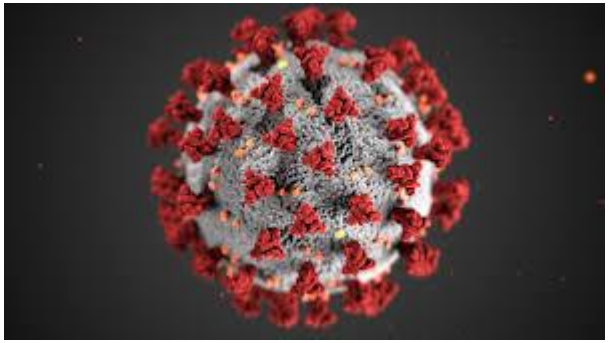
Analysis of COVID-19’s Impact on NY Counties

Introduction

- COVID-19 has disrupted the daily lives of eight billion across the globe, which has prompted political instability, global anxiety, and economic chaos.
- In the early days of the pandemic, New York City was one of the hardest hit cities as there was a record surge in cases, hospitalizations, and deaths.
- Generally, local and state municipalities were unprepared for the wave of Coronavirus.



Why did COVID-19 hit the Big Apple the hardest?

While COVID-19 first appeared in the United States in Washington State, New York City and the state was one of the hardest hit places in the United States. There are several reasons for this reality:

1. New York's massive transportation system is a breeding ground for transmission
2. "Close the Front Door but leave the Back Door wide open"
3. New York's shared spaces
4. Lack of reliable information

Why this Matters & Purpose

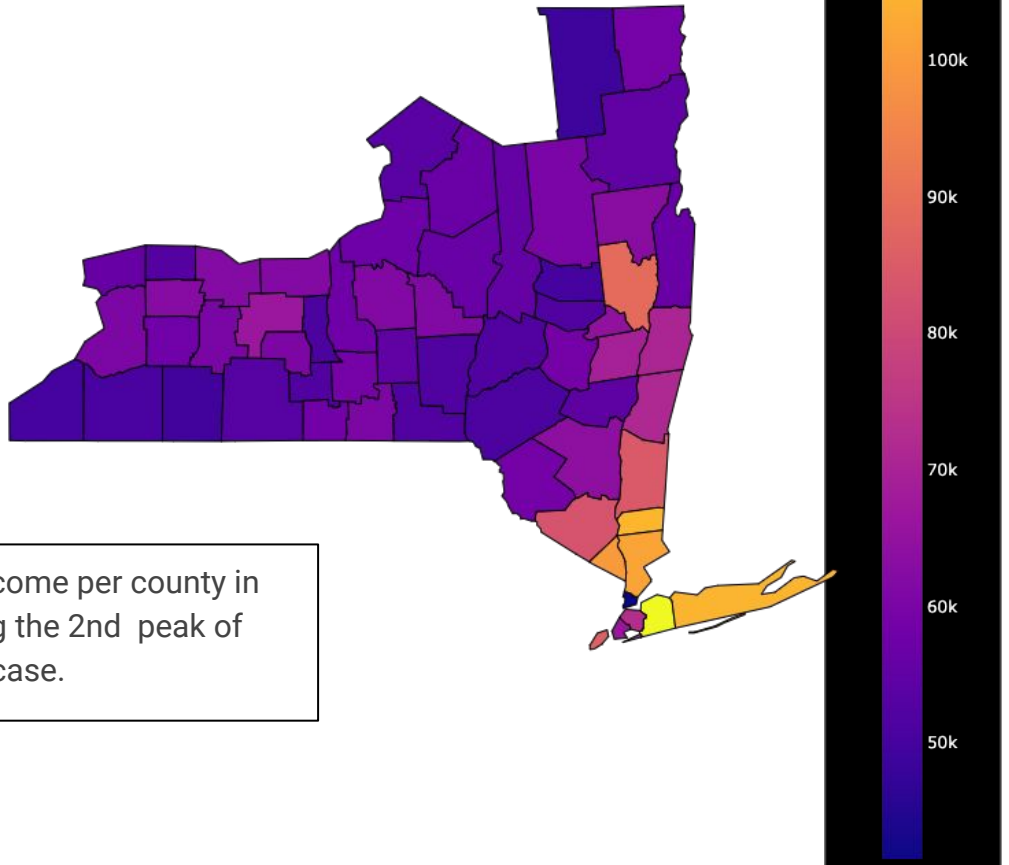
- A lack of a coordinated federal response from government officials resulted in mixed opinions which were often reflected on Twitter
- We are seeking to show a connection between twitter data and policy initiatives, economic activity, and other measures of behavior

