* The first feature here is the number of covid cases ([New York Times (2020)](https://github.com/nytimes/covid-19-data) and [NYC Health Department (2020)](https://github.com/nychealth/coronavirus-data). Daily covid case count. The NYT dataset puts all of NYC into a single number, so I supplement it with NYC DoH data for individual boroughs.
* Other features are Diagnosed diabetes, obesity and hypertension ([Badawi et al](https://www.sciencedirect.com/science/article/pii/S1201971216311006?via%3Dihub) and [Al-Sabah, S., Al-Haddad, M., Al-Youha, S., Jamal, M., &amp; Almazeedi, S. (2020, December). Covid-19: Impact of obesity and diabetes on disease severity](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7645952/)). Using data from 12 studies, Badawi et al. [2] conducted a thorough review of the incidence of comorbidities in MERS and discovered that diabetes and hypertension were present in 50% of cases.

According to the article (2nd one), those who have underlying medical issues including obesity and diabetes are more likely to develop severe illness from COVID-19 and pass away from it. Due to the immune system's impairment and increased inflammation caused by these disorders, people are more susceptible to infections like COVID-19.

Specifically, the article notes that obesity and diabetes are associated with an increased risk of hospitalization, admission to the intensive care unit, and death from COVID-19. The article also suggests that these conditions may contribute to the higher rates of COVID-19 seen in certain populations, such as African Americans and Hispanics, who have higher rates of obesity and diabetes compared to other racial and ethnic groups. The paper specifically mentions the link between obesity and diabetes and a higher risk of hospitalization, admission to the intensive care unit, and COVID-19 mortality. In addition, the study makes the case that these ailments may be to blame for the higher COVID-19 rates observed in some populations, such as Black Americans and Hispanics, who also have greater rates of obesity and diabetes than those of other racial and ethnic backgrounds.

* Next feature is % Adults 65 and Older ([Zimmermann, P., & Curtis, N](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8740029/). and [The isolated effect of age on the risk of COVID-19 severe outcomes](https://gh.bmj.com/content/6/12/e006434)). The association between COVID-19 and age is examined in the paper "COVID-19 and Age: The Effect of Older Age on Community Vulnerability." Age is cited in the article as a major risk factor for COVID-19-related severe illness and mortality.

According to the reports, older folks (defined as those aged 65 years and above) are more likely than younger people to experience hospitalization and mortality from COVID-19. The chance of developing a serious illness or passing away from COVID-19 rises with aging, with people 85 years of age and older having the highest risk.

According to the papers, immunosenescence—an immune system alteration brought on by aging—may be a factor in older persons' higher risk of developing serious illness and passing away from COVID-19. Age-related immunosenescence can reduce the immune system's ability to fight off infections, rendering older people more susceptible to serious illness and COVID-19-related consequences.Ultimately, the articles emphasize how crucial it is to safeguard senior citizens against COVID-19 through immunization, social seclusion, and other public health measures.

* Another feature is heart diseases (cardiovascular). Angiotensin-converting enzyme 2 (ACE2) is known to enhance COVID-19 infection, and individuals with diabetes, hypertension, and cardiovascular conditions have higher ACE2 expression levels.

[Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? The Lancet Respiratory Medicine.](https://www.sciencedirect.com/science/article/pii/S2213260020301168?via%3Dihub)

[Jia, X., Yin, C., Lu, S., Chen, Y., Liu, Q., Bai, J., &amp; Lu, Y. (n.d.). Two things about covid-19 might need attention.](https://www.preprints.org/manuscript/202002.0315/v1)

[Carlos del Rio, M. D. (2020, April 14). Covid-19-new insights on a rapidly changing epidemic. JAMA.](https://jamanetwork.com/journals/jama/fullarticle/2762510)

* COPD is another feature ([Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019 - United States, February 12-March 28, 2020. MMWR. Morbidity and mortality weekly report.](https://pubmed.ncbi.nlm.nih.gov/32240123/))

Here the study investigates the medical features of COVID-19 patients who died. The study discovered that COVID-19 patients who passed away frequently had respiratory symptoms and complications, including breathing issues.

The majority of individuals with fatal outcomes, according to the report, had respiratory issues. Shortness of breath, dyspnea (difficulty breathing), and hypoxia were among these respiratory issues (low oxygen levels). According to the paper, respiratory failure was a common factor in these individuals' deaths.

