CSC 605 Project Stage 4

This Document was edited and written by

Praneeth Aluru

Apoorva Gnana Saraswati Tangirala

Venkat Narayanan Reddy Cheedu

Nikhil Bolisetty

Akash Suresh

Tasks:

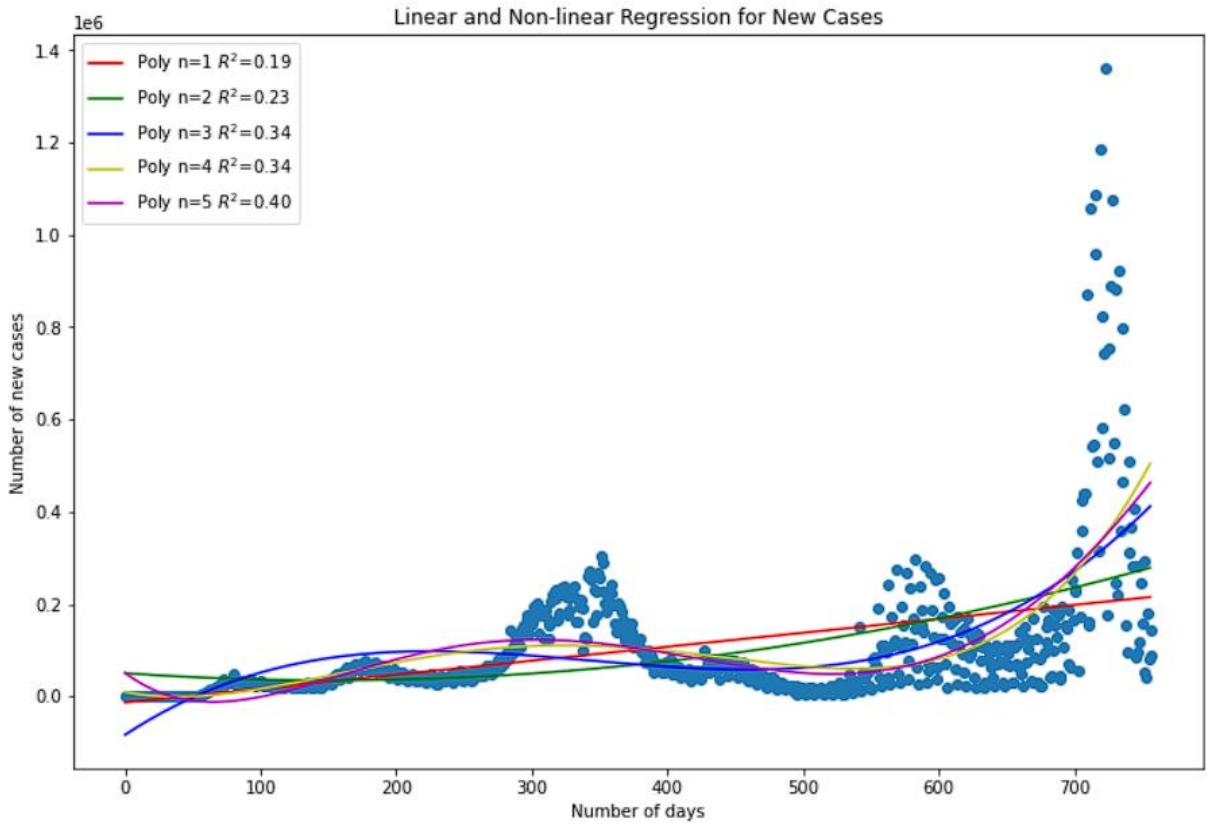
Task 1:

Develop Linear and Non-Linear (polynomial) regression models for predicting cases and deaths in US.