Data Analysis & Computation

Datasets + Data Wrangling & Cleaning

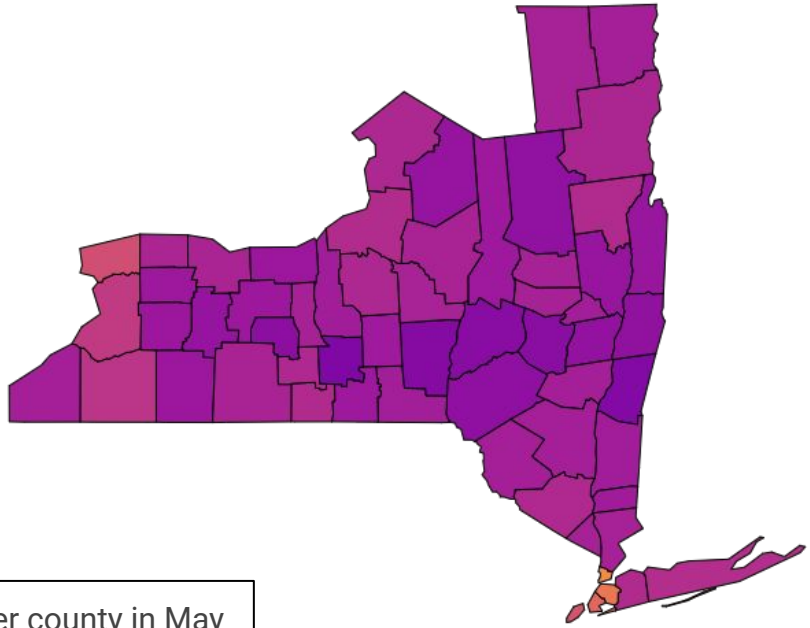
- Twint (Twitter scraping program)
  - Keywords: Covid-19, coronavirus, corona
  - Centroid & radius for each county
  - February - November 2020
- Additional Covid data collection sources
  - County level Gender / sex breakdown- Census & census estimations
  - County-level Economic Data- USDA Economic Research Service
  - Policy/stringency information - Oxford dataset
  - Unemployment rates- FRED (Fed Reserve Bank of St. Louis)
- Our end dataset consisted of an observation from each county for each day from February to November 2020 containing a mixture of demographic, economic, and covid related data.

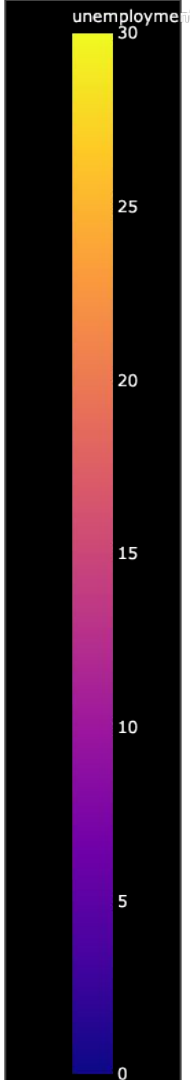
Median Income



Median Income per county in May during the 2nd peak of daily new case.

