

Hierarchical_forecasting

Janice Hsu

2023-09-15

Contents

1	Preliminary Analysis	2
1.1	Data Introduction	2
2	Exploratory Data Analysis	3
2.1	Number of patients entering ED under different hospital hierarchy	3
2.2	Group the changed Local Health Board together	4
2.3	Number of patients who enter ED under 6 different local health boards	4
2.4	Seasonality of number of attendances	5
2.5	Change the Age_Code structure into different groups	6
2.6	Plot Number of Patients in different age groups	6
3	Forecast	6
4	Forecast with ARIMA	10
4.1	Plotting using ARIMA	10
5	Forecasting with ETS	14
5.1	Plotting with ets	14
6	Accuracy assessment for ARIMA and ETS	17
6.1	Forecasting using ARIMA and ETS	17
6.2	Calculating Accuracy Metrics for ARIMA and ETS	17
7	Forecast after Covid-19	21
8	Reconciliation	29

1 Preliminary Analysis

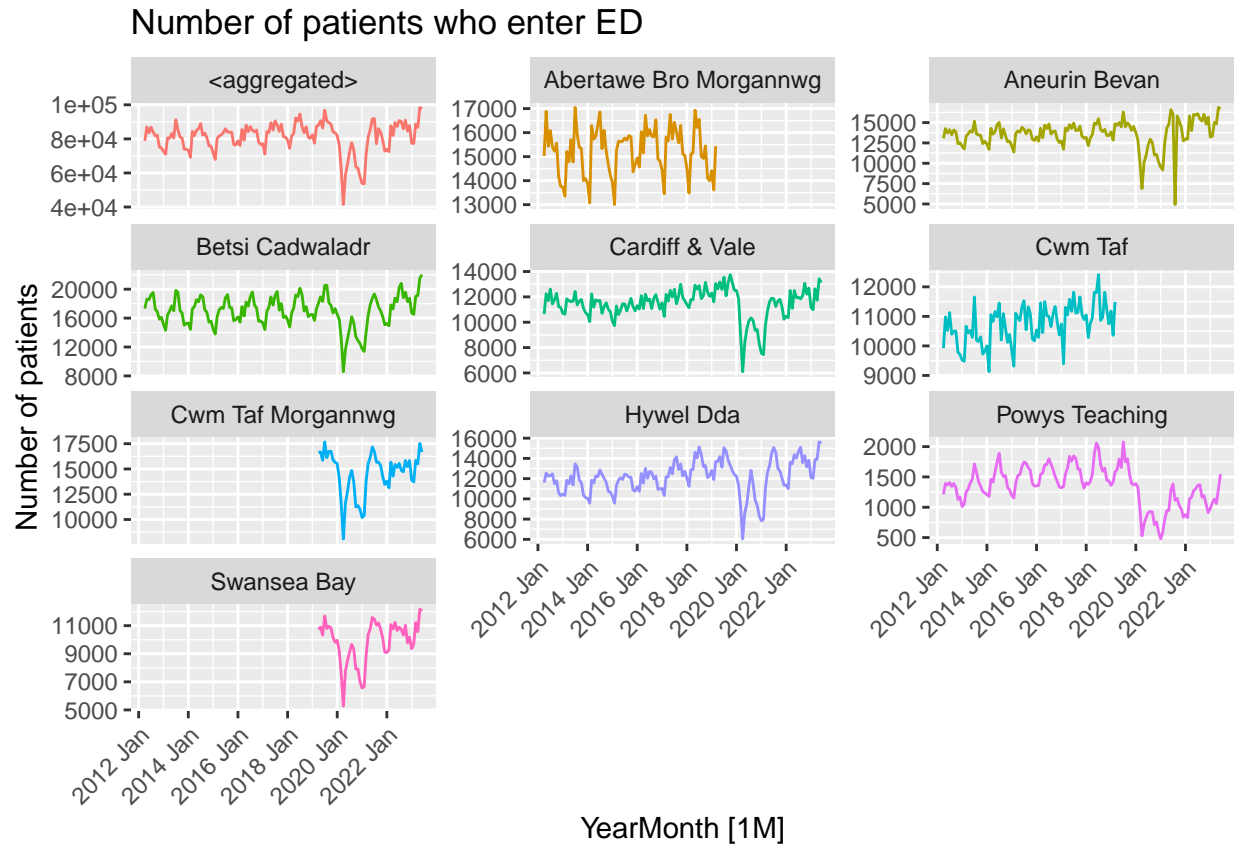
1.1 Data Introduction

The dataset contains 10 variables related to the hospitals and information of patients in Wales, UK. Here are the brief summary of the dataset:

- **Data:** This column represents the number of attendance in each emergency department.
- **YearMonth:** This column represents dates in the year-month format. Additionally, this dataset contains data from 2012 April to 2023 May.
- **Age_Code:** This column provides the age group that the patient is in. There are 17 different age groups. They are “0 to 4”, “5 to 17”, “18 to 24”, “25 to 29”, “30 to 34”, “35 to 39”, “40 to 44”, “45 to 49”, “50 to 54”, “55 to 59”, “60 to 64”, “65 to 69”, “70 to 74”, “75 to 79”, “80 to 84”, “85” and “Unknown”.
- **Sex_ItemName_ENG:** This column provides the information of patient’s gender.
- **Hospital_Code:** This column represents 42 different hospitals in Wales.
- **Hospital_ItemName_ENG:** This columns refers to the name of the 42 different hospitals in Wales.
- **Hospital_Hierarchy:** This column represents the code for the health board that the hospital belongs to.
- **Hospital_AltCode1:** This column provides an alternate code for the hospital.
- **Organisation:** This column represents the health board.
- **Organisation_Code:** A code for the organisation as well as the health board.
- There are three hierarchies in this dataset. On the top level, there is all the hospitals in Wales, while on the second hierarchy, there are 6 different health boards which also shown as the organisations. At the bottom level, there are 42 hospitals in total.

2 Exploratory Data Analysis

2.1 Number of patients entering ED under different hospital hierarchy



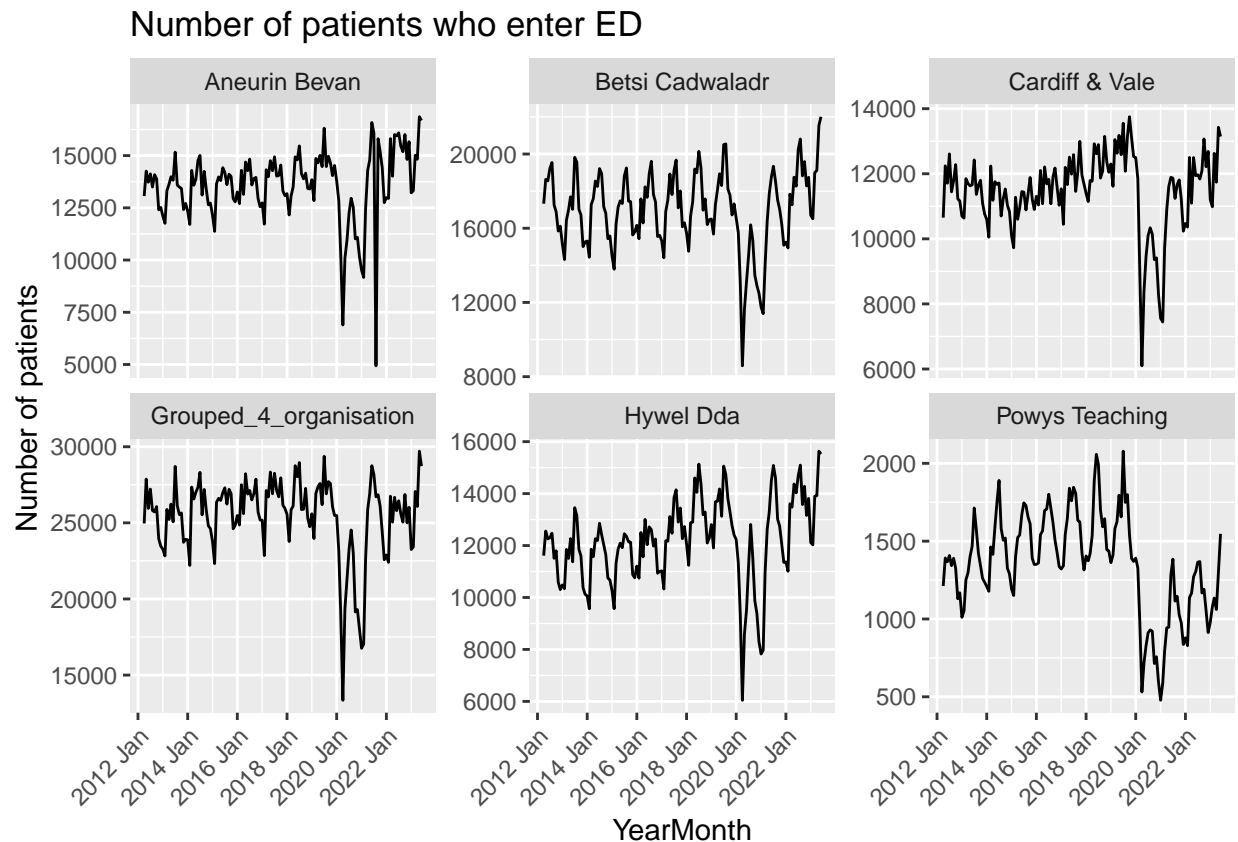
- A couple of Local Health Boards (LHBs) were redefined from the 1st of April 2019 onwards: Cwm Taf (27) → Cwm Taf Morgannwg (30) // Abertawe Bro Morgannwg (26) → Swansea Bay (31). Therefore, if you decide to forecast at LHB resolution, you might want to consider these 4 as a unique one.
- A the Princess of Wales Hospital changed its Local Health Boards
- So we analyse these 4 as one organisation

2.2 Group the changed Local Health Board together

2.2.1 There are 6 Local Health Boards

```
## [1] "Betsi Cadwaladr"      "Hywel Dda"      "Grouped_4_organisation"  
## [4] "Cardiff & Vale"      "Aneurin Bevan"  "Powys Teaching"
```

2.3 Number of patients who enter ED under 6 different local health boards



Findings:

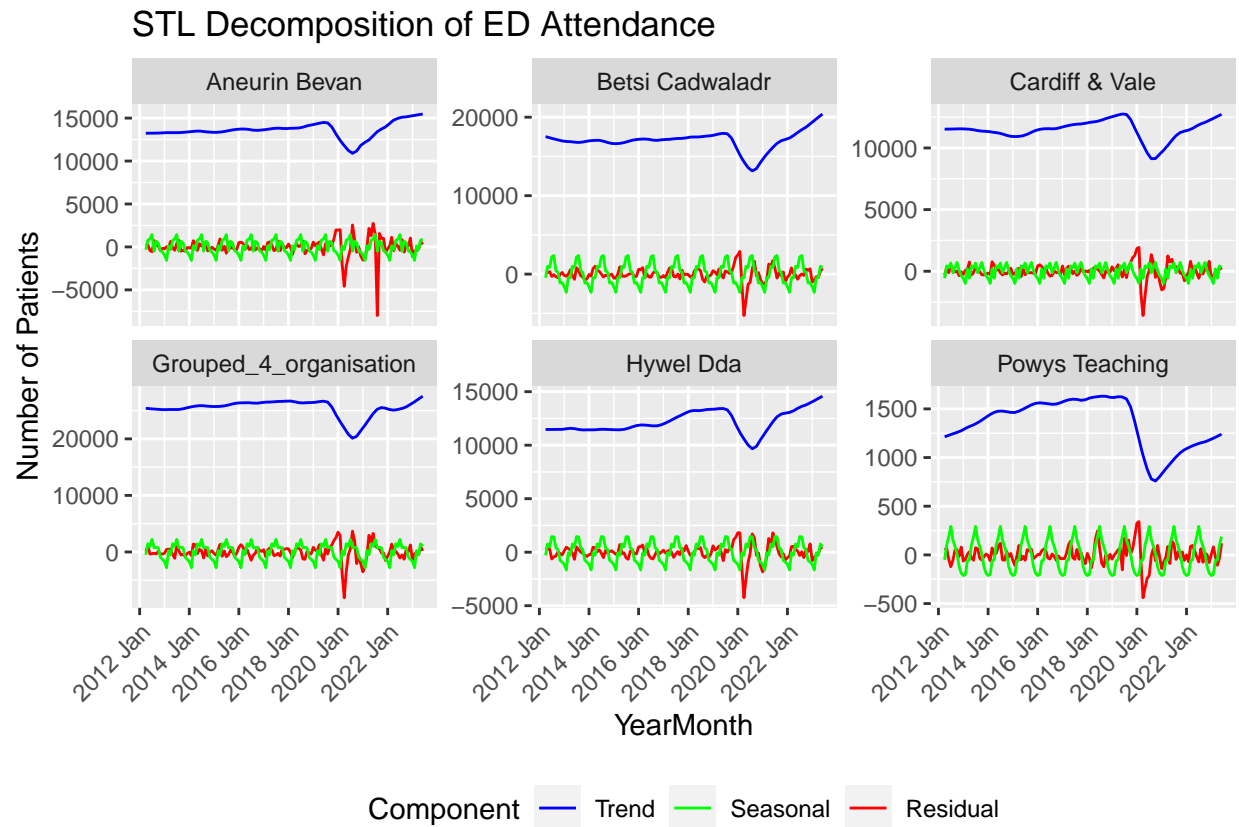
- There is a big decline during the Covid-19 period, and after the Covid-19, except the Powys Teaching, other local health boards have increased the number of attendance back to its previous years.
- There seems to be seasonality in the data for each health board, and I would like to investigate more on top of this.

2.4 Seasonality of number of attendances

To investigate deeper into the potential seasonality in the data, I would like to decompose the time series according to each health board. It allows us to discover the trend, seasonality and the residual components.

