3.3 Prediction

3.3.1 Malaria Report

The rising burden of disease counts as one of the most prominent concerns of a warming climate. These risks are particularly severe in rapidly growing cities. Surat is located on the banks of River Tapi that has temperature and humidity patterns that can be described as ideal mosquito genetic conditions. Surat has a long history of river floods and usual water logging during the peak rainy season. It makes Surat prone to endemic vector-borne diseases and morbidity. In the past, most cases in Gujarat were reported in Surat but due to the preventive actions taken by SMC that number has started to deteriorate. This decline has been reported despite an increase in population over time.

Climate change causes a possible increase in relative humidity and rainfall would increase Malaria risk in the city. We tried to develop an urban climate impact assessment model with public health as our focal point. We are using past data of the number of Malaria cases registered and meteorological data (rainfall, relative humidity) to predict Malaria risk. This helps SMC health and hospital organizations to take preventive steps that can be beneficial from an economic point of view.

Disease Incidence and Climate Interactions

Since Mosquito breeding and their disease transmission efficiency is highly influenced by climate patterns like rainfall, temperature, and relative humidity, most vector-borne cases have a seasonal trend. Thus, predicting the future spread of disease provides an opportunity for health officials to be prepared for a possible outbreak. According to the annual report by SMC Vector-borne diseases control department [1] , relative humidity above 60%, mean temperature around 25-30° along with continuous and dense rainfall is an ideal climate for the rise in Malaria Cases.

We can see in Figure-3.19 that with the increase in Humidity, the significant increase in Malaria cases occurs. Rainfall follows almost the same seasonal pattern as relative humidity. We have taken all cases registered from 2010 to 2019 and added monthly cases together. For relative humidity data, we have taken the average. The highest increase is between July-October and humidity is greater than 60% in those months.

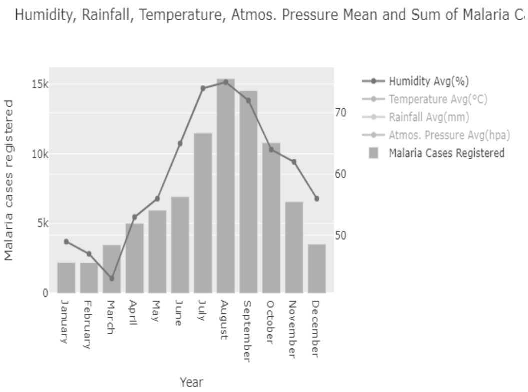

Figure 3.19 Avg Humidity and Monthly Malaria cases reported

Here, the decrease in air temperature plays a vital role in the increase in malaria cases. In summer, when the temperature is at the highest, there is not a significant increase in malaria cases which suggests that air temperature is in the negative correlation of malaria cases reported. Though, change in malaria cases is continuous throughout the year, the relative decrease in temperature suggests that it contributes to the increase in malaria cases. As you see in the Figure-3.20, the increase in malaria cases occurs during June-October and the temperature is between 25-30°.

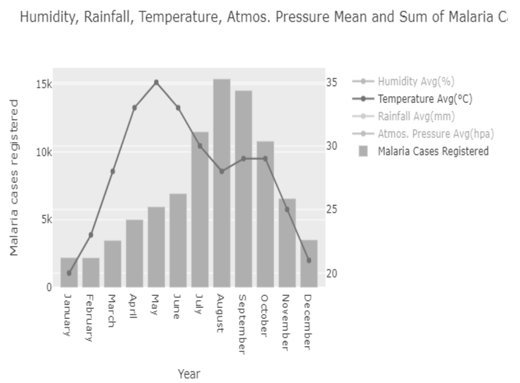

Figure 3.20 Avg Temperature and Monthly Malaria Cases

Here, the increase in rainfall correlates with the increase in malaria cases [17]. Though rainfall around non-monsoon months is around zero, when it rains, the cases are increased as there are potholes and spots with still water in which Anopheles mosquito lays eggs. Heavy rainfall can have a diverse range of effects on disease. In tropical and subtropical regions with crowding and poverty, heavy rainfall and flooding may trigger behavioral changes such as an increase in the morbidity of malaria.

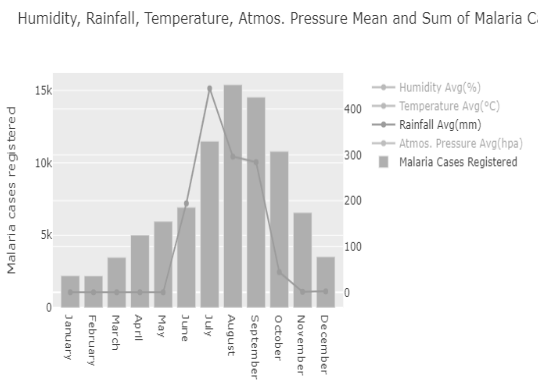

Figure 3.21 Avg Rainfall and montly reported cases

Table 3.Pearson Correlation between Meteorological elements and Malaria cases

Pearson Correlation between Meteorological elements and Malaria cases:				
	Relative Humidity(%)	Rainfall	Temperature	Atmospheric Pressure
Malaria Cases	0.23	0.5	0.62	-0.75

Correlations are calculated between [-1,1]. A correlation of -1 indicates that data points are negatively correlated which means that if one variable increases other decreases. A correlation of +1 indicates that data points are positively correlated which means that if one variable increases another increase as well. A correlation value 0 indicates that there is no correlation between two variables [2]. A Pearson correlation between two variables X and Y is calculated by

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

$\bar{X}$ - mean of X
$\bar{Y}$ - mean of Y
n - number of variables

Figure 3.22 Pearson's correlation matrix equation

Prediction

We started predicting with conventional machine learning models and though the results were adequate, we were looking for models that could help us with the time series forecasting. Conventional machine learning models do not train models in a sequential manner. Thus, we needed models that could perform satisfactorily on sequential data. We decided to use the SARIMA model which is Seasonal Auto-regressive Integrated Moving Average [3].

