COVID-19 Data Analysis Stage III Report

Jason Manning
Francis Perez
Jamison Valentine
Raiana Zaman
Xinrui (Sam) Zhang

CSC 405-01: Data Science

University of North Carolina Greensboro

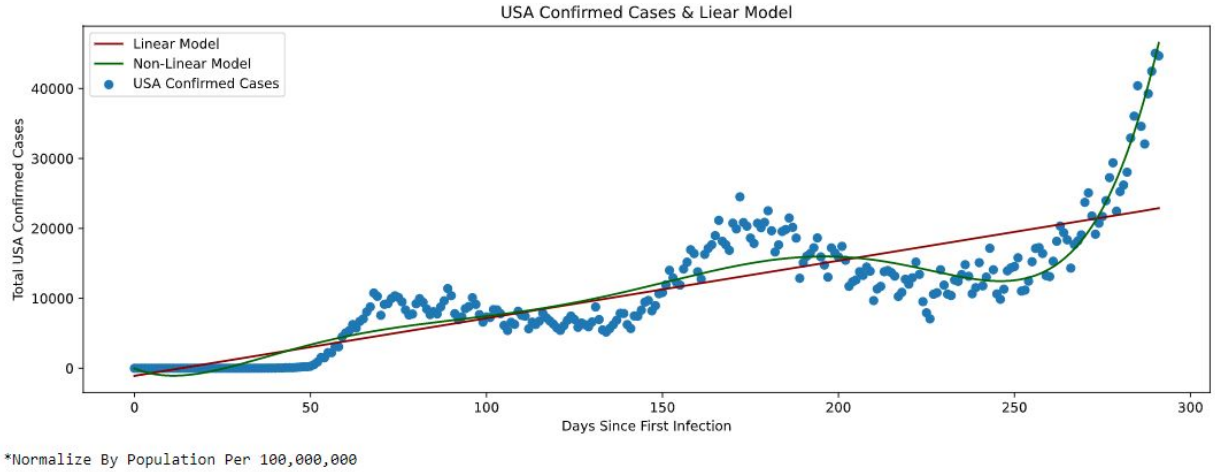
Fall: 2020

Team Task 1:

Part 1.

Develop Linear and Nonlinear (polynomial) regression models for predicting cases and deaths for the USA.

Confirmed Cases:

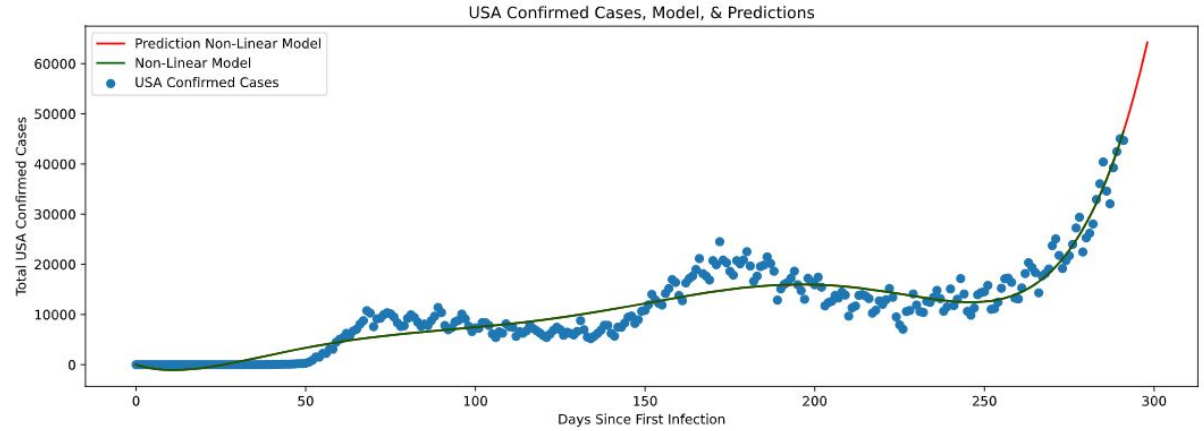


Confirmed Cases - RMSE - Linear Model: 4828.14 | Non-Linear Model: 2799.03

We used a NonLinear Model for the predictions on a lower RMSE value (above).

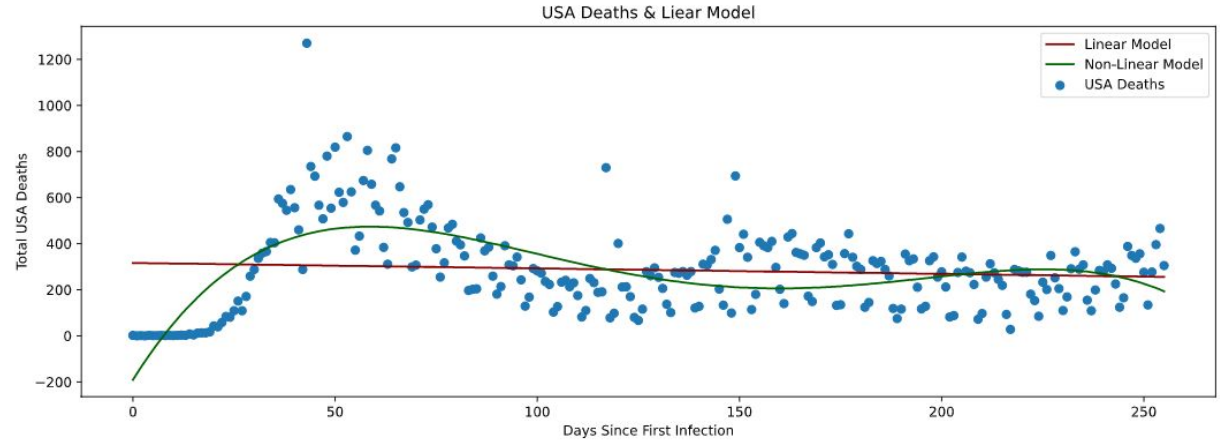
******Number of New Confirmed Cases 7-Days in the future.*******

