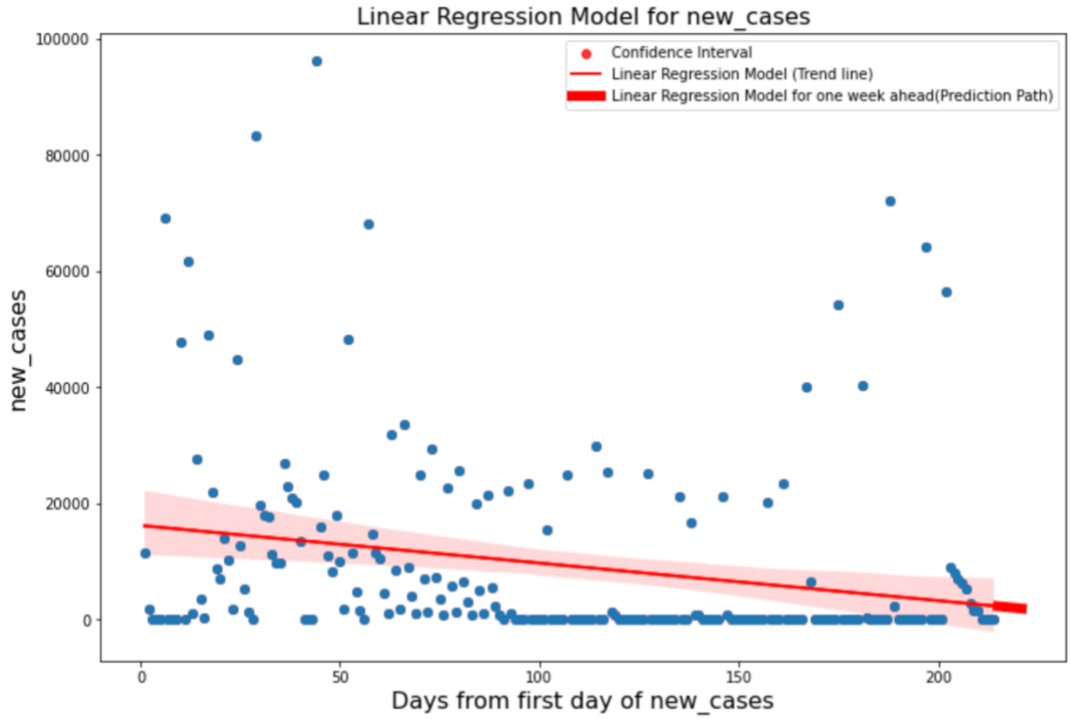
PROJECT_STAGE_4_REPORT

1) **Utilize Linear and Non-Linear (polynomial) regression models to compare trends for a single state (each member should choose different state) and its counties (top 5 with highest number of cases).**

