Drug Overdoses in Connecticut 2012-2018

Joosep Albre, Allar Kalina, Mihkel Järviste

Goals

- Find out area based drug problems
- Try to predict the type of drug the person used to overdose
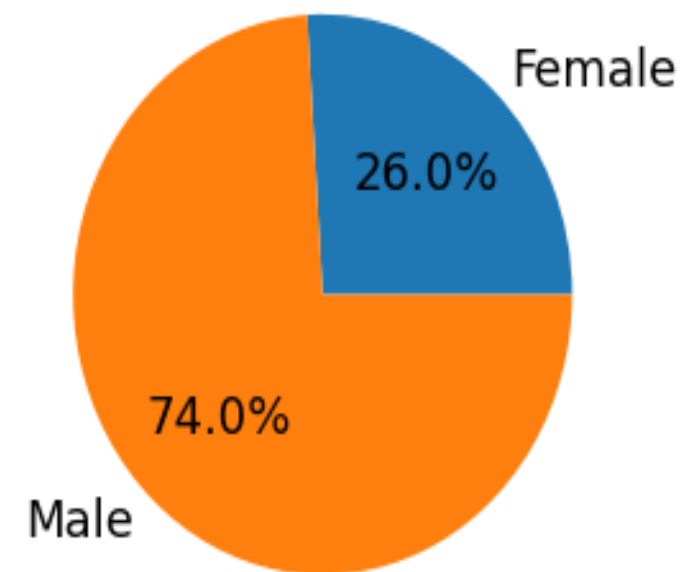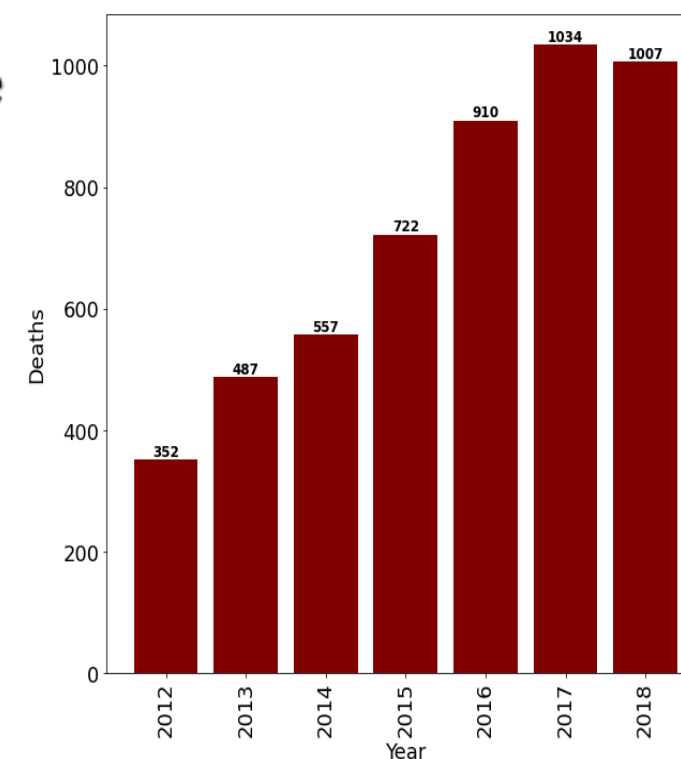- Analyze and visualize the data

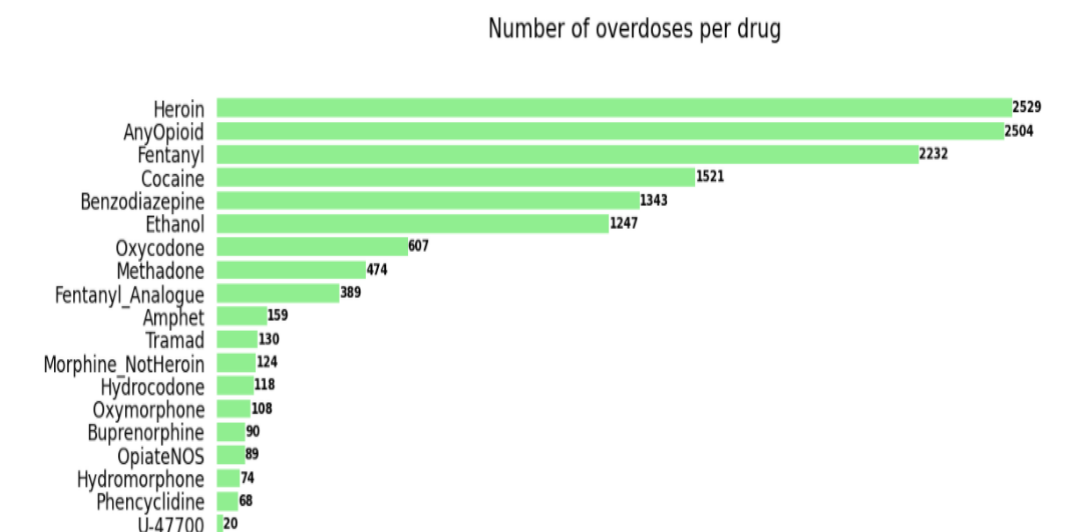Figure 1. Gender ratio

Figure 2. Yearly deaths

Figure 3. Overdose deaths per drug

Data

- DODC(Drug Overdose Database of Connecticut)
- Publicly available on Kaggle
- Collected between 2012-2018
- 5105 reported cases
- 42 initial features

Data Analysis

- Analyzing the data brought up some clear trends linked to drug overdoses
- Over two thirds of drug overdose related death victims were men.
- Recent trends show drug usage and drug fatalities rising
- Majority of deaths were caused by mixing different substances

Results

- We greatly misjudged the time it took to clean the database
- Best model was Random forest, which consistently predicted with the highest accuracy
- Creating the model became almost impossible the more we got familiar with the database, because of the amount of drugs needed to predict

Prediction Models

Even though the dataset was unbalanced, with some drugs killing 1 and others killing more than 2000 people, balancing the data made no accuracy improvements. We trained 3 different models, which all had similar accuracy of 44-45%.

Prediction Model Accuracies

	Name	Accuracy
0	Random Forest Classifier	0.453821
1	Random Forest on balanced data	0.313621
2	K-Nearest Neighbours	0.448505
3	C-Support Vector Classification	0.453821