BUSINESS INTELLIGENCE SYSTEM APPLICATION AND DEVELOPMENT

Urological Cancer Referral Forecast Model

TABLE OF CONTENTS

INTRODUCTION

The impact of the Covid-19 pandemic was horrifying as it made the whole world standstill. It adversely affected the social and economic progress of a country and it also became a threat to human lives. Several problems were also confronted by health care in NHS Scotland as they faced a shortage of staff, and bed availability due to the accumulation of cancer treatment referrals in the context of this pandemic integrated with the need for 62-day cancer waiting time target for treatment which had not been met since 2012. Urological cancer is one of them which needs immediate diagnosis. This report elucidates how NHS management can be improved by discussing the backlog in Urological cancer treatment by observing the variation in the percentage of referrals across time. And also the prospect of Business Intelligence to forecast and predict Urological cancer referral treatment and how they are functional to NHS Scotland is considered. This report comprehends Time series scatter plots on different Health Boards, forecasted time series plots of four Health Boards namely Highland, Lothian, Fife and Lanarkshire by finding the most appropriate forecasting method on the basis of MSE calculated and finally dashboard consists of four different charts which provide better awareness about the data.

SECTION A: FORECASTING AND EVALUATION

The aftermath of COVID-19 was extremely uncontrollable throughout the globe, especially in industrial and health sectors like NHS Scotland. It is a big setback for NHS Scotland and they require an effective forecasting strategy to plan the functioning of health services. Especially considering Urological cancer which has become an alarming issue whose percentage of referrals increased exponentially. So it is desirable to analyse the current status of the operation of health boards and also examine study whether COVID-19 has any influence on these statistics by excluding the years 2020 and 2021. For this, four questions have to be considered when it comes to estimating the success of forecasting procedures.

- A case study has developed?
  In order to acquire an accurate forecast we have to conduct a proper case study to realise the factor that influences the result.
- Use of various operational indicators and external data?
  This has an important role in determining the performance of a business. By improving these measures, we can enhance the performance also.
- Research automation that allows exploratory and causal data analysis?
  Forecasting is done without much effort using Excel. The in-built functions in Excel facilitate the calculations. But we cannot make assurance about its accuracy.
- Evaluating its efficacy in depth?
  The model that has been developed should be free from errors. So proper evaluation should be done to ensure the accuracy and precision of the forecast.

As a result of this, the organisation would confront some challenges by implementing these forecasting techniques. There are no conventional forecasting methods, it varies in accordance to companies and some of them question its accuracy. Forecasting becomes complex if using big data. It also failed to notice the possibilities to study business and generate appropriate forecasts.

SECTION B: DATA ACCESS, CLEANING AND PREPARATION

The NHS dataset is downloaded from the 'Scottish Health and Social care open data' platform managed by Public Health Scotland'. The accessed data is then cleaned by filtering, sorting, splitting, removal of rows and duplicates. After undergoing the cleaning processes, the Excel data contains 441 rows and 7 columns and in addition to this, a new column of the percentage of referrals treated is also calculated. For eleven Health Boards each has a Period of 40 observations of the cancer type 'Urological Cancer' comprised from the year 2012-2022 is included in the cleaned data which is prepared for analysis. From this data 'Time Series Plot' for 11 Health Boards is displayed as follows:

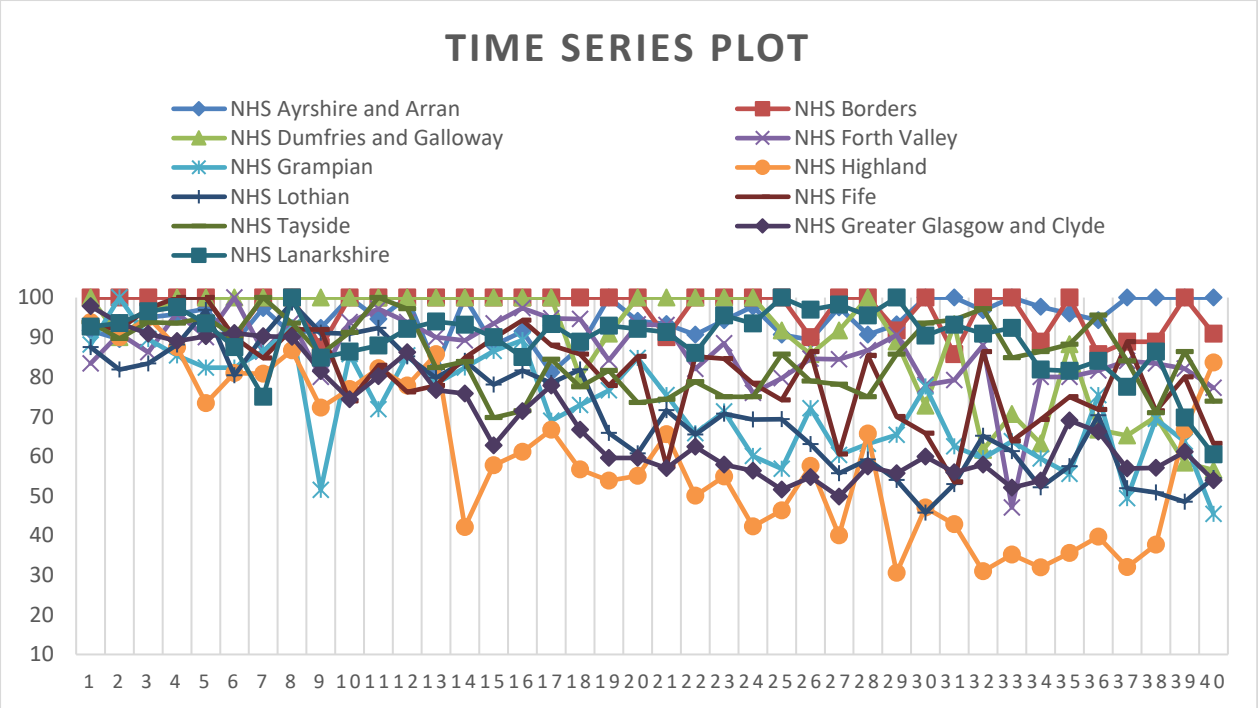

Figure 1.1 Times Series Plot

In this plot, each scatter line defines the Percentage of referrals treated of the corresponding 11 Health Boards. Here NHS Borders shows the highest 'Percentage of referrals treated'. Till the first quarter of the year 2015, this percentage was almost constant with slight variation for all Health Boards. But after that, it displays a decreasing trend for all of them. Among them, NHS Highland has the lowest percentage of referrals treated when coming to the year 2021 and there is a sudden increase in the percentage at the end of 2021 for the same Health