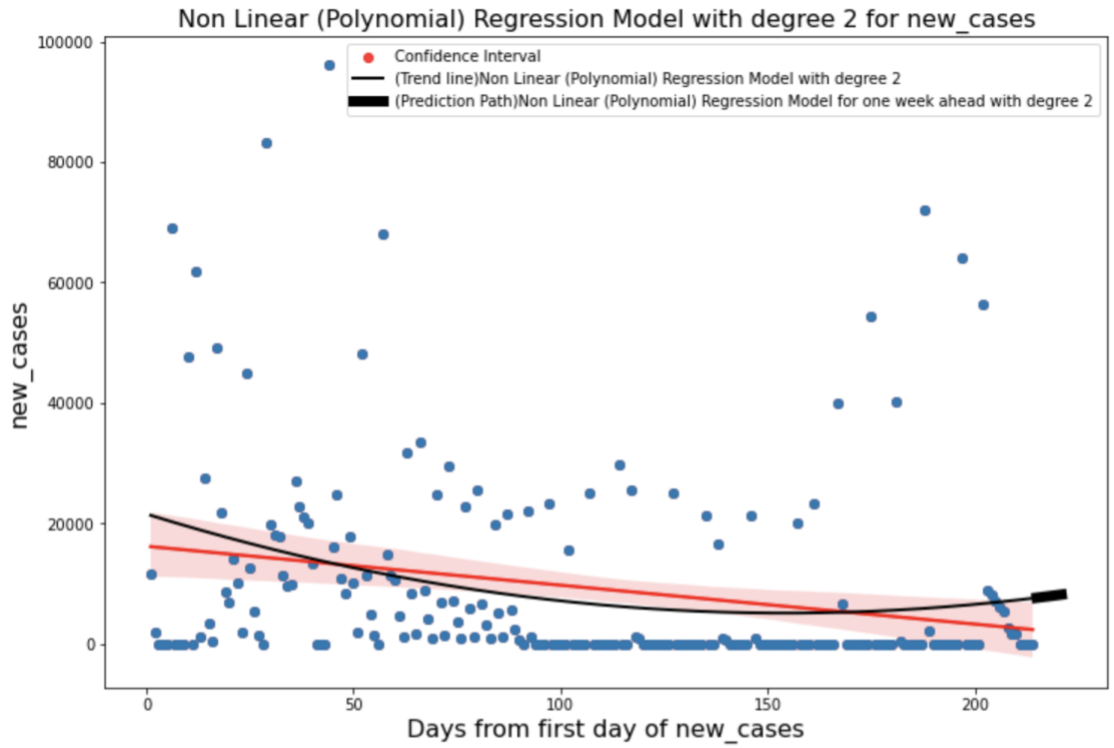
I have chosen California State.

Linear Regression Model for new cases across the California State



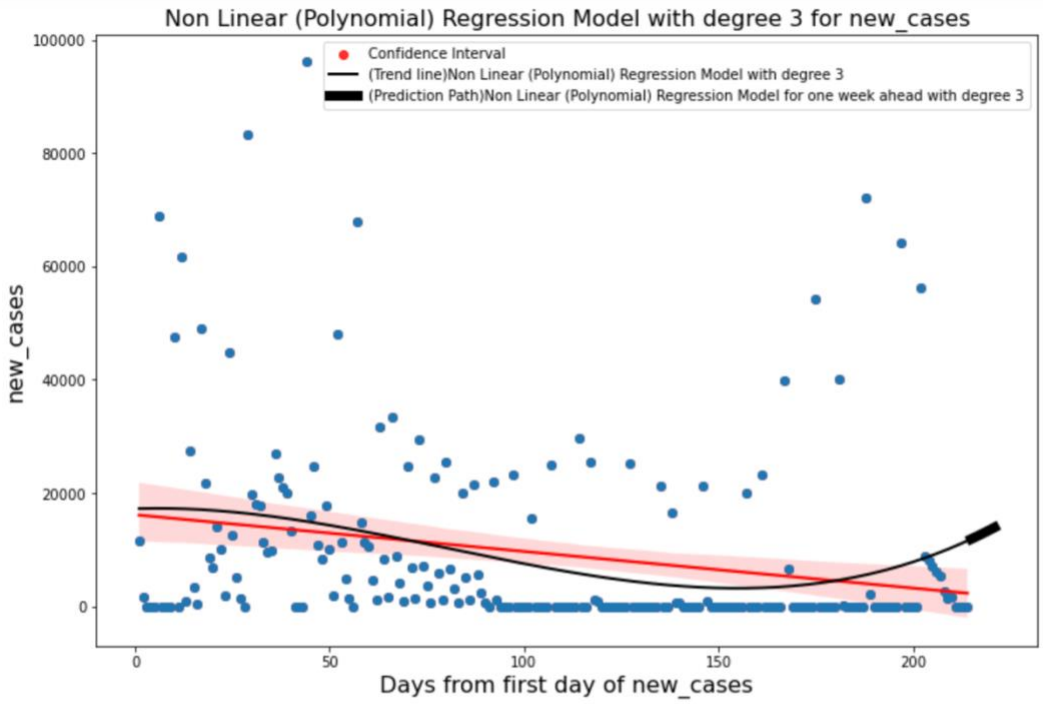

```
RMSE of Linear Regression on new_cases 16137.333352803971
```

