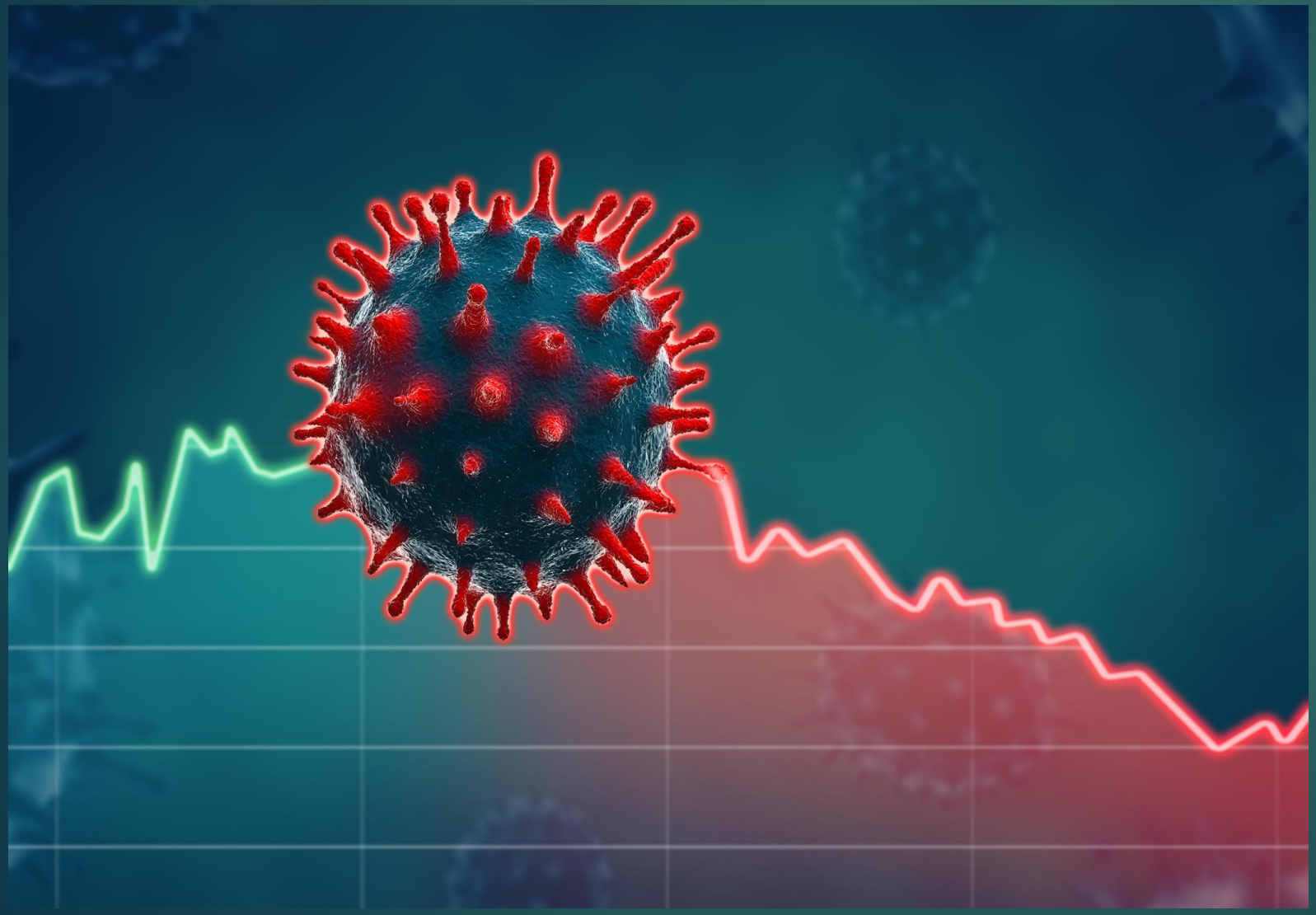