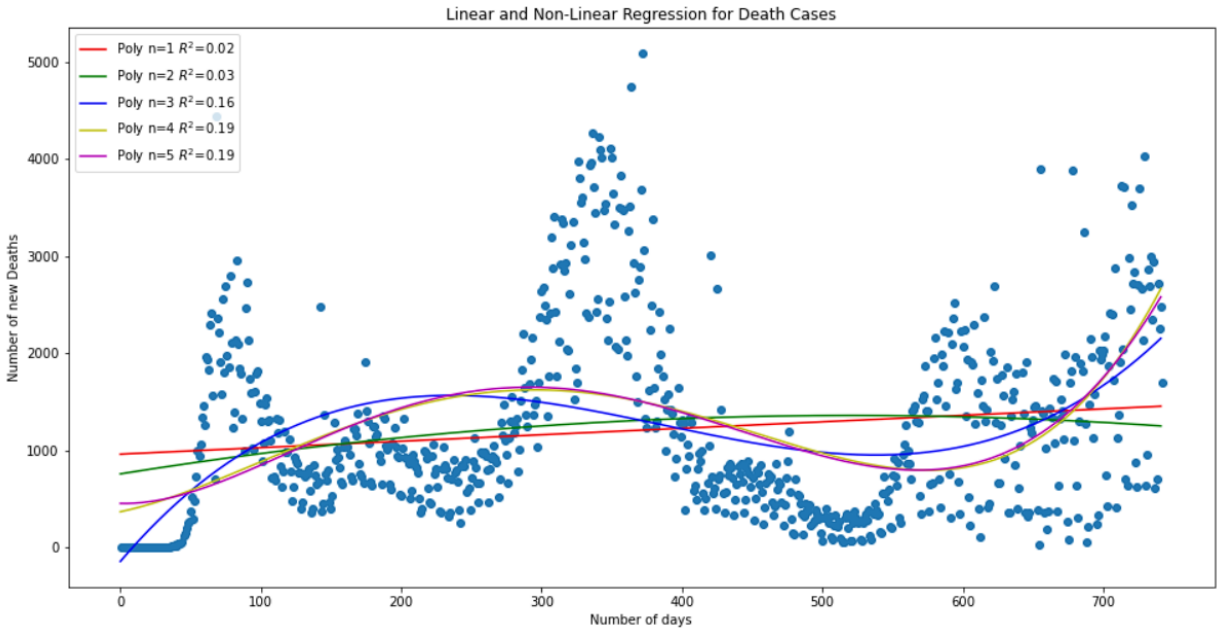
- Start your data from the first day of infections in US. X-Axis - number of days since the first case, Y-Axis - number of new cases and deaths.
- Calculate and report Root Mean Square Error (RMSE) for your models (linear and non-linear).
- Discuss bias versus variance tradeoff.
- Describe the trends as compared to other countries.

Linear and Non-linear Regression Models for predicting Cases and Deaths in US

Cases:



Deaths:



From the rsquared line(red) we generated above is definitely not a Linear model. There is some dependency between number of days and number of new cases. But, it is not linear. This is where Polynomial or non-linear modeling is useful to find a best fit regression line.