Non-linear regression model with degree 2 for new cases across California State



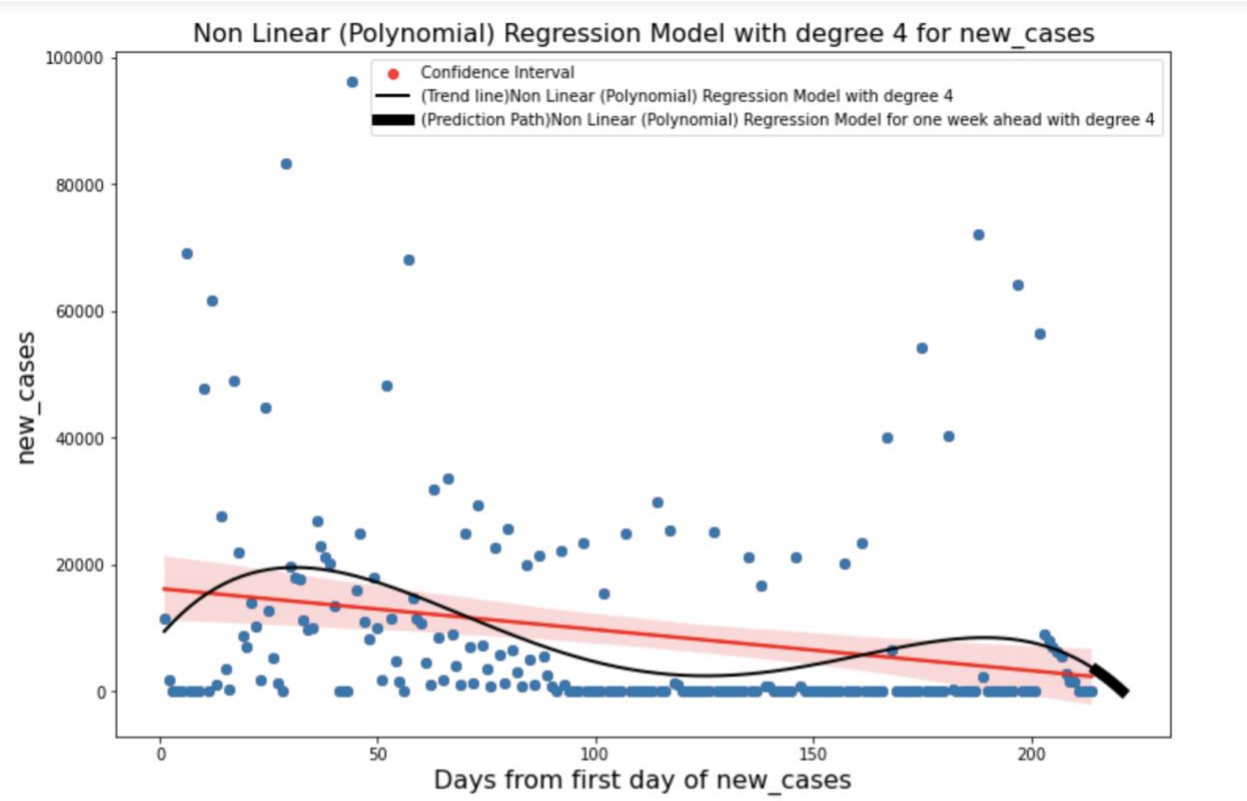

```
One week ahead forecast using Non-linear Regression Model with degree 2 [[7626.5207269 ]
 [7711.12460406]
 [7797.11051336]
 [7884.4784548 ]
 [7973.22842838]
 [8063.3604341 ]
 [8154.87447197]]
RMSE of Non Linear (Polynomial) Regression with degree 2 on new_cases 15964.037613310162
```