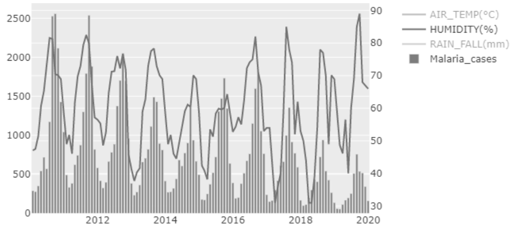

Figure 3.23 Humidity and Malaria cases of last 10 years

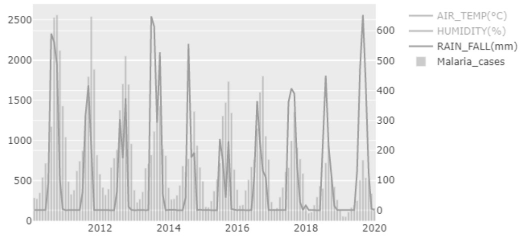

Figure 3.24 Rainfall and Malaria cases of last 10 years

From the Figure-3.23 and 3.24, we can depict that a recent drop in the number of malaria cases is not especially related to exogenous variables like Rainfall and Relative Humidity. There are no significant changes in either Humidity or Rainfall in the last few years that we can correlate with the recent changes in the trend of a number of cases registered. In Rainfall, the last year 2019 was an exceptional year but no changes in malaria cases occurred which was surprising considering the rainfall effect on malaria cases mentioned earlier in the report. Though Rainfall might affect malaria cases registered on a seasonal basis, it certainly is not helpful in predicting the trend of the graph. The same case can be made for Relative Humidity as well. The graph below can show you the decomposition of the time series which is malaria cases registered from 2010-2019.

Time series decomposition involves thinking of time series as a combination of Trend, Seasonality, and Residual which is known as noise as well. Decomposition provides a useful abstract model for thinking about time series generally and for better understanding problems during time series analysis and forecasting [4].

Here, our seasonal component is considerably high, which means that there are more similarities between the seasonal values of the time series. If we follow the trend graph, we can see that it is moving downward that infers that the number of malaria cases being recorded is decreasing due to the SMC's initiative. Starting from 2016, the graph has been in a downward trend.

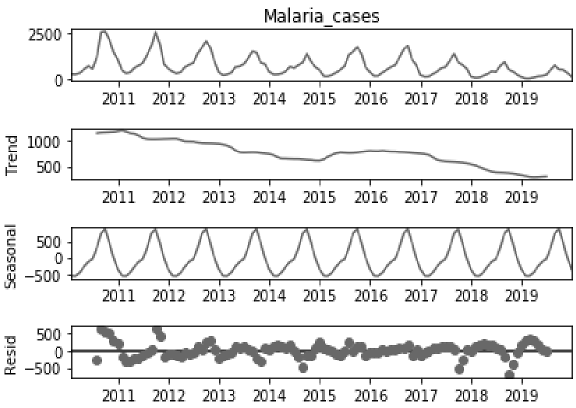

Figure 3.25 Decomposition of malaria cases time series

If we analyse the Figure-3.25, we can depict that graph is continuous through the start of the year to the end. Cases start to rise from January to September and start to decrease after September. Almost every line which represents a whole year suggests the same. Thus, we are using the SARIMA model without using any external variables. SARIMA is one of the most widely used forecasting methods for univariate time series data forecasting [5]. This method can handle the trend as well as the seasonality of the time series. Before we start training our model, we need to make sure that our data is stationary.

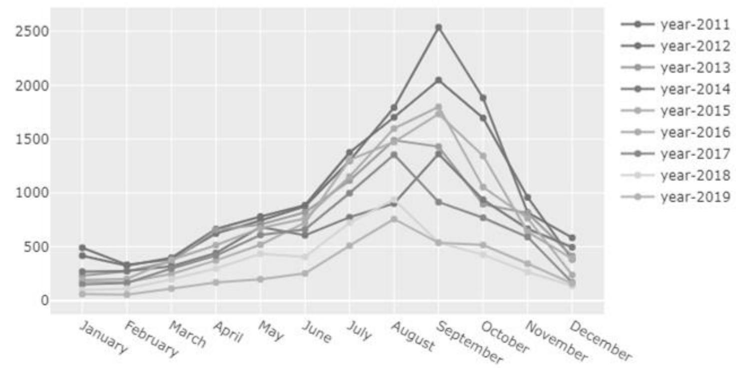

Figure 3.26 Malaria cases in last 9 years split year-wise

There are a myriad of ways to check for time series stationary but the easiest way is to check whether the mean and variance are constant over certain time periods. In our case, the time series we are dealing with does not have a constant mean over the years. One other way is through Unit test roots such as the Augmented Dickey-Fuller test [6]. There are numerous ways to make your time series stationary. We have achieved Stationary through Log Transformation [7] of the time series and then, taking a difference of the time series [8]. The change can be seen in the table-4.

