COVID-19 COMBATANT UNIFICATION COMPETITION

Coronavirus (COVID-19) Visualization & Prediction in South Asian Association for Regional Cooperation (SAARC) Countries

10.5.20

Adittya Soukarjya Saha

17101148

Mahmudul Haque

17101394

Ramisa Sadeque Khan

16241004

Md. Shahriar Rahman Rana

18201195

Abstract

An outbreak of coronavirus disease 2019 (COVID-19) caused by the novel coronavirus (SARS-CoV-2) began in Wuhan, China in December 2019, and has spread throughout China and to 31 other countries and territories. In this paper we talk about how it affected the South Asian Association for Regional Cooperation (SAARC) Countries, and further compare the number of confirmed cases, number of deaths, number of recoveries, number of active cases and the mortality rate amongst these countries. This report summarizes the aggressive measures collected from the country's local health departments and multiple other agencies, while the countries implemented to slow down the transmission of COVID-19.

Table of Contents

1. Introduction

This paper introduces an objective approach on predicting the spread of COVID-19 using a simple, but powerful method to do so. The data that we used is reliable and useful in many ways. The cause of this study is to stop overspending and being over careful . We focus on the cumulative daily figures aggregated country wise of the three main variables of interest: confirmed cases, deaths and recoveries. This paper describes the timeline of a live forecasting exercise with massive potential implications for planning and decision making and provides objective forecasts for the confirmed cases of COVID-19.

2. System Model Description and Algorithms used

We developed a graphical framework for the data representation of COVID-19. The proposed method in this paper is to predict and have a clear visualization of the coronavirus outbreak in the SAARC countries.

The methods used are:

1. Support Vector Mechanism (SVM)
2. Bayesian Ridge Regression
3. Polynomial Regression

1. The Support Vector Machines (SVM)

The Support Vector Machines (SVM) is a classification technique used as a mathematical function named as a classifier to predict the right class of unknown observations of a testing data set. SVM is a powerful classifier in biomedical science, image processing and data mining for the detection and classification purposes. SVM is an efficient classifier to classify two different sets of observations into their relevant class. It has the means to handle high-dimensional and nonlinear data excellently.

```
#setting up the SVM model

svm_confirmed = SVR(shrinking=True, kernel='poly', gamma=0.01, epsilon=1, degree=6, C=0.1)
svm_confirmed.fit(X_train_confirmed, y_train_confirmed)
svm_pred = svm_confirmed.predict(future_forcast)

# check against testing data using a plot and printing out the mean_absolute_error as well as mean_squared_error

svm_test_pred = svm_confirmed.predict(X_test_confirmed)
plt.plot(y_test_confirmed)
plt.plot(svm_test_pred)
plt.legend(['Test Data', 'SVM Predictions'])
plt.show()
```

Figure 2.0: Code snippet of the used SVM method

2. Bayesian Ridge Regression

Bayesian regression can be implemented by using regularization parameters in estimation. The Bayesian Ridge estimator applies Ridge regression and its coefficients to find out a post estimation under the Gaussian distribution.

```
# bayesian ridge polynomial regression

bayesian_poly = PolynomialFeatures(degree=4)
bayesian_poly_X_train_confirmed = bayesian_poly.fit_transform(X_train_confirmed)
bayesian_poly_X_test_confirmed = bayesian_poly.fit_transform(X_test_confirmed)
bayesian_poly_future_forcast = bayesian_poly.fit_transform(future_forcast)
tol = [1e-6, 1e-5, 1e-4, 1e-3, 1e-2]
alpha_1 = [1e-7, 1e-6, 1e-5, 1e-4, 1e-3]
alpha_2 = [1e-7, 1e-6, 1e-5, 1e-4, 1e-3]
lambda_1 = [1e-7, 1e-6, 1e-5, 1e-4, 1e-3]
lambda_2 = [1e-7, 1e-6, 1e-5, 1e-4, 1e-3]
normalize = [True, False]
bayesian_grid = {'tol': tol, 'alpha_1': alpha_1, 'alpha_2' : alpha_2,
                 'lambda_1': lambda_1, 'lambda_2' : lambda_2,'normalize' : normalize}

bayesian = BayesianRidge(fit_intercept=False)
bayesian_search = RandomizedSearchCV(bayesian, bayesian_grid,
                                     scoring='neg_mean_squared_error', cv=3,
                                     return_train_score=True, n_jobs=-1, n_iter=40,
                                     verbose=1)
bayesian_search.fit(bayesian_poly_X_train_confirmed, y_train_confirmed)

bayesian_confirmed = bayesian_search.best_estimator_
test_bayesian_pred = bayesian_confirmed.predict(bayesian_poly_X_test_confirmed)
bayesian_pred = bayesian_confirmed.predict(bayesian_poly_future_forcast)

plt.plot(y_test_confirmed)
plt.plot(test_bayesian_pred)
plt.legend(['Test Data', 'Bayesian Ridge Polynomial Predictions'])
plt.show()
```

Figure 2.1: Code snippet of the used Bayesian Ridge Regression

3. Polynomial Regression

Polynomial regression, like linear regression, uses the relationship between the variables x and y to find the best way to draw a line through the data points. Python has methods for finding a relationship between data-points and to draw a line of polynomial regression.

```
#setting up the Polynomial Regression

poly = PolynomialFeatures(degree=5)
poly_X_train_confirmed = poly.fit_transform(X_train_confirmed)
poly_X_test_confirmed = poly.fit_transform(X_test_confirmed)
poly_future_forcast = poly.fit_transform(future_forcast)

linear_model = LinearRegression(normalize=True, fit_intercept=False)
linear_model.fit(poly_X_train_confirmed, y_train_confirmed)
test_linear_pred = linear_model.predict(poly_X_test_confirmed)
linear_pred = linear_model.predict(poly_future_forcast)
print('MAE:', mean_absolute_error(test_linear_pred, y_test_confirmed))
print('MSE:', mean_squared_error(test_linear_pred, y_test_confirmed))

plt.plot(y_test_confirmed)
plt.plot(test_linear_pred)
plt.legend(['Test Data', 'Polynomial Regression Predictions'])
plt.show()
```

Figure 2.2: Code snippet of the used Polynomial Regression

3. Visual Representation (Graphs)

Afghanistan's Graphical Data

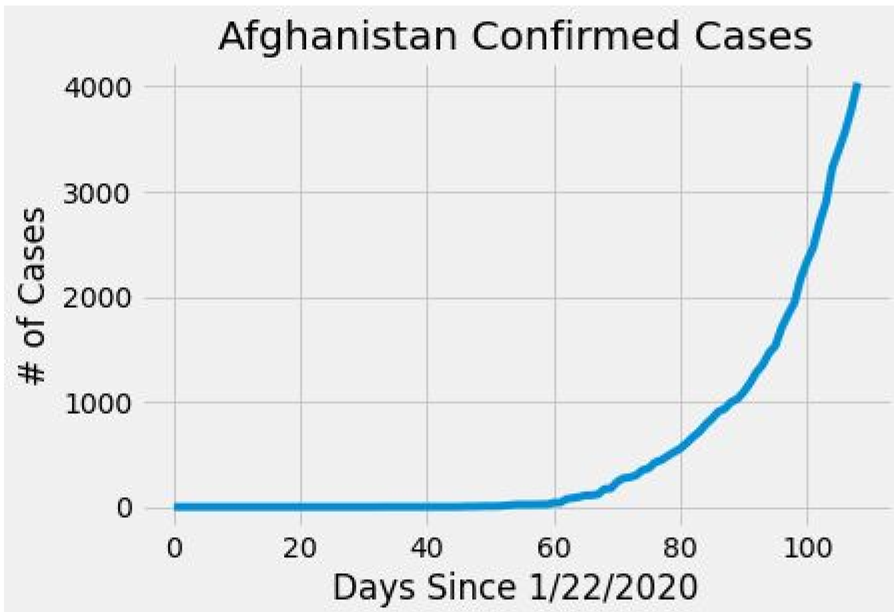

Figure 3.1: Daily total number of confirmed cases found in Afghanistan since 22 January 2020

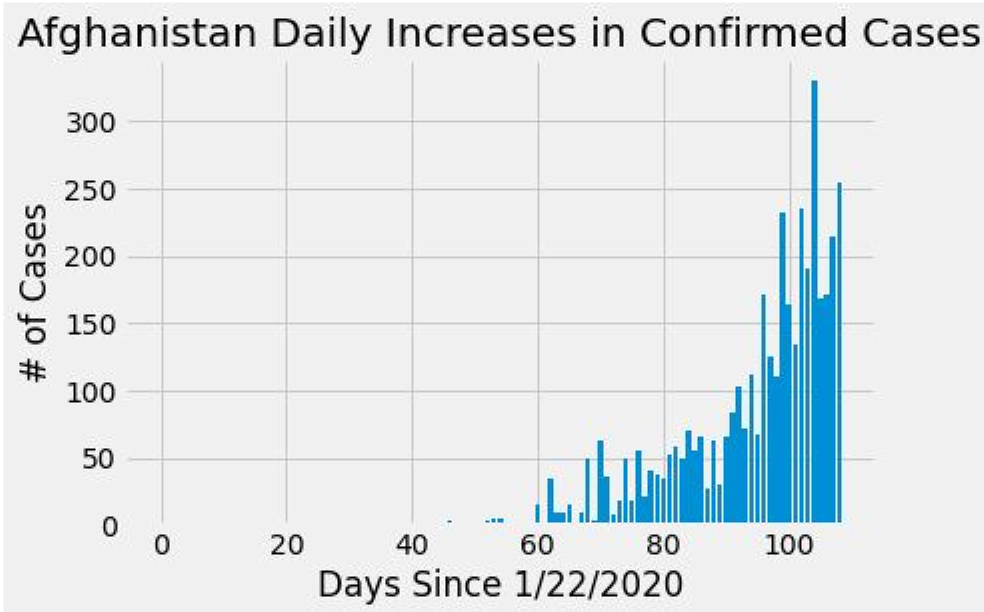

Figure 3.2: Daily increase in confirmed cases found in Afghanistan since 22 January 2020

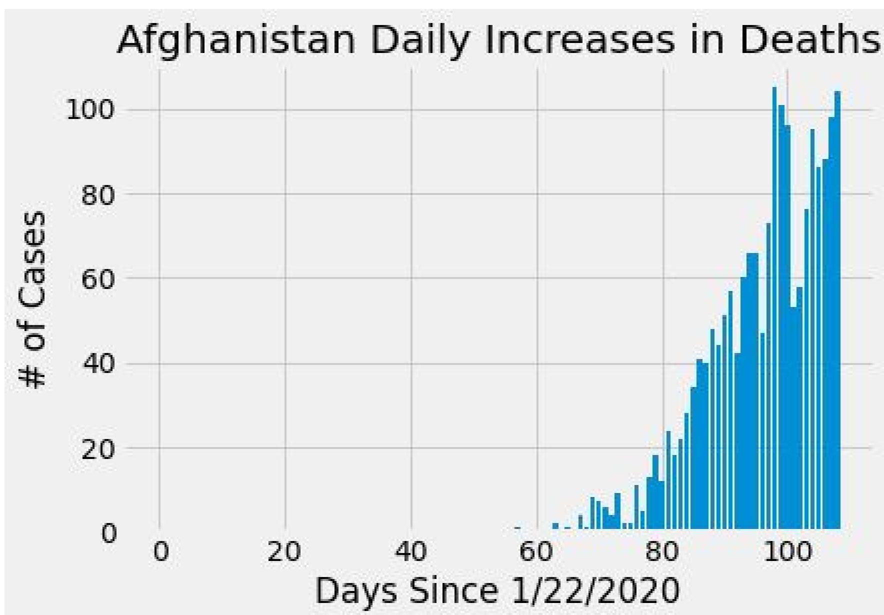

Figure 3.3: Daily increase in deaths found in Afghanistan since 22 January 2020

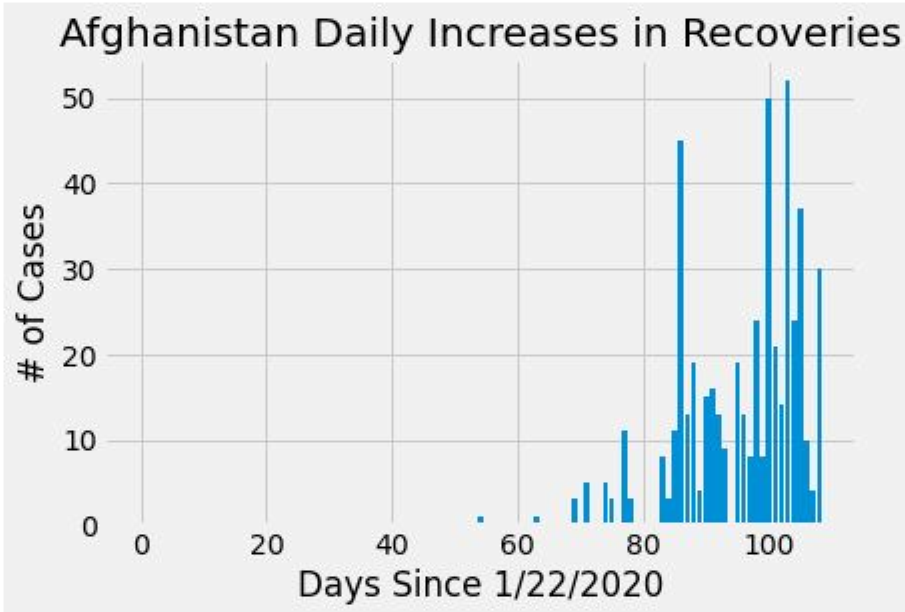

Figure 3.4: Daily increase in recoveries found in Afghanistan since 22 January 2020