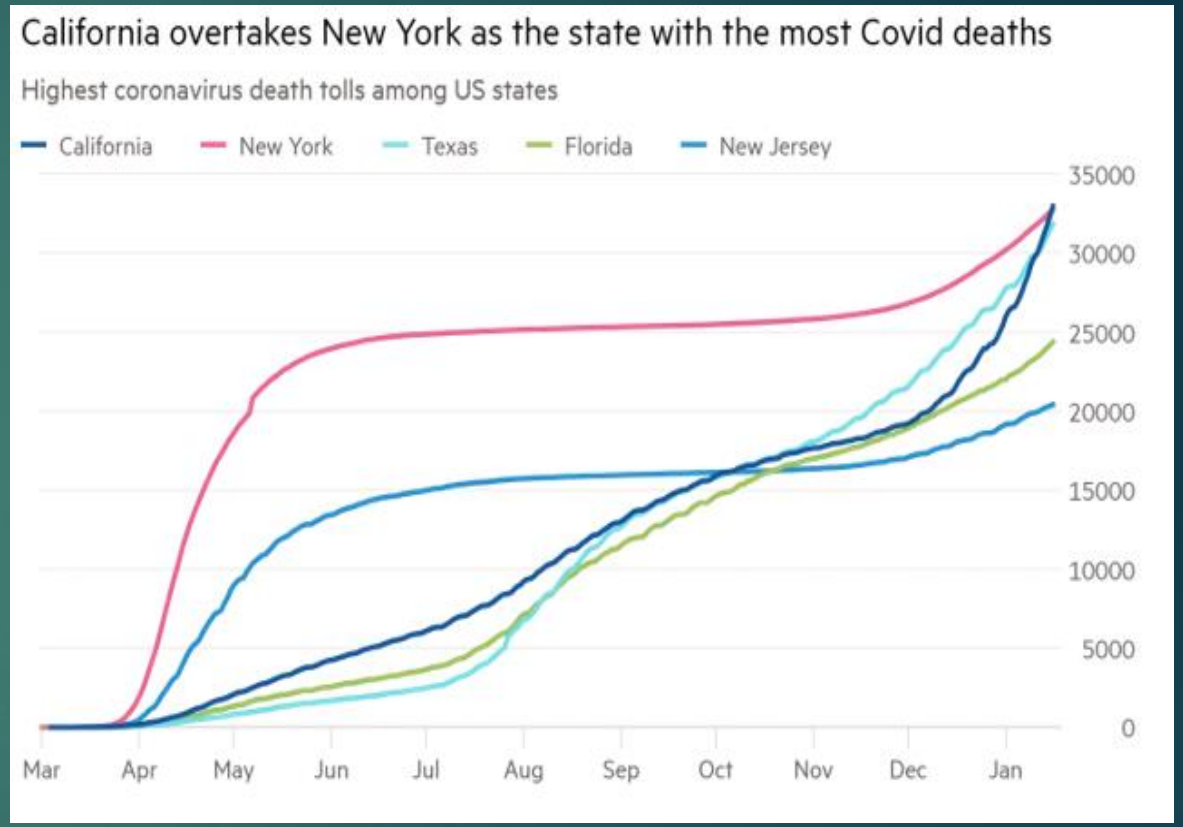
COVID-19 effects on USA States