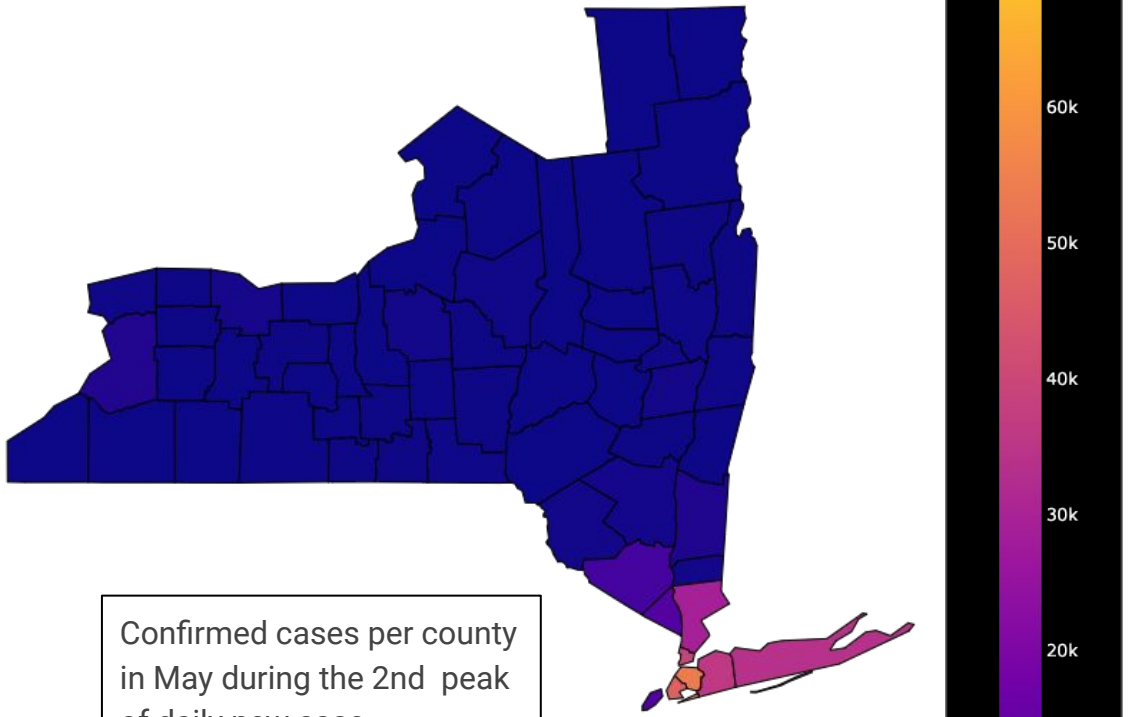
Unemployment Rate





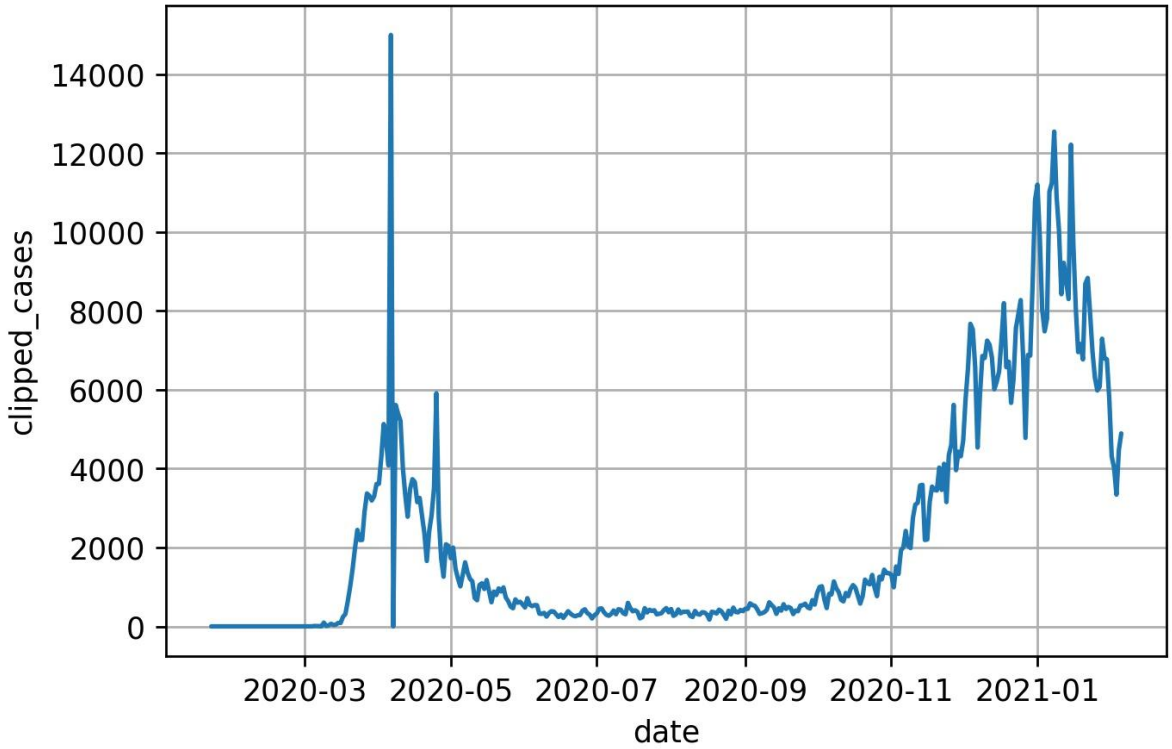
Unemployment rate per county in May during the 2nd peak of daily new case.