2.4.1 Decompose Time Series

2.4.2 Plotting



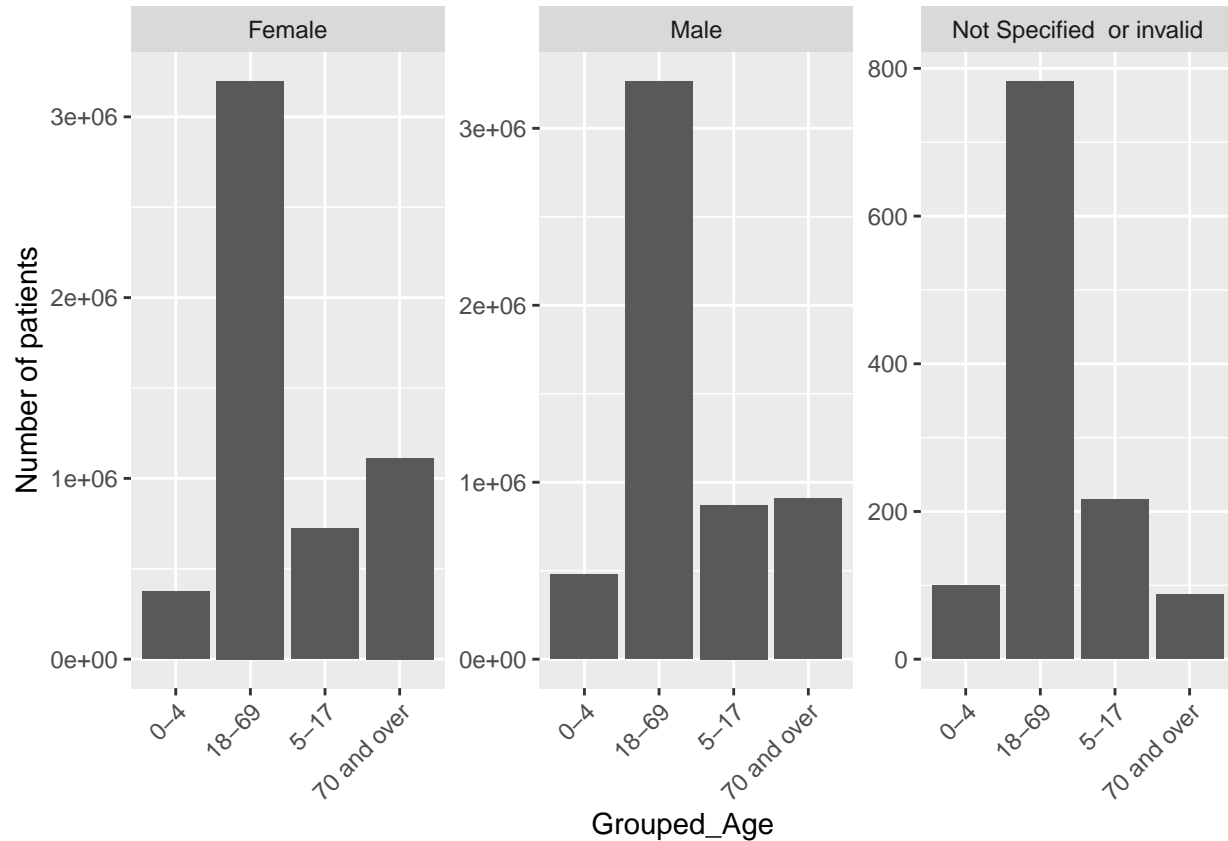
Based on the STL decomposition, it is evident that each health board follows a similar trend. Moreover, it is obvious that there is a significant downturn during the Covid-19 era. As for seasonality, there is a pronounced surge in the number of patient's attendance in the middle of the year (approximately in June or July). This seasonal pattern underlines the recurrent nature of patient admissions.

2.5 Change the Age_Code structure into different groups

```
## [1] "0 to 4" "18 to 24" "25 to 29" "30 to 34" "35 to 39" "40 to 44"
## [7] "45 to 49" "5 to 17" "50 to 54" "55 to 59" "60 to 64" "65 to 69"
## [13] "70 to 74" "75 to 79" "80 to 84" "85" "Unknown"
```

2.5.1 Age group: “0-4”, “5-17”, “18-69”, “70^”

2.6 Plot Number of Patients in different age groups



Findings:

- The observation that the age group 18-69 has the most amount of patient attendance is expected, as it is the biggest group among all. However, it is noteworthy that the second biggest group are from the oldest age bracket, aligning with the general understanding of the health care need for the elders.

3 Forecast

```
## $'Aneurin Bevan'
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jul 2023      15327.46 13401.32 17253.59 12381.69 18273.22
## Aug 2023      14758.62 12683.77 16833.47 11585.41 17931.83
## Sep 2023      14331.89 12174.36 16489.43 11032.23 17631.56
```

```

## Oct 2023      14077.53 11891.77 16263.28 10734.70 17420.35
## Nov 2023      13913.04 11715.52 16110.55 10552.23 17273.84
## Dec 2023      13809.83 11607.72 16011.95 10441.99 17177.68
##
## $'Betsi Cadwaladr'
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## Jul 2023      22520.72 21208.70 23832.74 20514.16 24527.28
## Aug 2023      22195.49 20282.25 24108.74 19269.44 25121.55
## Sep 2023      20359.95 18117.18 22602.73 16929.92 23789.98
## Oct 2023      19284.65 16835.70 21733.59 15539.31 23029.99
## Nov 2023      17985.58 15400.91 20570.24 14032.67 21938.49
## Dec 2023      17199.82 14523.44 19876.20 13106.64 21293.00
##
## $'Cardiff & Vale'
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## Jul 2023      12894.49 11932.17 13856.81 11422.743 14366.23
## Aug 2023      12453.99 11210.08 13697.89 10551.600 14356.38
## Sep 2023      12491.75 11090.31 13893.20 10348.425 14635.08
## Oct 2023      12649.40 11151.53 14147.27 10358.603 14940.20
## Nov 2023      12125.90 10566.68 13685.12  9741.274 14510.52
## Dec 2023      11831.67 10232.62 13430.73  9386.128 14277.22
##
## $Grouped_4_organisation
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## Jul 2023      28201.58 26101.77 30301.39 24990.20 31412.96
## Aug 2023      27165.96 24527.91 29804.02 23131.41 31200.52
## Sep 2023      26742.12 23837.94 29646.29 22300.56 31183.67
## Oct 2023      26782.76 23735.25 29830.27 22122.00 31443.52
## Nov 2023      25645.29 22517.89 28772.70 20862.34 30428.25
## Dec 2023      25286.03 22113.33 28458.73 20433.81 30138.26
##
## $'Hywel Dda'
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## Jul 2023      15984.72 14923.747 17045.70 14362.099 17607.35
## Aug 2023      16088.04 14582.199 17593.88 13785.055 18391.02
## Sep 2023      14280.20 12518.452 16041.94 11585.841 16974.55
## Oct 2023      13551.66 11633.763 15469.55 10618.491 16484.82
## Nov 2023      12548.97 10532.122 14565.81  9464.468 15633.47
## Dec 2023      11964.73  9883.705 14045.76  8782.076 15147.39
##
## $'Powys Teaching'
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## Jul 2023      1593.479 1421.3226 1765.635 1330.1887 1856.768
## Aug 2023      1476.309 1242.9241 1709.693 1119.3776 1833.240
## Sep 2023      1391.293 1116.9363 1665.651  971.7004 1810.887
## Oct 2023      1260.690  956.2214 1565.159  795.0455 1726.335
## Nov 2023      1188.828  861.2557 1516.400  687.8494 1689.806
## Dec 2023      1020.005  674.2630 1365.747  491.2381 1548.772
##
## $Total
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## Jul 2023      96354.13 89558.55 103149.71 85961.19 106747.1
## Aug 2023      89505.44 80770.23  98240.66 76146.08 102864.8
## Sep 2023      90104.99 80309.35  99900.63 75123.86 105086.1

```

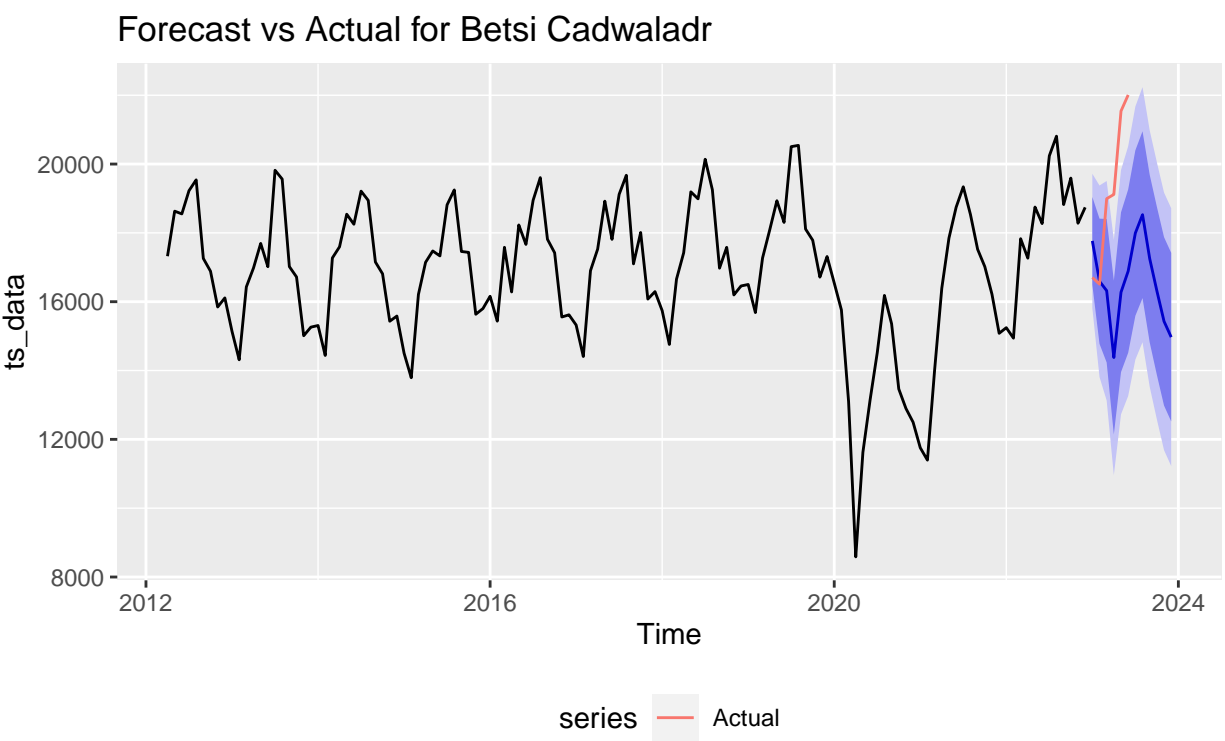
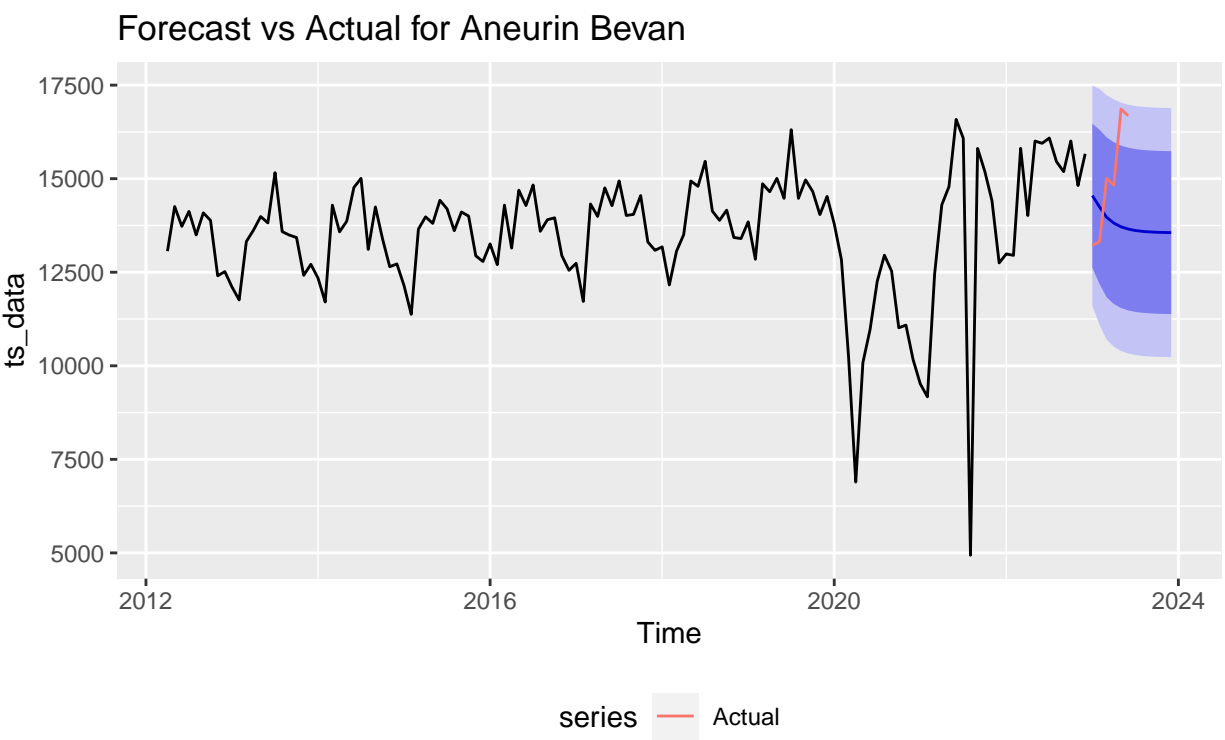
## Oct 2023	89432.47	79003.05	99861.90	73482.04	105382.9
## Nov 2023	85475.78	74652.91	96298.65	68923.62	102027.9
## Dec 2023	83342.11	72270.11	94414.11	66408.95	100275.3

