

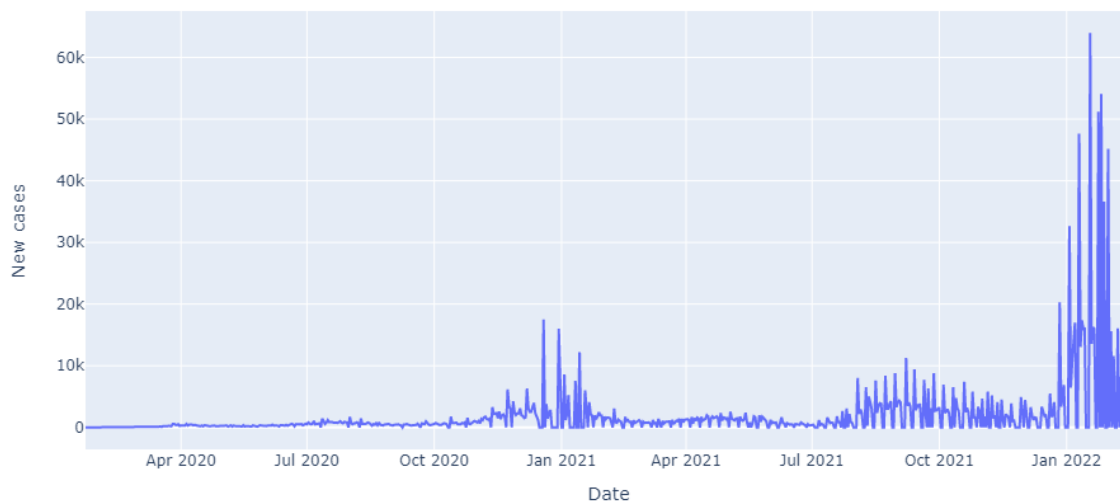
Task 3

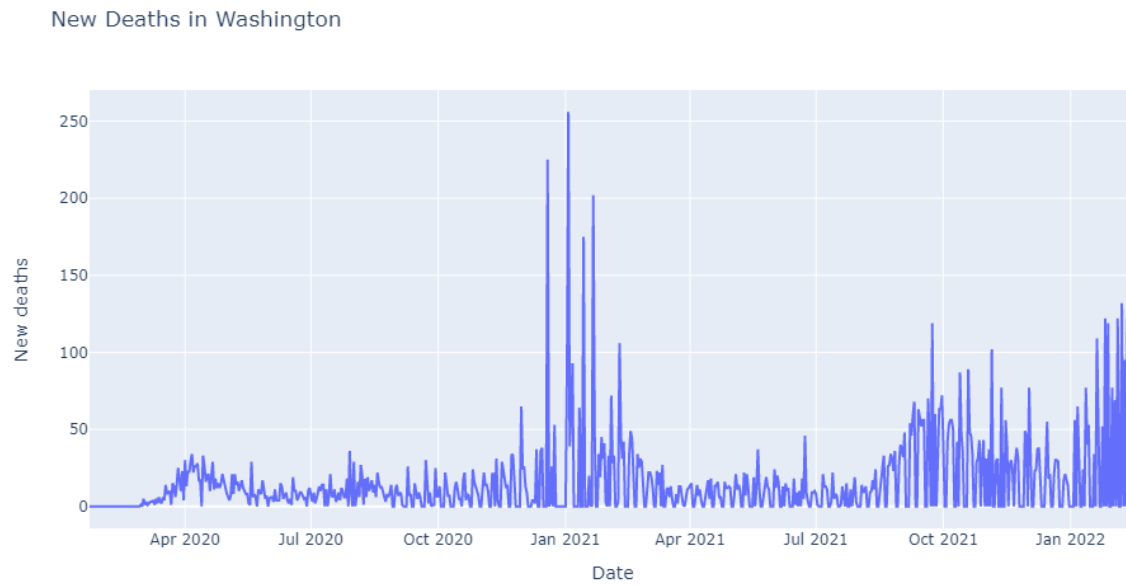
As for Task 3 we were required to complete the following

- To utilize Linear and Non-Linear (polynomial) regression models to compare trends for a single state and its counties (top 5 with highest number of cases). Start your data from the first day of infections.
 - X-Axis - number of days since the first case, Y - Axis number of new cases and deaths. Calculate error using RMSE.
 - Identify which counties are most at risk. Model for top 5 counties with cases within a state and describe their trends.
 - Perform hypothesis tests on questions identified in Stage II. (Can determine new questions as well)
 - e.x. *Does higher employment data (overall employment numbers) lead to higher covid case numbers or more rapid increase in covid cases.* Here you would compare the covid cases to the state or county level enrichment data to prove or disprove your null hypothesis. In this case there will be a two tail - two sample t-test to see if there is a difference and then one-tail - two sample t-test to show higher or lower.
 - Depending on your type of data you can also perform Chi-square test for categorical hypothesis testing.

