

Study of Correlation and Impact of Chronic Diseases

Aafaq Iqbal Khan
x20108851

National College of Ireland
Dublin, Ireland
x20108851@student.ncirl.ie

Bharti Mehataari
x20175825

National College of Ireland
Dublin, Ireland
x20175825@student.ncirl.ie

Neha Suryawanshi
x20169531

National College of Ireland
Dublin, Ireland
x20169531@student.ncirl.ie

Laiba Rehman
x20144032

National College of Ireland
Dublin, Ireland
x20144032@student.ncirl.ie

Abstract— In the past few decades, there has been an increasing dominance of chronic disorders, with a large number of people being adversely affected by them as well as increase in the number of deaths thus deteriorating the quality of life. Various factors like lack of awareness, negligence, lack of medical facilities, etc. have caused this situation. The aim of the present paper is to identify people with such ailments and find linkages between them. To support this objective, data of health surveys conducted by NHANES is used. This study also mentions the statistical analysis of various chronic diseases like Diabetes, Renal issues, Influenza, HIV and AIDS and also observes the prevalence occurring in them. It also highlights the geographic distribution of diseased people in the United States. The study was performed using four datasets based on health domain. For the purpose of storing and managing this data, MongoDB and Postgresql are used. Exploratory data analysis and visualisations are performed to examine and explain the underlying topic with the help of python libraries like Seaborn, Plotly and Matplotlib.

Keywords—Health, disease, COVID-19, United States, diagnoses, deaths.

I. INTRODUCTION

Any damaging fluctuation from the normal structure or the functional state in the human body can be defined as a disease. The abnormal state can be identified by the common set of symptoms displayed or exhibited by people. In order to identify the abnormal state, it is necessary to understand the normal condition of the human body. [1] With the increasing advancement in technology, medical research has visibly modified the paradigm of health care in the twenty-first century. With the help of health assessment, it has become easy to monitor and prevent any kind of disease development using the measurements of the body composition as well as the values of chemical compositions within the body. Number of researches is carried out by gathering and analysing health data. The analysis is further linked to the humans and their health to improve healthcare for all. The health data analysis helps us gain knowledge about the disease, their cause, symptoms and the number or people affected by it. This enables us to identify people which are at risk of a particular disease in an early stage. Early discovery of the risk of disease then enables one to get better medical care and treatment.[2]

Health data can be retrieved from many resources and one of them is the surveys conducted. Health surveys are a critical resource which are conducted to measure and study the health status, potential risk factors and overall behaviour of a

population. This data, upon further analysis, is critical to various factors like implementation and evaluation of health policies and improvement of healthcare facilities of an economy. [3] Data collected from health surveys and statistical reports of severe diseases diagnosis have been widely used in planning health programs and assessing the public's needs in a given area. This perspective shows a big picture of the population's health, its subjects and the various risks for poor health that exist for different members of the community. This data may also be useful for evaluating the impact on the health of affected groups. For example, data of people having a positive diagnosis of HIV can be assessed between different groups of people to attain any meaningful pattern of the spread of disease.

The study aims to find useful insights and relationships within the deaths caused by different diseases like pneumonia, influenza, Covid-19, and a combination of them state-wise in the United States. As per recent studies, the person who died due to Covid-19 is more likely to have pneumonia and other fatal respiratory distress so there is a separate column in data that represents the number of deaths that mention one or more of the conditions reported while the collection of death data. So, while exploration of the insights, a combination of disease-caused deaths is also considered. In our combined dataset, we have a number of features that we can use to answer the research question like state-wise distribution of deaths caused by influenza.

HIV is also a chronic disease around more than 1.1 million people in the US have HIV. Nearly 1 out of 7 of them are unaware they have HIV. The size of the population that was affected by HIV is relatively small compared to the overall population or but is heavily concentrated among several location/state or race populations. So as per data, we try to identify the states that have more HIV deaths. We also try to find if there is any relationship of HIV/AIDS with any other human disease.

Influenza is a viral infection that is highly contagious, and its peak mostly comes in the winter. Influenza is a common cause of pneumonia but not for all of them. Pneumonia is a lung infection and is particularly deadly for young babies and the elderly. Pneumonia and Influenza are often considered as one because many cases of pneumonia are actually caused by influenza. Moreover, respiratory infections are also a consequence of both influenza and pneumonia just like Covid-

19. Only in the United State, Influenza and pneumonia caused around 14.9 deaths per 100k population in 2018 [4]. In 2019, around 1.7 percent of all deaths in the United States are due to both Influenza and pneumonia. As influenza is a contagious disease and also causes pneumonia so we need to identify which US states are hot pockets for this and have higher deaths due to influenza and pneumonia.

