COVID-19 Deaths in the United States

Lauren Hoshiko

Problem Statement

The Coronavirus disease (COVID-19) has tremendously impacted public health and economic and social well-being [1][2]. State to state, populations had varying pandemic experiences whether it be more dramatic loss of human life or greater difficulty meeting healthcare needs. There are many demographic, macroeconomic, travelling and general health indicators that could play a role. Can we group the states together that were similarly challenged by the pandemic to create a model for high, medium and low level impact?

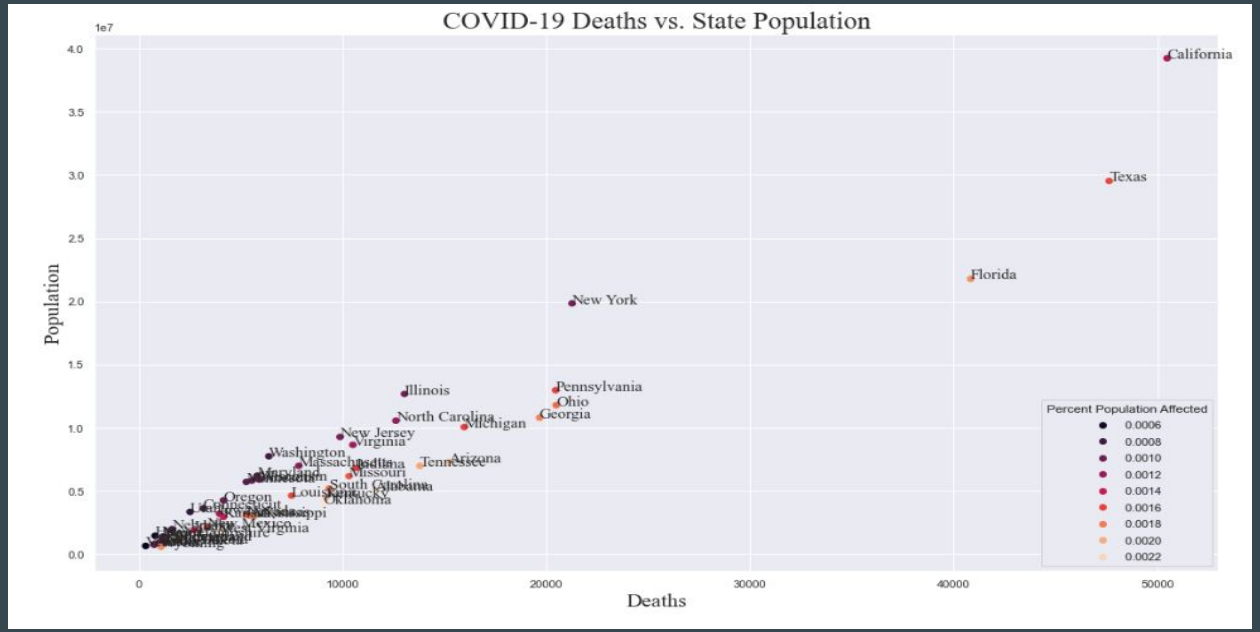
Data Wrangling

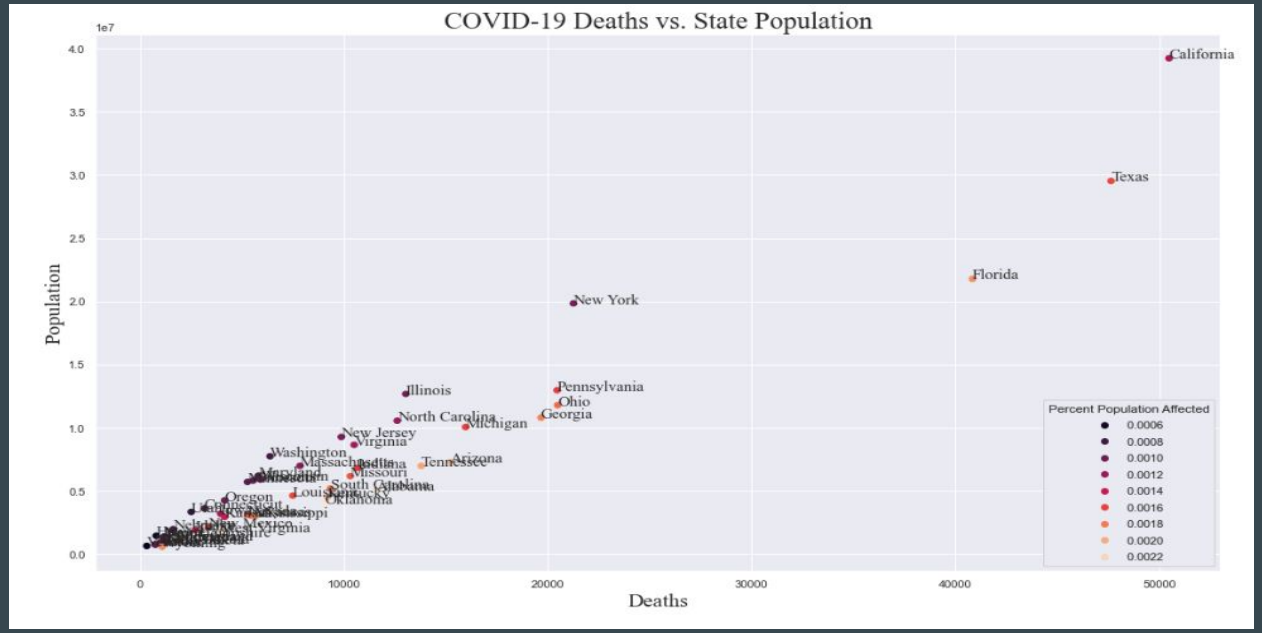
	Date	State	Cases	Deaths	Percent African American	Percent American Indian And Alaska Native	Percent Asian	Percent Native Hawaiian And Other Pacific Islander	Percent Two Or More Races	Percent Hispanic Or Latino	Percent White Not Hispanic Or Latino	Median Household Income	Percent Persons Below Poverty Level
