**Madakini**

**Q1. Which States reacted most/least aggressively to COVID (largest change in travel patterns before and after COVID)? Compare individual state’s pattern with nationwide data.**

**Graphs:**

1.Time Series of COVID Cases, total death, population staying at home, population not staying at home (Overlay 2019 and 2020) in USA

Chart, line chart

Description automatically generated

Observations :

Above graph is showing changes in number of trips one year before and one year after COVID started at national level.

* We can see a overall reduction of trips after covid.
* Specifically when COVID was picking up we can see dip in number of trips.
* After some time when COVID was stabilizing we can see increment of trips although COVID cases was happening.

1. Time Series of COVID Cases, total death ,population staying at home, population not staying at home (Overlay 2019 , 2020 and 2021) in California

Chart, line chart

Description automatically generated

**Observation** : We picked CA as one of highly COVID impacted state ,we can see similar pattern of reducing trips as national trend .But here intensity is high.

3)Time Series of COVID Cases, total death ,population staying at home, population not staying at home (Overlay 2019 , 2020 and 2021) in Vermont

Chart, line chart

Description automatically generated

**Observation**: We picked VT as one of the lowest COVID impacted state ,we can see similar pattern of reducing trips as national trend .But here intensity is low.

4) State Specific Bar Graph showing travel before / travel after , potentially all 50 states

Chart, bar chart

Description automatically generated

5) Change in percentage of travel after COVID-State wise

Chart, bar chart

Description automatically generated

**Observations**:

We plotted absolute and percentage change in number of trips comparing before and after covid situation. The graphs clearly show CA is the highest impacted state where trips are reduced significantly after COVID. And WI is the least impacted state.

**Q2. How did COVID change the nature / length of travel?**

5) Time Series of Trip Type over Time (Nationally)

Chart, histogram

Description automatically generated

**Observations**:

From the graph We can see less impact of COVID on long range travels (>=500). There is an impact on the medium range travel (>= 100 <250) when COVID started .But we can see significant impact on short trips (<100) due to COVID.

GUS

This part of the project aim to answer two specific questions in regards to the changes in travel pattern.

1. How long does it take for travel pattern to be affected after new cases or new deaths are reported.
2. If the new cases or new death reported has the same effect on every state or if each state react differently.

In order to look at the travel pattern, we must first determine how long it take after cases and deaths are reported before travel pattern changes. To do that we will assume that travel pattern is not affected by the reporting on the day of and that the higher number of people staying at home reflect change in travel pattern.

To determine the number of lag days to use in calculation, a series of lags for number people staying at home are created (1,3,5,7,9,11&13). From the lag data Pearson-R value are calculated for both **“New Case vs Population Staying at Home”** and **“New Death vs Population Staying at Home”**

The table below show the result of the Pearson-R Value calculation

Table

Description automatically generated

From the table, we determined that there are moderate correlations between the population travel pattern and new cases/new deaths reported. The strongest relations are between 11 days after reporting and New Death. Therefore, we will use 11 days after new death is reported for calculation for the other states.

Using the information we have, we plot a series of scatter plot with trend line to see if there is any correlation for the other states as well as calculate the Pearson R value. Note that for the initial calculation, we have selected California for the calculation. The scatter plots show a varying degree of change in travel pattern for differences states as well as in our R-value tables.

Graphical user interface, application

Description automatically generated

Based on the R value tables and the scatter plot, we can conclude that while there reported death affect travel pattern, not all states’ travel pattern are affected by the new death reported. The state that travel pattern is affected by new death reported are Illinois, California, and New York. While the states that travel pattern are less affected or not affected by the death reported are Texas and Florida.

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

PRINA

**Hypothesis: Reducing number of trips reduced the subsequent COVID-19 Case Count – Regression**

**Steps taken for this analysis:**

**1) after\_covid data is used for this analysis**

**2) Heatmap is created to show that highly corelated data should not be paired together. Result saved in output\_data/Corr\_heatmap.png**

**Chart, histogram

Description automatically generated**

**3) Regplot to show that the data is Linear.The straight line shows that the data is linear. Result saved in output\_data/Plot\_Linear\_regression.png**

Chart, scatter chart

Description automatically generated

**4) Linear Regression is run on "New COVID Cases" and "Number of Trips". Train data is 80% and Validation (Test) data is 20%**

**The train data R square is 0.217 and validation data R square is 0.211. The coefficient result is 0.0005 which shows that the data is not corelated. Less value of R square means that the model is not very accurate**

**5) Residual plot to prove the Linear Regression is created. The result is saved as is saved as output\_data/Residual\_Plot.png**

**The predictor value(X-axis) is "Number of Trips" and " residual value(Y-axis) is "New Cases".**

**Residual = Observed – Predicted**

**The negative value on Y-axis means predictor is too low**

Chart, scatter chart

Description automatically generated

**6) Fancy residual plot using Visualizer is created to prove the Linear Regression. The result is saved as is saved as output\_data/visualizer\_residual\_plot.png**

**The predictor value(X-axis) is "Number of Trips" and " residual value(Y-axis) is "New Cases".**

**Residual = Observed – Predicted**

**The negative value on Y-axis means predictor is too low**

Chart, scatter chart

Description automatically generated

**7) QQ-Plot and Histogram is created to show that the results are left-skewed. The results is saved as output\_data/QQ-Plot.png**

**As the data is left skewed (negative skew), it proves that the data is not normally distributed**

COLLIN

Summary: We evaluated the impact of COVID-19 on Travel Patterns, as well as the impact of travel patterns on COVID-19, using two publicly available datasets (1) The CDC COVID-19 dataset, which captures daily COVID cases and deaths at the state level (2) The BTS Trips dataset, which captures state level data on the number of trips taken on a daily basis, by trip distance, as well as the population staying at home (no trips >1 mile)

Key Question: How Did People react to COVID-19, in terms of travel patterns, and how was this reflected in subsequent infection and death rates?

Hypothesis: Reducing travel after COVID-19, or not travelling much to begin with, had a statistically significant negative effect on COVID-19 infection rates.

Additional Observations:

- Some states (e.g. IL, CA) where highly responsive to lagged COVID-19 death rates, while others were not

- Roughly 4% more of the population stayed at home on a given week after COVID, compared to pre-COVID levels of staying at home

- Using a simple linear regression, staying at home, and increasing the liklihood of staying at home, did not appear to have a statistically effect on COVID-19 infection rates

- The dataset don't appear to have any statistical properties that would make linear regression a bad choice (e.g. standard errors which aren't normally distributed)

Conclusion:

- Population staying at home, defined in several different ways (population at home, % of population at home, incremental % staying at home), did not appear to have a statistically significant effect on COVID-19 rates.

Limitations / Future Research: (1) The CDC dataset only reports at the state level, which limits our ability to tie more granular trip data to state level infection and death rates (2) The trips dataset doesn't provide any information about origin, destination, or duration of travel, which limits our ability to understand the manner in which type of travel impacted infection rates (3) BLS definition of "stayed at home" is based on travel distance, which biases results for places dominated by high density urban area (e.g. DC, NY)