

hierarchical_forecasting_report

Janice Hsu

2023-09-30

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	6
2.5	Change the Age_Code structure into different groups for simplicity and interpretability . . .	7
2.6	Plot Number of Patients in different age groups	7
3	Forecast with ARIMA	8
3.1	Plotting using ARIMA	8
4	Forecasting with ETS	12
4.1	Plotting with ets	12
5	Forecasting with ETS	16
5.1	Plotting with ets	16
6	Accuracy assessment for ARIMA and ETS	19
6.1	Calculating Accuracy Metrics for ARIMA and ETS	19
7	Reconciliation	24
7.1	Aggregate data	24
7.2	Time series cross validation	24
7.3	Reconcile	29
7.4	Include Age Group	30

1 Preliminary Analysis

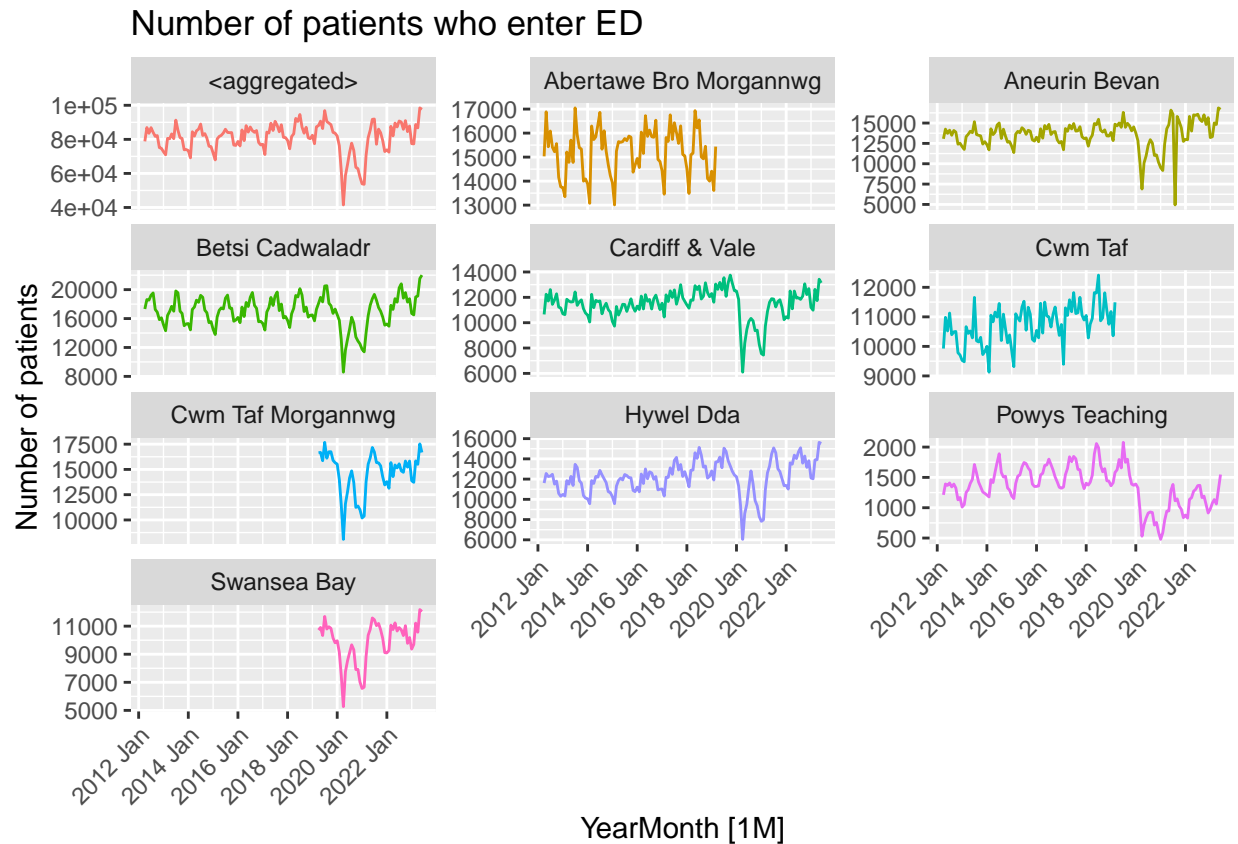
1.1 Data Introduction

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



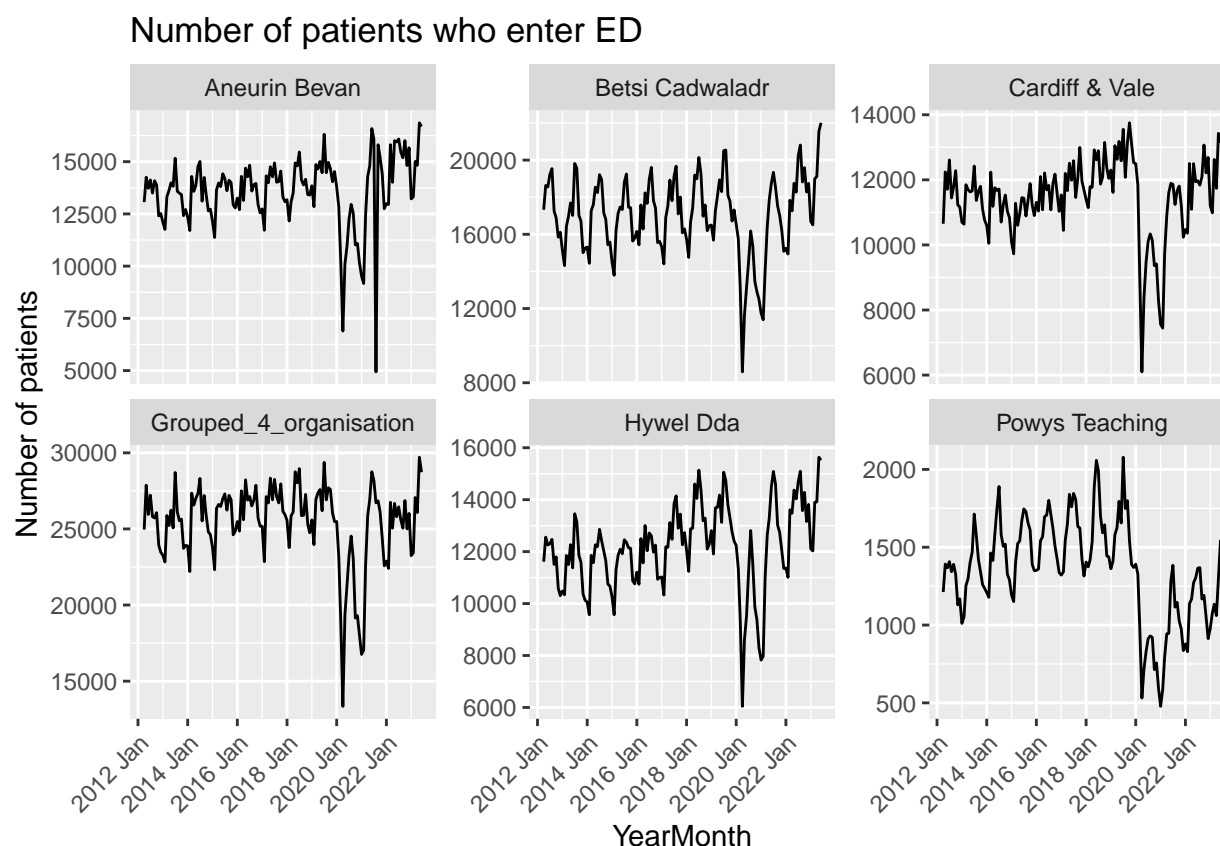
- A couple of Local Health Boards (LHBs) were redefined from the 1st of April 2019 onwards: Cwm Taf (27) → Cwm Taf Morgannwg (30) // Abertawe Bro Morgannwg (26) → Swansea Bay (31). Therefore, if you decide to forecast at LHB resolution, you might want to consider these 4 as a unique one.
- A the Princess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



Analysing Health Board Attendance Trends Amidst and Post-COVID-19:

Healthcare systems around the world faced significant difficulties during the COVID-19 epidemic, which caused them to quickly shift into crisis-response mode. Intriguing trends during and after the epidemic era have been revealed by a thorough review of attendance data from local health boards.

- During the pandemic:

Surprisingly, several local health boards experienced a large reduction in attendance throughout the epidemic period. This significant drop in attendance, which may have been brought on by reasons including stringent public health regulations, widespread concern over virus exposure, and maybe changes in healthcare delivery strategies, was an anomalous departure from historical attendance trends.

- Post pandemic:

After the pandemic's intensity started to lessen, there was a noticeable increase in attendance, however there were notable differences amongst the local health boards. In contrast to most boards, which showed a strong recovery and raised their attendance figures to levels prior to the pandemic, Powys Teaching stood out as an outlier and defied this recovery trend. To fully understand the subtleties affecting these disparate courses and influence upcoming strategic planning, more research is necessary.

- Findings on seasonality:

Initial data analysis reveals the existence of seasonality, which is characterized by repeated swings in attendance across all health boards. This pattern calls for a perceptive investigation into:

Identify Underlying Causes: Researching seasonality's probable triggers, such as public health initiatives, flu seasons, or holiday seasons, in order to comprehend the factors influencing these patterns.

Determine Seasonal influence: Establishing whether the observed seasonality has an equal influence on all health boards or whether there are differences, which may be a sign of regional causes or policy changes.

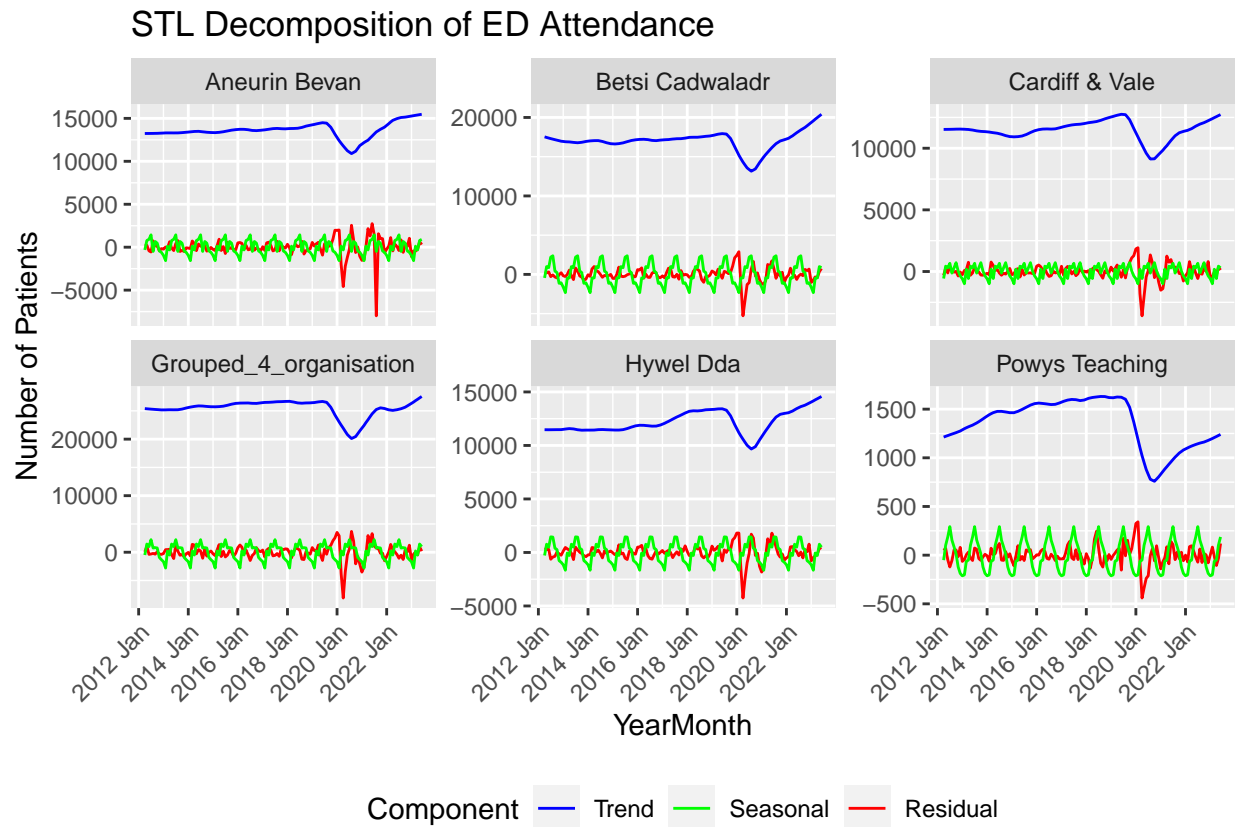
Planning and Forecasting: By using this understood seasonality to improve forecasting models and ensuring that they are calibrated to account for these seasonal changes, more precise and useful predictive insights are made possible.