0	2022-10-20	Alabama	1531305.0	20533.0	26.8	0.7	1.5	0.1	1.8	4.6	65.3	50536	15.5
1	2022-10-20	Alaska	306062.0	1350.0	3.7	15.6	6.5	1.4	7.5	7.3	60.2	77640	10.1
2	2022-10-20	Arizona	2283073.0	31514.0	5.2	5.3	3.7	0.3	2.9	31.7	54.1	58945	13.5
3	2022-10-20	Arkansas	957218.0	12408.0	15.7	1.0	1.7	0.4	2.2	7.8	72.0	47597	16.2
4	2022-10-20	California	11338044.0	96697.0	6.5	1.6	15.5	0.5	4.0	39.4	36.5	75235	11.8

	date	state	fips	cases	deaths	deaths_permonth	Year	Month	Month_year
586	2020-03-13	Alabama	1	6	0	NaN	2020	3	2020-03
637	2020-03-14	Alabama	1	12	0	0.0	2020	3	2020-03
689	2020-03-15	Alabama	1	23	0	0.0	2020	3	2020-03
742	2020-03-16	Alabama	1	29	0	0.0	2020	3	2020-03
795	2020-03-17	Alabama	1	39	0	0.0	2020	3	2020-03
849	2020-03-18	Alabama	1	51	0	0.0	2020	3	2020-03