Table 4. Results of Augmented Dickey-Fuller test

Results of Dickey-Fuller Test:		
	Before applying any changes	After applying log transformation and difference
Test Statistic	-0.830413	-3.556707
p-value	0.810025	0.006645
Number of Observations Used	106	106
Critical Value (1%)	-3.493602	-3.493602
Critical Value (5%)	-2.889217	-2.889217
Critical Value (10%)	-2.581533	-2.581533

From the Table-4, we can analyse that Test Statistic in the latter, is smaller than Critical Value(1%) and the p-value is smaller than the significance level of 0.05, which fundamentally means that we are rejecting the null hypothesis with the confidence level of 99%. Consequently, we can say that our time series is now stationary. The difference can be seen in the Figure 3.27.

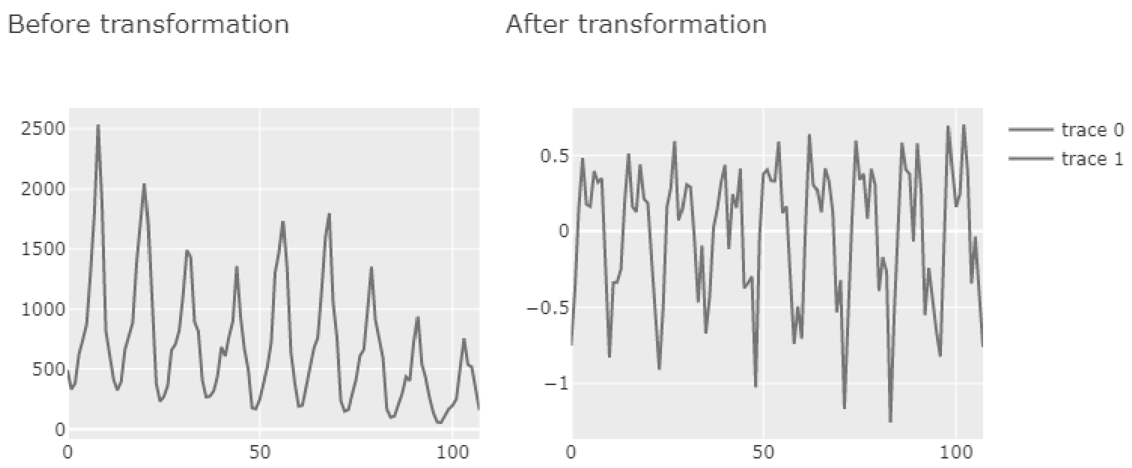

Figure 3.27 Malaria Cases time series

We can see in the figure above that the latter image has a constant mean and variance over different periods. Now, we can start predicting using the modified time series. Now, let us understand the SARIMA model letter by letter. SARIMA(p,d,q)(P,D,Q,s):

- AR(p) - Autoregression model i.e. regression of the time series onto itself. The basic assumption is that the current series values depend on its previous values with some lag. It can be determined from the PACF plot.
- MA(q) - Moving average model. Without going into detail, it helps the model to analyze the error of the time series. This can be determined using the ACF plot.
- I(d) - Order of Integration. This is a number of nonseasonal differences needed to make the time series stationary.
- S(s) - This is responsible for seasonality and it gives season period length of the given time series.
- P - Order of Autoregression for the seasonal component of the model. It can be determined from the PACF plot.
- Q - Same as P, but using ACF plot.
- D - Order of seasonal integration.

Configuring the SARIMA model requires selecting the above-mentioned hyperparameters for both trend and seasonal elements of the time series. These hyperparameters can be analyzed through ACF and PACF plots[9]. ACF stands for Autocorrelation function and PACF stands for Partial Autocorrelation function. These are plots that graphically summarize the strength of a relationship with an observation in a time series with observations at prior time steps.

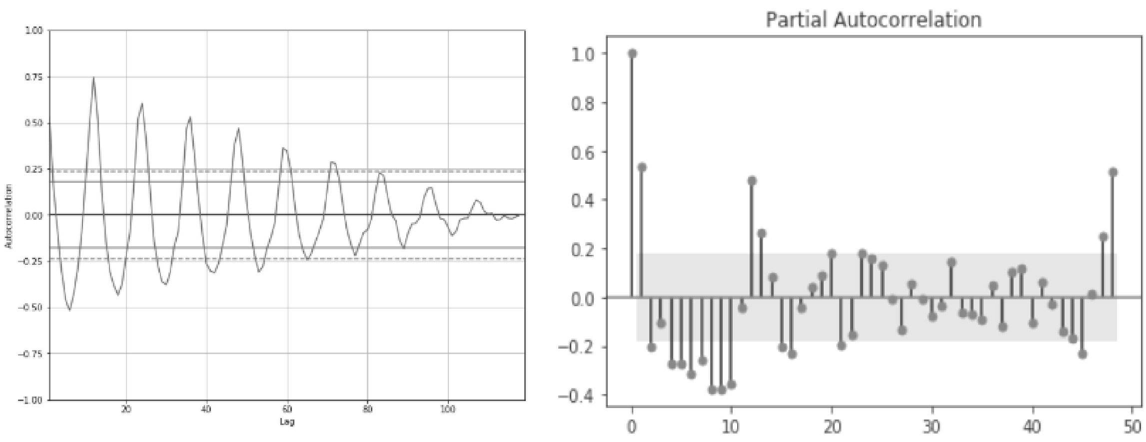

Figure 3.28 Autocorrelation and Partial Autocorrelation

Let us analyse the hyper parameters from ACF and PACF graphs:

- p is probably 1 because it is the most significant lag in the PACF graph.
- q is 1 as well because it is the most significant lag before it starts decreasing in the ACF graph.
- d should be 1 because we are taking differencing once.
- s should be 12 as we have a monthly data of malaria cases reported.