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



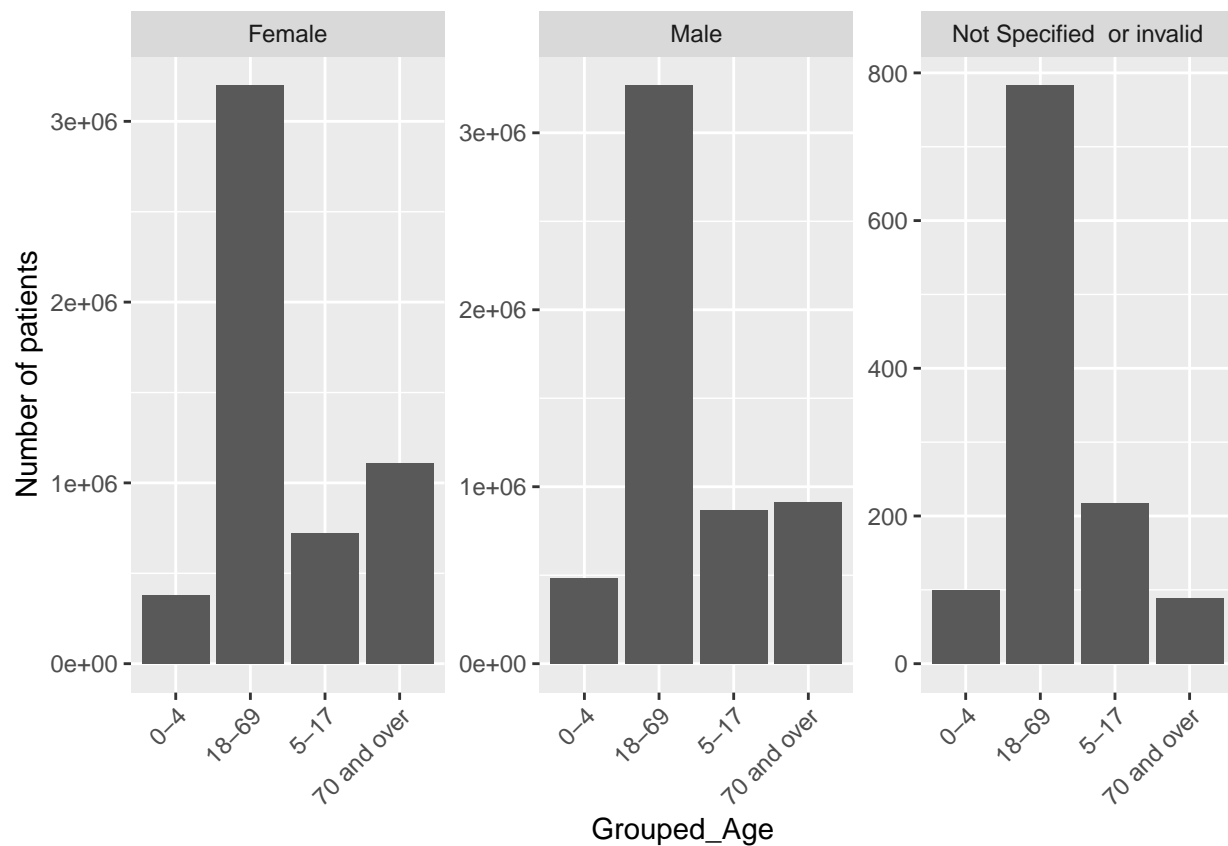
Based on the STL decomposition, it is evident that each health board follows a similar trend. Moreover, it is obvious that there is a significant downturn during the Covid-19 era. As for seasonality, there is a pronounced surge in the number of patient's attendance in the middle of the year (approximately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