Board. From the overview of the plot, it is understood that the treatment of cancer referrals are not been properly met in most of the Health Boards since 2015 due to some reasons. Its reason may include the influence of COVID-19 from the year 2020. According to the BBC news report , there is a massive reduction of cancer referrals due the insufficient and inadequate diagnosis during the lockdown because screening diagnosis and clinical trials of cancer were temporarily suspended due to restrictions of the pandemic. And some of them were reluctant for going to hospitals concerned about virus infection. These may be the reasons for the decrease of cancer referrals followed by the decrease in the percentage of referrals treated. By analysing these patterns in the scatter line ideal forecasting approach can be found which can be used for future studies.

For reference: https://www.nature.com/articles/s41571-020-00446-0

SECTION C: TIME SERIES FORECASTING

NHS HIGHLAND(S08OOOO22)

In this section, forecast analysis is carried out for each of the four Heath Boards in order to plot Forecasted Time Series Plot

The time series plot for Urological cancer referrals treated for NHS Highland is obtained as follows:

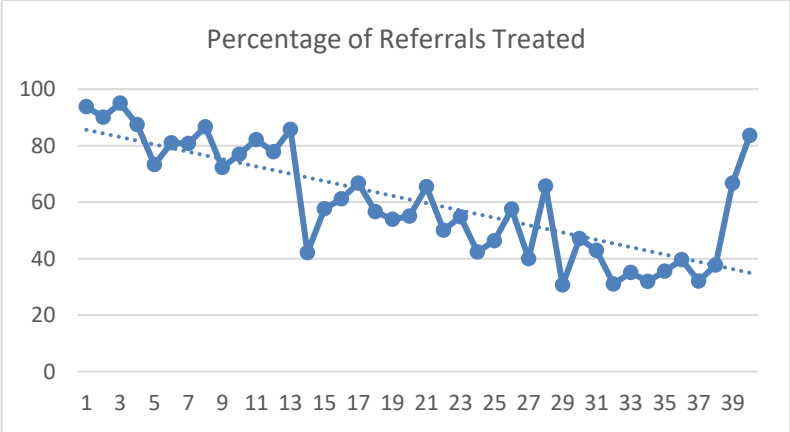

Figure 2.1 Time series plot for Highland

From this figure it can be assumed that the scatter line of percentage of referrals treated shows a decreasing trend but have a sudden increase at the end of the year 2021. And as a result of analysis, the scatter line appears to follow 'Seasonality'. To ensure that, five different forecasting methods are employed and 'Seasonality with trend' is obtained as the best approach with minimum Mean Squared Error, **MSE=156.8616.**This is the forecasted value and it is plotted in the Figure2.2. Here the red line represents the **Forecast.**

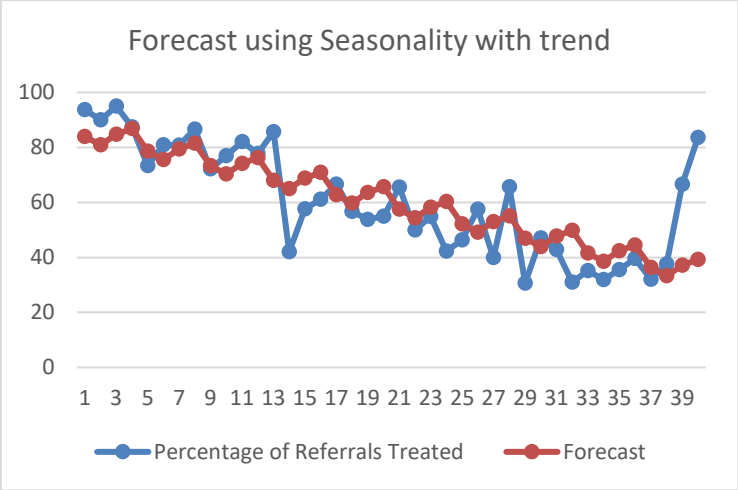

Figure 2.2 Forecasted time series plot for Highland

Now, Figure 2.3 and Figure 2.4 are charts obtained when the year 2020-2021 is excluded. While inferring them it can be understood that there is only slight changes in the trendline of the percentage of referrals and forecast compared to the previous graphs which include 2020 & 2021. And here also the forecast method used is Seasonality with trend which provide the minimum **MSE= 76.3941**. This MSE is lower than the MSE obtained including 2020 &2021.