According to the paper, COVID-19 patients who have pre-existing respiratory disorders such as interstitial lung disease, asthma, or chronic obstructive pulmonary disease (COPD) may be more susceptible to fatal respiratory consequences.

* Last feature is smokers ( [Smoking and covid-19 outcomes: An observational and mendelian ... - thorax.](https://thorax.bmj.com/content/77/1/65) And [British Heart Foundation. (n.d.). Does smoking increase or reduce your risk from coronavirus? BHF.](https://www.bhf.org.uk/informationsupport/heart-matters-magazine/news/behind-the-headlines/coronavirus-and-smoking))

The paper investigates the connection between COVID-19 and smoking. According to a meta-analysis of numerous research presented in the article, smokers may be more likely to experience severe COVID-19 results.

According to the report, smokers with COVID-19 have a larger chance of dying than non-smokers, are more likely to have severe symptoms, and need mechanical breathing. According to the article, smoking can impair the immune system, harm the respiratory system, and increase a person's risk of developing severe COVID-19 results.

The article also discusses how smoking could contribute to COVID-19 transmission. Smokers may touch their faces and mouths more frequently, which might increase the chance of viral transmission. Smoking may also lead to more mucus and coughing, which might result in more virus particles being released into the air.

The article's overall thesis is that smoking may raise the likelihood of catastrophic COVID-19 outcomes, and it underlines how crucial quitting smoking is to lower that risk. The report also underlines the need for more studies to clarify the connection between smoking and COVID-19 and to create efficient therapies to lessen smoking's negative effects on COVID-19 results.

ML PREDICTION RESULTS: (THE FINAL TABLE I HAVE GOT AFTER ML TECHNIQUES -UNSUPERVISED)

A black screen with white text

Description automatically generated with low confidence



THE RESULTS OF REGRESSION:

A screenshot of a computer

Description automatically generated with medium confidence

These are evaluation metrics for a regression model. Here's what they mean:

MAE (Mean Absolute Error) - This is the average absolute difference between the predicted and actual values. In other words, it gives you an idea of how far off your predictions are on average. A lower MAE is better, and a MAE of 3.098 means that on average, the model's predictions are off by about 3.1 units.

MSE (Mean Squared Error) - This is the average of the squared differences between the predicted and actual values. Squaring the differences gives more weight to larger errors. A lower MSE is better, and a MSE of 17.862 means that the model's predictions are off by about 17.9 units (on average, squared).

R2 (R-Squared) - This is a measure of how well the model fits the data. It represents the proportion of variance in the dependent variable that is explained by the independent variables. An R2 of 0.93 means that about 93% of the variance in the dependent variable is explained by the independent variables in the model. A higher R2 is better, and a perfect fit would have an R2 of 1.0.

Overall, these metrics suggest that the regression model has a good fit to the data and is making accurate predictions, with a relatively small average error (MAE) and a high proportion of variance explained (R2)

FURTHER PLANS:

1. I plan to better illustrate the results on map and make a simple web-app, so that all users can investigate it.
2. Plan to further include other features and train the models on them. (Economic and Geographic factors) and illustrate them on web-app(map)

MY CONTRIBUTION:

* Based on my investigations, there are not any models that uses regression and other ML techniques to predict the vulnerability index for each county of USA. In all models the vulnerability index was estimated by manual math and statistic calculations, no machine learning techniques were used to predict.
* Other users will be able to use my model to predict the vulnerability index for other countries. So, the only thing they will need is to prepare their own dataset and train it on my model.
* People can be better prepared for further covid transmission, since my model is adaptive, and they can apply updated datasets for investigation which parts of the world has more risks of Covid Cases.

1. Importing necessary libraries and the dataset.
2. Dropping unnecessary columns from the dataset.
3. Normalizing the data by scaling it between 0 and 1.
4. Performing KMeans clustering to group the counties based on their characteristics.
5. Calculating the distance of each county from the cluster centroids and using this distance to calculate the vulnerability index of each county.
6. Splitting the data into training and testing sets.
7. Building a linear regression model using the training data.
8. Evaluating the model using metrics like mean absolute error (MAE), mean squared error (MSE), and R-squared (R2).
9. Plotting the feature importance of the linear regression model.