Assignment 1: Visualizations

7PAM2000 Applied Data Science 1

ASHIKA MOHAN MUNGATH

MSc Data Science

Student ID: 22063061

Repository Link: ASHIKAMOHAN/ADS--Assignment1 at master (github.com)

INTRODUCTION:

This Visualization assignment consists of datasets ‘Drug use by Age’ and ‘Drug Deaths’ obtained from publicly available, open-source, free database “Kaggle” which are analyzed and visualized using three different visualization techniques . Pandas tools have been used for data processing, and the `pyplot` function has been used to create visualizations.

The visualization techniques used here are line plot,pie chart and bar chart.The pyplot library in matplotlib is used to draw plots.

DATASET

Datasets ‘Drug use by Age’ and ‘Drug Deaths’ are downloaded from Kaggle in which the first dataset consist of age, frequency and usage of different drugs and the second dataset consist of Age and drug deaths.The first dataset covers 13 drugs across 17 age groups.

The source of the dataset is:

https://www.kaggle.com/datasets/tunguz/drug-use-by-age

https://www.kaggle.com/datasets/joebeachcapital/drug-overdose-deaths

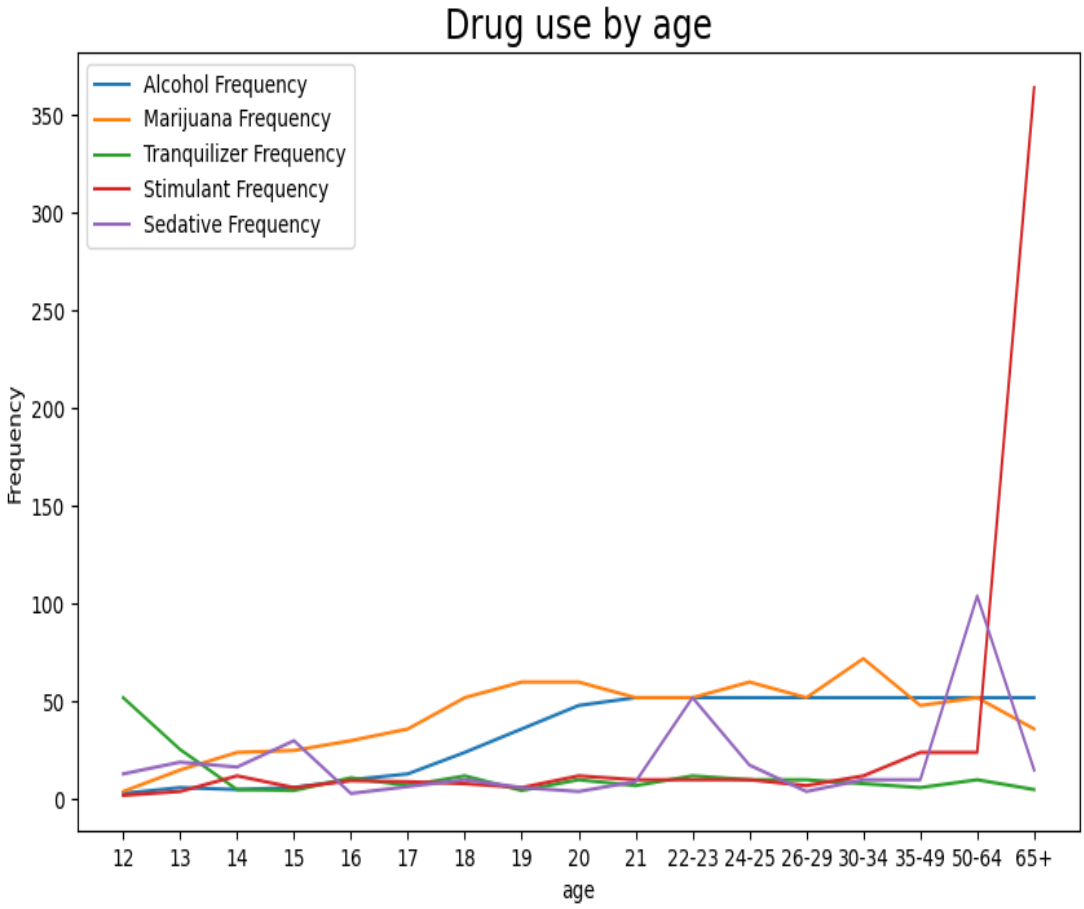
Visualization 1: LINE PLOT (DRUG FREQUENZY BY AGE GROUP)

The first visualization method used here is Line plot. Line plot is one of the simplest and most fundamental graphical analysis approach.

It is a good choice to show the drug frequency over different age groups.

This line graph shows the Frequency of different drugs('alcohol-frequency', 'marijuana-frequency', 'cocaine-frequency’, ‘crack-frequency', 'oxycontin-frequency') over different age groups. X-axis is Age and in the Y-axis, the Frequency of different drugs are plotted.

The plotted line graph shows the death frequency from the age group 12 to 65 above.