Confirmed Cases



Confirmed cases per county in May during the 2nd peak of daily new case.

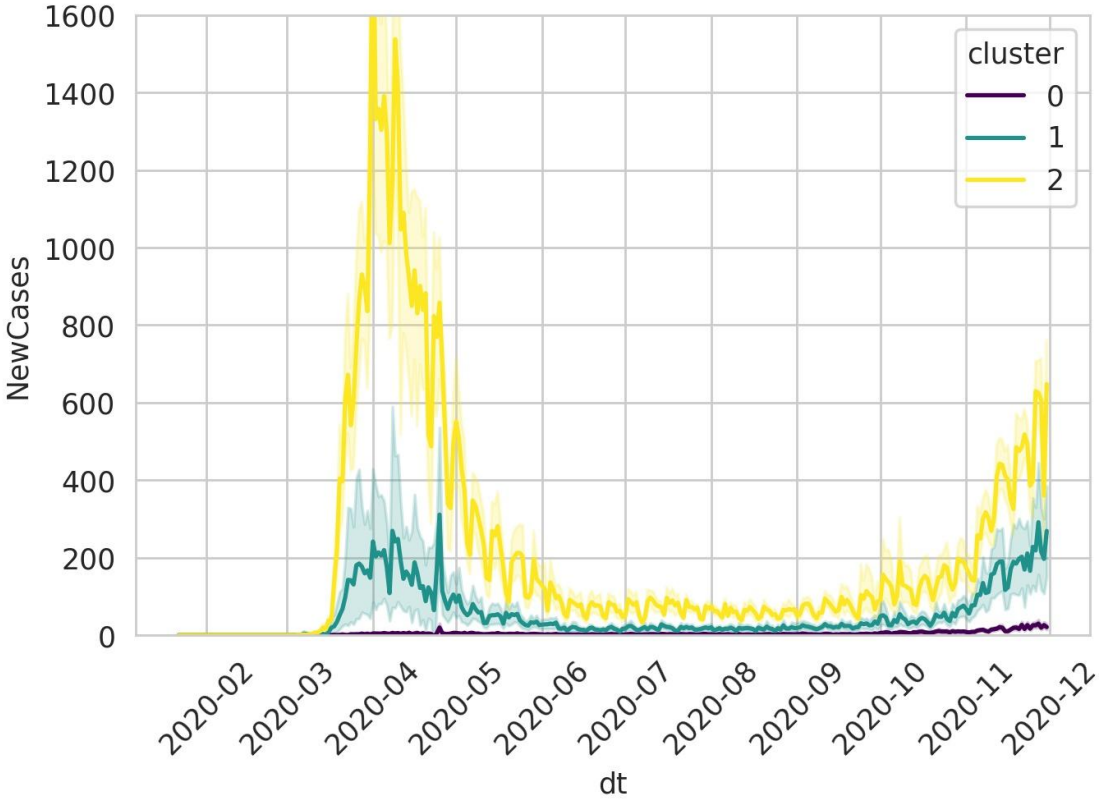
New Cases over Time in New York State



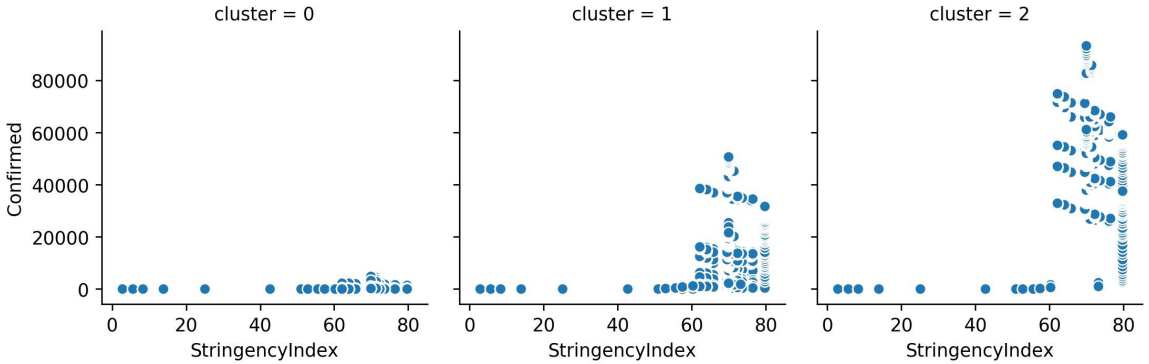
Statistical Analysis

- To simplify describing counties we use K-means clustering to group each county based on demographic and economic information.
- Selected a 3 cluster solution
  - The clusters broke out primarily by population.
    - Cluster 0: Smaller Counties
    - Cluster 1: Medium Sized Counties
    - Cluster 2: NYC + Long Island