- P should be 1 as we get the second significant lag at lag number 12 in the PACF graph.
- Q should be 3 or 4 as we get significant lags at lag number 12, 24, 36 and 48 in the ACF plot.
- D is 1 because we are taking differencing once.

Thus, we get the SARIMA(1,1,1)(1,1,3,12).Using these hyper parameters in the SARIMA model, we got quite an accurate result. A comparison between actual and predicted lines of the last two years can be seen in the Figure-3.29. We kept the trend constant as our time series had no significant trend after its transformation.

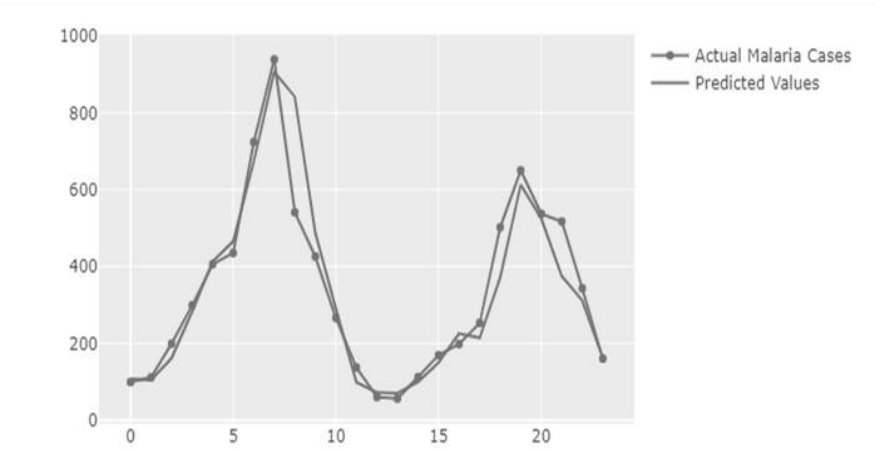

Figure 3.29 Actual cases and Predicted cases

After the prediction, it is time to check the accuracy of our model. We are using two measures to check the prediction accuracy of our model:

1. MAPE - Mean Absolute Percentage Error [10] is a statistical measure of how accurate a forecast system is. It measures this accuracy as a percentage and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. We got the MAPE of 14.0215 which means that our model was wrong by 14% on average which is quite an impressive result considering we are working on a relatively small time series and each value causes a significant change in our results.

$$\mathbf{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \qquad \begin{array}{l} A_t = \text{Actual Value} \\ F_t = \text{Forecasted Value} \end{array}$$

Figure 3.30 Mean Absolute Percentage Error

2. AIC - Akaike Information Criterion [11] can be used to determine the quality of the model. It is an estimator of the out-of-sample prediction error; a lower prediction score indicates a more predictive model. We got the AIC score of -5.398 which is a relative measure which indicates that we have taken the accurate hyper parameters to train our model.

$$AIC = -2/N * LL + 2 * k/N$$

N- Number of examples in dataset
LL- Log-likelihood of the model
K- Number of parameters in dataset

Figure 3.31 Akaike Information Criterion

Conclusion

In this report, we have only taken the weather effect into consideration. Supplementary data is required to comprehend the effect of malaria on the human immune system. There are few non-climatic factors that affect the Malaria transmission. The type of vector, the type of parasite, environmental development and urbanisation, population movement and migration, the level of immunity to malaria in the human hosts, insecticide resistance in mosquitoes, and drug resistance in parasites, all have a role in affecting the severity and incidence of malaria [18].

Malaria-free Gujarat 2022 campaign is the initiative started by the Health and family welfare department, Government of Gujarat [12]. Malaria is a major public health problem in Gujarat but is preventable and curable. Malaria interventions are highly cost-effective and demonstrate one of the highest returns on investment in public health. In regions where the disease is endemic, efforts to control and eliminate malaria are increasingly viewed as high-impact strategic investments that generate significant returns for public health, help to alleviate poverty, improve equity and contribute to overall development. Gujarat state has made significant achievements in malaria control as the state could keep the Annual Parasitic Incidence (API) less than 1.0 during the last three years. The lowest overall state API was recorded in 2015 and 2016 since 1961. Govt. of Gujarat is committed to making the state free from the burden of malaria. This commitment is reinforced by the National Framework for Malaria Elimination and also the target to be achieved in the health sector under Sustainable Development Goals. Gujarat State with good infrastructure and resources can take rapid strides in the plan to achieve malaria elimination by 2022.

We hope that our project helps the government in achieving its target. Our SARIMA model is capable of detecting a rise in the reported number of cases and can give pretty accurate forecasts that can help the government to take decisive precautionary actions in the future.

3.3.2 Typhoid Report

Typhoid Fever is a gastrointestinal infection caused by Salmonella Enterica Typhi bacteria. It is transmitted from person to person through the fecal-oral route where an infected or asymptomatic individual (who does not exhibit symptoms) with poor hand or body hygiene passes the infection to another person when handling food and water. The bacteria multiply in the intestinal tract and can spread to the bloodstream. Paratyphoid fever, a similar illness, is caused by Salmonella Enterica Paratyphi A, B, and C [19]. Typhoid fever is an important cause of avoidable mortality in regions without adequate access to safe water and sanitation. Although contaminated food and water, have been identified as the major risk factors for typhoid prevalence, a range of other factors have been reported in different endemic settings such as poor sanitation, close contact with typhoid cases or carriers, level of education, larger household size, closer location to water bodies, flooding, personal hygiene, poor lifestyle and travelling to endemic areas. In addition, climatic variables such as rainfall, Vapour pressure, and humidity have an important effect on the transmission and distribution of typhoid infections in human populations [20].