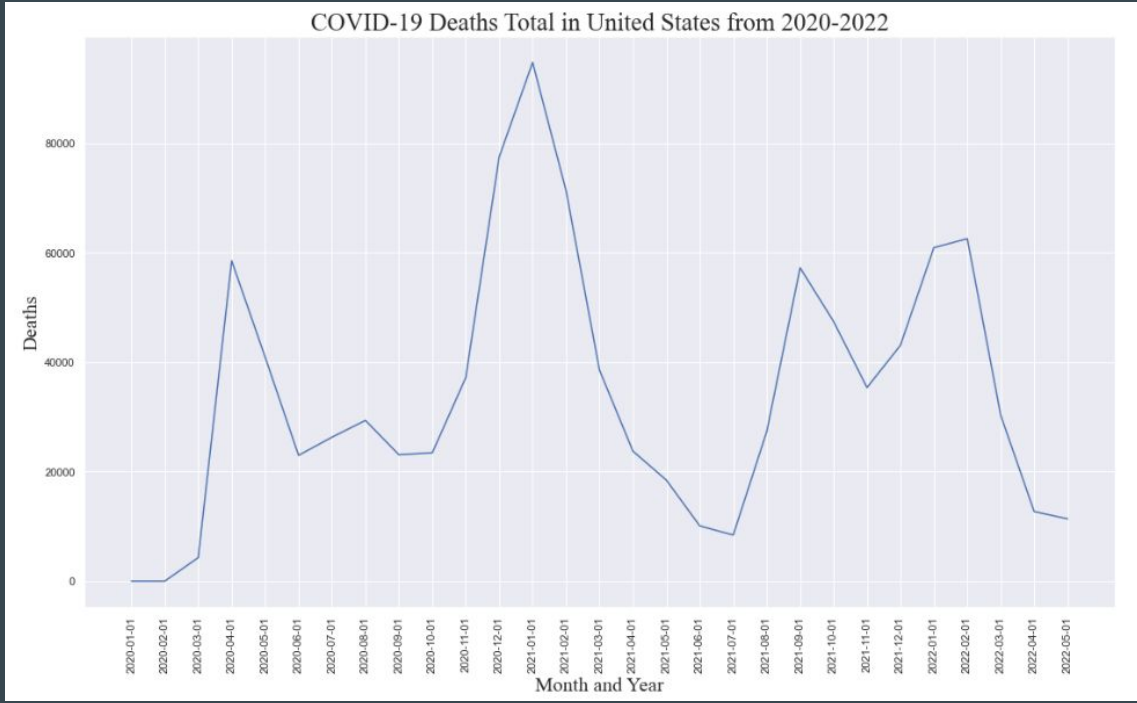
	Date	State	Fips	Cases	Deaths
53262	2022-10-20	Alabama	1	1531305	20533
53263	2022-10-20	Alaska	2	306062	1350
53265	2022-10-20	Arizona	4	2283073	31514
53266	2022-10-20	Arkansas	5	957218	12408
53267	2022-10-20	California	6	11338044	96697
53268	2022-10-20	Colorado	8	1674460	13588
53269	2022-10-20	Connecticut	9	908015	11423
53270	2022-10-20	Delaware	10	311850	3136
53272	2022-10-20	Florida	12	7145833	82065
53273	2022-10-20	Georgia	13	2832383	38986