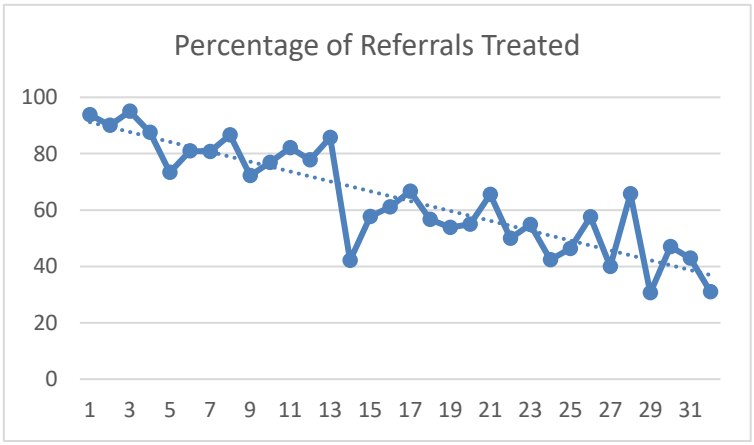

Figure 2.3 Time series plot without year for Highland

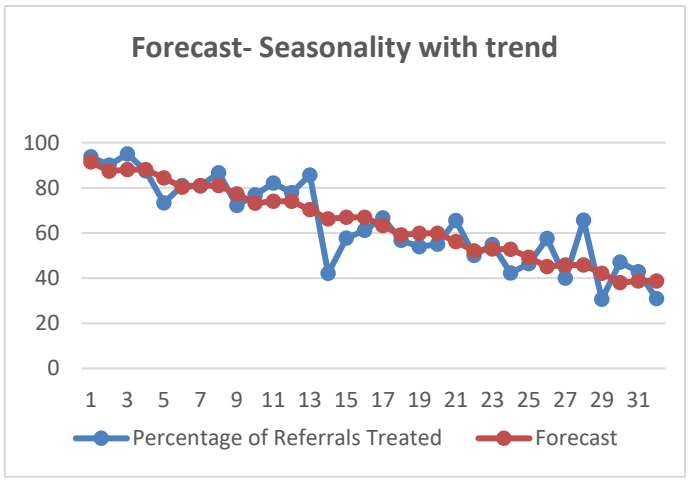

Figure 2.4 Forecasted Time series plot without year for Highland

NHS LOTHIAN(S08OOOO24)

For NHS Lothian, the time series graph of percentage of referrals versus period is given below:

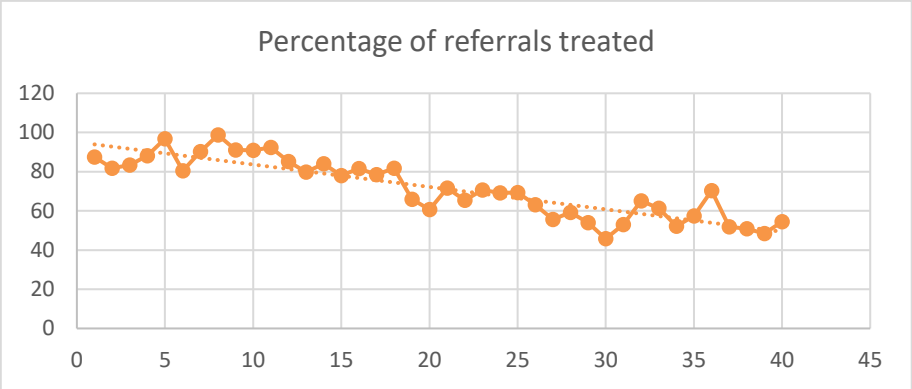

Figure 3.1 Time series plot for Lothian

Figure 3.1 it can be assumed to follow Seasonality as its trend line seems to decrease over time. So the minimum **MSE=41.5750** is obtained when Seasonality with trend forecast method is used and it is plotted as follows:

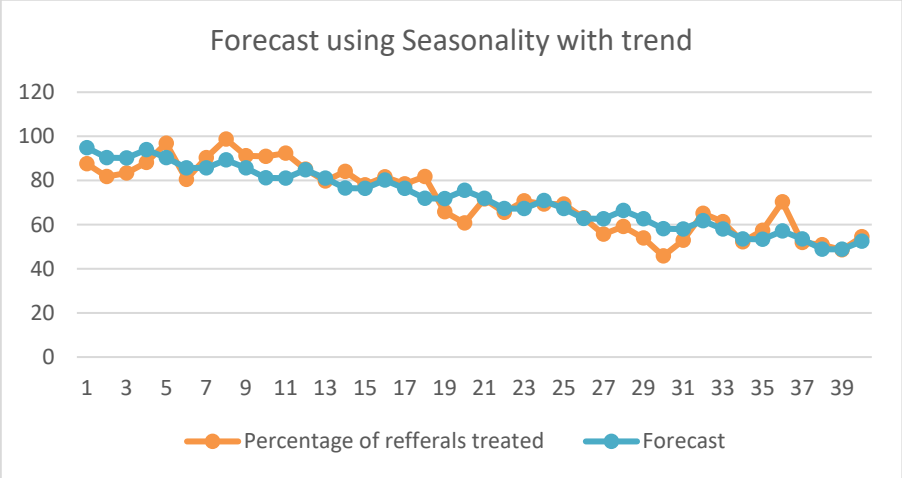

Figure 3.2 Forecasted Time series plot for Lothian

Similarly, when 2020 and 2021 are excluded, no considerable changes in the trendline is visible. And the best forecast approach is Seasonality with trend with minimum **MSE=42.6439.**

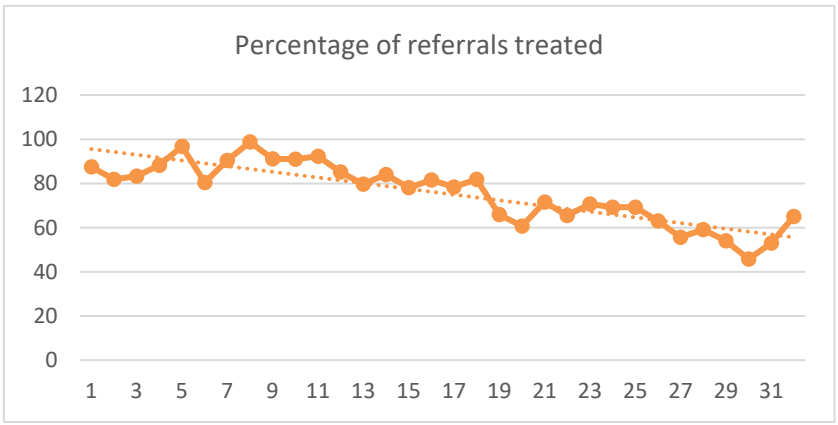

Figure 3.3 Time series plot without year for Lothian

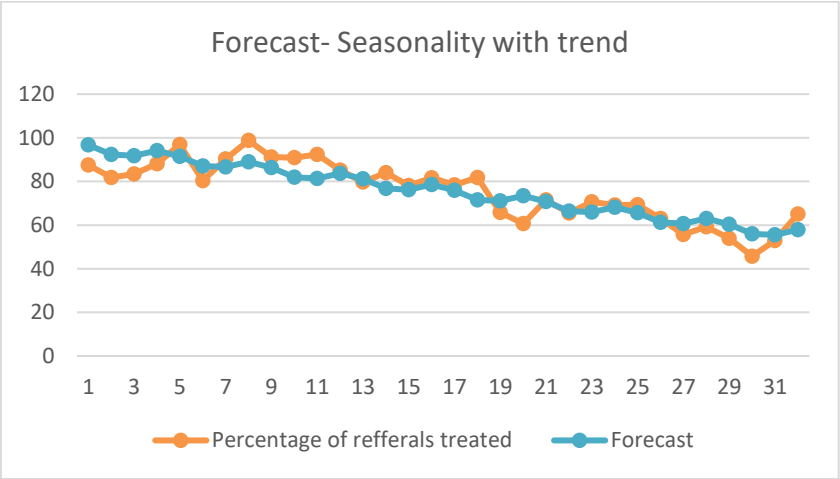

Figure 3.4 Forecasted Time series plot without year for Lothian

Upon evaluating the graphs there is no much difference in the result when these years are excluded.