days_since_infection_start	daily_count
292	48,740
293	51,043
294	53,452
295	55,970
296	58,599
297	61,344
298	64,209



*Normalize By Population Per 100,000,000

Deaths:



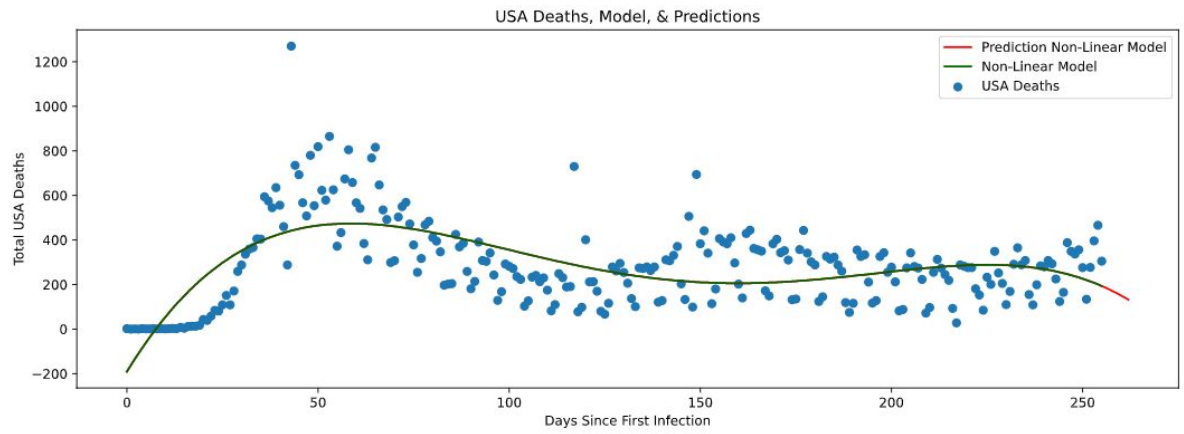
*Normalize By Population Per 100,000,000

Deaths - RMSE - Linear Model: 190.42 | Non-Linear Model: 150.32

We used a NonLinear Model for the predictions on a lower RMSE value (above).

******Number of New Deaths 7-Days in the future.*******

days_since_infection_start	daily_count
256	186
257	178
258	169
259	161
260	151
261	142
262	132



*Normalize By Population Per 100,000,000

Trends Compared USA to Other Countries

USA

- Confirmed Cases: The USA is predicted to see a sharp increase in the number of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- Deaths: The USA is predicted to see a slight decrease in the number of deaths in the week of 11-13-2020 to 11-20-2020.

Brazil

- Confirmed Cases: Brazil is predicted to see very little change, holding at around 7200 per 100,000,000 population, of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- Deaths: Brazil is predicted to see very little change, holding at around 150 per 100,000,000 population, of new deaths in the week of 11-13-2020 to 11-20-2020.

Japan

- Confirmed Cases: Japan is predicted to see very little change, holding at around 300 per 100,000,000 population, of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- Deaths: Japan is predicted to see very little change, hold at around 4 per 100,000,000 population, of new deaths in the week of 11-13-2020 to 11-20-2020.

Mexico

- Confirmed Cases: Mexico is predicted to see a slight increase of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- Deaths: Mexico is predicted to see a slight increase of new deaths in the week of 11-13-2020 to 11-20-2020.

Nigeria

- Confirmed Cases: Nigeria is predicted to see an increase of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- Deaths: Nigeria is predicted to see no change, hold at around 2 per 100,000,000 population, in new deaths in the week of 11-13-2020 to 11-20-2020.

Russia

- Confirmed Cases: Russia is predicted to see a sharp increase of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- Deaths: Russia is predicted to see an increase in new deaths in the week of 11-13-2020 to 11-20-2020.