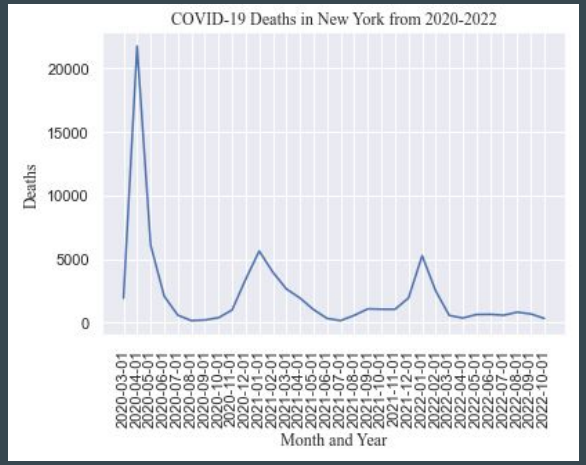
Exploratory Data Analysis

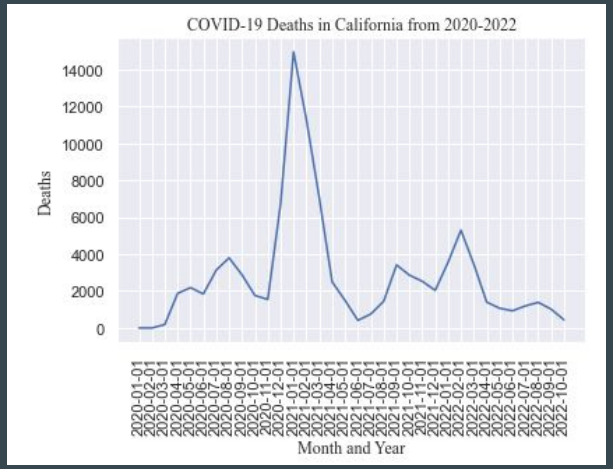


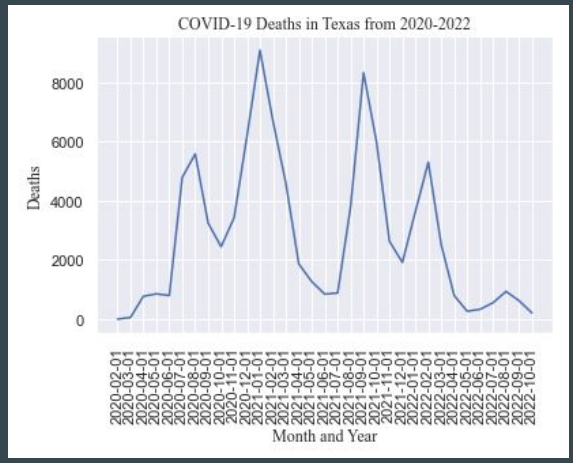


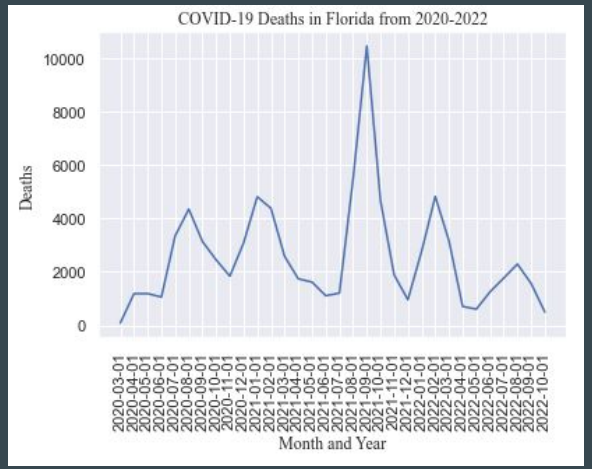
Deaths	1.000000
Percent African American	0.284419
Percent American Indian And Alaska Native	-0.233846
Percent Asian	0.152693
Percent Native Hawaiian And Other Pacific Islander	-0.141103
Percent Two Or More Races	-0.138449
Percent Hispanic Or Latino	0.509375
Percent White Not Hispanic Or Latino	-0.456957
Median Household Income	0.017543
Percent Persons Below Poverty Level	0.160476