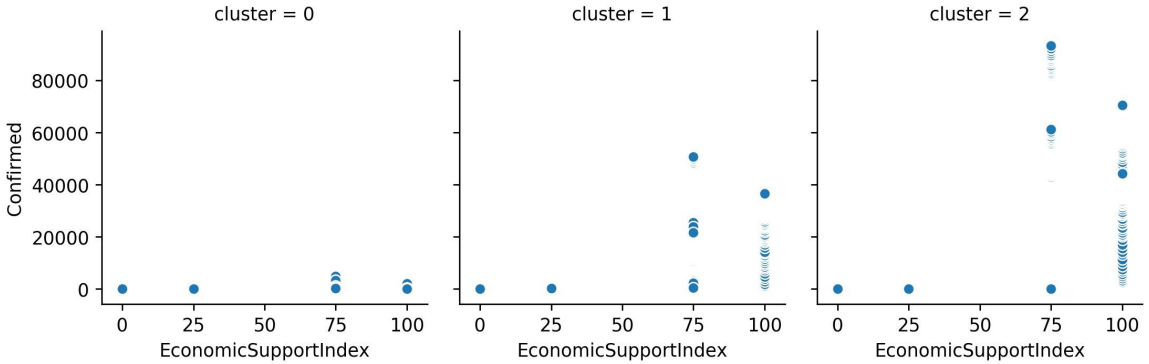
New Cases over Time by Clusters



Exploratory Data Analysis

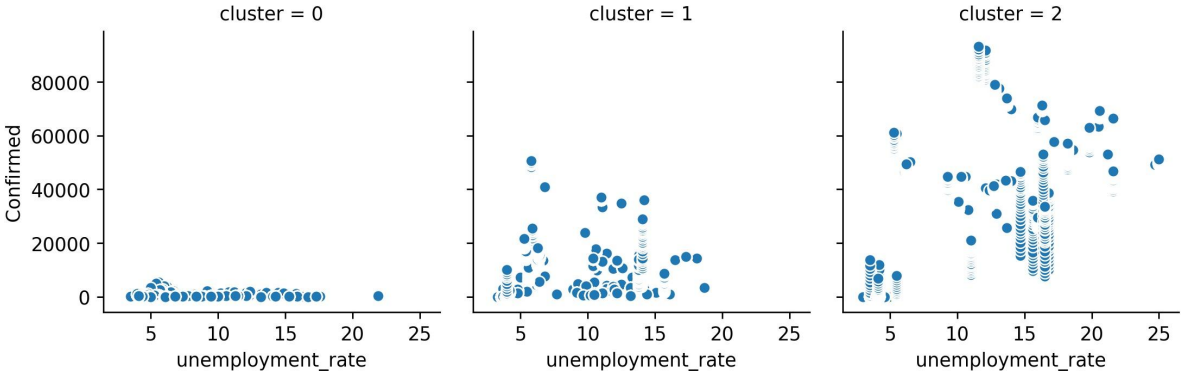


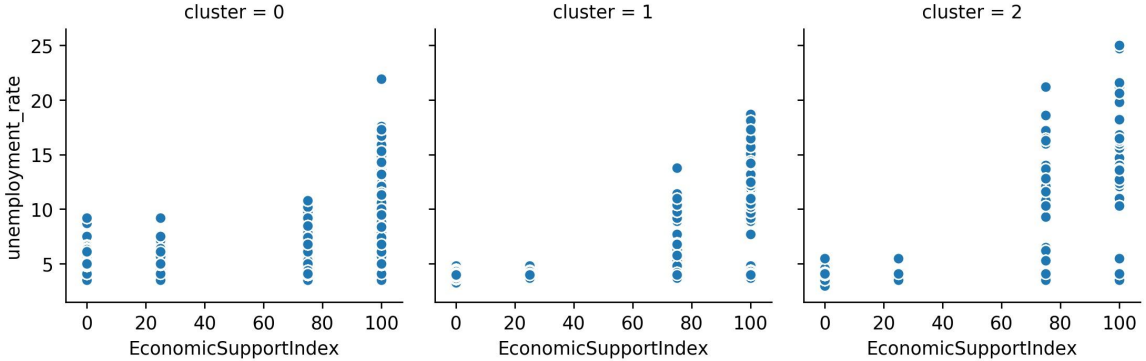
Stricter covid policies were found in counties with the highest confirmed cases