Disease Incidence and Climate Interactions

Though typhoid depends on a myriad of other factors, we are going to focus on the climate changes that affect the transmission of the bacteria that spread Typhoid. We are using the meteorological data taken from MOSDAC Meteorological and Oceanographic Satellite Data Archival Centre [14] is a data repository for the mission of the ISRO. We collected Typhoid cases from the site of Surat Municipal Corporation (SMC) [13]. Our meteorological data, including typhoid case data, is month-wise distributed. Let us start by analyzing the typhoid graph.

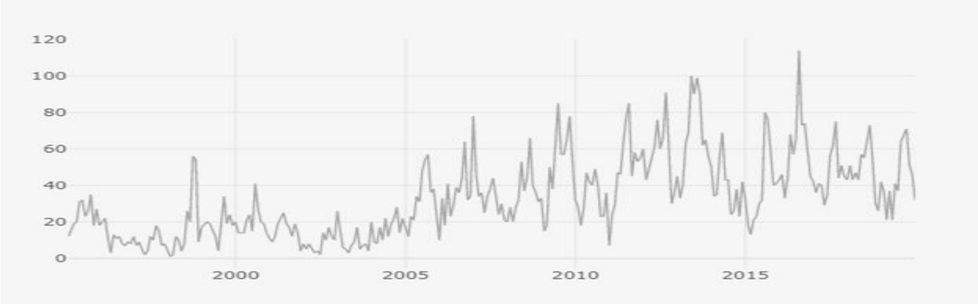

Figure 3.32 Typhoid cases 1995-2019

Here, you can depict that a significant pattern can be seen from around 2005. Before that, there was no obvious pattern to train our model on. Let us decompose our data into trend, seasonality, and residuals. We will take data from 2010 onwards.

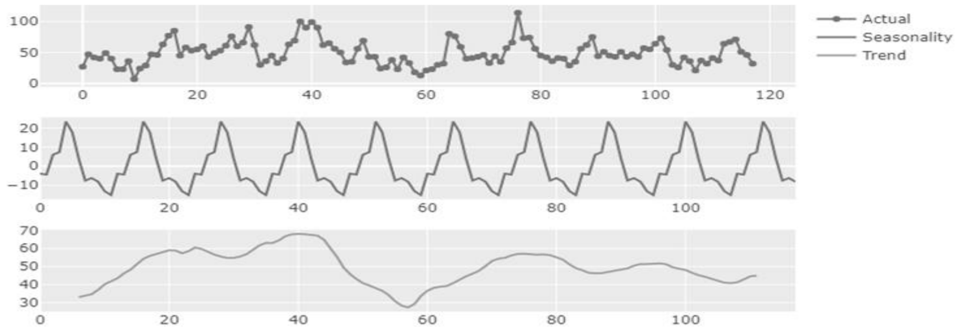

Figure 3.33 Decomposition of time series of Typhoid cases

We can say that our trend is not constant but, we have serious seasonality explaining ¼ of the actual time series. This is a positive sign because we have a significant seasonal pattern in our time series. Time series with significant seasonality are easier to model because it follows the same pattern as prior records. We still have to worry about the volatile trend, but we can deal with it later. Now that we have seen that our time series has significant seasonality, we can plot the monthly sum of the last 5 years to check seasonality throughout the year.

It can be seen from the Figure-3.34 beside that most cases are registered in the monsoon season. We can say that apart from the monsoon season, the graph is pretty balanced. Therefore, we can categorize annual typhoid data into two distinct sections. We will consider June to September as one section and other months as another. Now that we have seen the annual distribution of typhoid cases, we can start analyzing the change in climate affecting Typhoid cases reported.

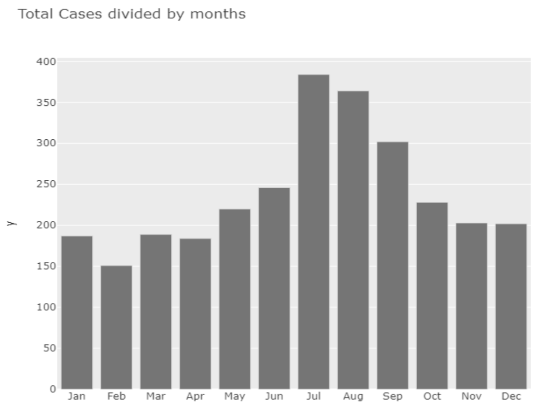

Figure 3.34 Typhoid cases total by Months

In Figure-3.35, Typhoid cases of the last 3 years along with rainfalls and rainy days can be seen. We can depict that rainfall and rain days have significant associations during the monsoon season (Jun-Sep). Furthermore, it can be said that a lag of 1 month in rainfall and rain days is more correlated with our time series compared to the same month’s rainfall and rain days. Though we require daily data to further analyze that supposition as that lag could be of three weeks or possibly 2 weeks as well.

