

## Data Understanding

The data consists of covid-19 information of each state of the USA. The data is collected till March 7, 2021. All dates and times are in US eastern time (ET). Each state has its own set of caveats. The data consists of 41 attributes and 20780 entries in total. We have information like the date, state, number of deaths, number of hospitalized patients, negative test, on ventilator, positive test etc about each covid case in a particular case in the USA. The dataset is collection of information about COVID-19 pandemic for the US. The dataset is used to study and analyze the spread of Covid-19 throughout various states of the United States.

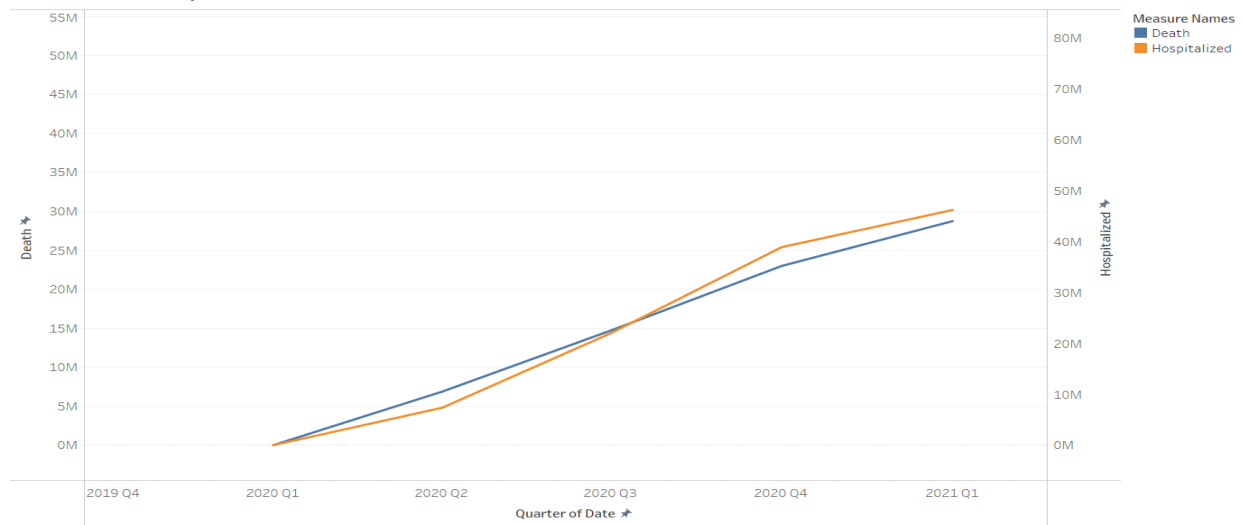
## Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	date	state	death	deathConf	deathIncr	deathProb	hospitaliz	hospitaliz	hospitaliz	hospitaliz	inlCuCum	inlCuCurre	negative	negativeI	negativeT	negativeT
2	#####	AK	305		0		1293	1293	33	0				0		
3	#####	AL	10148	7963	-1	2185	45976	45976	494	0	2676		1931711	2087		
4	#####	AR	5319	4308	22	1011	14926	14926	335	11		141	2480716	3267		
5	#####	AS	0		0					0			2140	0		
6	#####	AZ	16328	14403	5	1925	57907	57907	963	44		273	3073010	13678		
7	#####	CA	54124		258				4291	0		1159		0		
8	#####	CO	5989	5251	3	735	23904	23904	326	18			2199458	0	369995	
9	#####	CT	7704	6327	0	1377			428	0				0		
10	#####	DC	1030		0				150	0		38		0		
11	#####	DE	1473	1337	9	136			104	0		13	545070	917		
12	#####	FL	32266		66		82237	82237	3307	92			9339038	19166	864153	816231
13	#####	GA	17906	15598	1	2308	56797	56797	2008	35	9263			0		
14	#####	GU	133		0				2	0		1	112887	0		
15	#####	HI	445	445	1		2226	2226	27	0		5		0		
16	#####	IA	5558		6				167	0		35	1044418	1299		93984
17	#####	ID	1879	1652	3	227	7184	7184	150	5	1245	33	505964	535		
18	#####	IL	23014	20763	12	2251			1141	0		255		0		
19	#####	IN	12737	12310	11	427	43217	43217	656	64	7600	104	2483156	4643		
20	#####	KS	4812		0		9387	9387	235	0	2554	50	974686	0		
21	#####	KY	4819	4369	13	450	19457	19457	558	6	4034	156		0		
22	#####	LA	9748	9033	32	715			532	0			5261679	20578		
23	#####	MA	16417	16085	43	337	10713	10713	665	0		174	4404702	8078		