Bangladesh's Graphical Data

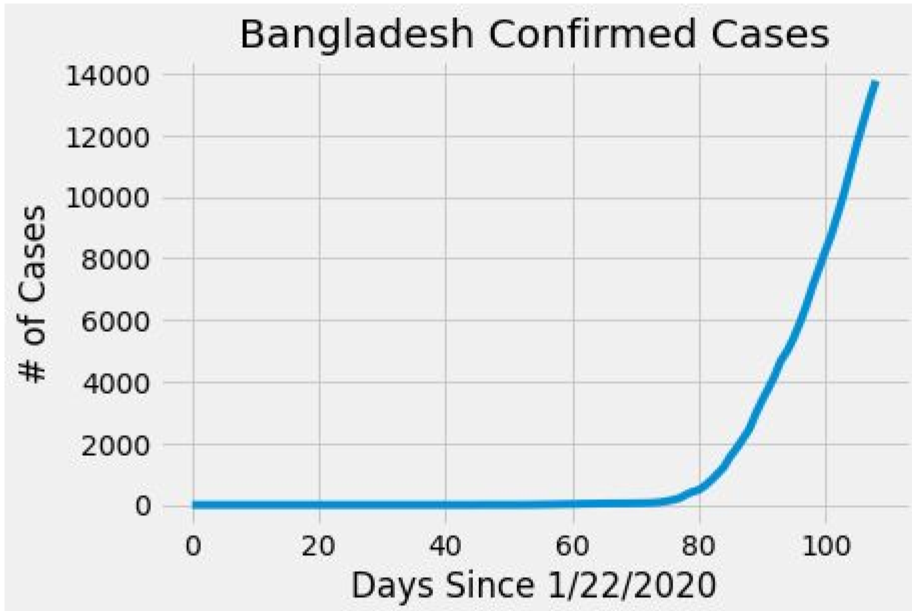

Figure 3.5: Daily total number of confirmed cases found in Bangladesh since 22 January 2020

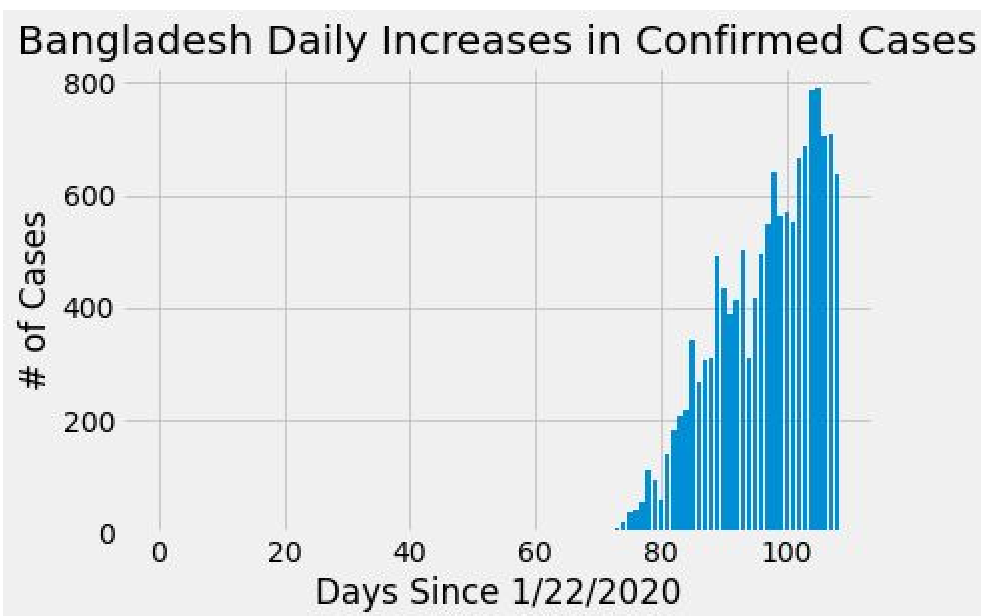

Figure 3.6: Daily increase in confirmed cases found in Bangladesh since 22 January 2020

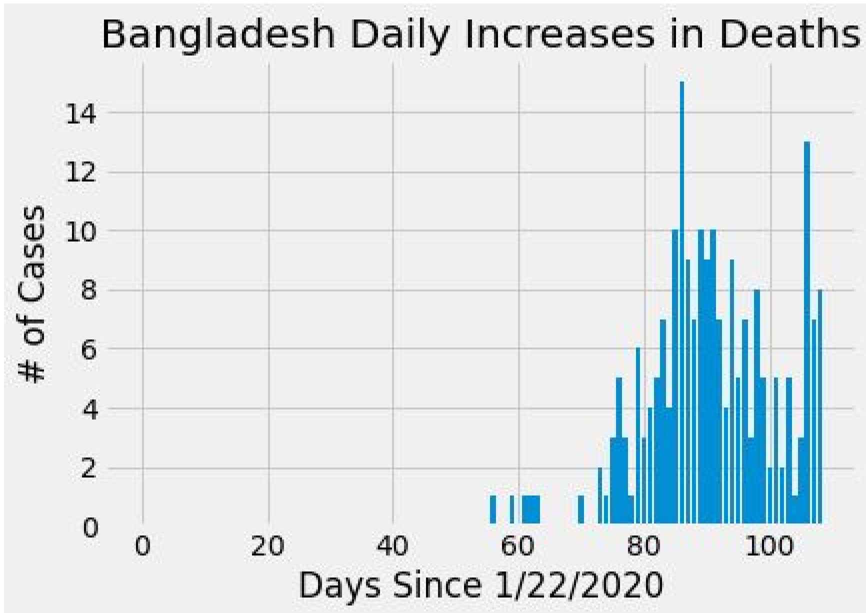

Figure 3.7: Daily increase in deaths found in Bangladesh since 22 January 2020

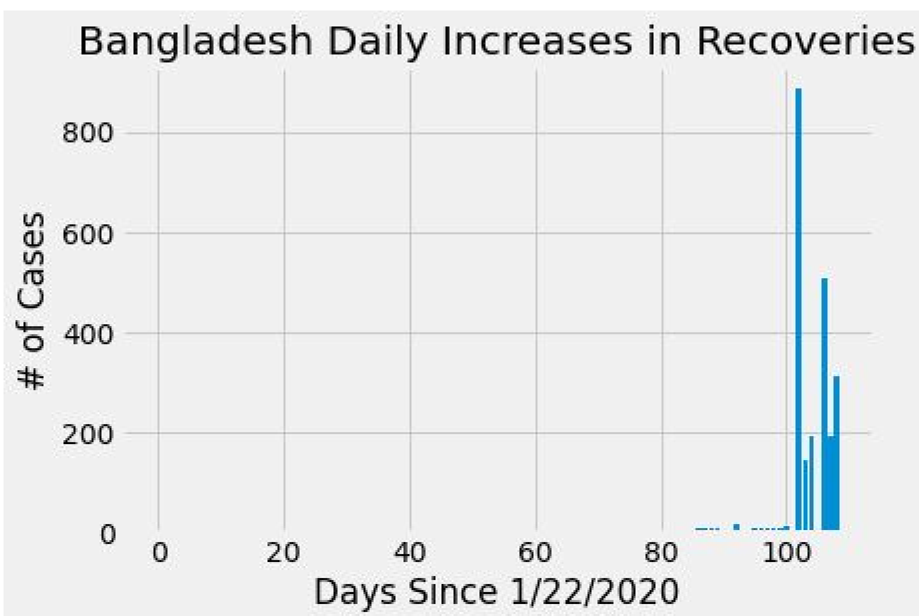

Figure 3.8: Daily increase in recoveries found in Bangladesh since 22 January 2020

Bhutan's Graphical Data

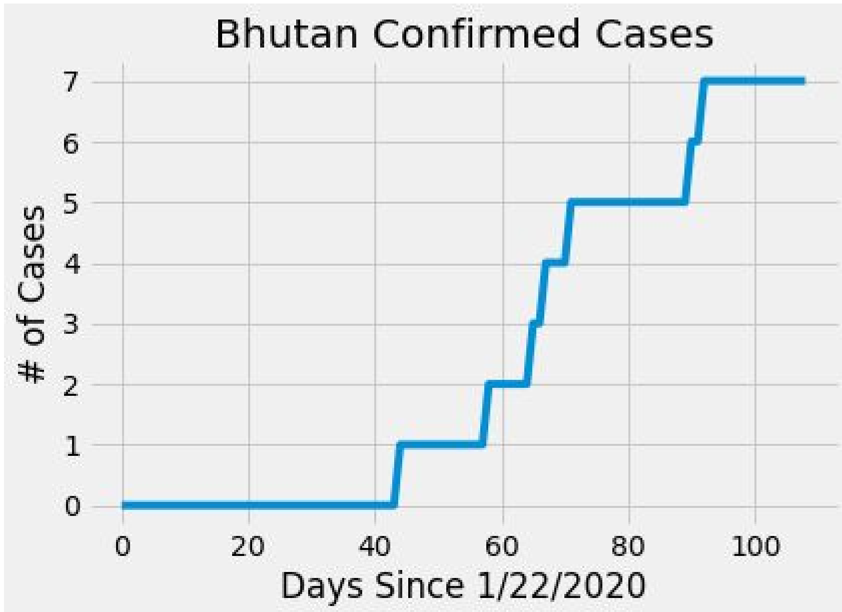

Figure 3.9: Daily total number of confirmed cases found in Bhutan since 22 January 2020

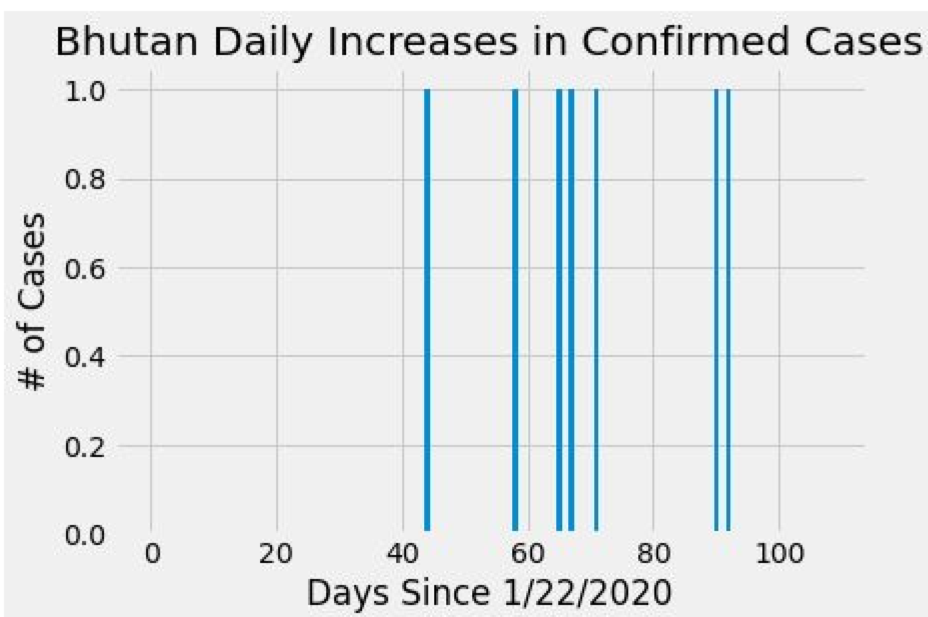

Figure 3.10: Daily increase in confirmed cases found in Bhutan since 22 January 2020

