Competitive Analysis on the U.S.A. health care system: Analysing Depression, Death & Obesity

Linu Anil
(x23183853)
Master of Science in Data Analytics
National Collage of Ireland
Dublin, IRELAND
x23183853@student,ncirl.ie

Dipesh patel
(x23197609)
Master of Science in Data Analytics
National Collage of Ireland
Dublin,IRELAND
x23197609@student.ncirl.ie

R Anantha Padmanabha
(x22242520)
Master of Science in Data Analytics
National Collage of Ireland
Dublin, IRELAND
x22242520@student.ncirl.ie

Abstract—This project aims to analyze key aspects of health care in the United States , with a focus on issues pertinent to the elderly population. Utilizing datasets issued by the U.S Health & Human services, This study involves Three primary sources of data set a csv file containing data of old age related health topic and two Json data set containing data related to nutrition and accidental drug related death . Due to the findings of our analysis we got to know about U.S citizen mental health and the group most affected by it, which drug overdose killed most of the people at what age and depression is real thing

Keywords—Python, Mongo DB,MYSQL, Github ,ETL

I. INTRODUCTION

After pandemic health and health consciousness became trending topic and death is the end of health .In this project we are doing analysis on 3 different factor that are affecting human life one of them is effect of age on human health (age group above 50 years) ,bad habit like drugs or people life choice. All 3 Data set fetched from DATA.GOV it's U.S.A government portal where all the government open data set is provided . Based on those 3 data set research include competitive analysis based on 3 different branches of U.S health care. First will be on challenges based on old age, self-inflicted death because of a bad habit and ignorance towards persons own health. Based on this analysis in report its shown some stereotypes are correct and some were not.

II. LITERATURE REVIEW

Freya Tyrer, Francesco Zaccardi, Kamlesh Khunti, Richard Morriss[2] where studied depression and obesity are interlink with each other in their studies they did there research on 519,513 adults, incidence of depression was 9.2 per 1,000 person-years and was higher in women and in 40- to 59-year-old men who had severe obesity.

Borne, Jessica MD*; Riascos, Roy MD[3] they did their research about opioids and morphine derivatives where they found out heroin acts on three types of brain receptors.George A. Bray[4] in his study he studied the effect that can happen on human health related to obesity like it might be environmental or genetic he sorted out all the probable cause of obesity in his research.

in research on interlink between obesity and depression[2] the conclusion was they guided antidepressant drug and specific service for people how are obesity and depression but this report conclusion is little bit different.

in drug abuse[3] its stated that how drugs effect brain based on that this report specifically focuses on death related to drugs.medical consequences of obesity enlighten the information about how many factors can effect obesity[4] in this report analysis is based on U.S citizen obesity rate per state.

III. METHODLOGY

*A.*Detailed description of data

1) Dataset 1: The data set "Alzheimer's disease and Healthy aging"**.**This data set contain 31 column with information done by American federal government about diseases old age American people are facing ,it sorted through years , state , gender ,race and data value in percentile where the research done by them on every topic like "Percentage of older adults with a lifetime diagnosis of depression".the data set contain research done by federal government from 2015 to 2021.To do the analysis on data collected in CSV format.

2) Dataset 2: The "Accidental Drug Related Deaths 2012-2022"provide information about deaths associated with overdosing of different type of drugs in United States. The dataset contains 48 columns and 10,653 rows. It includes details about death rate, location, cause of death, caused drug, race, gender, age and other attributes. This information can be used to analyze which drug is more dangerous, what category of people are consuming it the most, like gender, age group, city and so on.

3) Dataset: "Nutrition, Physical Activity, and Obesity - Behavioural Risk Factor Surveillance System". The dataset provided is sourced from DATA.GOV a US government website and it provides insights about nutrition, physical activity, and obesity behaviours among adults in US. The dataset contains 34 features which includes Start_year, End_Year, State, Topic, Question, Race/Ethnicity and may more with 9000 rows. The dataset contains informations from the year 2020 to 2023.

*4)*Data precessing

In this research Data per-processing includes Extract,Transform and Load (ETL) method. Extraction is a process of collecting data from from different sources and in different file format.the file format is divided in 3 types as shown in fig 1. Structured, semi-structured and unstructured.Then the transformation process involves removing garbage or null values from data set, combining data, rename or edit data entries and then the modified data loaded into database in structured format to perform analysis .