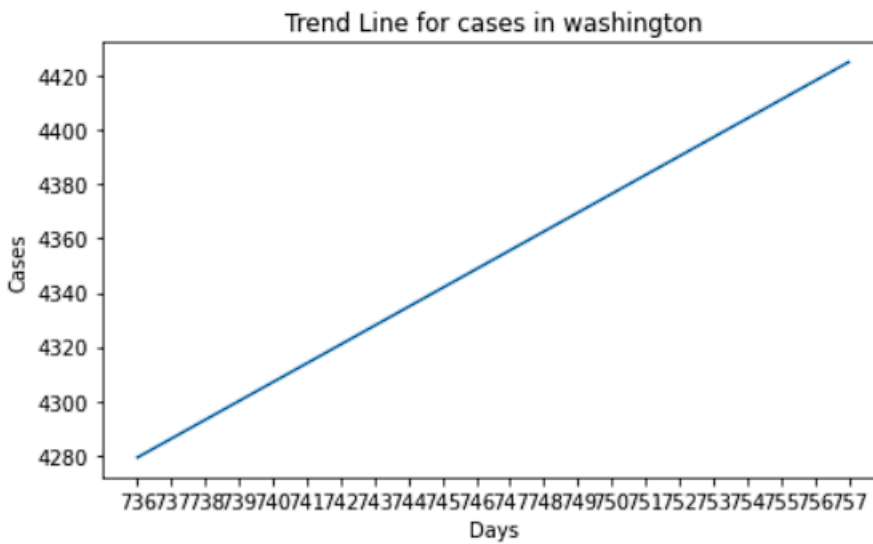
Washington

New Cases in Washington

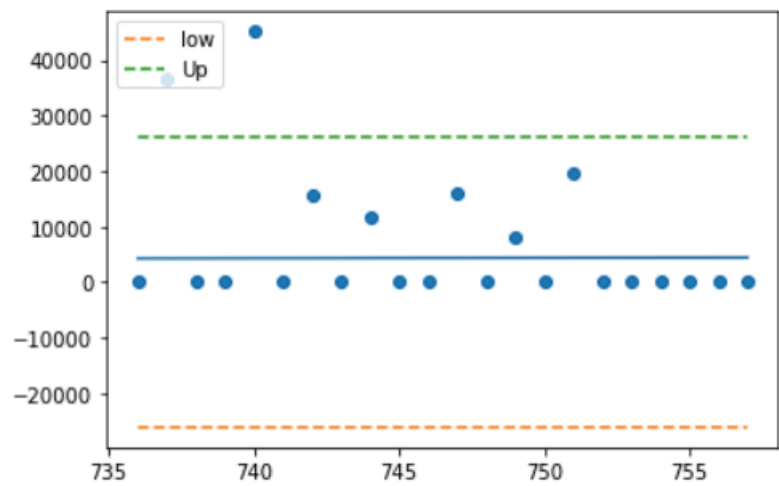




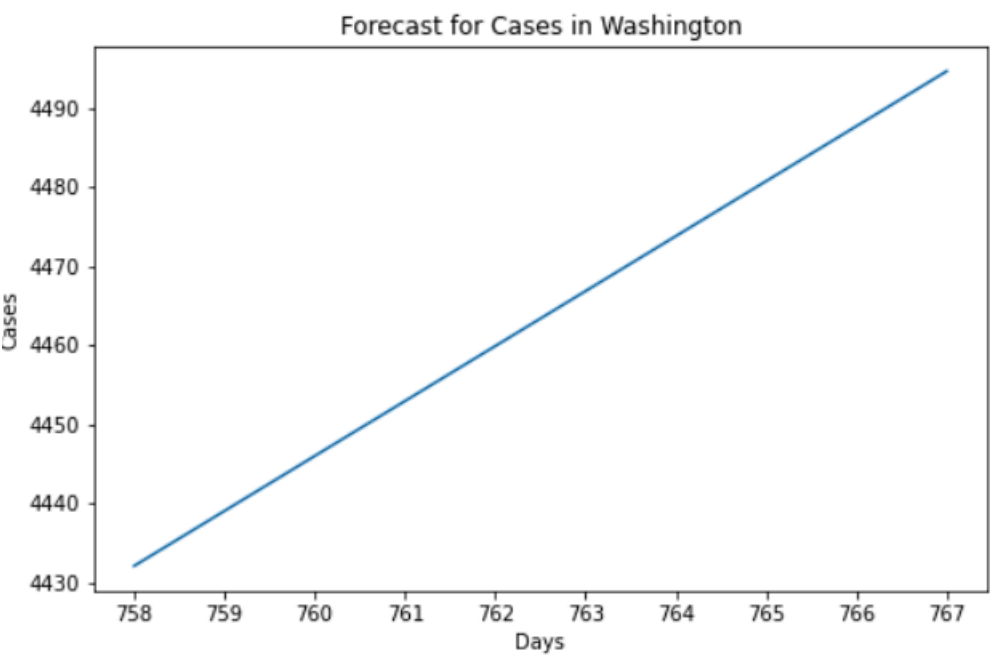
Linear Regression RMSE and Trends



Confidence Intervals

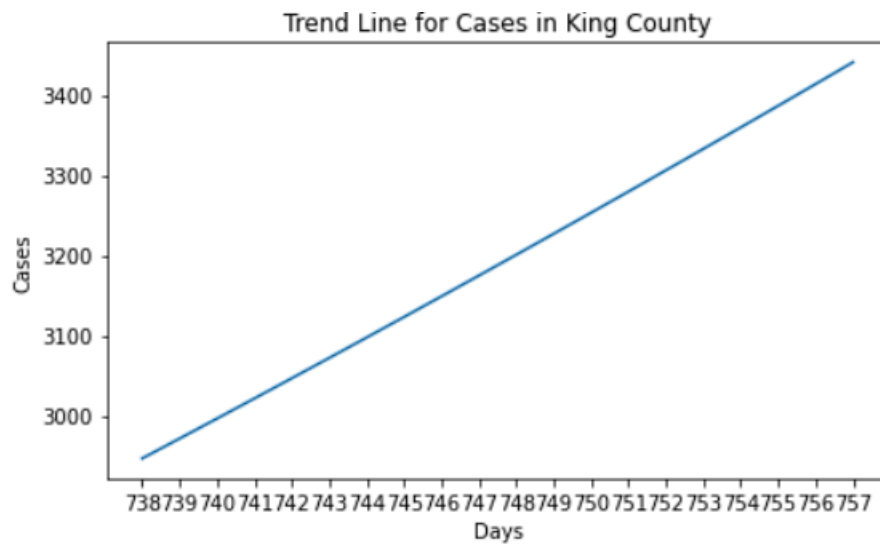


Forecast for Washington