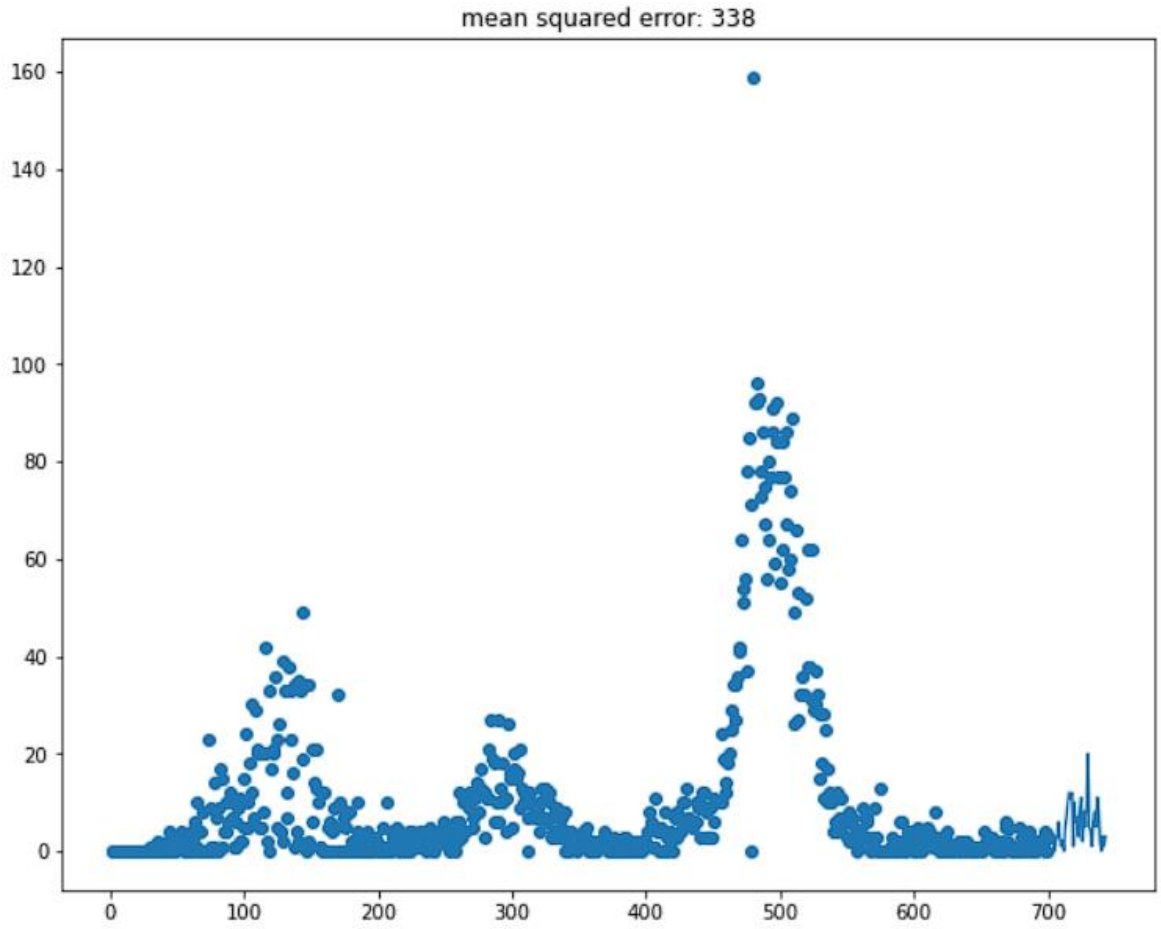
Bias VS Variance

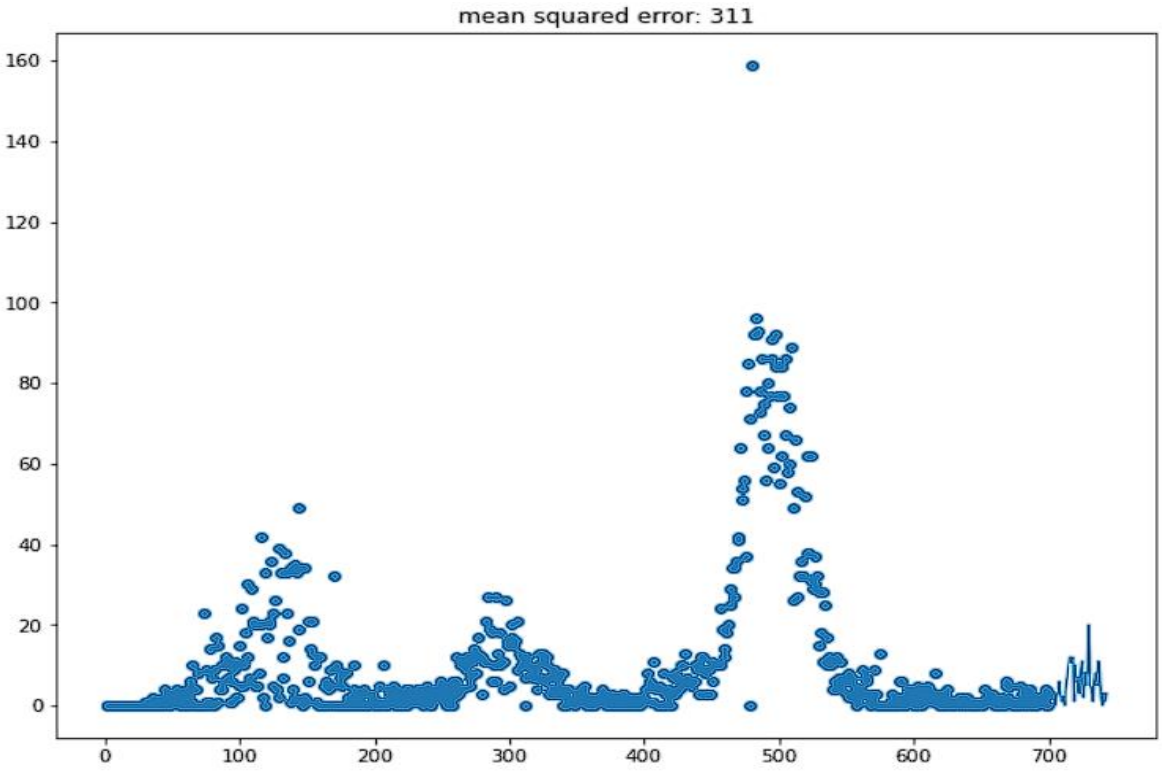
For the cases model, the polynomial regression model has more error than the linear regression model therefore using a high bias, low vairance model is better here.