NHS FIFE(S08OOOO29)

Figure 4.1 illustrates the time series scatter line for percentage of referrals treated when 2020 and 2021 are included. Comparing all models, the minimum **MSE=74.0528** is given by the Seasonality with trend method. This is the forecast value for percentage of referrals and it is plotted in the Figure 4.2

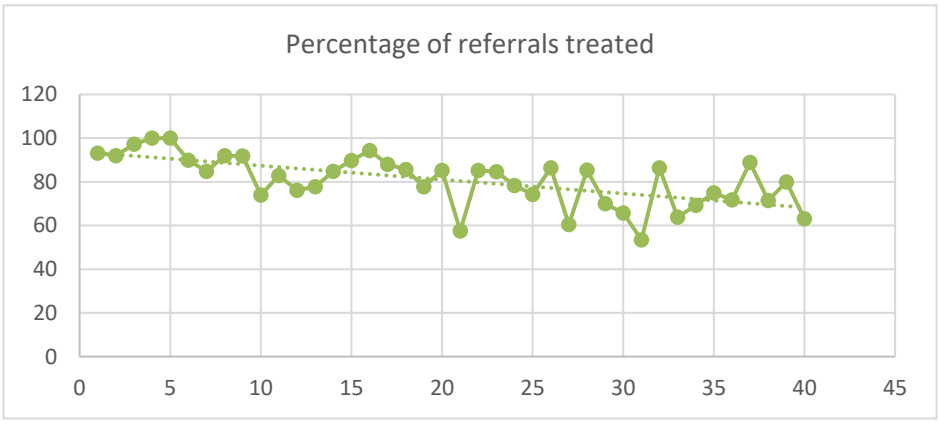

Figure 4.1 Time series plot for Fife

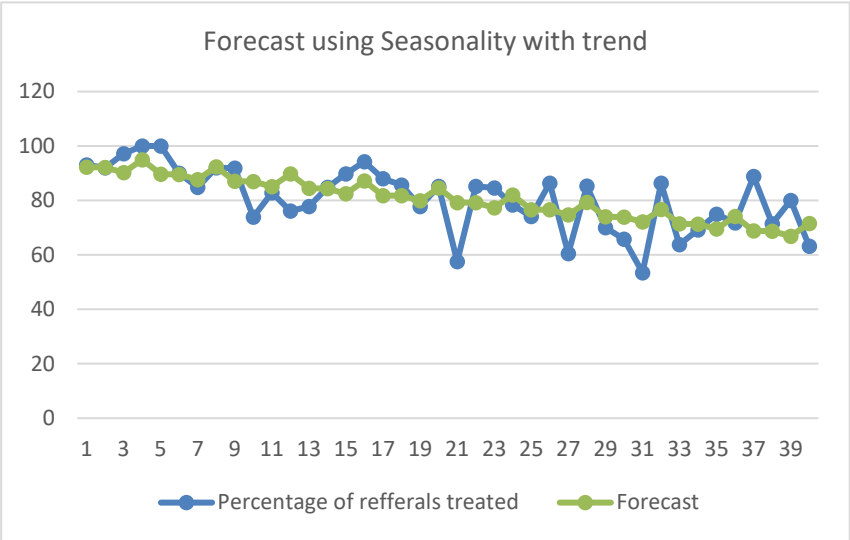

Figure 4.2 Forecasted Time series plot for Fife

Now without the years 2020 and 2021, the plots of the percentage of referrals and forecast is display in Figure 4.3 and Figure 4.4.There is no significant difference noticed in the graphs when the years are excluded. And the suitable forecast approach for NHS Fife is Seasonality with trend and the minimum MSE is attained as **63.9749**. Here also there is no significant difference in the pattern and MSE value when years are excluded.