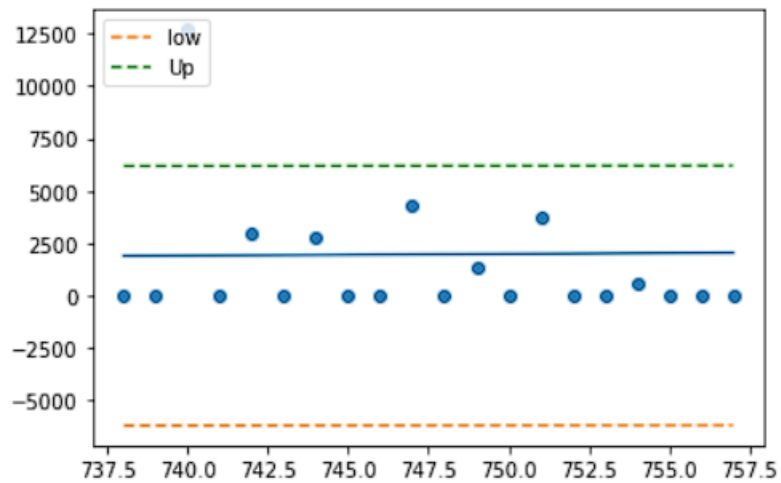


Top 5 Counties in Washington with High Count of cases aand death

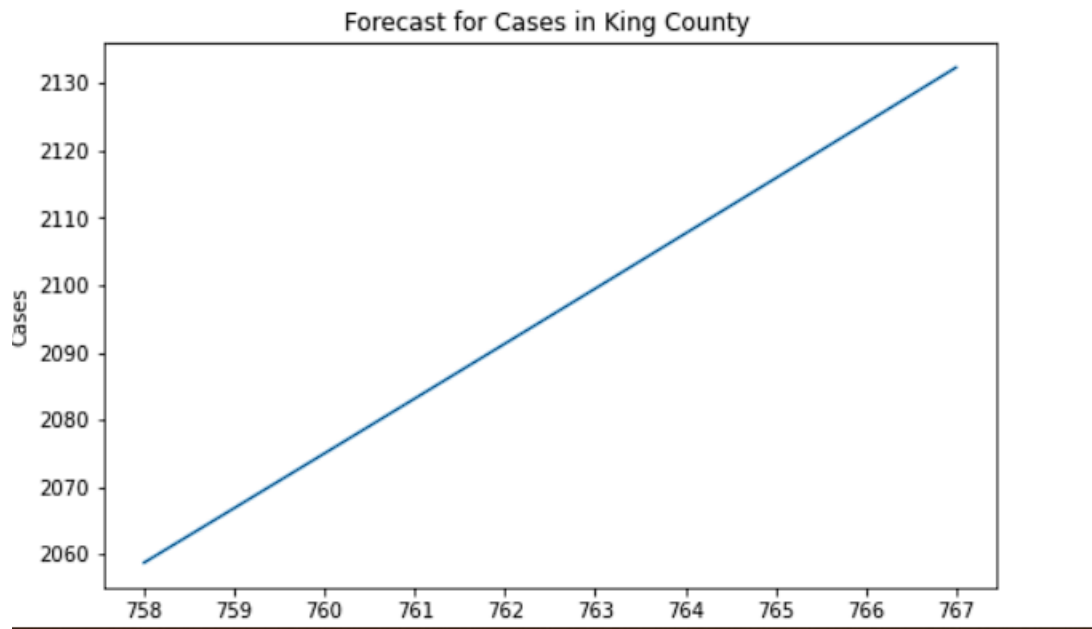
King County



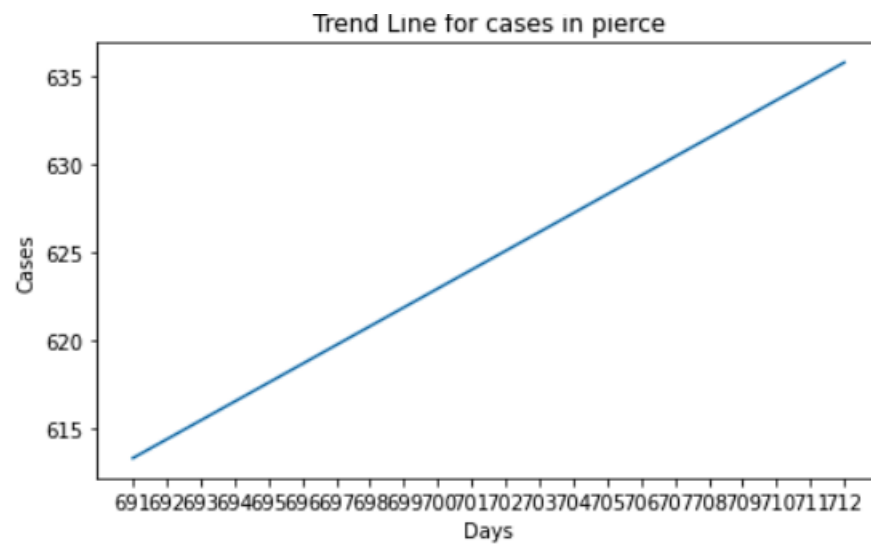
Confidence Interval



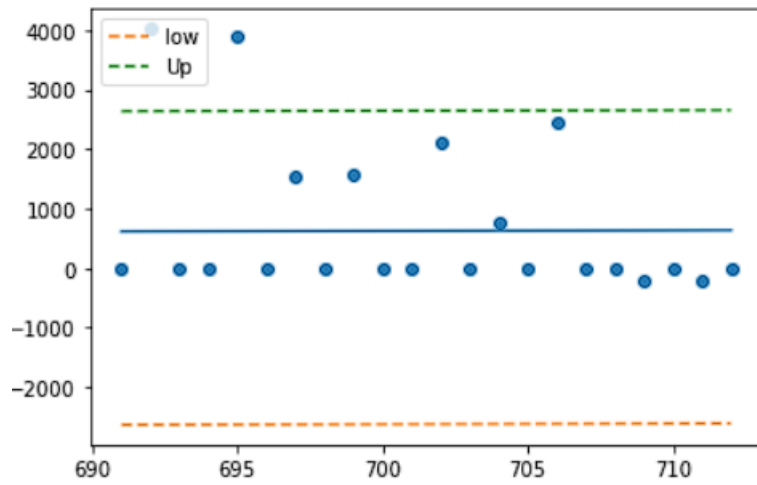
ForeCast



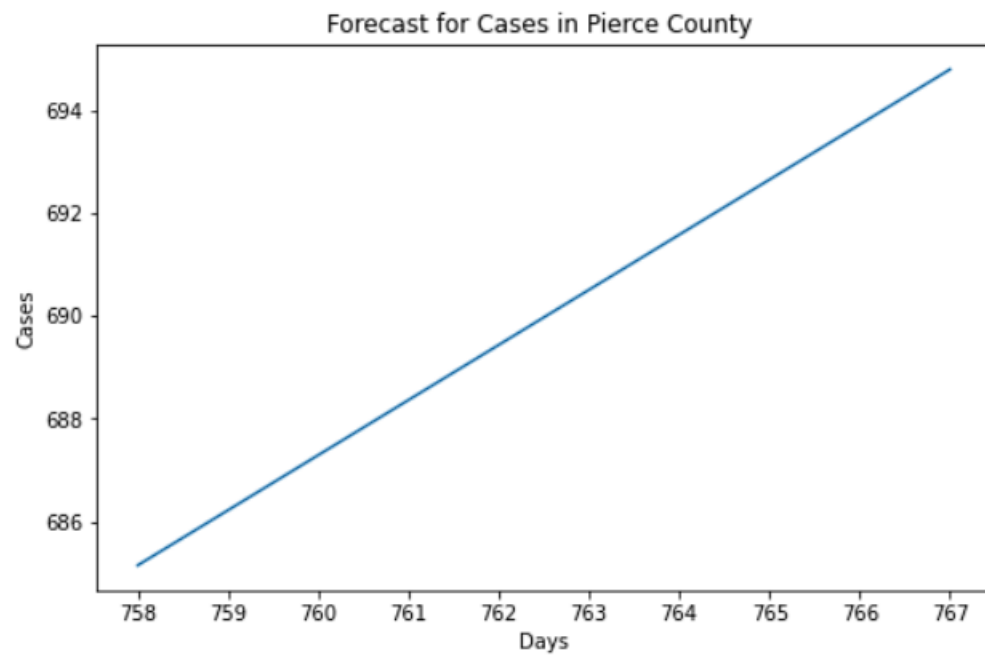