One of the objectives of this project is to get insights from already existing and new data in order to help the researchers, policy makers, health professionals and most importantly the patients to get the idea of eye care services, scope of vision loss and eye disorders services available in the United States. The main objective of the NHANES (from where this dataset has been taken from) is to recognize and gather knowledge on eye and vision health from already existing sources and develop new measures of the prevalence of eye disease and vision loss. The data is analyzed to evaluate; the occurrences of eye disabilities and disorders, the eye-health services usage and at last the discrepancy between the treatment and outcomes of visual health

II. RELATED WORK

A study conducted by Clodagh Toomey et al explains the measurements of body composition in assessing the health of individuals using current technological advancements and few machine learning models are applied to support the conclusion. While the extensive research paper carried out the examination of components relating to adiposity, body mass, bones and other highly somatic terms; it proved useful for gaining knowledge about body mass parameters and how an increase or decrease in the metrics can lead to a possibility of health risk factor. [5]

Another research performed by Syed Wasif et al aimed at examining the lack of knowledge of physically disabled patients suffering from diabetes mellitus type II. The data was collected by conducting interviews and recording the answers of a small population of mixed races in Malaysia. Several factors like the age, socio-economic status, previous medication and check-up details were analysed in this research while the BMI parameter which is highly correlated to diabetic condition was not included in the study. They concluded the research with the objective of requesting special attention and care to be taken by the determinants and gaining knowledge of the disease in addition to the treatment. [6]

A study conducted by CHES (Chinese American Eye Study) motive was to identify the visual impairment and eye disease prevalence based on population amid Chinese Americans. Study was conducted on the age group of 50 years and older who lived in the urban areas. Participants were included in the part of study by going door-to-door of people's houses, through the examined setting in the clinics. They used a modified Grading system of Wisconsin Age-Related Maculopathy, age-related macular degeneration (AMD) was defined as having soft uncleared drusen with the abnormalities of retinal pigment epithelium. After the census, they found that the prevalence rate was 8.1% and that the Chinese Americans

were more prone to AMD when compared with Chinese population living in China. [7]

A study was conducted in Los Angeles by the MEC (Mobile Eye Clinic) of the University of California. In this screening of free vision was conducted by a bus which used to reach Los Angeles residents. They examined the test on around 431 patients for the age group 55-94. Age-related macular degeneration was examined on the patients based on the International Classification of Diseases which includes codes to calculate the several eye diseases on the prevalence estimates. Overall, 5.1% AMD prevalence rate was found and majority of the cases were of age group 75 years and older. [8]

Arthi Ramachandran and colleagues used the electronic medical records and geospatial data to develop a machine learning model. The main objective of this model was to anticipate the patients at risk for dropping out of retained care in a HIV care clinic. It was observed that the HIV patients which are in retained care were more likely to receive the antiretroviral medication and attain HIV viral suppression. This led to effective elimination of the risk of transmitting HIV to others. There was no systematic way to identify the patients, which may fall out of the retained care. [9] Authors drew motivation from this aspect and studied HIV related data to build a machine learning model. The model helped HIV care clinics proactively identify the patients at risk of dropping out of the retained medical care.

Dandan Tang and colleagues worked on 3 datasets which consisted of data of people exposed to HIV/AIDS between 2009 to 2015. This data was collected from Urumqi in Xinjiang, one of the key areas of HIV/AIDS infection. Their main goal was to establish a model which helped in identifying the people with HIV infection. Data was pre-processed and balanced using SMOTE. Post attribute selection, various machine learning models were built. Random Forest model performed the best in identifying the people with HIV infection. This model is helpful in quick screening and diagnoses of HIV/AIDS from a large amount of information [10]

Seung-Ji Kang and Sook In Jung [11] tried to find any consistent trends between coronavirus and age group associated morbidity or mortality, they worked on epidemiology covid-19 data from different countries. In any pandemic, one of the main issues faced by clinical health professionals is to understand the spectrum of illness and pathogenic condition in a vulnerable population according to age. They also reviewed how different hosts according to age responded to Covid-19 infection. However, at that time, the biggest problem that they faced was inconsistent data because of multiple reasons like laboratory authenticity and different medical systems in each country. They concluded that the dynamics of age-specific mortality for coronavirus is almost the same with influenza. In other words, the increasing case fatality with age is not a specific finding for COVID but they mentioned that this finding is entirely based on the available dataset. Their findings will be useful in developing the treatment policies for coronavirus.

III. METHODOLOGY

ETL – Data plays an important role in nearly every business operation for the data to be valuable. We need to gather the data from an unlimited number of sources, organize it together and centralize it into a single repository that indirectly means that we need ETL processes. In the first step of this process, the raw data is extracted from an array of sources and then after applying some calculations it is fitted into a required format which is known as the Transformation phase. The last step includes the formatted records which are stored in structured format in the data warehouse system, this process is the ETL methodology loading phase. For the analysis of data, we have applied the ETL methodology which can be seen as we progress through the paper.