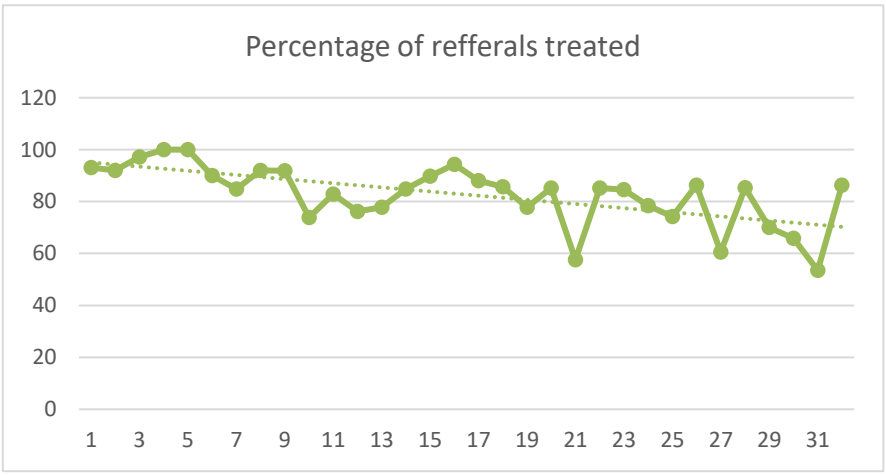

Figure 4.3 Time series plot without year for Fife

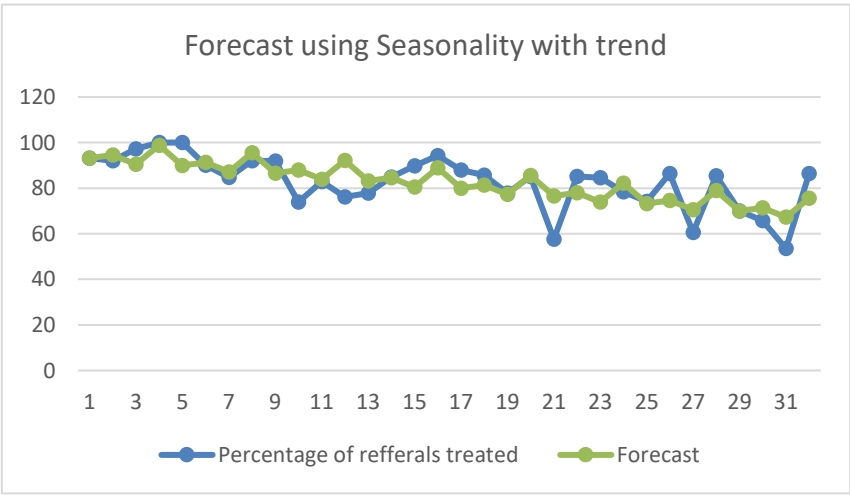

Figure 4.4 Forecasted Time series plot without year for Fife

NHS LANARKSHIRE(S08OOOO32)

For Lanarkshire, it is inferred that the scatter line in Figure 5.1 follows a Horizontal Pattern and various models is used to calculate least MSE and it is obtained by Exponential smoothing with damping factor 0.5 as shown in Figure 5.2. The forecast value is **43.4252**.

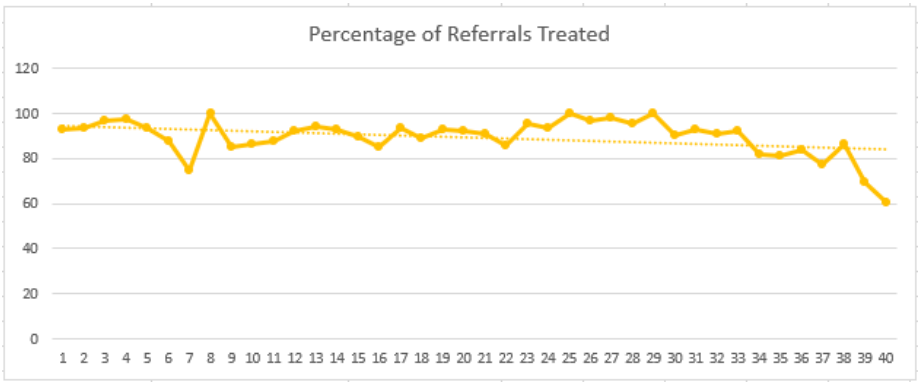

Figure 5.1 Time series plot for Lanarkshire

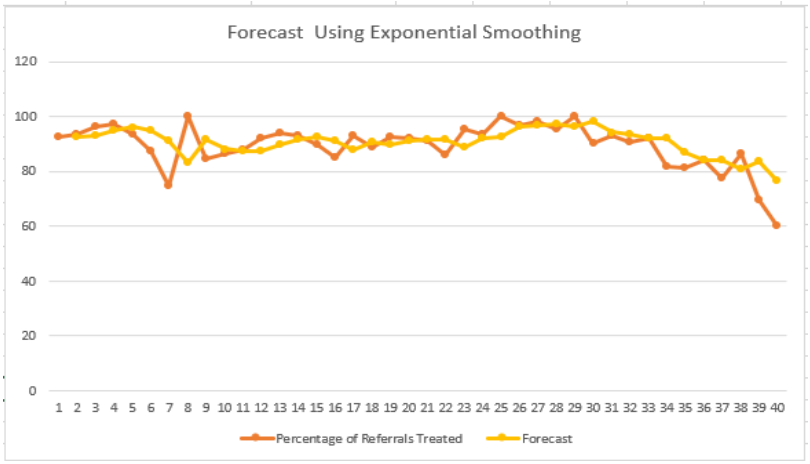

Figure 5.2 Forecasted Time series plot for Lanarkshire

In the next step, the years 2020 and 2021 are excluded and plots are displayed in Figures 5.3 & 5.4. Here, the trendline seems like a straight line and after employing different forecasting methods, Seasonality with trend is attained as appropriate with least **MSE= 23.2784.** In Figure 5.4 the yellow line indicates the Forecast.

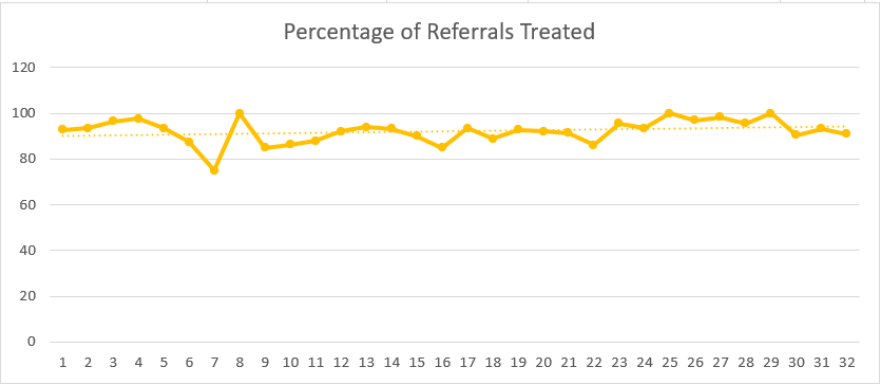

Figure 5.3 Time series plot without year for Lanarkshire

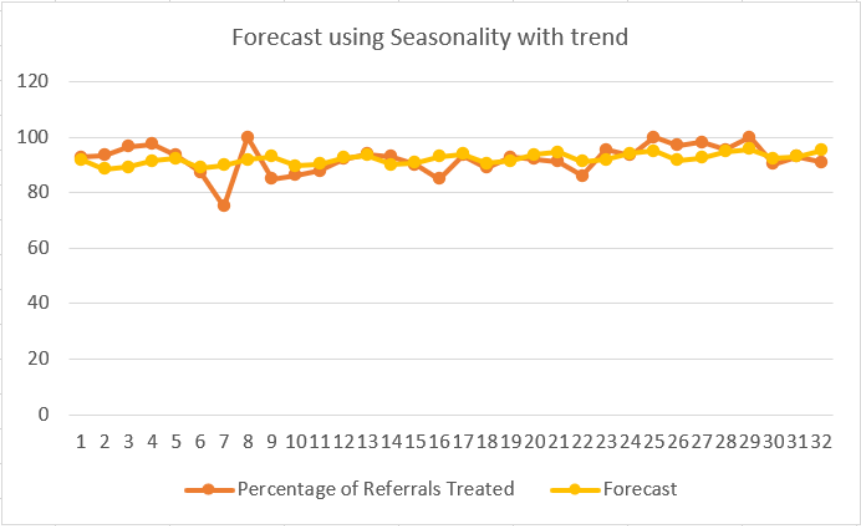

Figure 5.4 Forecasted Time series plot without year for Lanarkshire

When 2020 & 2021 are excluded , no noticeable changes visible in the graphs.