From this plot we can conclude that,

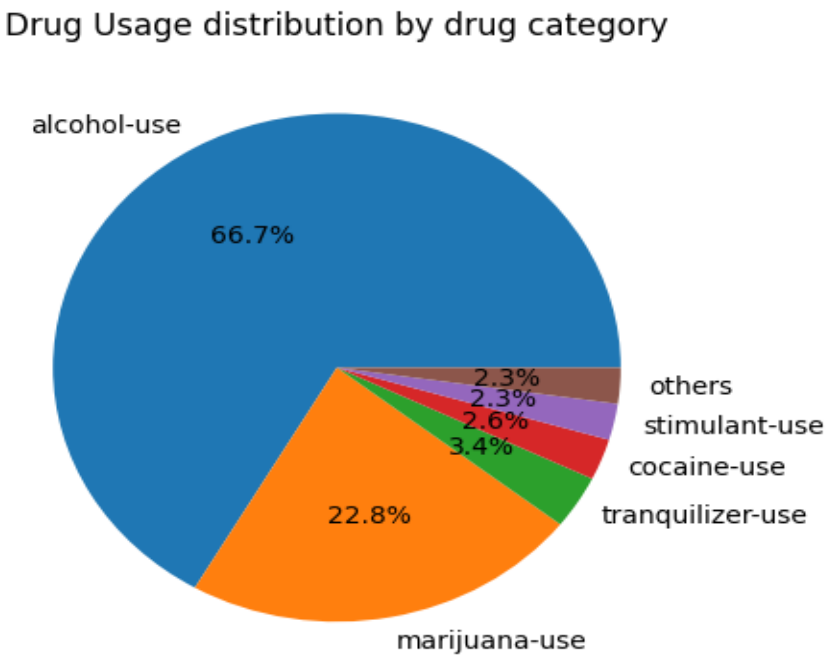
- Use of alcohol and marijuana is higher and consistent than other drugs.
- Alcohol use maintains a high frequency over all the age groups.
- For the most part, drug use tends to decline with age.

Visualization 2: PIE CHART(DRUG USAGE DISTRIBUTION BY DRUG CATEGORY)

The second visualisation method used is pie chart.

A piechart is a circular statistical plot used to depict a set of data.Here piechart is used as the best way to plot the drug use and their percentage over other drugs.

This plot focuses on the top most used five drugs and their relative percentage over the total drug usage.Additionally ‘others’ represents the summation of drug usage which falls out of the top 5 most used drug category. Each slice shows the size of the drug usage distribution by drug category



From this plot we can conclude that,

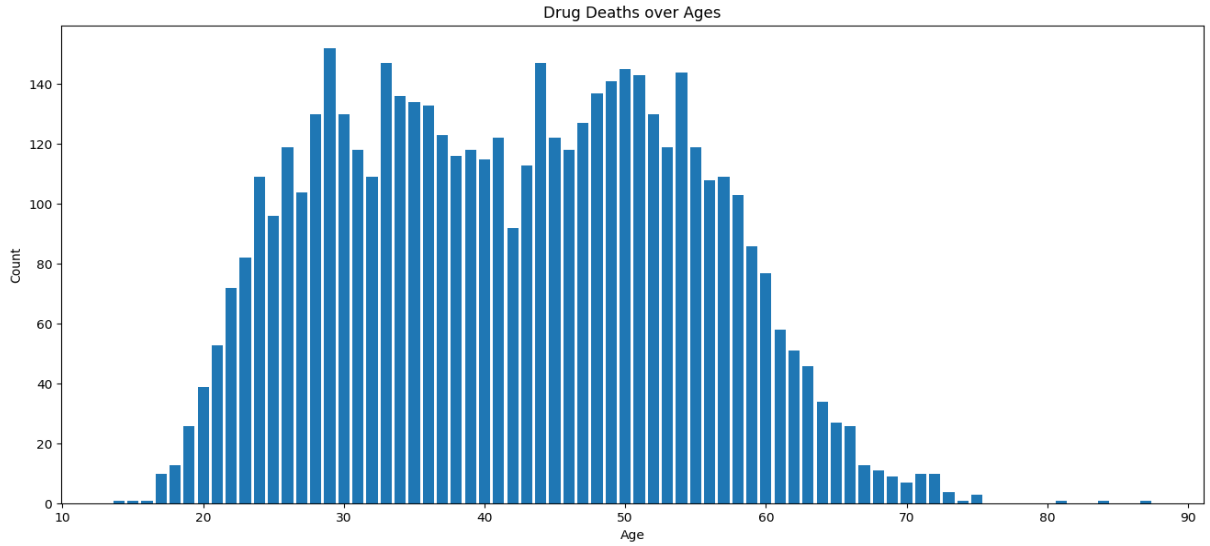
- The use of Alcohol is more than the 50% of total drug usage
- Alcohol and marijuana are the topmost used drugs among other drugs with a percentage of 66.7 and 22.8 % respectively
- This plot aids in determining the top five most used drugs and their contribution to the overall drug consumption
- The pie chart's many slices reflects to various drug categories, giving viewers a visual representation of drug usage over other drugs

Visualization 3: BAR PLOT (DRUG DEATHS OVER AGES)

The third visualization method used is bar plot.

The bar plot shows the death due to the use of various drugs over different age groups from the 'drug_deaths.csv' dataset.

Bar plots are best used to depict the data using bars of different heights.



It is the best way to illustrate the drug deaths over different age groups. The Bar plot summarizes,visualizes and compares the drug deaths across all age groups.

From this plot we can conclude that,

- In this graph Age data is plotted across the x-axis and drug deaths along the Y-axis.
- From the graph we can find out that highest deaths fall in the age group 29,33,44 and 54s
- Death rates are least in early 20s and after 60s

CONCLUSION

In this assignment three visualization methods are used ie,line plot,pie chart and bar plot.

These various visualization datas offer various perspectives on drug frequencies, drug usage and drug deaths over different age groups and drug categories.

CSV file is read using pandas and graphs are plotted using matplotlib.

Each visualization provides unique information ,percentage and comparative values about the specified datasets.