

Analyzing the factors affecting the COVID-19 risk level in the US counties

Samira Ziyadidegan ^{1,*}, Moein Razavi ², Homa Pesarakli ³, Amir Hossein Javid ⁴

1 Zachry Department of Civil Engineering, Texas A&M University, College Station, TX, 77843, samirazyadg@tamu.edu

2 Department of Computer Science and Engineering, Texas A&M University, College Station, TX, 77843, moeinrazavi@tamu.edu

3 Department of Architecture, Texas A&M University, College Station, TX, 77843, Homa.pesarakli@tamu.edu

4 Department of Statistics, Oklahoma State University, Stillwater, OK, 74074, Amir.javid@okstate.edu

* Correspondence: samirazyadg@tamu.edu.

Abstract: The COVID-19 disease spreads swiftly, and nearly three months after the first positive case was confirmed in China, Coronavirus started to spread all over the United States. Some states and counties reported an extremely high number of positive cases and deaths, while some reported too few COVID-19 related cases and mortality. In this paper, the factors that could affect the transmission of COVID-19 and its risk-level in different counties have been determined and analyzed. Using Pearson Correlation, K-means clustering, and several classification models, the most critical ones were determined. Results showed that mean temperature, percent of people below poverty, percent of adults with obesity, air pressure, percentage of rural areas, and percent of uninsured people in each county were the most significant and effective attributes.

Keywords: multinomial logistic regression, K-means clustering, COVID-19, SARS-CoV-2, meteorological variables

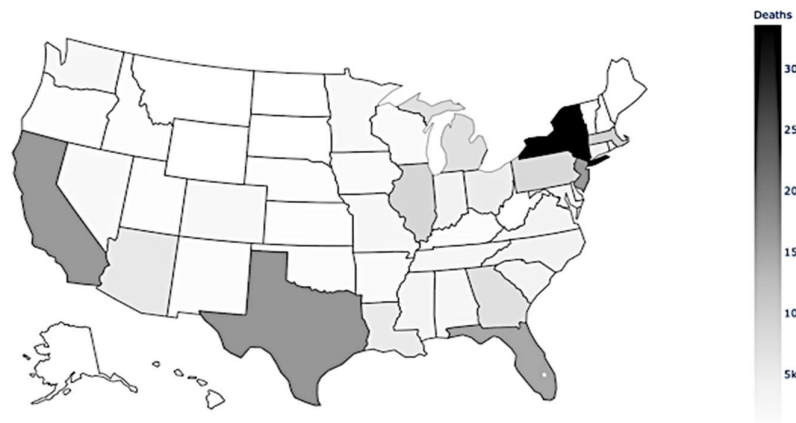
1. Introduction

COVID-19 disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with common symptoms of fever, dry cough, shortness of breath, and other signs of respiratory-related infections. World Health Organization (WHO) reported that 80% of patients experienced these symptoms mildly. However, older people (>60 years old) and the ones with co-morbid diseases are at a higher risk for severe symptoms and death [1,2]. Besides, younger patients with no underlying disease might also experience severe symptoms or even die [3–5].

The first positive case of COVID-19 in the United States was reported in the state of Washington on January 20, 2020. By March 17, 2020, Covid-19 has spread across all US states [6,7]. Figure 1 shows the aggregated COVID-19 positive case and death count maps for all US states until November 6, 2020. Reports showed that on November 6, 2020, the top five number of positive COVID-19 cases belonged to the states of California, Texas, Florida, New York, and Illinois, while the top 5 death cases belonged to New York, Texas, California, New Jersey, and Florida. [8,9].



(a)



(b)

Figure 1 COVID-19 positive case and death counts maps for all states of the US, (a) COVID-19 positive case counts map, (b) COVID-19 Death counts map

Previous studies showed that meteorological factors could have an essential role in the spread of a virus [10]. Temperature can affect the life cycle and proliferation of viruses; precipitation and rainfall can impact water-borne virus diffusion, while humidity and the wind speed can affect the dissemination of air-borne viruses [10–14].

Different studies showed contradictory effects regarding the meteorological attributes on the transmission or life duration of Coronavirus. Liu et al. [15] studied the correlation between confirmed case counts and weather-related attributes. The results showed that diurnal temperature, ambient temperature, and humidity are significant attributes and have negative correlations with the transmission of COVID-19. Moges Menebo [16] revealed that the maximum temperature and the normal temperature (average of high and low temperatures) have positive, but precipitation has negative impacts on COVID-19 case counts. Ma et al. [17] evaluated the effects of meteorological and air pollutant factors on COVID-19 death rates. They concluded that diurnal temperature range and absolute humidity are positively and negatively associated with COVID-19 mortality rate, respectively. He et al. [18] studied the effects of temperature and relative humidity on daily COVID-19 confirmed positive cases in different Asian cities. Their study revealed that in different cities, the impact of temperature on COVID-19 confirmed cases could differ. In Beijing, Shanghai, and Guangzhou, the correlation between temperature and the number of positive cases was negative, while in Japan, it was positive. The study conducted by Pahuja et al. [19] demonstrated a positive correlation between the wind speed and the COVID-19 case counts. In contrast, Coccia [20] showed that the high wind speed could reduce the number of COVID-19 cases.