For the deaths model, the polynomial regression model has more error than the linear regression model therefore using a high bias, low vairance model is better here.

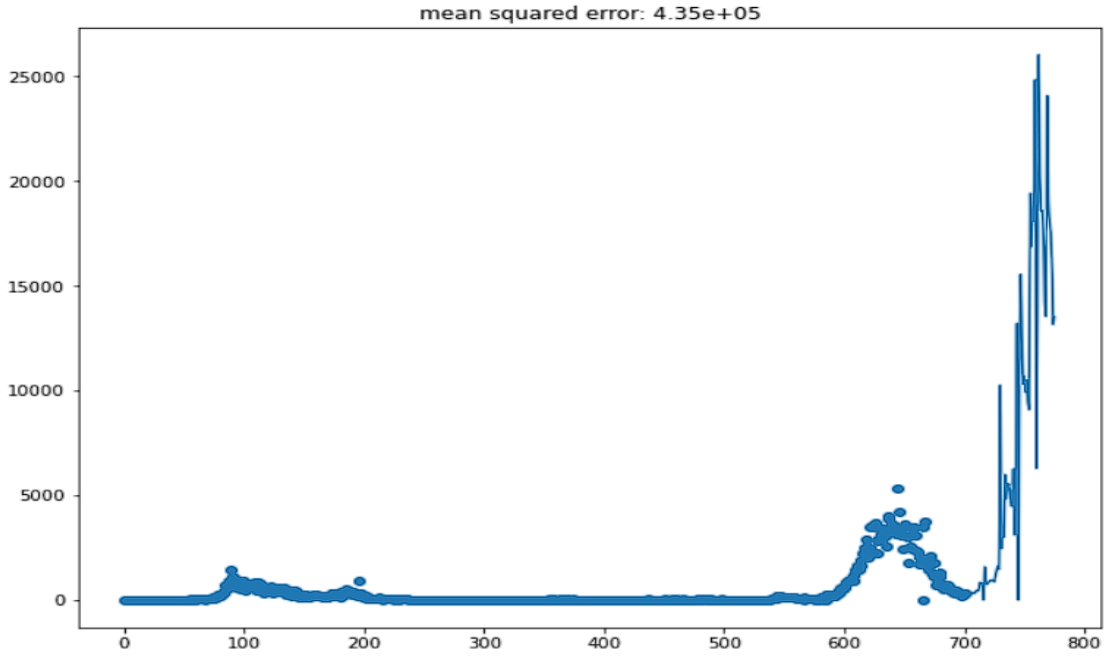
Trends of other countries :