4 Forecast with ARIMA

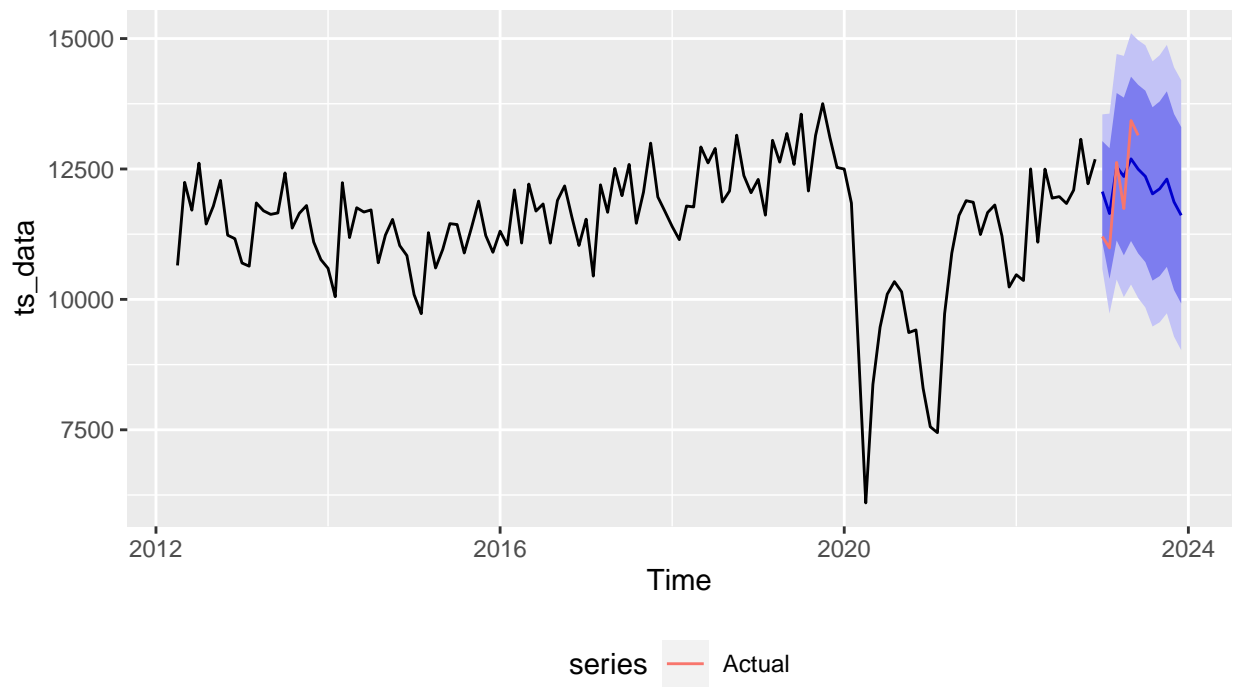
4.0.1 Define the forecast horizon and validation period

4.0.2 Splitting the data into training and testing sets

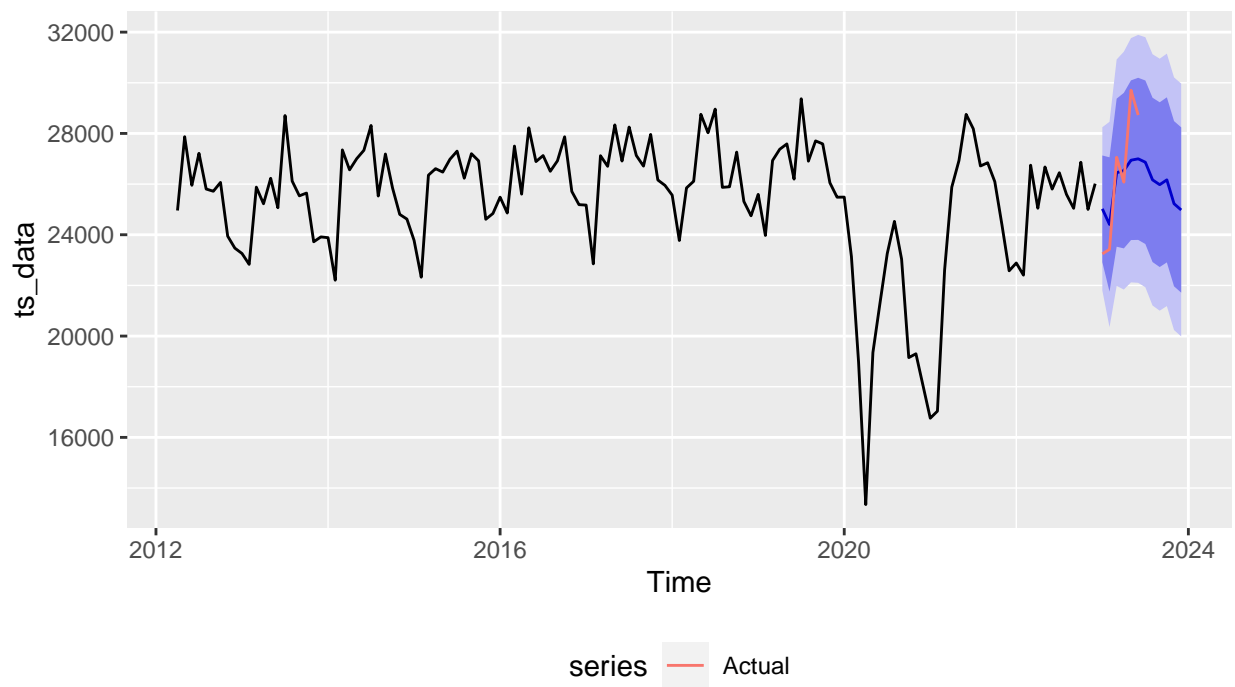
4.1 Plotting using ARIMA



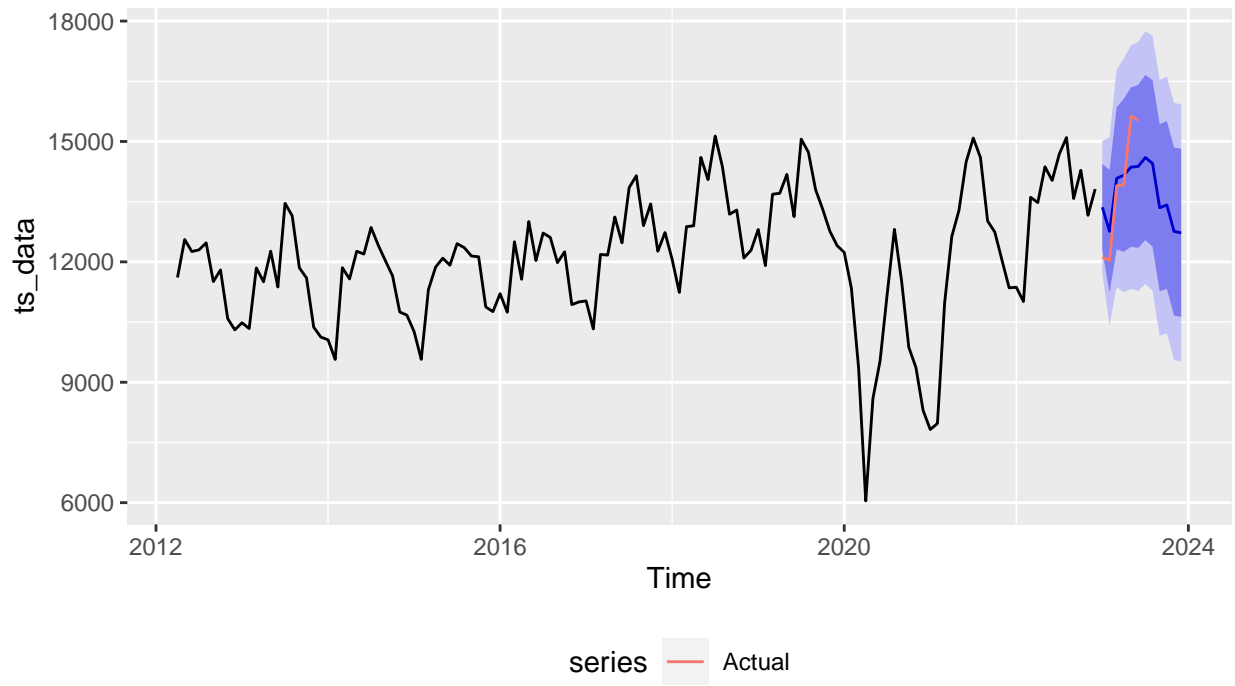
Forecast vs Actual for Cardiff & Vale



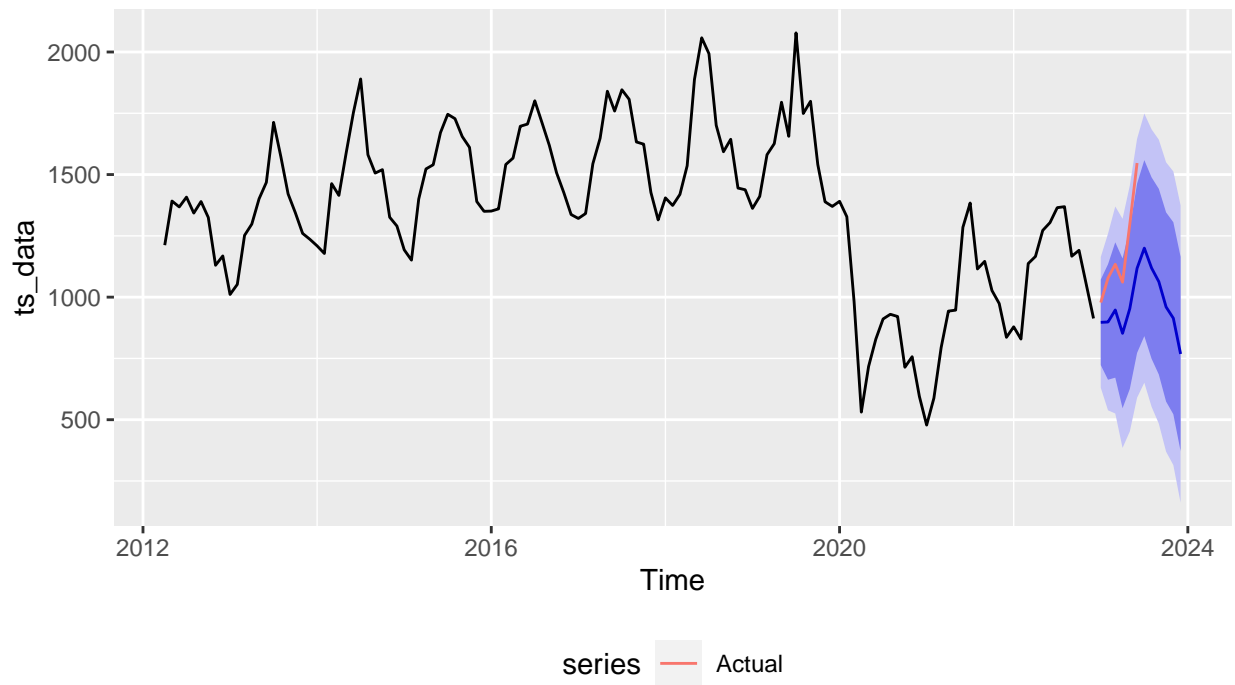
Forecast vs Actual for Grouped_4_organisation