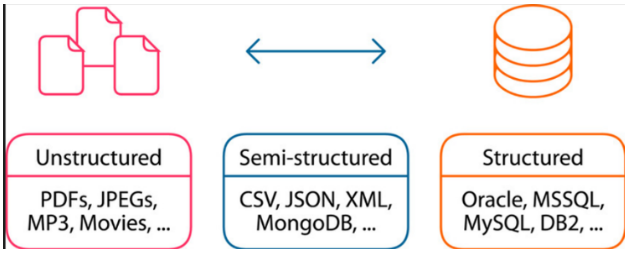

Fig 1 Collectible data

Table 1 Data Set

Data set	Name of data set	Type
1	ALZHEIMER'S DISEASE & HEALTHY AGING DATA	CSV
2	ACCIDENTAL DRUG RELATED DEATHS	JSON
3	LEADING CAUSES OF DEATH: UNITED STATES	JSON

1) Data Collection: For this process first data set collected 3 data set from DATA.GOV in 2 different file format as shown in Table 1.DATA.GOV is federally funded open data portal where all the federal government agencies load their data in semi-structured or unstructured format. Based because of that its one of the best site to gather data for data analysis

*2)*Data Loading:- To load the data MongoDB Atlas is used its a cloud-hosting part of MongoDB. 1 of the 3 data set had different file types and MongoDB accepts JSON format and records are too high in the CSV data set before Loading the data 10 thousand random records were created and then it converted into dictionary format after the conversion. The connection was established with MongoDB Atlas using pymongo Library and connection code given by cloud platform after creating the cluster data set loaded from individual system to the cloud platform in the same cluster data can be seen in Fig 2.

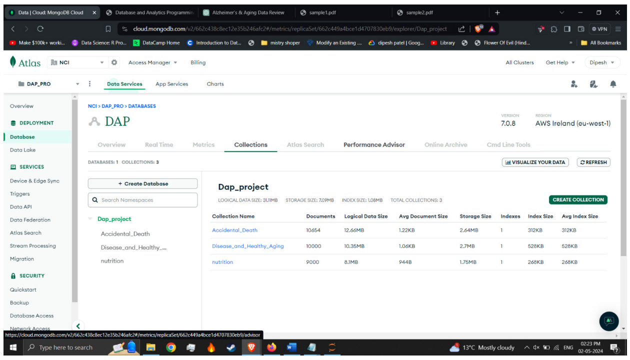

Fig 2 MongoDB Atlas

*3)*ETL pipeline: As shown In Fig 3 of data flow Architecture after loading the data it was extracted from the cloud after extraction, transformation is done in it data set were check.

In data set 1 in transformation process basic analysis were performed basic analysis were like checking information, shape which defines rows and column of data, is.null() to check null values of data set.

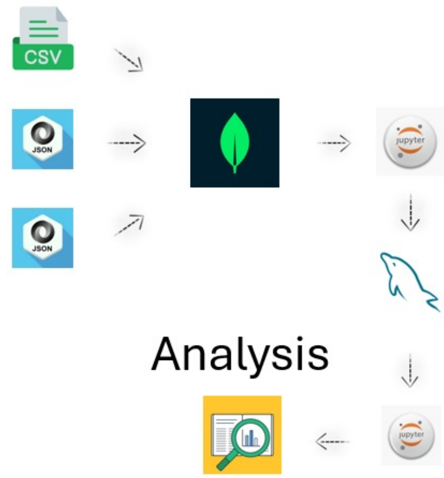

Fig 3 Data Flow Architecture

After this initial analysis high null values column were removed and column having data types of another column were merged like Data_value_type had “%” that column merged with data value and the output were similar like “24%”in dataset before transformation there are 31 column and 10 thousand rows but after transformation there are only 13 left with close to 7 thousand rows. In data set 2 the transformation process started with some basic analysis like checking information, shape, head, names of columns to get an overall idea about the dataset. After these initial analysis, it became clear that some transformation must be done in the dataset. So as a first step the ‘Y’ values in each drug columns are changed to its actual abbreviations of drugs and then combined different drug columns to one column and name the combined column as ‘Caused_Drugs’. After that dropped the different drug columns from the DataFrame. Secondly, check for missing values in each column using ‘isnull()’ method. Then it was observed that many columns have missing values. Moreover, four columns contain more than 50% missing value. So drop the four columns that has more than 5o % missing values and for the other columns use ‘Mode Imputation’ to fill the missing values because all the columns are in object datatype. Then rename the columns as a percussion because some of the column names has space in between them so when the data frame load to MySQL it will show some error. In data set 3 the Transformation process undergone through some basic analysis using ‘info(),’ ‘shape’, and ‘isna()’ methods and understand the information about data. The dataset’s columns contains three different type of datatype like int, float, and object and also many missing values and irrelevant columns. After basic analysis the irrelevant columns were eliminated and drop the rows with missing values. Additionally, renamed some column names to appropriate name for easy analysis.