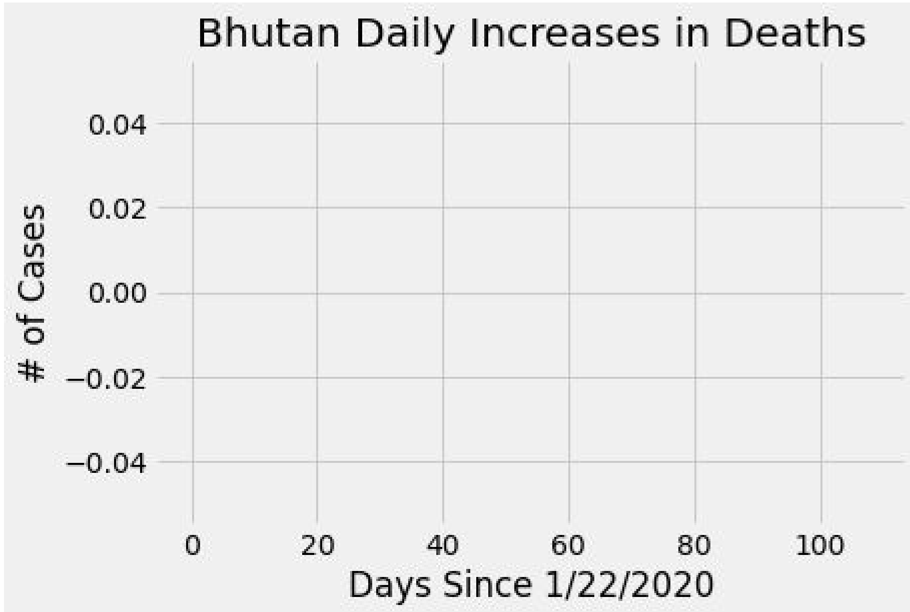

Figure 3.11: Daily increase in deaths found in Bhutan since 22 January 2020

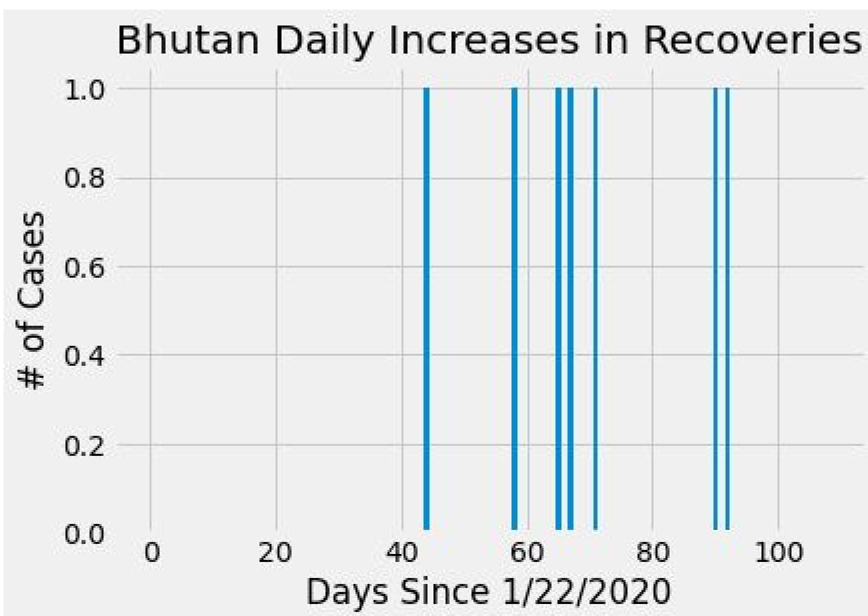

Figure 3.12: Daily increase in recoveries found in Bhutan since 22 January 2020

India’s Graphical Data

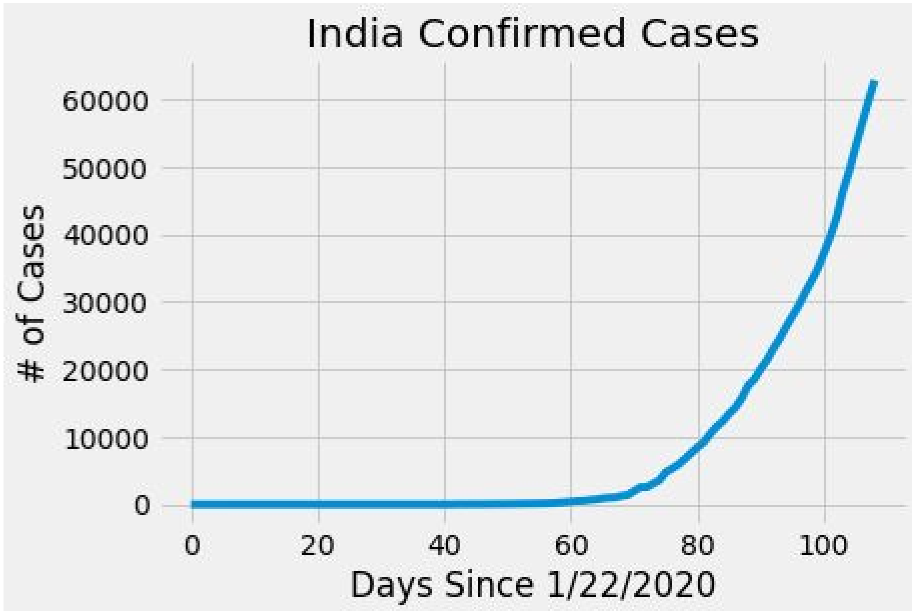

Figure 3.13: Daily total number of confirmed cases found in India since 22 January 2020

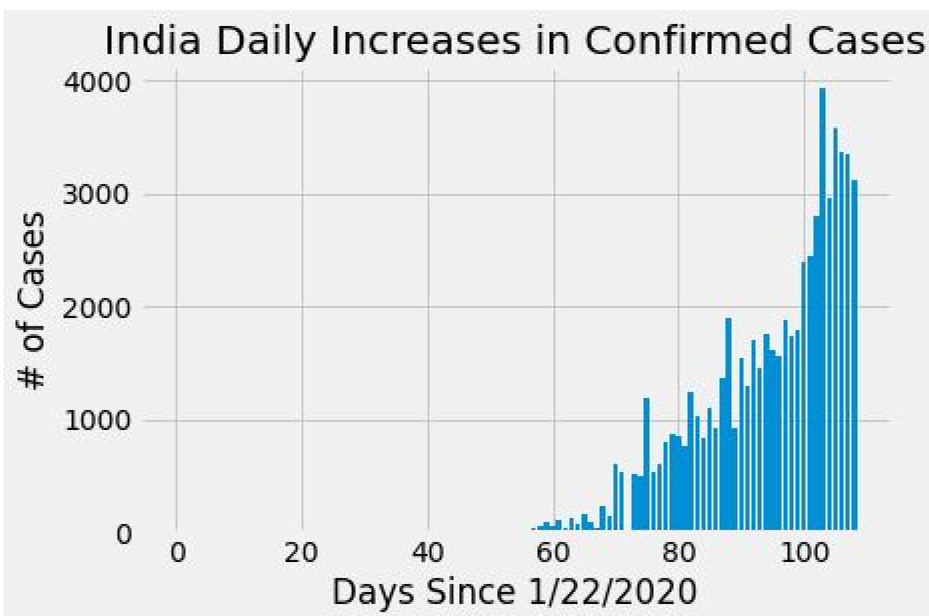

Figure 3.14: Daily increase in confirmed cases found in India since 22 January 2020

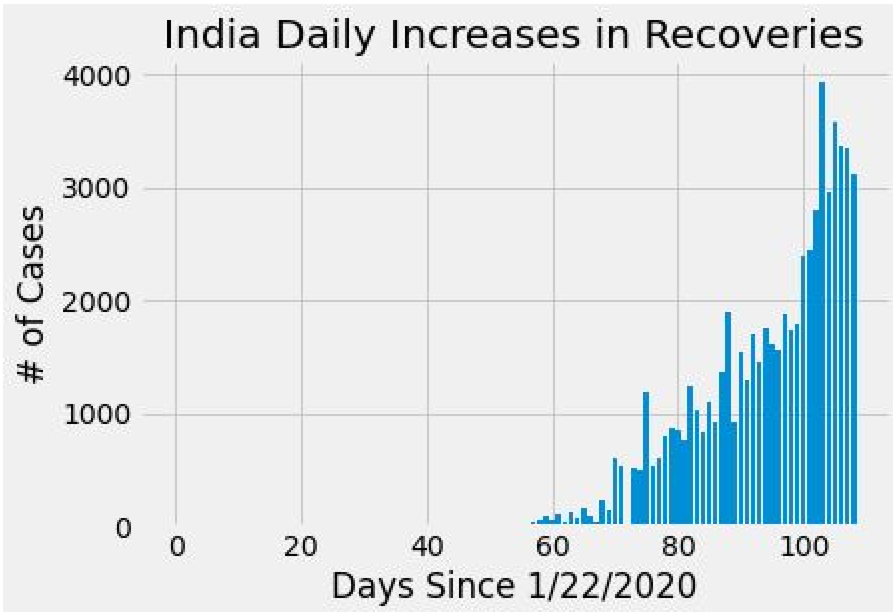

Figure 3.15: Daily increase in recoveries found in India since 22 January 2020

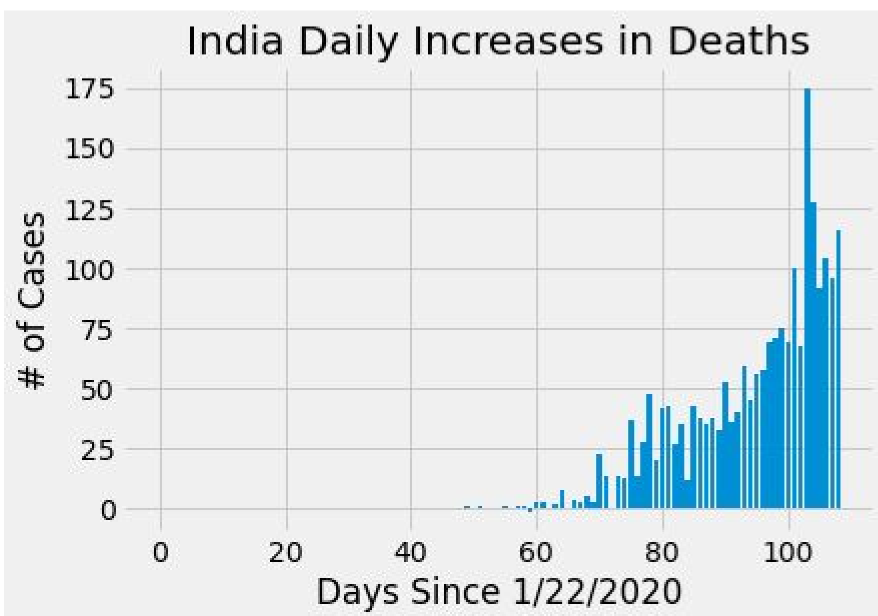

Figure 3.16: Daily increase in deaths found in India since 22 January 2020

Maldives's Graphical Data

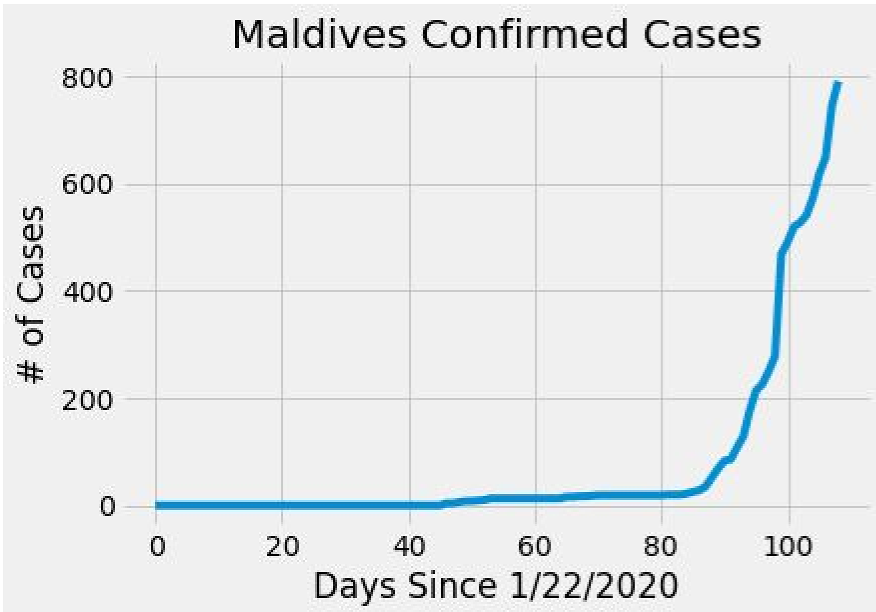

Figure 3.17: Daily total number of confirmed cases found in Maldives since 22 January 2020

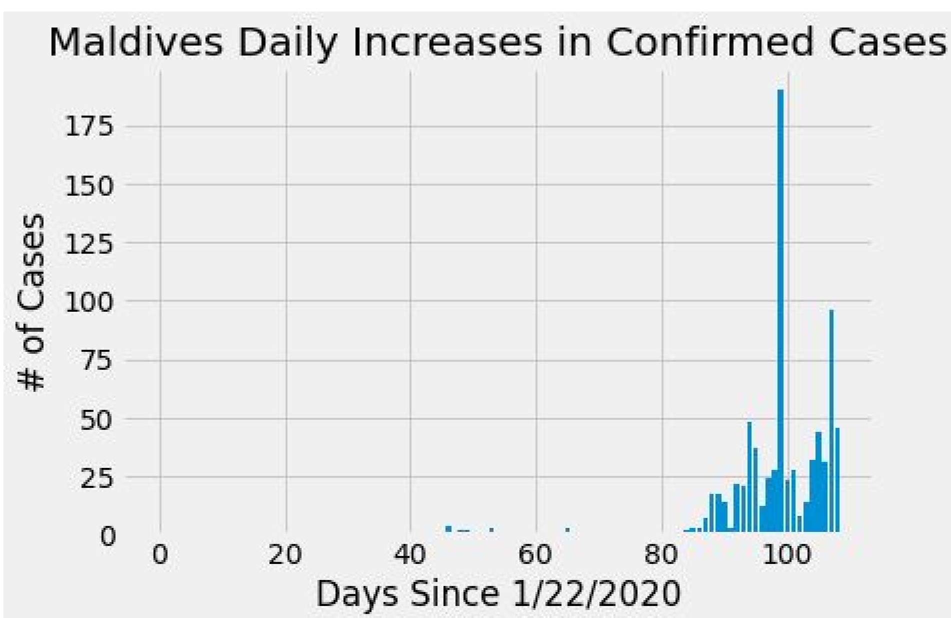

Figure 3.18: Daily increase in confirmed cases found in Maldives since 22 January 2020

