Forecasting In-Patient & Day Case Waiting Times in Ireland

Ankit Malik
Master of Computing
Dublin City University
Dublin, Ireland
ankit.malik3@mail.dcu.ie

Hao Io Cheong
Master of Computing
Dublin City University
Dublin, Ireland
hao.cheong3@mail.dcu.ie

Sakthi Somaskandan
Master of Computing
Dublin City University
Dublin, Ireland
sakthignana.somaskandan2@mail.dcu.ie

Abstract – This research paper aims at predicting the waiting times for patients across Ireland for In-Patient and Day Case facilities. Data used for forecasting is from NPTF and spans over 8 years to provide a thorough view. There can be a number of precedented and unprecedented parameters which can affect wait time, however, for the purposes of this study, historic wait time data were chosen. Based on literature reviews and related work in this domain, several time series algorithms such as ARIMA, SARIMA, Prophet and TBATS were evaluated for this study, out of which ARIMA model produced the most accurate results.

keyword: ARIMA, Time series, SARIMA, TBATS, Hospitals waiting list.

I. Introduction

Today we live in a world where food, daily needs, consumer goods and electronics are delivered to your doorstep in a single touch on your smart phone. However, there is one crucial area where not enough progress has been made – providing health care in times of need. According to HSE data each year 1 million people have a planned day case procedure and approximately 94,000 patients have an elective inpatient procedure [1]. The National Treatment Purchase Fund has released waiting time data across In-Patient, Outpatient & Day Case facilities. The area focused on this study is limited only to In-Patient & Day Case. Historic In-patient & day case data informs that waiting times can range up to 20 months for certain specialties and age groups. The study analyses the data and targets forecasting of waiting times based on historic data available from hospitals for multiple specialties and age-groups in Ireland. The research also analyses time bands to understand the highest volume of patients impacted.

A. Scope

The hospital waiting time although spans across four areas such as:

- In-Patient: Patient who stays overnight for treatment
- Out-Patient: Patient who visits for a short appointment which could be a consultation or test
- Day Case: Patient who stays for a single day treatment
- Emergency Department

To keep the study effective given the time frame and huge volume of data, the scope would be limited to In-Patient and Day Case only.

The scope is limited to using historic data only without any external supporting dataset.

B. Objective & Research Question

The objective of the study is to define a foundational model for forecasting waiting times for in-patient and day case facilities across hospitals for which treatment is required based on historic data.

The forecasting model identified by this study will then be used to extend the research by incorporating additional related data to increase the efficacy of the model and provide a holistic time series forecasting solution.

The aim of current research is to answer the below questions:

1. Identify which time series forecasting model is most suitable for predicting in-patient & day case waiting time

2. Study existing literature to understand the work done in this area and discover other factors involved in the evaluation of waiting times

C. Dataset

The dataset used for this research has been provided by the National Treatment Purchase Fund (NTPF) who work closely with public & private hospitals alongside HSE and Department of Health. The data is captured for every hospital across all specialties and categorized by age along with time banding the waiting time providing the count of patients in these bands by their age groups. The study has been chosen to span over 8 years to have meaningful insight [2].

The raw data itself is tabular and hence in readable format. For transparency, a record count break-up is displayed below.

#	Year	Record Count
1	2014	35049
2	2015	43884
3	2016	49695
4	2017	54303
5	2018	55377
6	2019	52415
7	2020	58457
8	2021	15887

Figure 1 - Dataset record count break-up

D. Methodology

The Methodology proposed for this study is Cross-Industry Standard Process for Data Mining as this allows to transition between stages in reverse order and allows to plan the research in a structured manner. This has been discussed further in the paper.

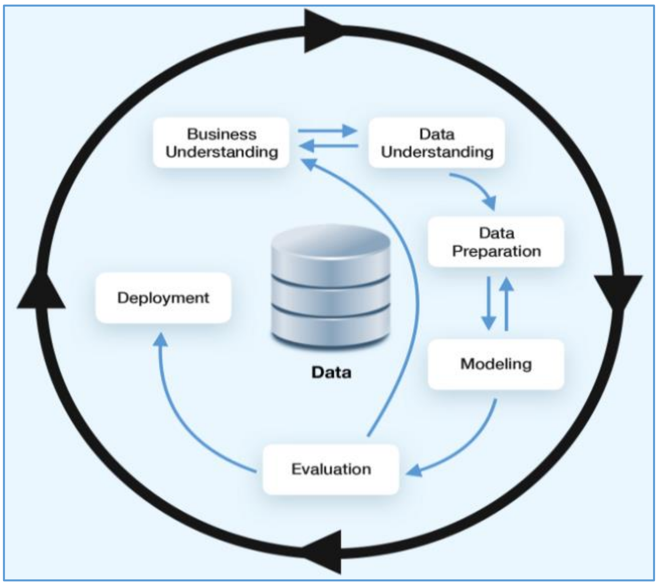

Figure 2 - CRISP-DM Methodology

II. RELATED WORK

The research has been divided into three sections starting from the identification of various forecasting techniques that could be useful in this scenario and then gathering further insight on the factors affecting in-patient and day case waiting times upon which a critical analysis is required to define the parameters needed to predict accurately.