SPARSH CHANDRA

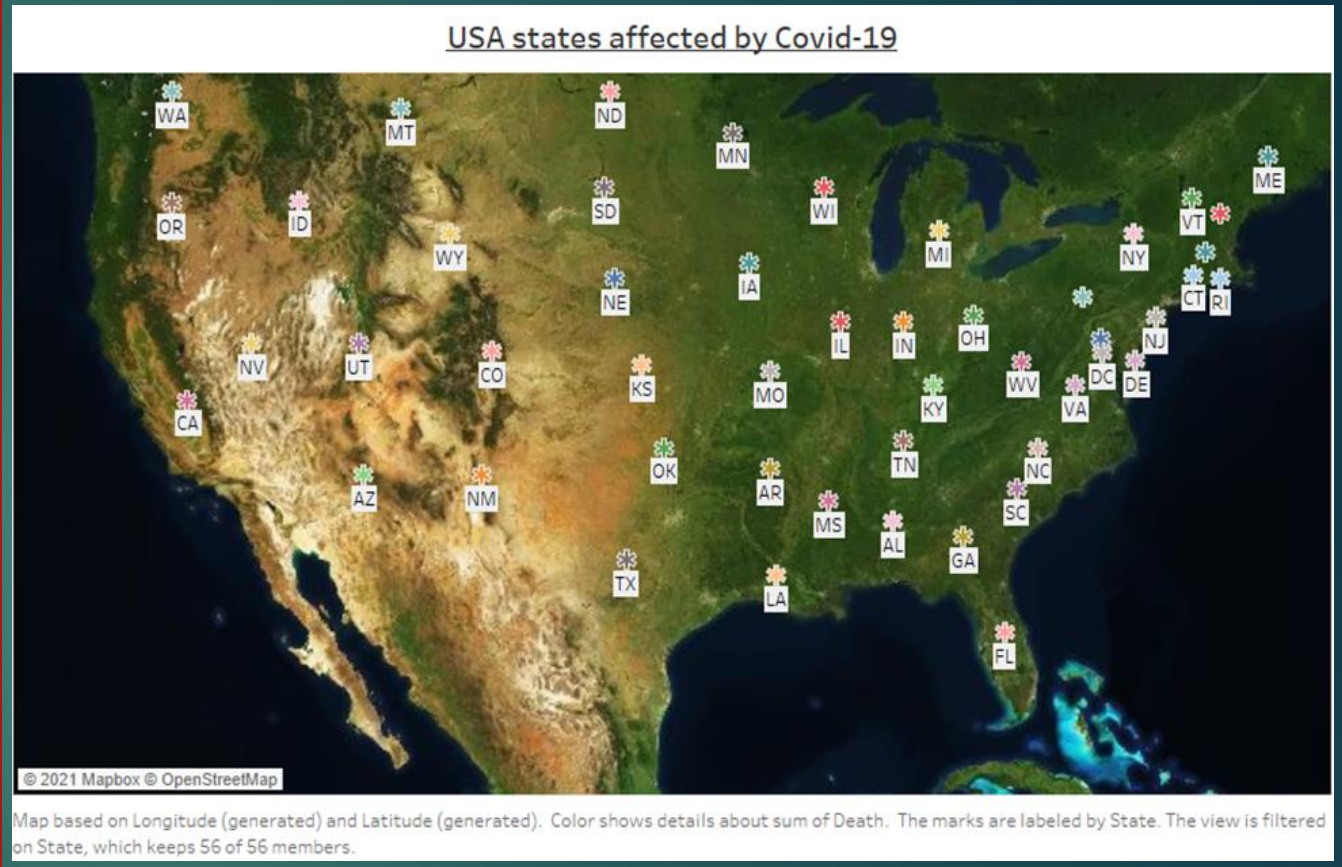
Problem Statement

- One event dominated in 2020: a deadly and previously unknown virus wreaked havoc across the globe, killing more than 1.5 million people, infecting many more and causing economic devastation.
- The pandemic set the course of science to an extraordinary degree
- Across the United States, COVID-19 infections and deaths were surging
- California, New York, Florida, Texas, New Jersey were amongst the worst hit states
- Based on the chosen dataset, we are trying to compare COVID-19 statistics for California state versus New York state