Non-linear regression model with degree 3 for new cases across California State



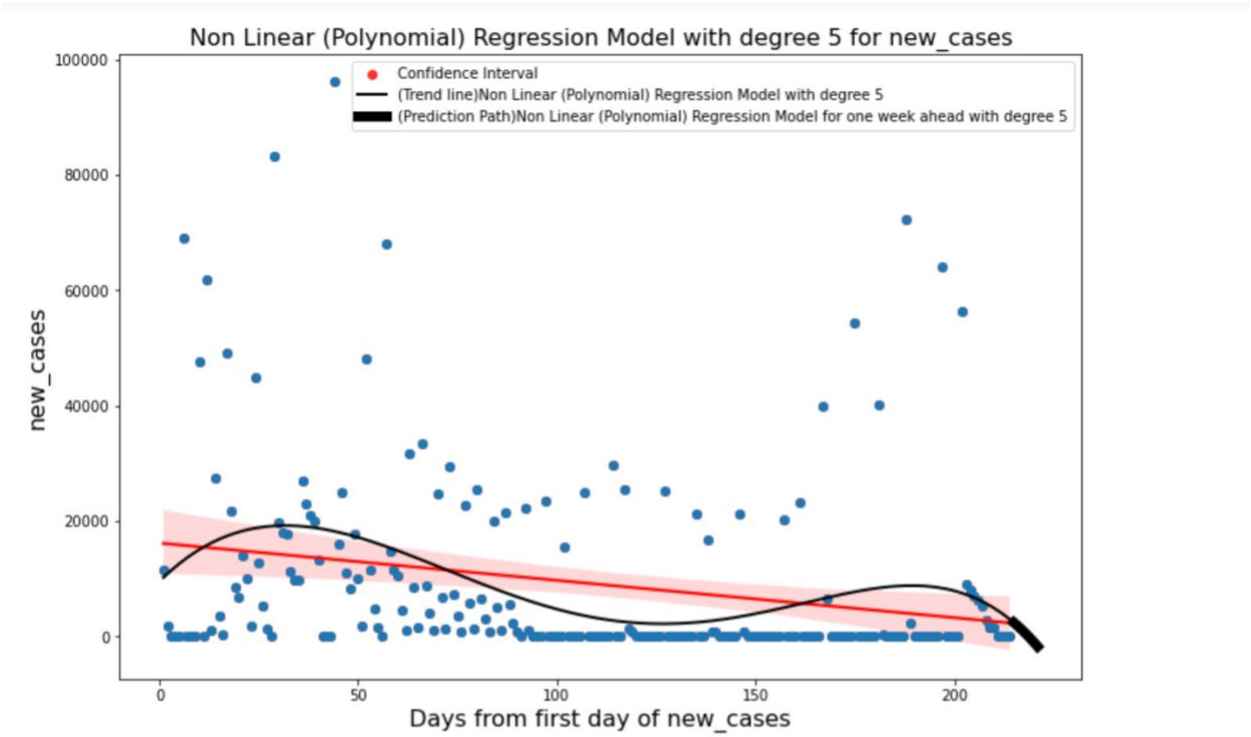

```
One week ahead forecast using Non-linear Regression Model with degree 3 [[11902.44464186]
 [12225.70473752]
 [12555.87131483]
 [12892.99529037]
 [13237.12758073]
 [13588.31910249]
 [13946.62077223]]
RMSE of Non Linear (Polynomial) Regression with degree 3 on new_cases 15886.503816265138
```