A. Forecasting Technique

We start by understanding time series forecasting and various models and we progress this study as a [3] quantitative forecasting where "data with patterns is available" and understand the various data decomposition techniques and learn that STL decomposition would be most suitable for our study as classical approach of additive and multiplicative decomposition discounts [4] first few and last few observations from the trend cycle, assumes that the seasonal component is constant throughout the entire series and is not responsive to sharp fluctuations. [5] STL decomposition on other hand handles any type of seasonality, the user can control the rate of change of the seasonal component and is robust to outliers. [6] Before moving on to forecasting implementation there was a study done on evaluation metrics such as:

- MAE (mean absolute error)
- RMSE (root mean squared error) $\sqrt{\sum_{i=1}^{N} w_i'(n_i - \overline{\mu}_n)^2}$
- MAPE (mean absolute percent error)

All three values for a good model are expected to be low. [7] We also review forecasting model techniques starting from classical forecasting methods such as Naive method, the Simple average method and the simple moving average method autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model and autoregressive integrated moving average (ARIMA) model.

Other models that were also explored in this analysis were Seasonal autoregressive integrated moving average (SARIMA).

Model	Characteristic	PACF correlogram	ACF correlogram	Data haracteristic
AR(p)	1. y_t depends on its own past values 2. P is computed using PACF function	Spikes till p^{th} lag then cuts off to zero	Spikes then decays to zero	Data should be stationary in nature
MA(q)	1. y_t depends on error term which follows a white noise process 2. q is computed using ACF function	Spikes then decays to zero	Spikes till p^{th} lag then cuts off to zero	Data should be stationary in nature
ARMA (p, q)	1. ARMA = AR+MA 2. Value of p and q are determined using AIC criteria	Spikes then decays to zero	Spikes then decays to zero	Data should be stationary in nature
ARIMA (p, d, q)	1. Data is made stationary by differencing it 2. Box-Jenkins approach is used to determine model	Spikes then decays to zero	Spikes then decays to zero	Data should be non-stationary in nature

Figure 3 - Characteristic of commonly used methods

B. Factors affecting waiting times

In parallel to identifying the most suitable time series algorithm there was research done on possible factors that could affect in-patient and day case waiting times. [8] A recent report published by Sinn Féin (an Irish republican political party) on "Understanding the Causes of Hospital Waiting Lists" it's depicted that the major causes of waiting lists are:

- Lack of Staff
- Remoteness of Location
- Limitation of services in certain areas
- Long recruitment process

Unplanned circumstances such as the COVID-19 pandemic has also caused a huge spike in the waiting list and is also acknowledged by the [9] Department of Health in their report "Health in Ireland: Key Trends 2021" saying that "There is an increase of over 25% from March 2020"

The Irish medical organisation which represents doctors in Ireland [10] have also acknowledged the report and have strongly responded with their feedback highlighting gaps in the proposed policies to lower waiting times. Other research also suggest that Insurance status is also a factor impacting the time it takes for a patient to receive medical service. [11] As mentioned in the paper, "A higher percentage of those with PHI were waiting less than three months relative to those without PHI.".

C. Predicting Accurately

After understanding all the factors, it raises the question whether predicting waiting time based on historic data is holistic or sufficient enough to produce an accurate forecasting model. Based on the analysis we have concluded that the research needs to be extended post defining a baseline model using only historic data. The baseline model can then be expanded to include more factors as we identify additional associated data such as:

- Average increase in waiting time due to COVID-19
- Average Staff to patient ratio across speciality for all hospital groups
- Introducing county as a dimension to predict waiting time by area

Analysis was also done to identify approach for reducing forecasting errors [11], [12] from evidence-based checklist which uses “forecasting methods consistent with forecasting principles and have been shown to provide out-of-sample forecasts with superior accuracy”.

In the essence of time, this study will aim to produce the baseline forecasting model on the historic data as that forms the backbone of the prediction and trend analysis.

III. Data Mining Implementation

The CRIPS-DM methodology is used to progress this research involving following steps:

Business Understanding: The aim of the research is to provide a forecast model to predict waiting times by age-group and specialty for which treatment is required.

Data Understanding: The data collected from NTPF is described below:

Archive Date	Date when the data was captured
Hospital Group	Group to which the hospital belongs
Hospital HIPE	Hospital In-patient Enquiry number
Hospital	Name of the Hospital
Specialty HIPE	Specialty In-patient Enquiry number
Specialty	Name of Specialty
Case Type	Classification of In-Patient or Day Case
Adult/Child	Classification of Adult and Child
Age Categorization	Age category of patient
Time Bands	Waiting time band for availing service
Count	Number of patients

Figure 4 - Dataset Attributes

Data Preparation: There are 8 years of hospital waiting time data used in this report. Since each individual dataset attributes are worded differently, the attribute names were renamed and unified accordingly before merging all datasets into one. Based on the project objective we have then analyzed the training dataset.

Data Pre-processing: Once merging process is completed, next step is to clean and get the dataset prepare for further modelling evaluation process.

1. Develop a function to avoid any case sensitive issue, by establish a function to set “Time Band” into lower case.
2. Consolidate all “Archive Date” related data rows and unify one row only for per date.
3. Shorten all time band related attributes into “0-3 Months”, “3-6 Months”, “6-9 Months” etc.
4. Dropping all unused columns and only keeping “date” and “patients” columns to prepare for upcoming time series forecasting evaluation processes.
5. Splitting into training and test data sets.

Data Selection: The following datasets were agreed as part of data selection activity -

- Training Data Records range from 2014 – 2016.
- Test Data records range from 2016 – 2018.
- Attributes selected for modelling: Archive Date, Count

Data Cleaning: Data from 2021 was removed as part of the cleansing activity due to insufficient volume. Upon analysis there was no missing data identified in the test data. Data was also aggregated by time-bands and to aggregate number of patients by archive date. This was done to make decomposition analysis easier which is explained further.