Health Board	Pattern Identified	Forecasting Method	MSE Value
NHS Highland	Seasonality	Seasonality With Trend	**156.8616**
NHS Highland Excluding 2020-2021	Seasonality	Seasonality With Trend	**76.3941**
NHS Lothian	Seasonality	Seasonality With Trend	**41.5750**
NHS Lothian Excluding 2020-2021	Seasonality	Seasonality With Trend	**42.6439.**
NHS Fife	Seasonality	Seasonality With Trend	**74.0528**
NHS Fife Excluding 2020-2021	Seasonality	Seasonality With Trend	**63.9749**
NHS Lanarkshire	Horizontal Pattern	Exponential Smoothing	**43.4252**

NHS Lanarkshire Excluding 2020-2021	Seasonality	Seasonality With Trend	**23.2784**

SECTION D : LINEAR PROGRAMMING/ LINEAR OPTIMISATION

Linear programming is one of the best approach for prescriptive analytics which is used to optimise (minimise or maximise) measurable objective which is expressed in the form of linear function. Here a problem is formulated to optimise the number of staff needed for beds so as to assign these staffs to required speciality departments such as emergency and accident in Tayside Health Board. The objective function is to minimise the staff number on the particular time shifts. The main limitation of this technique is that it is made on an assumed objective function. So a problem can arise regarding the accuracy. And LP is not suitable for the problems which are non-linear.

Time Period Shifts	Number of Staff Required
9:00-12:00	36
12:00-15:00	28
15:00-18:00	22
18:00-21:00	30
21:00-24:00	16
0:00-3:00	4
3:00-6:00	5
6:00-9:00	18

Figure 6.1 No. of Staffs needed per work shift

DECISION VARIABLES

1) Y_1 = No. of staff working 9:00-12:00 on time shift
2) Y_2 = No. of staff working on 12:00-15:00 time shift
3) Y_3 = No. of staff working on 15:00-18:00 time shift
4) Y_4 = No. of staff working on 18:00-21:00 time shift
5) Y_5 = No. of staff working on 21:00-24:00 time shift
6) Y_6 = No. of staff working on 0:00-3:00 time shift
7) Y_7 = No. of staff working on 3:00-6:00 time shift
8) Y_8 = No. of staff working on 6:00-9:00 time shift

OBJECTIVE FUNCTION

$Z = Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6 + Y_7 + Y_8$

Our objective is to minimise this function.

RESULT VARIABLE

Z is the total number of staffs required in 24 hours period.

CONSTRAINTS

$Y_1 + Y_7 + Y_8 >= 36$

$Y_1 + Y_2 + Y_8 >= 28$

$Y_1 + Y_2 + Y_3 >= 22$

$Y_2 + Y_3 + Y_4 >= 30$

$Y_3 + Y_4 + Y_5 >= 16$

$Y_4 + Y_5 + Y_6 >= 4$

$Y_5 + Y_6 + Y_7 >= 5$

$Y_6 + Y_7 + Y_8 >= 18$

$Y_i >= 0$ (i= 1,2,…8)

TABLEAU PREP

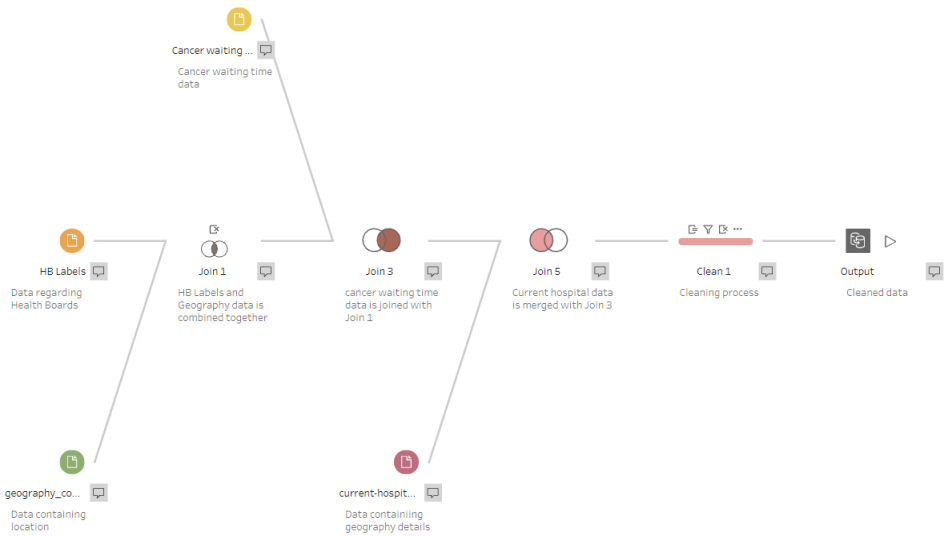

Figure 7.1 Flow diagram

Tableau Prep Builder is use to merge, arrange, format and clean the dataset and make it prepared for data analysis in tableau. To begin with this, required four data sets are connected to tableau prep builder. Now the data 'HB Labels' is joined with data 'geography codes and labels' to form 'Join1'.This inner joining incorporates matching entries from both datasets. After this 'Cancer waiting time' is right joined with 'Join1'. The right join is done to merge all entries from 'cancer waiting time' data along with matching entries from Join1 for every row. Finally, 'current hospital flagged' data is left joined with 'Join3' to form 'Join 5'.