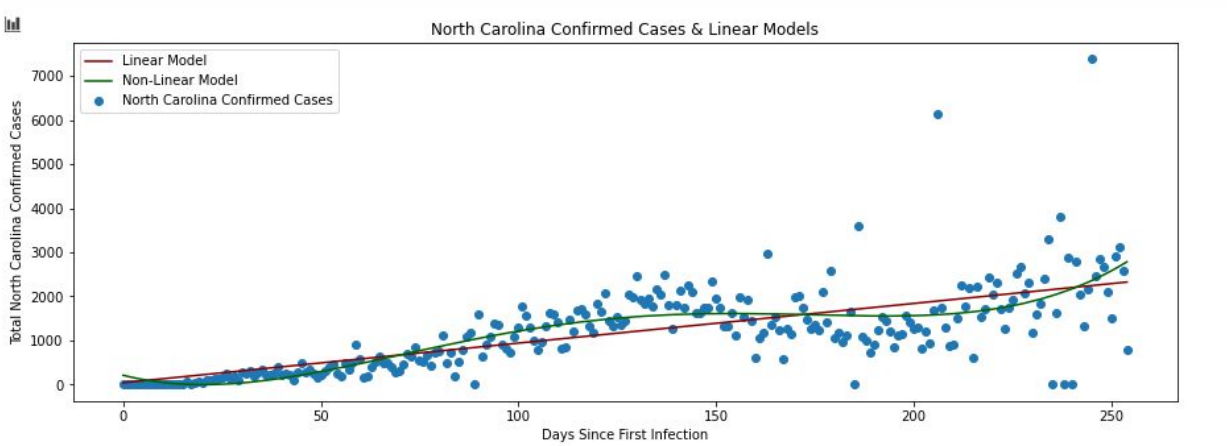
***Other Countries Graphs Can Be Viewed In Team Stage 3 Notebook**

Member Tasks:

Francis Perez:

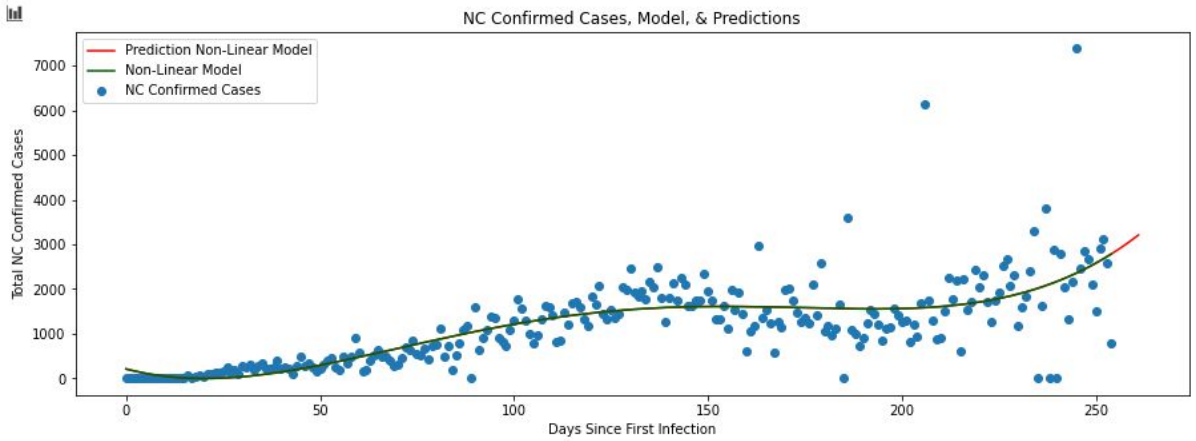
Develop Linear and Nonlinear (polynomial) regression models for predicting cases and deaths for North Carolina

Both Models:



NC Confirmed Case - RMSE - Linear Model: 697.95 | Non-Linear Model: 659.74

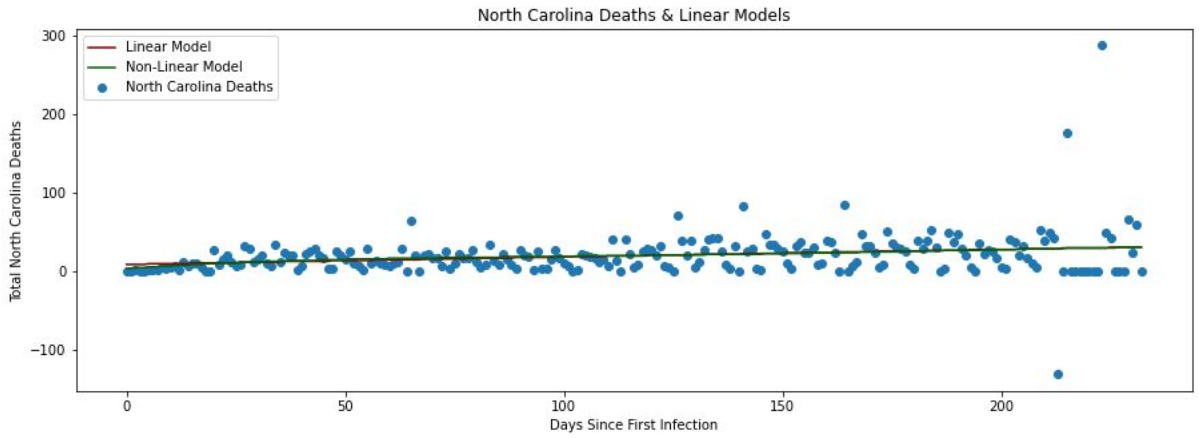
Used NonLinear Model due to lower RMSE Value (above)