Data Formatting: Data was sorted by archive date to feed to time series modelling.

Data Exploration helped us understand the trend in waiting time from 2014 to 2020.

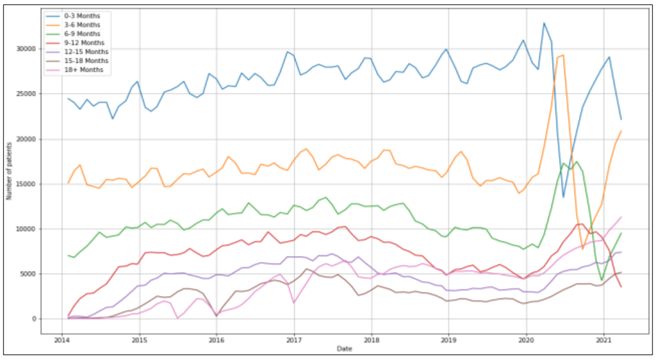

Figure 5 - Aggregated Plot

Data Decomposition was done to identify seasonality, stationarity, autocorrelation in the data.

- Seasonality refers to data fluctuations over different calendar cycles.
- Trend component provides the overall trend of time series data.
- Cycle Component helps to understand the non-seasonal decreasing and increasing pattern
- Noise component informs random fluctuations in the data

There are two varieties of techniques in decomposition – additive and multiplicative. Additive decomposition is preferred where there the variation in seasonality is relatively constant whereas in multiplicative decomposition is when the variation increases with time.

A. Seasonality

The study builds on top of STL (Seasonal and Trend decomposition using Loess) decomposition which is additive in nature as the seasonality observed is not varying with time.

Seasonality was observed across all time bands.

