ST344 Individual Report - 1632905

1632905

March 4, 2020

1 Executive Summary

This report has been created to discover if the demand for Coventry GPs has increased over the recent past and what factors influence it. Data used for it contains daily information about booked appointments for two full years: 2018 and 2019. Data comes from NHS Digital database.
The first part of the report investigates if the demand for Coventry GPs has increased. Initial analysis suggested that there is a positive trend in the data. However, after removing weekends and bank holidays, the trend disappears. Those days contributes to 98% of total appointments made. Therefore, weekends and bank holidays have a significant impact on the trend even if the number of visits occurring then is small. Further analysis confirms that there is not enough evidence to state that the demand for Coventry GPs has increased.
Furthermore, the impact of weather and astronomical seasons on the demand was investigated. The analysis showed that the demand strongly depends on the current season. Winter is the busiest period in the year with autumn, spring and summer coming next in descending order.
There is also a significant relationship between weekdays and number of booked appointments. The analysis showed that the demand is the highest on Mondays while it is the lowest on Thursdays. Overall, people tend to make fewer appointments as the week progresses, however they book more appointments on Fridays than on Thursdays. The probable explanation is that people prefer to see a doctor earlier instead of waiting three days to Monday.
Lastly, it can be said that the weather in the UK has a visible but weak effect on the demand for Coventry GPs. While in the rainy and cold months people tend to book more appointments, the correlation is too small to draw any stronger conclusions out of those results.

2 Introduction

2.1 Data Explanation

The dataset used in this report came directly from National Health Service databases. It contains data about appointments in General Practices collected across the UK and provides information about the NHS regional Office code, type and date of an appointment, date of the booking, and the type of healthcare professional who attended the appointment. The dataset contains information from January 2018 to December 2019.
This data collected by NHS is still in its experimental stage due to variations in practice coverage. The coverage in the chosen time frame ranges from 87% at the beginning of 2018 to more than 95% from the middle of the same year onward, when Vision and Microtest GP systems were included.
Weather data used in this report came from Statista website. It contains monthly data about the average temperature(in Celsius) and the total amount of rainfall(in mm) in the UK.

2.2 Research Questions

This report aims to answer two research questions. The first question was provided by the client and it is phrased as follows:

- Does the demand for Coventry's GP has increased over the recent past?

"Recent past" has not been thoroughly defined so it was important to choose a reasonable time frame for the study. The quality of the dataset became a breaking point since the coverage of GP in the data was not satisfactory in data from 2017. Choosing the frame of exactly two calendar years helps with a reliable exploration of the seasonal and monthly variations in the data.

The second problem was chosen from a variety of options as the one which could possibly help in estimating the demand for Coventry GPs.

- Do weather and seasonal variations influence the number of booked appointments in Coventry GPs in a significant way?

Weather and seasonal variations include features such as weekday, month, astronomical seasons, average temperature and rainfall in the UK. The report will look into the relations between those features and the demand for Coventry GPs over time.

3 Data Preparation

3.1 Merging data from three different releases

In order to obtain full two calendar years of collected data, it was necessary to merge data from three different releases. The majority of data was taken from the release from August 2019 which covers 18 months of data; four most recent months were taken from the December 2019 release; two earliest months were taken from the January 2018 release. In particular, the columns' names in all three releases must have been matched.

3.2 Pre-processing of the data

After the merger, the dataset with two full years of data was obtained. In order to investigate how the demand for appointments has changed over time, the count of appointments has been summed over the appointment date.
Furthermore, since we aim to investigate if any seasonal or weather variations influence the demand, the features indicating season, month and weekday have been added. For readability and clarity reasons, seasons were divided using the astronomical system for the northern hemisphere which is detailed below:

- **Spring**: 01 March - 31 May
- **Summer**: 01 June - 31 August
- **Autumn**: 01 September - 31 October
- **Spring**: 01 December - 28 February

Weather data was downloaded from Statista website in the xlsx file. The file required some initial processing. It was not ready to be imported to R environment directly because of its structure.

3.3 Missing values

NHS data has contained a lot of missing values. After qualitative analysis, we conclude that the missing days are Sundays and bank holidays. In other words, the days when GPs are closed. Therefore, it was decided to include those days in the dataset and populate *Appointment_Count* column with the mean(13) obtained from Sundays which were originally included in the dataset.