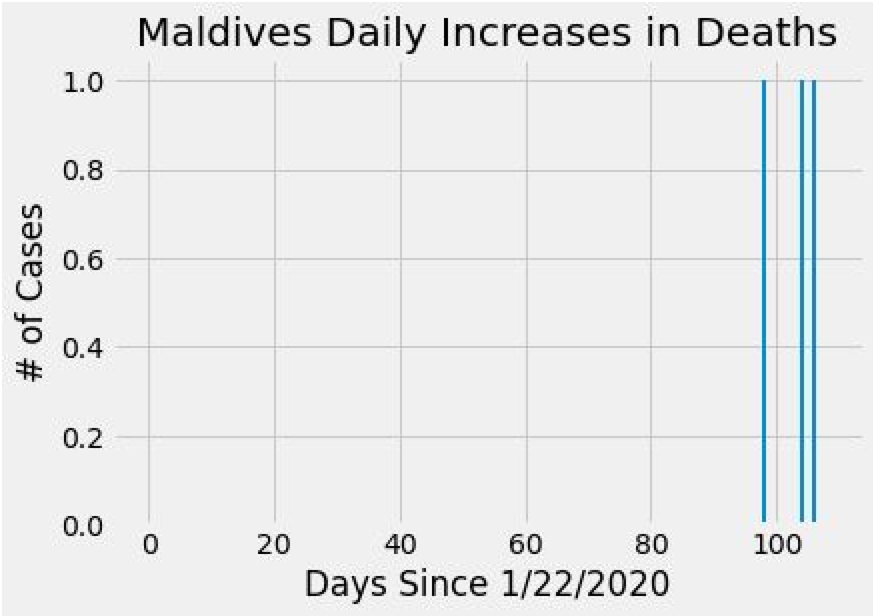

Figure 3.19: Daily increase in deaths found in Maldives since 22 January 2020

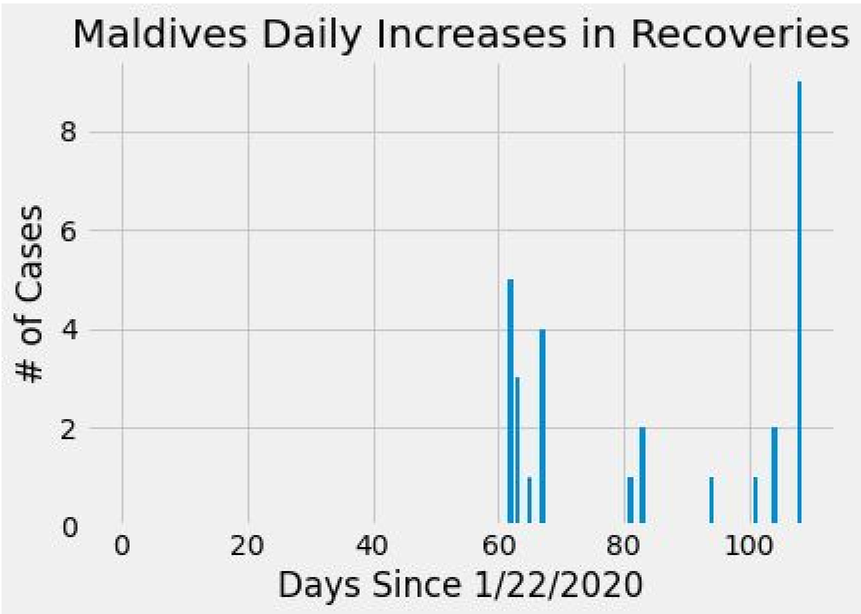

Figure 3.20: Daily increase in recoveries found in Maldives since 22 January 2020

Pakistan's Graphical Data

Figure 3.21: Number of confirmed Coronavirus cases in Pakistan

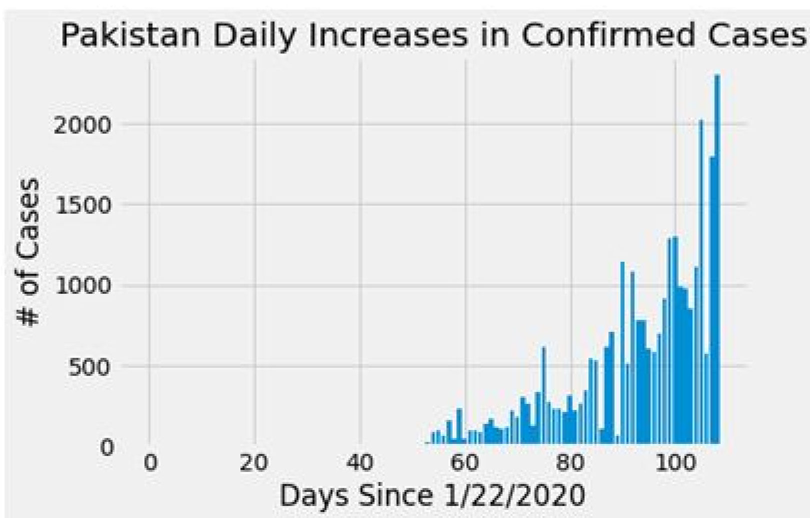

Figure 3.22: Number of daily increase in confirmed cases of Coronavirus in Pakistan