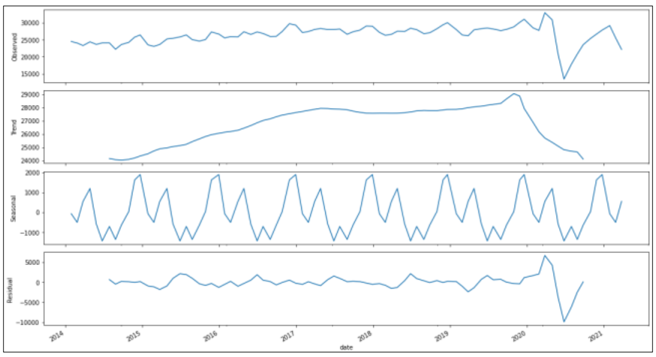

Figure 6- Decomposition for patients in waiting time band 0-3 months

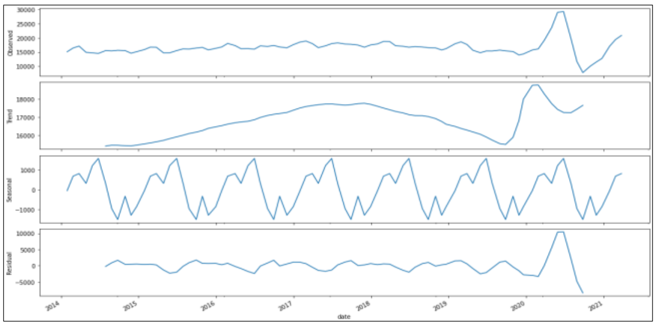

Figure 7- Decomposition for patients in waiting time band 3-6 months

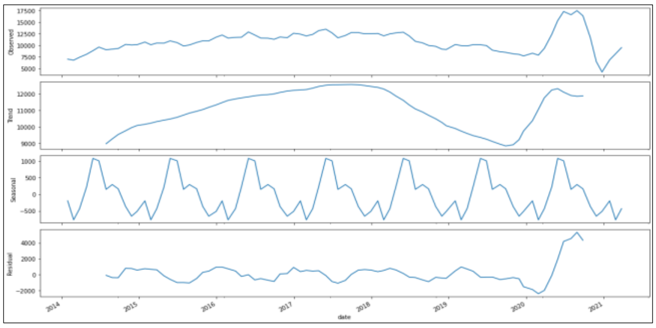

Figure 8- Decomposition for patients in waiting time band 6-9 months

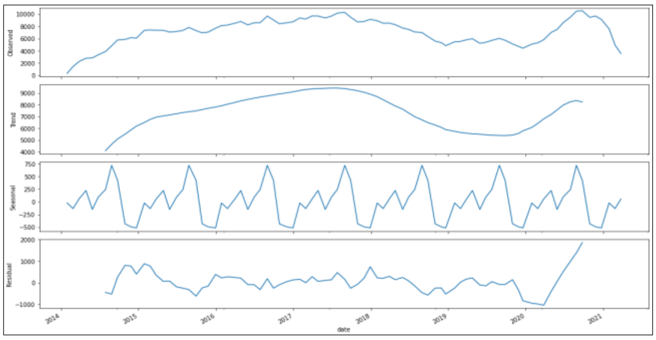

Figure 9 - Decomposition for patients in waiting time band 9-12 months

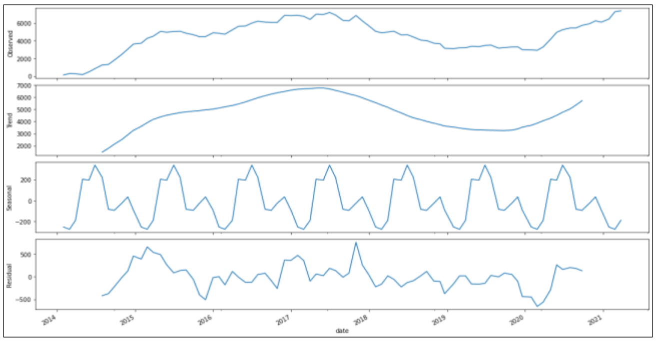

Figure 10 - Decomposition for patients in waiting time band 12-15 months

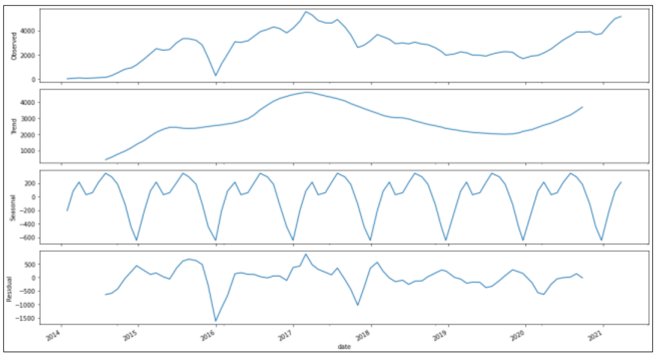

Figure 11 - Decomposition for patients in waiting time band 15-18 months

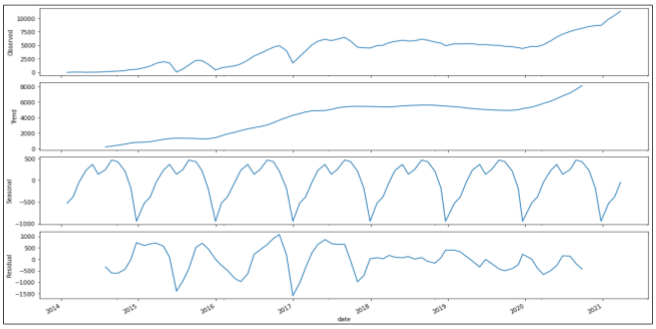

Figure 12 - Decomposition for patients in waiting time band 18+ months

B. *Stationarity*

Stationarity tells if a time series has its mean, variance and co-variance independent of time. Stationarity can be tested by applying the Augmented Dickey-Fuller test (ADF). It tests the null hypothesis that a unit root is present. This is measured by the p-value.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

If the p-value > 0.05, the process is not stationary
If the p-value = 0, the process is considered stationary and the null hypothesis is rejected.

The ADF test confirms that the process below is not stationary as the mean is not constant through time. P-values for all time series plots are mentioned below:

- P-value for waiting time band 0-3 months: 0.01751
- P-value for waiting time band 3-6 months: 0.06036
- P-value for waiting time band 6-9 months: 0.21890
- P-value for waiting time band 9-12 months: 0.51775
- P-value for waiting time band 12-15 months: 0.31246
- P-value for waiting time band 15-18 months: 0.24616
- P-value for waiting time band 18+ months: 0.97205