3.4 Adjusting the total count of appointments

One of the main issues with the dataset is the variation of GP's coverage which ranges from 87% to 98%. It would be misleading to model the number of appointments without taking this variation into account. During the whole data collection process, there were more than 6 500 GP's in the UK, and the coverage has been over 95% for 22 out of 24 months in our dataset. Therefore, it was assumed that the location of the included practices is uniformly distributed. It allowed us to calculate the estimated total number of appointments using the following formula:

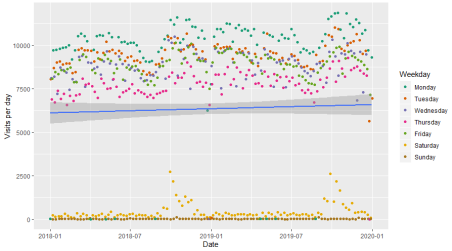
$$Adjusted\,total\,count\,of\,appointment\;=$$

$$(Total\,count\,of\,appointments\,from\,included\,GP's)\;*\;\frac{Patients\,in\,included\,GP's}{Patients\,in\,open\,GP's}$$

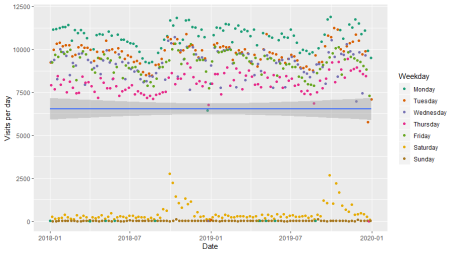
4 Data Analysis

4.1 Research Question 1: Does the demand for Coventry's GP has increased over the recent past?

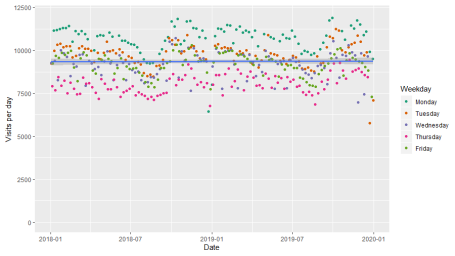
For the first research question, in the beginning we look at various time series containing *Adjusted total count of appointments* on the Y-axis. Different colors on the graphs indicate different weekdays. It allows to look at how each of them influences the overall trend.



(a) Figure 1: Total number of appointments before adjusting for the coverage ratio



(b) Figure 2: Adjusted total number of appointments

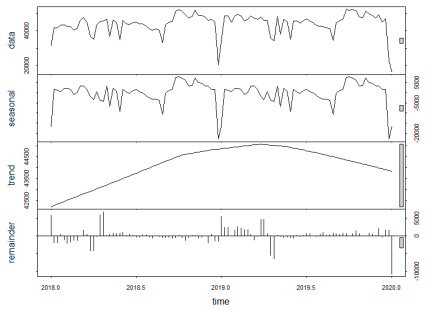


(c) Figure 3: Adjusted total number of appointments using only data from working days

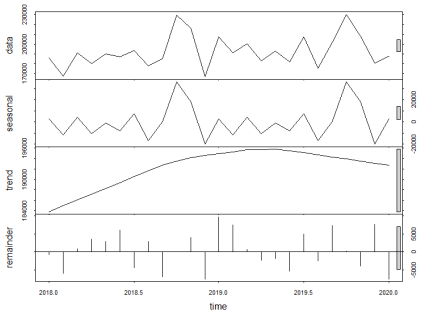
Figure 1 shows how the time series looks before the process of adjusting and is used to evaluate how this process has changed data. The biggest increase in the number of appointments is observed in the beginning of 2018. Adjusting process influenced data from 2019 much less.
In Figure 2 there is still a positive trend, however it has lost half of its intercept. As we see in Figures 1 and 2, there are usually almost no doctor GP appointments on weekends. Weekends and bank holidays were deleted from the dataset used in Figure 3 to examine the impact of those days on the overall trend. As we see, the positive trend completely disappears in this scenario. The fitted line looks completely flat.

In further studies, it would be interesting to look at the Saturdays which have over 1000 visits and check how they, in particular, impact the overall trend in the time series.