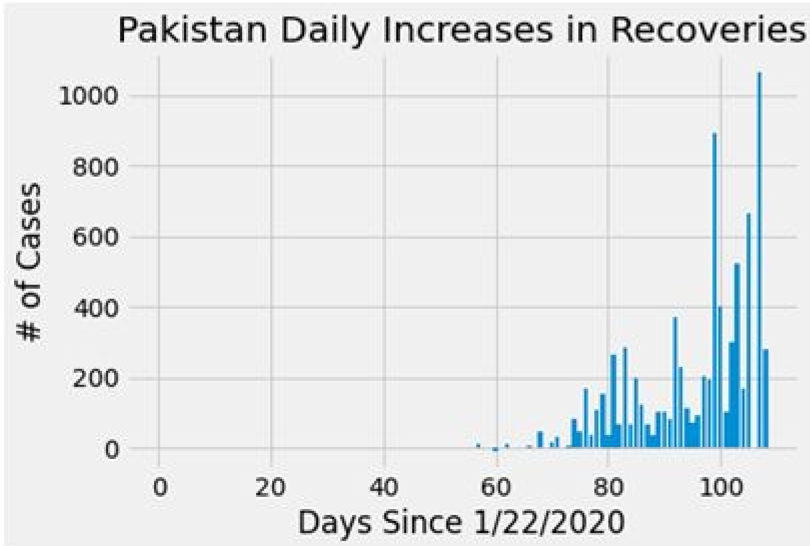

Figure 3.23: Number of daily increase in recoveries from Coronavirus in Pakistan

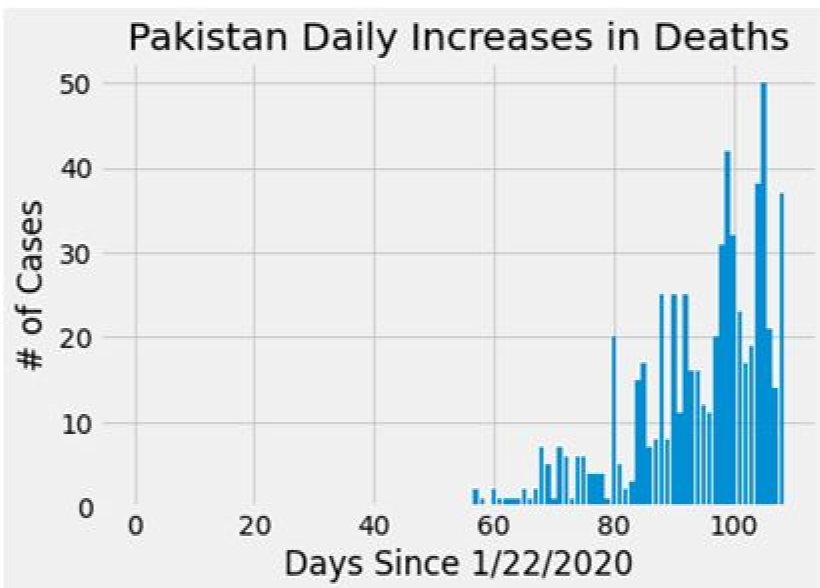

Figure 3.24: Number of daily increase of deaths from Coronavirus in Pakistan

Nepal's Graphical Data

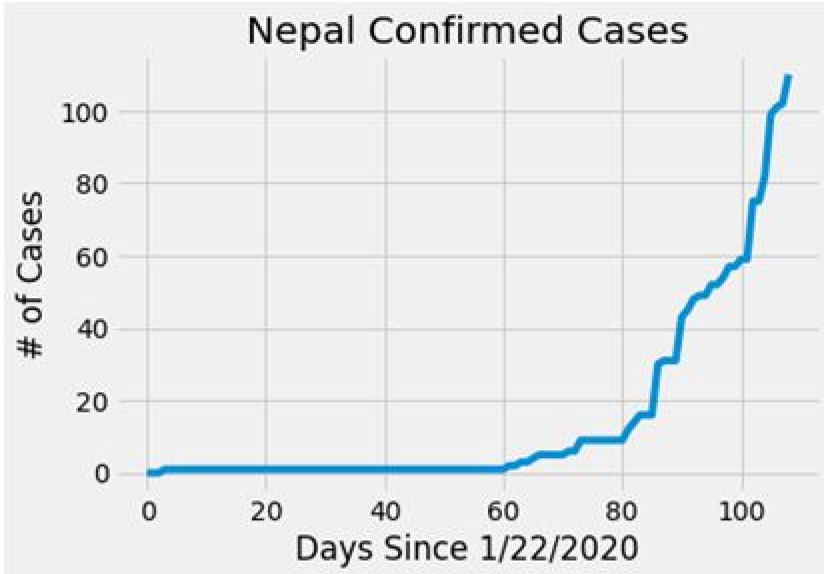

Figure 3.25: Number of confirmed cases of Coronavirus in Nepal

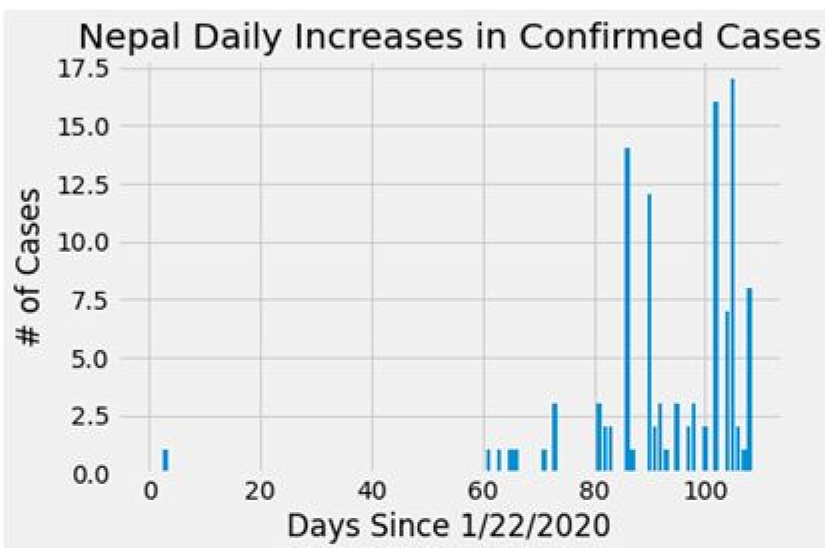

Figure 3.26: Number of daily increase in confirmed cases of Coronavirus in Nepal

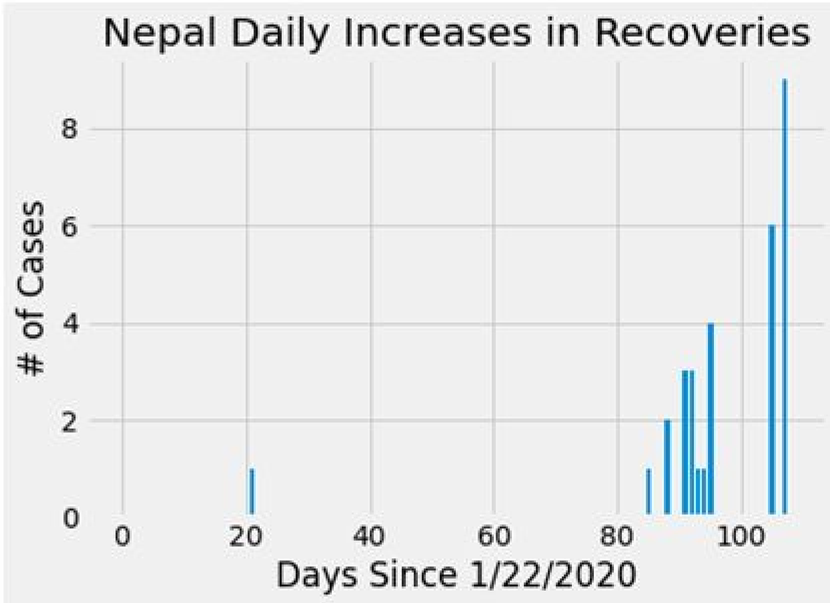

Figure 3.27: Number of daily increase in recoveries from Coronavirus in Nepal

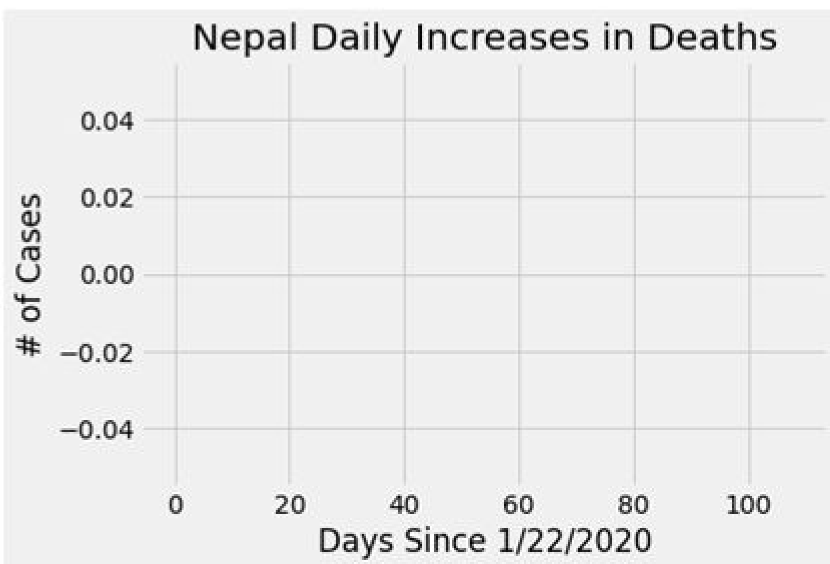

Figure 3.28: Number of daily increase of deaths from Coronavirus in Nepal

Sri Lanka's Graphical Data

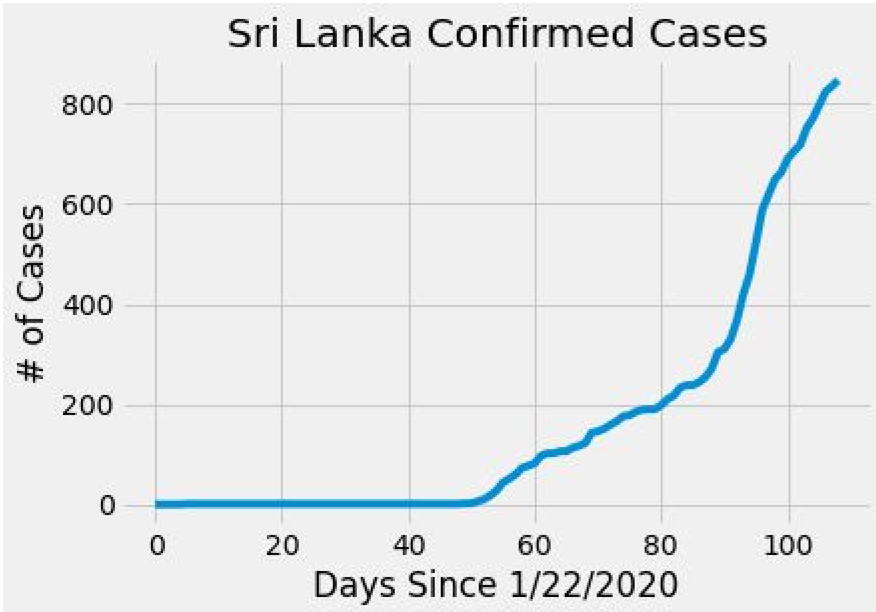

Figure 3.29: Number of confirmed cases of Coronavirus in Sri Lanka

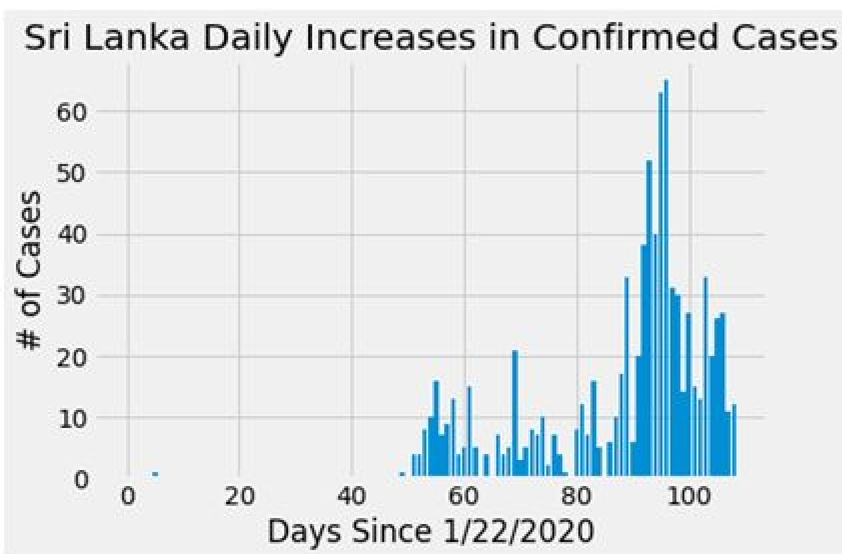

Figure 3.30: Number of daily increase in confirmed cases of Coronavirus in Sri Lanka

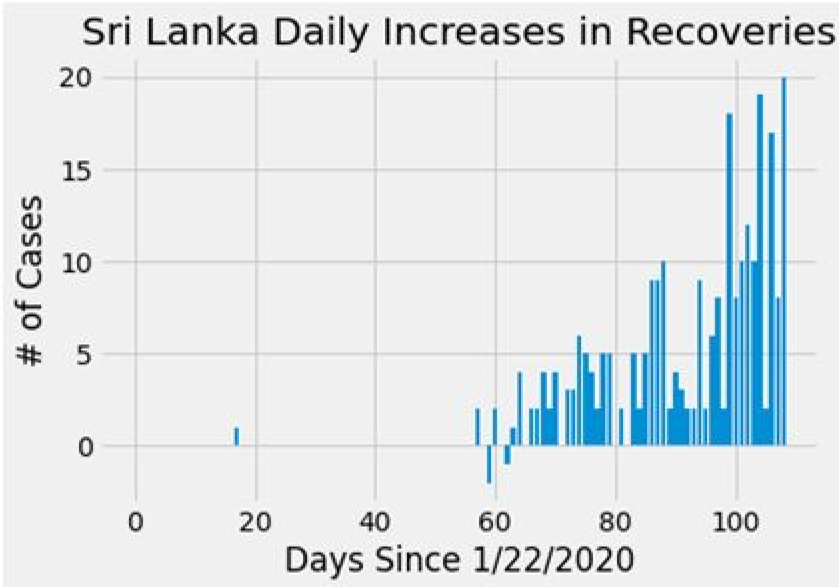

Figure 3.31: Number of daily increase in recoveries from Coronavirus in Sri Lanka

Some of the recovery values are seen to be negative which implies that the patients who were recovering, were again tested for Coronavirus positive.

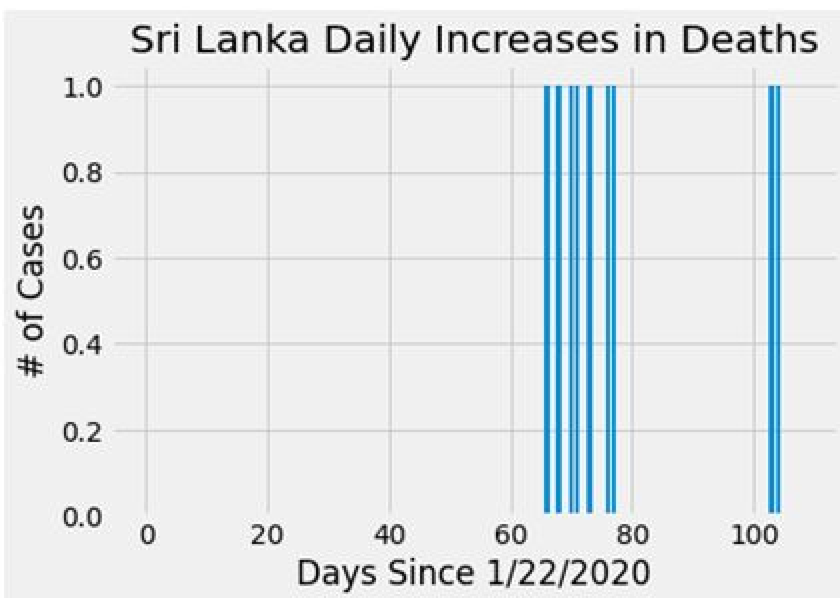

Figure 3.32: Number of daily increase in deaths from Coronavirus in Sri Lanka

Overall visualization of SAARC Countries

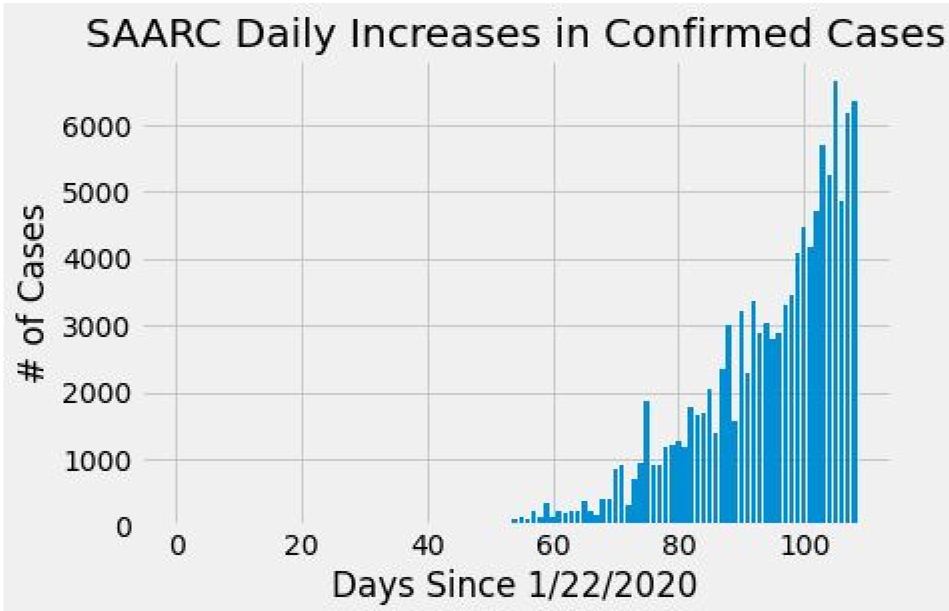

Figure 3.33: Daily total number of confirmed cases found in SAARC countries since 22 January 2020

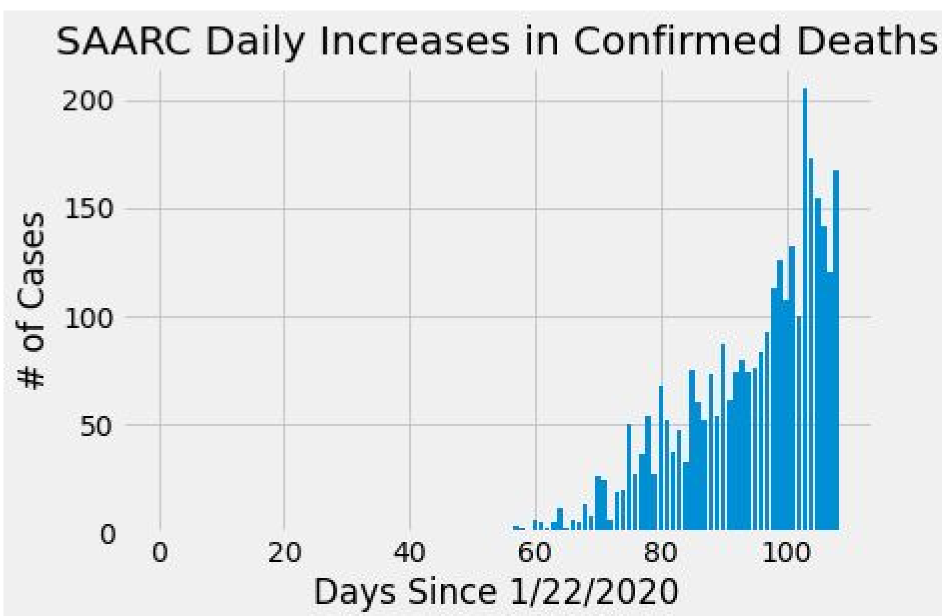

Figure 3.34: Daily total number of deaths found in SAARC countries since 22 January 2020