Confirmed Cases Predictions 7-Day Future:

days_since_start	total
255	2,839
256	2,896
257	2,955
258	3,016
259	3,079
260	3,143
261	3,209

Deaths Models

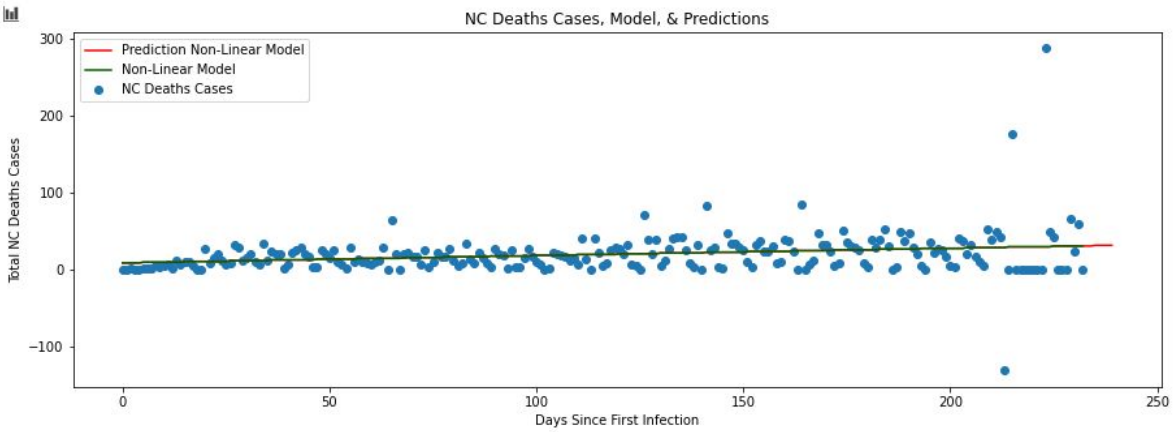


NC Confirmed Case - RMSE - Linear Model: 27.16 | Non-Linear Model: 27.13

Using the Linear Model for making predictions as the RMSE Values are very close and Linear has less variance.

Deaths Predictions 7-Day Future:

days_since_start	total
233	31
234	31
235	32
236	32
237	32
238	32
239	32



North Carolina Trends

- **Confirmed Cases**
  NC is predicted to see an increase in the number of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- **Deaths**
  NC is predicted to see no change in the number of deaths, holding at 32 deaths, in the week of 11-13-2020 to 11-20-2020.

Using These Top - 5 Most Infected Counties From Stage II Per Population

- Duplin County
- Hertford County
- Jones County
- Robeson County
- Scotland County

These Counties still hold the top spot for most Infected bases on current data by re-running Fperez_Task1.ipynb

Duplin County, NC

- **Confirmed Cases:**
  Duplin is predicted to see no change, holding at around 17 per 1,000,000 population, of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.

- **Deaths:**
  Duplin is predicted to see a slight increase of new deaths in the week of 11-13-2020 to 11-20-2020.

Hertford County, NC

- **Confirmed Cases:**
  Hertford is predicted to see no change, holding at around 0 per 1,000,000 population, of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- **Deaths:**
  Hertford is predicted to see no change, holding at around 0 per 1,000,000 population, of new deaths in the week of 11-13-2020 to 11-20-2020.

Jones County, NC

- **Confirmed Cases:**
  Jones is predicted to see an increase of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- **Deaths:**
  Jones is predicted to see hold, holding at around 0 per 1,000,000 population, of new deaths in the week of 11-13-2020 to 11-20-2020.

Robeson County, NC

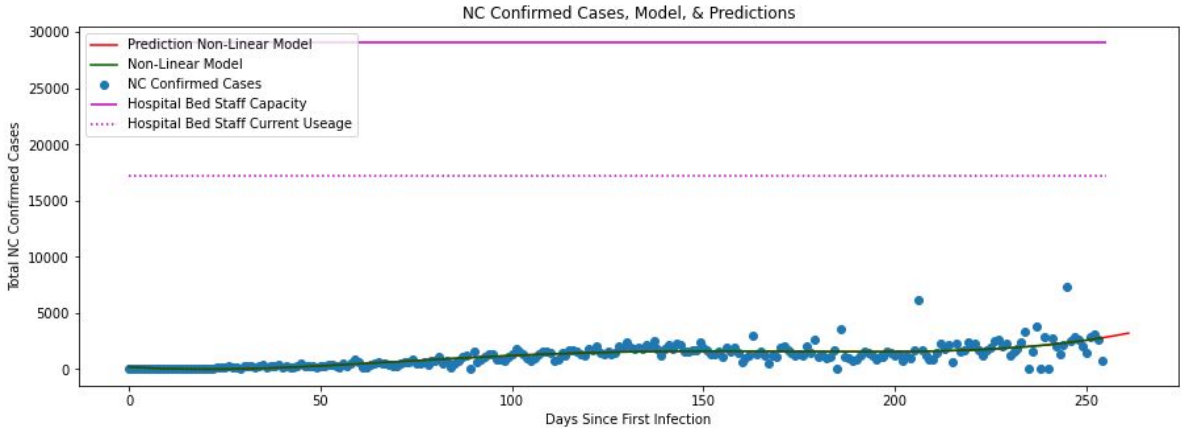
- **Confirmed Cases:**
  Robeson is predicted to see an increase of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- **Deaths:**
  Robeson is predicted to hold, holding at around 3 per 1,000,000 population, of new deaths in the week of 11-13-2020 to 11-20-2020.