Confirmed cases did not appear to increase economic support

Exploratory Data Analysis



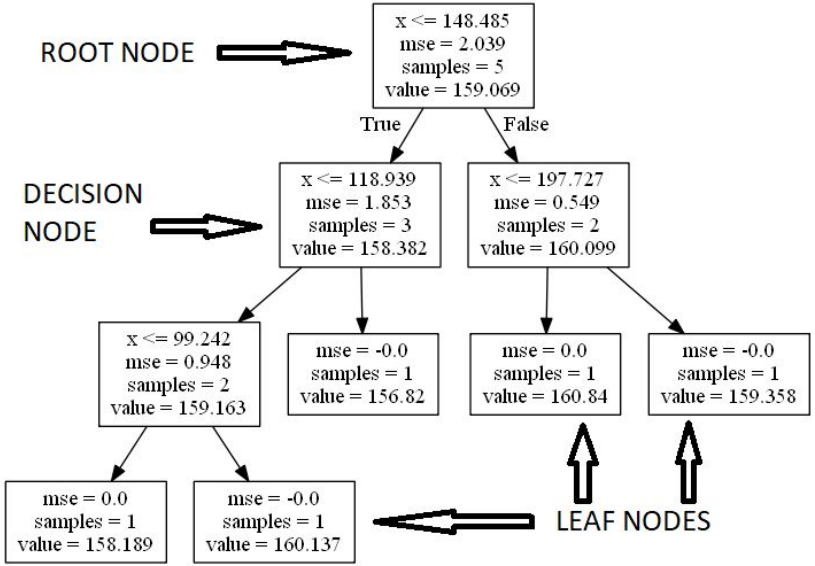


Larger counties with high unemployment saw more cases

Economic support seemed to not be closely connected with unemployment

Predictive Modeling

We used a DecisionTreeRegressor because of its ability to model non-linear relations without implicitly adding presuppositions of the degree of the “line” of best fit.