In addition to the meteorological factors, there are other factors that impact the transmission and spread of COVID-19 positive case and death case counts. They include air pollutant factors such as the concentration of air pollutants (e.g., PM_{10} and SO_2) [17,20], demographic factors (e.g., elderly people index, population density, etc.) [20], adults smoking history [21,22], location-based factors (e.g., latitude) [20,23–25], and disease-related factors (e.g., cardiovascular or respiratory disease experience) [20,22,26,27].

Reviewing the literature indicated that an analysis that took into account a comprehensive set of factors affecting the rate of positive COVID-19 cases and its death rate was missing. Besides, as Figure 1 shows, states with a high number of COVID-19 cases (e.g., Florida, New York, and California) are located in areas that are different in terms of geographical, health, demographical, and meteorological characteristics. However, the majority of the previous studies analyzed the factors affecting COVID-19 transmission in the areas with similar characteristics (e.g., geographical, demographical, etc.). This might lead to identifying biased influential factors. The mentioned issue could be addressed by considering more extensive areas (e.g., all US counties) with different geographical, health, demographical, and meteorological characteristics.

In this study, a comprehensive set of attributes that have a high impact on the transmission of the COVID-19 and its risk level in all US counties have been determined. This research conducts analysis to find out why some areas have higher COVID-19 positive cases and death rates than other areas.

2. Materials and Methods

Materials

The data used in this study was taken from various online sources from March 2020 to November 2020, including COVID-19 positive cases and deaths, demographic, meteorological, health, and location-based data separated by each US county (Table 1). To evaluate the risk level of each county, we used the COVID-19 positive cases and death rates (positive cases and death counts divided by the population of the county) instead of their actual numbers. For meteorological data, we took the average of the data over all days of each month to have the average monthly data because the time-series nature of the meteorological data was not match with other databases used. The maximum and minimum temperatures for each county were driven from the maximum and minimum temperature across all days from March to November. Population density along with the population of elderly people (>65 years old) and young people (<65 years old) have also been added to the dataset. All the parameters used in this study are shown in Table 1.

TABLE 1 Parameters used in the study

Category	Parameters
COVID-19	positive case rate, death rate
location-based	lateral coordinate, longitudinal coordinate, county Area, percent of rural areas
meteorological	minimum temperature, maximum temperature, mean temperature, wind speed, air pressure, precipitation, dewpoint
Health	number of ICU beds, number of staffed beds, percent of smokers, percent of adults with obesity, percent of people uninsured, percent of adults with diabetes
Demographic	elderly population, young population, white population, non-white population, eightieth percentile income, twentieth percentile income, percent of people below poverty, total population, population density

Methodology

Figure 2 indicates the steps taken in this article to determine the key factors with higher effects on Coronavirus transmission and can mostly determine the risk level of each county. In the first step, we found and corrected the inaccurate records from the dataset (data cleaning). Then, we calculated the correlation coefficients between different parameters, identified highly correlated parameters, and among the ones with high correlation, we chose only one variable. We then applied clustering with the two key parameters of positive case rate and death rate as features. Counties were then labeled based on clustering results such that each label was considered as a class. In the next step, we applied different classification models on the complete dataset, and the best model was determined based on the accuracy rates. Finally, the significant parameters were selected for the analysis.

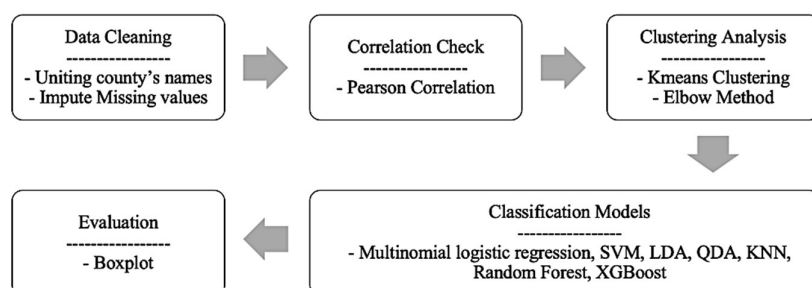


Figure 2 Methodology used in the current study

3. Results

Data Cleaning

After combining the data from various sources, it was revealed that many values were missing (Figure 3). One reason was the name mismatch between different counties in different datasets. Moreover, the number of ICU beds for some counties was missing. So, they have been imputed by zero. Missing values for the demographic data were replaced with data from the United States Census Bureau website [28]. If no data was available for a specific county, values were imputed using the average of the non-missing values of that parameter across the neighboring counties. For example, the missing precipitation, air pressure and dewpoint values for the “District of Columbia” county were replaced with the corresponding data for “Arlington” county in Virginia state. The total number of counties analyzed in the current study was 3131.