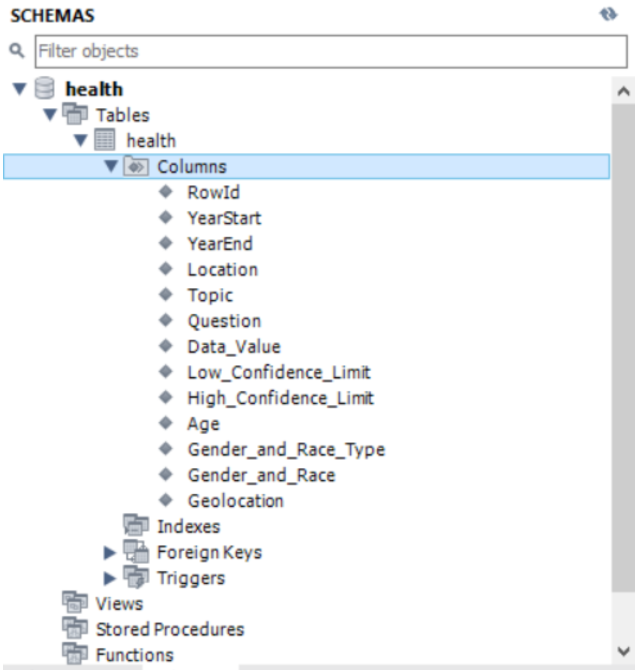

Fig 4 MySQL TABLE

After transformation data Loading is performed to load the data sets to MySQL .first connection were created to My Sql localhost, the cursor object created to interact with database by executing Sql commands with the help of python . Using cursor data base is created and loaded the filtered data from data set to MySQL final output shown in Fig 4

IV. RESULT AND EVALUATION

Data set 1:-

*4)*In this fig 5 report shows the top 5 state with adult with a lifetime diagnosis of depression across U. S state from 2015 to 2021.which includes state like Oregon ,Ohio, Michigan, Alabma, Washinton, Road, Island, the sharpest decline is stars from 2019 to 2021 in it Washington Shows dramatic decrease. The reasons behind this include mental health services, or societal shifts affecting mental health (like he COVID-19 pandemic) because of quarantine most of the family sated at home and grandchildren or family can spend more time towards their family member. Fig 5 shows different coloured line per state so is easy to track. Understanding this trend could help policymaker and health professionals design plan for old adult facing mental health issues [1]

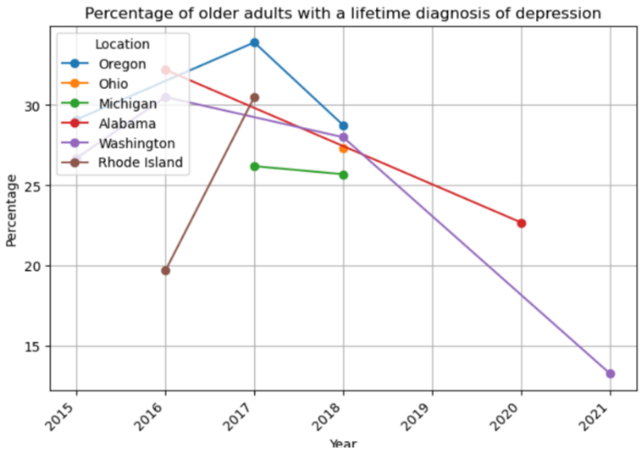

Fig 5 Data set 1 Chart 1

*5)*After doing the analysis on states wise mental health analysis report went one step further to do analysis of gender of old adult are as shown in fig 6 . females

*6)*percentile with depression is much more than male from 2015 to 2022.