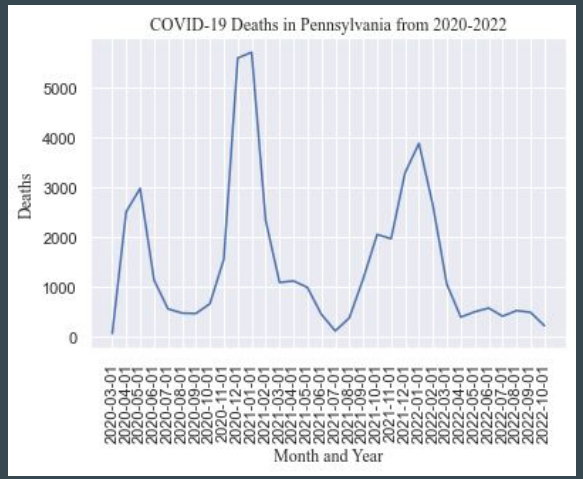

COVID-19 Deaths Total in United States from 2020-2022
Deaths
0
20000
40000
60000
80000
2020-01-01
2020-02-01
2020-03-01
2020-04-01
2020-05-01
2020-06-01
2020-07-01
2020-08-01
2020-09-01
2020-10-01
2020-11-01
2020-12-01
2021-01-01
2021-02-01
2021-03-01
2021-04-01
2021-05-01
2021-06-01
2021-07-01
2021-08-01
2021-09-01
2021-10-01
2021-11-01
2021-12-01
2022-01-01
2022-02-01
2022-03-01
2022-04-01
2022-05-01
Month and Year


COVID-19 Deaths in New York from 2020-2022
Deaths
Month and Year

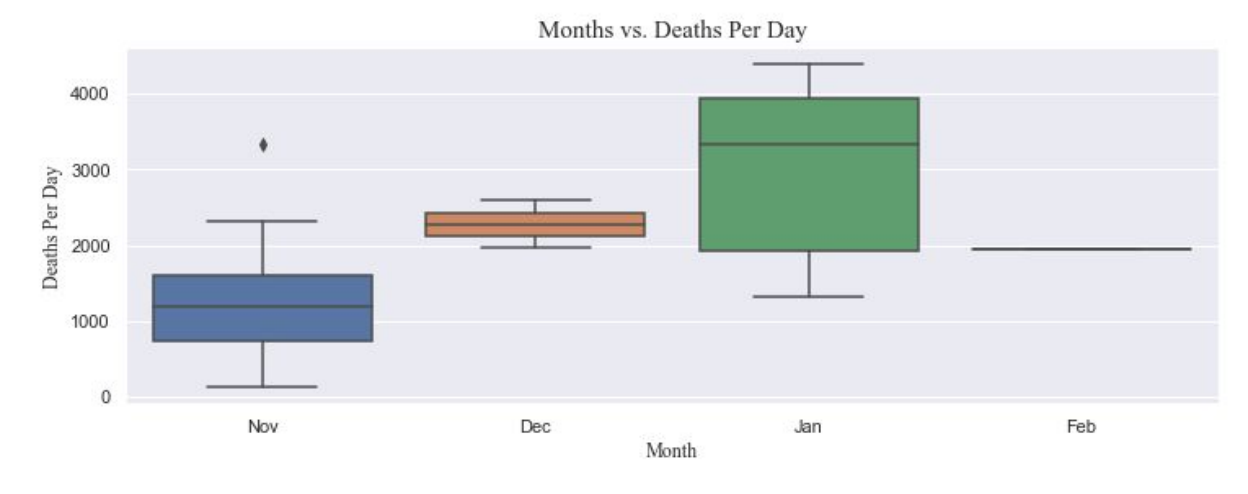

COVID-19 Deaths in California from 2020-2022
Deaths
Month and Year


COVID-19 Deaths in Texas from 2020-2022
Deaths
Month and Year


COVID-19 Deaths in Florida from 2020-2022
Deaths
Month and Year


COVID-19 Deaths in Pennsylvania from 2020-2022
Deaths
Month and Year

Null Hypothesis: The holiday season did not affect the number of COVID-19 deaths in the United States.