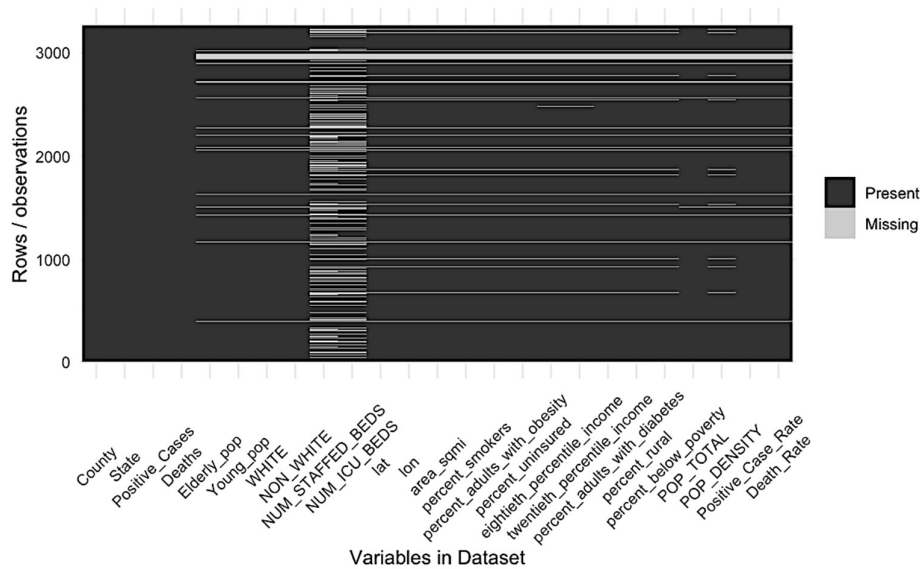


Figure 3 Missing values plot

Correlation check

After normalizing all independent parameters, Pearson Correlation was applied to the data, and the parameters with high linear correlation (more than 0.8) were removed. The remaining parameters for further analyses are shown in Table 2.

TABLE 2 Final parameters used in this study

Category	Parameters
COVID-19	positive case rate, death rate
location-based	longitudinal coordinate, county area, percent of rural areas
meteorological	mean temperature, Wind Speed, precipitation, air pressure
health	number of ICU beds, percent of smokers, percent of adults with obesity, percent of people uninsured, percent of adults with diabetes
demographic	eightieth percentile income, percent below poverty, population density

Clustering Analysis

To determine the COVID-19 risk level of each county, K-means clustering was done based on COVID-19 positive case rates and death rates. The cluster number for each county was used as its label (dependent variable) for the classification analysis.

To determine the optimal number of clusters that can define the risk level of each county, the Elbow method is used (Figure 4). The Elbow method is a visual method that can determine the optimal number of clusters considering the total within-cluster sum of squares of Euclidean distances (the cost). The optimized k value (k is the number of clusters) is such that adding another cluster (k+1) does not significantly decrease the cost. In the Elbow method plot, the optimal k value is located at the elbow of the curve. [29,30]

Figure 4 shows that k=3 could be the optimal number of clusters. As it can be seen in Figure 5 (clustering results), counties were clustered into three clusters of low-positive-case low-death, medium-positive-case medium-death, and high-positive-case high-death. The county located far apart from the other points belongs to New York County, which had a positive case rate of 15.5% and a death rate of 1.5%.