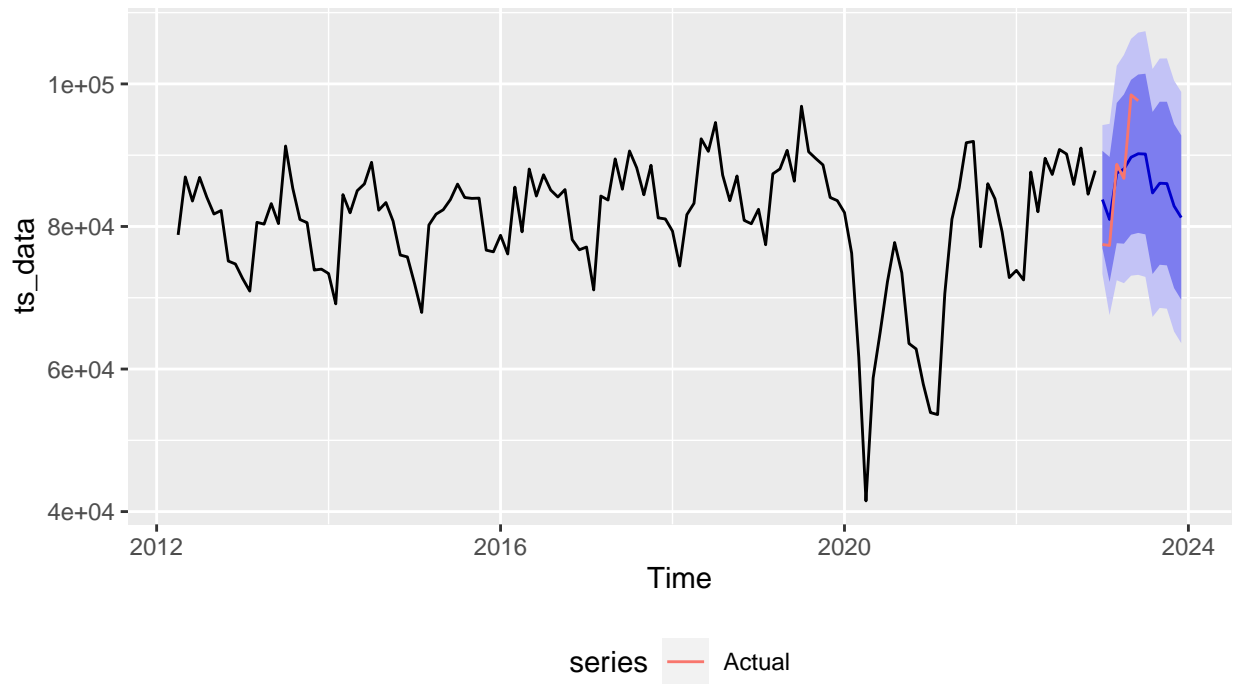
Forecast vs Actual for Hywel Dda



Forecast vs Actual for Powys Teaching



Forecast vs Actual for Total

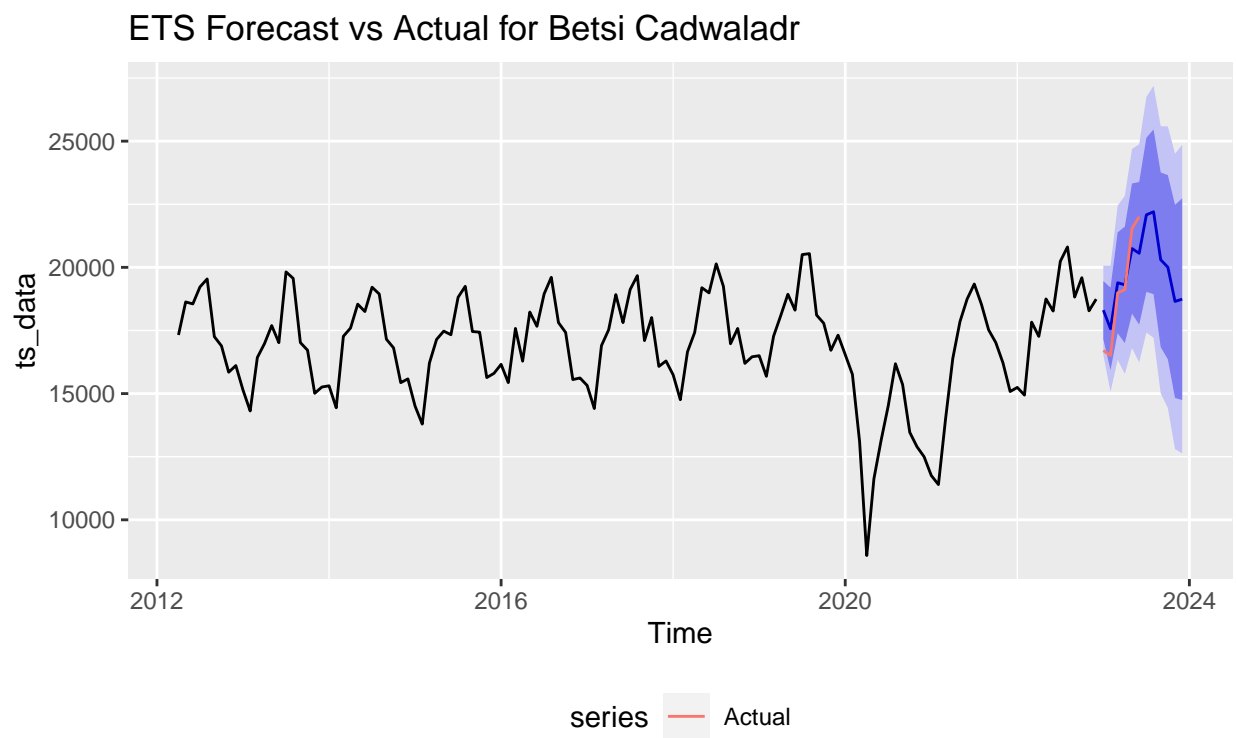
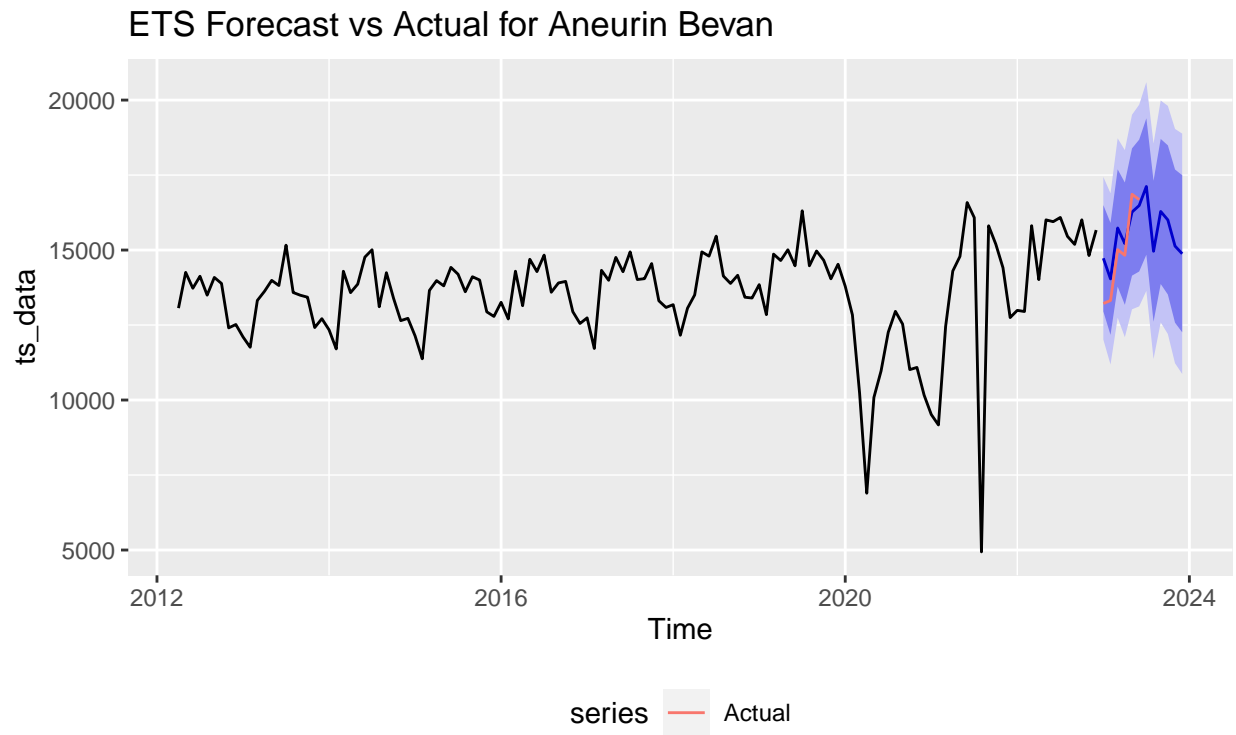


Due to the impact of Covid-19, the trend of forecast in the future appears to be lower than expected. However, the ARIMA model seems to suggest the consistent direction as the actual data for Cardiff & Vale, Grouped_4_Organisation, Hywel Dda, Powys Teaching and Total (All-Wales). Noticeably, they all have a downturn in the latter half of 2023.

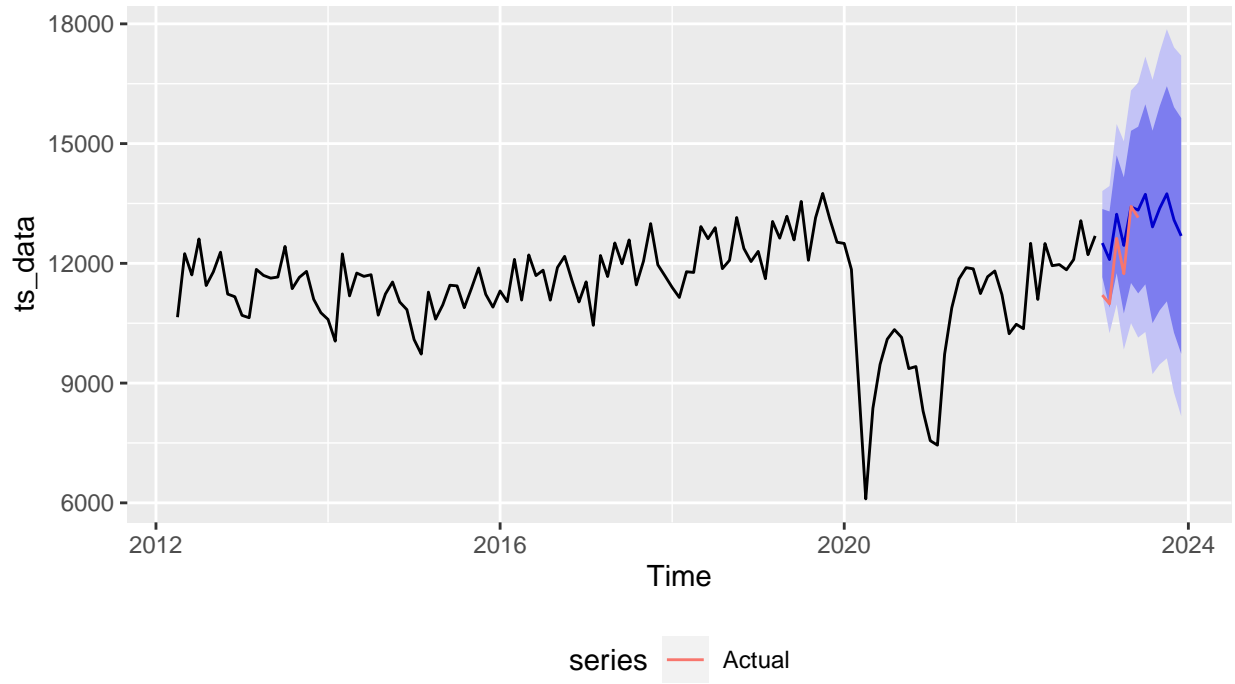
However, the forecast for Aneurin Bevan and Betsi Cadwaladr are not that accurate according to the divergence between the predicted value and the actual data.

5 Forecasting with ETS

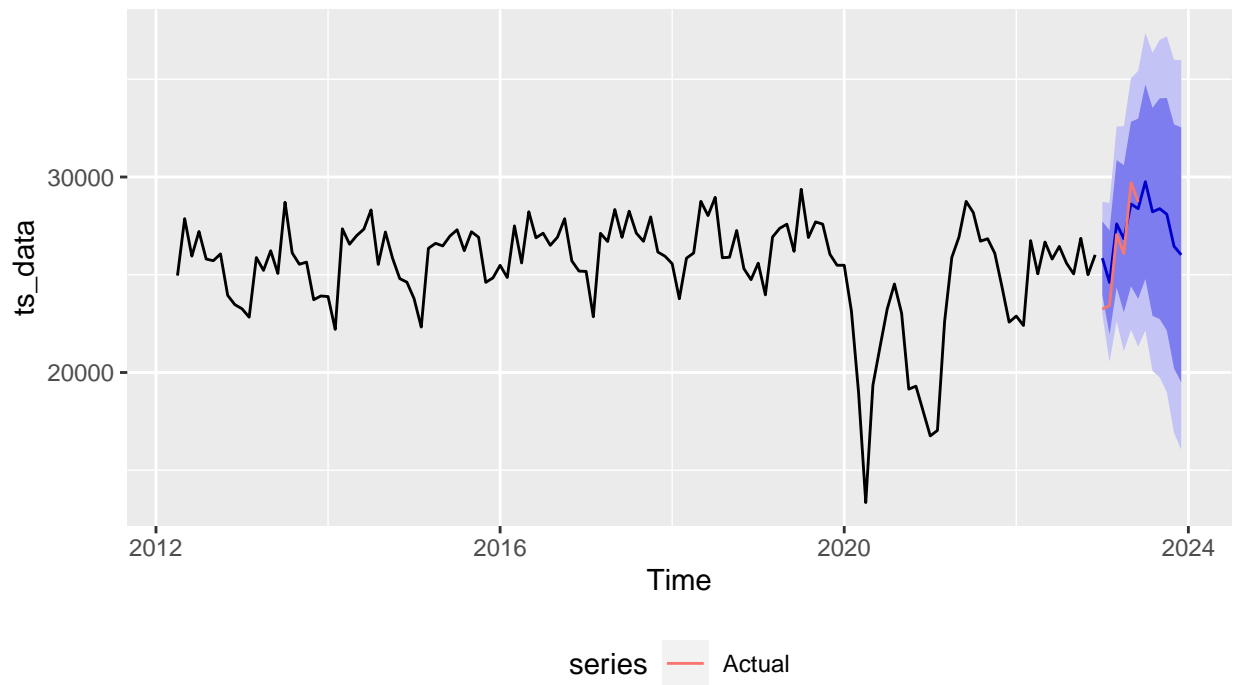
5.1 Plotting with ets



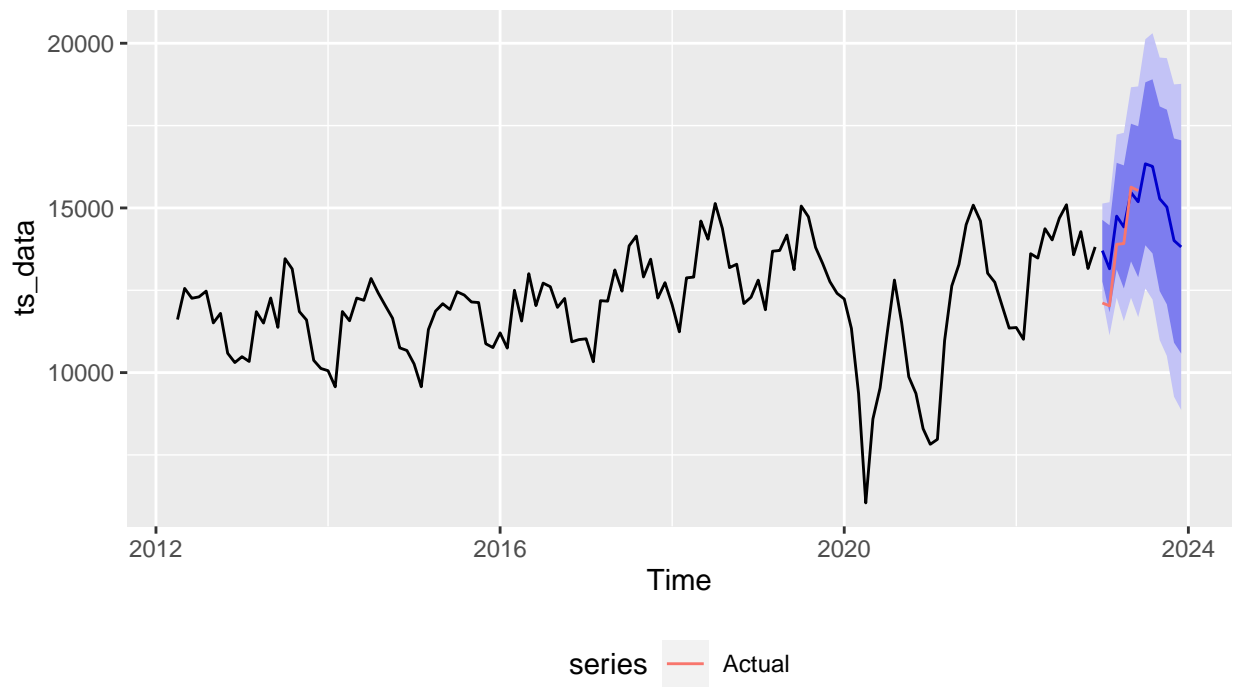
ETS Forecast vs Actual for Cardiff & Vale