We will perform multiple visualizations to analyze this data and derive some useful insights regarding covid in the United States.

## Visualization 1

Time Series Analysis

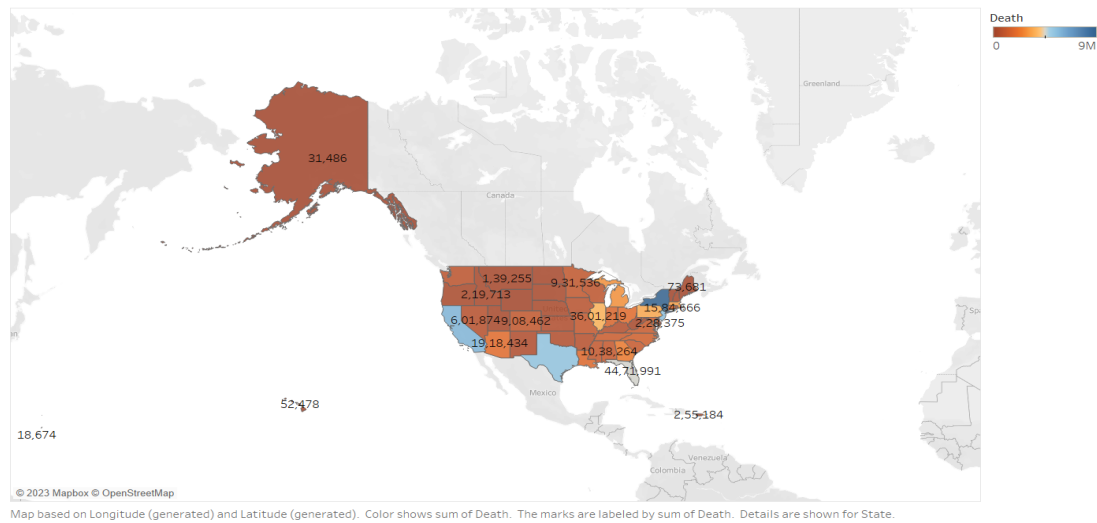


The trends of Death and Hospitalized for Date Quarter. Color shows details about Death and Hospitalized.

- The above graph is filtered on the attribute year.
- We have divided the time period for the year 2020 to 2021 into 4 quarters to perform time series analysis on the data.
- Here we can see a trend for Number of deaths and number of hospitalized cases.
- Let's first see the number of deaths, the number of deaths start from 20739 deaths in the start of 2020 and increases linearly throughout and we can see the number of death increases to 28743995 till 2021 in the same linear pattern.
- Now let's see the graph for number of hospitalized cases for the same period of time. The number of hospitalized cases starts from 32221 cases at the start of 2020. We can see a peak increase in the number of hospitalized cases from quarter 1 to 2 then there's a linear increase in the cases from quarter 2 to 3 for a large amount of time.
- On comparing the number of cases hospitalized with deaths we can observe that overall till the year 2021 both the numbers have shown an increase and the effects of covid-19 on the life of the people in USA have increased throughout.
- In quarter three the both the numbers are very close to each other but before and after the meeting point it is observed that the number of hospitalized are 1.5 times (approx) more than the number of deaths.
- There's a direct relation between number of cases hospitalized and deaths. More the number more the deaths.
- The doctors were able to save some patients but still the deaths had an adverse effect.