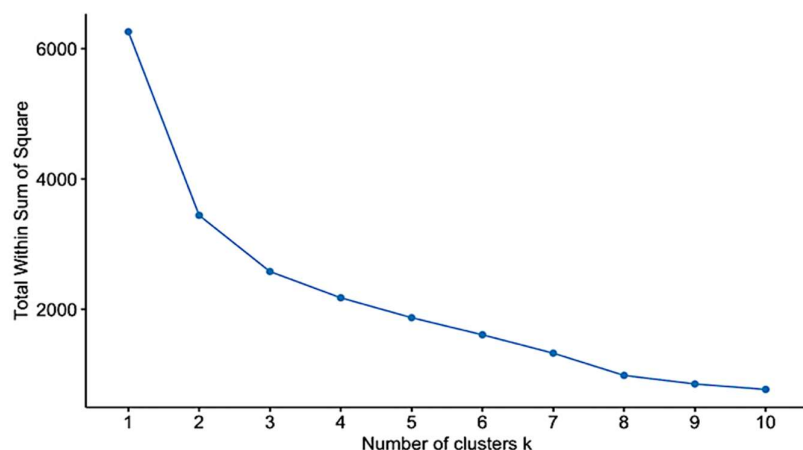


Figure 4 Elbow Method

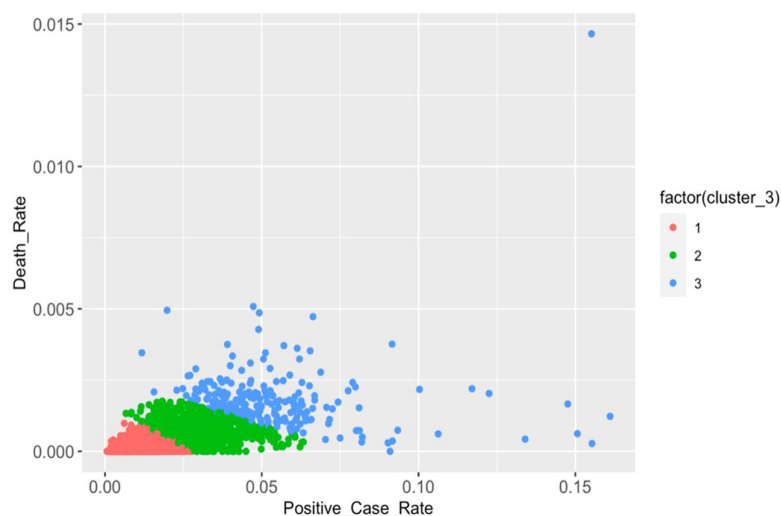


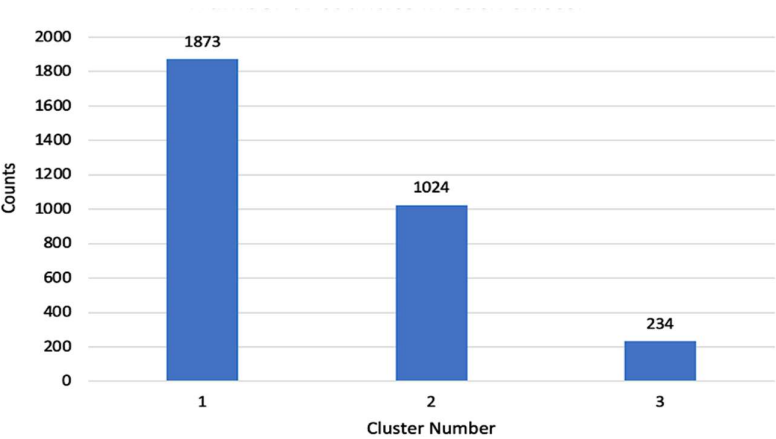
Figure 5 Clustering output plot

Figure 6 and Table 3 show some characteristics of each cluster. As results showed, cluster 1 seemed to be the low-risk cluster. It contained the highest number of counties with the lowest positive case rate and death rate on average. On the other hand, cluster 3 had the lowest number of counties but the highest positive case rate and death rate on average, which referred to high-risk counties. Cluster 2 is considered as the medium-risk cluster.

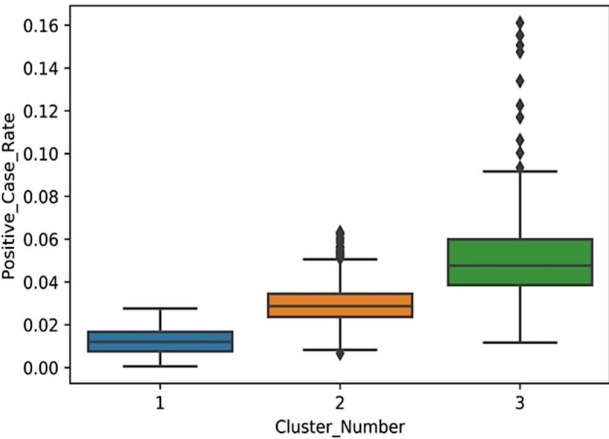
In the next section, the result of clustering analysis was used to find significant parameters that affect the COVID-19 risk level of each county by means of classification analysis.

