Covid-19 Visualizations

Machine learning is being used to learn more about the coronavirus pandemic, or COVID-19. By using machine learning algorithms, researchers can analyze large datasets of patient data to better understand the disease and identify possible treatments. For example, machine learning can be used to identify common symptoms of the disease accompanying the covid itself , predict outcomes for patients, and identify risk factors for severe illness. Machine learning can also help researchers better understand the long-term effects of the virus and the effectiveness of treatments. By utilizing machine learning, researchers can continue to improve our understanding of the virus and develop more effective strategies for fighting it.

dataset link on kaggle :

https://www.kaggle.com/datasets/meirnizri/covid19-dataset

The dataset was provided by the Mexican government .

This dataset contains an enormous number of anonymized patient-related information including pre-conditions.

The raw dataset consists of 21 unique features and 1,048,576 unique patients.

In the Boolean features, 1 means "yes" and 2 means "no". values as 97 and 99 are missing data.

Data Description as per Kaggle:

• sex: 1 for female and 2 for male.

• age: of the patient.

• classification: covid test findings. Values 1-3 mean that the patient was diagnosed with covid in different degrees. 4 or higher means that the patient is not a carrier of covid or that the test is inconclusive.

• patient type: type of care the patient received in the unit. 1 for returned home and 2 for hospitalization.

• pneumonia: whether the patient already have air sacs inflammation or not.

• pregnancy: whether the patient is pregnant or not.

• diabetes: whether the patient has diabetes or not.

• copd: Indicates whether the patient has Chronic obstructive pulmonary disease or not.

• asthma: whether the patient has asthma or not.

• inmsupr: whether the patient is immunosuppressed or not.

• hypertension: whether the patient has hypertension or not.

• cardiovascular: whether the patient has heart or blood vessels related disease.

• renal chronic: whether the patient has chronic renal disease or not.

• other disease: whether the patient has other disease or not.

• obesity: whether the patient is obese or not.

• tobacco: whether the patient is a tobacco user.

• usmr: Indicates whether the patient treated medical units of the first, second or third level.

• medical unit: type of institution of the National Health System that provided the care.

• intubed: whether the patient was connected to the ventilator.

• icu: Indicates whether the patient had been admitted to an Intensive Care Unit.

• date died: If the patient died indicate the date of death, and 9999-99-99 otherwise.

Preprocessing:

Text

Description automatically generated

-Using Value counts function to get the count of how many entries are present for each unique feature per column.

Text

Description automatically generated

-Since it was declared in the Dataset description that 97,98 &99 meant “missing values” ,also columns as ‘INTUBED’&ICU’ had more than 85% (total above 1M ) of their entries containing nulls ,therefore they were dropped with the exception of pregnancies as it will be dealt with later

Text

Description automatically generatedChart, waterfall chart

Description automatically generated

-As shown in the figure above, You can see that 97 (green bar) is always the value of pregnancy feature for males, So we changed 97 into 2( which means no pregnancy)

-For further assurance both Males & Males with 97 values have the same number of entries (511936)

Text

Description automatically generated

-Since “9999-99-99’” meant that the patient was discharged from the hospital it was figured it’s easier to use ‘alive’, also ‘Dead’ if a death date is present in a new column to be used later.

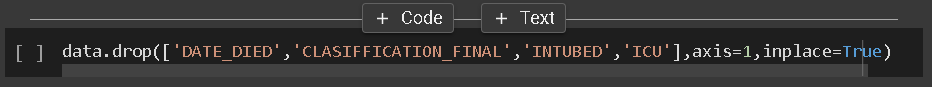
-Same goes for “Classification” column 1,2&3 meant that the patient is a carrier of the virus while any number above 3 meant that the patient is safe.

-Segmenting the DATE into days, months & years is much easier for extracting times where the infection rate was the highest.