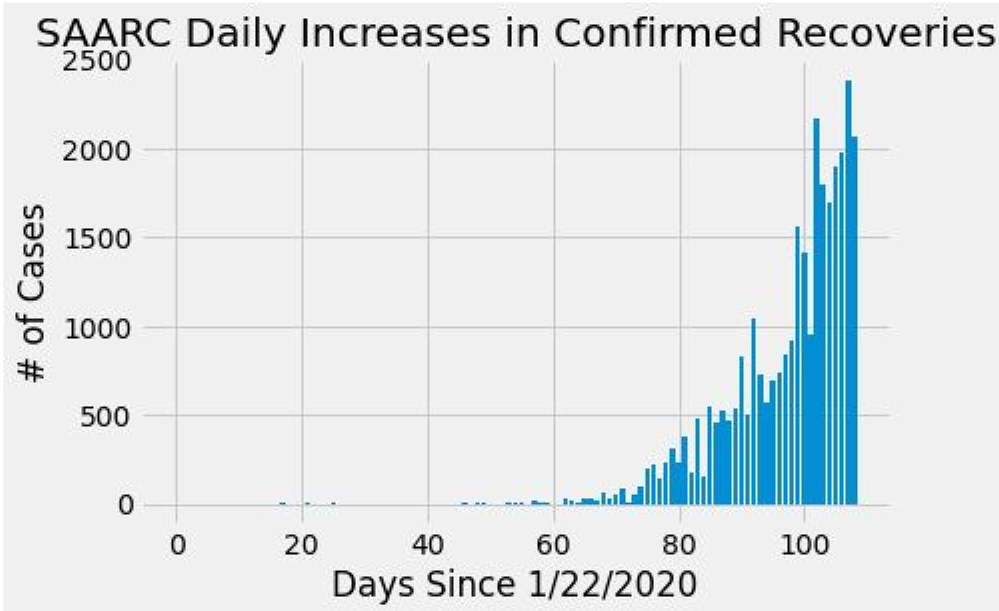

Figure 3.35: Daily total number of confirmed recovery cases found in SAARC countries since 22 January 2020

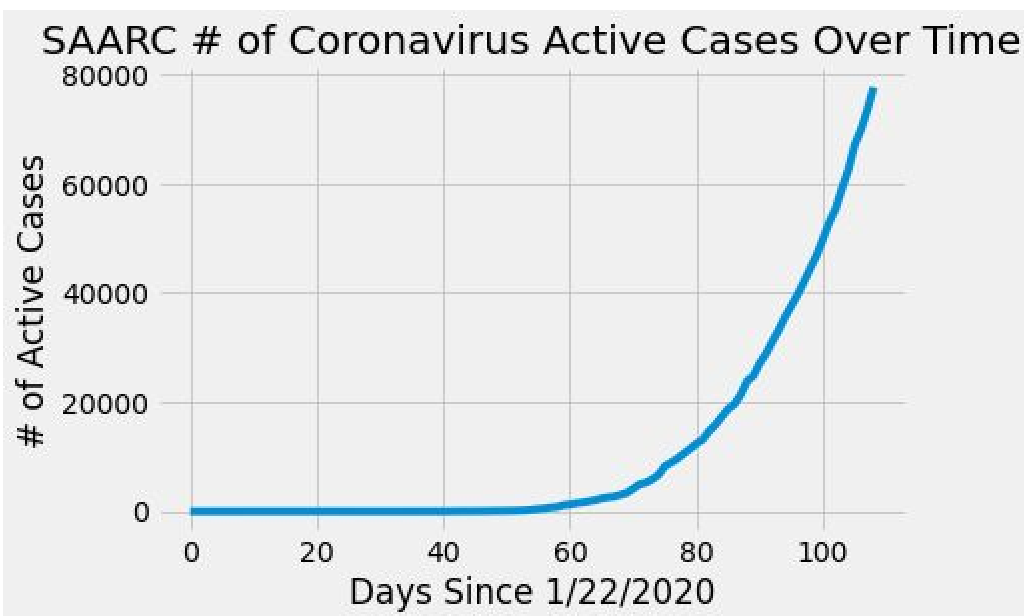

Figure 3.36: Daily total number of cases found in SAARC countries since 22 January 2020

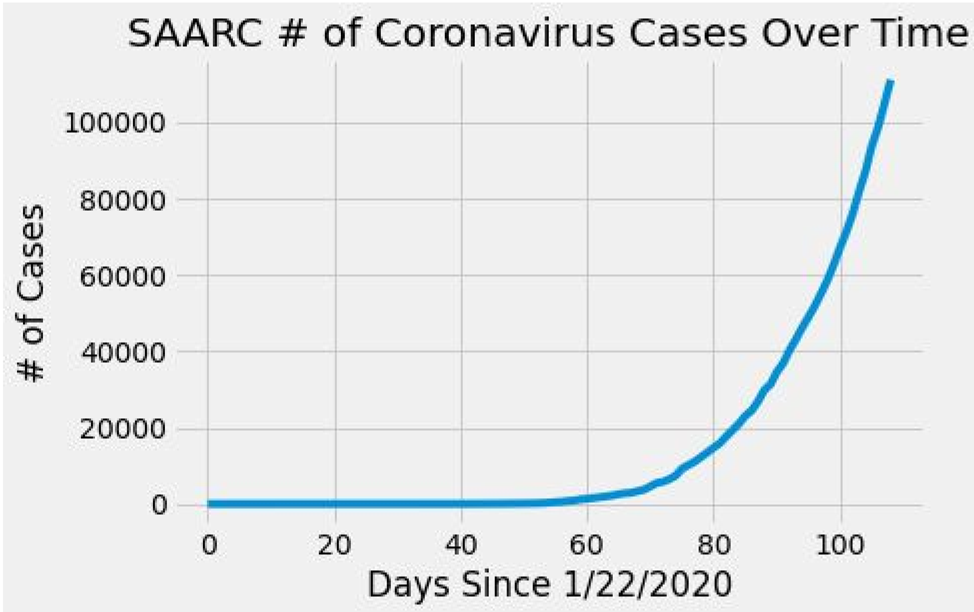

Figure 3.37: Daily total number of deaths found in SAARC countries since 22 January 2020

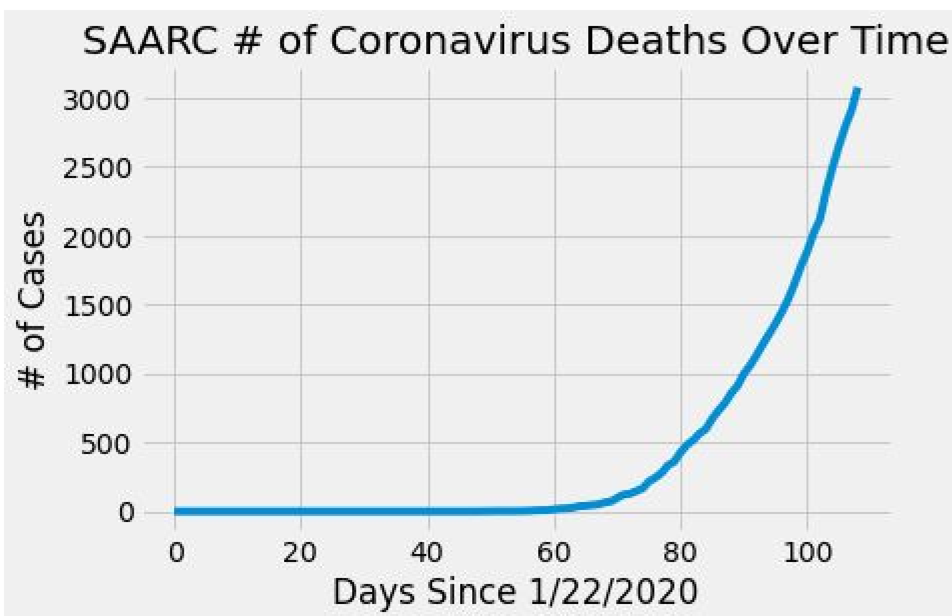

Figure 3.38: Daily total number of recoveries found in SAARC countries since 22 January 2020

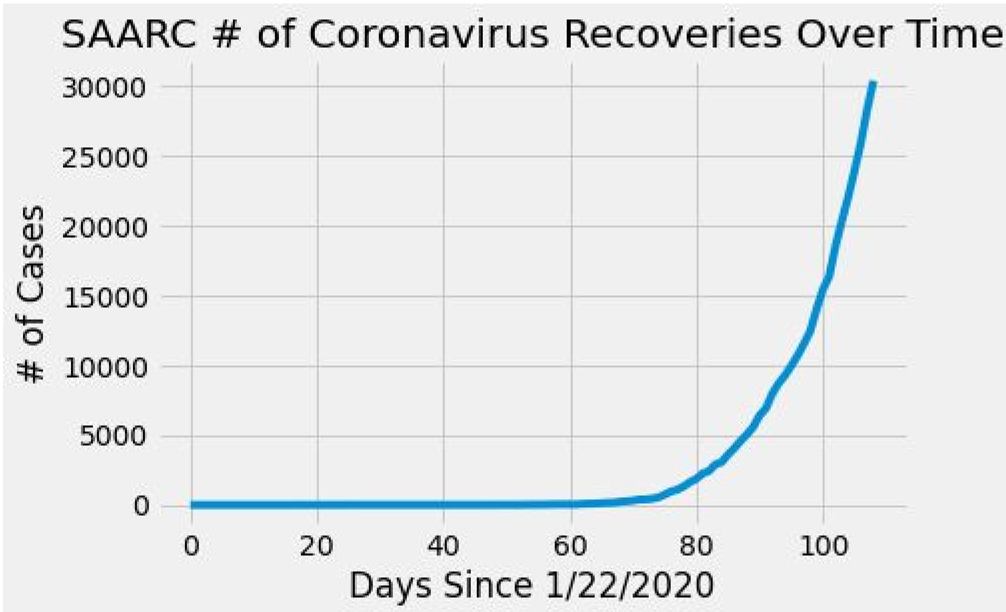

Figure 3.39: Daily log of total number of cases found in SAARC countries since 22 January 2020

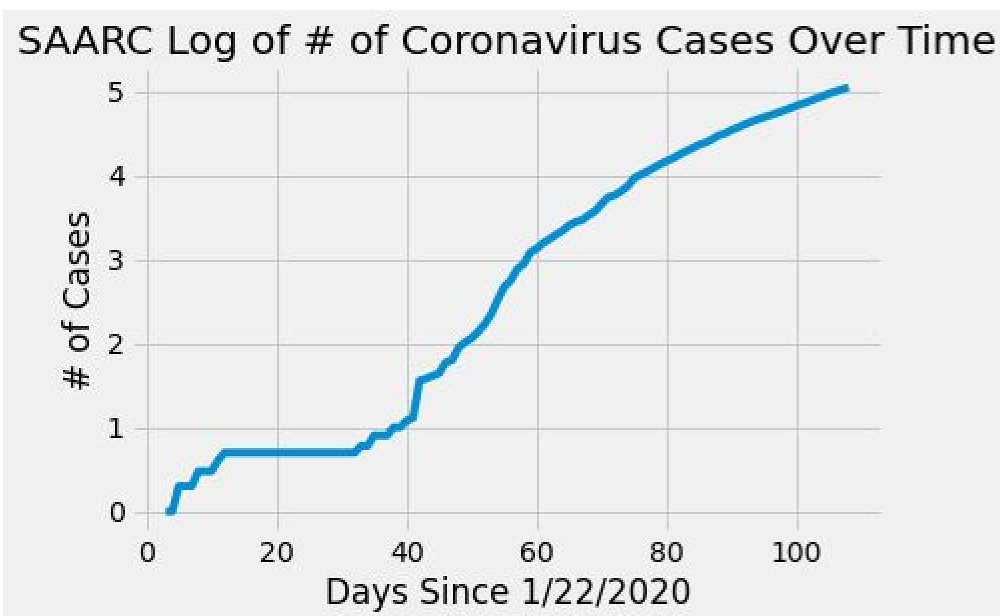

Figure 3.40: Daily total number of active cases found in SAARC countries since 22 January 2020

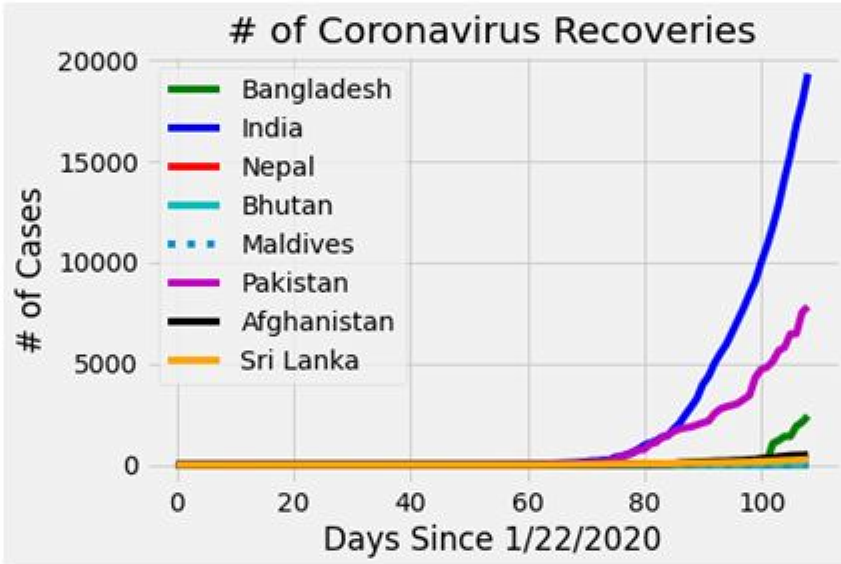

Figure 3.41: Number of Coronavirus Case Recoveries in SAARC countries

Here we see the number of recoveries which is the highest in India with around 20000 cases, then comes Pakistan at a very low number of recovery cases and lastly the rest.