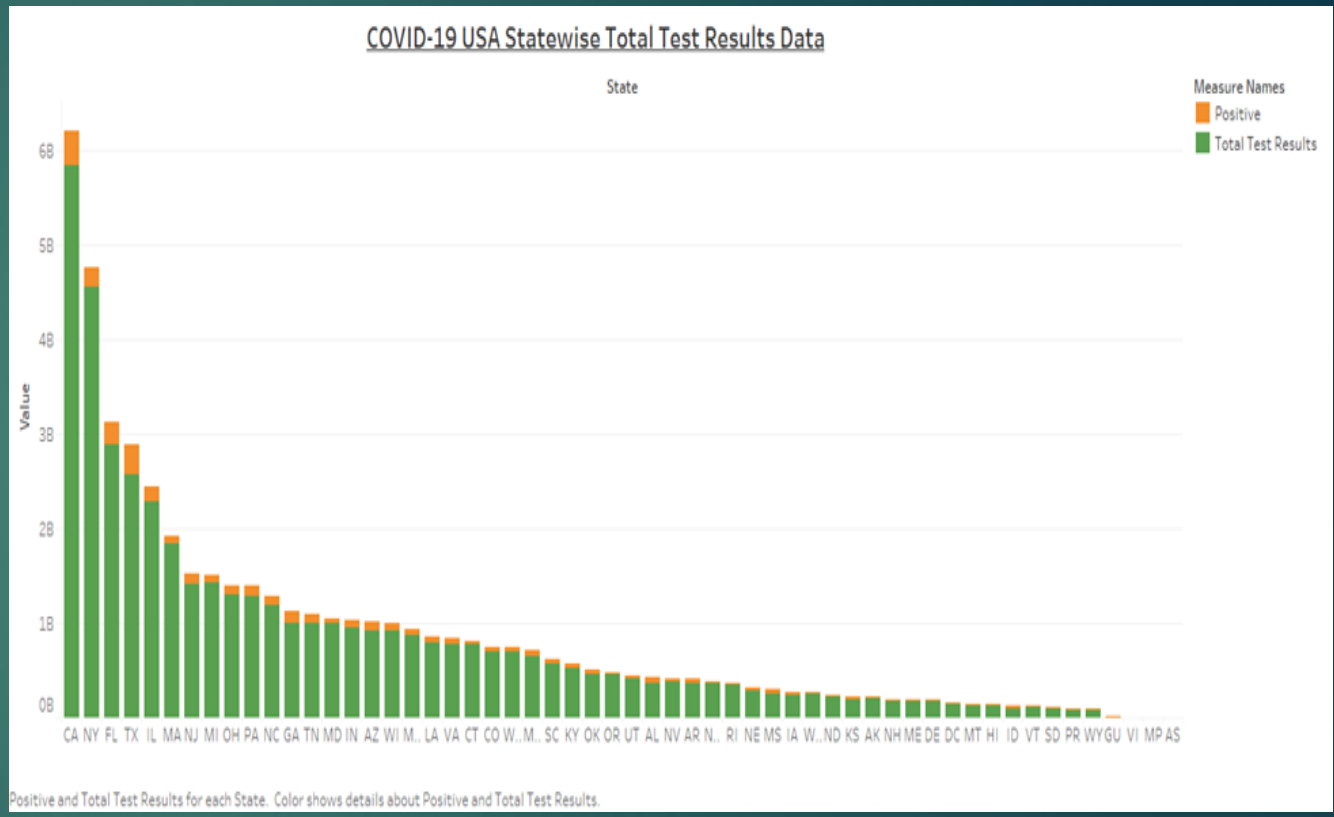
Dataset Overview

- The dataset used in this analysis is from January 2020 to March 2021, and it depicts how COVID-19 affected all 50 states in the US throughout the epidemic
  - **Dataset volume : 3 MB**
  - **Number of rows : 20,780**
  - **Number of columns : 41**
- Kaggle was used to obtain this dataset
- The fundamental reason for selecting this dataset is that COVID-19 outbreak had a huge impact on the world for past two years and it continues to do so
- From vaccines and treatments to lockdowns and virtual reality, the coronavirus epidemic altered the year in research



Dataset Overview

- As the graph suggests, maximum number of tests done, and positive cases were found in California and New York
- Comparing these two states we are evaluating the relationship between following parameters:
  - **Number of total tests results**
  - **Number of positive cases**
  - **Number of hospitalized currently**
  - **Number of deaths**