Seasonal Decomposition of Time Series by Loess(STL)



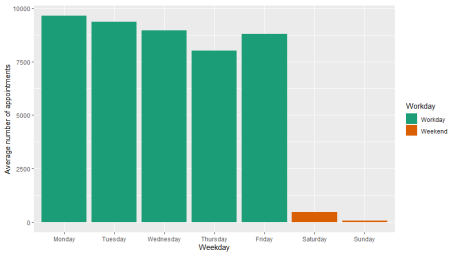
(d) Figure 4: STL produced using weekly sums of appointments



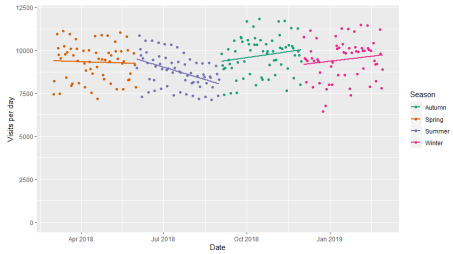
(e) Figure 5: STL produced using monthly sums of appointments

Seasonal Decomposition of Time Series is a statistical method of decomposing a time series into 3 components containing seasonality, trend and residual. The type of decomposition used above is called STL(Seasonal and Trend decomposition by Loess). Loess is a type of regression which uses weighted regression to fit a smooth curve through points in the time series.
As a result of STL, we see graphs above which show the visual interpretation of the three elements of decomposition. The plot on the left shows STL applied to the time series with weekly sums of appointments while the right one uses monthly sums. On both graphs, the seasonal component looks very similar to the original time series. The small range bars on the right of the plots confirm that the seasonal signal is very strong and can explain a large part of the variation in the original data.
Looking at the trend panel, we see that the range bars are very large on both plots which implies that the trend signal is relatively weak and a small portion of the variation in the data can be explained by the trend component.
Furthermore, the trend line in both graphs indicates the trend in data is ambiguous and it can't be concluded that there is a significant increase in demand.

4.2 Research Question 2: Do weather and seasonal variations influence the number of booked appointments in Coventry GPs in a significant way?



(f) Figure 6: Average number of visits in each weekday



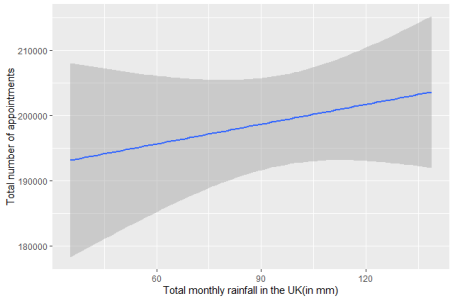
(g) Figure 7: Adjusted total count of appointments with indicated seasons

In this part, I focus on exploring the relationship between the demand for Coventry GPs and variables such as Weekday, Month, Season, Rainfall and Average Temperature in the UK. Figure 6 shows how the distribution of visits on different weekdays looks like. As expected, Mondays are the busiest days. It can be explained by the fact that the majority of people, who get ill during weekends, must wait until

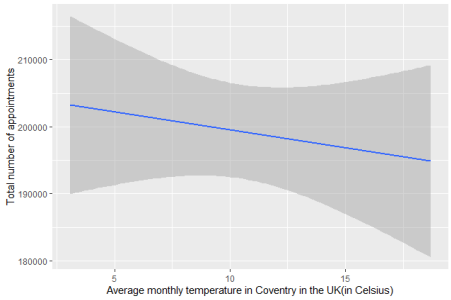
Monday to see a doctor. After Monday we see a gradual decrease in demand with a significant increase on Friday. The possible reason for this can be that people prefer to see a doctor earlier rather than to wait three days and possibly see a doctor on Monday.
Weekends are by far the least busy period in Coventry GPs. General Practices are usually closed on Saturdays and Sundays which implies that only emergency appointments take place during weekends. While each Saturday has a small but significant number of visits, Sundays have almost no appointments with an average of only 13.

Figure 7 displays our main time series in which the colors are used to indicate the four astronomical seasons with the fitted trend line for each season. In order to create a readable and insightful graph, I used data from working days such that it contains one period of each season. This allows to present how the demand for Coventry GPs change during each season. As we see, winter and autumn are the busiest seasons in the year with a significant increase in the demand. The number of visits is the lowest during summer while in spring it is only slightly lower than in the winter period.