2.5 Change the Age_Code structure into different groups for simplicity and interpretability

```
## [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

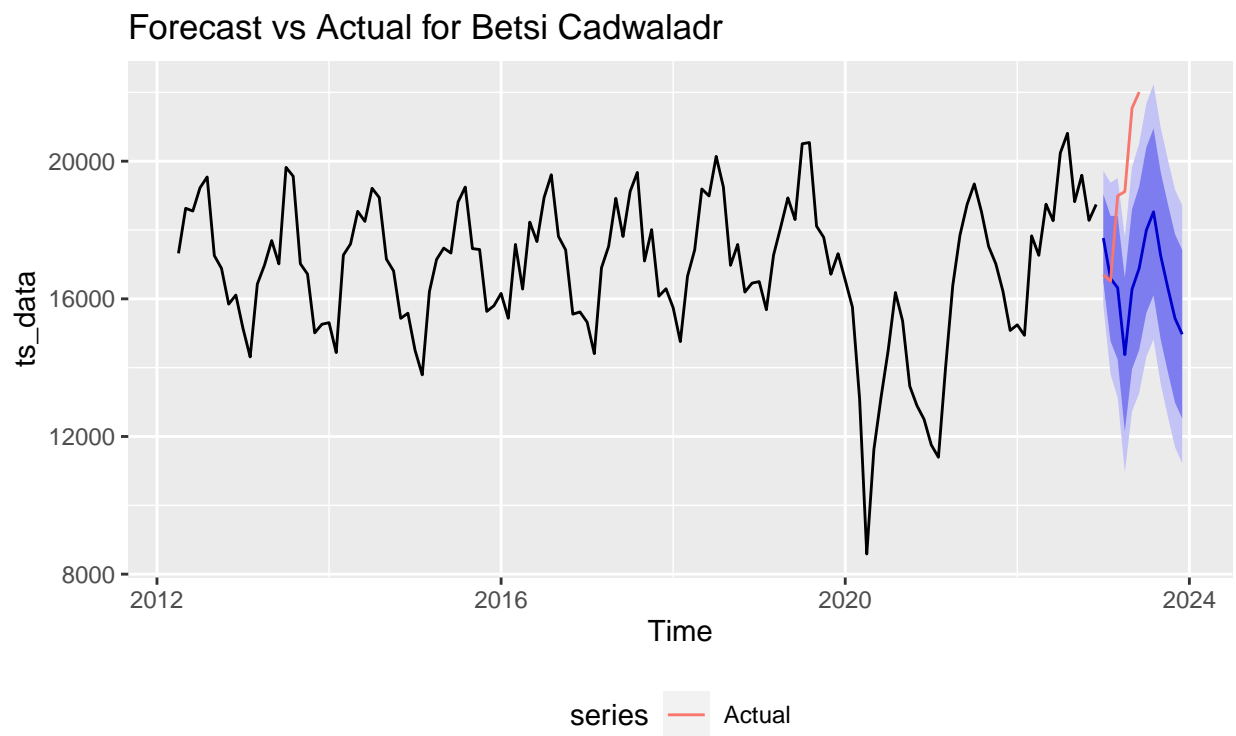
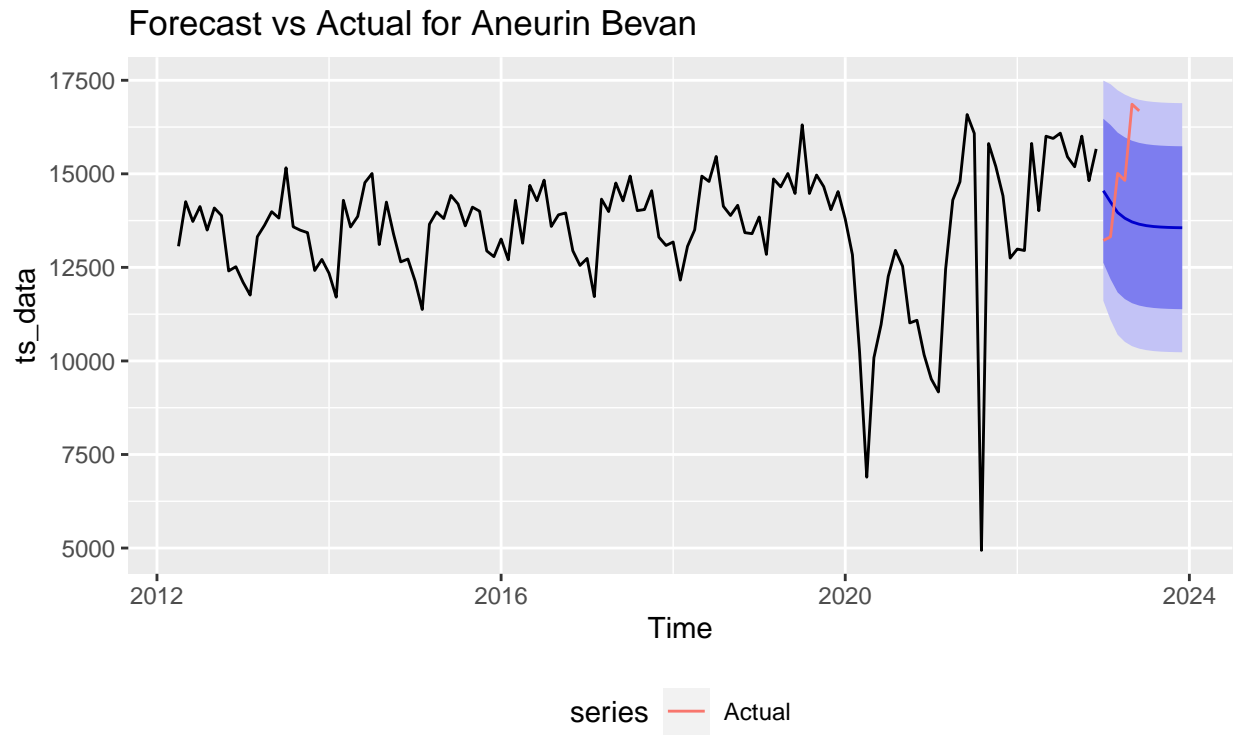


Findings:

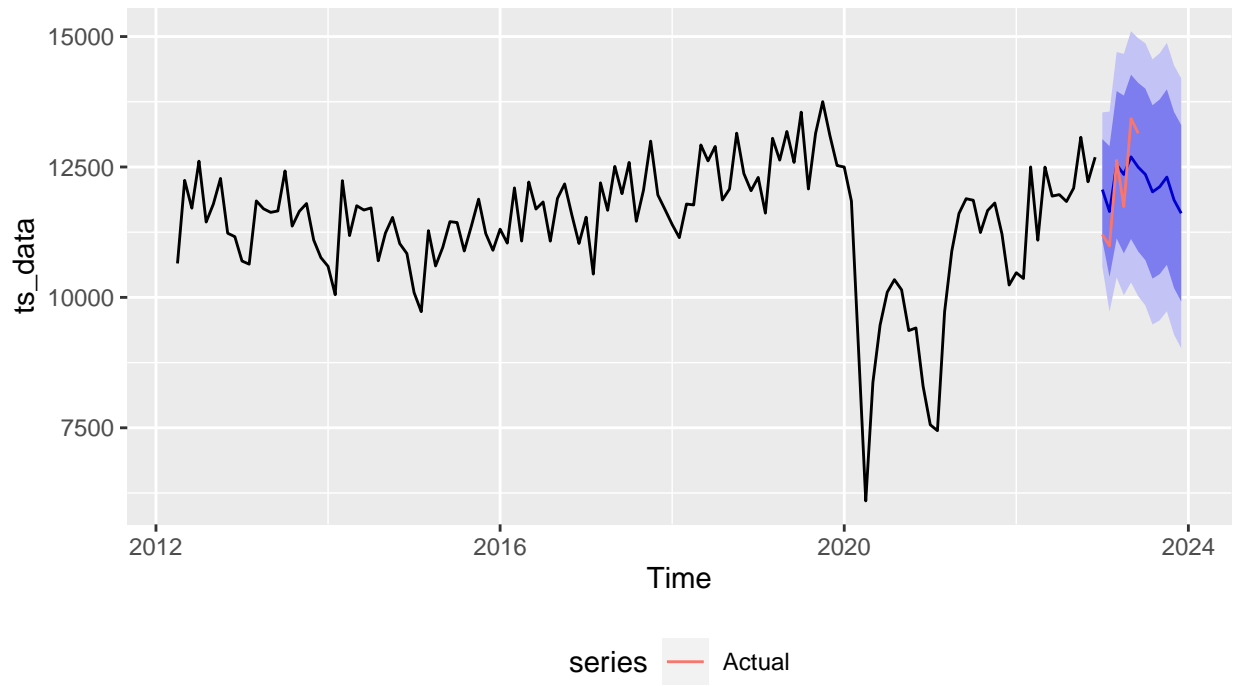
- The observation that the age group 18-69 has the most amount of patient attendance is expected, as it is the biggest group among all. However, it is noteworthy that the second biggest group are from the oldest age bracket, aligning with the general understanding of the health care need for the elders.

3 Forecast with ARIMA

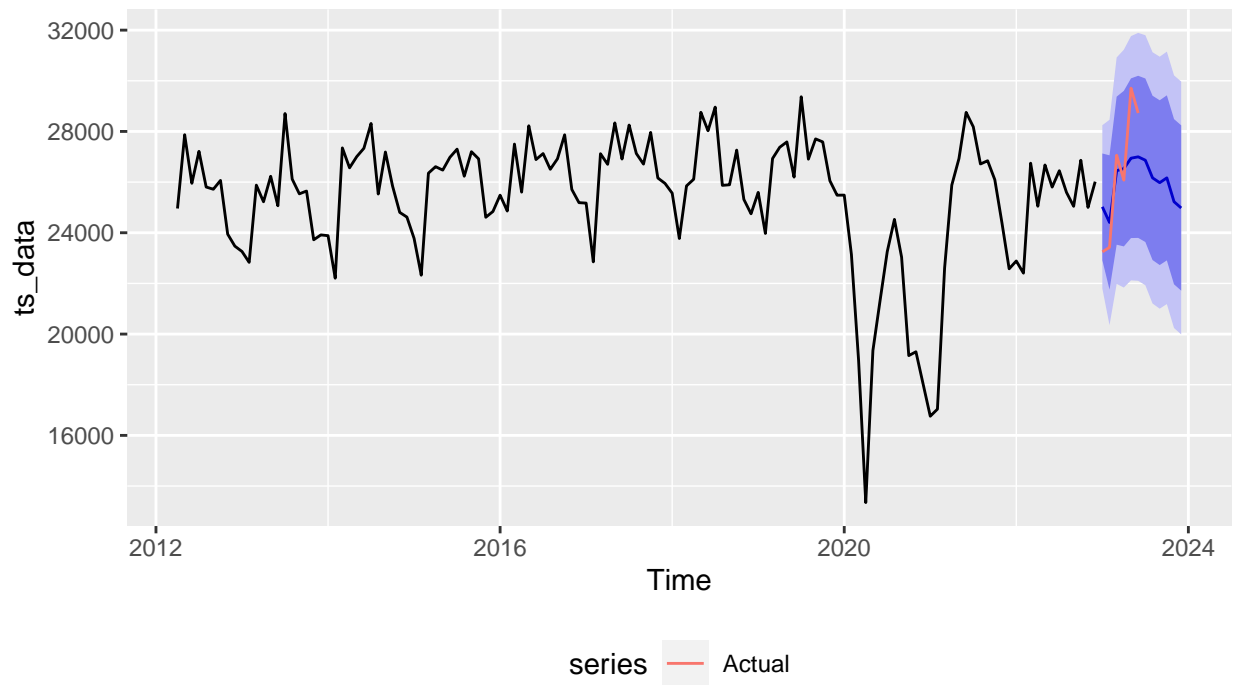
3.1 Plotting using ARIMA



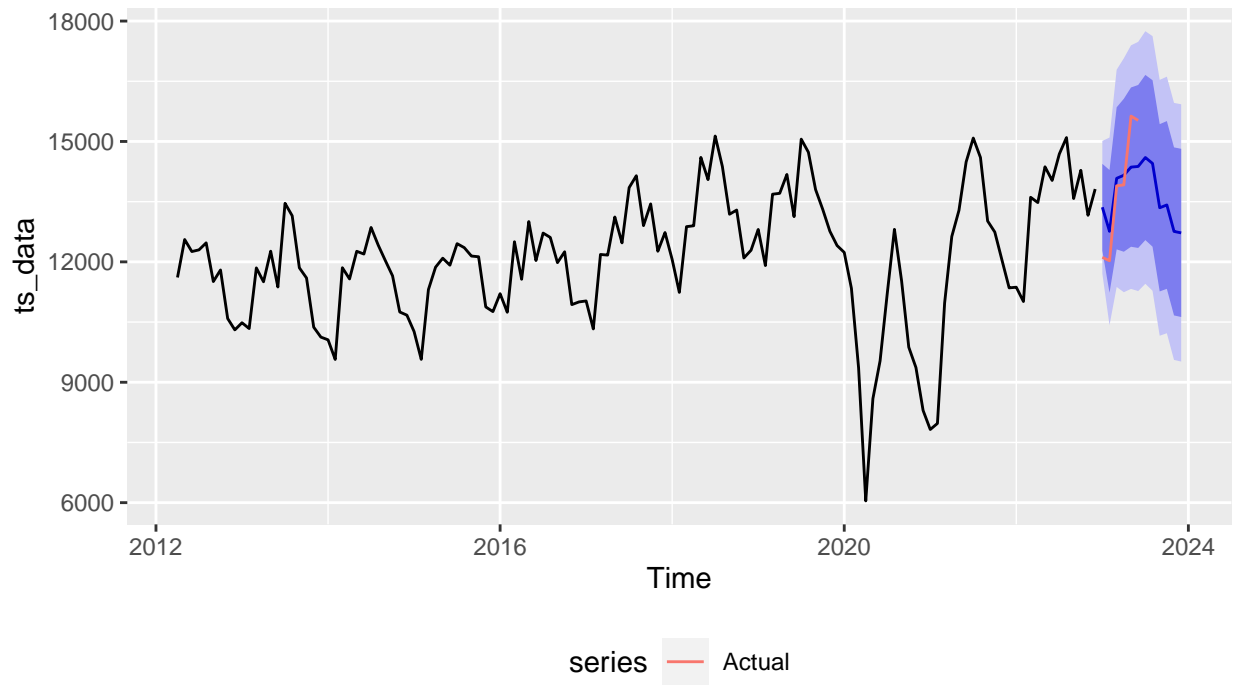
Forecast vs Actual for Cardiff & Vale



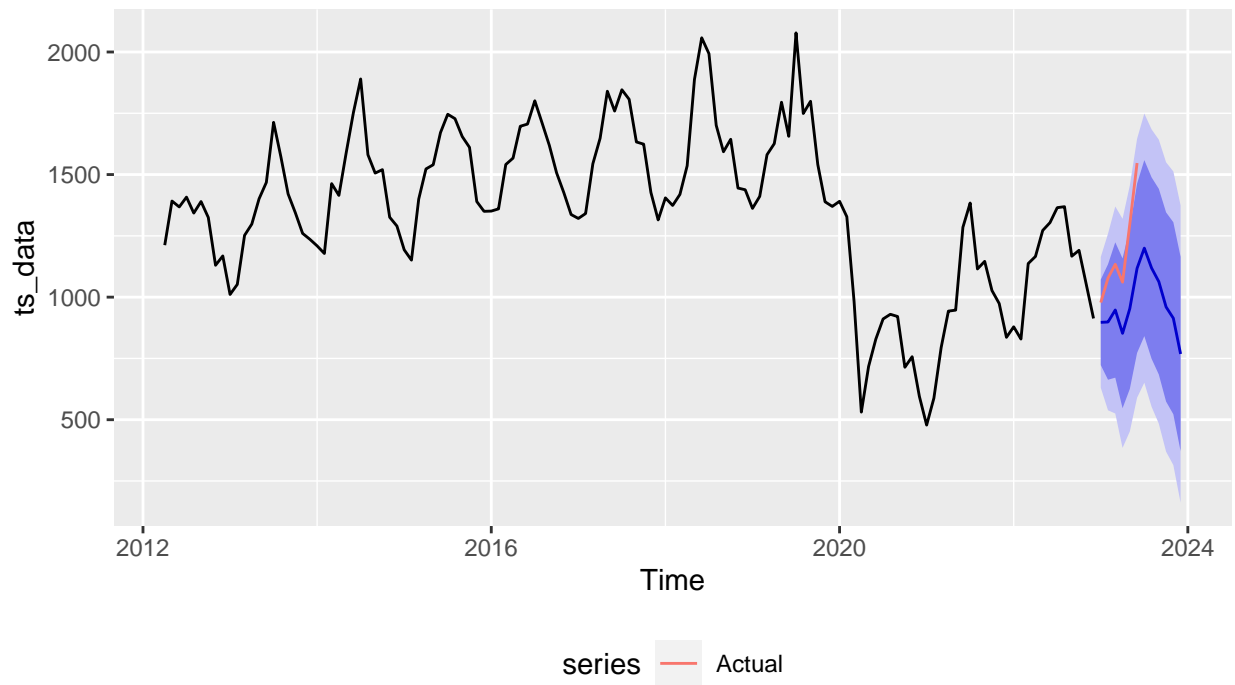
Forecast vs Actual for Grouped_4_organisation



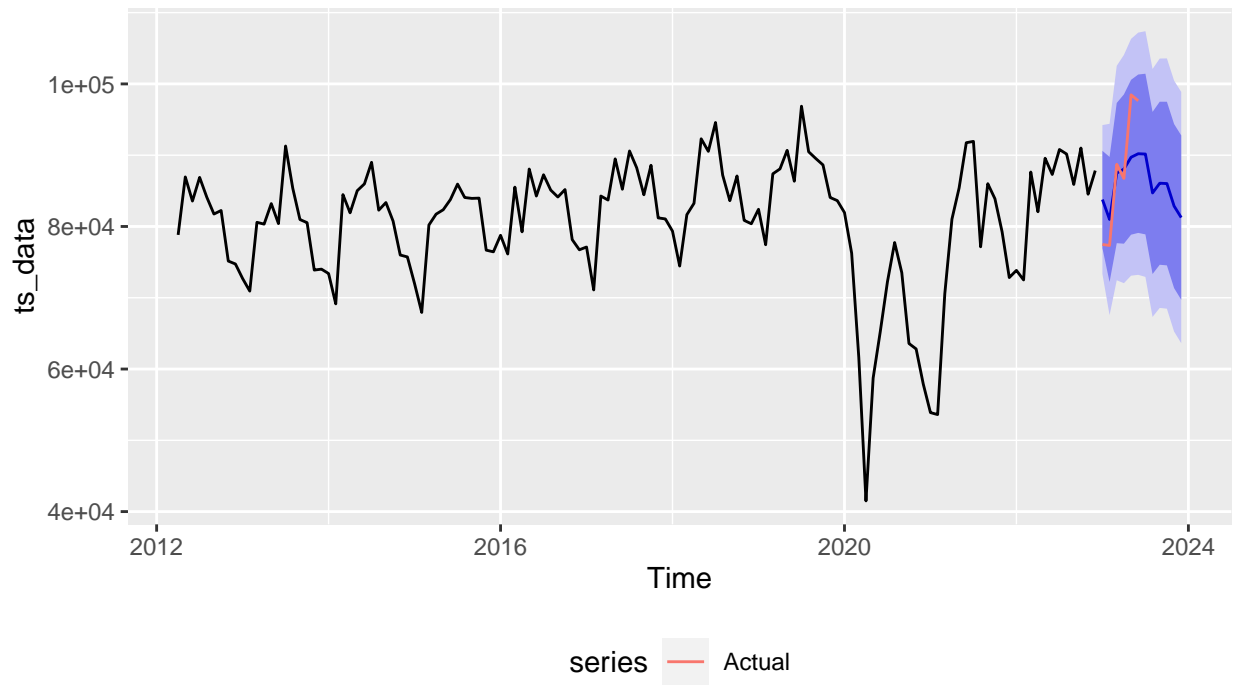