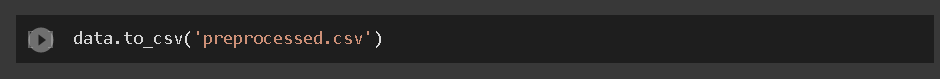
Text

Description automatically generated

-Grouping the ages of people in known discrete groups helps in identifying which age group is at most risk and must be prioritized while giving out vaccines .

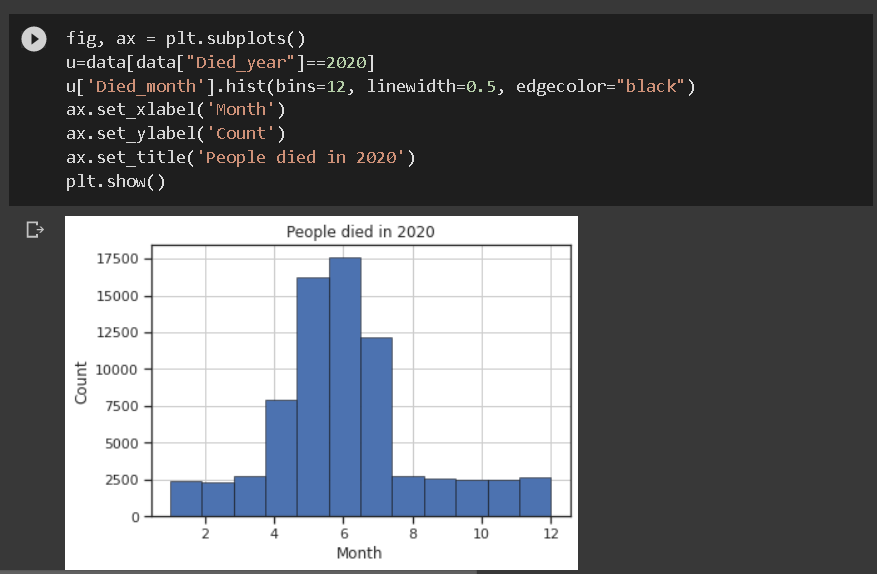


-Dropping unnecessarily columns that are not being used or replaced with more informative columns



-Exporting the preprocessed Data into a new CSV file to be used in visualizations.

Visualizations:



* The figure shows that most deaths were in months May -> July for year 2020.

Chart

Description automatically generated

* The figure shows that most deaths were in April for year 2020.

Chart, histogram

Description automatically generated

* The figure shows that most deaths caused by covid-19 were in months May -> July for year 2020.

Chart, histogram

Description automatically generated

* The figure shows that most deaths caused by covid-19 were in months February & April for year 2021.

Chart

Description automatically generated

* - The figure shows all deaths either from covid or other medical reasons.

Chart, bar chart, treemap chart

Description automatically generated

-During the 2 years Men were more prominent to be infected by covid.

Chart

Description automatically generated

-People that were discharged from the hospitals were far more than people died in hospitals by nearly 7 times.

Graphical user interface, chart, pie chart

Description automatically generated

-79% of people that were diagnosed with covid didn’t need medical hospitalized attention & were assigned home treatment.

