN	NAN	Residence	Drug Use	BRIDGEPORT	NAN	NAN	Hospital	NAN	Acute Fentanyl Intoxication	Accident	NAN	NAN	NAN	NA

```
In [3]: df.describe().T
```

```
Out[3]:
```

```
Age    556.0    43.02137    12.92594    14.0    33.0    50.0    79.0    87.0
```

```
In [4]: df.describe(include=object).T
```

```
Out[4]:
```

	count	unique	top	freq
Date	8202	3163	05/27/2021	12
Date Type	3002	2	Date of death	7435
Sex	1334	2	Male	1005
Race	9175	13	White	7951
Ethnicity	1287	5	Hispanic	972
Residence City	8813	485	HARTFORD	596
Residence County	8160	135	NEW HAVEN	2289
Residence State	7435	32	CT	7171
Injury City	8024	289	HARTFORD	775
Injury County	8014	16	NEW HAVEN	1763
Injury State	6290	6	CT	6259
Injury Place	9100	91	Residence	6188
Description of Injury	8411	426	Substance Abuse	2640
Death City	9197	238	HARTFORD	1083
Death County	8990	10	HARTFORD	2441
Death State	6972	2	CT	6872
Location	9180	13	Residence	4759
Location if Other	839	441	Friend's Residence	52
Cause of Death	8022	6017	Acute Fentanyl Intoxication	416
Manner of Death	9183	5	Accident	9364
Other Significant Conditions	815	279	Recent Cocaine Use	146
Heroin death certificate (DC)	740	1	Y	740
Cocaine	3171	1	Y	3171
Fentanyl	5672	3	Y	5670
Fentanyl Analogue	728	1	Y	728
Oxycodone	877	1	Y	877
Oxycodone	148	1	Y	148
Ethanol	2470	2	Y	2469
Hydrocodone	155	1	Y	155
Benzodiazepine	2238	1	Y	2238
Methadone	814	1	Y	814
Methamphetamine	127	1	Y	127
Amphetamine	255	1	Y	255
Tramadol	258	1	Y	258
Hydromorphone	56	1	Y	56
Morphine (Not Heroin)	49	4	Y	45
Xylazine	441	1	Y	441
Gabapentin	182	1	Y	182
Opiate NOS	119	1	Y	119
Heroin/Morph/Codeine	1952	1	Y	1952
Other Opioid	90	1	Y	90
Any Opioid	6281	2	Y	6273
Other	654	145	PCP	118
ResidenceCityGeo	9037	348	HARTFORD,CTV41.765775,-72.673556	596
InjuryCityGeo	8991	275	HARTFORD,CTV41.765775,-72.673556	775
DeathCityGeo	9021	238	Hartford,CTV41.765775,-72.673556	1083

```
In [5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8202 entries, 0 to 8201
Data columns (total 48 columns):
 0      Column              non-Null Count  Dtype  
---  --
 1      Date                8202 non-null object
 2      Date Type            3002 non-null object
 3      Sex                  1334 non-null object
 4      Age                  556 non-null float64
 5      Race                 9175 non-null object
 6      Ethnicity            1287 non-null object
 7      Residence City       8813 non-null object
 8      Residence County     8160 non-null object
 9      Residence State      7435 non-null object
10     Injury City          8024 non-null object
11     Injury County        8014 non-null object
12     Injury State         6290 non-null object
13     Injury Place         9100 non-null object
14     Description of Injury 8411 non-null object
15     Death City           9197 non-null object
16     Death County         8990 non-null object
17     Death State          6972 non-null object
18     Location             9180 non-null object
19     Location if Other     839 non-null object
20     Cause of Death       8022 non-null object
21     Manner of Death      9183 non-null object
22     Other Significant Conditions 815 non-null object
23     Heroin death certificate (DC) 740 non-null object
24     Cocaine              3171 non-null object
25     Fentanyl             5672 non-null object
26     Fentanyl Analogue    728 non-null object
27     Oxycodone            877 non-null object
28     Oxycodone            148 non-null object
29     Ethanol              2470 non-null object
30     Hydrocodone          155 non-null object
31     Benzodiazepine       2238 non-null object
32     Methadone            814 non-null object
33     Meth/Amphetamine     127 non-null object
34     Amphetamine          255 non-null object
35     Tramadol             258 non-null object
36     Hydromorphone        56 non-null object
37     Morphine (Not Heroin) 49 non-null object
38     Xylazine              441 non-null object
39     Gabapentin           182 non-null object
40     Opiate NOS            119 non-null object
41     Heroin/Morph/Codeine 1952 non-null object
42     Other Opioid          90 non-null object
43     Any Opioid           6281 non-null object
44     Other                 654 non-null object
45     ResidenceCityGeo     9037 non-null object
46     InjuryCityGeo        8991 non-null object
47     DeathCityGeo         9021 non-null object
memory usage: 3.4 MB
```

```
In [6]: df['Date Type'].value_counts()
```

```
Out[6]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [7]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[7]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [8]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[8]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [9]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[9]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [10]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[10]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [11]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[11]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [12]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[12]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [13]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[13]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [14]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[14]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [15]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[15]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [16]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[16]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [17]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[17]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [18]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[18]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [19]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[19]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [20]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[20]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [21]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[21]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [22]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[22]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [23]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[23]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [24]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[24]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [25]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[25]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [26]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[26]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [27]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[27]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [28]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[28]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [29]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[29]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [30]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[30]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [31]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[31]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [32]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[32]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [33]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[33]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [34]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y':
            return col
    return None
df['Drug Used'] = df.apply(get_drug_used, axis=1)
df['Drug Used'].value_counts().plot(kind='bar')
```

```
Out[34]:
```

```
Date of death      7435
Date reported      1767
Name: Date Type, dtype: int64
```

```
In [35]: df['Heroin death certificate (DC)'].value_counts()
```

```
Out[35]:
```

```
Y      748
Name: Heroin death certificate (DC), dtype: int64
```

```
In [36]: drug_columns = [
    "Heroin",
    "Heroin death certificate (DC)",
    "Cocaine",
    "Fentanyl",
    "Fentanyl Analogue",
    "Oxycodone",
    "Hydrocodone",
    "Ethanol",
    "Benzodiazepine",
    "Methadone",
    "Meth/Amphetamine",
    "Tramadol",
    "Hydromorphone",
    "Morphine (Not Heroin)",
    "Xylazine",
    "Gabapentin",
    "Opiate NOS",
    "Heroin/Morph/Codeine",
    "Other Opioid"
]
```

```
def get_drug_used(row):
    for col in drug_columns:
        if row[col] == 'Y
```