Impact of the weather on the demand

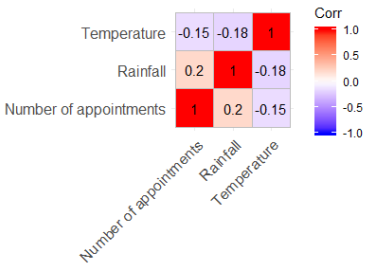


(h) Figure 8: Relationship between monthly sum of visits and total monthly rainfall in the UK



(i) Figure 9: Relationship between monthly sum of visits and average monthly temperature in the UK

The plots above were created using monthly weather data for the UK. They may help in assessing if this kind of information is useful in predicting the demand for Coventry GPs. Figure 8 shows the relationship between the number of visits and the monthly average temperature in the UK. We observe patterns in them which could be expected in this kind of graphs. Amount of rain is positively, and average temperature negatively, correlated with the number of appointments made. We must be cautious if we want to draw any conclusions out of those findings because the confidence intervals are very high on both graphs.



(j) Figure 12: Correlation heatmap the demand and weather features

In order to further analyze the impact of weather on demand, correlation heatmap with Number of Appointments, Rainfall and Average Temperature was created. The correlation between the number of appointments and Rainfall is 0.2 which implies that there is a positive correlation between them but it is weak and likely unimportant. A similar situation happens with Rainfall where the correlation is -0.18 which indicates a weak negative correlation between demand and the average temperature in the UK.

5 Conclusions

The first research question asked if there is a positive trend in the number of appointments made over time. As a count of appointments could be biased because of the different coverage in GPs across time, its values were adjusted using formula explained earlier in the report. The initial analysis suggested that there may be some positive trend in the data. However, after looking at the data for working days, the trend has disappeared. Weekends and bank holidays impacted the overall trend heavily even if less than 2% of appointments were made then. STL graphs show that seasonal signal is very strong in the data while a trend can explain very little variation in the original time series. Trend lines are inconclusive and show a positive trend for 2018 with a decrease in demand in 2019.
Concluding, there is not enough evidence to state that the demand for Coventry GPs has increased over the last 2 years.
The second research question asked if seasonal and weather variation influence the demand in a significant way. Starting from the seasonal changes, there are significant differences in the demand for Coventry GPs in different astronomical seasons. The demand increases in November and stays high until February. Then it experiences a slight drop throughout spring with the largest decrease in the summer. Furthermore, the impact of the different weekdays on the demand was investigated. It was found that Mondays are by far the busiest days while Thursdays are the least busy. There is a negative trend visible from Monday to Thursday. However, then the demand increases on Friday. Analysis of an impact of weather has given much more inconclusive results. Regardless of the quality of the data, it was shown that rainfall has a weak positive correlation with the number of visits while average temperature has a weak negative correlation with the demand.

6 Limitations

There is a number of factors which would help in calculating the trend more precisely. The major one was that it was possible to get data only from the last 2 years. In order to calculate long-term trends, it would be advantageous to get data for at least last 5 years. Strong seasonalities in data meant that we got only two full periods of data which made the analysis more difficult. Furthermore, the possibility to look at how the population and demography of Coventry have changed over time would give very interesting insights.
The process of adjusting used in this report is also one of its weaknesses since its results are only estimates. Even if the adjusted numbers are more reliable than those before adjusting, they may still be biased. Ideally, I would like to have data with consistent coverage ratio above 95%.
The first and biggest limitation of the second part was the quality of the weather data. If more detailed data(daily measurements for Coventry) were used, there would be possible to see the impact of the rain on a particular day on the number of attended appointments. Using monthly data for the UK gave some general insights about the influence of the weather but with better data, the results could be much more insightful.
Furthermore, the seasonal analysis might suffer from the fact that the Christmas period happens in December which means that there are less working days in December than in any other month. Also, people tend to stay at home during Christmas which may result in less number of infections.

7 References

1. NHS Digital. (2020). Appointments in General Practice - NHS Digital. [online] Available at: https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice [Accessed 4 Mar. 2020].

2. Statista. (2020). UK: average monthly temperatures 2015-2019 — Statista. [online] Available at: https://www.statista.com/statistics/322658/monthly-average-daily-temperatures-in-the-united-kingdom-uk/ [Accessed 4 Mar. 2020].

3. Statista. (2020). UK: Monthly rainfall 2014-2019 — Statista. [online] Available at: https://www.statista.com/statistics/584914/monthly-rainfall-in-uk/ [Accessed 4 Mar. 2020].

4. Gurarie, E. (2020). Decomposition and forecasting with loess. [online] Faculty.washington.edu. Available at: http://faculty.washington.edu/eliezg/teaching/StatR503/CourseMaterials/Week7/Lab07c_SeasonalDecompositionWithLoess.html [Accessed 4 Mar. 2020].