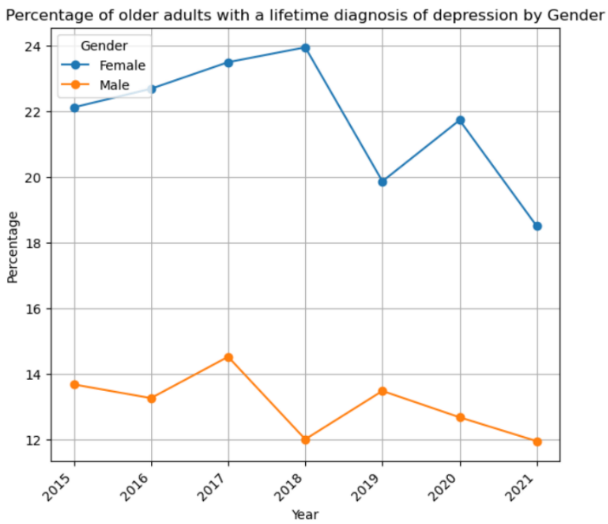

Fig 6 Data set 1 Chart 2

*7)*From highest 23% in 2015 to lowest 18% in 2021 on the other hand man depression diagnostic starts from 14% which is 4% lower than the man 2nd highest percentile. There is one interesting fact in this analysis that it’s goes against stereotype that man are having more stress and depression than female and another fact was shown in Fig 6 is that every year Females are having drastic change in the chart but man are having stable between 14 to 12 percentage like nothing can change their depression. Further studies are need to explore behind the volatility in depression rates, especially among older females.

*8)*As shown in fig 7 This bar chart depicts the percentage of older adults with a lifetime diagnosis of depression, segmented by race and year from 2015 to 2021. The bars represent two racial groups: Hispanic (in blue) and White, non-Hispanic (in Orange). Over the seven years displayed, both groups show fluctuations in the percentages of depression diagnoses. The trend for both groups appears to peak in in 2018, followed by a general decline in 2019 and a slight rebound in 2020. Overall, White, non-Hispanic adults consistently report higher percentages of depression diagnoses compared to their Hispanic counterparts, with the gap between the two groups varying across the years. This visualization helps in understanding the racial disparities in mental health diagnosis among older adults over time.

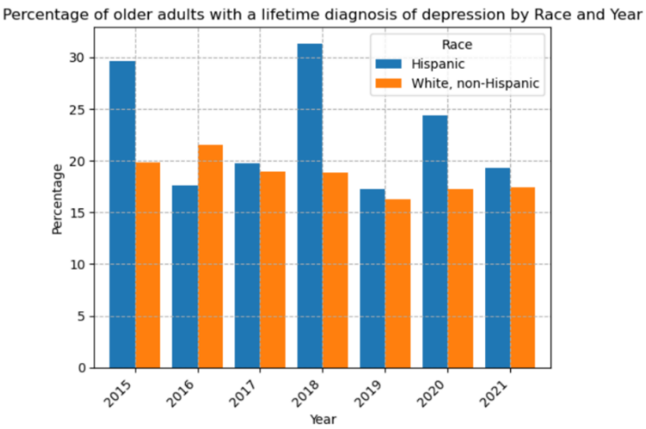

Fig 7 Data set 1 Evaluation 3

B. **Data set 2:-**

*1)*In first visualization of data set 2 shown in fig 8 useed a bar plot to compare the number of deaths by drug intoxication among genders. With the help of this chart, Report observed that there is a dramatic difference between males and females' death rate. When comparing the death count more than twice the number of males is dying as compare to female death count by drug toxicity.

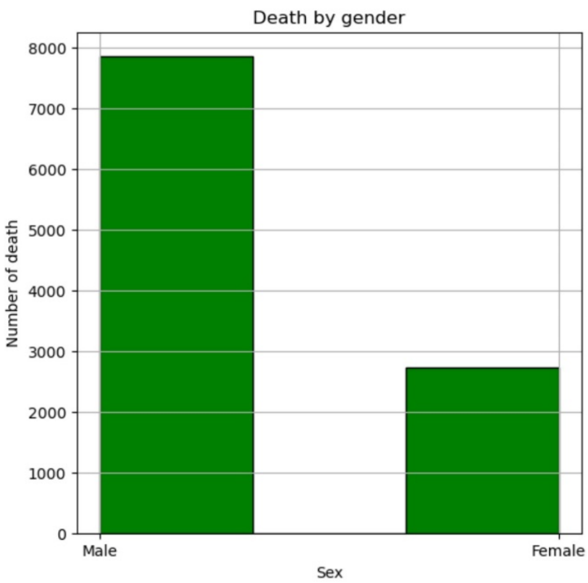

Fig 8 Data set 2 Evaluation 1

*2)*Analyzing the death rate by gender in different age group I used bar plot. For this analysis as shown Fig 9, first grouped every age into different age categories from 10 years old to 80+ years old. Using For loop inside a function assign each age to its appropriate category. From this bar plot also, it is evident that males are huge in number of deaths compared to female in drug related accidents. From this age group 30-39, 40-49 and 50-59 are showing majority of death rate in males and female, although in these categories of age females are only less than half the percentage of men. It the fact that the there is a small portion of children from the age 10-19 and adults above the age of 70 years are also have died because of drug intoxication.