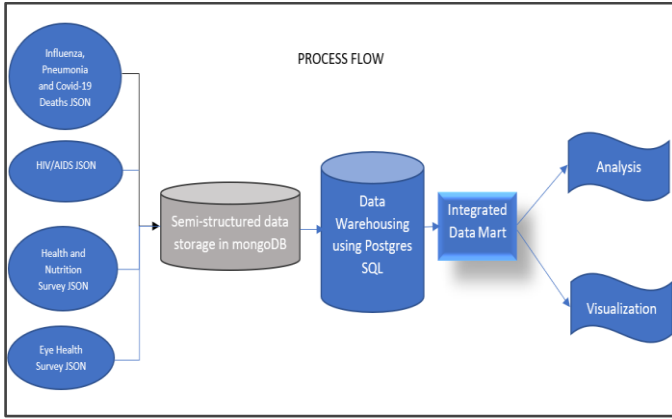


Fig. 1- ETL Methodology

Fig. 1 shows the various steps which are included in our analysis. According to this ETL methodology, from the various data sources, the data is extracted and then utilized for the purposes of prediction and visualization.

A. Data Description

The Eye Health Survey dataset is taken from NHANES website in a semi-structured format in JSON. It is a summary data frame made from vision and eye health data which was gauged from NHANES. It is a unique survey that includes both physical and interview examinations. This dataset is based on all types of combinations that were available such as risk factors which have values like diabetes, hypertension and smoking, age group, gender and race/ethnicity. This dataset includes examinations and questions based on Service Utilization, Eye Health Conditions, Vision Exam Measures, Visual Function, and Examination Measures.

Dataset Health and Nutrition used for this research is collected from datahub.io in a semi-structured format (json). It is based on the respondents of the National Health and Nutrition Examination Survey which involves a combination of anthropometric measurements like height, weight, arm and waist circumference, and other chemical components like Albumin, Blood Urea Nitrogen, Glycohemoglobin and Creatinine which can be used to learn and establish the assessment of an individual's health and nutritional status.

HIV stands for Human Immunodeficiency Virus. As the name suggests the virus attacks the immunity system in the human body which makes it very difficult for the body to fight against infections. If HIV is not treated correctly it may lead to extreme weakening of the immunity system and development of AIDS. The HIV/AIDS diagnosis dataset reported to New York city is used in this study. This data consists of the new diagnoses for the year 2010 through 2013 and is as per the neighbourhood, age group and race/ethnicity. [14]

To address the research questions related to the deaths we need a comprehensive dataset that contains US state-wise deaths statistics of the Covid-19, Influenza, and Pneumonia last but not least, data should have the information of medical conditions like respiratory diseases and circulatory diseases etc. We managed to arrange the 2 datasets, published by the Centers for Disease Control and Prevention. On Data.gov, CDC have published the 2 two JSON datasets. The first JSON dataset, has information of deaths involving Covid-19, pneumonia, and influenza reported to (National Center for Health Statistics) NCHS by race, age, and state. The dataset also has the information of deaths caused by multi diseases like one of the features is *Pneumonia and COVID-19 Deaths*. The second dataset shows health conditions given in conjunction with deaths involving Covid-19. In data 20 features are categorical and 3 are numerical variables. There are a total 235,980 listings define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Data Extraction

The datasets in this research are originally in a semi-structured json format. This kind of file contains the information as key-value pairs. So, in order to read such files, a database platform-MongoDB was used. Each key value pair is stored as a document in the collection of MongoDB. The structure of these documents is quite similar to that of the json files with an addition of a unique ID value, hence it was easy to comprehend the data when put through a particular database platform. We chose MongoDB which is a schema-less database to store our datasets. All datasets are programmatically stored in MongoDB using Python. Once the files were loaded into the collection, pymongo library was used to establish a connection with the database. Pymongo is a native Python driver that allows interaction with MongoDB databases through Python, we were able to retrieve the documents as a dataframe in the Jupyter Notebook and go on with further analysis.