Forecast vs Actual for Hywel Dda



Forecast vs Actual for Powys Teaching

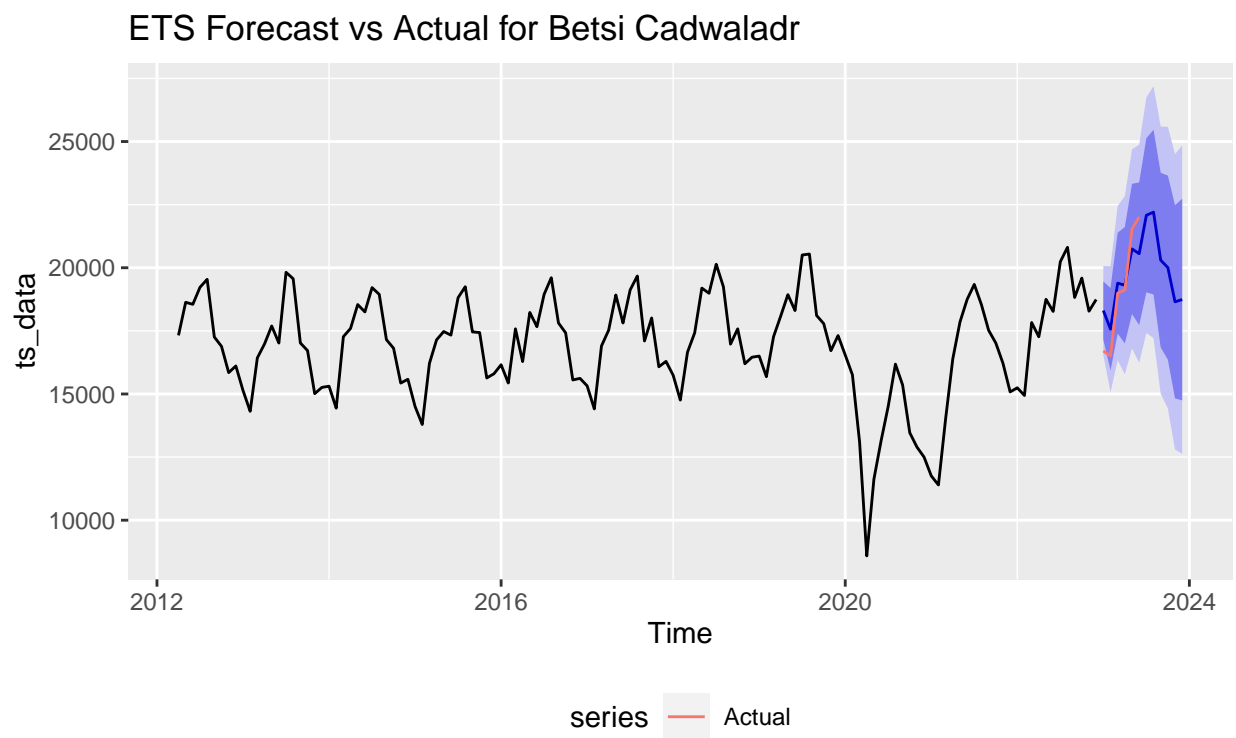
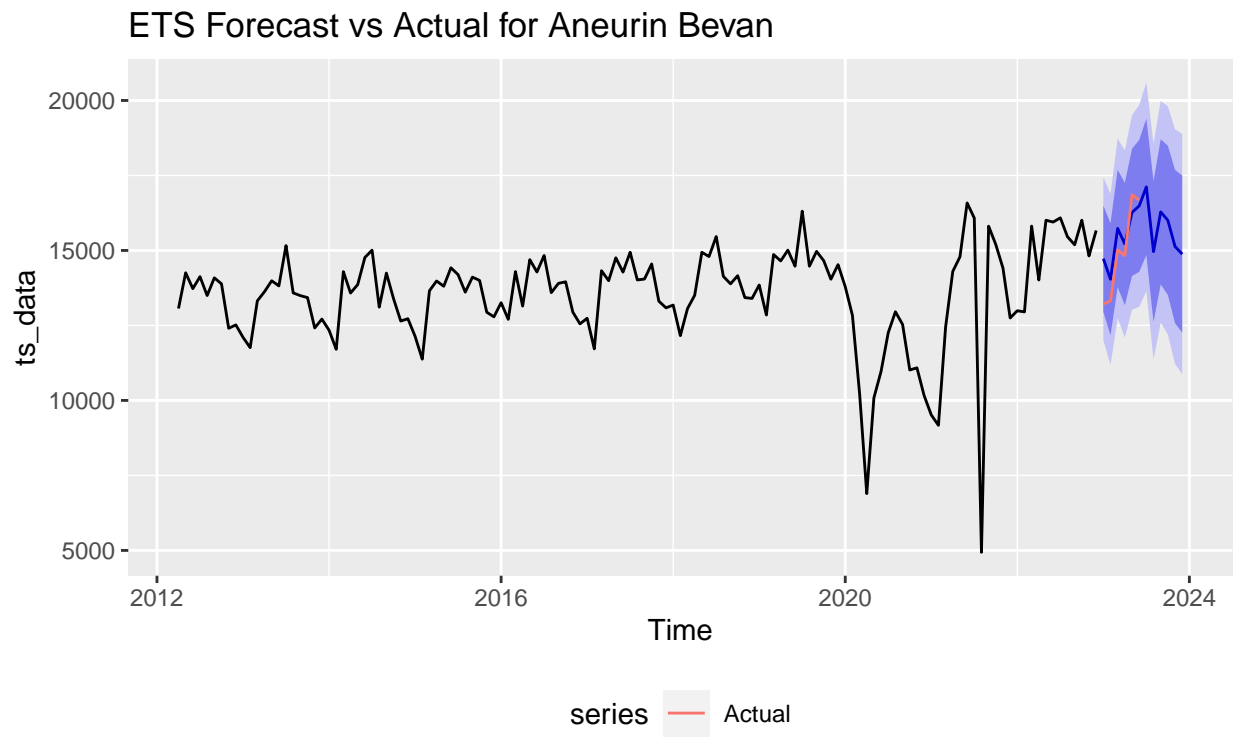


Forecast vs Actual for Total

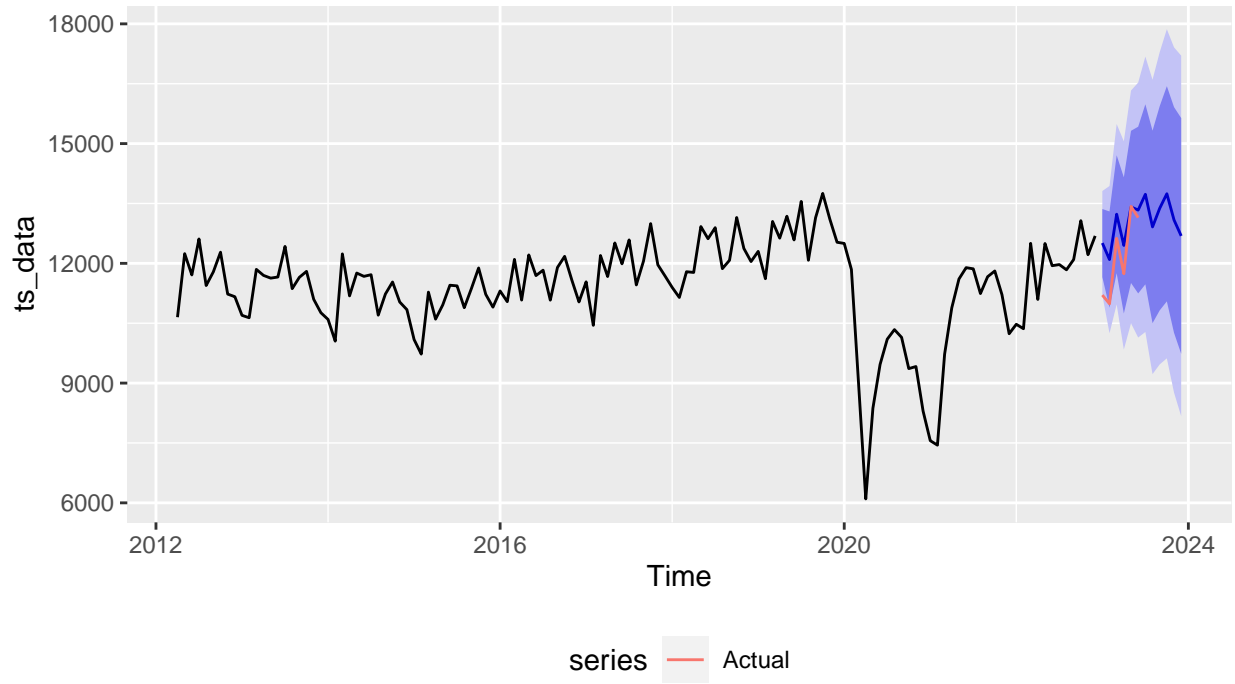


4 Forecasting with ETS

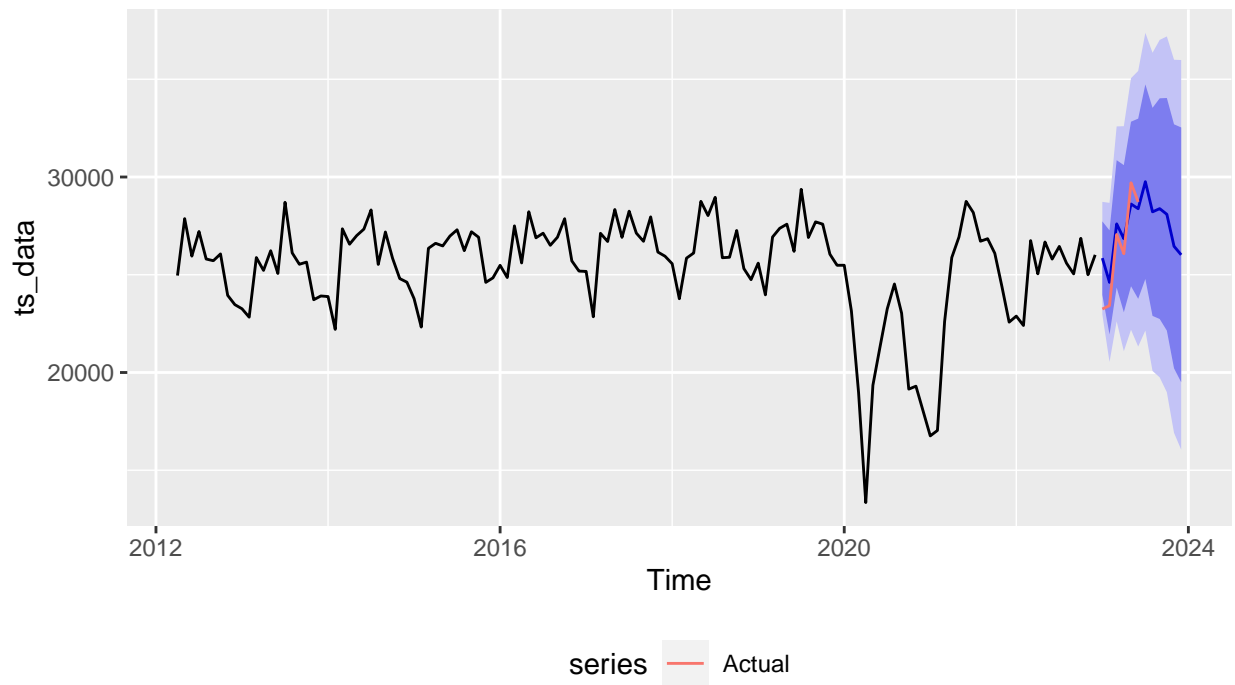
4.1 Plotting with ets



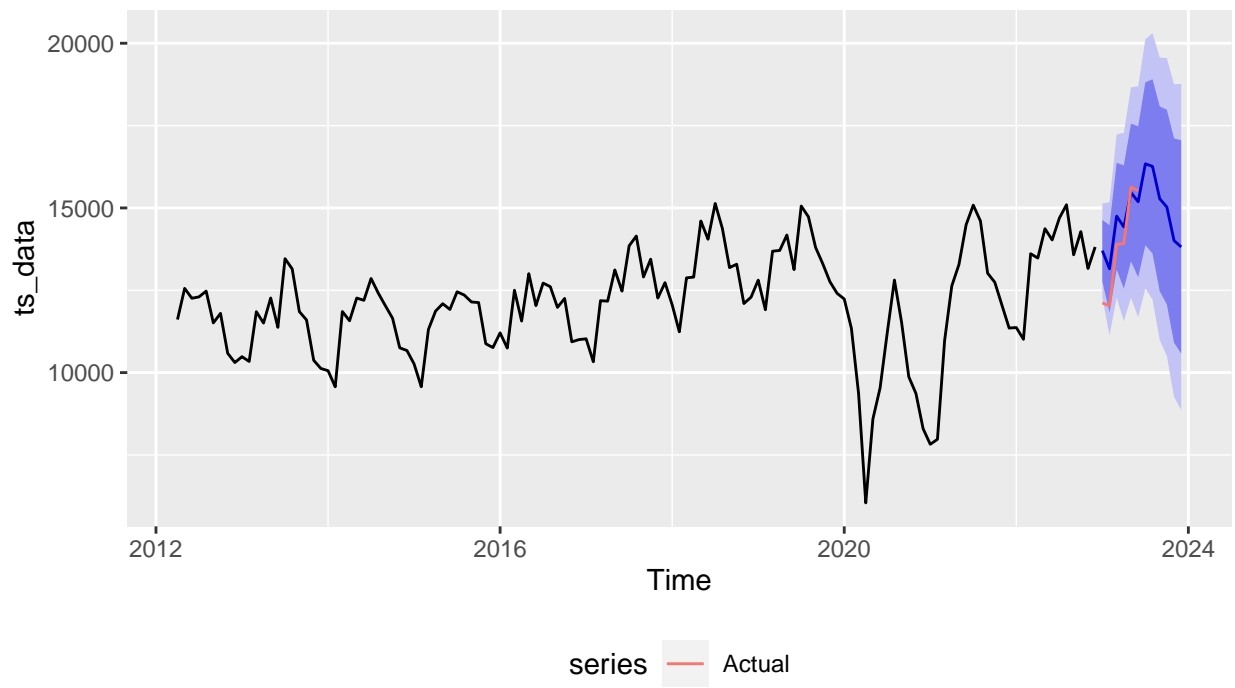
ETS Forecast vs Actual for Cardiff & Vale



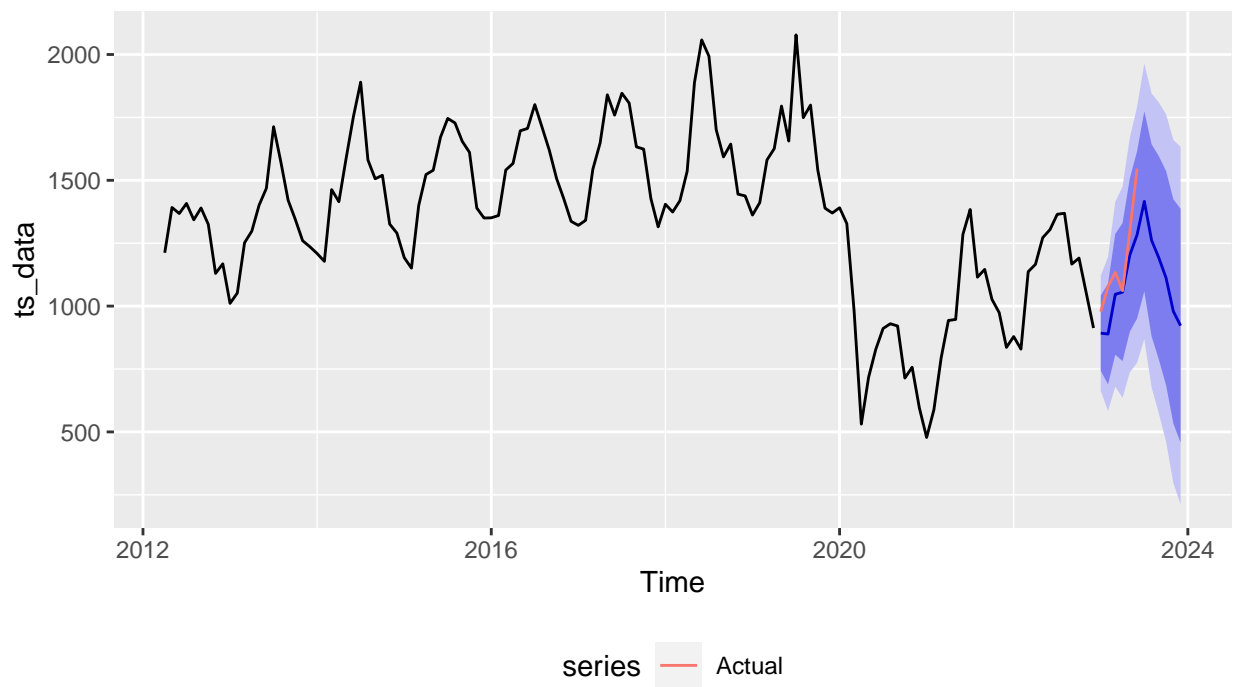
ETS Forecast vs Actual for Grouped_4_organisation

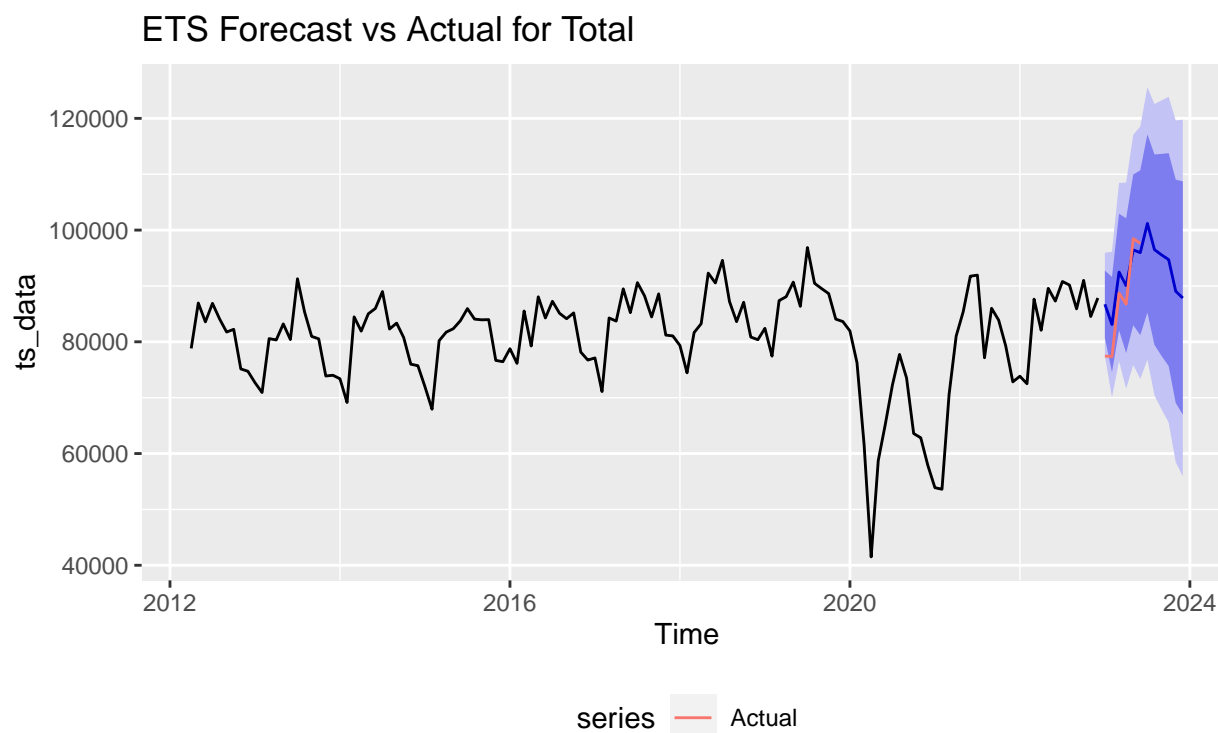


ETS Forecast vs Actual for Hywel Dda



ETS Forecast vs Actual for Powys Teaching





Explanation for the ARIMA model:

In time series forecasting, the ARIMA (AutoRegressive Integrated Moving Average) model is frequently employed. It belongs to a class of models that can reproduce a variety of common temporal structures in time series data.

- Conformity:

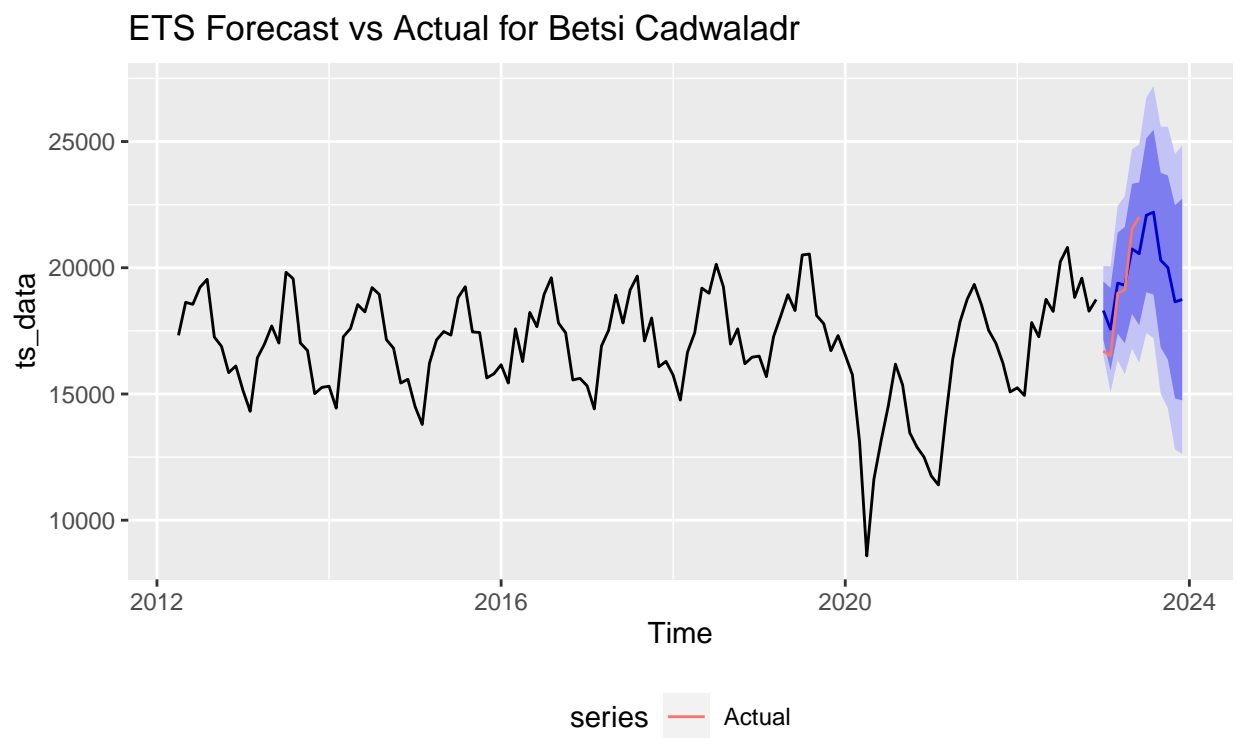
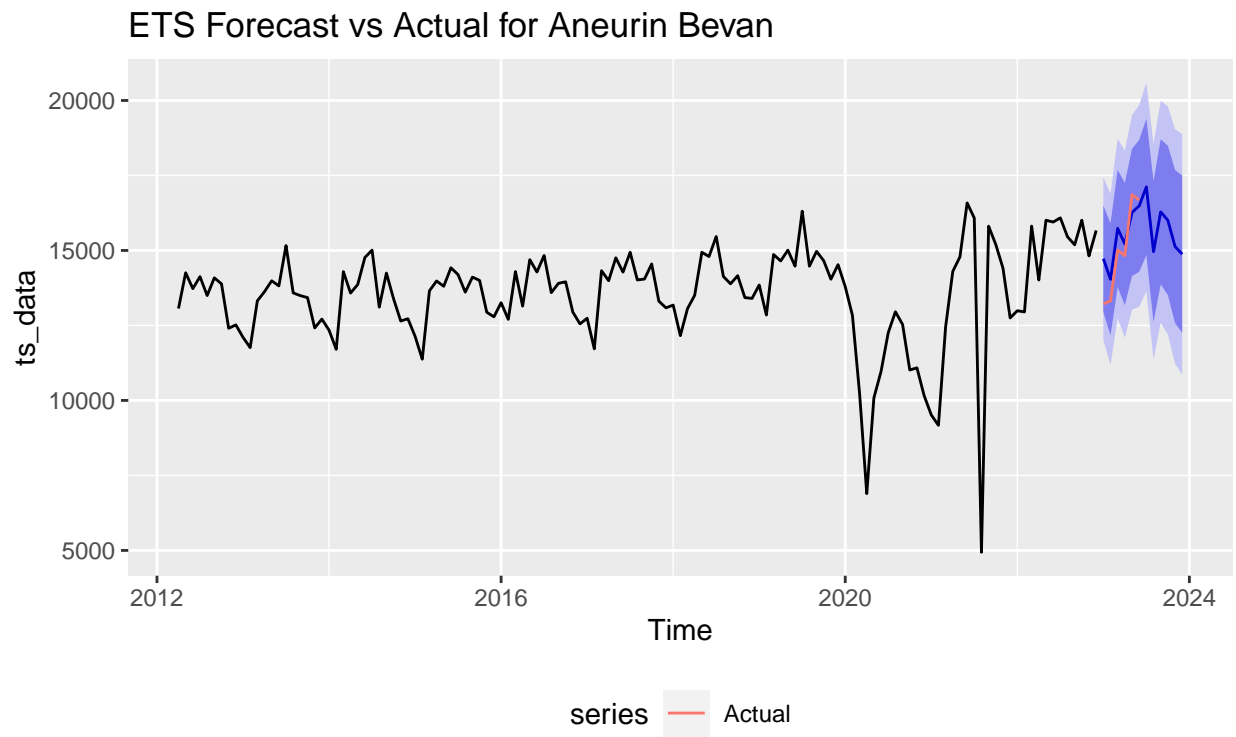
Hywel Dda, Powys Teaching, Cardiff & Vale, Grouped_4_Organisation, and Total (All-Wales): For these health boards, the ARIMA model effectively predicted the downturn that was observed in the latter portions of 2023 and successfully reflected the actual data, indicating a cogent direction. This demonstrates ARIMA's ability to effectively model and anticipate time-series data for these specific entities, despite the unpredictably fluctuating conditions brought on by the epidemic.

- Contradictions:

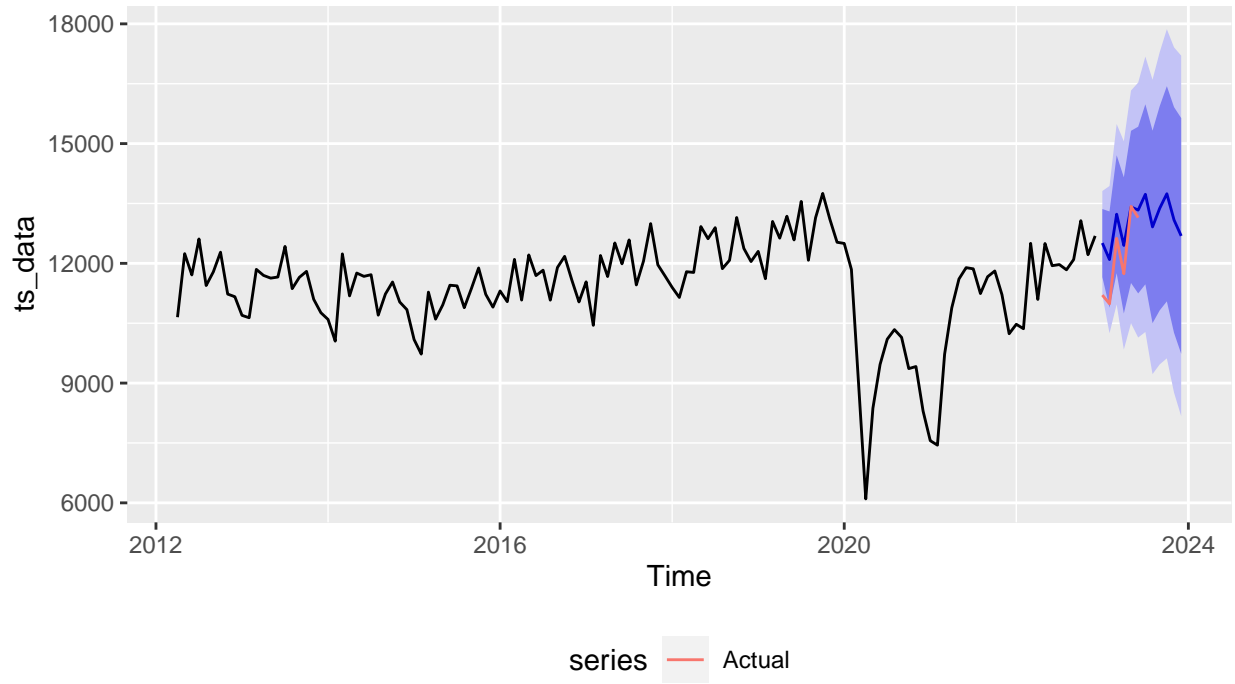
Aneurin Bevan and Betsi Cadwaladr: By contrast, the prediction statistics for Aneurin Bevan and Betsi Cadwaladr show a clear departure from the actual numbers. It is necessary to assess the ARIMA model's applicability and accuracy with regard to these specific datasets because there is a discrepancy between the predicted value and the real-time data.