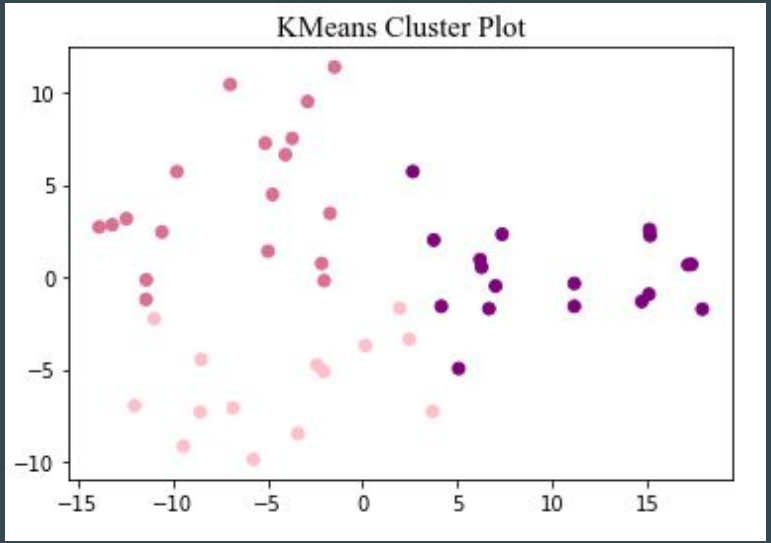


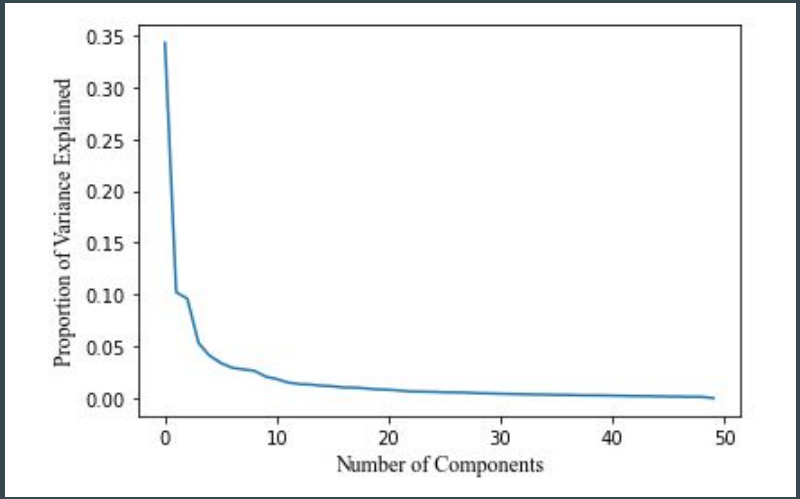
Pre-processing

In order to be able to group states based on their shared macroeconomic, demographic and health features, a dataframe with all this information needed to be aggregated on the state level.

Measure Name State	Access to Care - AnnualRank	Adverse Childhood ExperiencesRank	Air PollutionRank	Air and Water Quality - AnnualRank	All Determinants - AnnualRank	ArthritisRank	AsthmaRank	Avoided Care Due to CostRank	Behavioral Health - AnnualRank	BehaviorsRank	...
Alabama	45.0	30.0	33.0	29.0	45.0	49.0	16.0	45.0	33.0	47.0	...
Alaska	20.0	44.0	16.0	43.0	31.0	8.0	13.0	31.0	25.0	26.0	...
Arizona	42.0	35.0	46.0	41.0	32.0	12.0	26.0	40.0	27.0	22.0	...
Arkansas	43.0	48.0	23.0	19.0	48.0	43.0	18.0	44.0	42.0	46.0	...
California	18.0	9.0	50.0	46.0	28.0	1.0	20.0	13.0	31.0	15.0	...

Modeling


KMeans Cluster Plot
10
5
0
−5
−10
−15
−10
−5
0
5
10
15


Proportion of Variance Explained
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
Number of Components

Summary

Overall, the features used in the final dataframe covered a comprehensive list of demographic, macroeconomic and health indicators. Future analyses would benefit from additional data on each state's compliance to various mandates as well as the effectiveness of different responses to COVID-19 such as vaccine/booster availability.