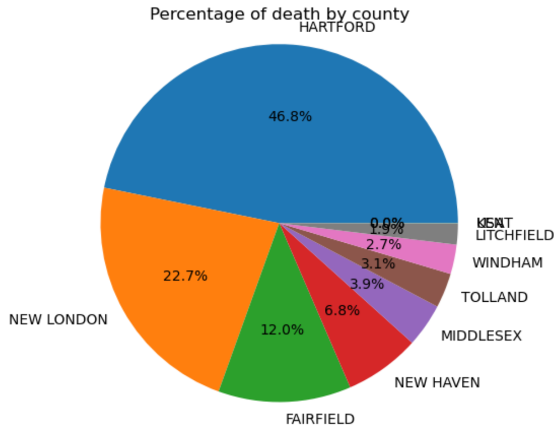

Fig 9 Data set 2 Evaluation 2

*3)*Further in the second analysis I used pie chart to visualize the percentage of death by county. From pie chart in Hartford has most number death from drug toxicity compared to other states about 46.8%. New Haven and Fairfield is showing 22.7% and 12% of drug accident death rate, respectively.

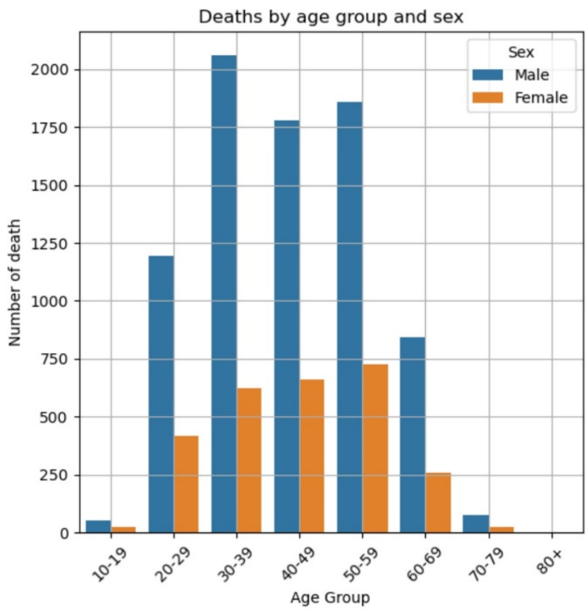

Fig 10 Data set 2 Evaluation 3

6.8% percentage of people were dying by drug usage in New London. For Litchfield, Middlesex, Windham, and Tolland is showing less than 5% of death rate. Overall, Hartford state having a huge percentage of death by drug while Tolland having the least percentage.

*4)*For the further analysis use two graphs, one illustrates the comparison of death rate between male and female among different kind of drugs combinations and the second chart depicts the death rate between genders in intoxication of drug. Males are always showing a substantial number than female in death rate.