It should be added that the incubation period for typhoid is around 1-2 weeks [21].We also analyzed that typhoid cases registered post-monsoon have little to no association with rainfall registered that year. There does not seem any variation in the number of cases registered annually with the change in annual rainfall registered. This means we can use rainfall and rain days to predict seasonality but it won't be helpful to analyze the trend of our data. Monsoon-related flooding increases the risk of waterborne diseases, including typhoid fever and related diseases such as paratyphoid and invasive non-typhoidal Salmonella disease. Clearly, the monsoon season creates the ideal climate for the transmission of typhoid: plentiful waters facilitate the growth of typhoid bacteria; intense rains can damage water and sewage lines, releasing sewage into the environment and contaminating water sources used for cleaning, cooking and drinking. The monsoon particularly exacerbates the risk of cross-contamination and causes open drains in the community to overflow with sewage, potentially contaminating otherwise clean water [22].

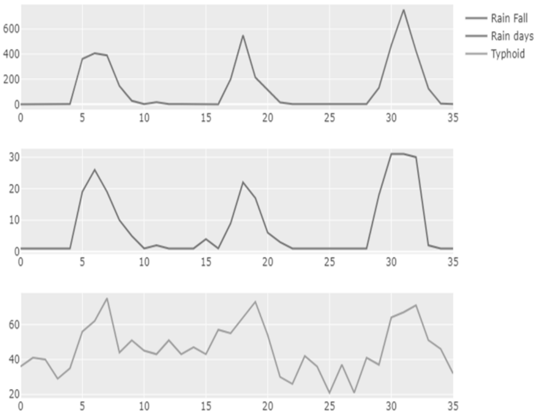

Figure 3.35 Rainfall, Rain days and Typhoid cases of last 3 years

The Figure-3.36 shows the average relative humidity graph along with relative humidity and min humidity of the last 36 months. We can see that Typhoid cases are in apparent positive correlation with relative humidity. Though there have not been any significant studies that suggest how humidity has a direct effect on typhoid cases, one report can be found here [23]. Humidity is relatively higher during monsoon because the longer it rains, the air becomes more humid. That is why we can use relative humidity to train our model. Plus, humidity is lowest during winters, which helps because typhoid cases are relatively low during winter.

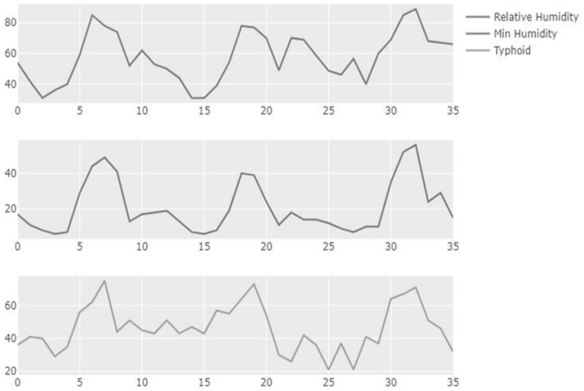

Figure 3.36 Relative Humidity, Min Humidity and Typhoid cases of last 3 years

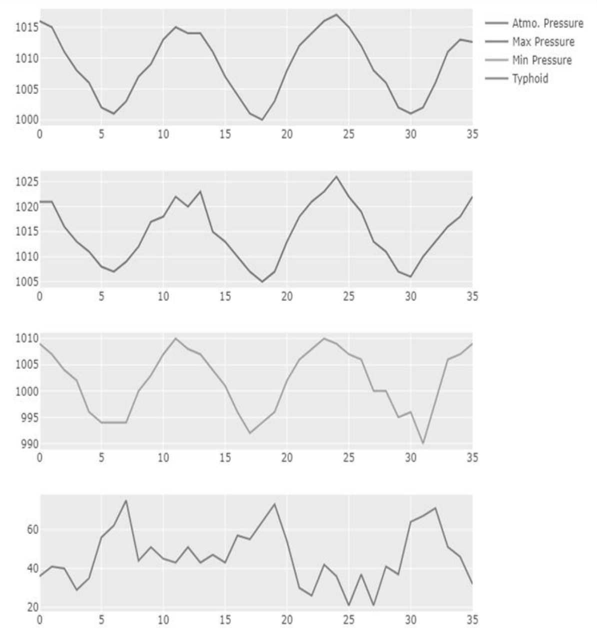

Figure 3.37 Atmo. Pressure, Max Pressure, Min Pressure and Typhoid cases of last 3 years

The Figure-3.37 shows the average atmospheric pressure, minimum pressure, maximum pressure, and typhoid cases of the last 36 months. It is evident that typhoid cases are negatively correlated with pressure. Rain comes down in varying intensities, so long, steady rain isn't always what you'll see. When a long, steady rain does happen, it's because of the location of the low-pressure system in relation to a warm front. Warm, moist air enters the area of low pressure and is pulled up and over the mass of cool air ahead of the warm front. This results in longer, steadier periods of rain. Average atmospheric pressure produces better results when we take a lag of 1 month. For which, we require daily data to analyze further.