## Visualization 2:

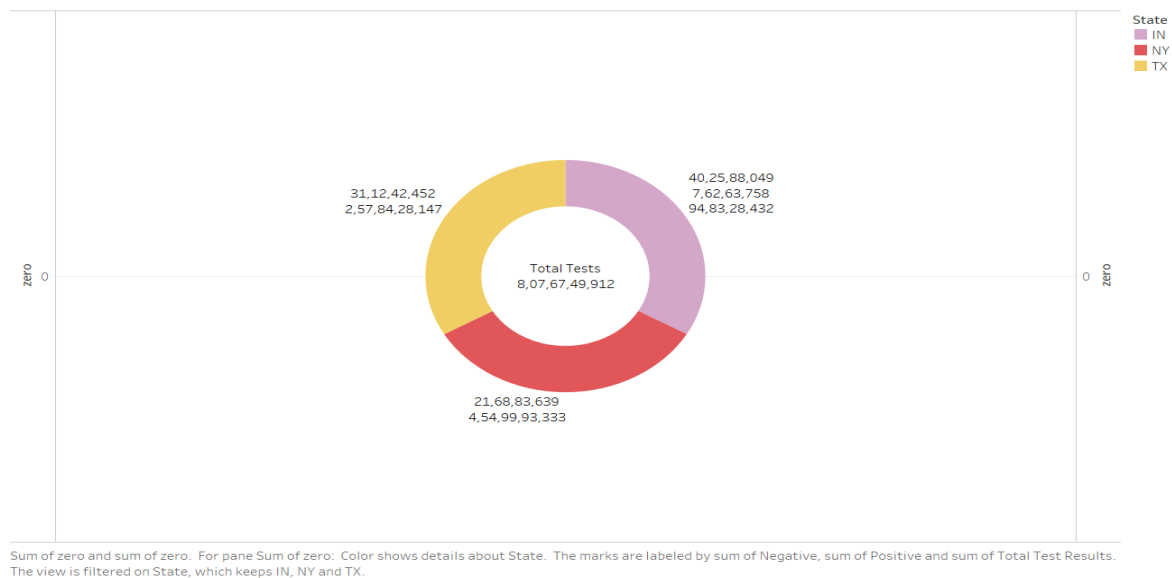
Sheet 9



- In the above visualization we observe geographical distribution of the covid spread across the USA.
- The graph is color coordinated according to the number of deaths in a particular state.
- We can see that New York has maximum deaths during covid followed by Texas and CA. It can be said that big states with more population have been more affected due to covid and have more number of deaths and number of hospitalized cases.
- The trend continues with states like IL, CT, MI and AZ with comparatively less death cases but still more than states like MT, WY, OR and NE which have much less death cases.
- The maximum hospitalized cases were in FL, NJ and AZ. But more deaths were in NY, TX and CA. The people in FL, NJ and AZ received a better medical health or the patients there were able to recover with the medical health provided.
- It is observed that state NV had 601874 death cases and zero hospitalized case saying that the people in this state didn't receive any medical help and died due to covid.