Non-linear regression model with degree 4 for new cases across California State



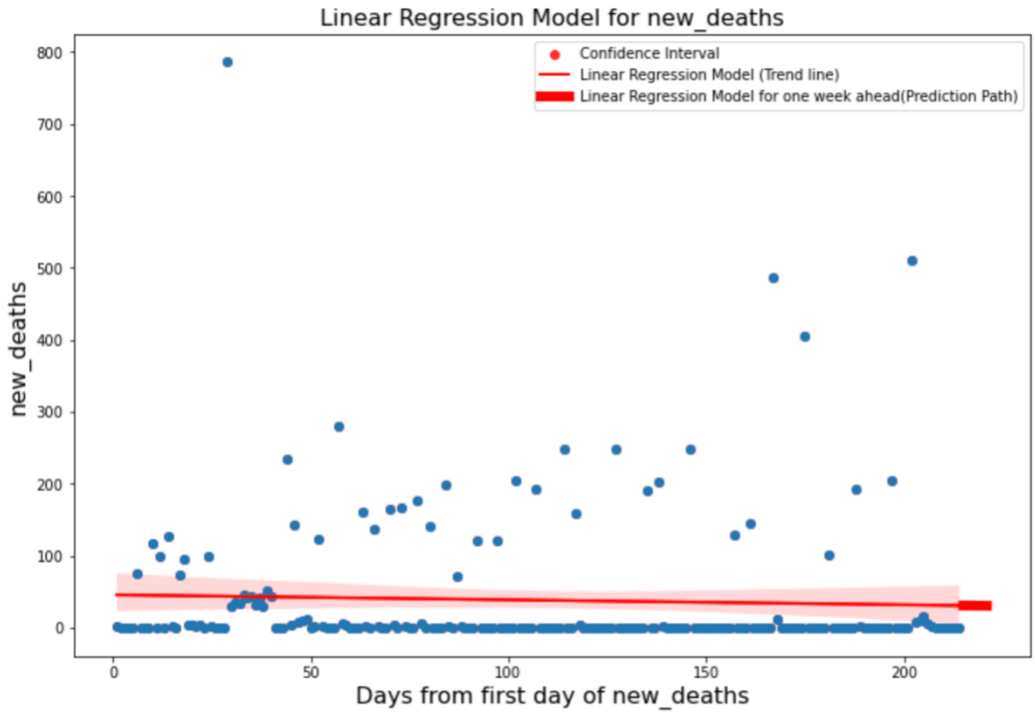

```
One week ahead forecast using Non-linear Regression Model with degree 4 [[3293.4337172 ]
 [2815.85558667]
 [2311.81567834]
 [1780.64731181]
 [1221.67722323]
 [ 634.22556529]
 [  17.60590722]]
RMSE of Non Linear (Polynomial) Regression with degree 4 on new_cases 15648.667818651118
```

Non-linear regression model with degree 5 for new cases across California State



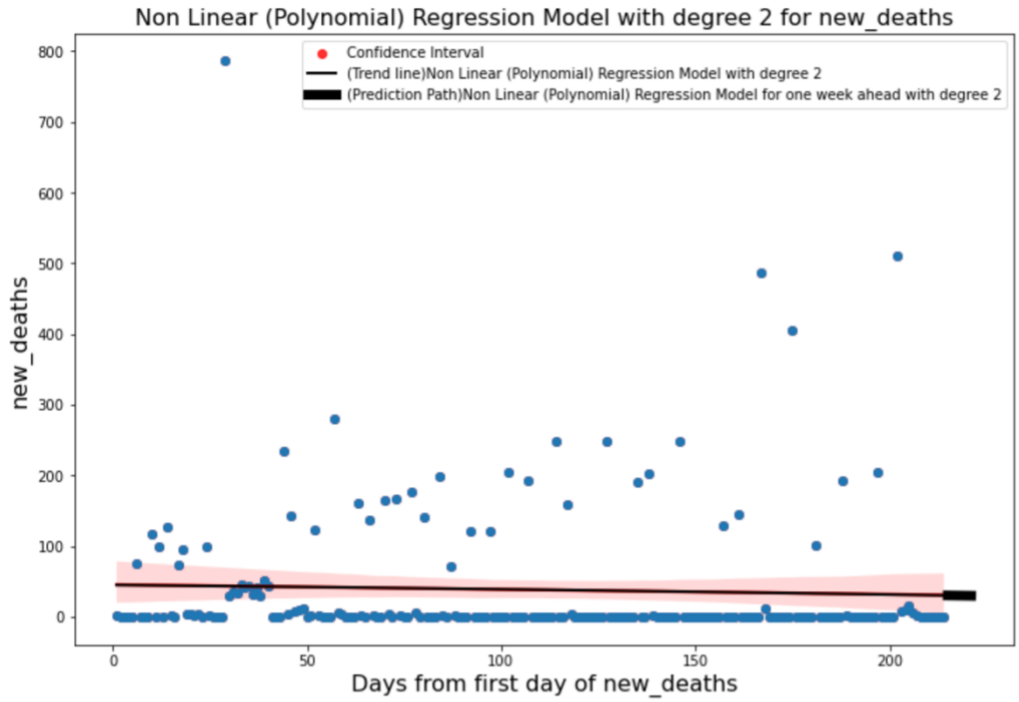

```
One week ahead forecast using Non-linear Regression Model with degree 5 [[ 2404.44410639]
 [ 1802.82387422]
 [ 1166.70213543]
 [  495.11581442]
 [ -212.91086612]
 [ -958.36644108]
 [-1742.25225881]]
RMSE of Non Linear (Polynomial) Regression with degree 5 on new_cases 15646.672533720646
```