Descriptive Statistics : Key Variables

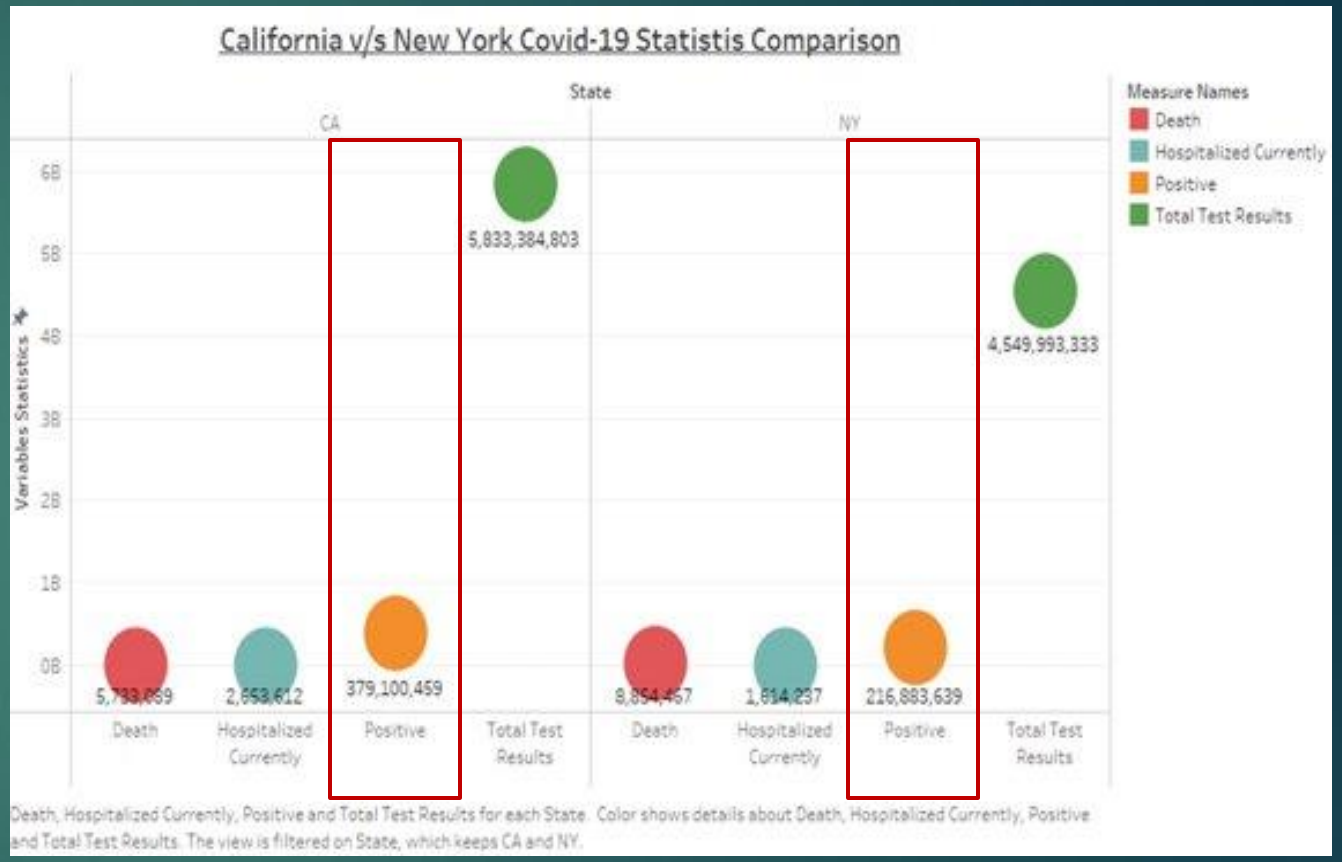
TOTAL TEST RESULTS

- Ratio data
- Counts total number of covid test conducted on each state population
- CALIFORNIA:
  - Mean: 15808631
  - Standard deviation: 15138231
- NEW YORK:
  - Mean: 12264133
  - Standard Deviation: 11600514

Descriptive Statistics : Key Variables

POSITIVE CASES

- Ratio Data
- Measures the number of affected people with covid19 virus
- CALIFORNIA:
  - Mean: 1027373
  - Standard Deviation: 1110252
- NEW YORK:
  - Mean: 584592
  - Standard Deviation: 426828.2