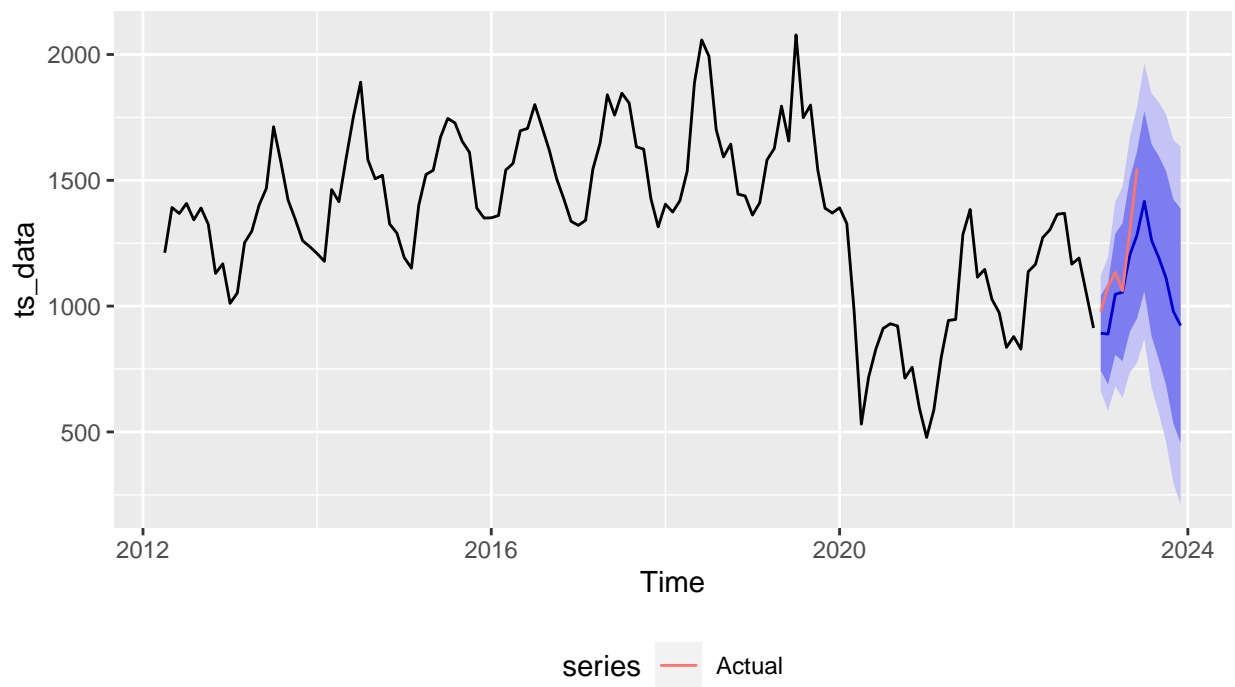
ETS Forecast vs Actual for Grouped_4_organisation

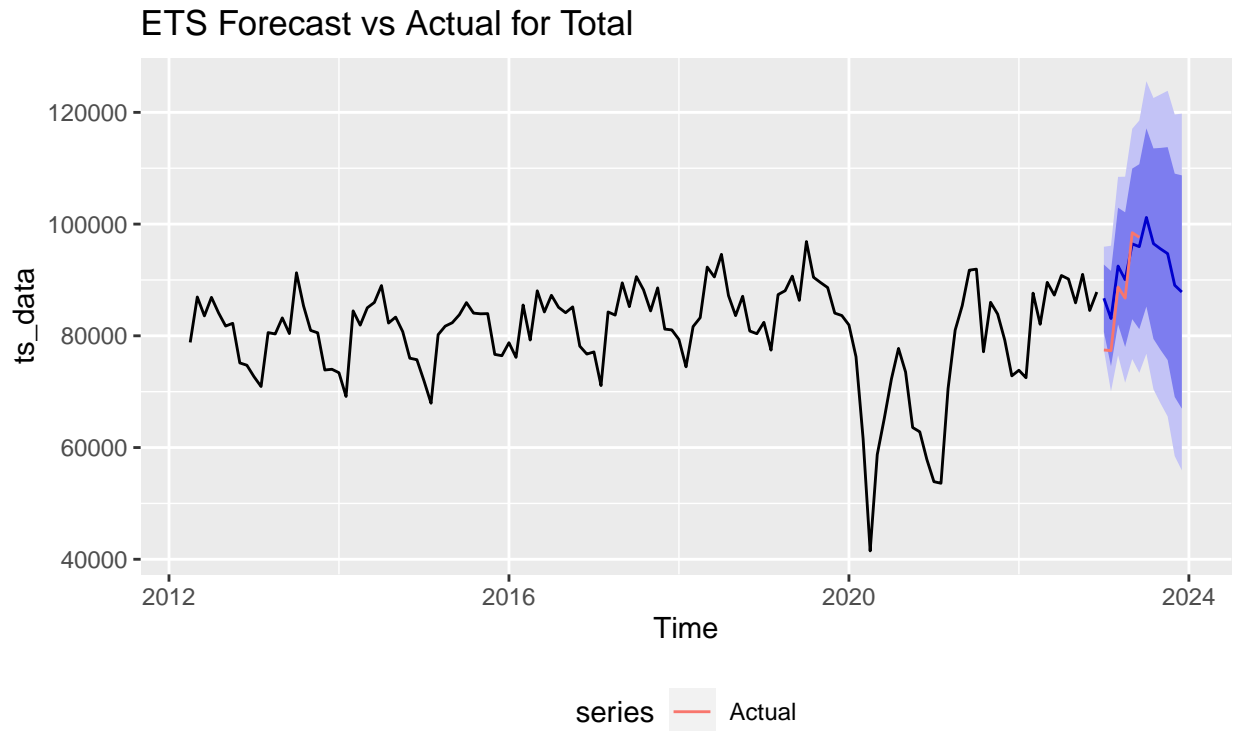


ETS Forecast vs Actual for Hywel Dda



ETS Forecast vs Actual for Powys Teaching





The prediction from the ETS model aligns closely with the actual data upon the visual inspection. Therefore, It might suggest that a higher accuracy for the ETS model compared to the ARIMA model.

Additionally, all the predictions have a downturn of trend in the latter half of 2023. According to the historical data, this might be due to that the peak appears in the middle of the year followed by a trough at the end and the start of the year.

6 Accuracy assessment for ARIMA and ETS

6.0.1 Define the forecast horizon and validation period

6.0.2 Splitting the data into training and testing sets

6.0.3 Lists to store forecasts and error metrics

6.1 Forecasting using ARIMA and ETS

6.2 Calculating Accuracy Metrics for ARIMA and ETS

6.2.1 Displaying Accuracy Metrics for each Column

```
##
## Accuracy metrics for: Aneurin Bevan
## -----
## ARIMA:
## $MAE
## [1] 1748.634
##
```

```

## $RMSE
## [1] 1992.543
##
## $MAPE
## [1] 0.1127431
##
## ETS:
## $MAE
## [1] 688.4172
##
## $RMSE
## [1] 801.6437
##
## $MAPE
## [1] 0.04825847
##
##
##
## Accuracy metrics for: Betsi Cadwaladr
## -----
## ARIMA:
## $MAE
## [1] 3157.257
##
## $RMSE
## [1] 3756.403
##
## $MAPE
## [1] 0.1557319
##
## ETS:
## $MAE
## [1] 915.7354
##
## $RMSE
## [1] 1050.122
##
## $MAPE
## [1] 0.04895248
##
##
##
## Accuracy metrics for: Cardiff & Vale
## -----
## ARIMA:
## $MAE
## [1] 600.4287
##
## $RMSE
## [1] 648.3758
##
## $MAPE
## [1] 0.04999195
##

```

```

## ETS:
## $MAE
## [1] 654.8113
##
## $RMSE
## [1] 799.0302
##
## $MAPE
## [1] 0.05682325
##
##
##
## Accuracy metrics for: Grouped_4_organisation
## -----
## ARIMA:
## $MAE
## [1] 1386.277
##
## $RMSE
## [1] 1598.559
##
## $MAPE
## [1] 0.0519391
##
## ETS:
## $MAE
## [1] 1089.509
##
## $RMSE
## [1] 1315.071
##
## $MAPE
## [1] 0.0434871
##
##
##
## Accuracy metrics for: Hywel Dda
## -----
## ARIMA:
## $MAE
## [1] 804.5973
##
## $RMSE
## [1] 923.8532
##
## $MAPE
## [1] 0.05832447
##
## ETS:
## $MAE
## [1] 763.0849
##
## $RMSE
## [1] 905.503

```

```

##
## $MAPE
## [1] 0.05915982
##
##
##
## Accuracy metrics for: Powys Teaching
## -----
## ARIMA:
## $MAE
## [1] 238.2648
##
## $RMSE
## [1] 264.6673
##
## $MAPE
## [1] 0.1922351
##
## ETS:
## $MAE
## [1] 120.9989
##
## $RMSE
## [1] 146.7962
##
## $MAPE
## [1] 0.09808414
##
##
##
## Accuracy metrics for: Total
## -----
## ARIMA:
## $MAE
## [1] 4781.817
##
## $RMSE
## [1] 5601.875
##
## $MAPE
## [1] 0.05378115
##
## ETS:
## $MAE
## [1] 4297.46
##
## $RMSE
## [1] 5014.551
##
## $MAPE
## [1] 0.05203606

```

- Most of the error metrics for ETS model are lower than the ARIMA model, except for Cardiff & Vale.
- The ARIMA model is based on the autoregressive and moving average components of the time series,

in line with the differencing method in order to make the data stationary. On the other hand, the ETS model focuses on the error, trend and seasonality components of the time series.

- According to the interest based on the NHS team, they would like to understand what is the best error metric to assess the quality of the forecast for the non-zero count data. Therefore, I would like to provide some insights for the error metrics:

1. Root Mean Squared Error (RMSE)

- **Description:** Measures the square root of the average squared differences between forecast and actual values.
- **Pros:** It gives more weight to large errors.
- **Cons:** It can be influenced significantly by outliers. RMSE might be inflated, if the count data is prone to significant spikes or declines, which should not be the problem for this data except for the Covid-19 era.

2. Mean Absolute Percentage Error (MAPE)

- **Description:** Measures the average of the absolute percentage errors.
- **Pros:** It is a relative metric, and is scale-independent and easy to interpret.
- **Cons:** MAPE can overemphasize relative errors on small counts.

3. Mean Absolute Error (MAE)

- **Description:** Measures the average of the absolute differences between forecast and actual values.
- **Pros:** It's less sensitive to outliers than RMSE. It provides a straightforward average error size.
- **Cons:** It does not emphasize large errors.

Recommendation:

- If you want to penalize larger errors more severely, **RMSE** might be the preferred choice if the larger errors need to be considered, such as the drops in the Covid-19 era.
- For a more balanced view of the average error, **MAE** can be considered.
- **MAPE** is useful and insightful if a relative metric is needed to as it gives error as a percentage of the actual counts.

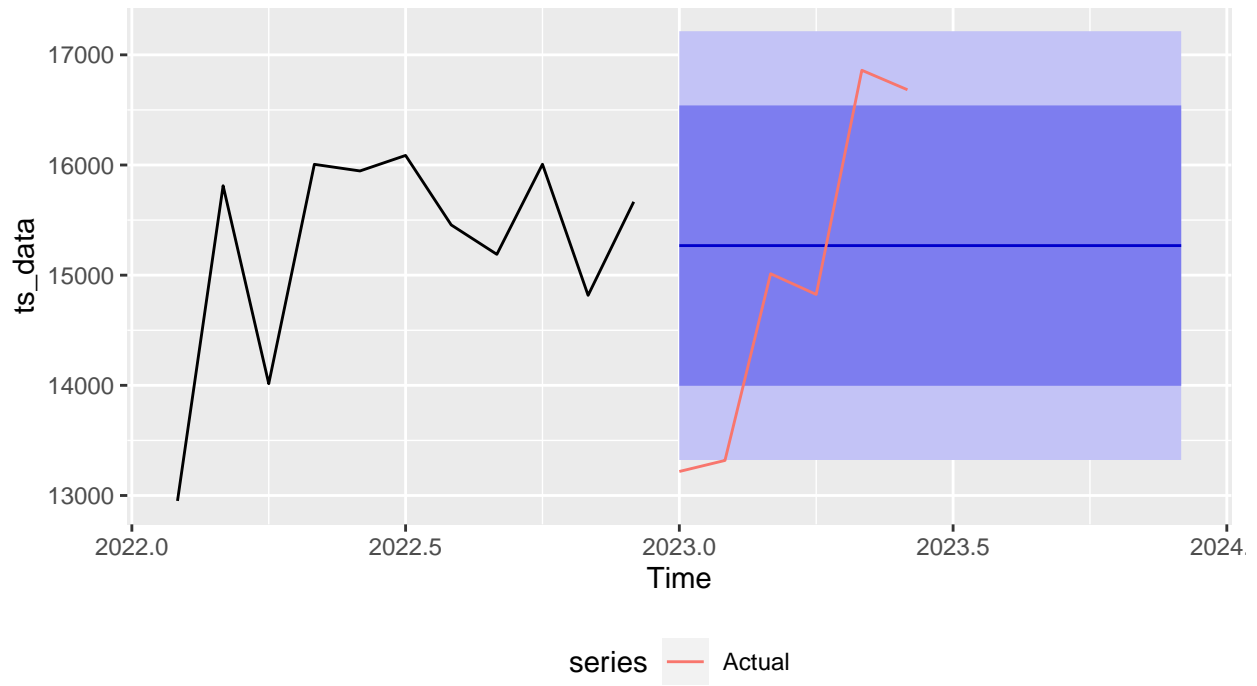
However, to get a comprehensive overview of the forecast accuracy, it is beneficial to examine on multiple error metrics. Additionally, the choice of the metrics also depend on the specific objectives.