Linear Regression Model for new deaths across the California State



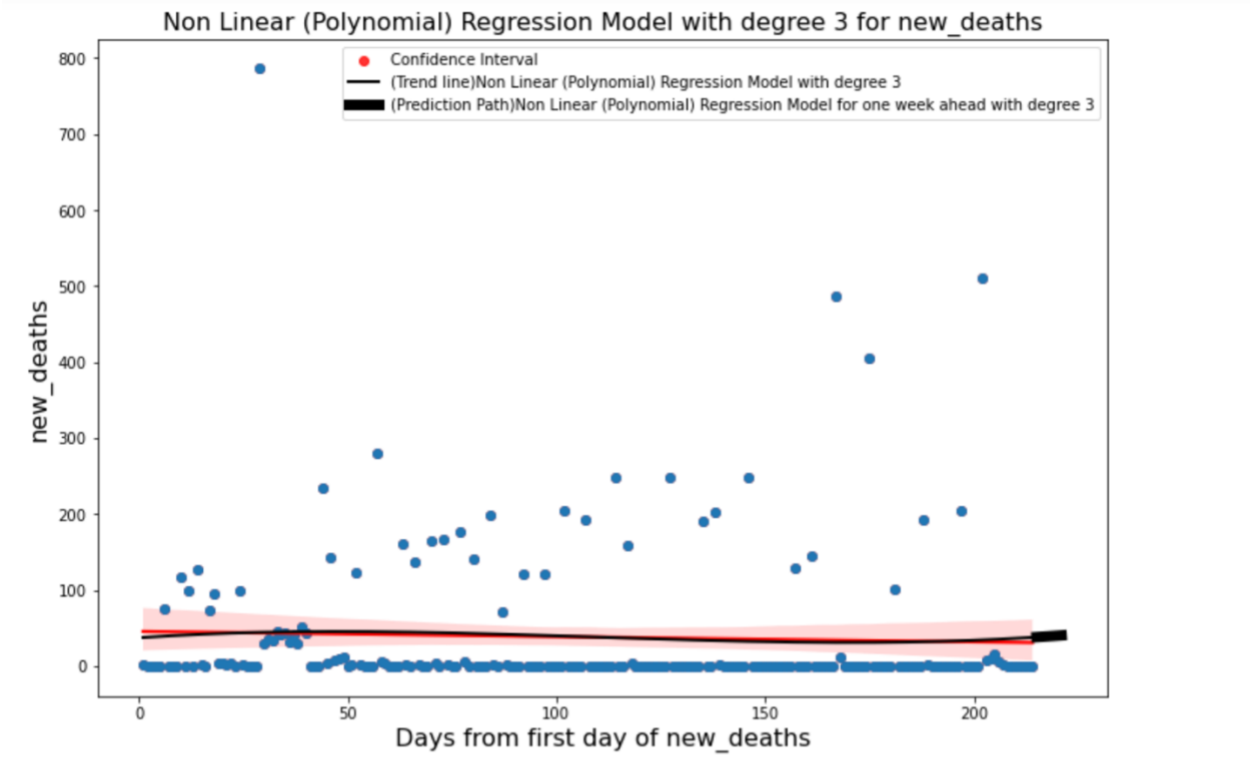

```
RMSE of Linear Regression on new_deaths 95.21051925043596
```