Now, the cleaning process is intiated by removing the unwanted fields from the data and create a calculated field and calculate 'Percentage of referrals treated' using the corresponding equation. Next the current data is filtered by excluding unwanted entries from the fields 'Quarter' and 'HBT'. After that 'Quarter' is changed its data type to 'Date' and the field is renamed to 'Time'. Split is done in the field 'Postcode' and its role is changed to 'Geographic'. In order to obtain the final cleaned data 'run flow' is applied and this is used for creating Dashboard

DASHBOARD

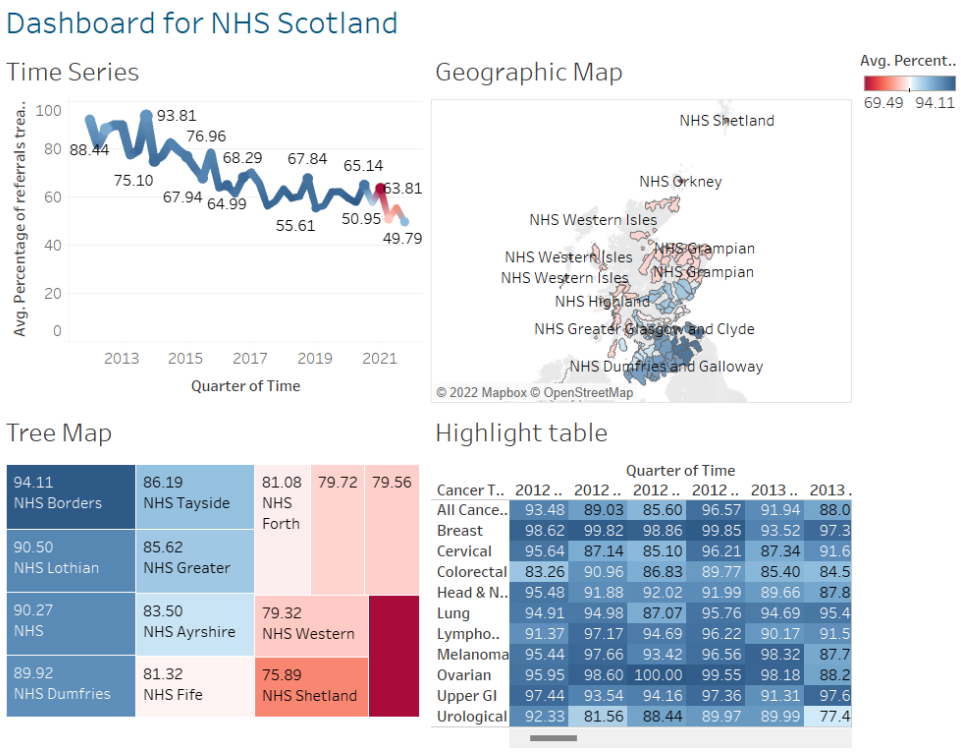

Figure 8.1 Dashboard for NHS Scotland

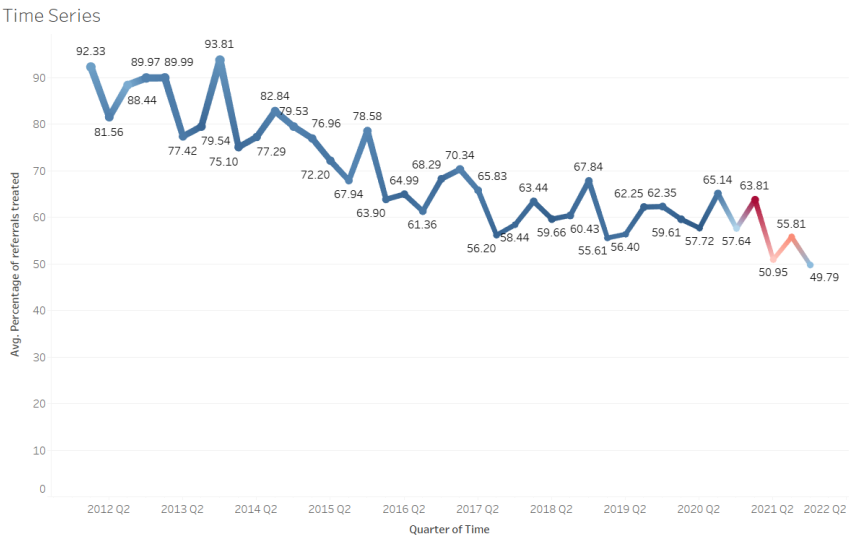

Figure 8.2 Time series graph

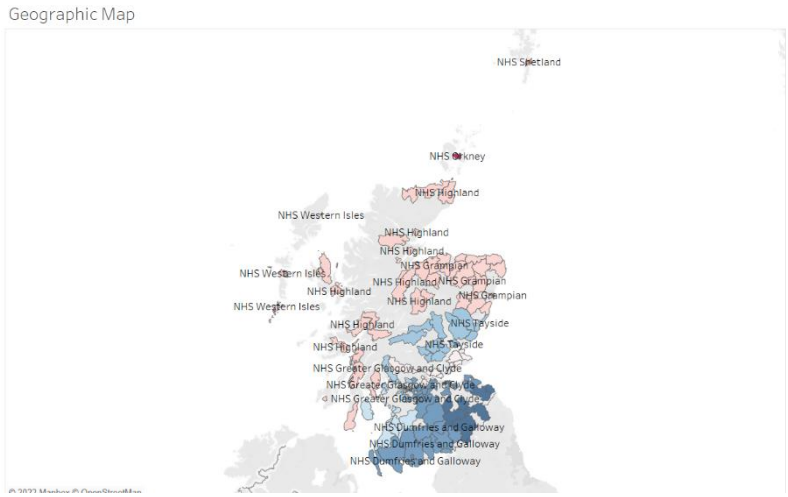

Figure 8.3 Geographic map

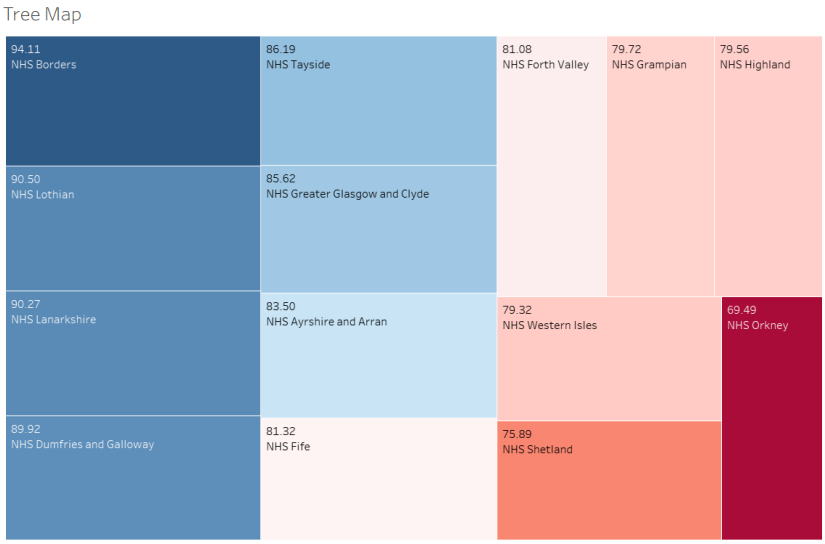

Figure 8.4 Treemap

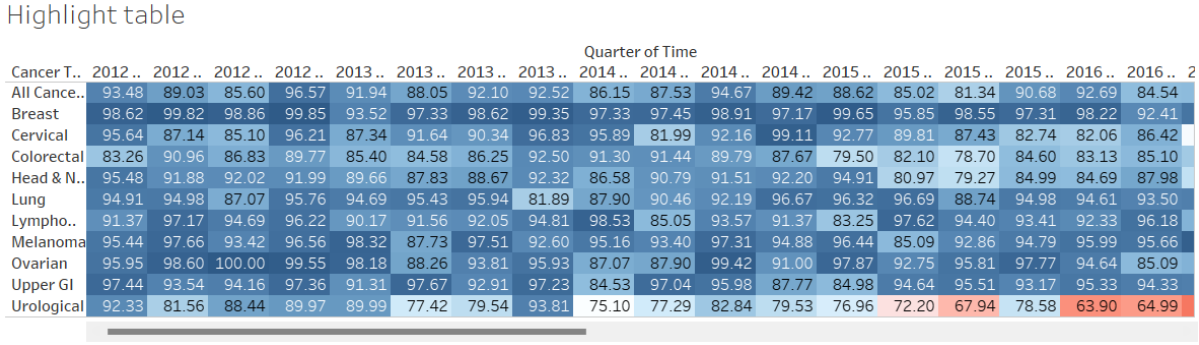

Figure 8.5 Highlight table

In the Time series Chart the quarter-wise average percentage of referrals treated during 2020-2021. Here 92.33 is the highest percentage and its in the year 2012 and in the following years, a seasonal decreasing trend is observed. This consistent decrement is continuously exhibited throughout the years. Here the lowest percentage is 49.79 in 2021 and only with this it cannot be concluded that there is an influence of Covid-19.

Now, considering the Geographic map it can inferred that the dark blue portions in the map denote the NHS with a higher average of percentage, lighter red portions represent the ones has a lower average of percentages and the red ones indicate the regions of least percentage. So it can be inferred that more referrals are treated in south end side of the map and less number of referrals are treated in the northern side. The highest percentage is 94.11 and it's for NHS Borders. And the lowest percentage is for NHS Orkney which is 69.49.

Likewise the Geographic map, when observing the Tree map, NHS Borders has the highest percentage and it is denoted by the dark blue. Then NHS Orkney has the least percentage and it is denoted by red colour. By colour variation, the percentage deviation can be easily identified. From these two maps, it is understood that, the no. of referrals that get treated is higher in NHS Borders and lower in NHS Orkney.

When studying the Highlight table, the percentage of referrals treated for various cancer types in the period 2012-2020 is considered. Among these, Breast cancer is the most treated and Urological is the least treated cancer type. And year after year the percentage is decreasing and the lowest percentage for Urological is 49.79 and its in the year 2021. So it can be predicted that the percentage will continue to decrease irrespective of the COVID scenario.