C. Data Preprocessing

1. For Eye Health dataset

According to the ETL methodology, the second phase after completion of the data extraction is the transformation which is implemented using various sources. Once the data is loaded using python programming in the Jupyter notebook few records which had numeric values are converted to numeric dtype by applying the pandas library to_numeric function. I had

categorical data which needed to be converted to numeric in order to use them in the graphs. Thus, by importing the preprocessing module from sklearn library I imported the LabelEncoder() class. I chose two of my categorical variables “RiskFactor” and “RiskFactorResponse” to be converted to numerical variables. An additional column is added in the dataset named “Actual Deaths” which is calculated by dividing the percentage value of “Data_Value” with the Sample_Size and then multiplied by 0.01. This is done to get the death counts of the number of people. The sum of all the null values is taken to know how many missing values are present and in which variables. The null values were checked using is_null function. The dataset is suppressed if the Relative Standard Error is greater than the 30% of the mean or if the sample size is less than 30.[\[12\]](#)

2. For Health and Nutrition dataset

The transformation phase required a lot of data cleaning and preprocessing efforts to avoid any kind of anomalies within the analysis. To start with, the Family_Income variable is converted from a range value into continuous value to get the respondent’s fixed income. Missing values for all the columns are identified and imputed using the median values. Then a median method is selected as it is not influenced by outliers and provides an almost accurate representation of the samples. A user-defined function proved helpful in finding the outliers within the entire dataframe. As the number of outliers in the samples were quite less, they are discarded from the analysis rather than imputation. The dataset consisted of chemical composition values providing information of the different organs of the body and ailment characteristics, hence it seemed significant to learn about each feature and get some prior domain knowledge before transforming them. An increase or decrease in these values could affect the health of the human body or add to the risk of developing any kind of illness. Thus, a significant amount of time was spent to learn these compositions and come up with the valid creation of new features so as to avoid any sort of misinterpretation or miscalculation of the substances. This analysis also proved useful in getting rid of some of the dimensions. An example of this is explained below:

Urea nitrogen is a waste product created by our body to process food. While the liver breaks down the proteins in our food it creates blood urea nitrogen (BUN) and releases it into the blood, eventually ending up in the kidneys. The kidney processes this substance and flushes it out of the body through urine. Creatinine is another waste product released from muscle and protein metabolism which also tells us about the state of the kidneys. Hence, a specified level of BUN:creatinine ratio is used to check for any problems in the kidney. The ratio below or above a certain range may cause intrarenal or prerenal issues.

3. For HIV/AIDS Diagnoses Dataset

On checking for any missing values in the dataset it was observed that there were no values missing but few columns had value ‘*’. This data with ‘*’ values was 0.01% of the total data hence was eliminated from the study. Few of the numeric columns had ‘NA’ value and were 15% of the total data. This data was imputed using the mean of the respective columns. Outliers in the dataset were identified using the statistical method z-score. Calculating the z-score using the mathematical function numpy.abs in python was a challenge as the dataset had categorical features apart from the numerical feature. To handle this issue a data dictionary was created where a key value pair of the data was stored with its respective encoded value. The encoded value was obtained by using label encoding which was also used to convert the categorical columns into numeric values. After calculating the z-score and identifying the outliers it was observed that 984 records were to be dropped. Post removing the outliers, the categorical variables were again decoded using earlier mentioned dictionaries.

4. For Influenza, Pneumonia, and Covid-19 Deaths Dataset

In this step, the datasets were transformed and preprocessed into a structured format. For this, the data was extracted from MongoDB server and converted into a data frame first by Pandas library. Both datasets have some irrelevant columns that need to be dropped, after that to join the both datasets into a single data we have to merge them. We selected only those records that have the same starting date to synchronize them. After joining, the joined data was checked for missing values. There are 6 features that have missing values. But to replace the NAN value with some appropriate imputed value we need to first convert into numeric form as all these features contain a number of deaths which has to be integer. We replaced the NaN value with means. For outlier detection, we checked feature’s distribution and plot Q-Q plot to check the normality within the feature (Figure 2). Although, there are some extreme values detected but we cannot say them as an outlier because of the deaths statistics so there might be some age group or race or region that may be a hot pocket of specific disease so these extreme values are important for us.

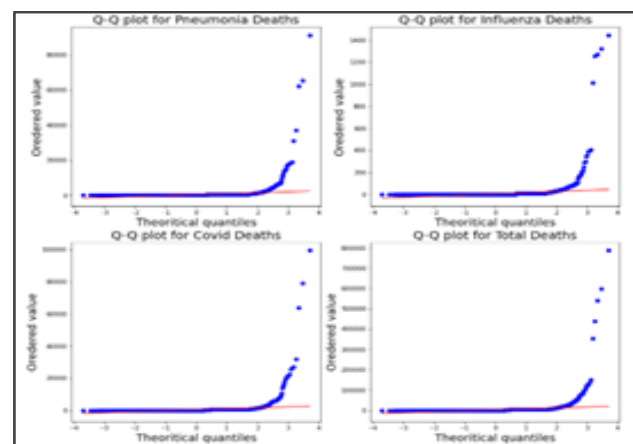


Fig. 2-. Q-Q plots

D. Data Transformation

Post the cleaning and pre-processing of individual datasets, the data was stored into a relational database management system- Postgresql. One of the main reasons to use PostgreSQL is that it is more secure and sophisticatedly stores the data. The database is hosted on Oracle Virtual machine and acts as a central hub for the integration of individual datasets and also enhances the efficiency. The initial step is to define and create the structure of a suitable data object that can hold the data, hence four tables are created on a common database. After inserting data, all the tables have a common attribute “Age” which is used for the integration. However, it was found out that its values were not identical in all the tables, hence, the values needed to be standardised.