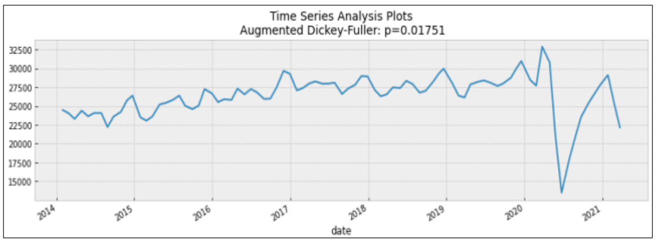

Figure 13 - Time Series Analysis Plot for time band 0-3 months

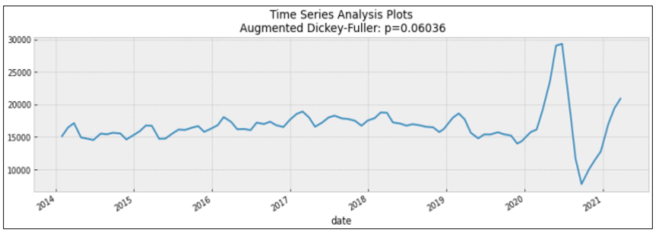

Figure 14 - Time Series Analysis Plot for time band 3-6 months

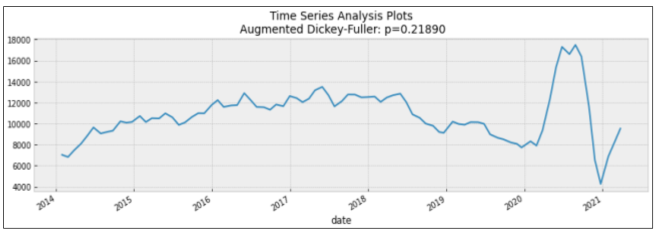

Figure 15 - Time Series Analysis Plot for time band 6-9 months

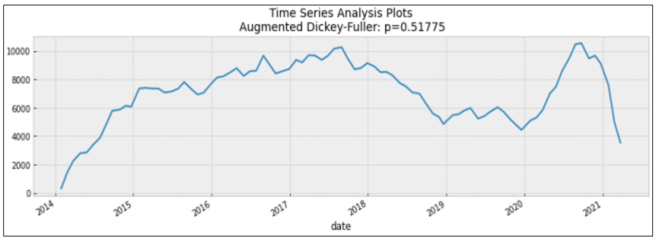

Figure 16 - Time Series Analysis Plot for time band 9-12 months

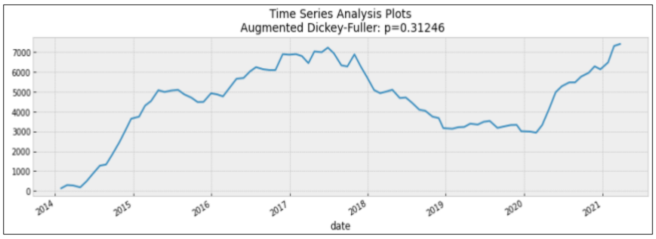

Figure 17 - Time Series Analysis Plot for time band 12-15 months

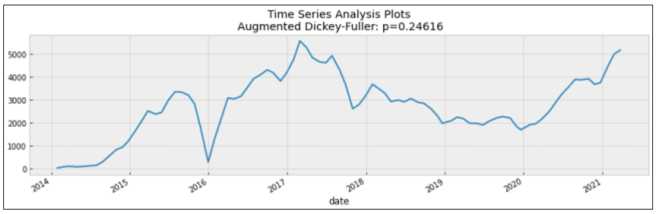

Figure 18 - Time Series Analysis Plot for time band 12-15 months

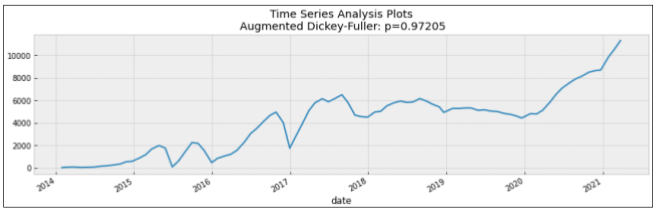

Figure 19 - Time Series Analysis Plot for time band 18+ months

C. Autocorelation

Auto Correlation Function (ACF) explains how the current value of a given time series is correlated with the past values in the unit of time (1-unit past, 2-unit past, ..., n-unit past). This autocorrelation value can vary between +1 & -1

- x-axis - correlation coefficient
- y-axis - number of lags

Partial Auto Correlation Function (PACF) explains the partial correlation between the series and lags of itself.