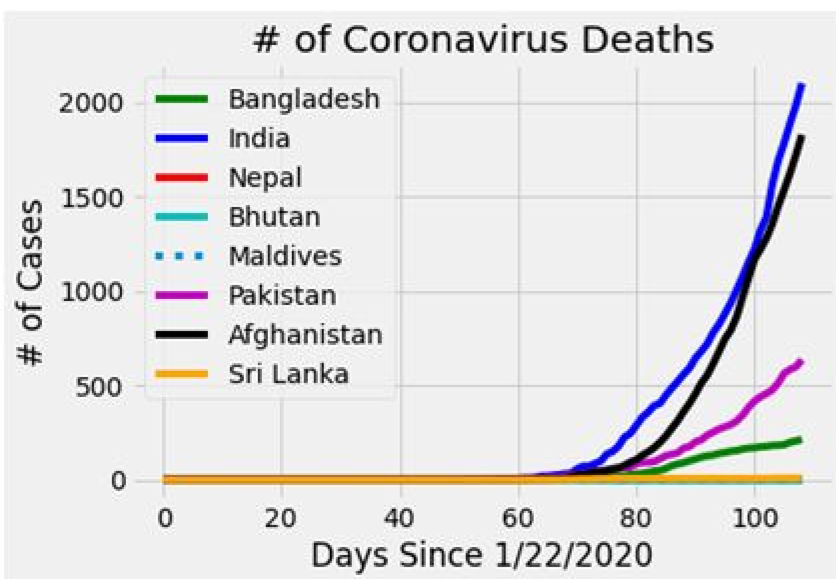

Figure 3.42: Number of deaths from Coronavirus in SAARC countries

Here we see India and Afghanistan has the highest number of deaths compared to the rest of the countries.

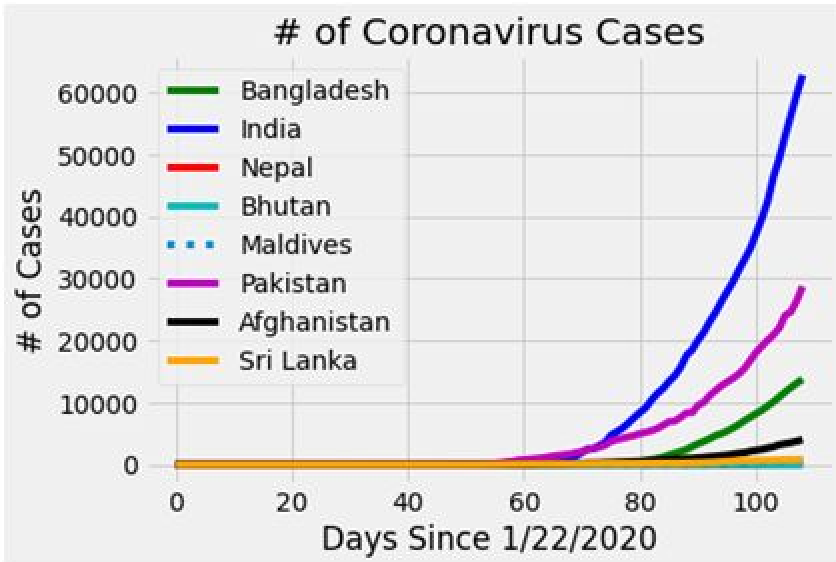

Figure 3.43: Number of Coronavirus cases in SAARC countries

The number of coronavirus cases is the highest in India and Pakistan, Bangladesh being close to 15000 and the rest being below 10000.

SAARC Log of Coronavirus Death and Recoveries

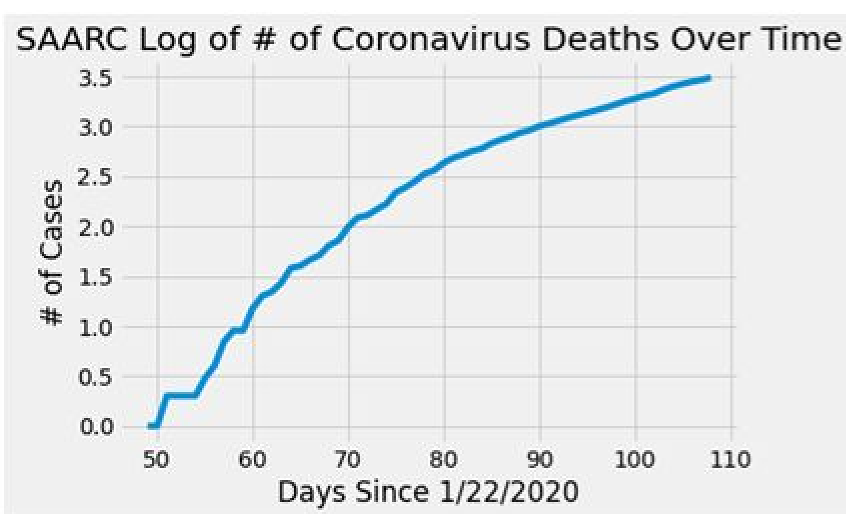

Figure 3.44: Log of number of deaths from Coronavirus over time for SAARC

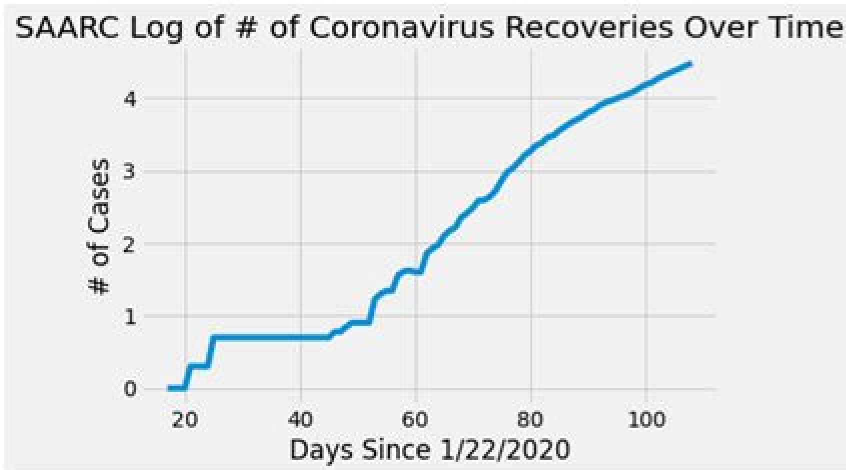

Figure 3.45: Log of number of recoveries from Coronavirus over time for SAARC

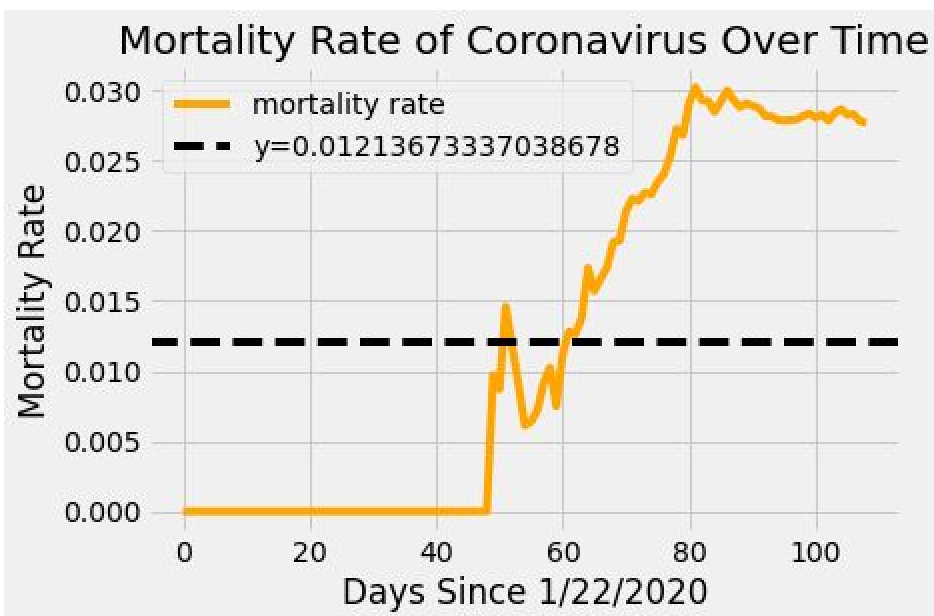

Figure 3.46: Daily mortality rate of SAARC countries since 22 January 2020

The yellow line describes the trend in the mortality rate whereas the dashed black line signifies the mean mortality rate (y value = 0.0121 approx.) of Coronavirus overtime.

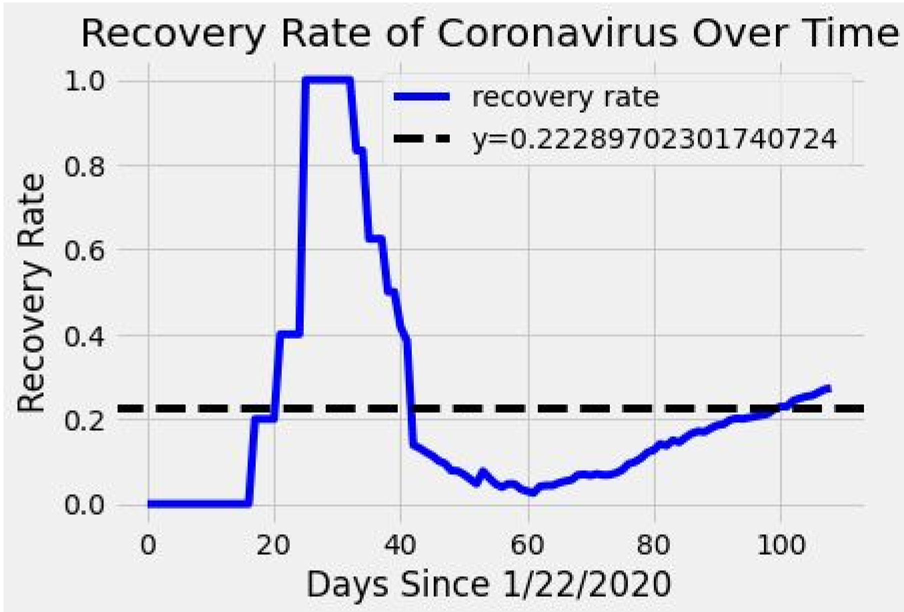

Figure 3.47: Daily recovery rate of SAARC countries since 22 January 2020

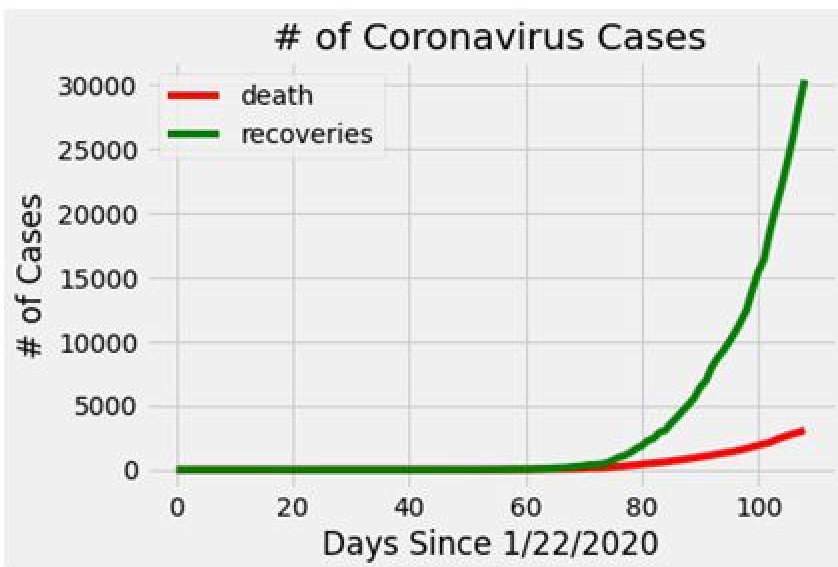

Figure 3.48: Number of cases of deaths and recoveries from Coronavirus in SAARC

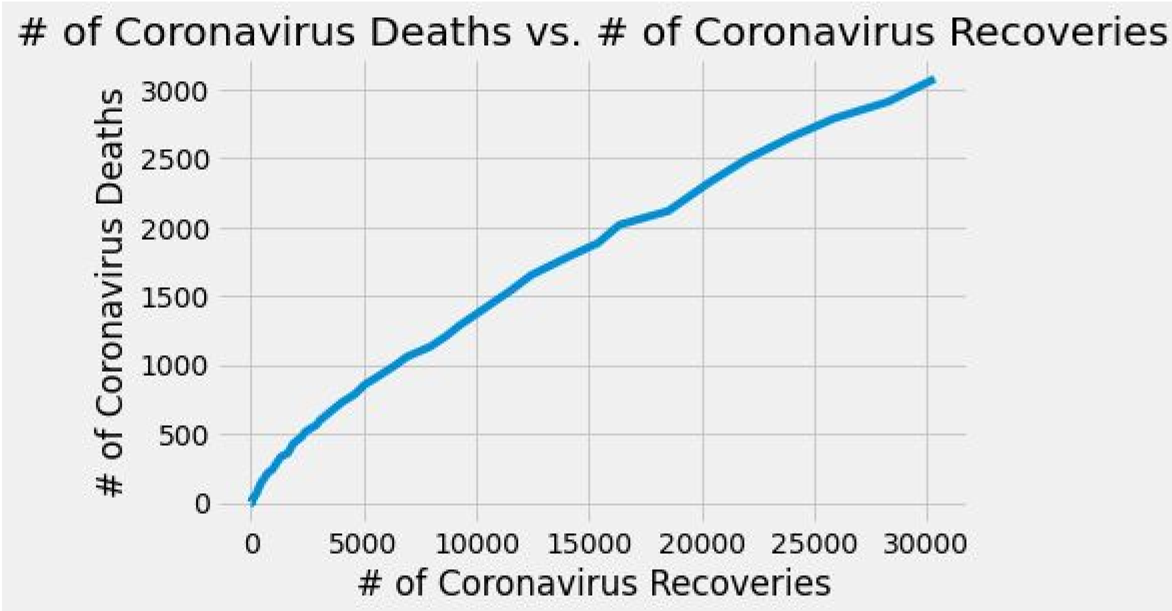

Figure 3.49: Visualizes the number of deaths in comparison to number of recoveries in SAARC countries