TABLE 3 Cluster attributes

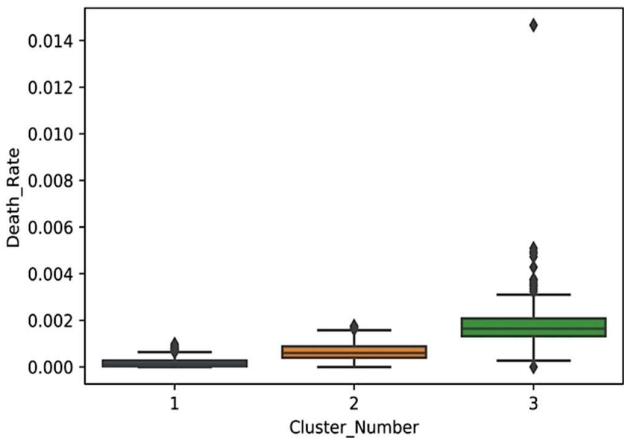
Cluster number	Counts	Mean Positive Case Rate	Mean Death Rate
1	1873	1.231e-2	1.838e-4
2	1024	2.966e-2	6.639e-4
3	234	5.257e-2	18.293e-4



(a)



(b)



(c)

Figure 6 Clustering Results: (a) Number of counties in each cluster; (b) Positive Case Rate; (c) Death Rate

Classification Analysis

To determine the significant factors, different classification models have been applied to the data. Based on the accuracy values attained, the best model was used to select the factors with the highest importance in the transmission of the COVID-19.

To extract significant factors, several classification models shown in Table 4 were applied to the dataset. The data was divided into train and test sets (80% and 20%, respectively). Table 4 shows the train accuracy and the test accuracy for each model. Among the classification models, Random Forest was overfitted on the data. XGBoost obtained a very low performance. SVM models, KNN, QDA, LDA, and MLR performed almost similarly on the test data. SVM and KNN coefficients are not interpretable. QDA performed poorly compared to LDA and MLR since the QDA fitted the more flexible classifier than necessary. LDA and MLR performed very similarly in this study. LDA makes more restrictive Gaussian assumption. However, the normality assumption was not satisfied on the current dataset. Therefore, to balance between the accuracy and complexity of the model, we chose the multinomial logistic regression (MLR) model as the best describing model.

Evaluation

Further analyses were finally performed to answer the research question: What makes a county be high-risk while another county has low risk against COVID-19 disease.

To extract the significant factors, a backward selection approach was used based on the p-values of the coefficients in the multinomial logistic regression model. Considering 95% of confidence interval, final significant factors were determined. The final factors were mean temperature, percent of people below poverty (people with income lower than the threshold determined by the United States Census Bureau [31]), percent of adults with obesity (BMI > 30 [32]), air pressure, percent of rural areas, and percent of people uninsured. It should be noted that other models with acceptable accuracy values (e.g., LDA, KNN, and SVM) suggested almost the same significant factors as MLR.

TABLE 4 Train and test errors of the classification models used

Method	Train Accuracy	Test Accuracy
Multinomial Logistic Regression	71.69%	68.58%
LDA	71.01%	68.74%
QDA	68.29%	66.35%
KNN	75.20%	70.18%
SVM Linear	70.41%	68.74%
SVM Radial	77.84%	70.02%
SVM Polynomial	77.20%	69.06%
Random Forest	100.00%	71.61%
XGBoost	12.22%	26.16%

Significant Variables

In this step, the MLR model is applied only to the significant variables. The cross-validation accuracy and test accuracy were 70.45% and 66.83%, respectively. Table 5 shows the coefficients of significant variables driven from the MLR model. The second columns are the coefficients (log odds units) of variables of cluster=2 compared to cluster=1, and the coefficients (log odds units) of the third columns are for cluster=3 compared to cluster=1. As an illustration, one unit increase in mean temperature increases the odds by $\exp(0.5662)=1.7616$ of being a county in cluster=2 vs. cluster=1.

TABLE 5 Log odds unit of significant variables

Variable	cluster=2	cluster=3
Intercept	-0.64	-3.0006
mean temperature	0.5662	0.7778
percent below poverty	0.2861	0.9019
percent adults with obesity	0.2517	0.2671
air pressure	-0.3616	-0.7727
percent of rural areas	-0.4293	-0.2039
percent of people uninsured	0.2892	0.3852

Table 5 demonstrated that increasing the average temperature, percent of people below poverty, percent of adults with obesity, and percent of people uninsured would enhance a county's chance to be in a cluster with a higher level of COVID-19 risk. On the other hand, increasing the air pressure and percent of rural areas would decrease that chance.

Previous studies concluded that there is a positive relationship between temperature and COVID-19 cases [16,17,33]. The results of this study were in line with their conclusion. Higher average temperature belonged to clusters 2 and 3, which had higher COVID-19 positive case rate and mortality rate.

Results revealed that the percentage of people below poverty in a county was positively associated with belonging a county to a cluster with a higher level of COVID-19 risk, as shown in Table 5. Low-income people might have limited access to health products such as masks and sanitizers, which have high effects on virus transmission and vitality [34,35]. They are more likely to work out of their homes due to unstable jobs and income or less likely to have reliable and valid information about the COVID-19 disease [36,37].