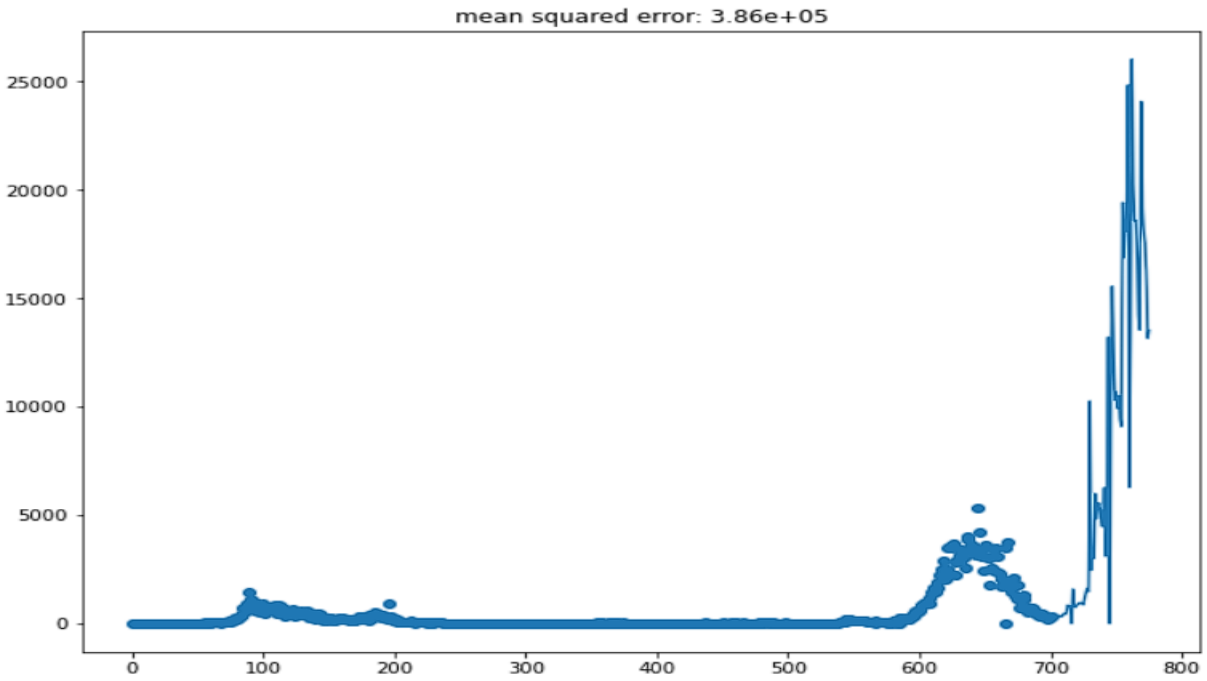
Afganisthan :