Table 5 Pearson Correlation between Meteorological elements and Typhoid cases

Pearson Correlation between Meteorological elements and Typhoid cases:					
	Relative Humidity(%)	Rainfall	Temperature	Atmospheric Pressure	Rain Days
Typhoid Cases	0.47	0.4	0.32	-0.53	0.49

Prediction

We are using the prophet package developed by Facebook. Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. It provides us with the ability to make time-series predictions with good accuracy using simple intuitive parameters and has support for including the impact of custom seasonality. We use a decomposable time series model with three main model components: trend, seasonality, and holidays. They are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Figure 3.38 Prophet model definition

- g(t): piecewise linear or logistic growth curve for modeling non-periodic changes in time series
- s(t): periodic changes (e.g. weekly/yearly seasonality)
- h(t): effects of holidays (user-provided) with irregular schedules
- εt: error term accounts for any unusual changes not accommodated by the model

Using time as a regressor, Prophet is trying to fit several linear and nonlinear functions of time as components. The reason why we are using the prophet package and not SARIMA as we did in Malaria prediction is because the prophet allows us to configure each external regressor we use such as rainfall, rain days, etc. It allows us to configure how each regressor will affect our time series by setting their prior scales [24]. We are adding a yearly seasonality to our model and also specifying the Fourier order for added seasonality. Fourier order specifies how well we want our model to fit with our time series. A Fourier series is an expansion of a periodic function. In terms of an infinite sum of sines and cosines. Fourier series makes use of the orthogonality relationships of the sine and cosine functions [25]. We need to be careful while choosing the Fourier order because we want to avoid over fitting or under fitting our model [26].

We can also add more parameters to make our model more accurate in which we categorize our time series into two parts: Months with the monsoon season and other months. These added regressors should be in True or False form. We can also specify each month-wise parameter and their Fourier order as well. After we add these regressors, our dataset has large dimensions. The Prophet traditionally uses Monte Carlo simulation to calculate the uncertainty interval [27]. When we predict a value of something using the Prophet, we get not only the estimated value but also the lower and upper bound of the uncertainty interval. The problem with Monte Carlo is it does not perform well with high dimension dataset. We are using Markov Chain Monte Carlo (MCMC) [28] samples to train and predict. We are using Bayesian inference using MCMC samples which takes a long time to run but provides a better result in our case. Here are the actual vs. values our model predicted.

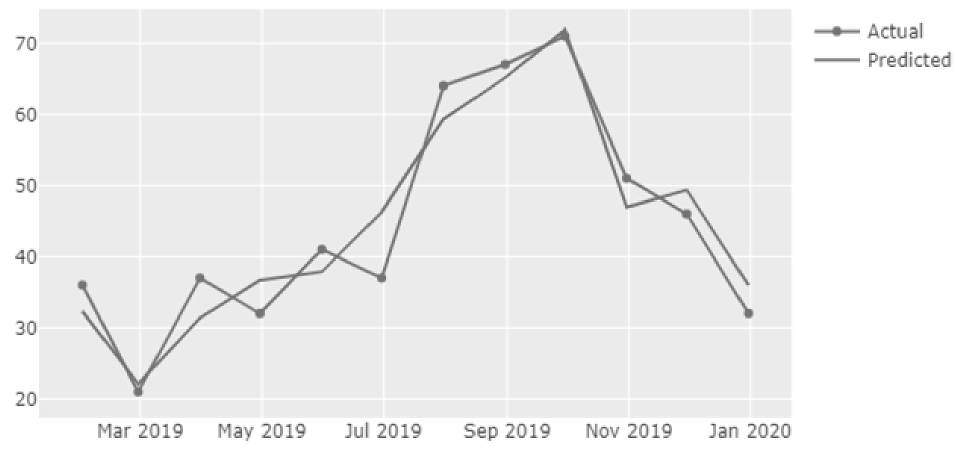

Figure 3.39 Actual Cases and Predicted cases of Typhoid of last 12 months

After the prediction, now we have to check the accuracy of our model. We are using the Mean Absolute Percentage Error (MAPE). Mean Absolute Percentage Error [10] is a statistical measure of how accurate a forecast system is. It measures this accuracy as a percentage and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. Equation for MAPE can be seen in Figure-3.30.

We got the MAPE of 9.7594 which means that our model was wrong by9.75% on average which is quite an impressive result considering we are working on a relatively small time series and each value causes a significant change in our results.

Conclusion

Though we were able to achieve a pretty good result, we still haven't considered all factors that affect Typhoid like level of education, personal hygiene, and poor lifestyle. We have only done our analysis of Surat. We were not able to get the disease data from any other cities. Our research would have been more accurate if we had more data available for other cities as well. Though we might need more accurate meteorological data from the Ahmadabad Meteorological department under which the Surat Meteorological department comes. We had In Situ data from MOSDAC but some data was still missing so we had to do some preprocessing.

Primary strategies for typhoid fever control involve safe water supply, adequate sanitation facilities, and proper hygienic practices. However, these require sustainable investments, huge financial outlays, and long-term political commitment. The introduction of the WASH program has helped to improve the WASH situation in India although there still remains a significant gap in achievements. To make the best use of the control measures in such resource-poor endemic settings, there is a need to develop disease burden extrapolation models to choose the sites that need to be prioritized for routine intervention.

There might be shortcomings or errors of fact or interpretation in our report but we hope that it helps the government in handling possible epidemic in the future.

3.3.3 Gastroenteritis Report

We tried analyzing Gastroenteritis, but due to insufficient data, we were not able to achieve adequate results. There are three major causes of gastroenteritis:

- **Viral**
  It is a cause in 60% of the cases

  Table 6 Types of Viral gastroenteritis

Types of Viral gastroenteritis	
Virus	Seasonality
Rotavirus	Predominantly in winter and occasionally in fall
Norovirus	Year-round, but especially in Winter
Sapovirus	Year-round
Astrovirus	Predominantly in Winter
Enteric adenovirus	Predominantly in Summer

  Here in Table-6, we can see that it different viruses have different seasonality. Thus, it is difficult to analyze the effect of various weather parameters with the data of monthly reported gastroenteritis cases without any knowledge about its cause.
- **Bacterial**
  10% to 20% of total cases are bacterial gastroenteritis [29]. Bacterial gastroenteritis is common in summer.
- **Parasitic**
  Another 10% to 15% of total gastro cases have parasitic cause.