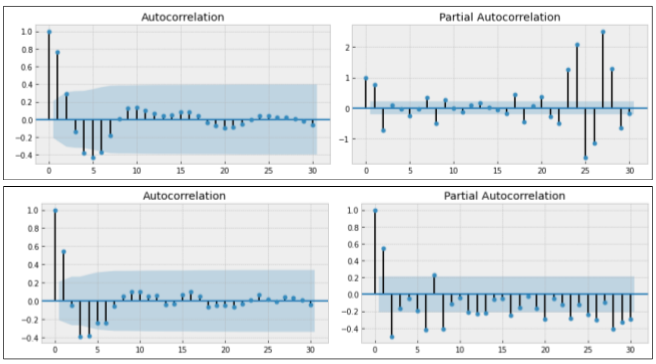

Figure 20 - Autocorrelation Plot for time band 3-6 months with 1 unit lag

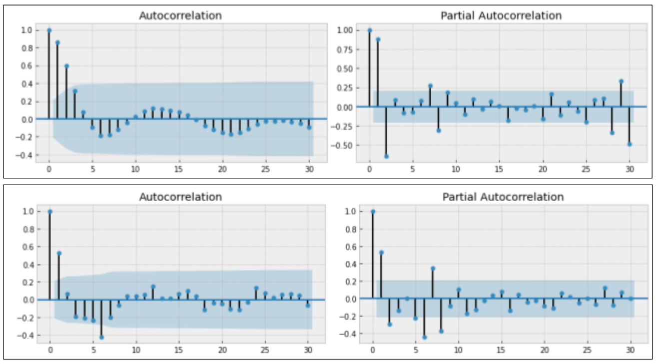

Figure 21 - Autocorrelation Plot for time band 6-9 months with 1 unit lag

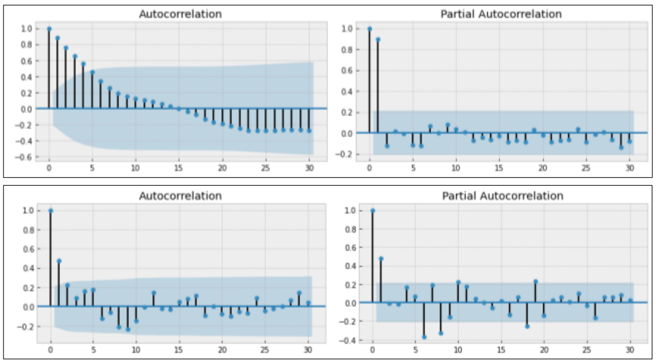

Figure 22 - Autocorrelation Plot for time band 9-12 months with 1 unit lag

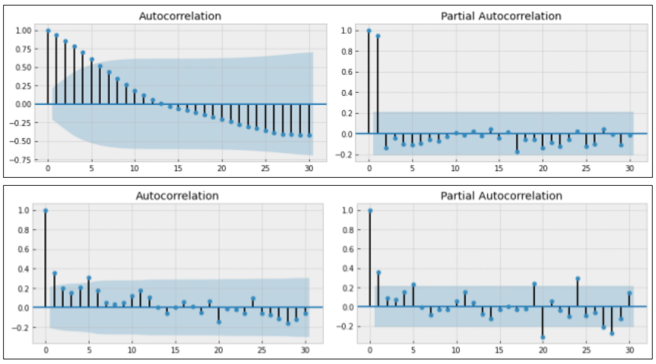

Figure 23 - Autocorrelation Plot for time band 12-15 month with 1 unit lag

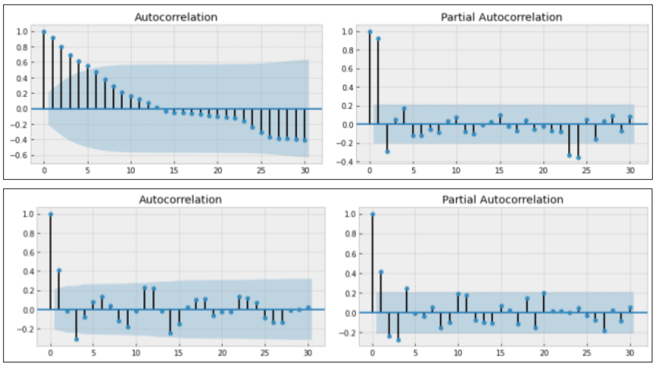

Figure 24 - Autocorrelation Plot for time band 18+ months with 1 unit lag

D. Modelling

There were multiple time series model considered out of which few mentioned below were shortlisted to progress with the study [13]:

- ARIMA
- SARIMA
- TBATS

1) ARIMA

Auto-retrogressive integrated moving average also known as ARIMA is a time series forecasting model which consists of three major components or parameters:

- Auto-regression (AR): Relationship between an observation and its lagged observation
- Integrated (I): Subtracting an observation from a lagged observation to make the time series stationary
- Moving Average (MA): Series of average calculated from historic data

These parameters are defined as ARIMA(p,d,q):

- Lag Order (p): Number of lag observations
- Degree of differencing (d): Number of times raw observations are differenced
- Order of Moving Average (q): Size of moving average window

In order to define an ARIMA model [14], the methodology suggested by Box-Jenkins is used which suggests to breakdown the process as follows:

- Identification: Estimation of values – p,d,q using auto-correlation and partial correlation. Grid search method was used to identify the most suitable values for these parameters.
- Estimation: Evaluate the coefficient of regression model. Based on multiple iterations the combination with the lowest root mean squared error values is selected
- Diagnosis: Run statistical tests on residual errors to evaluate the model

2) SARIMA

Seasonal Auto-Regressive Integrated Moving Average is an extension of the ARIMA model with seasonality as the differentiating factor. It is measured by P,D,Q,s parameters where P is the lag order for the seasonality, D is the Seasonal Integration order, Q is the moving average window of the seasonality and s is seasonal periodicity. Again P, D, Q is identified using grid search method and the seasonal periodicity of our data is 12 as the data is aggregated by month.

3) TBATS Model

TBATS model is used for time series forecasting for seasonal data and is inspired from exponential smoothing. TBATS is short for Trigonometric Box Cox ARMA trend seasonality. For our study we have implemented TBATS as

TBATS(seasonal_periods=[12], use_box_cox=True, use_arma_errors=True, use_damped_trend=True, use_trend=True)

The final model will be chosen based on the Akaike information criterion (AIC) which is a mathematical method to understand how well a model fits the data it is generated from.

IV. Evaluation

This research utilizes several forecasting models to evaluate the best fit foundational model for further analysis and examination. Due to the complexity of the data with seasonality three models were finalized such as Auto-retrogressive integrated moving average also known as ARIMA with grid search method to identify parameters, Seasonal Auto-Regressive Integrated Moving Average (SARIMA) which accounts for the seasonality component of the data and finally TBATS. The performance of these models is evaluated by selecting the lowest value of Root mean squared error (RMSE). A reference for the time band 0-3 month is given below:

Model	RMSE Value	MAPE
ARIMA	5580	0.205
SARIMA	5919.37	0.217
TBATS	5905.44	0.215

For running the ARIMA model, grid search method was used to identify the best fit parameters for generating forecast. Grid search method was iteratively executed for multiple combinations of parameters ranging from (0,0,0) to (10,2,2) with highest to low values of RMSE ranging from 1857.03 to 696.48. ARIMA (4,0,1) turned out to be the most suitable with a rounded RMSE value of 696.48.

Post grid search, ARIMA (4,0,1) was plotted for residuals and a forecast was run in comparison with actual waiting times. ARIMA was identified as the most performant model when compared with others which had the RMSE value of 5580 and MAPE of 0.205.

The RMSE value for TBATS & SARIMA model was found to be significantly high and hence will not be considered for future work unless new evidence emerges.

RMSE & MAPE values for other time bands also suggest similar pattern and suitability for ARIMA model except when comparing for time band 15-18 months which shows TBATS model as a better implementation. Few of the time-bands and there corresponding RMSE & MAPE values are shown below with graph plots.

The SARIMA model for the 15-18 month time band has extremely high values of RMSE and hence will not be considered for any future analysis. On the contrary, the TBATS model performed significantly well compared to the ARIMA model when was the most performant across all other time bands.

There was also consideration given to the Prophet model which is an open-source model shared by Facebook but was not analyzed further due to time restrictions.