Pierce County



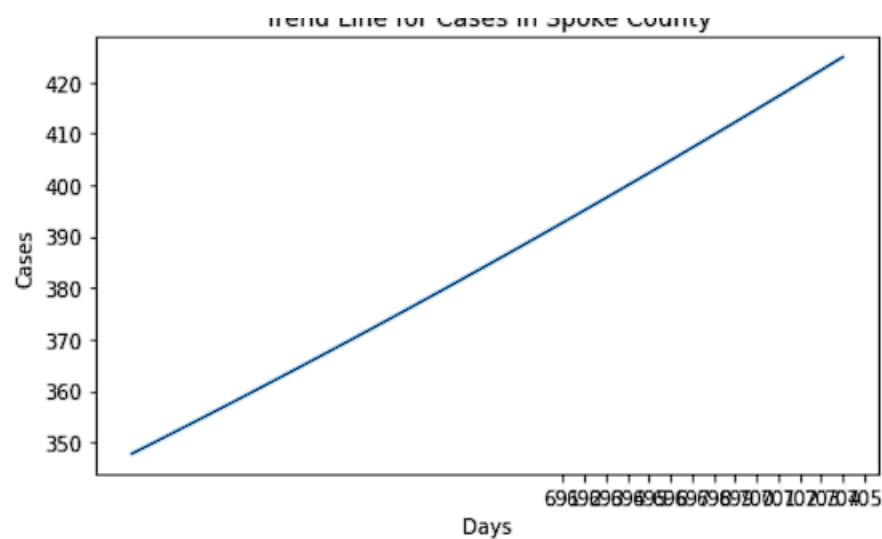
Confidence Intervals



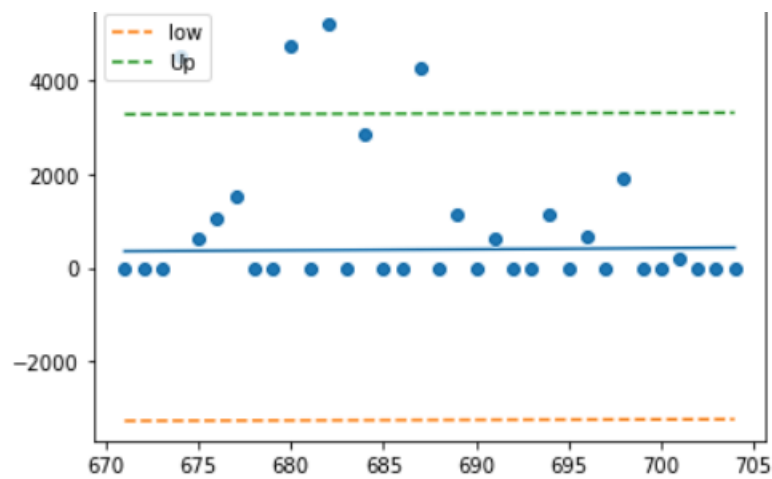
Forecast



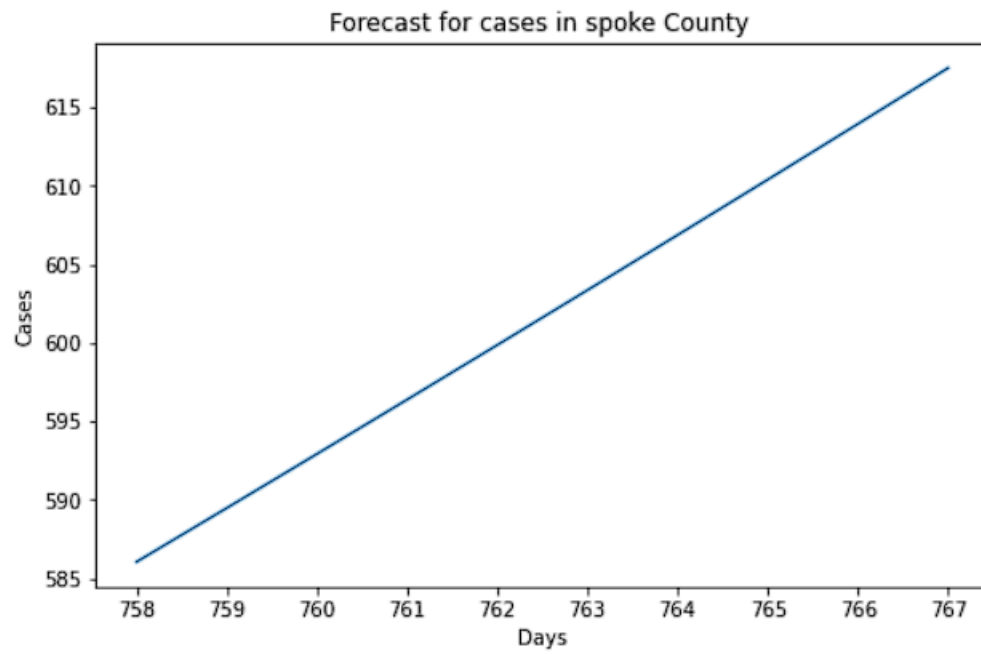
Spoke County



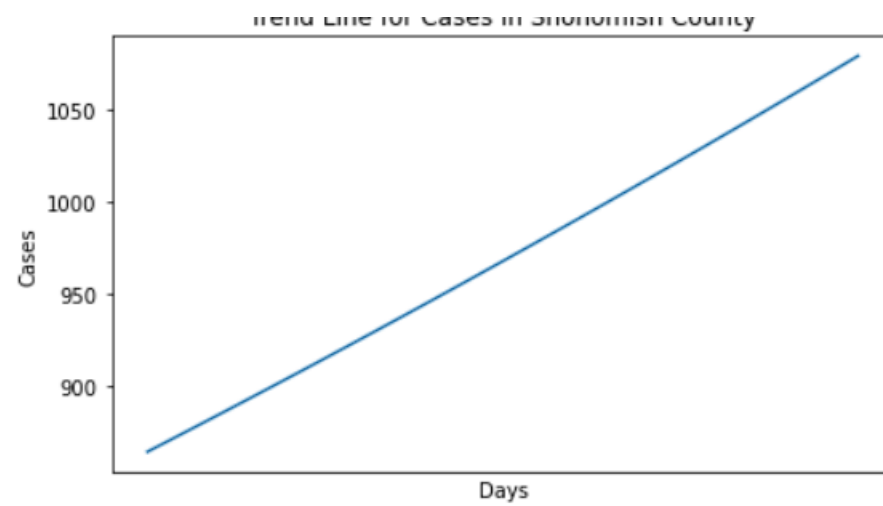
Confidence intervals