We faced a lot of challenges while merging the tables using INNER JOINS as the identified common column does not have matching values throughout the tables. Due to this reason, the inner join keeps on increasing the rows exponentially as it has no primary key common amongst all the four datasets or a unique key constraint. To overcome this problem, we add a new identity index column in all tables and use the combination of age and the newly introduced index column to successfully merge the data frames into a single dataframe

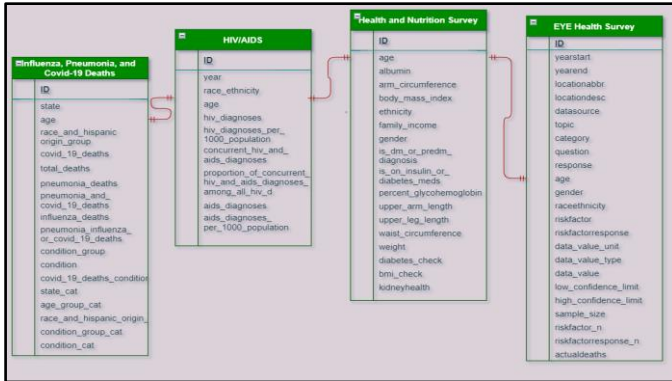


Fig. 3- E-R diagram of the four datasets

An Entity Relationship diagram is made using the draw.io website. Fig. 3 visualizes the tables that are developed in the data warehouse and depicts the relationships that exist amidst them which is mainly based on the one common column i.e. Age.

E. Data Warehousing

Transformed data was loaded into PostgreSQL, an open-source object-relational database system. The steps of connecting to PostgreSQL involved connecting to the database using the credentials, port number and the host name. After a successful connection was established, a new database and table were created to load the transformed data into it. The new table was created keeping in mind the datatypes of the features of transformed data

IV. RESULTS

This section explains the descriptive and statistical analysis and insights gained from the merged dataset. Python offers a

wide range of libraries for the visualisation of data out of which Seaborn, Matplotlib and Plotly are used.

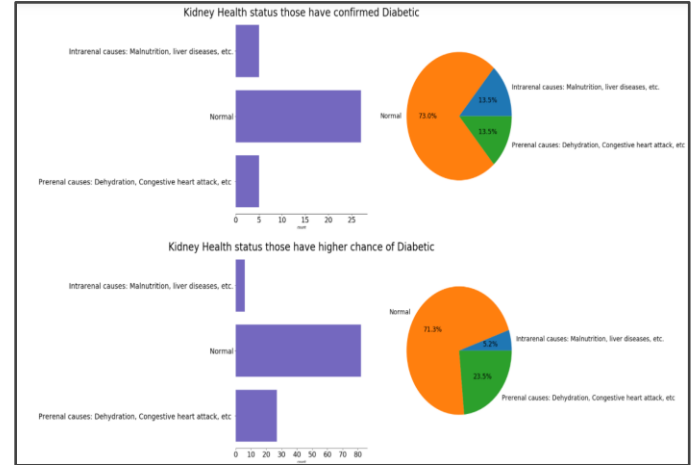


Fig. 4- Status of Kidney Health for Diabetic people

The above graph (Fig. 4) is based on the percentage and counts of people who come in the category of Diabetes but were divided into two further sub categories; one who had confirmed diabetes and another who had a higher chance of having diabetes. Based on these two factors the kidney health status was diagnosed. The insight taken was that for the confirmed diabetes diagnosed people there was an equal number of percentages of kidney failure it did not matter whether they were from the Malnutrition and liver disease category or Dehydration and Congestive heart failure disease category. For the other category of people who had higher chances of getting Diabetes it was seen that Dehydration and Congestive heart failure category people were more prone to Kidney failure. The reason behind this could be that these people had a risk for Type 2 diabetes in which the sugar level of blood is increased and the body becomes insufficient to pump blood throughout the body, thus the chances for heart failure is two to four times higher.[\[13\]](#)