Model	0-3 Month		6-9 Month		15-18 Month	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ARIMA	5580	0.205	5451	0.297	1510	0.326
SARIMA	5919.37	0.217	8521.6	0.615	5313304	1649.4
TBATS	5905.44	0.215	4292.53	0.327	968.9	0.255

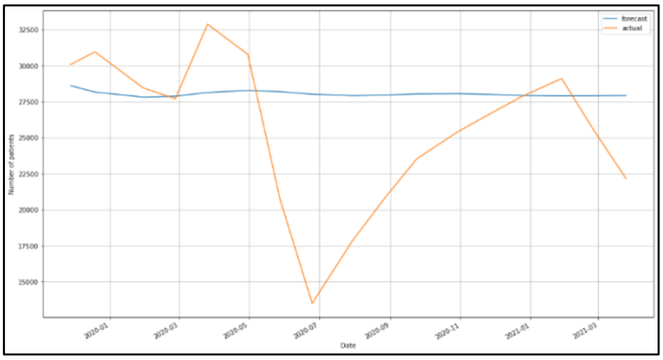

Figure 25 - ARIMA Model Forecast for 0–3 month time band

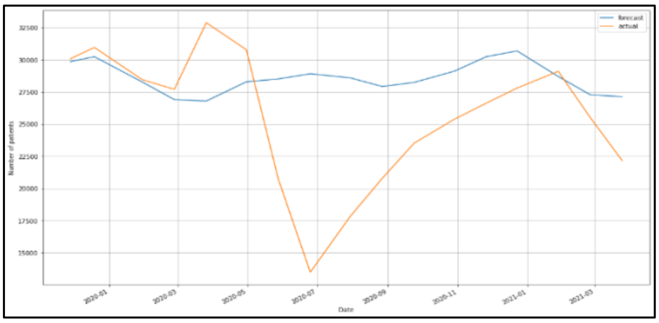

Figure 26 - SARIMA Model Forecast for 0-3 month time band

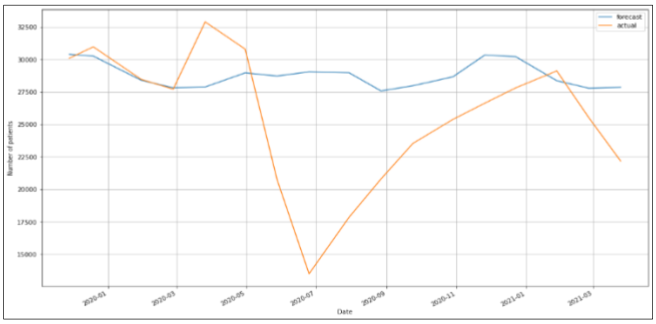

Figure 27 - TBATS Model Forecast for 0-3 month time band

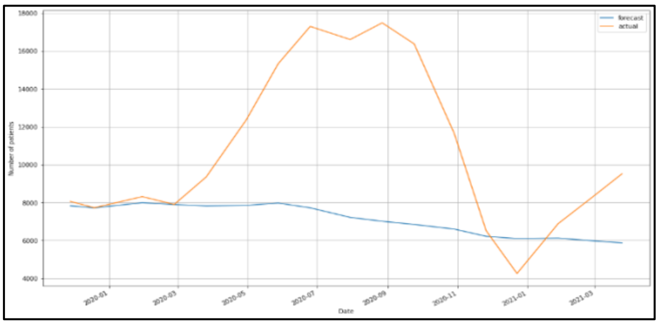

Figure 28 - ARIMA Model Forecast for 6-9 month time band

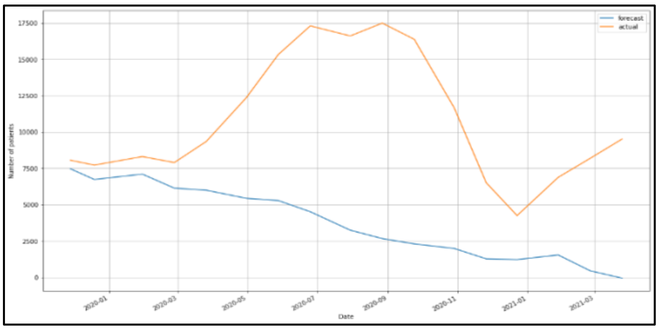

Figure 29 - SARIMA Model Forecast for 6-9 month time band

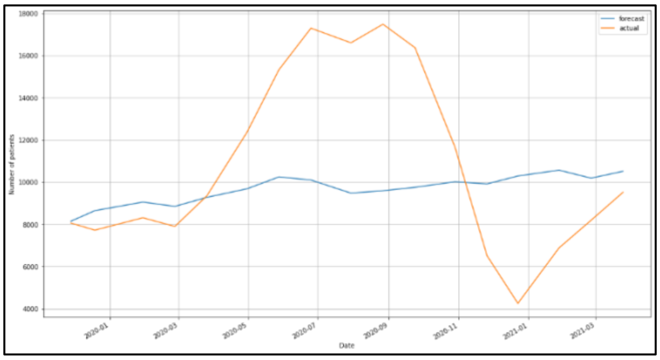

Figure 30 - TBATS Model Forecast for 6-9 month time band

V. Conclusion

Upon analysis of various time series model and techniques we realized that ARIMA model has the most overall accuracy of predicting the in-patient & day case waiting time.

Although, additional factors were discovered that could be responsible for influencing waiting times. It was understood that historical data alone is not sufficient to accurately predict future waiting times. As an example, if we slip into another pandemic in the future, the waiting time prediction will not be accurate based on this model. To accurately predict this, we will need additional data such as:

- Average increase in waiting times during pandemic
- Average increase or decrease in waiting time due to changing staff capacity
- Average Staff to patient ratio

Also being cognizant of the fact that every situation is different and there can be new emerging factors in the future which we either have not considered or are unknown yet. This also cannot confirm 100% accuracy, but we can get closer to reality by including this.

As an extension of this study, we can alter the forecasting model for a targeted specialty or age-group e.g., elder people are likely to need healthcare access earlier than younger people excluding emergency cases. We can use the data to target specifically the 65+ age group and even extend it for most impacted or critical specialties such as Neurosurgery, Vascular Surgery and Cardiology.