Descriptive Statistics : Key Variables

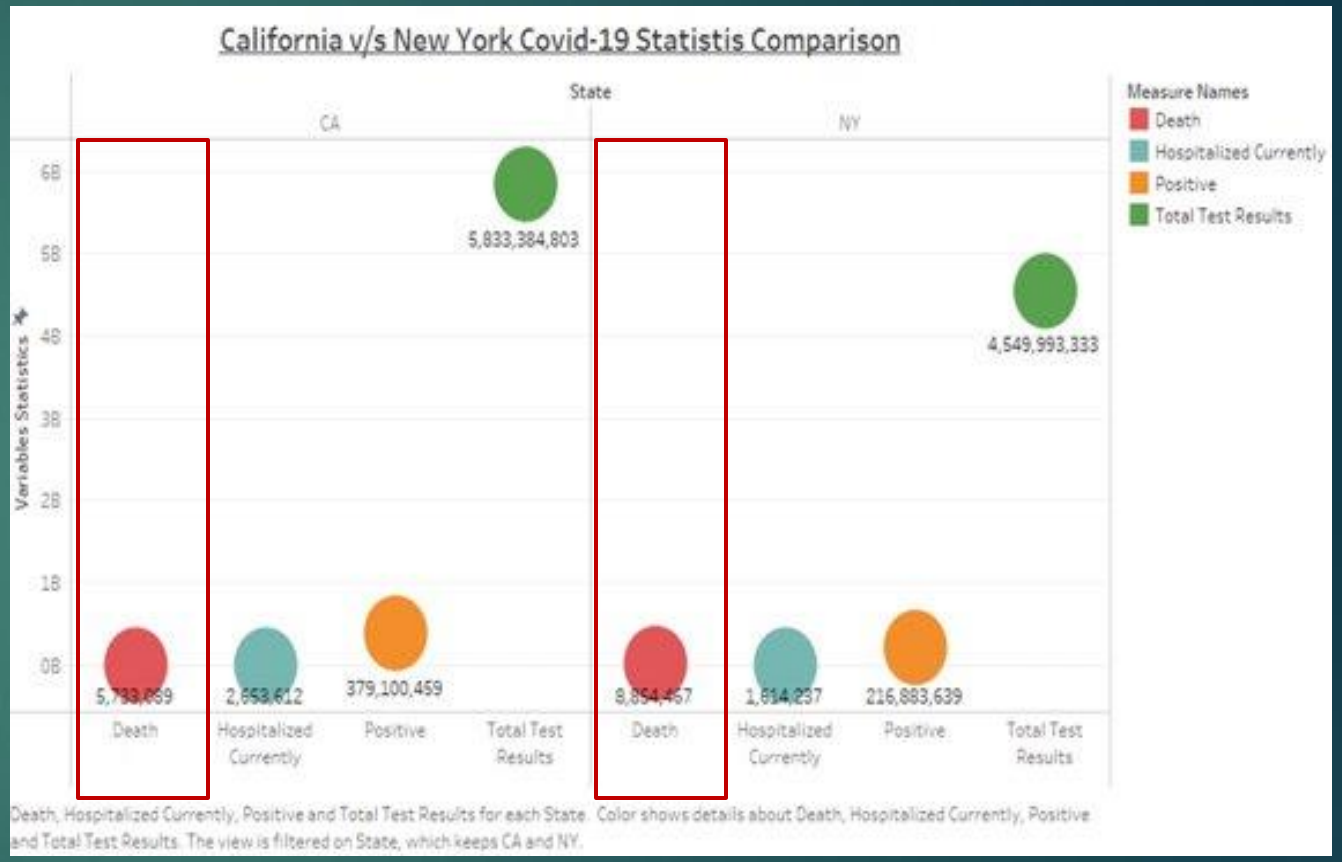
HOSPITALIZED CURRENTLY

- Ratio Data
- Count Of Positive Cases That Are Admitted For Further Treatment
- CALIFORNIA:
  - Mean: 7191.36
  - Standard Deviation: 5741.391
- NEW YORK:
  - Mean: 4351.043
  - Standard Deviation: 4602.676

Descriptive Statistics : Key Variables

DEATH

- Ratio Data
- Measures the number of people died because of covid19 cases
- CALIFORNIA:
  - Mean: 15536.83
  - Standard Deviation: 14050.95
- NEW YORK:
  - Mean: 23866.49
  - Standard Deviation: 9413.505



Analytical Approach

We have used our key variables to conduct following two experiments:

- **Experiment 1**:
  - t-test was conducted to compare the total test results for the states of California (CA) and New York (NY)
- **Experiment 2**:
  - Regression analysis was conducted to check which key variables affect the total test results

Experiment 1 Findings

Total Test Results in CA and NY

H0: There **IS NOT** a significant difference in the total test conducted in California and New York