7 Forecast after Covid-19

- Use the data after **2022 Jan** (after the significant drop appears in the graph)
- WHO says Covid-19 is no longer a global health emergency (2023 May)

Gumbrecht, J., Howard, J., & McPhillips, D. (2023, May 5). WHO says Covid-19 is no longer a global health emergency. CNN. <https://edition.cnn.com/2023/05/05/health/who-ends-covid-health-emergency/index.html>

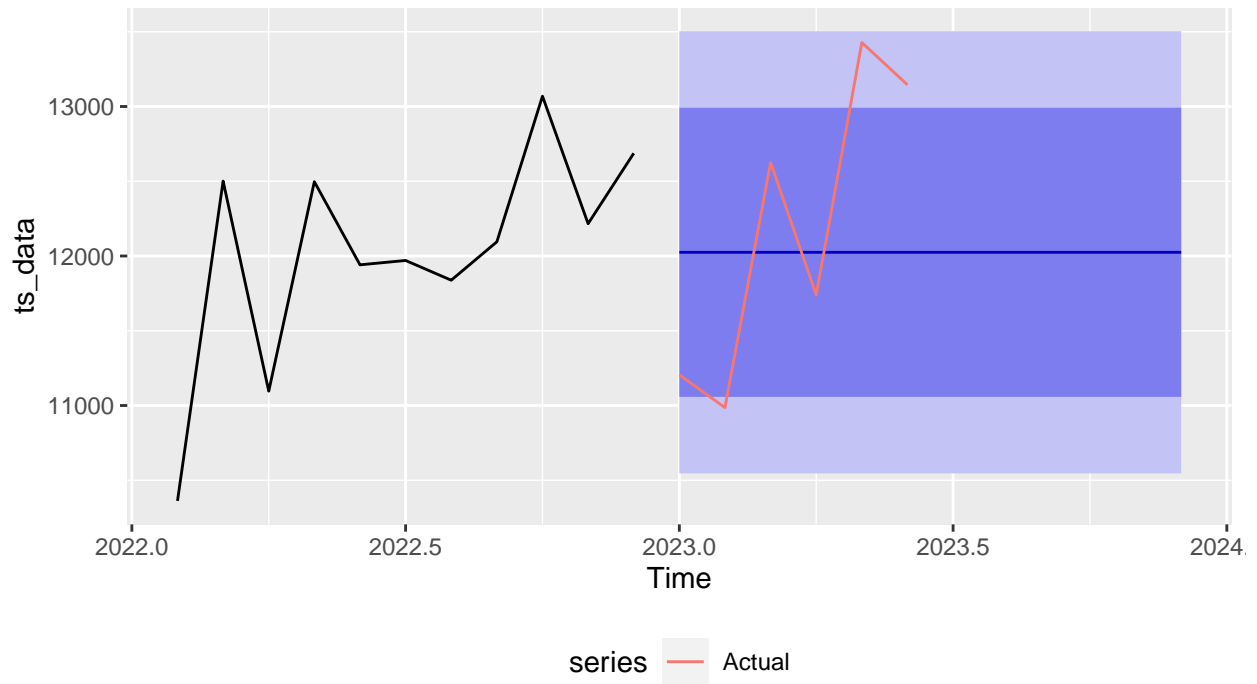
Forecast vs Actual for Aneurin Bevan



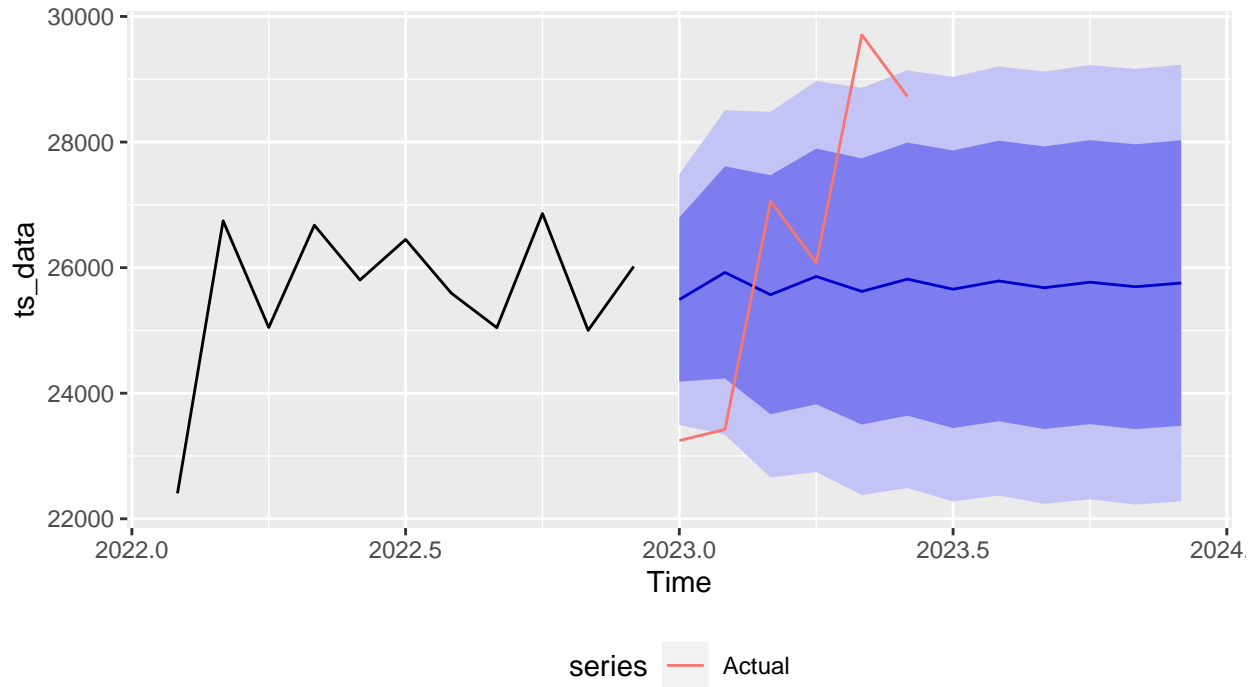
Forecast vs Actual for Betsi Cadwaladr

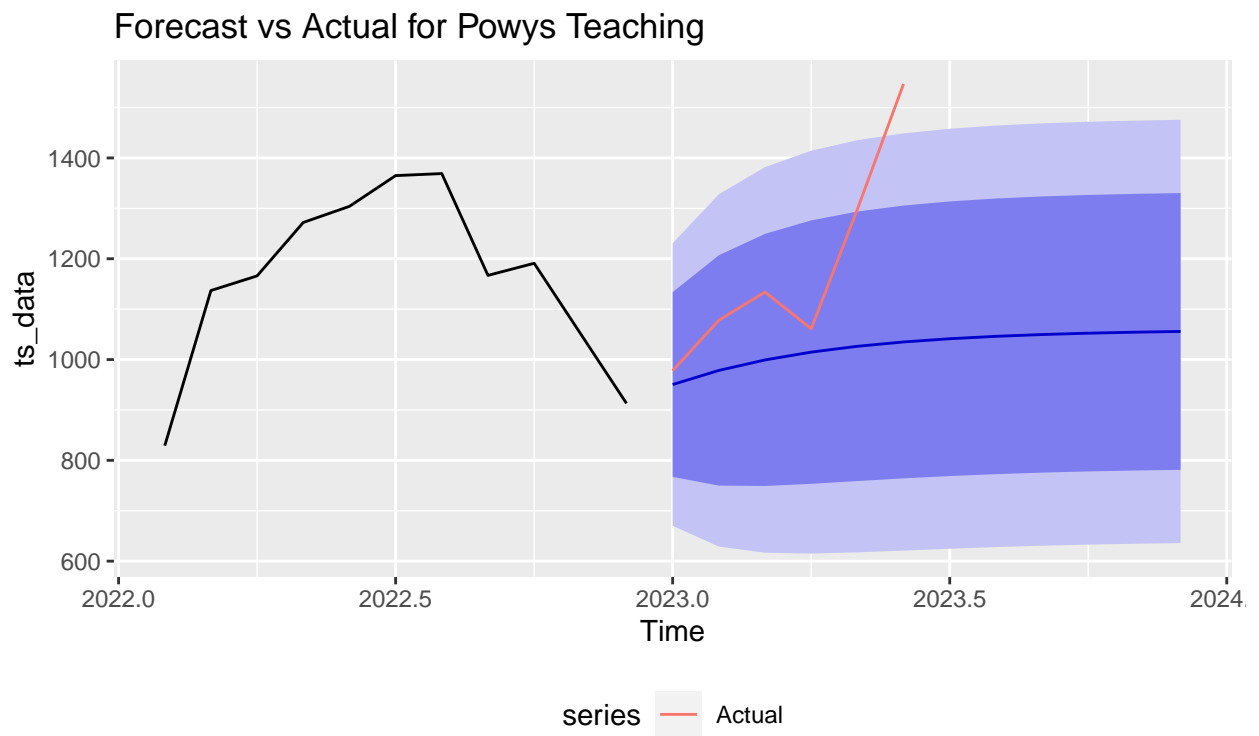
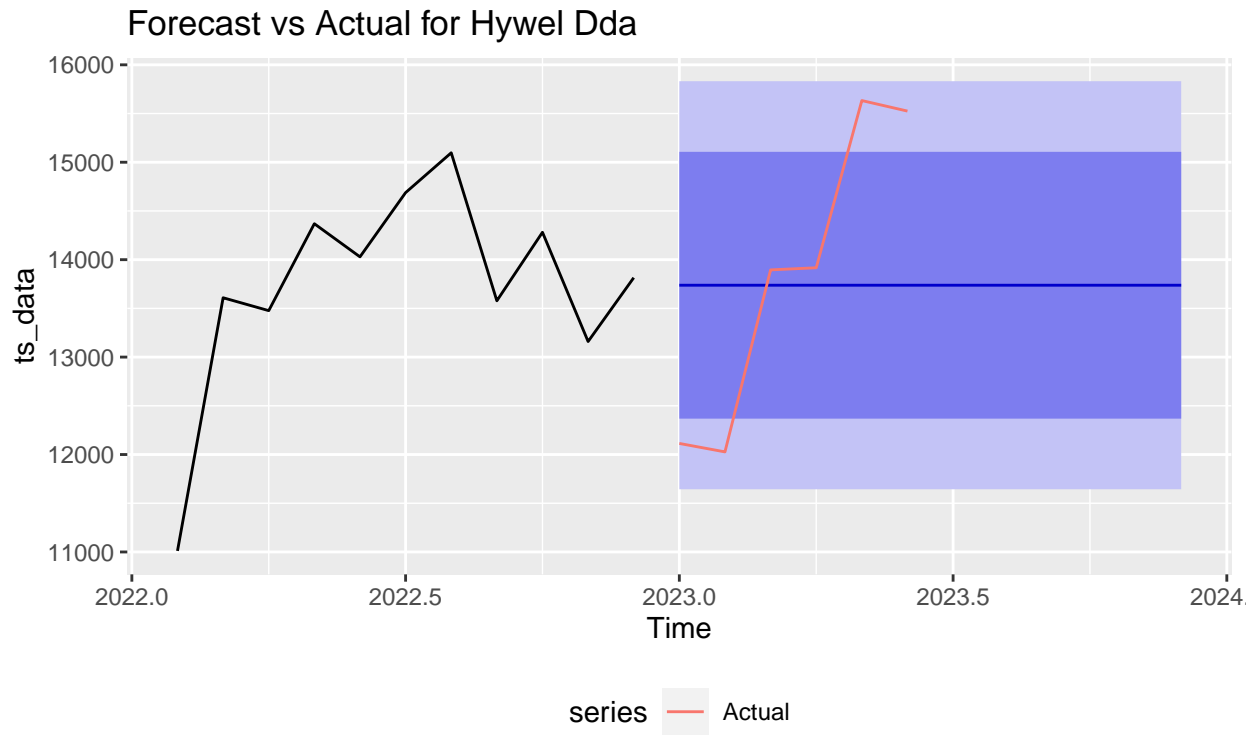