Chart, pie chart

Description automatically generated

-It was discovered that smoking wasn’t a huge factor on a person dying or not

Chart, pie chart

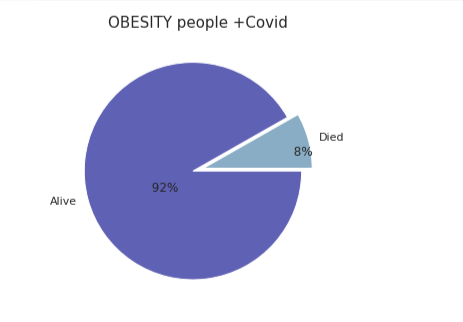
Description automatically generated

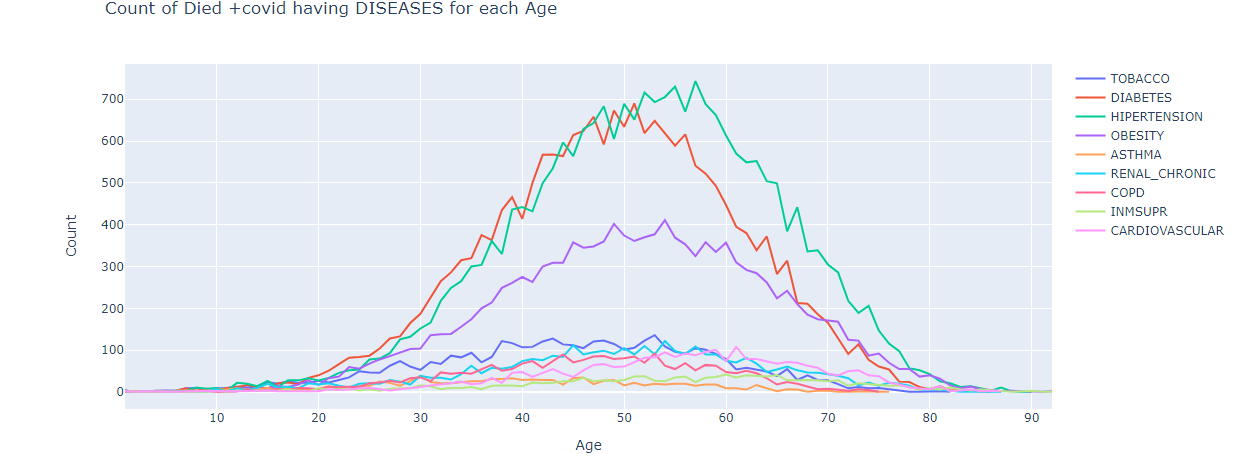
-To be more certain only 8% of smokers died

Chart, pie chart

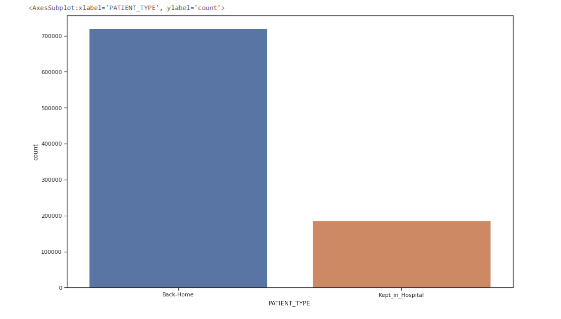
Description automatically generated

-16% of Positive Covid died people suffered from Obesity

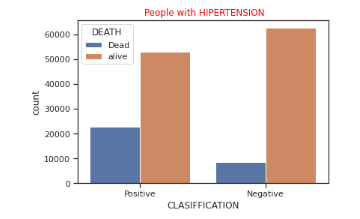
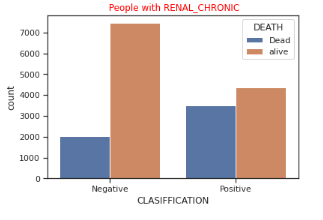
-8% of Positive Covid who suffer from Obesity died which means Obesity doesn’t affect death of a person 



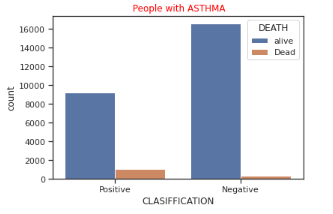
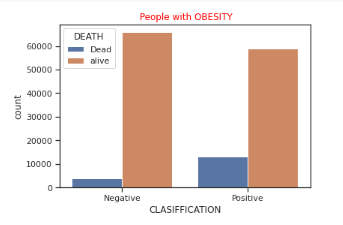
This one shows that old-Aged died people (with +covid)suffered with Hypertension and Diabetes.



-Most of People returned their home ,that is logically due to the diversity in cases and also because hospital can’t have space for all people .



If you were diagnosed with hypertension or Renal Chronic and infected with covid , Unfortunately You are in danger to die.



-In the other hand If you were diagnosed with Asthma or Obesity and infected with covid , fortunately You have a great chance to live.