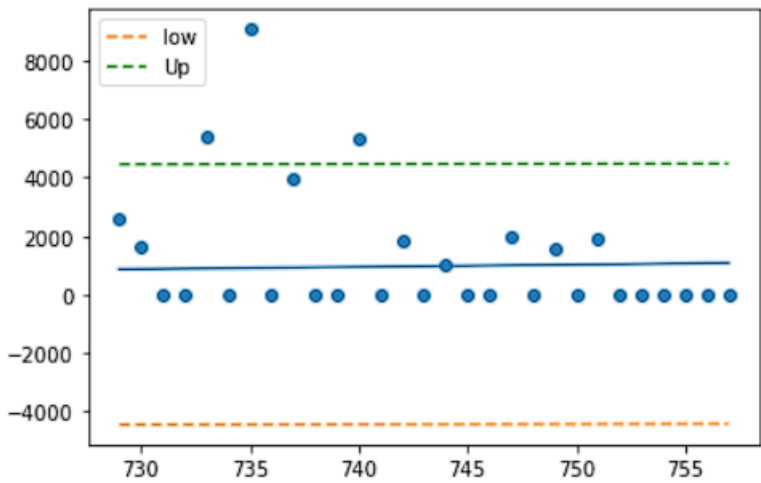
Forecast



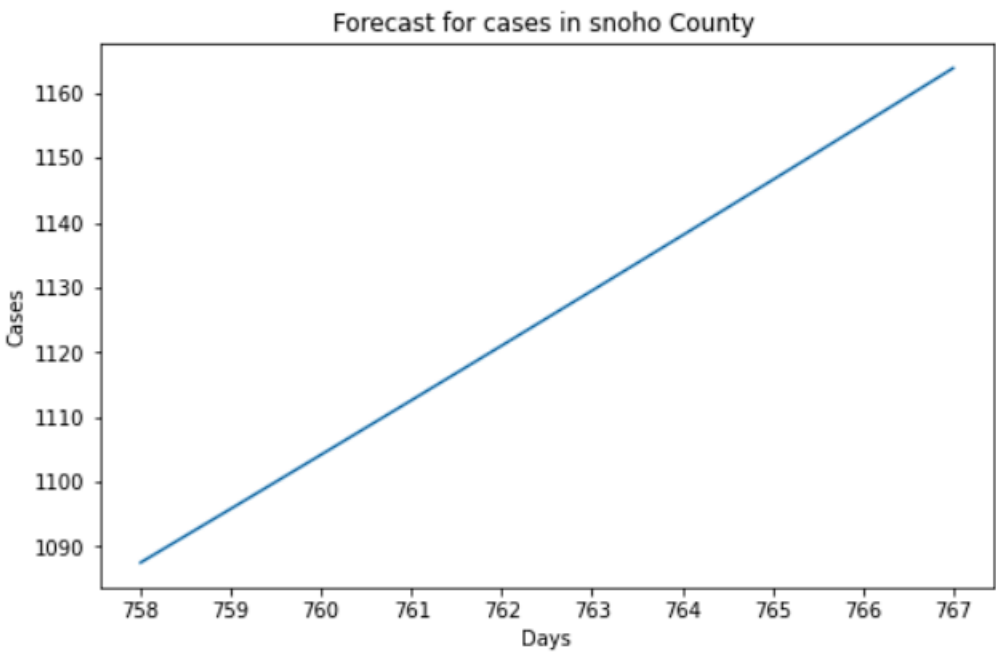
Snohomish County



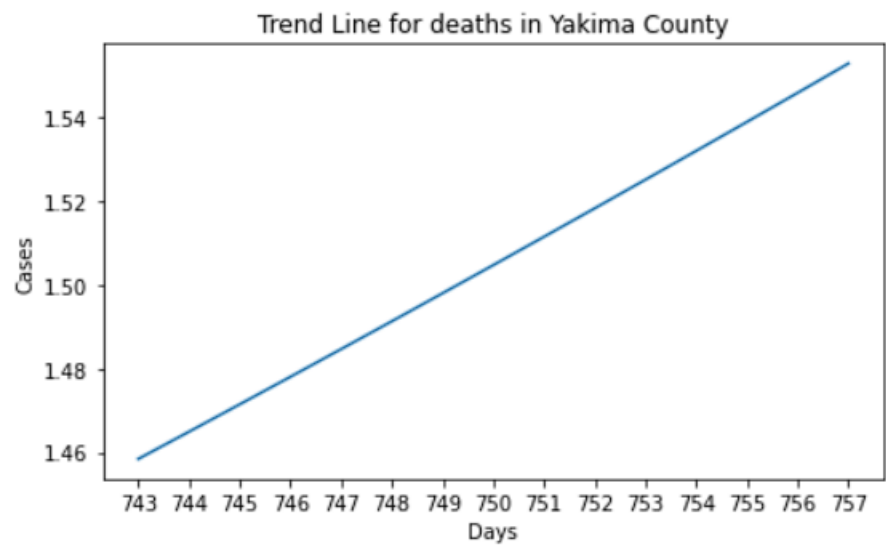
Confidence Intervals



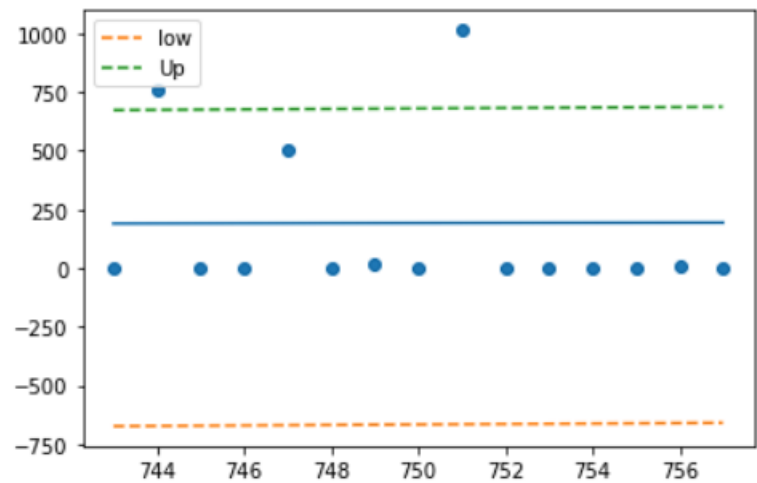
Forecast



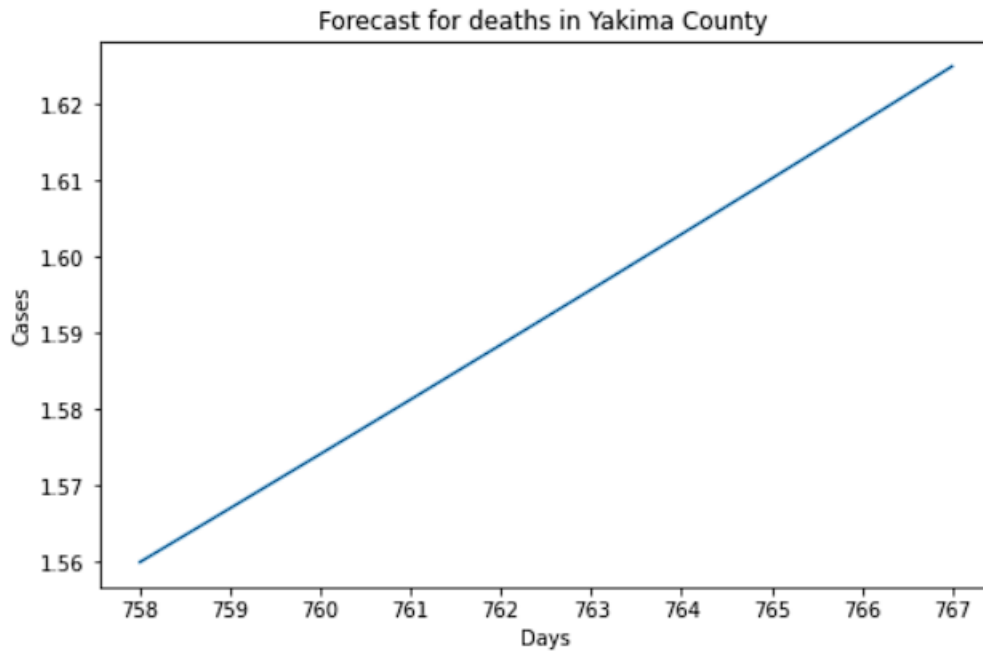
Yakima County



Confidence Intervals



Forecast



Hypothesis

1. Does the number of votes depend on the Cases of each County?
2. The States with Higher populations have higher Voting Rates

1. Does the number of votes depend on the Cases of each County

```
] res = stats.ttest_ind(merged['Cases'], merged['current_votes'], equal_var = False)
display(res)
```

```
Ttest_indResult(statistic=-495.2759049474959, pvalue=0.0)
```

Here the P value is astronomically less hence the T test returns the P value as 0 which is less than 0.05 which we can assume The hypothesis has failed

2. The States with Higher populations have higher Voting Rates

```
] res = stats.ttest_ind(merged['total_votes'], merged['population'], equal_var = False)
display(res)
```

```
Ttest_indResult(statistic=-531.7188701618219, pvalue=0.0)
```

the T test returns the P value as 0 which is less than 0.05 which we can assume The hypothesis has failed

3.