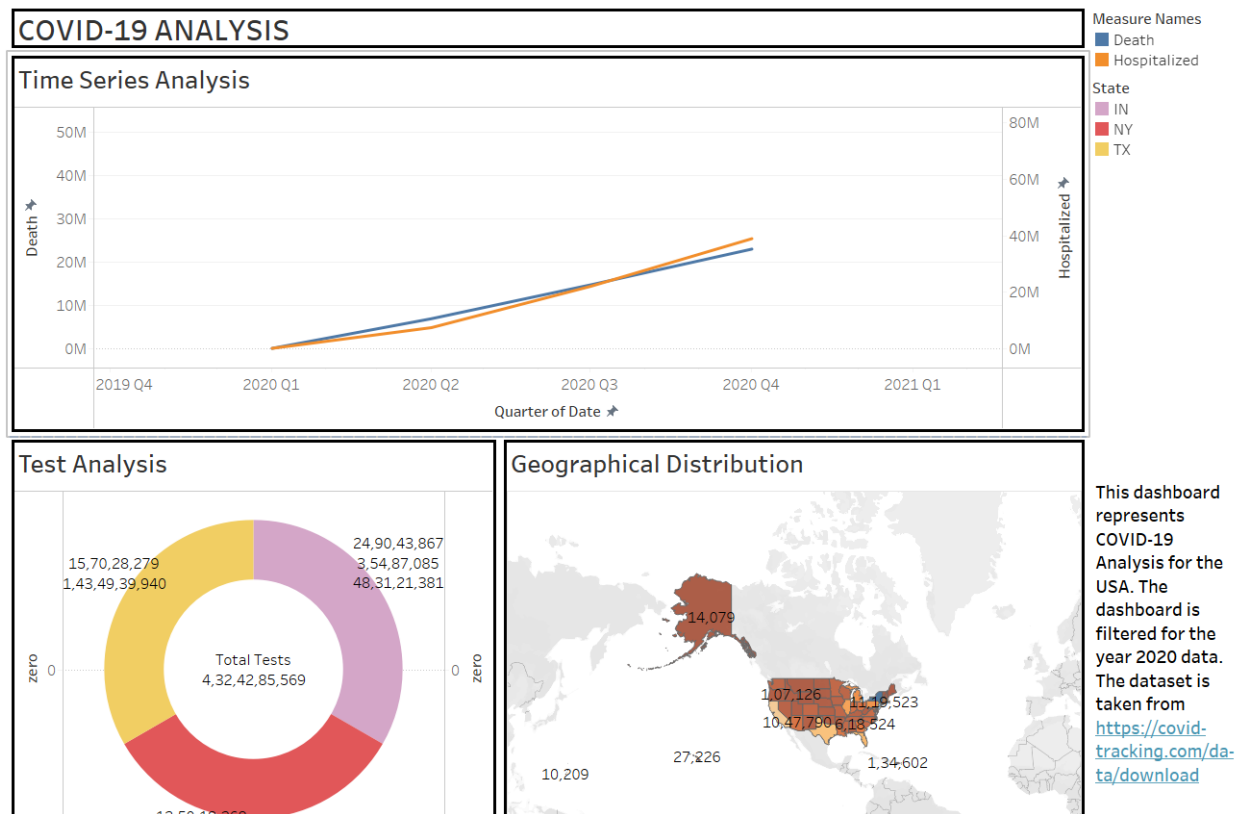
### Visualization 3:

Sheet 7



- In the above visualization we have analyzed three parameters: Total tests, positive reports and negative reports.
- Here I have concentrated on three states of USA: Indiana, Texas and new york. We can see the total tests that were carried out in the entire USA were 8076749912 tests. From which 3981749912 tests were from these three states.
- Lets analyse for each state now, In new york there were zero negative cases reported. The total tests carried out in were 4549993333 and out of which 216883639 were positive cases reported. The rest tests if they are not negative there are possibility that the data is missing for them or there is a loss of data or the results for those tests are unclear. There can be multiple possibilities.
- For the state Texas the total number of tests carried out were 2578428147, out of which 311242452. Here again there were no negative cases reported in the data. There's a possibility of loss or missing data just as above.
- For the state Indiana the total number of tests carried out were 948328432 out of which 76263758 were positive and 402588049 were negative. Here there is comparatively less missing data.
- The maximum tests that were carried out were in New York followed by Texas and then Indiana. We can infer that the state with more population or the busy state like new york had way more tests than a small state with less population like Indiana.
- In total it was observed that we have more negative cases reported than positive. It was surprising to observe that state like new york and texas had no positive case reported in the data.

## Dashboard



For the dashboard I have filtered the values for all the sheets for the year 2020. And based on that the dashboard is created.

- From this dashboard we can observe that there is an increase in the number of deaths and number of hospitalized people for the year 2020. The graph linearly increases throughout the year.
- Number of deaths and hospitalized cases are directly proportional to each other.
- In the year 2020 the number of deaths in the states like CA and TX was comparatively less as observed from the color saturation but the places saw an adverse effect in the following year 2021 as well leading to a large number of deaths.
- The total tests in the year 2020 are half of the total tests for the entire period.
- In total we can conclude that covid 19 effect gradually increased from 2020 to 2021. It was observed that all the parameters only increased from first year to second.