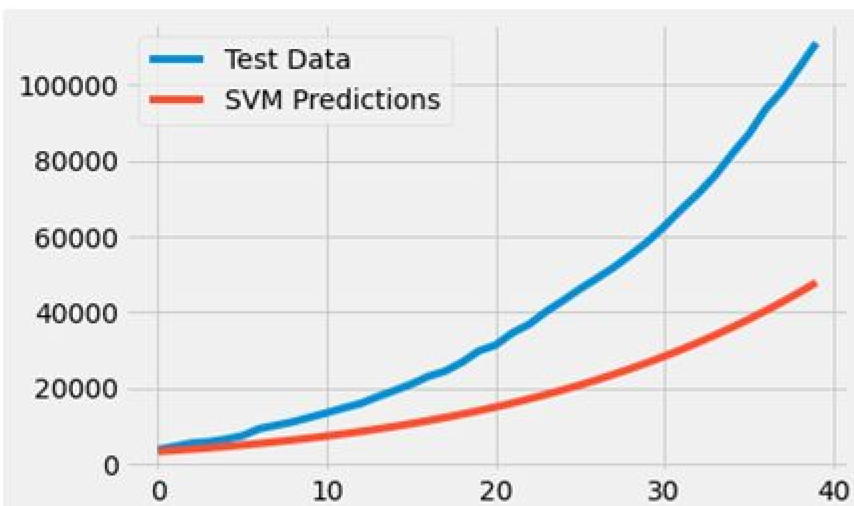

Figure 3.50: SVM predictions with Test data of the Coronavirus cases in SAARC

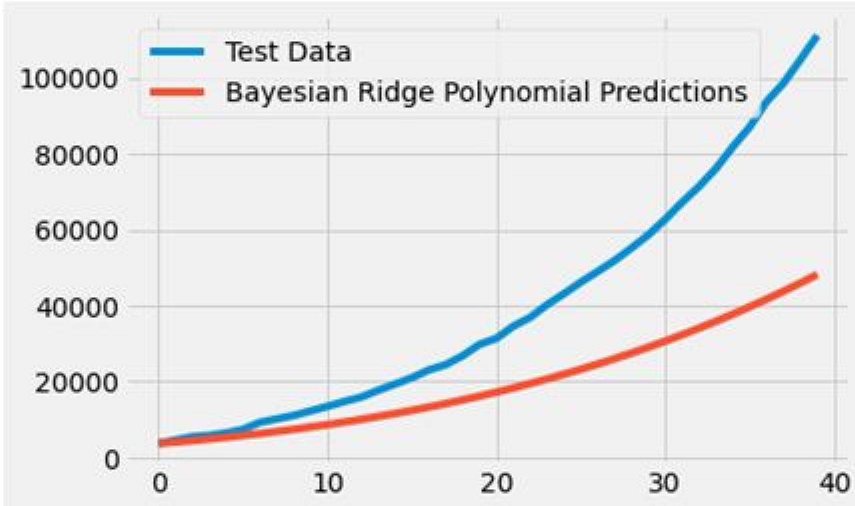

Figure 3.51: Bayesian Ridge Polynomial Predictions with Test data for the Coronavirus cases in SAARC

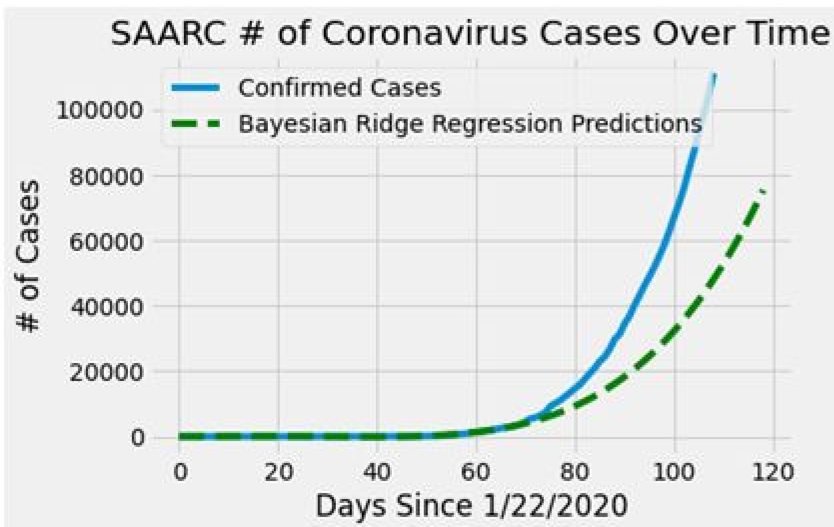

Figure 3.52: Bayesian Ridge Regression Predictions alongside the actual Confirmed Cases over time in SAARC

	Date	Bayesian Ridge Predicted # of Confirmed Cases SAARC
0	05/10/2020	50507.0
1	05/11/2020	52912.0
2	05/12/2020	55400.0
3	05/13/2020	57973.0
4	05/14/2020	60633.0
5	05/15/2020	63383.0
6	05/16/2020	66223.0
7	05/17/2020	69156.0
8	05/18/2020	72183.0
9	05/19/2020	75308.0

Figure 3.53: Bayesian Ridge Predicted Confirmed Cases of Coronavirus in SAARC

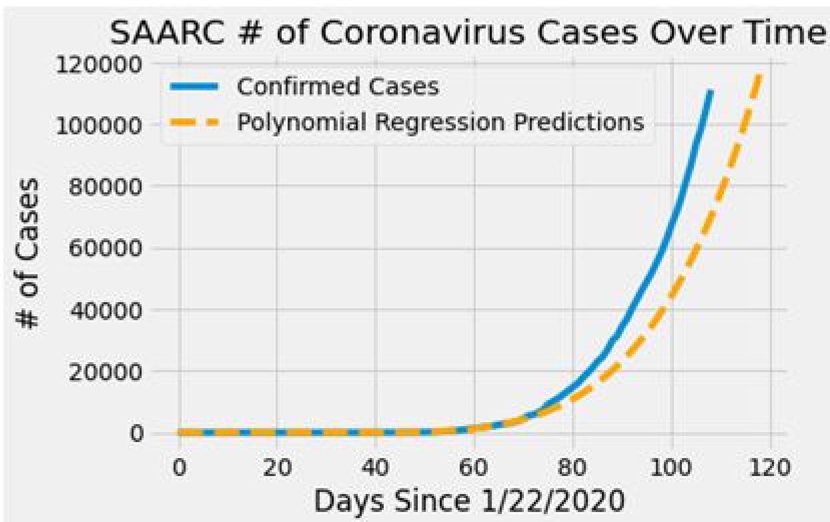

Figure 3.54: Polynomial Regression Predictions alongside actual Confirmed Cases over time in SAARC

	Date	Polynomial Predicted # of Confirmed Cases SAARC
0	05/10/2020	73849.0
1	05/11/2020	77895.0
2	05/12/2020	82112.0
3	05/13/2020	86504.0
4	05/14/2020	91078.0
5	05/15/2020	95838.0
6	05/16/2020	100791.0
7	05/17/2020	105941.0
8	05/18/2020	111294.0
9	05/19/2020	116857.0

Figure 3.55: Polynomial Predicted Confirmed Cases of Coronavirus in SAARC

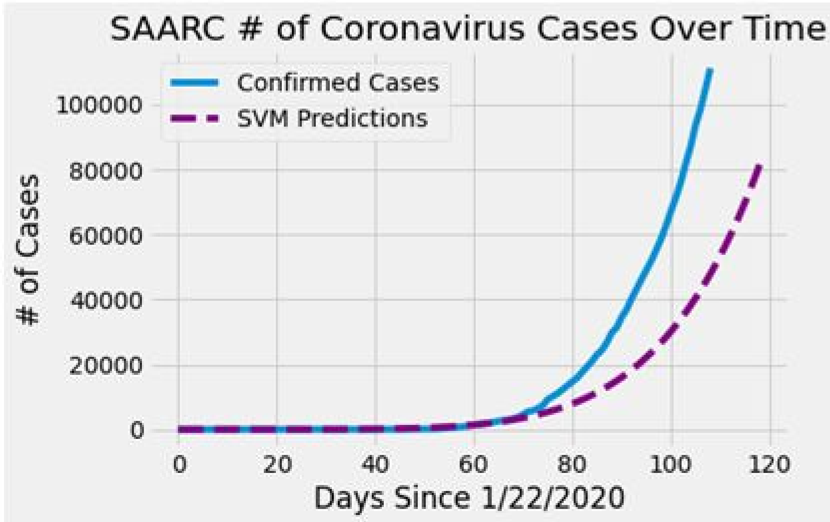

Figure 3.56: SVM Predictions alongside the actual Confirmed Cases over time in SAARC

	Date	SVM Predicted # of Confirmed Cases SAARC
0	05/10/2020	50582.0
1	05/11/2020	53431.0
2	05/12/2020	56413.0
3	05/13/2020	59533.0
4	05/14/2020	62794.0
5	05/15/2020	66204.0
6	05/16/2020	69766.0
7	05/17/2020	73486.0
8	05/18/2020	77371.0
9	05/19/2020	81425.0

Figure 3.57: Data Table of SVM Predicted Confirmed Cases of Coronavirus in SAARC

4. Code

The code required for this system to work is nearly about 600 lines, hence the code's github link has been attached below:

click here to redirect to the link

A readMe file has been attached to the Git repository which explains how the code works as well as how the same results can be reproduced using the code. The readMe file link has been attached below:

Read me

5. Results and Conclusion

Polynomial regression provides the best approximation of the relationship between the dependent and independent variable. Since a broad range of the function can be fit under it so it fits a wide range of curvature. The COVID-19 data set that is being used comprises exactly that.Taking into consideration the different countries in SAARC as well as the different types of data including the deaths, recoveries, confirmed cases, etc. gives a perfect wide range of curvature. Since the model works well on that type of data, it is significantly better than the SVM prediction model as well as the Bayesian Ridge Prediction model. This information can be clearly inferred from the graphs plotted showing the relationships between the models and the test data that were used. As the curve of the Polynomial Regression based on the training data upon which the model was trained, models most accurately with the test data that was provided. The notable difference with the other models can also be derived based on numerical analysis from the predicted data tables of the models. The Polynomial Regression model gives the confirmed cases of Coronavirus to be ‘116857’ on ten days of predicted values whereas that of SVM and Bayesian Ridge models are ‘81425’ and ‘75308’ respectively. From this inferential analysis we can come to the conclusion that the Polynomial Regression model works best on the data chosen for SAARC countries.

To conclude, from the comparative analysis of the datasets it is observed that the effect of COVID-19 in India is much greater than other SAARC countries. India has maximum confirmed cases, number of deaths and recoveries compared to the other countries. Furthermore, Afghanistan follows India in the number of deaths due to this recent pandemic. On this note, Indian government should drastically take measures to prevent the spread of coronavirus as they are affected the most within SAARC.

6. References

1. CSSEGISandData. “CSSEGISandData/COVID-19.” *GitHub*, 8 May 2020, github.com/CSSEGISandData/COVID-19.

2. kp, Devakumar. “COVID-19 Dataset.” *Kaggle*, 10 May 2020, www.kaggle.com/imdevskp/corona-virus-report

3. Therealcyberlord. “Coronavirus (COVID-19) Visualization & Prediction.” *Kaggle*, Kaggle, 9 May 2020, www.kaggle.com/therealcyberlord/coronavirus-covid-19-visualization-prediction?fbclid=IwAR1iMwbWcUr9nFmZdGJpqauacbiQbnquC1jbWeUvcfD8WkC-c9PAVpvEFgY

4. Kaggle.com. 2020. COVID-19 Dataset. [online] Available at: <https://www.kaggle.com/imdevskp/corona-virus-report> [Accessed 10 May 2020].