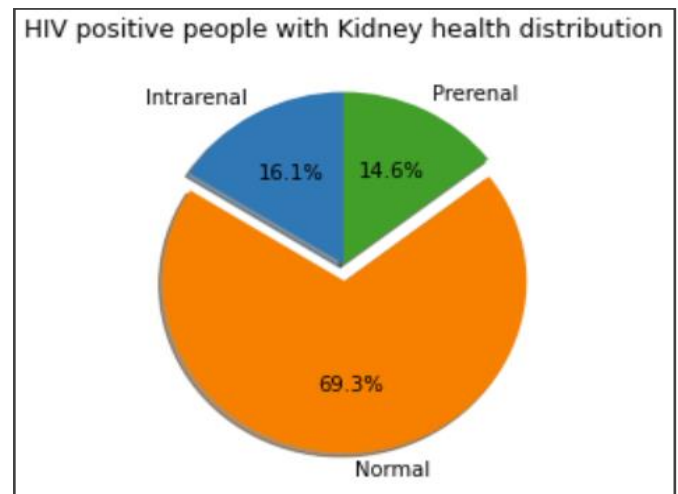


Fig. 5- The distribution of HIV positive people along with their Kidney condition

The above pie chart (Fig. 5) indicates the distribution of HIV positive people and states the condition of their kidneys. It is seen that people having an HIV infection are at greater risk of getting kidney diseases or severe impairment. This can happen because the virus infects the cells in the kidneys and also tries to harm the nephrons thus preventing them to function properly. In fact, some medicines used for the treatment of HIV can adversely affect the kidneys. Hence, health care providers carefully consider the risk of kidney damage when recommending specific HIV medicines to include in an HIV regimen.

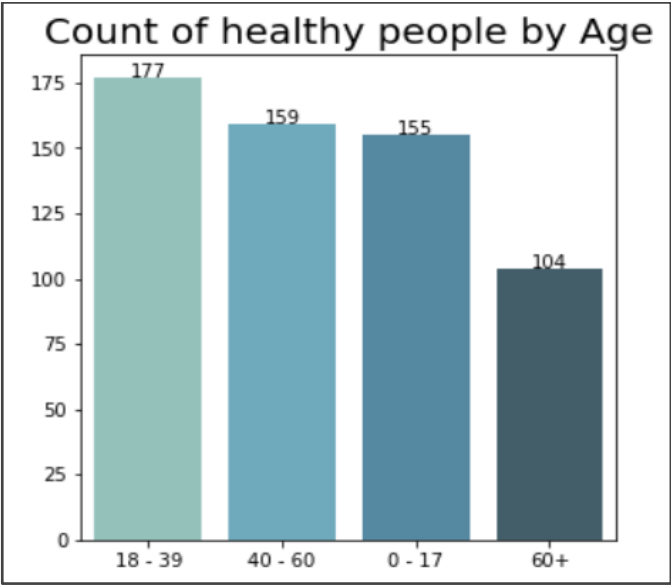


Fig. 6- People who do not have any disease

Amongst the data of people with some or the other kind of ailment, we also try to identify people who have not been diagnosed with any disease or illness thus presuming that they follow proper dietary measures and lead a healthy lifestyle. As per the graph above (Fig. 6), people between the ages of 18 to 39 accounted for the majority that did not suffer from any adverse health diseases as compared to other age groups. The skew in the distribution seems ideal as health deteriorates with age however the low count of healthy/normal people in 0-17 age-group is due to the factors of being underweight (low BMI as per the data) and early smoking habits in teenagers.



Fig.7- Number of HIV Diagnoses State-wise

As HIV is a contagious disease, there some regions that are more affected by the diseases and subsequently have more diagnoses depend upon the spread of the chronic diseases. Although in the US, there are significantly less deaths caused by HIV than other lethal contagious diseases but HIV is concentrated among several location/state or race populations. Fig. 7 shows Maine, Virginia, and South Carolina are the most affected by HIV and have the highest HIV deaths in the US. A report published by the National Centre of HIV/AIDS [14] mentioned that, in eastern and southern states more people are diagnosed with HIV and living with it, so our resultant insight from the below graph supports the report because it is more likely these states have more deaths by HIV.