H1: There **IS** a significant difference in the total test conducted in California and New York

t-Test: Two-Sample Assuming Unequal Variances		
	CA_totalTestResults	*NY_totalTestResults*
Mean	15808630.9	12264132.97
Variance	2.29166E+14	1.34572E+14
Observations	369	371
Hypothesized Mean Difference	0	
df	689	
t Stat	3.573609758	
P(T<=t) one-tail	0.000188325	
t Critical one-tail	1.647068188	
P(T<=t) two-tail	0.00037665	
t Critical two-tail	1.963413002	

Experiment 2 Findings

Regression Analysis for California (CA)

H0: Death, Hospitalized currently and positive **IS NOT** a significant predictor of Total Test Results

H1: Death, Hospitalized Currently And Positive **IS** a Significant Predictor Of Total Test Results

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.991749114							
R Square	0.983566306							
Adjusted R Square	0.983431234							
Standard Error	1948588.8							
Observations	369							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	8.29472E+16	2.76491E+16	7281.821825	0			
Residual	365	1.3859E+15	3.797E+12					
Total	368	8.43331E+16						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1673389.168	300579.9506	-5.567201553	5.02419E-08	-2264475.011	-1082303.325	-2264475.011	-1082303.325
death	853.9021907	54.76788339	15.59129435	2.30331E-42	746.2019916	961.6023897	746.2019916	961.6023897
hospitalizedCurrently	351.6020335	36.16373116	9.72250435	5.0826E-20	280.4866133	422.7174538	280.4866133	422.7174538
positive	1.641656609	0.782913693	2.096855149	0.036694102	0.102068891	3.181244327	0.102068891	3.181244327

Experiment 2: New York (NY)

Regression Analysis for New York(NY)

H0: Death, Hospitalized currently and positive **IS NOT** a significant predictor of Total Test Results

H1: Death, Hospitalized Currently And Positive **IS** a Significant Predictor Of Total Test Results

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.967681724							
R Square	0.936407918							
Adjusted R Square	0.935888092							
Standard Error	2937288.234							
Observations	371							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	4.66253E+16	1.55418E+16	1801.386257	3.9394E-219			
Residual	367	3.16635E+15	8.62766E+12					
Total	370	4.97916E+16						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	572449.9779	557103.7984	1.027546356	0.304840023	-523066.1995	1667966.155	-523066.1995	1667966.155
death	-207.5184668	31.99910587	-6.485133294	2.86719E-10	-270.4430749	-144.5938588	-270.4430749	-144.5938588
hospitalizedCurrently	-266.2080208	38.25851577	-6.958137696	1.59282E-11	-341.4414388	-190.9746027	-341.4414388	-190.9746027
positive	30.45320834	0.715318163	42.57295552	4.8802E-144	29.04657169	31.85984499	29.04657169	31.85984499

Interpretations

- As per the dataset, California had its first COVID-19 case a month before New York. By March, New York had eclipsed California, and cases exploded. Making challenges in both the states
- According to the results of experiment 1, which is the t-test done on the data set, there is a statistically significant difference between the total test conducted in California and New York
- With regards to California-specific regression analysis, H1 was highlighted, indicating a significant predictor. Therefore, a review of the data set reveals that fatalities, hospitalizations, and the presence of positive outcomes are all important indicators for the state of California
- While the statistical regression analysis conducted for the state of New York plainly indicates that there is no significant predictor in the data set that might provide insight into the increase in covid instances occurring in the state, thus emphasizing the H0

Recommendations

- As per the data set and analysis performed by us, the quantity of testing done in New York is considerably less than that done in California resulting in a higher number of positive instances in California
- According to us, COVID-19 may be regulated in New York by increasing the testing with help from the government officials
- This will help identify the positive cases at earlier stages and take necessary precautions by frontline workers to save their people

References

- https://www.kaggle.com/ramjasmaurya/covid-in-usa-states
- https://dph.georgia.gov/covid-19-daily-status-report
- https://publichealthmdc.com/blog/a-message-of-thanks-from-public-health-madison-dane-county
- https://abc7news.com/feature/health/coronavirus-how-close-was-california-to-becoming-a-new-york-level-crisis/6108354/


THANK
YOU


BE HEALTHY, BE CLEAN
CLEAN & DISINFECT
SOCIAL DISTANCE
STAY
6
FEET AWAY