Moreover, low-income people are less likely to have health insurance (the higher percentage of people uninsured). So, due to high medical expenses, they prefer not to go to clinics/hospitals or use medications, which might increase the COVID-19 death rate and, as a result, increase the association of a county to a higher risk cluster.

As the center for disease control and prevention [38] claimed, obesity would increase the risk of COVID-19 death. Besides, obesity would affect the immune system adversely. These are in line with the findings of this study. As the obesity percentage in a county increases, the chance of being in the high-risk cluster would increase as well.

Results demonstrated that air pressure was another significant factor which negatively associated with belonging a county to a higher risk cluster. Air pressure determines the precipitation, wind, and weather condition. High air pressure is associated with mild wind and calm weather [39]. So, as the study of Coccia [20] claimed, it decreased the transmission of COVID-19. So, it lowered the association of a county to higher risk clusters. Research conducted by Takagi et al. [40] also demonstrated that high air pressure would reduce the COVID-19 prevalence.

Considering the percent of rural areas in each county, Table 5 indicated that counties with higher percentages of rural areas almost belonged to cluster 1 (low-risk cluster). In rural areas, people can keep physical distance from others which is one of the most important factors to prevent the transmission of COVID-19. However, in rural areas, sometimes, accessing advanced medical centers is lower.

4. Discussion

In this paper, several demographic, meteorological, geographic, and health factors have been analyzed to determine the critical parameters which highly affected the transmission of COVID-19 in the US counties and to determine each county's risk level. To do so, aggregated COVID-19 positive cases and deaths for each county were driven, and then their rate with respect to population was calculated. Next, K-means clustering was applied to these two rates. Based on the Elbow method, three clusters were chosen titled low-risk (low-positive-case low-death), medium-risk (medium-positive-case medium-death), and high-risk (high-positive-case high-death) clusters. The labels obtained from clustering were then considered as the nominal dependent variable for classification.

In the next step, Pearson Correlation was applied to all factors to remove those highly correlated ones. Sixteen variables were kept for further analyses.

Using cluster labels, several classification models with 10-fold cross-validation were applied to the data. Based on the train and test accuracy, multinomial logistic regression (MLR) was finally chosen as the best model to determine the significant variables. Finally, six factors of mean temperature, percent of people below poverty, percent of adults with obesity, air pressure, percent of rural areas, and percent of people uninsured were chosen. Increasing the mean temperature, percent of people below poverty, percent of adults with obesity, and percent of people uninsured would increase the belonging of a county to the higher risk cluster. On the contrary, increasing the air pressure would increase the association of a county to the lower risk cluster. Counties with higher percentages of rural areas seemed to belong to the low-risk cluster mostly.

Author Contributions:

Conceptualization, Ziyadidegan. S., Razavi. M., and Pesarakli. H.; methodology, Ziyadidegan. S., Razavi. M., and Pesarakli. H., Javid. A.; software, Ziyadidegan. S., Razavi. M., and Pesarakli. H.; formal analysis, Ziyadidegan. S., Razavi. M., and Pesarakli. H., Javid. A.; data curation, Ziyadidegan. S., Razavi. M., and Pesarakli. H.; writing—original draft preparation, Ziyadidegan. S.; writing—review and editing, Razavi. M., and Pesarakli. H., Javid. A. "All authors have read and agreed to the published version of the manuscript."

Data Availability Statement:

Publicly available datasets were analyzed in this study. This data can be found here:

Data combined and used in this paper by authors:

<https://github.com/SamiraZiyadg/COVID-19>

Covid Data

<https://github.com/nytimes/covid-19-data>

Health Data

https://opendata.dc.gov/datasets/1044bb19da8d4dbfb6a96eb1b4ebf629_0?selectedAttribute=ADULT_ICU_BEDS

Demographic, Meteorological, and Location-based data

<https://www.kaggle.com/johnjdavisiv/us-counties-covid19-weather-sociohealth-data>

Acknowledgments:

The first three authors would thank Dr. Darren Homrighausen from Texas A&M University for his valuable feedback on the preliminary draft of this study when he was the instructor of STAT 656: Applied Analytics in 2020.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Velavan, T.P.; Meyer, C.G. The COVID-19 epidemic. *Trop. Med. Int. Heal.* **2020**, *25*, 278–280, doi:10.1111/tmi.13383.
2. World Health Organization *Coronavirus Disease - 2019 (COVID-19) Situation report-41*; 2020;
3. Jahromi, R.; Avazpour, A.; Jahromi, M.; Alavi, J. Covid-19 With Positive Bronchoalveolar Lavage Fluid But Negative Nasopharyngeal and Oropharyngeal Swabs: a Case Report and Insights. *Indian J. Case Reports* **2020**, *6*, 380–382, doi:10.32677/ijcr.2020.v06.i07.010.
4. Yousefzadegan, S.; Rezaei, N. Case report: Death due to COVID-19 in three brothers. *Am. J. Trop. Med. Hyg.* **2020**, *102*, 1203–1204, doi:10.4269/ajtmh.20-0240.
5. The Washington Post Hundreds of young Americans have now been killed by the coronavirus, data shows Available online: <https://www.washingtonpost.com/health/2020/04/08/young-people-coronavirus-deaths/> (accessed on Mar 8, 2021).
6. Centers for Disease Control and Prevention, C. First Travel-related Case of 2019 Novel Coronavirus Detected in United States.
7. Saad B. Omer; Malani, P.; Rio, C. del The COVID-19 Pandemic in the US A Clinical Update. *Clin. Infect. Dis.* **2020**, *71*, 778–785, doi:10.1093/cid/ciaa310.
8. John Hupkins University & Medicine Covid-19 Map Available online: <https://coronavirus.jhu.edu/us-map> (accessed on Nov 13, 2020).
9. 1point3acres Global Covid-19 Trackers and Interactive Charts Available online: <https://coronavirus.1point3acres.com/en> (accessed on Nov 13, 2020).
10. Wang, G.; Minnis, R.B.; Belant, J.L.; Wax, C.L. Dry weather induces outbreaks of human West Nile virus infections. *BMC Infect. Dis.* **2010**, *10*, doi:10.1186/1471-2334-10-38.
11. Wu, X.; Lu, Y.; Zhou, S.; Chen, L.; Xu, B. Impact of climate change on human infectious diseases: Empirical evidence and human adaptation. *Environ. Int.* **2016**, *86*, 14–23, doi:10.1016/j.envint.2015.09.007.
12. Chen, P.S.; Tsai, F.T.; Lin, C.K.; Yang, C.Y.; Chan, C.C.; Young, C.Y.; Lee, C.H. Ambient influenza and avian influenza virus during dust storm days and background days. *Environ. Health Perspect.* **2010**, *118*, 1211–1216, doi:10.1289/ehp.0901782.
13. Pica, N.; Bouvier, N.M. Environmental factors affecting the transmission of respiratory viruses. *Curr. Opin. Virol.* **2012**, *2*, 90–95, doi:10.1016/j.coviro.2011.12.003.
14. Lowen, A.C.; Mubareka, S.; Steel, J.; Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* **2007**, *3*, 1470–1476, doi:10.1371/journal.ppat.0030151.
15. Liu, J.; Zhou, J.; Yao, J.; Zhang, X.; Li, L.; Xu, X.; He, X.; Wang, B.; Fu, S.; Niu, T.; et al. Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. *Sci. Total Environ.* **2020**, *726*, 138513, doi:10.1016/j.scitotenv.2020.138513.
16. Moges Menebo, M. Temperature and precipitation associate with Covid-19 new daily cases: A correlation study between weather and Covid-19 pandemic in Oslo, Norway. *Sci. Total Environ.* **2020**, *737*, 139659, doi:10.1016/j.scitotenv.2020.139659.
17. Ma, Y.; Zhao, Y.; Liu, J.; He, X.; Wang, B.; Fu, S.; Yan, J.; Niu, J.; Zhou, J.; Luo, B. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci. Total Environ.* **2020**, *724*, 138226, doi:10.1016/j.scitotenv.2020.138226.
18. He, Z.; Chin, Y.; Yu, S.; Huang, J.; Zhang, C.J.P.; Zhu, K.; Azarakhsh, N.; Sheng, J.; He, Y.; Jayavanth, P.; et al. The Influence of Average Temperature and Relative Humidity on New Cases of COVID-19: Time-Series Analysis. *JMIR Public Heal. Surveill.* **2021**, *7*, e20495, doi:10.2196/20495.
19. Pahuja, S.; Madan, M.; Mittal, S.; Pandey, R.M.; Nilima; Madan, K.; Mohan, A.; Hadda, V.; Tiwari, P.; Guleria, R. Weather Parameters and COVID-19: A Correlational Analysis. *J. Occup. Environ. Med.* **2021**, *63*, 69–73, doi:10.1097/JOM.0000000000002082.