Scotland County, NC

- **Confirmed Cases:**
  Scotland is predicted to hold, holding at around 0 per 1,000,000 population, see an increase of new Confirmed cases in the week of 11-13-2020 to 11-20-2020.
- **Deaths:**
  Scotland is predicted to see a sharp increase of new deaths in the week of 11-13-2020 to 11-20-2020.

Plot North Carolina Hospital Bed Data & Current Confirmed Cases & Model & Predictions

	STATE_NAME	NUM_STAFFED_BEDS	Beds_Used
27	NC	29052	17232



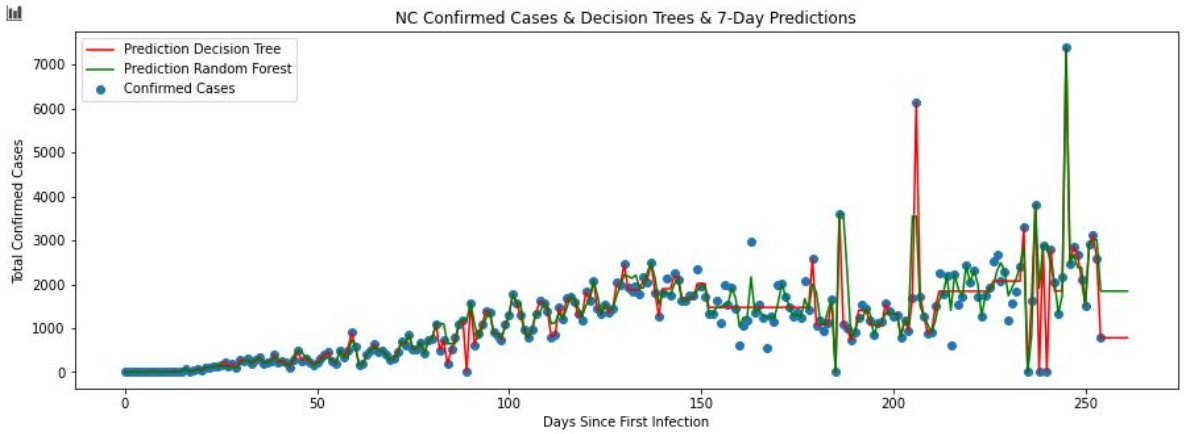
North Carolina is not close to "Point of No Return", as it has had plenty of Hospital Bed Capacity.

These State are close to being on the "Point of No Return". Due to them seeing confirmed cases on or above the "Current Usage" of hospital beds. When the "Current Usage" is then added to the Covid Confirmed cases and the next 7-Days predictions this could be a serious problem for these states.

- **North Dakota- ND:**
  The trend predicted to continue increasing in the next 7-Days. This state has had a day where the confirmed cases have exceeded the number of beds for the state.
- **Wyoming - WY:**
  The trend predicted to continue increasing in the next 7-Days.
- **Iowa - IA:**
  The trend predicted to continue increasing in the next 7-Days.
- **Michigan - MI:**
  The trend predicted to continue increasing in the next 7-Days
- **South Dakota - SD:**
  The trend predicted to continue increasing in the next 7-Days.
- **Utah - UT:**
  The trend predicted to continue increasing in the next 7-Days.
- **Minnesota - MN:**
  The trend predicted to continue increasing in the next 7-Days.
- **Rhode Island - RI:**
  The trend predicted to continue increasing in the next 7-Days. This state has had a day where the confirmed cases have exceeded the number of beds for the state.

