**Outcome of your EDA:**  
  
As part of this project, I have analyzed various techniques to perform the EDA of COVID19 Trends and Outbreak Prediction of Spread in the USA.  
  
The below are the outcomes of my EDA   
  
1. Calculated DeathRate Ratio - From Feb 29,2020 to Nov 18,2020, overall Death Count is 250K. Initially Death Ratio was increased, and it started gradually decreasing from July,2020  
  
2. Number of Death : Number of deaths is increasing day by day ( as of Nov 18 )  
  
3. Confirmed Cases : Number of positive Count is increasing day by day ( as of Nov 18 ) - 11.61 M   
  
4. State Level Cases : Created Animation plot for State Level counts on daily basis. ( Both Confirmed and Death count )  
   observed NY State count had highest counts.  
  
5. Based on the Data as of Nov 18,2020, The prediction of Death count on January 31,2021 is 280K  
    (If the same situation continuous, the count may reach more than 300K in Feb 2021)  
  
6. Based on the Data as of Nov 18,2020, The prediction of Confirmed Cases count on January 31,2021 is 18 Million  
    (If the same situation continuous, the count may reach more than 22 Million in Feb 2021)  
  
7. Verified various Analysis through various Regression, Charts and Probability function to verify the accuracy  
  
8. Compared the model with prediction and actual values to make sure the steps we followed  
  
9. Performed Various Trend Analysis and Hypothesis Test, P-Value findings  
  
10. Created Animation for State Level count with Date Level   
      
As of November 21,2020, We are hearing that vaccination is going to be provided to people and I hope this will help to stop the COVID Spread and deaths.

**What do you feel was missed during the analysis?**

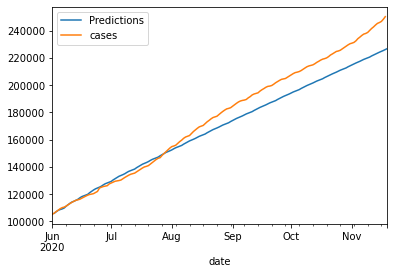
As part of Week 10, I have planned to perform  Prophet time series method for the forecasting but I couldn't  complete the time series forecasting through that. In that way, my forecasting could be more accurate.

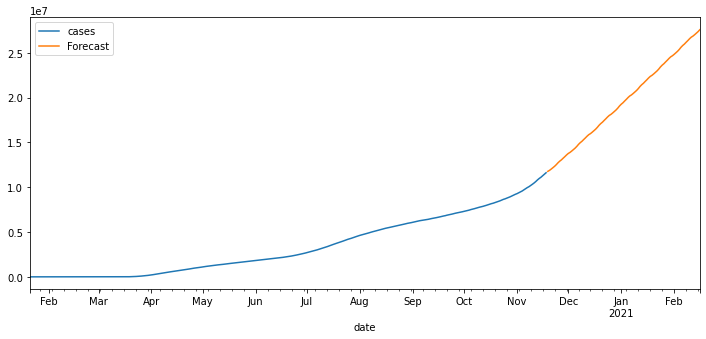
I tried to get more data on Age, ethnicity to perform detailed Analysis on Age,Sex and ethinitiy.    
  
Were there any variables you felt could have helped in the analysis?  
 I felt that if we get a variable for Negative Confirmed Cases of people with non-Testing, in that way I have shown the possibility of  rest of population analysis.  
  
**Were there any assumptions made that you felt were incorrect?**

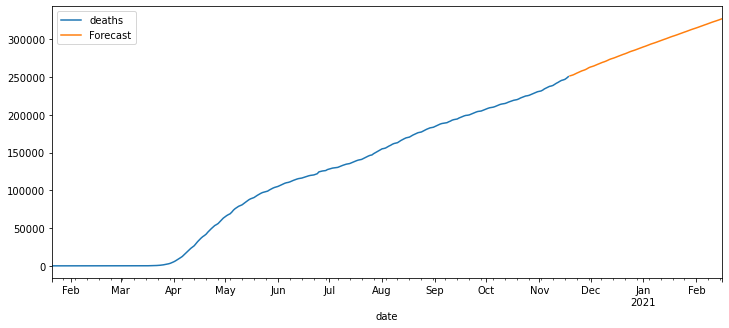
As part of model verification, I have performed the Prediction versus Actual comparison. As per Prediction, the death count in November is 220K but the actual count is 252K.

I have identified that my training dataset is up to May,2020. July onwards, we have seen a lot of deaths. Hence understanding that Training / Testing Dataset is very important.

Hence for future death count forecasting, I considered the data up to November,2020.







**What challenges did you face, what did you not fully understand?**

In this Project, I tried to utilize all the concepts which I learnt in this semester except Survival Analysis. ( both Hazard and Survival )

I thought of using Hazard Analysis to analyses based on the death event. I might need to focus on the analysis one more time to see both Hazard and Survival covered.

Project Output Charts as below