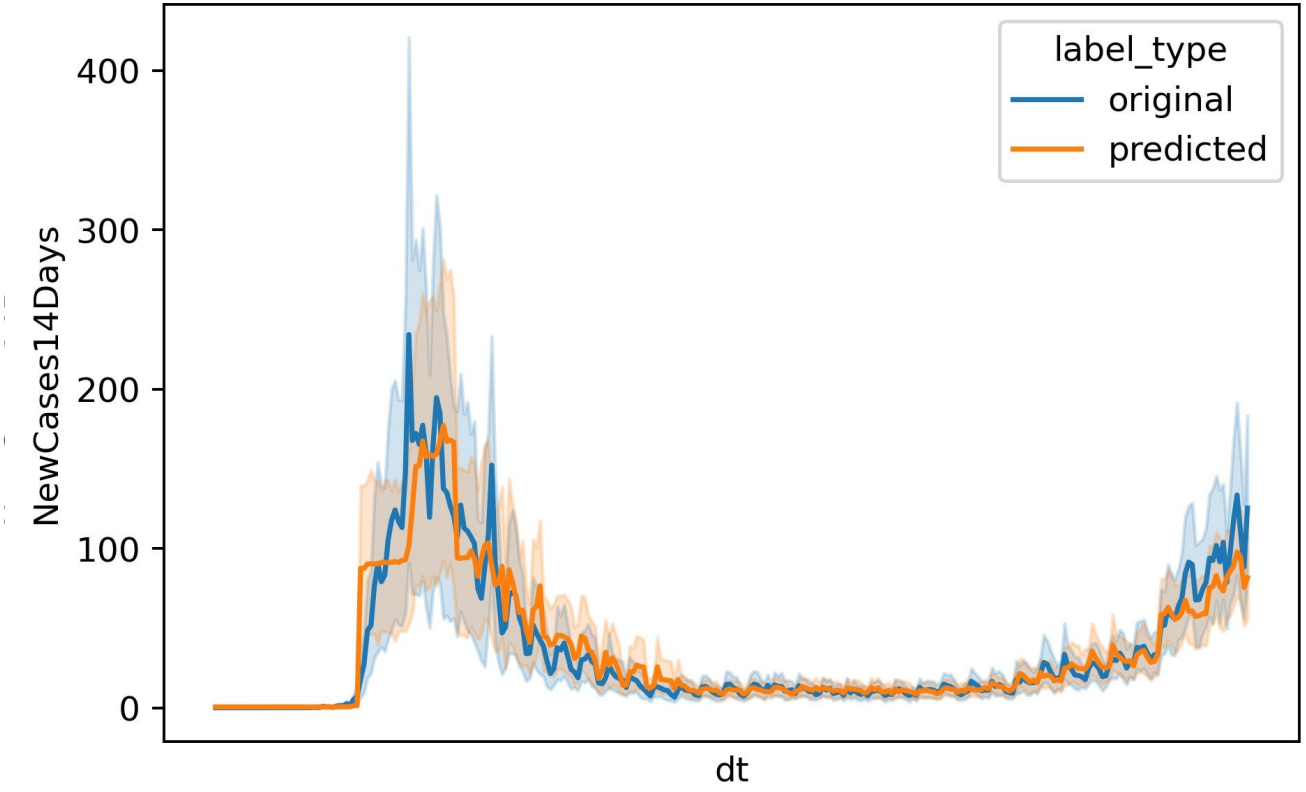
Predictive Modeling

All results presented with 5-fold cross validation

Model	7-Day New Cases	14-Day New Cases	30-Day New Cases
Always Predicting Mean	-0.0047	-0.0048	-0.0051
Always Predicting Median	-0.0048	-0.0625	-0.065
Decision Tree Regressor	**0.6665**	**0.5137**	**0.3**

R^2 score per model for the 7,14, & 30 day predictions. Best results are **bolded.**

Predictive Modeling



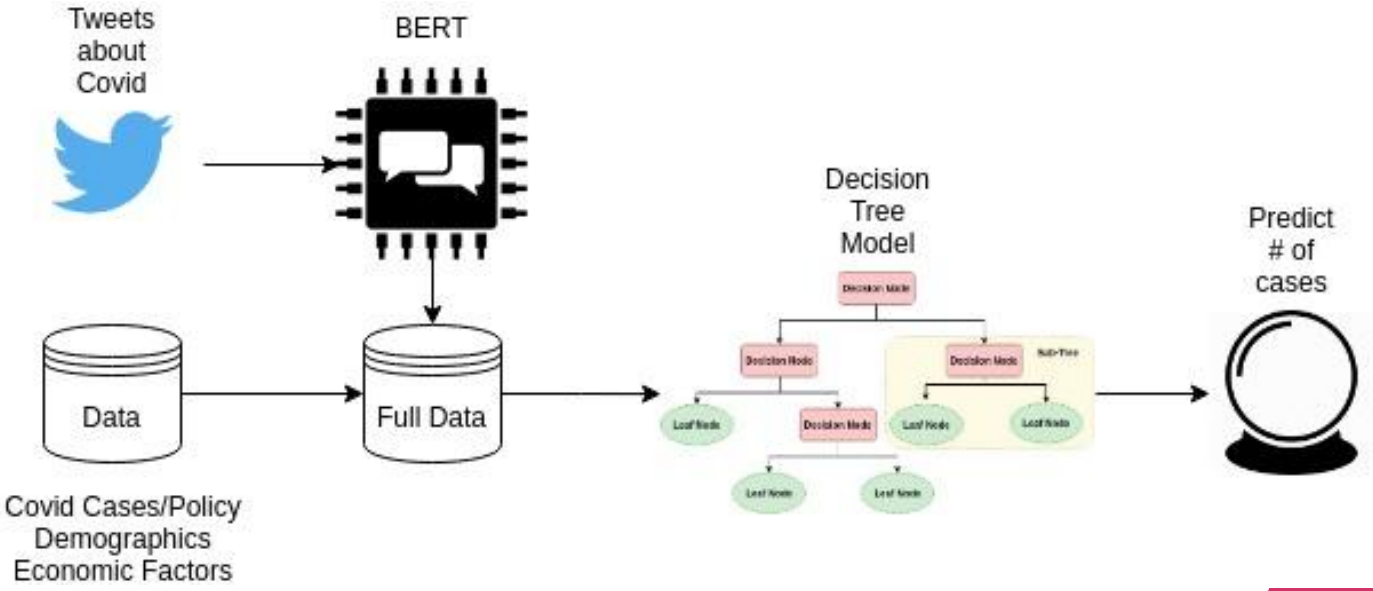
Application / Dashboard

What it does?	Use Case?	Interaction?
Takes inputs from the user and then spits out a prediction of how many cases you will have in the next 7, 14 or 30 days.	To be used by New York State **government officials** to determine how different counties are affected by various levels covid-related shut downs To be used by **citizens** in New York to determine how Covid-related policies will affect their counties	Visit the app and select the NY county of interest Toggle between the various options (cases, stringency lvls, dates, deaths, etc.) Observe the predicted changes in case numbers & plan accordingly