**** Graphs Can be Seen in Francis Perez Stage 3 Notebook.**

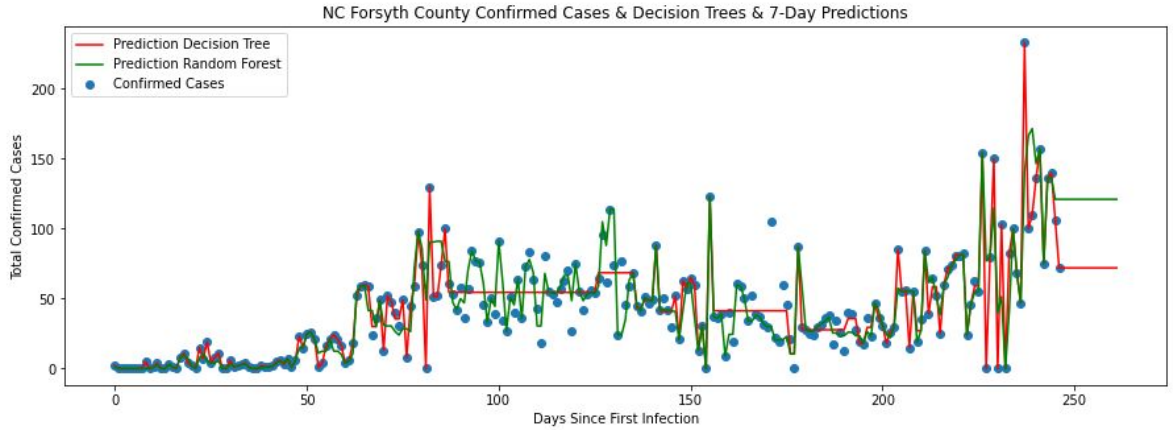
Decision Tree & Random Forest

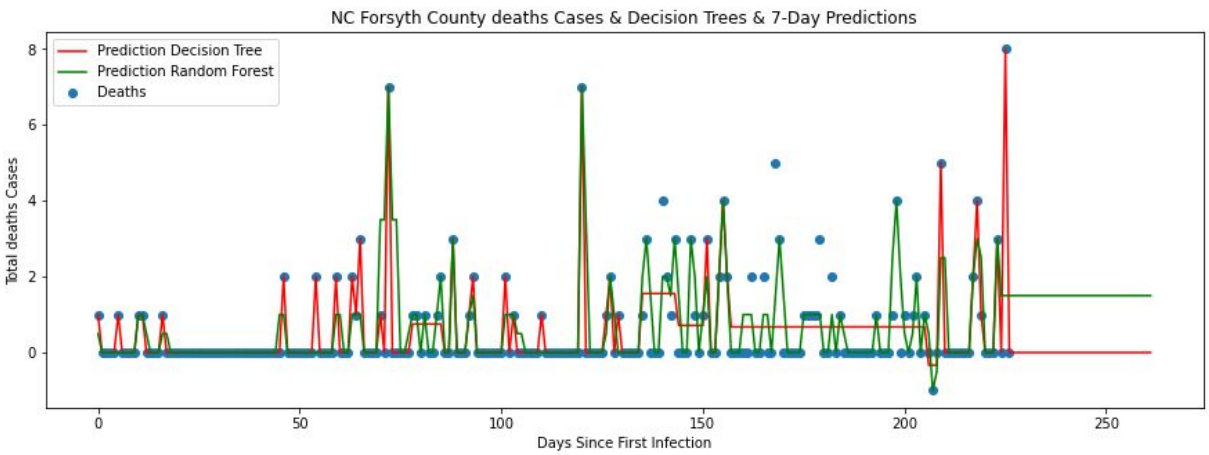


Enrichment data to predict

Data Used:

`employment_mar_gov_fed_subtotal`
`employment_mar_gov_fed_subtotal`
`employment_mar_gov_state_subtotal`
`employment_mar_priv_goods_subtotal`
`employment_mar_priv_services_subtotal`