We would like to also give a second look to the model and run more statistical tests and explore few more time series models such as NARNN, hybrid SARIMA-NARNN to forecast in-patient waiting times. Once the model is stabilized, we can also apply this for outpatient and emergency waiting times after aligning the exploratory data analysis with existing assumptions. We would like to continue the research and explore more data associated with other factors to increase precision as this can be beneficial not only for policy making & introducing reforms by government bodies but can also be incorporated by private healthcare providers to provide transparent information to their customers.

VI. Project Work

GitHub Repository Link: https://github.com/sur-sakthy/da-dm-project.git [15], [16], [17], [18], [19] Presentation Link: https://drive.google.com/file/d/1CEo3m1bLNp0piHtXfmeq5c_T0yNeypP9/view?usp=sharing

VII. References

[1] Z. Ouyang, P. Ravier, and J. Meryem, “STL Decomposition of Time Series Can Benefit Forecasting Done by Statistical Methods but Not by Machine Learning Ones,” *Eng. Proc.*, vol. 5, p. 42, Jul. 2021, doi: 10.3390/engproc2021005042.

[2] Data.gov.ie, “Inpatient/Day Case Waiting List 2014 - Mar 2021 - Datasets,” *The National Treatment Purchase Fund*, 2021. https://data.gov.ie/dataset/ipdc-waiting-list-by-group-hospital (accessed Apr. 07, 2022).

[3] P. P. Pathak, “Time Series Forecasting — A Complete Guide | by Puja P. Pathak | Analytics Vidhya | Medium,” *Analytics Vidhya*, 2021. https://medium.com/analytics-vidhya/time-series-forecasting-a-complete-guide-d963142da33f (accessed Apr. 07, 2022).

[4] A. Plummer, “Different Types of Time Series Decomposition | by Andrew Plummer | Towards Data Science,” *Towards Data Science*, 2020. https://towardsdatascience.com/different-types-of-time-series-decomposition-396c09f92693 (accessed Apr. 07, 2022).

[5] I. Cleveland, Robert B; Cleveland, William S; Terpenning, “STL A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion).pdf,” *J. Off. Stat.*, 1990.

[6] R. A. Collenteur, “How Good Is Your Model Fit? Weighted Goodness-of-Fit Metrics for Irregular Time Series,” *Groundwater*, vol. 59, no. 4, pp. 474–478, Jul. 2021, doi: 10.1111/GWAT.13111.

[7] Z. Liu, Z. Zhu, J. Gao, and C. Xu, “Forecast Methods for Time Series Data: A Survey,” *IEEE Access*, vol. 9, pp. 91896–91912, 2021, doi: 10.1109/ACCESS.2021.3091162.

[8] “Understanding the Causes of Hospital Waiting Lists,” *Sinn Féin*, 2021. https://www.sinnfein.ie/contents/62699 (accessed Apr. 07, 2022).

[9] “Health in Ireland: Key Trends 2021,” *Government of Ireland*, 2021. https://www.gov.ie/en/publication/350b7-health-in-ireland-key-trends-2021/# (accessed Apr. 07, 2022).

[10] “IMO responds to Government Plan for Waiting Lists,” *Irish Medical Organisation*, 2022. https://www.imo.ie/news-media/news-press-releases/2022/imo-responds-to-governmen/index.xml (accessed Apr. 07, 2022).

[11] R. Whyte, S. Connolly, and M. A. Wren, “Insurance status and waiting times for hospital-based services in Ireland,” *Health Policy (New. York).*, vol. 124, no. 11, pp. 1174–1181, Nov. 2020, doi: 10.1016/J.HEALTHPOL.2020.07.001.

[12] J. Armstrong and K. Green, “Forecasting methods and principles: Evidence-based checklists,” *J. Glob. Sch. Mark. Sci.*, vol. 28, pp. 103–159, Apr. 2018, doi: 10.1080/21639159.2018.1441735.

[13] A. S. Bharatpur, *A LITERATURE REVIEW ON TIME SERIES FORECASTING METHODS*. 2022.

[14] M. Din, *ARIMA by Box Jenkins Methodology for Estimation and Forecasting Models in Higher Education*. 2015. doi: 10.13140/RG.2.1.1259.6888.

[15] Nadeem, “Time Series Forecasting using TBATS Model | by Nadeem | Analytics Vidhya | Medium,” *Analytics Vidhya*, 2021. https://medium.com/analytics-vidhya/time-series-forecasting-using-tbats-model-ce8c429442a9 (accessed Apr. 07, 2022).

[16] A. M. de Livera, R. J. Hyndman, and R. D. Snyder, “Forecasting time series with complex seasonal patterns using exponential smoothing,” *J. Am. Stat. Assoc.*, vol. 106, no. 496, pp. 1513–1527, Dec. 2011, doi: 10.1198/JASA.2011.TM09771.

[17] M. Peixeiro, “The Complete Guide to Time Series Analysis and Forecasting | by Marco Peixeiro | Towards Data Science,” *Towards Data Science*, 2019. https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775 (accessed Apr. 07, 2022).

[18] J. Brownlee, “How to Create an ARIMA Model for Time Series Forecasting in Python,” *Machine Learning Mastery*, 2017. https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/ (accessed Apr. 07, 2022).

[19] J. Brownlee, “How to Grid Search ARIMA Model Hyperparameters with Python,” *Machine Learning Mastery*, 2017. https://machinelearningmastery.com/grid-search-arima-hyperparameters-with-python/ (accessed Apr. 07, 2022).