Data Engineering

The web app uses a combination of Dash, plotly, Bootstrap and Heroku.

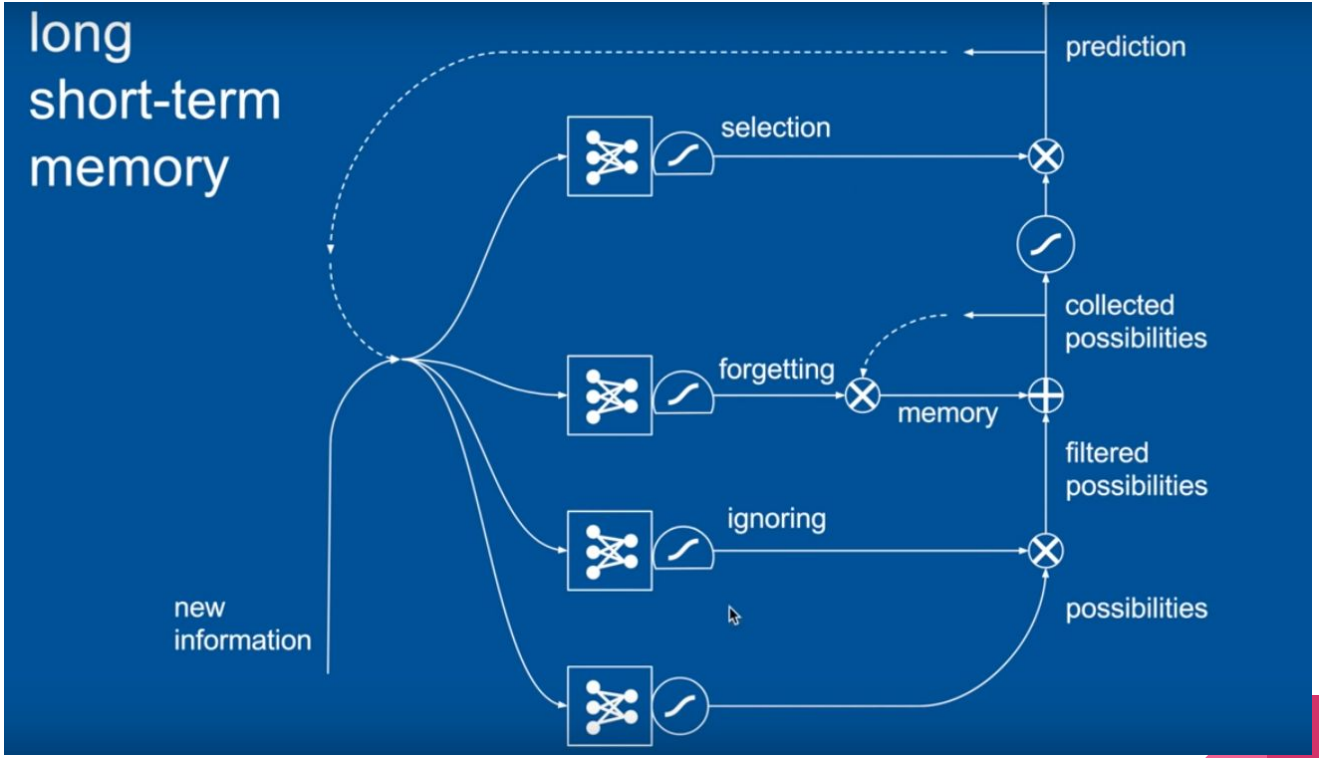
Data Engineering



Link to dashboard

https://ds4a-dashboard.herokuapp.com/

Conclusions & Future Work


long
short-term
memory
prediction
selection
collected
possibilities
forgetting
memory
filtered
possibilities
ignoring
possibilities
new
information

Thank you!

Now time for Q&A