Forecast vs Actual for Cardiff & Vale

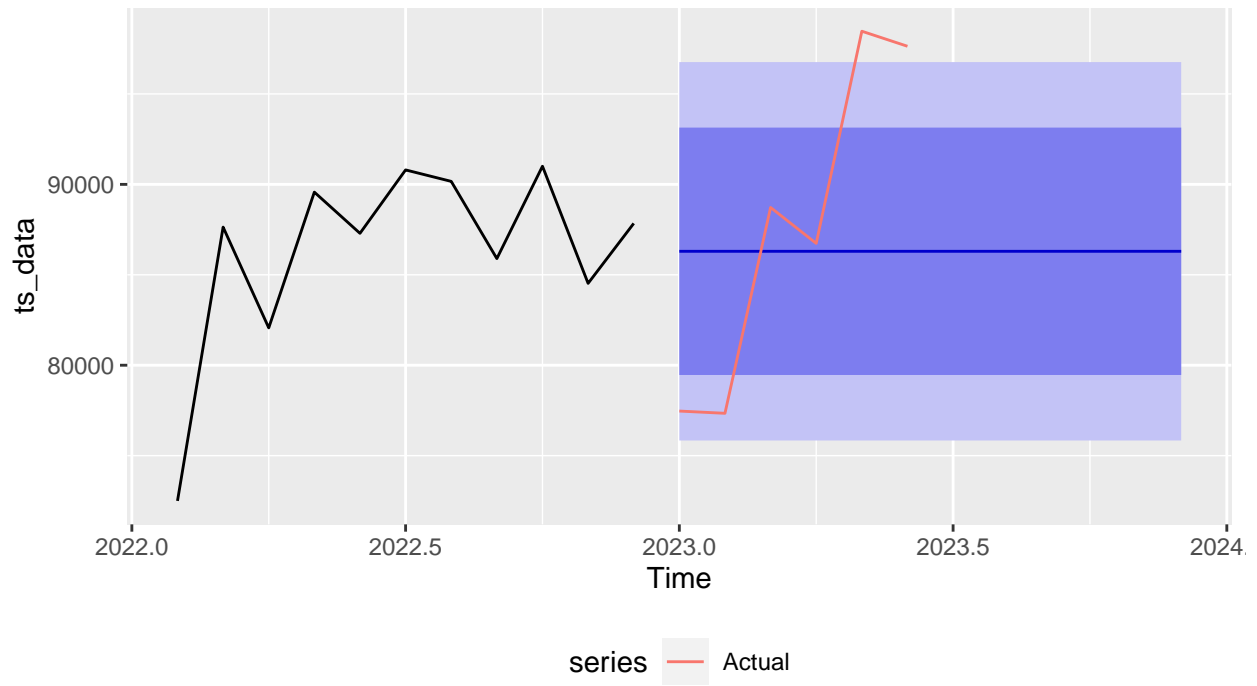


Forecast vs Actual for Grouped_4_organisation

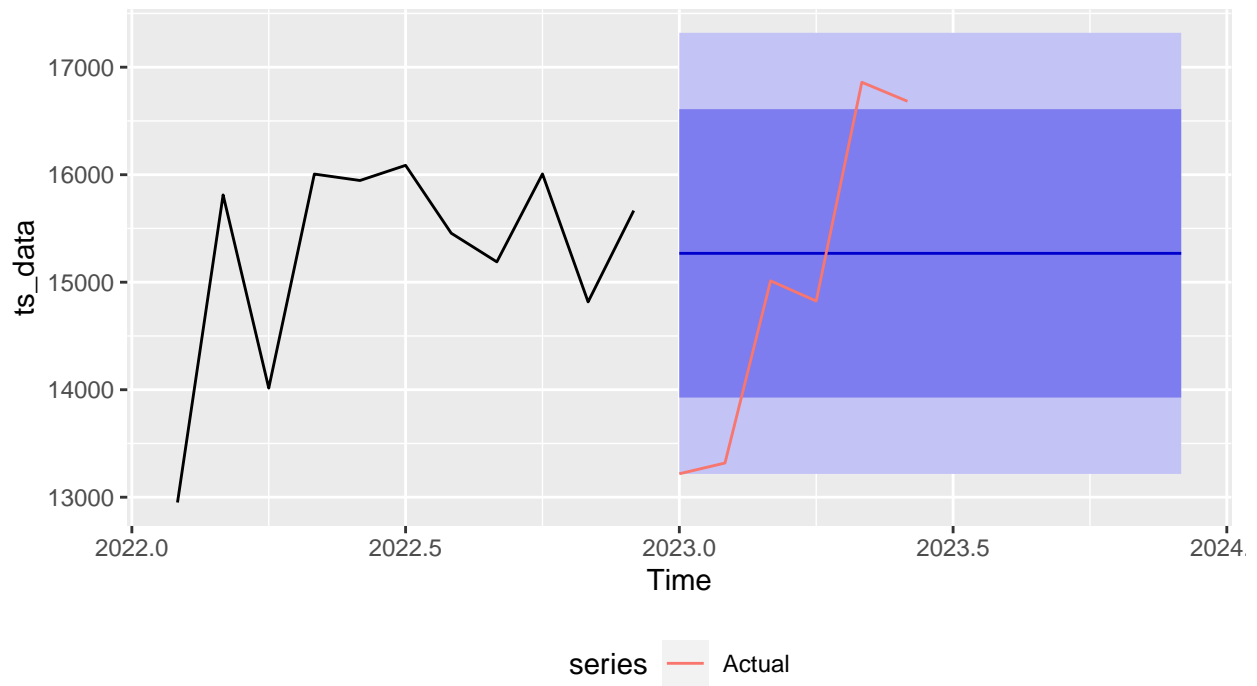




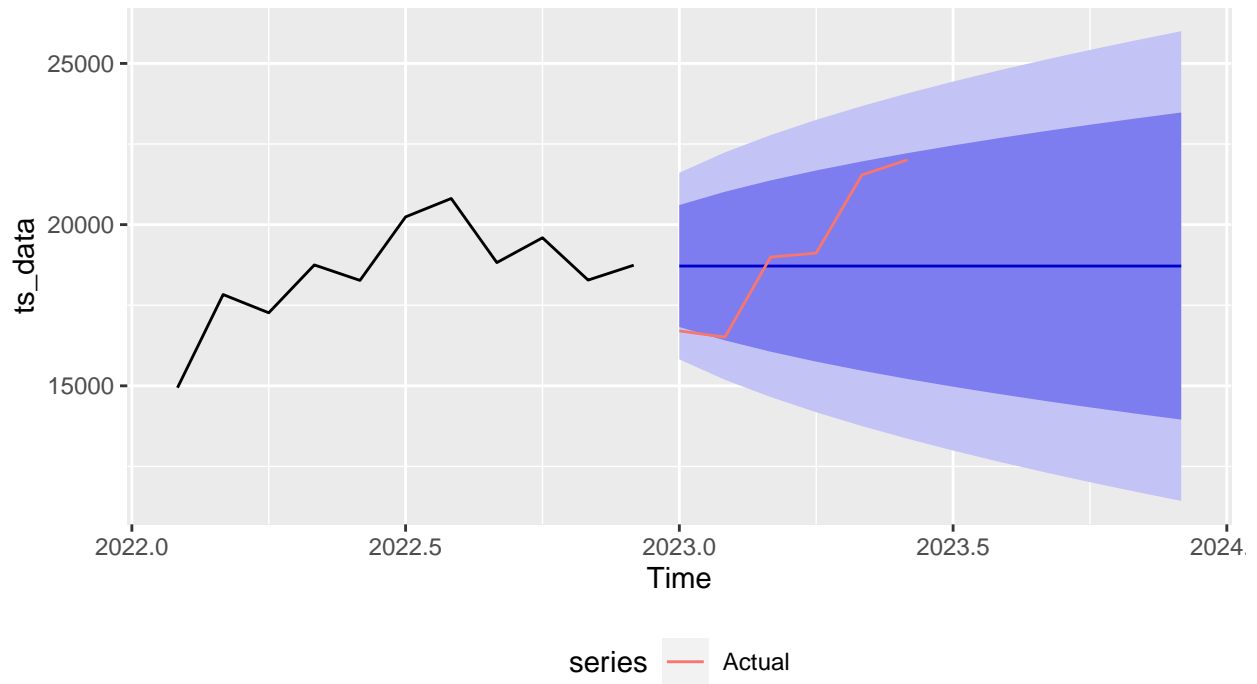
Forecast vs Actual for Total



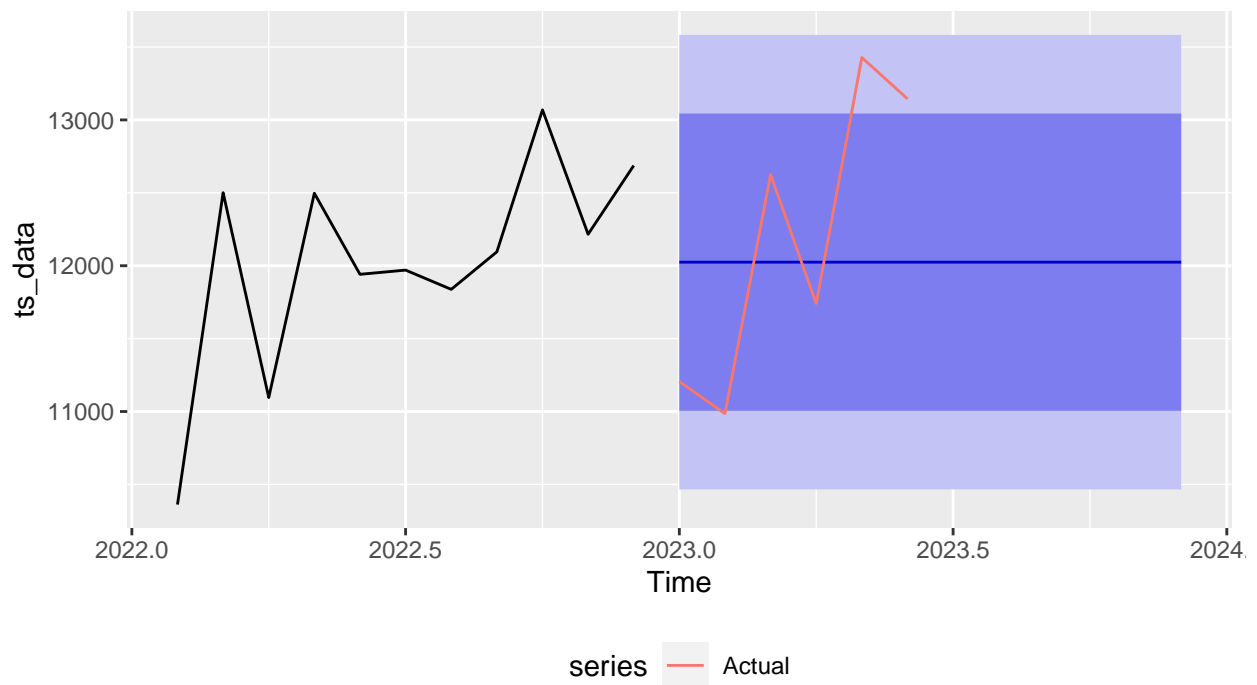
ETS Forecast vs Actual for Aneurin Bevan

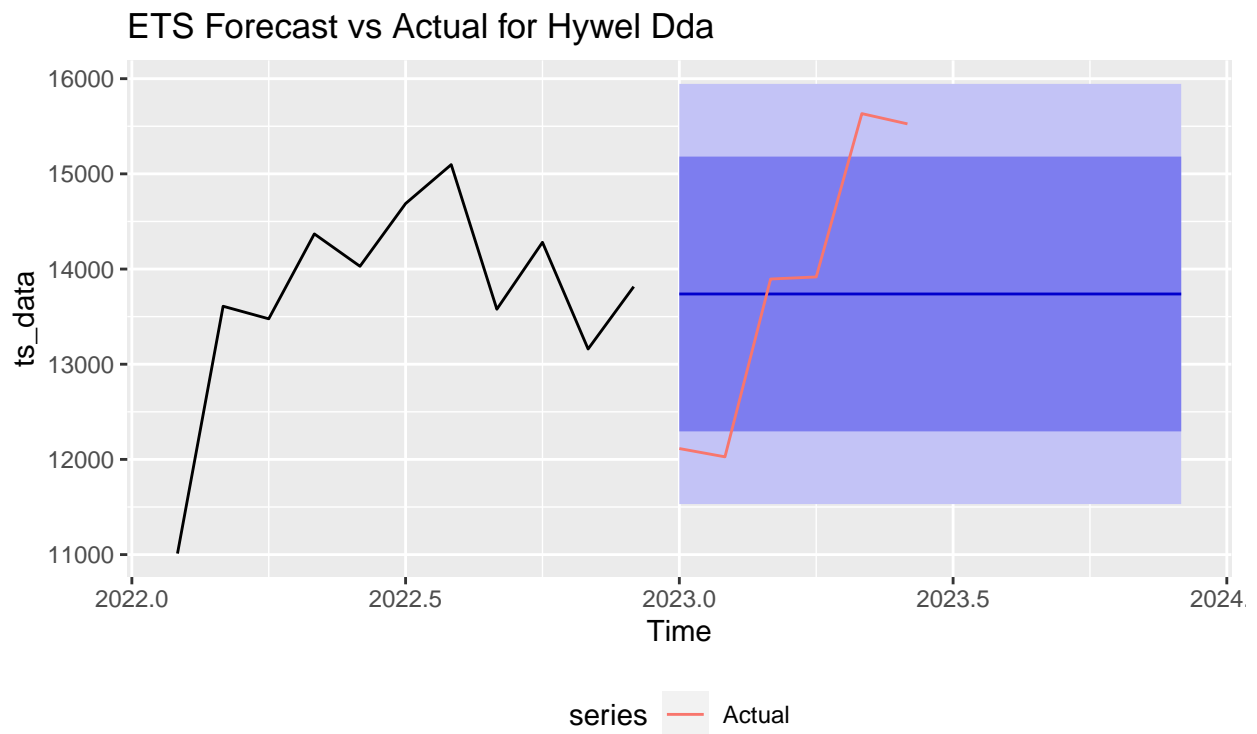
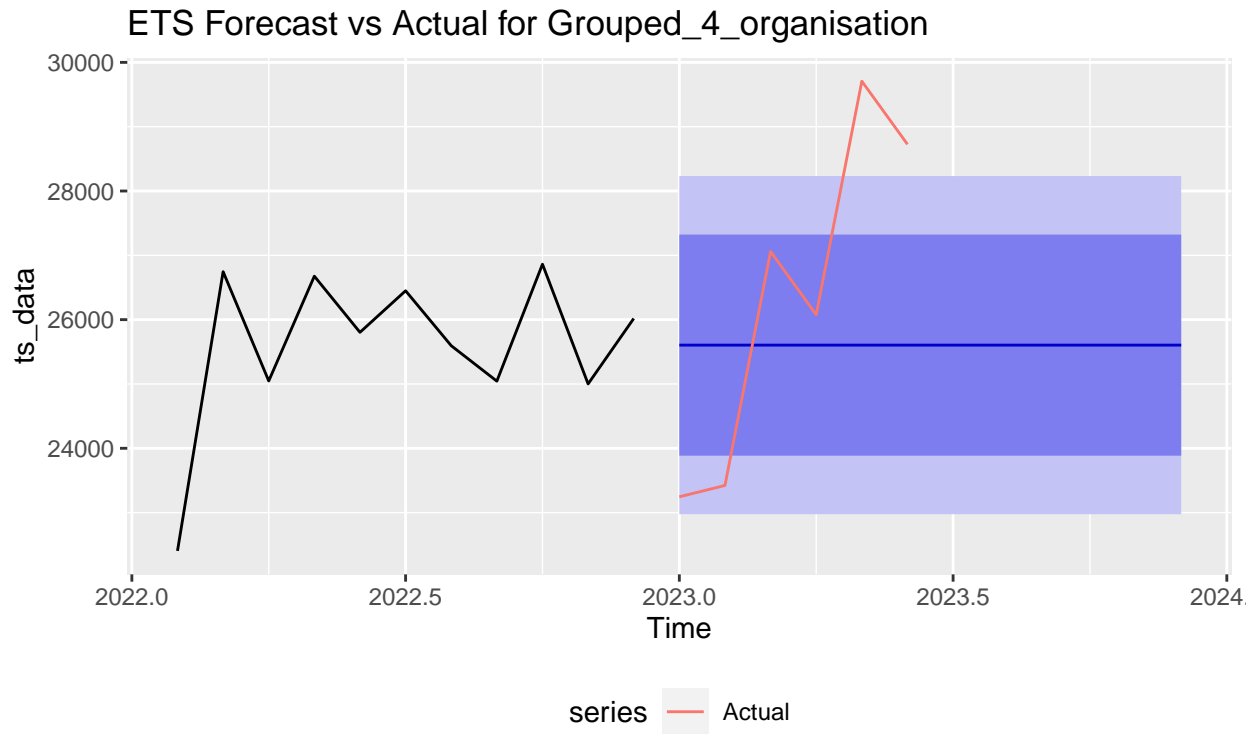