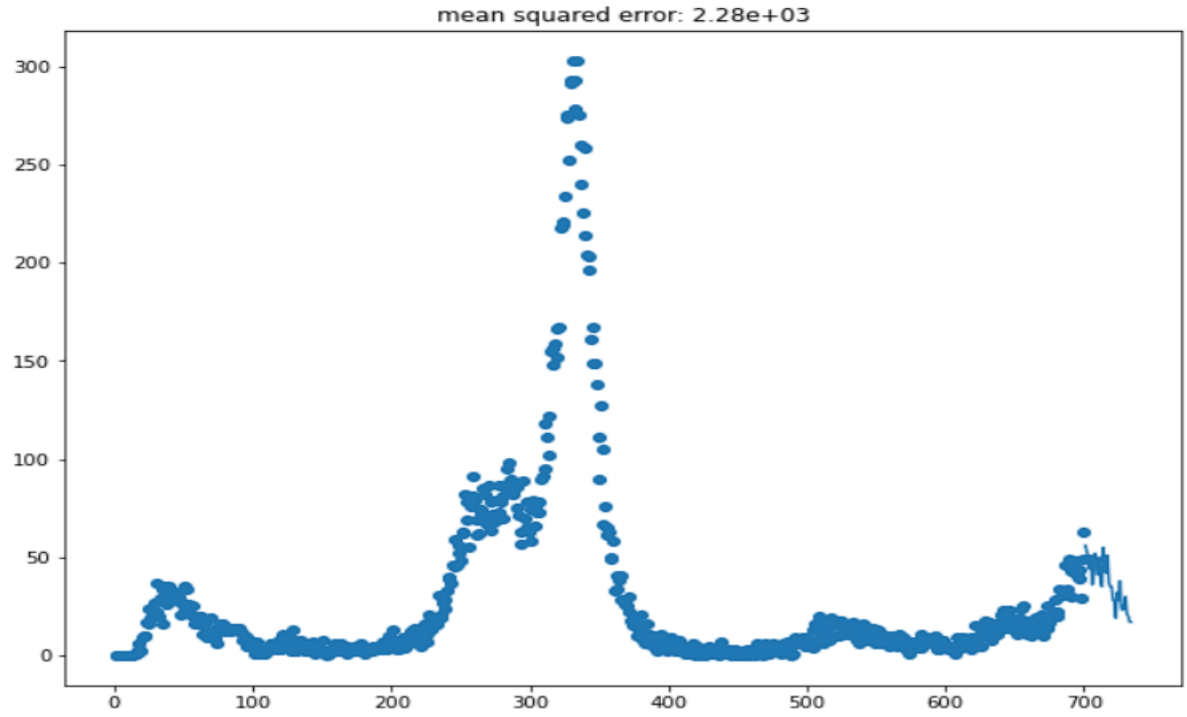

mean squared error: 311
160
140
120
100
80
60
40
20
0
0
100
200
300
400
500
600
700

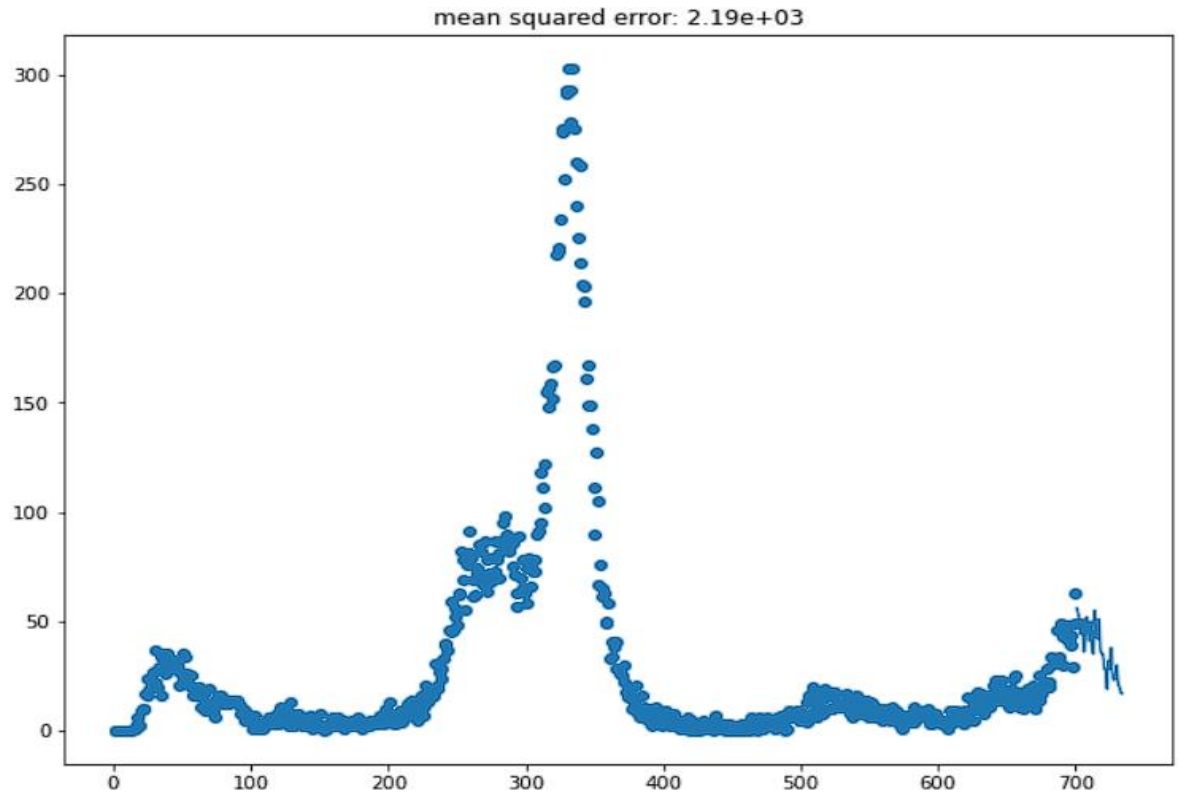
Singapore:


mean squared error: 4.35e+05
25000
20000
15000
10000
5000
0
0
100
200
300
400
500
600
700
800

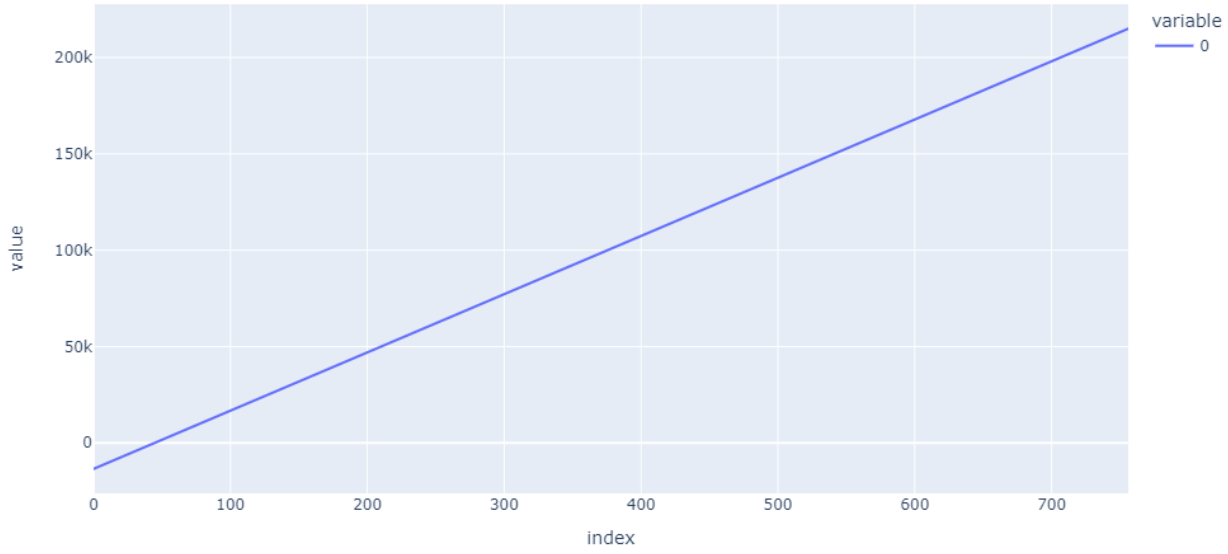

mean squared error: 3.86e+05
25000
20000
15000
10000
5000
0
0
100
200
300
400
500
600
700
800

Portugal:


mean squared error: 2.28e+03
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700



- Incorporate your best model prediction trend line - Linear / Non-Linear.



Best Prediction trend Line was USA

7 Day Rolling Average :