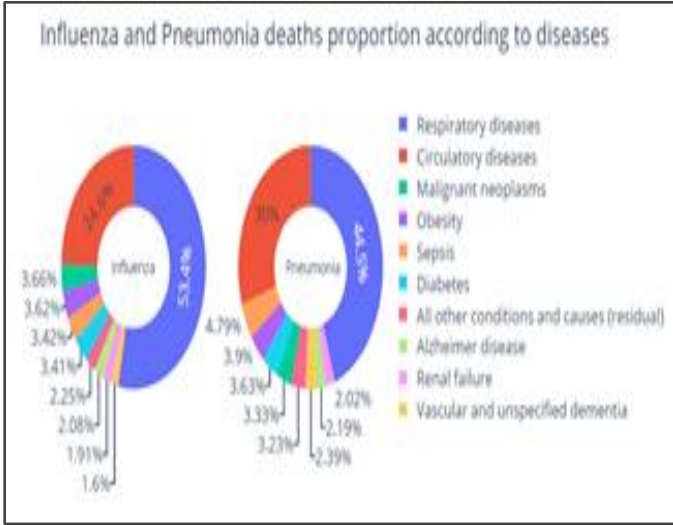


Fig. 8- Influenza and Pneumonia deaths proportion according to diseases

Our dataset has information of medical condition like respiratory diseases, circulatory diseases, sepsis, Diabetes, Renal failure etc. It helps to find the relationship between the condition group and deaths because these pre-existing long or short-term medical condition also play a vital role in specific deaths. A pre-existing condition can be defined as any medical condition that either paves the way for the disease of interest in the sequence of events leading to death, or it can be a contributory factor in the death. Fig. 8 illustrates that in both influenza and pneumonia deaths respiratory disease is the most frequent medical condition. Influenza and pneumonia both badly affect the human respiratory system and if a person already has long or short-term respiratory disease history that combination can be deadly and lead to death as per above visualization.

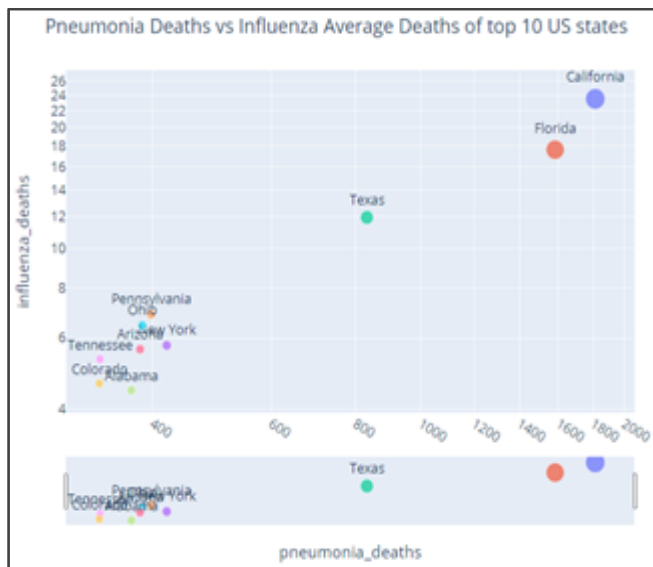


Fig. 9- Pneumonia vs Influenza Deaths of top 10 US states

As per recent studies, the person who has influenza is more likely to develop pneumonia and other fatal respiratory distress. As influenza may cause pneumonia so we need to identify which US states have higher spread and higher deaths due to influenza and pneumonia. The scatter plot shown in Fig. 9 demonstrates that California state has the highest number of deaths not only influenza but also pneumonia. The colour here represents the state and the size of the bubble is according to the pneumonia deaths, more deaths larger is the size of the bubble



Fig. 10- COVID-19 Deaths by Group condition

The above fig.10 demonstrates that people having Diabetes or Obesity or Alzheimer or sepsis diseases are more likely to recover from Covid-19 and develop their immune system to fight against coronavirus. Whereas, coronavirus is more deadly for those who have pre-existed respiratory or circulatory diseases.

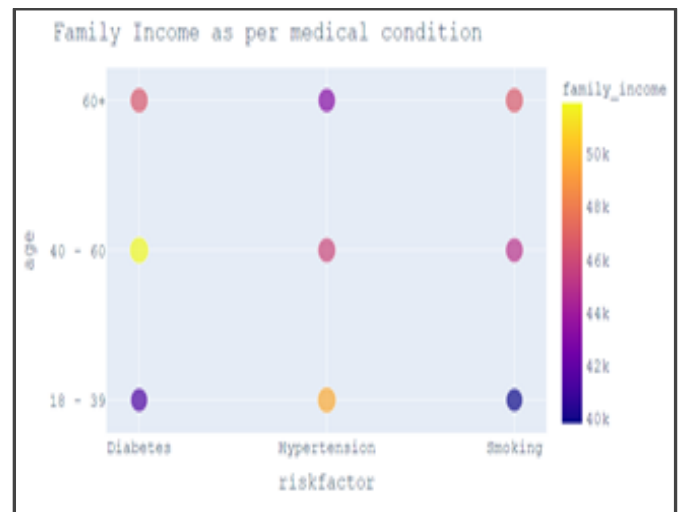


Fig. 11- Family Income as per medical condition

Fig. 11 illustrates the average family income of people as per their medical conditions. The people having more family income and aged between 40-60 mostly have diabetics. There is also one interesting fact that young people aged between 18-39 are more likely to have hypertension disease if they have higher income.