Non-linear regression model with degree 2 for new deaths across California State



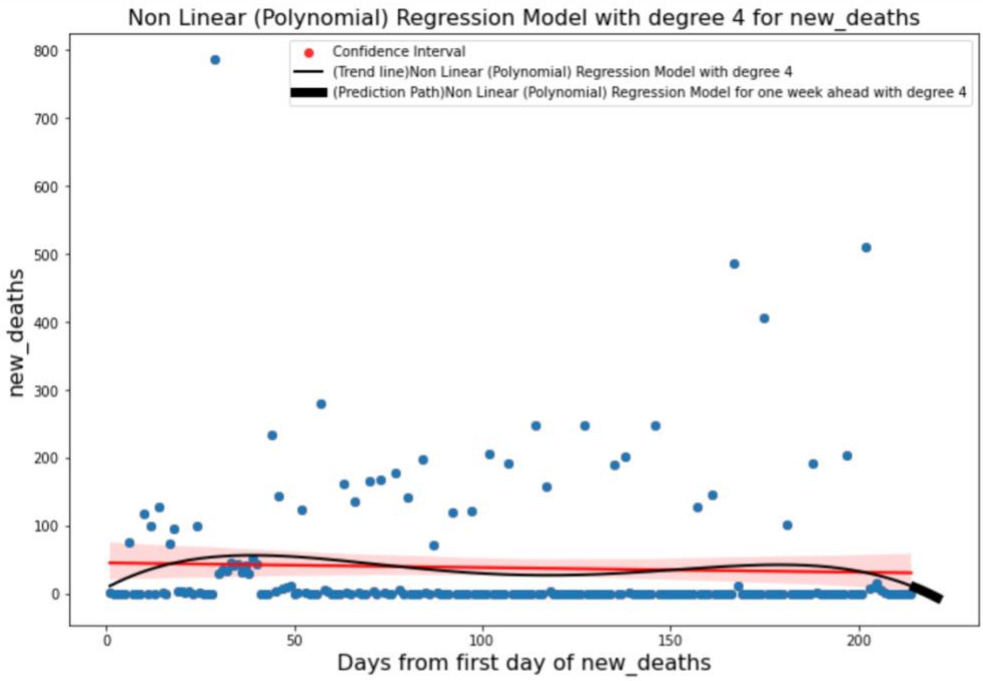

```
One week ahead forecast using Non-linear Regression Model with degree 2 [[30.03442521]
 [29.94987296]
 [29.86517734]
 [29.78033833]
 [29.69535594]
 [29.61023017]
 [29.52496102]]
RMSE of Non Linear (Polynomial) Regression with degree 2 on new_deaths 95.21020480366471
```