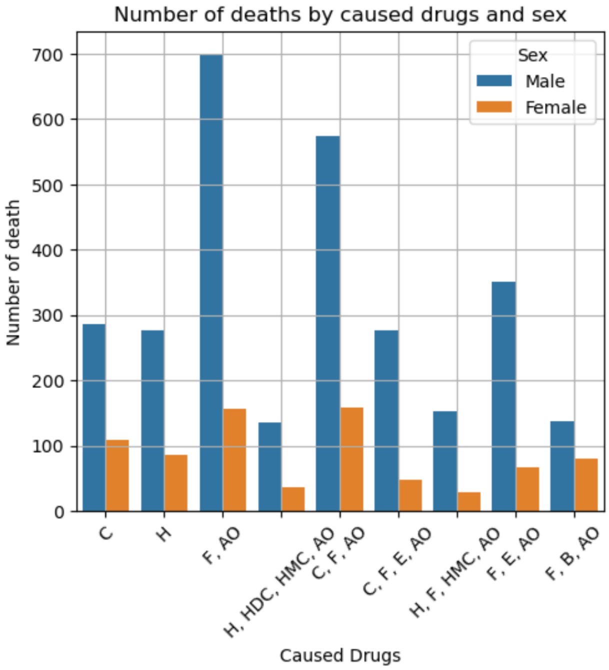

Fig 11 Data set 2 Evaluation 4

Among different type of drugs combinations most of the Males died because of Fentanyl and Any Opioid and Cocaine, Fentanyl and Any Opioid drug combination also Acute Fentanyl Intoxication is major reason of their death. similarly, females most of the died because of the same drug combination and the same intoxication as males but the numbers of death in males are huge in number compared to females. among the drug combinations the least affected drug combination for male death rate is Heroin, Heroin death certificate (DC), Heroin/Morph/Codeine, and Any Opioid, Fentanyl, Benzodiazepine and Any Opioid and Heroin, Fentanyl, Heroin/Morph/Codeine and Any Opioid and the Intoxication for death is Heroine Intoxication, Acute Heroine Intoxication, Acute cocaine Intoxication and Multiple drug toxicity. While in female Heroin, Fentanyl, Heroin/Morph/Codeine and Any Opioid, Heroin, Heroin death certificate (DC), Heroin/Morph/Codeine and Any Opioid and Cocaine, Fentanyl, Ethanol and Any Opioid from these drugs the intoxication that cause their death are Heroine Intoxication, Acute Heroine Intoxication and Acute Cocaine Intoxication.

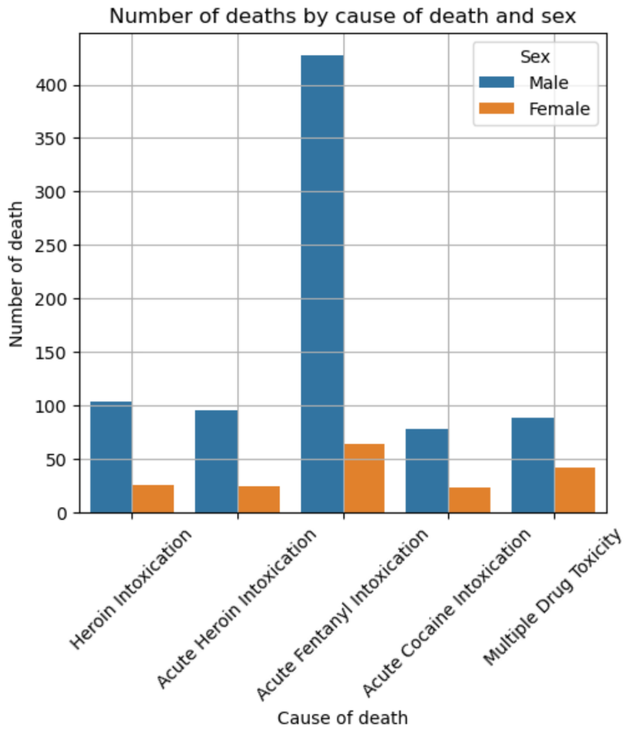

Fig 12 Data set 2 Evaluation 5

C. **Data set 3:-**

1) From years 2012 to 2022 the chart will display the percentage of adults aged 18 years and older who are obese within five regions: Virgin Islands; West Virginia;Guam;Louisiana;Kentucky. Each region is depicted by a separate colored line on the graph: Virgin Islands (purple), West Virginia (orange), Guam (green), Louisiana (red), and Kentucky (blue).