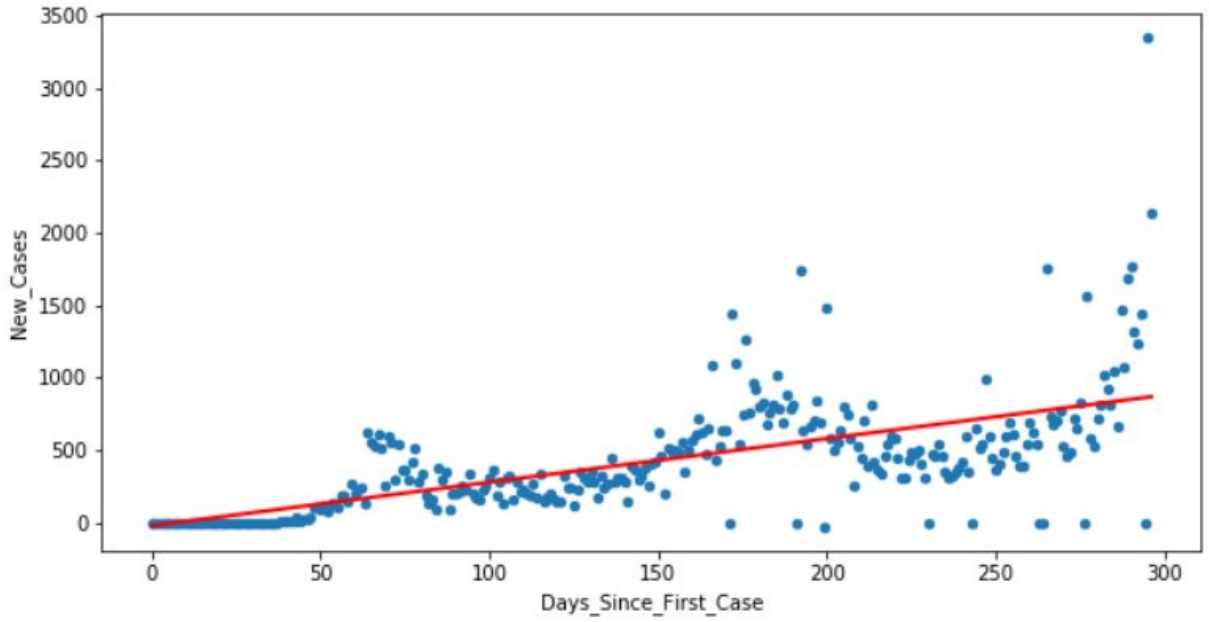


End of Francis Perez

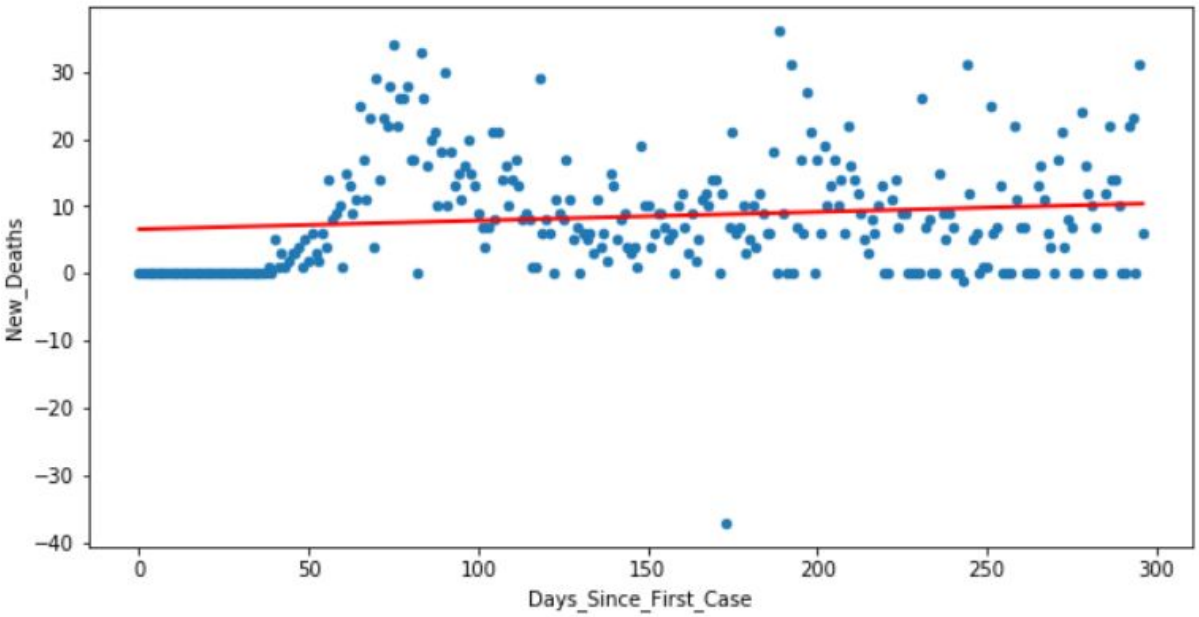
Jason Manning

-Utilize Linear and Non-Linear regression models to compare trends for a single state and it's counties