https://public.tableau.com/shared/TT3YX9N27?:display_count=n&:origin=viz_share_link

ETHICAL ISSUES IN BUSINESS INTELLIGENCE

While implementing business intelligence in the health sector, apart from exploring and forecasting data, privacy issues, ethical issues, and legal issues should be taken into account. Privacy is the major concern regarding this. The authorities should respect the privacy of the patients because detailed and complete data is collected from patients as a part of the investigation. And this data should not be misemployed for fraudulent purposes. For example, in the case of NHS Scotland, entire data is collected regarding the patients and it is uploaded as open data so that anybody can access it. They should also protect from the leakage of data. And also it is complicated to define privacy because it may vary in regions in accordance to circumstances.

In case of ethical issues, it is biased due to personal beliefs and moral views. Problems may arise related to the reliability, transparency, accuracy, quality, etc of the data. It is compelled to make certain decisions because of the data which will result in changes in our human nature.

Now legal issues are also an important thing that should be taken into consideration. Negligence of medical care from the medical professionals is a serious issue among them. The patients should be informed and aware of how the treatment and tests they are going to undergo.

CONCLUSION

In this report, an analysis of the increased accumulation of the percentage of referrals treated for cancer type 'Urological' using optimisation and forecasting procedures is carried out. The objective was to find a model suitable to improve the functioning of NHS Scotland. From the analysis, it can be concluded that the percentage of referrals treated has a consistent decrement in every subsequent year. And when the years 2020 and 2021 are excluded, there are no consequential changes in statistics due to the pandemic. Although lack of staff or shortage of beds due to the pandemic might be the reason for the enormous fall in the percentage of referrals. And it is recommended to make improvements in the data collection for NHS in order to develop accurate forecasting models.

REFERENCES

- CDP-2022-0053.pdf (parliament.uk)
- https://www.bbc.co.uk/news/uk-scotland-54337646
- https://www.bbc.co.uk/news/uk-scotland-57650825
- https://journals.sagepub.com/doi/full/10.1177/0141076820962447
- https://ifs.org.uk/publications/14798