20. Coccia, M. How do low wind speeds and high levels of air pollution support the spread of COVID-19? *Atmos. Pollut. Res.* **2020**, doi:10.1016/j.apr.2020.10.002.
21. Patanavanich, R.; Glantz, S.A. Smoking is associated with COVID-19 progression: A meta-analysis. *Nicotine Tob. Res.* **2020**, *22*, 1653–1656, doi:10.1093/ntr/ntaa082.
22. Zheng, Z.; Peng, F.; Xu, B.; Zhao, J.; Liu, H.; Peng, J.; Li, Q.; Jiang, C.; Zhou, Y.; Liu, S.; et al. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *J. Infect.* **2020**, *81*, e16–e25, doi:10.1016/j.jinf.2020.04.021.
23. Franch-Pardo, I.; Napoletano, B.M.; Rosete-Verges, F.; Billa, L. Spatial analysis and GIS in the study of COVID-19. A review. *Sci. Total Environ.* **2020**, *739*, 140033, doi:10.1016/j.scitotenv.2020.140033.
24. Jüni, P.; Rothenbühler, M.; Bobos, P.; Thorpe, K.E.; Da Costa, B.R.; Fisman, D.N.; Slutsky, A.S.; Gesink, D. Impact of climate and public health interventions on the COVID-19 pandemic: A prospective cohort study. *Cmaj* **2020**, *192*, E566–E573, doi:10.1503/cmaj.200920.
25. Walrand, S. Autumn COVID-19 surge dates in Europe correlated to latitudes, not to temperature-humidity, pointing to vitamin D as contributing factor. *Sci. Rep.* **2021**, *11*, 1–9, doi:10.1038/s41598-021-81419-w.
26. Jordan, R.E.; Adab, P.; Cheng, K.K. Covid-19: Risk factors for severe disease and death. *BMJ* **2020**, *368*, 1–2, doi:10.1136/bmj.m1198.
27. Bansal, M. Cardiovascular disease and COVID-19. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2020**, *14*, 247–250, doi:10.1016/j.dsx.2020.03.013.
28. United States Census Bureau Available online: <https://www.census.gov> (accessed on Nov 11, 2020).
29. Kodinariya, T.M.; Makwana, P.R. Review on determining number of Cluster in K-Means Clustering. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* **2013**, *1*, 2321–7782.
30. Purnima, B.; Arvind, K. EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *Int. J. Comput. Appl.* **2014**, *105*, 17–24.
31. United States Census Bureau How the Census Bureau Measures Poverty Available online: <https://www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html#:~:text=If a family's total income,it is considered in poverty.&text=The official poverty definition uses,Medicaid%2C and food stamps> (accessed on Mar 29, 2021).
32. Ogden, C.L.; Fryar, C.D.; Martin, C.B.; Freedman, D.S.; Carroll, M.D.; Gu, Q.; Hales, C.M. Trends in obesity prevalence by race and hispanic origin - 1999-2000 to 2017-2018. *JAMA - J. Am. Med. Assoc.* **2020**, *324*, 1208–1210, doi:10.1001/jama.2020.14590.
33. Islam, N.; Bukhari, Q.; Jameel, Y.; Shabnam, S.; Erzurumluoglu, A.M.; Siddique, M.A.; Massaro, J.M.; D'Agostino, R.B. COVID-19 and climatic factors: A global analysis. *Environ. Res.* **2021**, *193*, 110355, doi:10.1016/j.envres.2020.110355.
34. Jahromi, R.; Mogharab, V.; Jahromi, H.; Avazpour, A. Synergistic effects of anionic surfactants on coronavirus (SARS-CoV-2) virucidal efficiency of sanitizing fluids to fight COVID-19. *Food Chem. Toxicol.* **2020**, *145*, 111702, doi:10.1016/j.fct.2020.111702.
35. Ramesh, N.; Siddaiah, A.; Joseph, B. Tackling Corona Virus Disease 2019 (COVID 19) in Workplaces. *Indian J. Occup. Environ. Med.* **2020**, *24*, 16–18, doi:10.4103/ijoom.IJOEM.
36. Little, C.; Alsen, M.; Barlow, J.; Naymagon, L.; Tremblay, D.; Genden, E.; Trosman, S.; Iavicoli, L.; van Gerwen, M. The Impact of Socioeconomic Status on the Clinical Outcomes of COVID-19; a Retrospective Cohort Study. *J. Community Health* **2021**, doi:10.1007/s10900-020-00944-3.
37. Patel, J.A.; Nielsen, F.B.H.; Badiani, A.A.; Assi, S.; Unadkat, V.A.; Patel, B.; Ravindrane, R.; Wardle, H. Poverty, inequality and COVID-19: the forgotten vulnerable. **2020**.
38. Centers for Disease Control and Prevention, C. Overweight & Obesity Available online: <https://www.cdc.gov/obesity/data/obesity-and-covid-19.html> (accessed on Mar 4, 2021).

-
39. ThoughtCo Air Pressure and How It Affects the Weather.
 40. Takagi, H.; Kuno, T.; Yokoyama, Y.; Ueyama, H.; Matsushiro, T.; Hari, Y.; Ando, T. Higher Temperature, Pressure, and Ultraviolet Are Associated with Less COVID-19 Prevalence – Meta-Regression of Japanese Prefectural Data. *medRxiv* **2020**, doi:10.1101/2020.05.09.20096321.