Washington State Linear Regression for New Cases

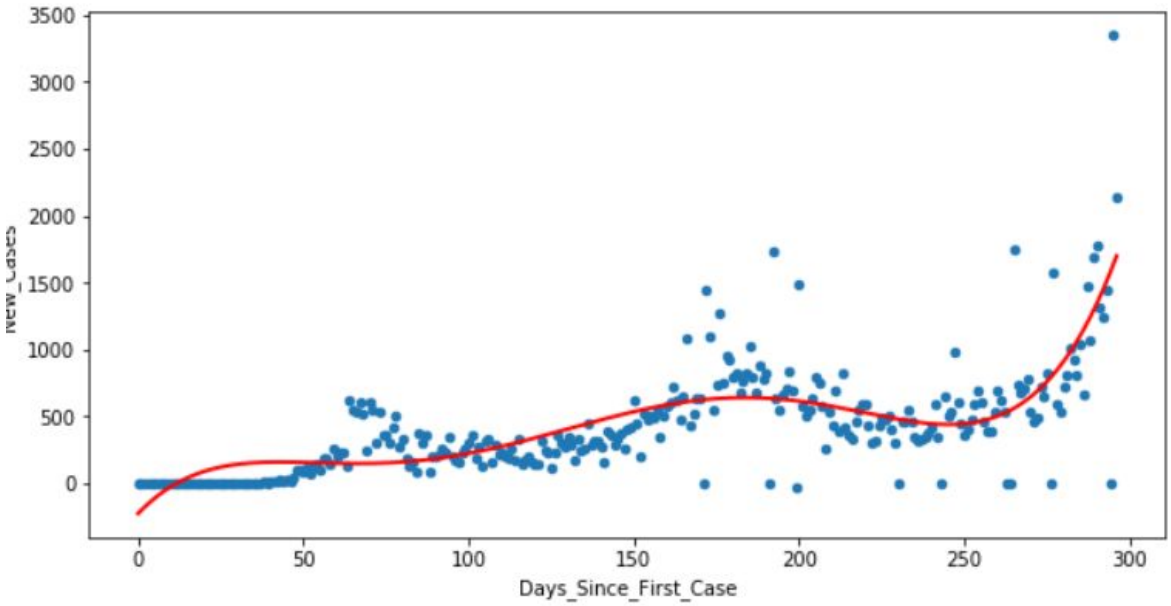


Washington State Linear Regression for New Deaths

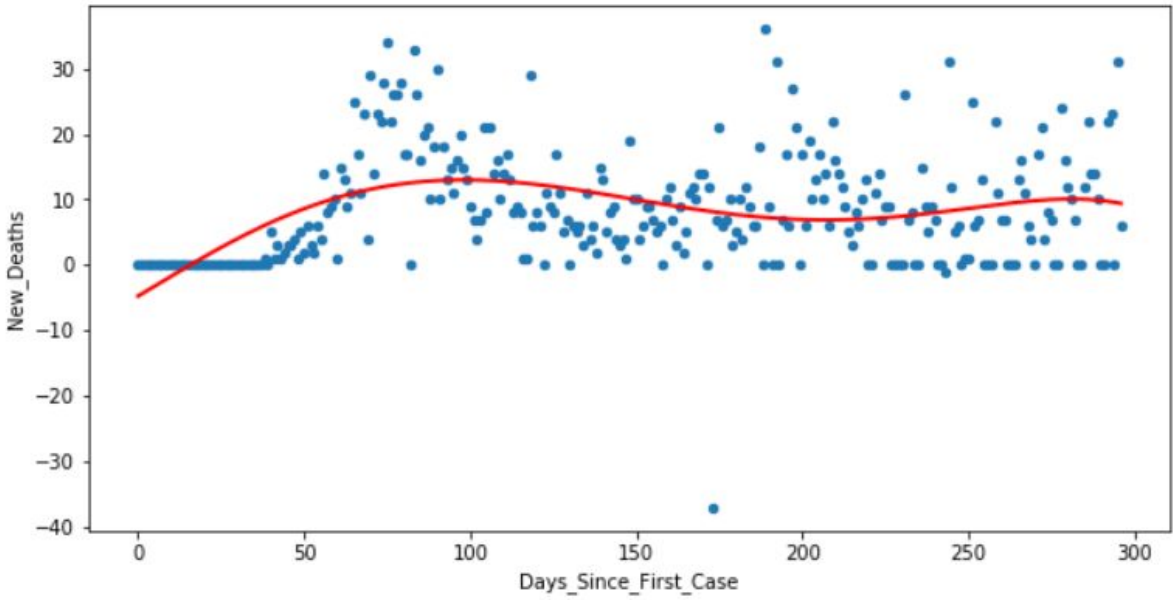


Linear Regression Model for New Cases RMSE: 312.04770774147676
Linear Regression MOdel for New Deaths RMSE: 8.589193930610767

Washington State Non-Linear Regression Model for New Case



Washington State Non-Linear Regression Model for New Deaths



Non-Linear Regression Model for New Cases RMSE: 266.17766900328564
Non-Linear Regression Model for New Deaths RMSE: 7.829011488762823

The lower RMSE for the non-linear regression models indicates a better fit than the linear regression model.

The trend for New Cases is spiking for Washington, while the New Deaths are trending slightly downward. Deaths lag behind the New Cases, so an upturn in Deaths is expected with the increase in New Cases.

Washington County Trends

Franklin, Adams, Yakima, Whitman and Grant counties continue to have the highest infection per 10,000 people.

Franklin county is showing a big upswing in New Cases while the Deaths are on a gradual downward slope. As the New Cases rise in the next week, a turn in the trend of Deaths will be sure to follow.

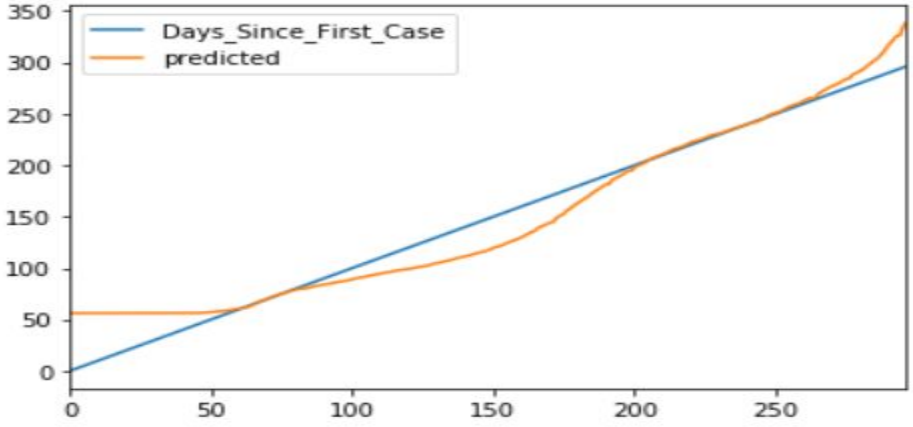
Adams county is showing a moderate upturn in New Cases which should continue over the next week, while Deaths have leveled off.

Yakima county has a moderate uptick in New Cases as well, but also has a slight increase in Deaths that should continue trending that way with the increase in New Cases.

Whitman county is actually trending downward on New Cases while the Deaths are going up. The Deaths should level off and drop with the downward trend of New Cases.

Grant county dipped a little, but it is on the rise again for New Cases. Deaths are trending downward slightly, but will start to move upward following the increase in New Cases.

Washington Hospital Bed Data



The hospital bed data for Washington shows that the state has 10,713 open beds remaining, which is about a third of the total available hospital beds in Washington. Based on trends for the state, they are 80 days out from the point of no return. Enough time to hopefully turn things around and not reach that point.

End Jason Manning