V. CONCLUSION

This study aims at depicting the causes of various chronic diseases and the amount of prevalence accounted in each of them.

On the basis of our findings, we observed that a person having any one of the chronic diseases was prone to catch another one. For instance, our study claims to prove that prolonged influenza would lead to pneumonia, an HIV infection if not taken care of; would eventually cause AIDS which is fatal, increase in the sugar levels has an adverse effect on the kidney and the HIV too accounted for deteriorating the kidney's health. The eastern and southern US states have more mortality of contagious disease like HIV, influenza, and pneumonia.

However, there were a few limitations in the research as the data related to medications taken by the people were not available hence it restricted us from finding any improvements in the population's health and life span. The future plan of this project is to extend the scope of this analysis to different continents of the world and apply the discussed approach to especially African countries where the spread of HIV and other contagious diseases is higher than the United States.

REFERENCES

- [1] D. G. Scarpelli, J. Ernest and H. H. Magerstadt, "Disease", [Online]. Available: <https://www.britannica.com/science/disease> [Accessed on: Apr 21, 2021].
- [2] Health Data Research UK, "Health Data Research Explained", Available: <https://www.hdr.uk/what-is-health-data-research/health-data-research-explained/> [Accessed on: April 25, 2021]
- [3] L. A. Aday, L. J. Cornelius and S. B. Cohen, *Designing and Conducting Health Surveys, A Comprehensive Guide*, 3rd Ed, 2006, pp xi, pp 221-223.

- [4] J. Elflein, "Deaths by influenza and pneumonia in the U.S." [online], Available:<https://www.statista.com/statistics/184574/deaths-by-influenza-and-pneumonia-in-the-us-since-1950/> [Accessed on: April 17, 2021]
- [5] C.Toomey, A. Cremona, K. Hughes and C. Norton, "A Review of Body Composition Measurement in the Assessment of Health", March 2015. Available:https://www.researchgate.net/publication/273459067_A_Review_of_Body_Composition_Measurement_in_the_Assessment_of_Health/link/5aa914350f7e9b88266e1af4/download/. [Accessed on: April 24, 2021]
- [6] S. W. Gillani, S. A. Syed Sulaiman, M. I. Mohammad Abdul, S. Y. Saad, "A qualitative study to explore the perception and behavior of patients towards diabetes management with physical disability" [Online], Available:<https://dmsjournal.biomedcentral.com/articles/10.1186/s13098-017-0257-6>, [Accessed on: April 16, 2021]
- [7] Varma, et al. (2016) "The Chinese American Eye Study", [Online]. Available:<https://www.norc.org/PDFs/VEHSS/EyeConditionExamLiteratureReviewVEHSS.pdf>", pp-21, 2010-2013
- [8] Haronian et al. (1993) "The University of California, Los Angeles Mobile EyeClinic" Available: <https://www.norc.org/PDFs/VEHSS/EyeConditionExamLiteratureReviewVEHSS.pdf>, pp 38, 1982-1990
- [9] A. Ramachandran, A. Kumar, H. Koenig, A. D. Unanue, C. Sung, J. Walsh, J. Schneider, R. Ghani and J. P. Ridgway, "Predictive Analytics for Retention in Care in an Urban HIV Clinic", Apr 2020, doi: 10.1038/s41598-020-62729-x
- [10] D. Tang, M. Zhang, J. Xu, X. Zhang, F. Yang, H. Li, L. Feng, K. Wang and Y. Zheng, Application of Data Mining Technology on Surveillance Report Data of HIV/AIDS High-Risk Group in Urumqi from 2009 to 2015, Dec 2018. doi: <https://doi.org/10.1155/2018/9193248>
- [11] S.Ji Kang and S. In Jung, "Age-Related Morbidity and Mortality among Patients with COVID-19", Infect Chemother. 2020 Jun;52(2):154-164, doi: 10.3947/ic.2020.52.2.154
- [12] J. Wittenborn, C. Davidson, D. Rein, "Survey Analysis Plan Vision & Eye Health Surveillance System", Available: <https://www.norc.org/PDFs/VEHSS/VEHSSSurveyAnalysisPlan.pdf>, pp-9, 2018 [Accessed on: April 21, 2021]
- [13] American Heart Association News, "Diabetes and heart failure are linked; treatment should be too" <https://www.heart.org/en/news/2019/06/06/diabetes-and-heart-failure-are-linked-treatment-should-be-too> [Accessed on: Apr 23, 2021]
- [14] Centre for Disease Control and Prevention, "HIV in the United States by Region", [Online]. Available : <https://www.cdc.gov/hiv/statistics/overview/geographicdistribution.html> [Accessed on: Apr 24, 2021]