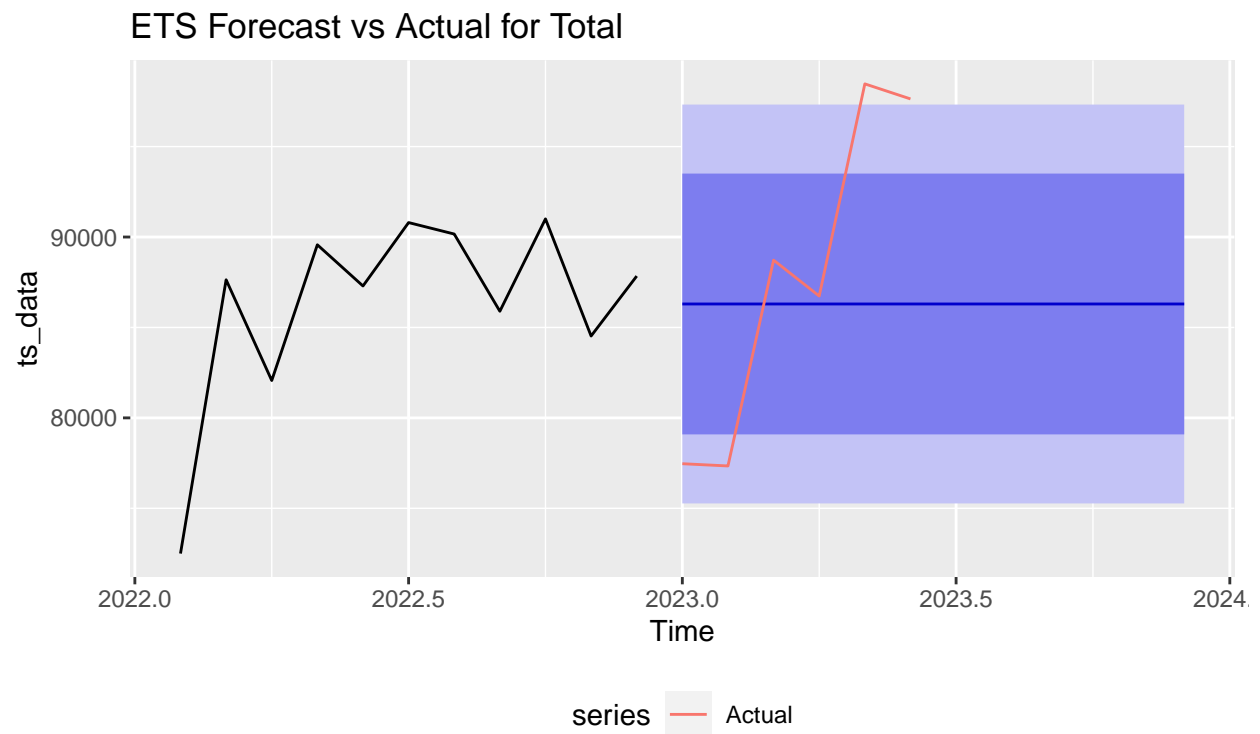
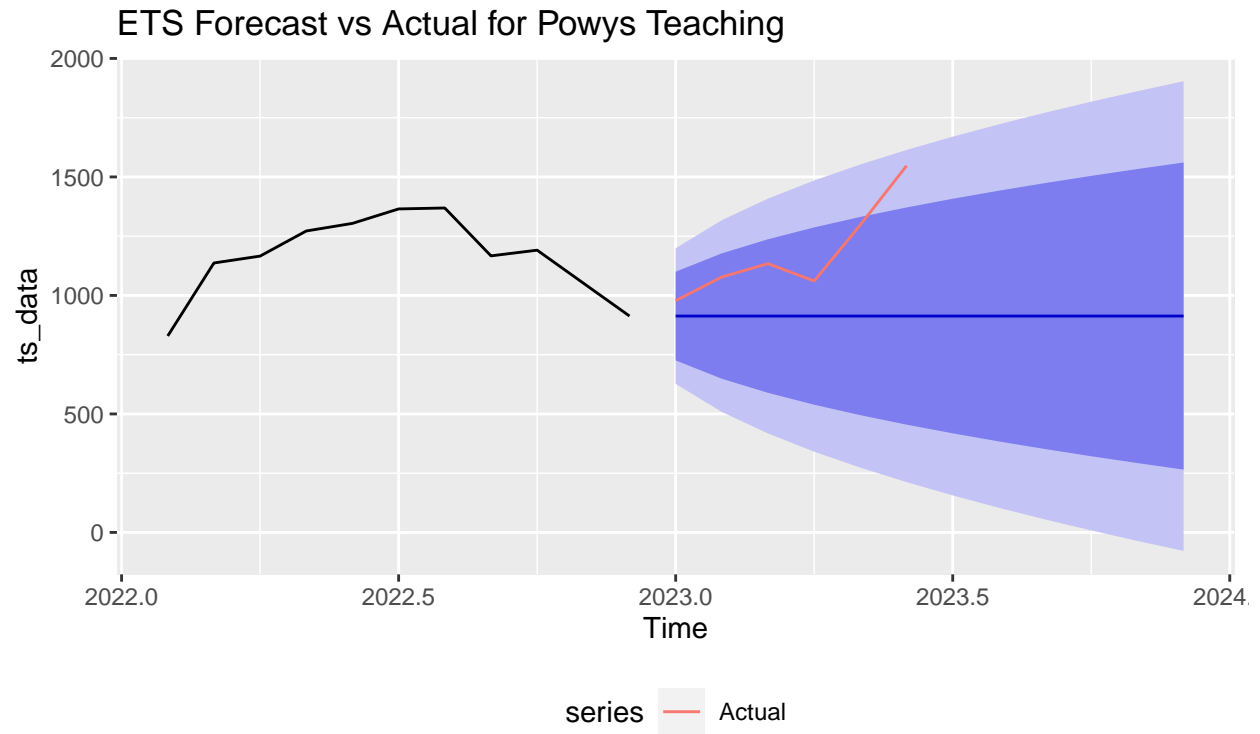
ETS Forecast vs Actual for Betsi Cadwaladr



ETS Forecast vs Actual for Cardiff & Vale

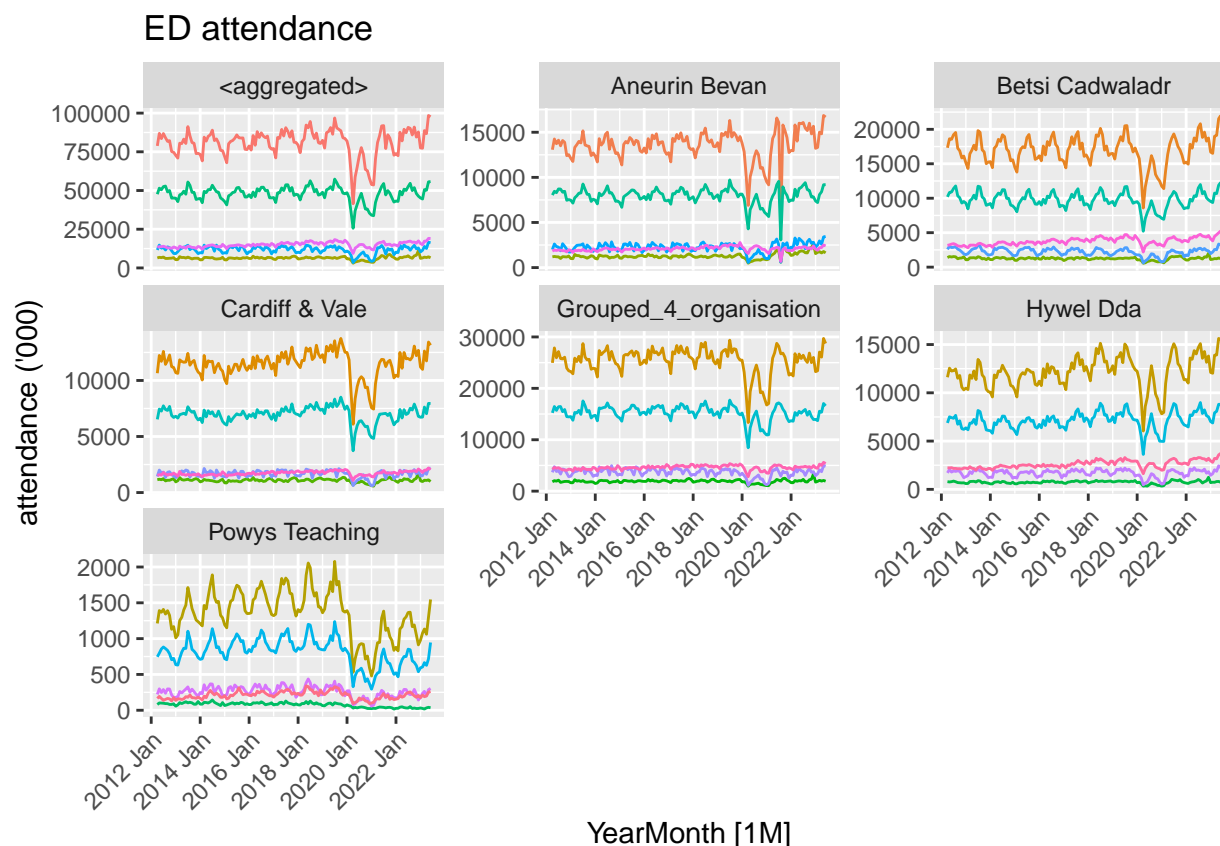






- A flat line: No trend, no seasonality, not sure if it's because of the forecast horizon.

8 Reconciliation



```
## # A fable: 336 x 5 [1M]
## # Key:      Aggregated_Organisation, .model [14]
##   Aggregated_Organisation .model YearMonth      attendance .mean
##   <chr*>                  <chr>      <mt>         <dist> <dbl>
## 1 Aneurin Bevan          ets        2023 Jul N(16995, 1845770) 16995.
## 2 Aneurin Bevan          ets        2023 Aug  N(14940, 2e+06) 14940.
## 3 Aneurin Bevan          ets        2023 Sep N(16156, 2227494) 16156.
## 4 Aneurin Bevan          ets        2023 Oct N(15976, 2418356) 15976.
## 5 Aneurin Bevan          ets        2023 Nov N(15069, 2609218) 15069.
## 6 Aneurin Bevan          ets        2023 Dec N(14907, 2800080) 14907.
## 7 Aneurin Bevan          ets        2024 Jan  N(14655, 3e+06) 14655.
## 8 Aneurin Bevan          ets        2024 Feb N(13971, 3181804) 13971.
## 9 Aneurin Bevan          ets        2024 Mar N(15665, 3372666) 15665.
## 10 Aneurin Bevan         ets        2024 Apr N(15149, 3563528) 15149.
## # i 326 more rows
```

```
## # A fable: 4,320 x 7 [1M]
## # Key:      Aggregated_Organisation, Grouped_Age, Hospital_ItemName_ENG, .model
## # [240]
##   Aggregated_Organisation Grouped_Age Hospital_ItemName_ENG .model YearMonth
##   <chr*>                  <chr*>      <chr*>         <chr>      <mt>
## 1 Aneurin Bevan          0-4          Nevill Hall Hospital base      2022 Jul
## 2 Aneurin Bevan          0-4          Nevill Hall Hospital base      2022 Aug
```

```
## 3 Aneurin Bevan      0-4      Nevill Hall Hospital base 2022 Sep
## 4 Aneurin Bevan      0-4      Nevill Hall Hospital base 2022 Oct
## 5 Aneurin Bevan      0-4      Nevill Hall Hospital base 2022 Nov
## 6 Aneurin Bevan      0-4      Nevill Hall Hospital base 2022 Dec
## 7 Aneurin Bevan      0-4      Nevill Hall Hospital base 2023 Jan
## 8 Aneurin Bevan      0-4      Nevill Hall Hospital base 2023 Feb
## 9 Aneurin Bevan      0-4      Nevill Hall Hospital base 2023 Mar
## 10 Aneurin Bevan     0-4      Nevill Hall Hospital base 2023 Apr
## # i 4,310 more rows
## # i 2 more variables: attendance <dist>, .mean <dbl>
```