Non-linear regression model with degree 3 for new deaths across California State



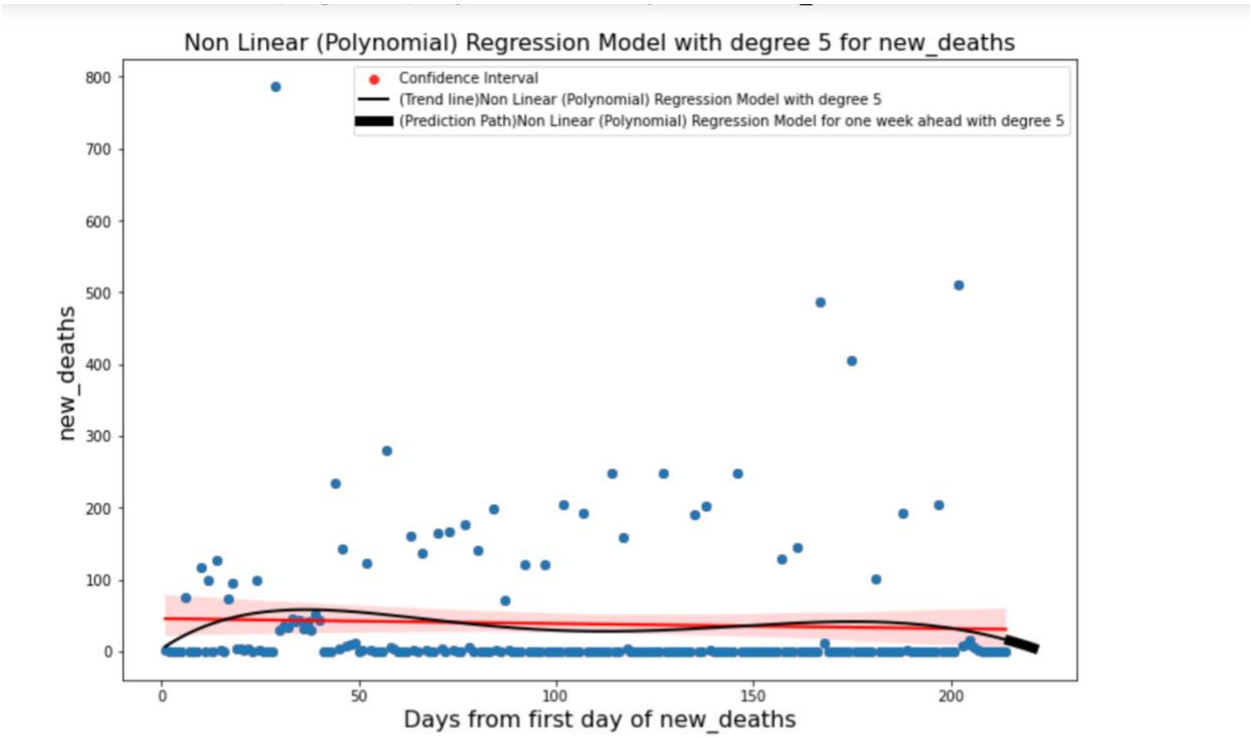

```
One week ahead forecast using Non-linear Regression Model with degree 3 [[37.89312386]
 [38.24719666]
 [38.61127943]
 [38.98546576]
 [39.36984922]
 [39.7645234 ]
 [40.16958187]]
RMSE of Non Linear (Polynomial) Regression with degree 3 on new_deaths 95.166388508328
```

Non-linear regression model with degree 4 for new deaths across California State




```
One week ahead forecast using Non-linear Regression Model with degree 4 [[ 9.35042414]
 [ 7.04936213]
 [ 4.64767947]
 [ 2.1430906 ]
 [-0.46671188]
 [-3.18405721]
 [-6.01129643]]
RMSE of Non Linear (Polynomial) Regression with degree 4 on new_deaths 94.73224380387285
```

Non-linear regression model with degree 5 for new deaths across California State




```
One week ahead forecast using Non-linear Regression Model with degree 5 [[14.66371375]
 [13.1040503 ]
 [11.49181993]
 [ 9.82650865]
 [ 8.1076172 ]
 [ 6.33466138]
 [ 4.50717241]]
RMSE of Non Linear (Polynomial) Regression with degree 5 on new_deaths 94.72046886108102
```

2) Identify which counties are most at risk. Model for top 5 counties with cases within a state and describe their trends.

Finding the top 5 infected counties across the CA State

```
selected_state_county_data_daily2 = selected_state_county_data_daily1.sort_values(['new_cases'],ascending = False)
selected_state_county_data_daily2.head(5)
```

	countyFIPS	County Name	StateFIPS	population	cases	deaths	new_cases	new_deaths
18	6037	Los Angeles County	6	10039107	685420148	7083040	576726.0	2461.0
36	6073	San Diego County	6	3338330	193002028	1171012	169096.0	338.0
29	6059	Orange County	6	3175692	139335973	1576808	120667.0	618.0
35	6071	San Bernardino County	6	2180085	139636983	1707439	114782.0	492.0
42	6085	Santa Clara County	6	1927852	89858008	523756	111182.0	267.0

Linear Regression Model trend lines and non-linear regression model trend lines for new cases and new deaths across these top 5 counties are modeled in ipynb file.

3) Perform hypothesis tests on questions identified in Stage II

- Does the increase in white population increases in the spread of covid-19 cases?
- Covid cases are more in the age group of 65 years and above.
- Covid cases spread is more in the Male population.
- Covid cases are more among the Asian population.

Applied two tail two sample t test and one tail two sample t test on each of these hypothesis questions in the ipynb file.