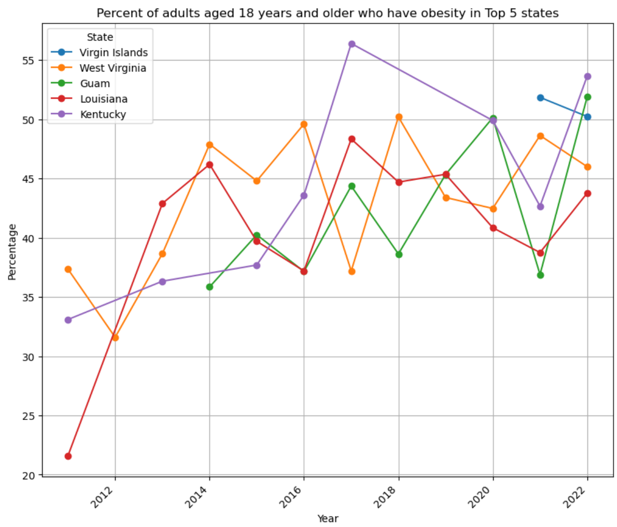

Fig 13 Data set 3 Evaluation 1

The graph reflects significant variations occurring in each region throughout different years.A good example of this would be West Virginia and Louisiana which both reveal significant peaks and troughs in the rate of obesity, indicating that there are various differences in obesity rate changes over time. Kentucky and the Virgin Islands witness a sudden rise before different levels of volatility take on. The trend in Guam is somehow constant albeit with slight differences. The situation changes however when it comes to 2022 data points as nearly all other regions register either an upward progression except for Louisiana which records a decrease.This graph aids in pinpointing the trends and comparisons in obesity rates between these areas spanning ten years emphasizing the ever-changing nature of obesity related public health problems. As shown fig 13.

2)The graph illustrates the percentage of adults in five regions- virgin islands, Puerto Rico, Guam, Utah and Nevada which reports consuming vegetables less than one time daily from 2017 to 2021. Each bar represents a year, segmented by colour to denote each region: Yellow for the Virgin Islands, Orange for Puerto Rico, Green for Guam, Purple for Utah, and red for Nevada. Over these five years. Nevada consistently shows the highest percentage of adults with low vegetable consumption, whereas Utah consistently shows the lowest. The percentages remain relatively stable for each region throughout the period, suggesting little change in dietary habits concerning vegetable intake. This stabilization highlights the ongoing challenges in promoting higher vegetable intake in these regions. As shown in fig 14.

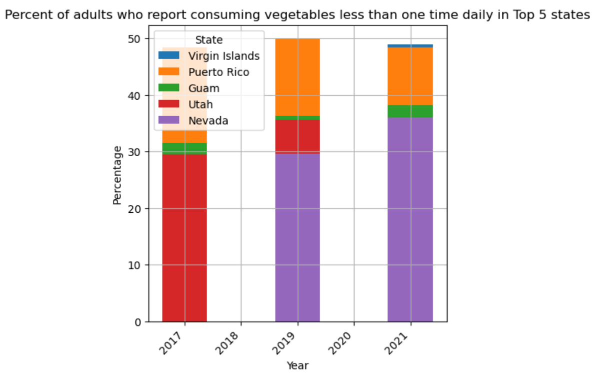

Fig 14 Data set 3 Evaluation 2

V. CONCLUSION

Dataset 1: The analysis of the depression among old adult across different demographic like state, race, gender over the years 2015 to 2021 reveals crucial insights into mental health disparities and trend. Female depression having more fluctuation as compare to male and which is to clear the old stereotype that male are more depressed than female, and location can affect your depression level. State data complicates the picture showing culture makes significant role in mental health outcomes. The analysis has sudden fluctuation because of covid 2019 if there were no pandemic the result of the analysis would have been different.

Dataset 2: The analysis collectively emphasize the need for targeted on drug accidents that consider gender, age, states, Type of drugs and specific drug intoxications. The level of death rate among males due to drug toxicity is always higher than females in every situations. According different locations the drug related accidents varies. These data driven insights can help in crafting more effective health policies.

Dataset 3: The visualizations highlights the current public health problems in improving healthy diet habits to manage obesity. The problems according to obesity is surging as years progress. The data suggested that there is a need for targeted interventions that consider regional dietary trends such as access to healthy foods and effective obesity prevention programs. These interventions should be adaptive to the changing dynamics observed in the graphs and tailored to address the specific needs and circumstances of each region to foster health outcomes.

VI. REFRENCE

[1]https://www.youtube.com/watch?v=RILt88vqEgY

[2]Freya Tyrer, Francesco Zaccardi, Kamlesh Khunti, Richard Morriss"Incidence of Depression and First-Line Antidepressant Therapy in People with Obesity and Depression in Primary Care" https://doi.org/10.1002/oby.22772

[3]Borne, Jessica MD*; Riascos, Roy MD†; Cuellar, Hugo MD†; Vargas, Daniel MD‡; Rojas, Rafael MD"Neuroimaging in Drug and Substance Abuse Part II Opioids and Solvents" https://journals.lww.com/topicsinmri/fulltext/2005/06000/neuroimaging_in_drug_and_substance_abuse_part_ii_.4.aspx

[4]George A. Bray"Medical Consequences of Obesity" https://doi.org/10.1210/jc.2004-0535