5 Forecasting with ETS

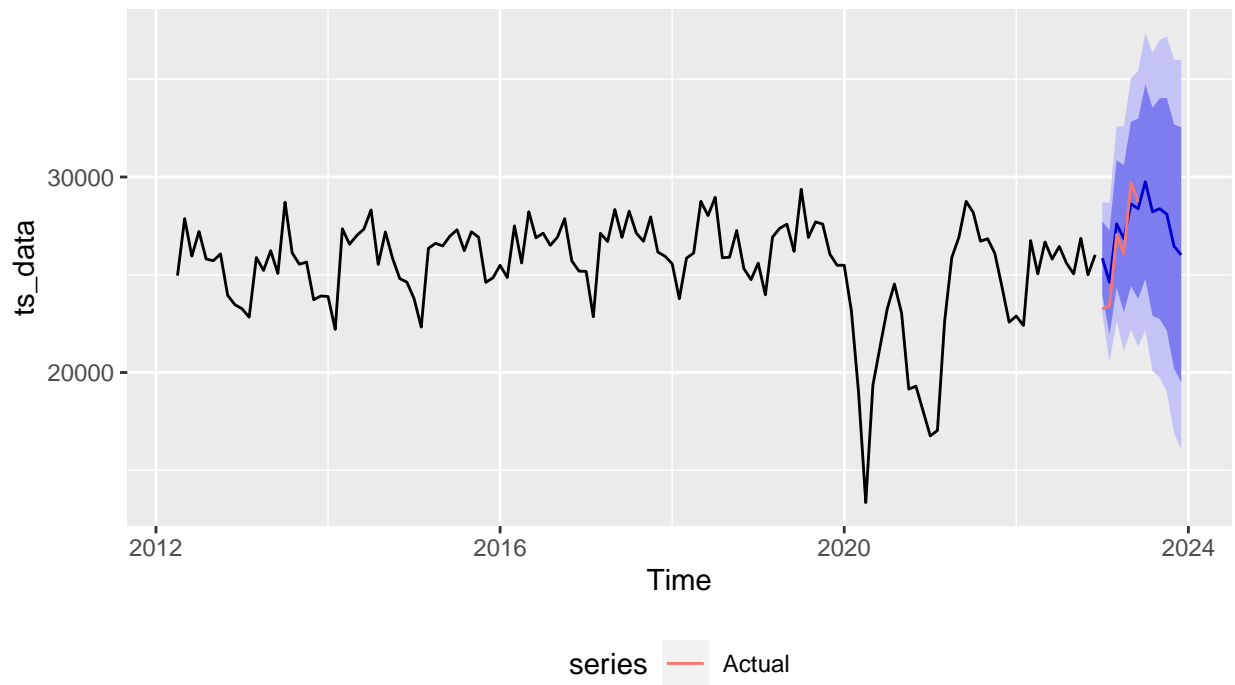
5.1 Plotting with ets



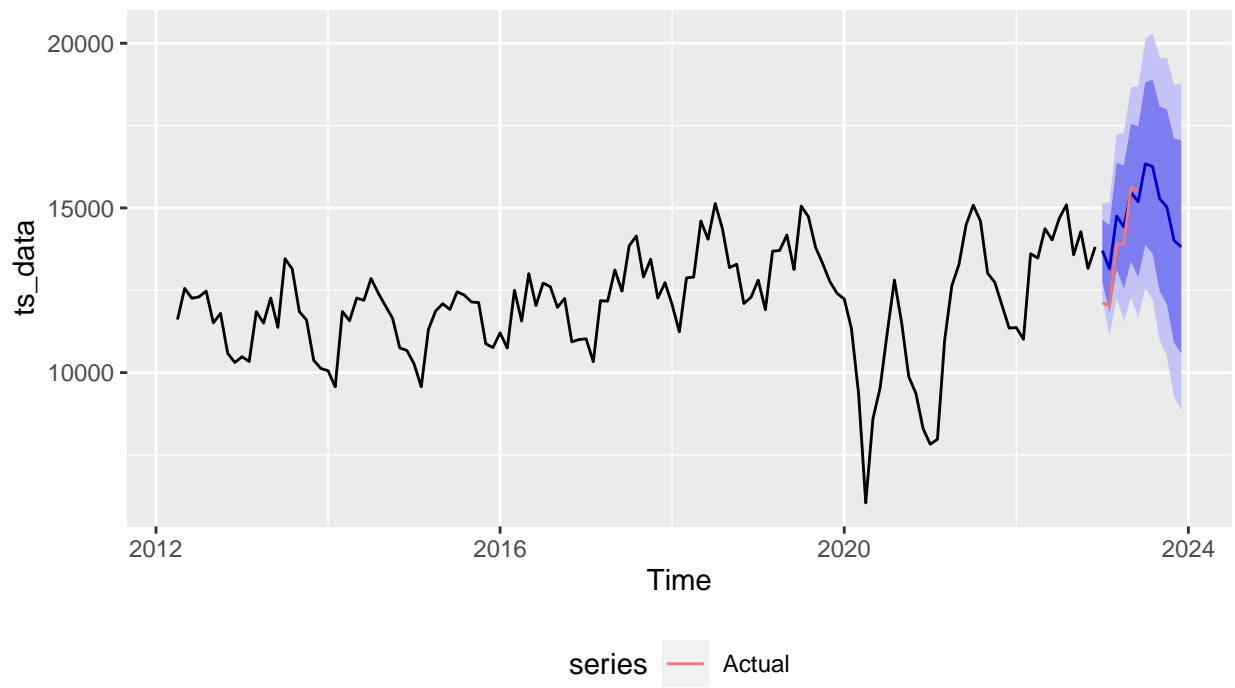
ETS Forecast vs Actual for Cardiff & Vale



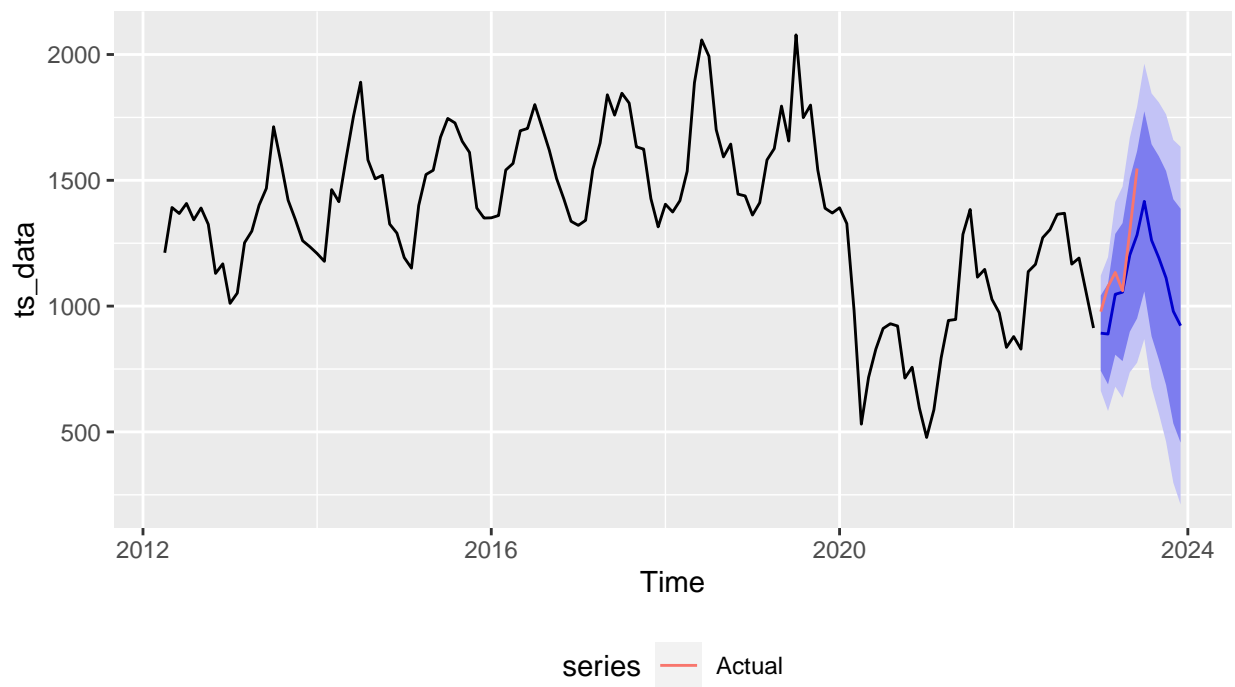
ETS Forecast vs Actual for Grouped_4_organisation

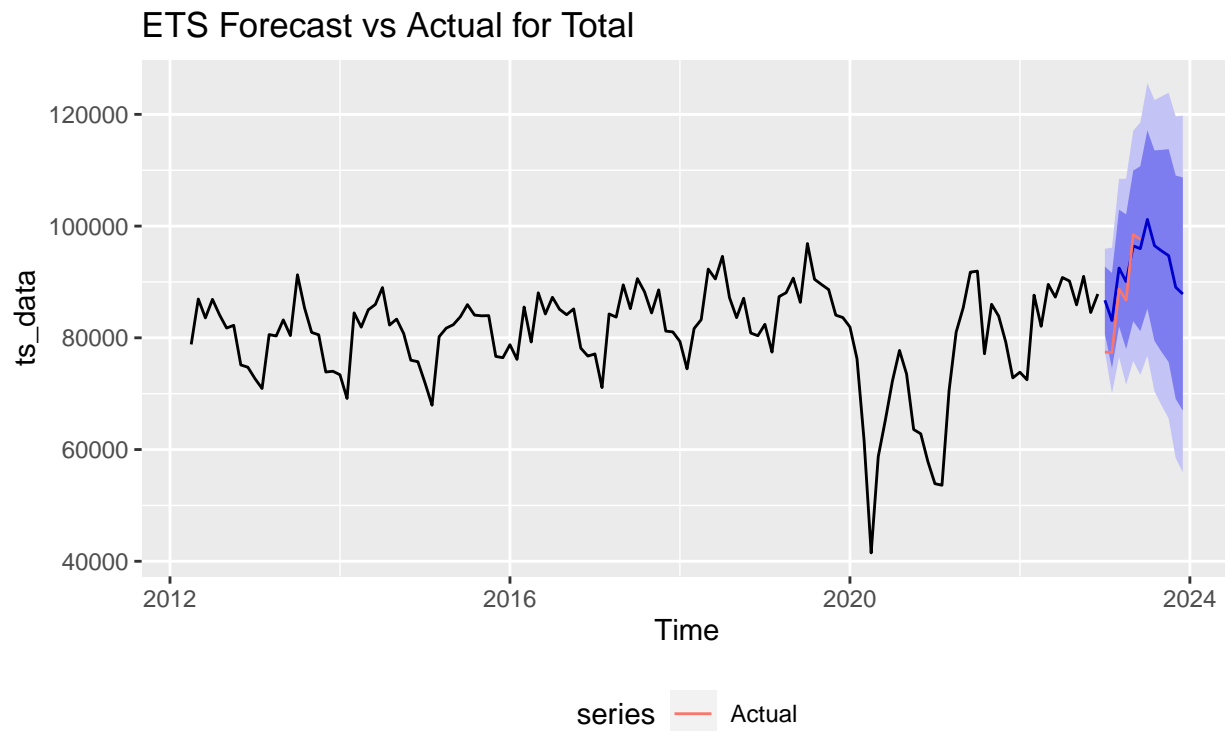


ETS Forecast vs Actual for Hywel Dda



ETS Forecast vs Actual for Powys Teaching





Upon first glance, it appears that the ETS model offers a relatively close alignment with the real data, demonstrating a strong predictive power. When compared to the ARIMA model, the projections closely mimic the actual data patterns rather than merely following them, which suggests a possibility for greater accuracy.

-Consistent Decline in Predictions:

Downtrend in 2023: Regardless of the model used, all forecasts uniformly point to a decline in trends during the second half of 2023. This might be explained by the historical data's apparent cyclical nature. In other words, the historical data reveals a pattern in which peaks frequently appear in the middle of the year and then descend into troughs at year's end and the beginning of the next year. This apparent phenomena may have a significant impact on the expected downturn in 2023.

In light of the above, it is crucial to approach model selection with a nuanced comprehension of the underlying data and impacting factors, even though the ETS model shows a potentially higher prediction accuracy within the given circumstances.

Therefore, I would like to conduct an accuracy analysis upon both arima and ets models.

6 Accuracy assessment for ARIMA and ETS

6.1 Calculating Accuracy Metrics for ARIMA and ETS

6.1.1 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
##
##
## -----
```