Let us analyze the time series data of Gastroenteritis.

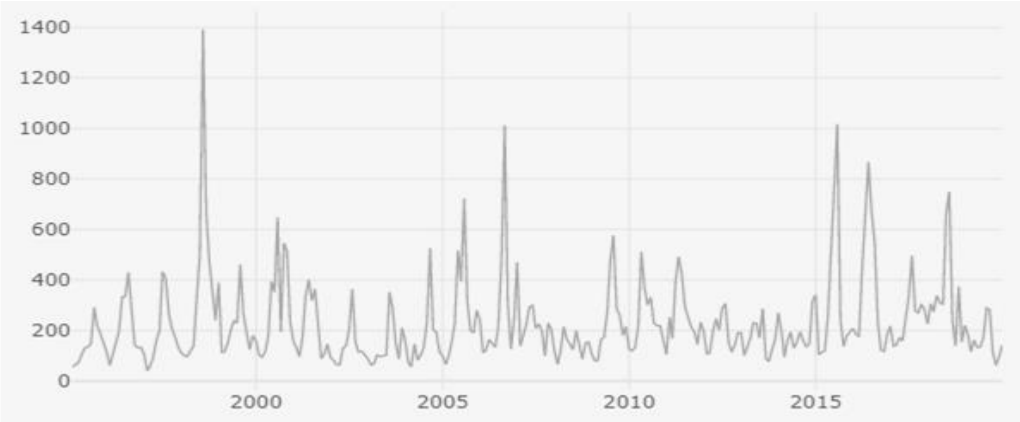

Figure 3.40 Gastro Cases 1995-2019

Here, we can see that sudden increase and decrease occur in gastroenteritis cases. Though using the boxcox transformation and the taking its difference, we were able to predict the trend of the time series but seasonality was difficult to analyze. Due to insufficient in data, we were unable to analyze the seasonality and relations with weather factors like Temperature, Humidity or Rainfall. We use many methods like RNN using LSTM, prophet, SARIMA, etc., but no acceptable result was achieved. Due to Lack of seasonality, no consistent trend or without any significant relation with other weather factors, it was extremely difficult to comprehend enough knowledge about gastroenteritis to train our model on.

References

1. SMC Vector-borne diseases control Annual Report - 2018[link]
2. SPSS-tutorials - Pearson Correlations – Quick Introduction[link]
3. Statsmodels - SARIMA[link]
4. Machine learning mastery - How to Decompose Time Series Data into Trend and Seasonality[link]
5. TowardsDataScience - How to forecast sales with Python using the SARIMA model[link]
6. Statisticshowto - what is the Augmented Dickey-Fuller test?[link]
7. NCBI - Log Transformation[link].
8. Otexts - Stationarity and differencing[link]
9. Machine learning mastery - A Gentle Introduction to Autocorrelation and Partial Autocorrelation[link]
10. StatisticsHowto- Mean absolute percentage error (MAPE)[link]
11. Machine learning mastery - Probabilistic Model Selection with AIC[link]
12. Malaria-free Gujarat 2022[link]
13. SMC - Disease records[link]
14. MOSDAC - Providing the meteorological data[link]
15. Timeanddate - Providing with additional Meteorological data[link]
16. Healthline - About Malaria[link]
17. Researchgate - Characteristics and trends of malaria in Surat district of Gujarat: a hospital-based study[link]
18. Open.edu - Communicable Diseases Module: 6. Factors that Affect Malaria Transmission[link]
19. IAMAT - India recommended vaccination - Typhoid fever[link]
20. Ashraf M. Dewan, Rober Corner, Masahiro Hashizume, Emmanuel T. Ongee (2013)- Typhoid Fever and Its Association with Environmental Factors in the Dhaka Metropolitan Area of Bangladesh: A Spatial and Time-Series Approach[link]
21. HealthVic - Typhoid and paratyphoid[link]
22. Dipika Sur, Mohammad Ali - Comparisons of predictors for typhoid and paratyphoid fever in Kolkata, India[link]
23. Times of India - Cases of typhoid on the rise as weather gets humid[link]
24. Prophet - Seasonality, Holiday Effects, And Regressors[link]
25. Mathworld - Fourier Series[link]
26. Machine learning mastery - Overfitting and Underfitting With Machine Learning Algorithms[link]
27. Bartosz Mikulski - Understanding uncertainty intervals generated by Prophet[link]
28. Machine Learning Mastery - A Gentle Introduction to Markov Chain Monte Carlo for Probability[link]
29. Contemporary Clinic - Best Approaches to Acute Gastroenteritis[link]
30. Literature review(comparing with current available application Surat Municipal Corporation App)[link] -
31. Medium - Prediction of epidemic disease dynamics using machine learning[link]
32. BMC infect disease - Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economicdata[link]
33. Researchgate - Malaria Outbreak Prediction Model Using Machine Learning[link]

34. NHP - Literature review(comparing with current available application NHP Health Directory Service App)[link]
35. NHP - Literature review(comparing with current available application NHP Swasth Bharat App)[link]
36. NHP - Literature review(comparing with current available application NHP MeraAspataal)[link]
37. The skeptical chemist - Predicting Viral Outbreaks With Machine Learning[link]
38. Google Maps Platform – Google places API[link]
39. SMC – Draft Budget 2019-2020[link]