

CSE572 Assignment - 1

Team members:

Manoj Kumar Yellojirao: 1215201117

Vishal Kumar: 1215200480

Ronak Maheshwari: 1215180408

Saswat Bastia: 1215053814

Abstract:

Analyze and experiment with opioid-related drug overdose fatalities of different states of the United States of America. Given population, death, state code of different states, we need to find the Pearson correlation between population, death. Analyze the opioid death density of all the states and compare how it is related to other state using similarity matrix and graphical analysis of death density of all states using a bar chart.

Keywords: Pearson correlation, Opioid, Death density, Cosine, Euclidean, Similarity matrix.

Introduction: Accidental deaths by fatal drug overdose is a rising trend in the United States The **overdoses.csv** dataset contains information on such opioid-related drug overdose fatalities. It has the 50 rows (one for each state) and the following four columns :

State: Names of states

Population: Population in a particular state

Deaths: Number of opioid casualties in that state

Abbrev: State abbreviation

Data Analysis: we can consider each row data as entities and the columns as different dimensionalities, in the given data we have 50 different entities and 3 dimensions. We have derived the Opioid death density attribute from the existing attributes population and death count.

Task1:

Problem statement: calculate the Pearson correlation coefficient between the **Population** and **Deaths** columns (you may use python/ Matlab libraries).

Implementation: Correlation is any statistical association that refers to how close two variables are to having a with each other. Pearson correlation coefficient is a measure of the linear between two variables X and Y. We used scipy.stats to calculate the Pearson correlation coefficient.

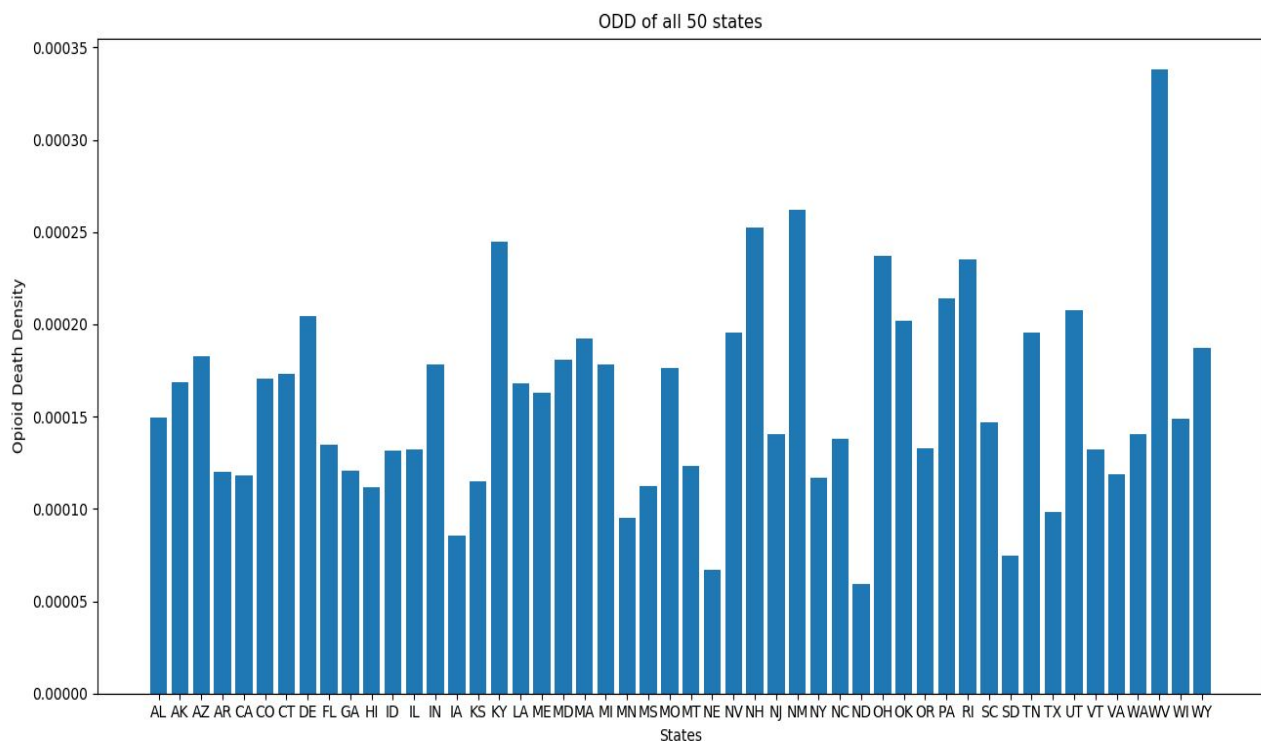
Output: Pearson correlation coefficient is - 0.9399210630849444

Task 2:

Problem statement: Construct a bar-graph representing the Opioid Death Density (ODD), Opioid Death Density = Number of deaths in the state/Population for that state, for each state. There will be 50 bars (one for each state) with the height of each bar representing death density for that particular state. Give proper labels to the x and y-axes.

Implementation: To construct the bar chart of states versus the Opioid death density we need to calculate the ODD based on the given data population and death. Opioid Death Density = Number of deaths in the state/Population for that state. Once we calculate this data. We have required data for the x-y axis. We have used the matplotlib.pyplot library in python to plot the bar chart.

Output:



Task 3:

Problem statement: Construct a similarity matrix representing the closeness of state pairs with respect to their ODD- a state pair will have a similarity value of 1 if the difference in their ODD values is 0, and will have a value of 0 if difference in their ODD values is maximum among the ODD values of all the given pairs.

Implementation: to construct the state-state similarity matrix based on ODD, we need to calculate the distance between each state's ODD. Criteria for calculating the distance between the two states was provided in the problem description. We made our calculations based on that criteria only.

Output:

(Non-truncated matrix is in the file named task3.csv)

Conclusion:

We successfully completed all given tasks and understood the basic concepts and methodologies required to complete them.