- With Cardiff & Vale being a significant exception, most places displayed reduced error metrics when modeled with ETS, signifying a higher predictive accuracy as compared to the ARIMA model.
- Model Differentiation:

ARIMA, A Time-Series Tradition Autoregressive, integrative (difference), and moving average features are included in the forecasting core components. The foundation is laid by differencing to make data stationary, and then autoregressive and moving average techniques are applied.

ETS: A Comprehensive Method for Time-Series Model Constituent Elements reflects the seasonality, trend, and error components in time series data. Consists of rigorous detection and modeling of error, trend, and seasonal patterns within the data, which tends to include entire pattern analysis.

- According to the interest expressed by the NHS executives, I would like to provide some insights for the error metrics according to the quality of the forecast for the non-zero count data:

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.

However, RMSE and MAE should not be used for the hierarchical time series data. And here is why:

- **1. Consistency Between Levels:** Granularity Errors are not consistently distributed or similar across all hierarchical levels due to the variance in data granularity between levels, and there will be an aggregation dilemma. Due to the variation in magnitude and volume at various hierarchical levels, direct aggregation of RMSE or MAE from bottom levels to top levels might produce inconsistent or biased judgments of forecast accuracy.

- **2. Problems with reconciliation:** It is possible for forecasts to not add up coherently in a hierarchical structure when they are independently generated for several levels. Since they assess mistakes without taking into account hierarchical dependencies, RMSE and MAE do not by default address this reconciliation. And the alignment issue will therefore be caused. The use of standard error measurements at various hierarchical levels may produce false conclusions regarding the precision and dependability of the forecasts in the absence of a cogent reconciliation process.
- **3. Metric incomparability:** Impact of a unit error is not constant across levels; for instance, a 100-unit error may be insignificant at a higher aggregate level but significant at a lower one. Additionally, due to the divergent scales of the errors at various levels, it is difficult to use a single statistic, such as RMSE or MAE, to fairly compare performance across all hierarchical levels.

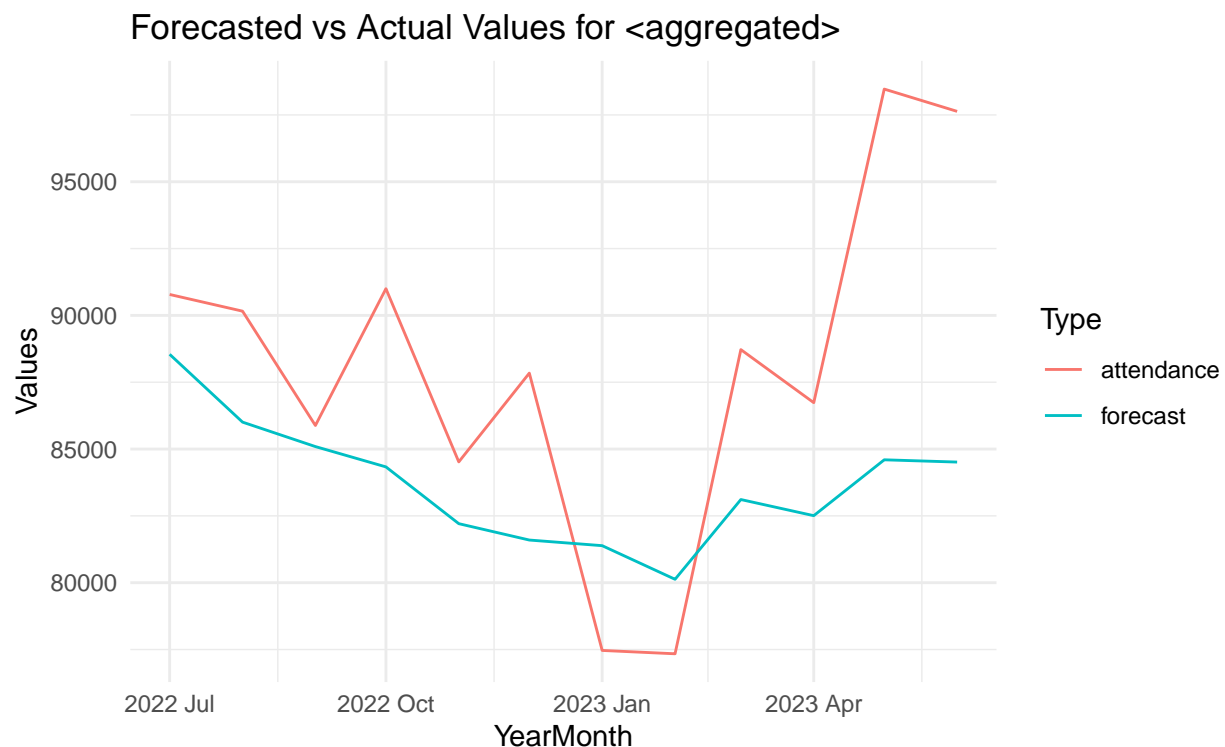
7 Reconciliation

7.1 Aggregate data

7.2 Time series cross validation

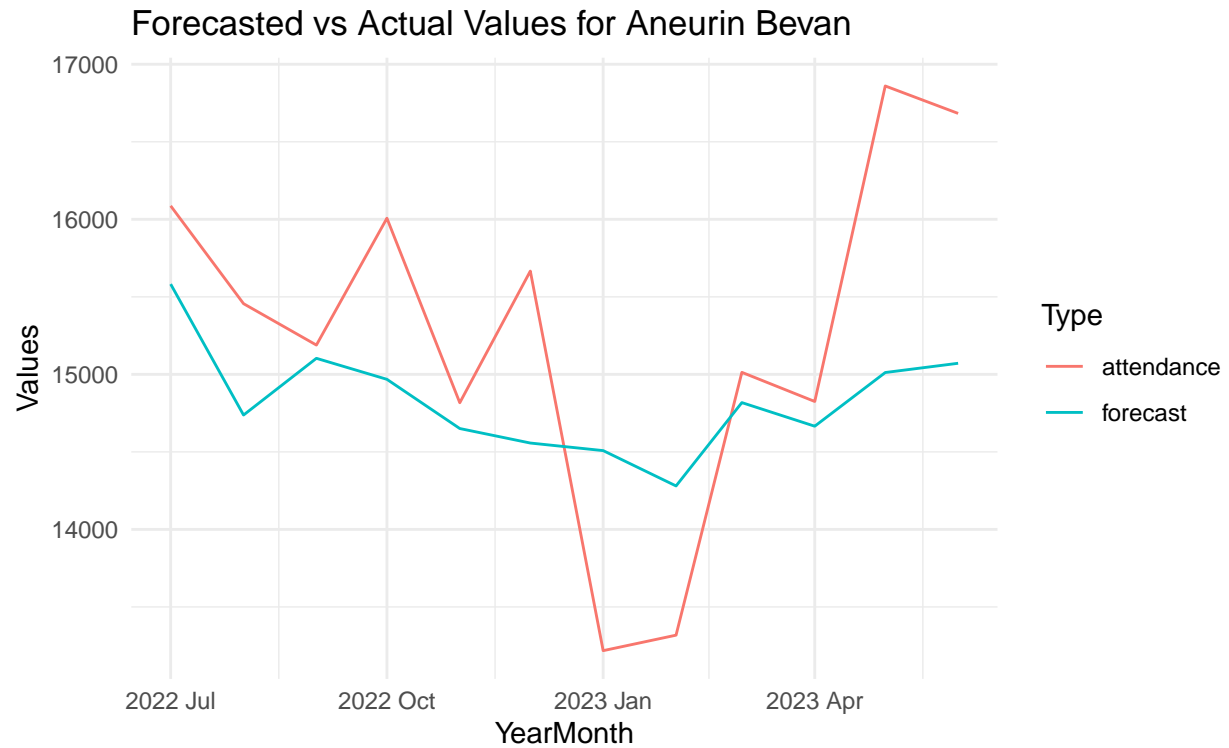
```
## [1] 0
```

```
## [[1]]
```

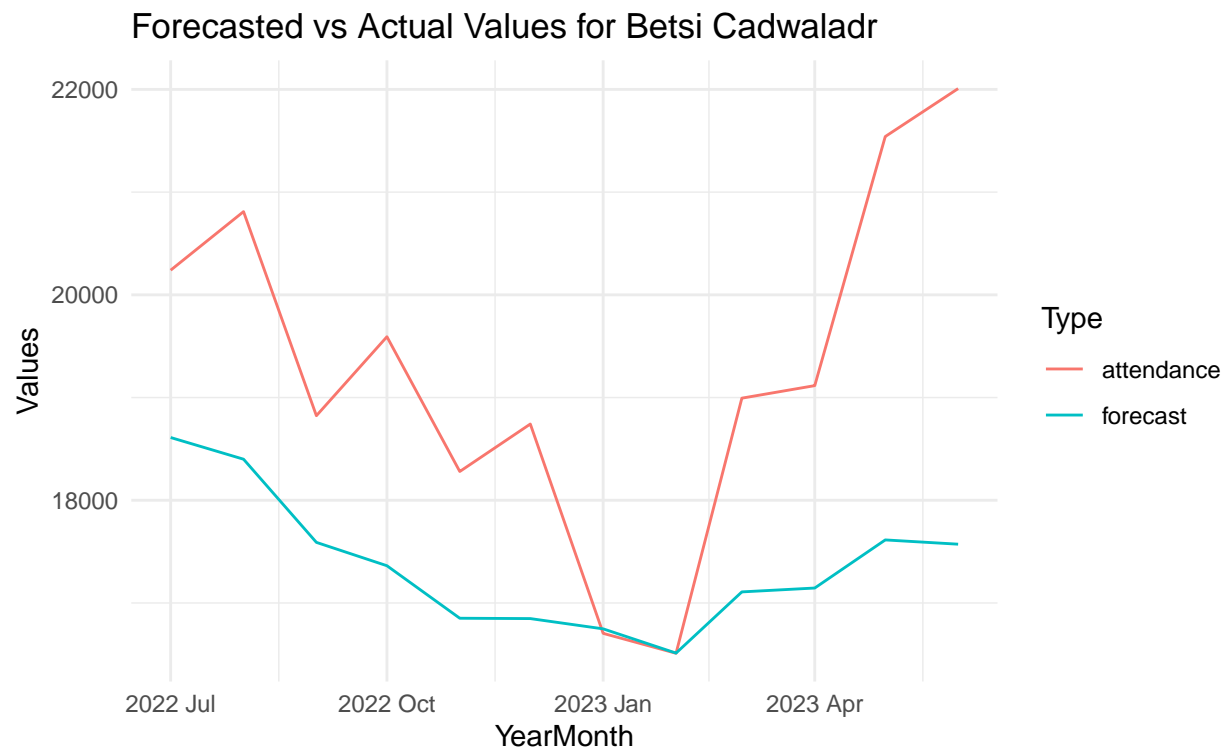


```
##
```

```
## [[2]]
```

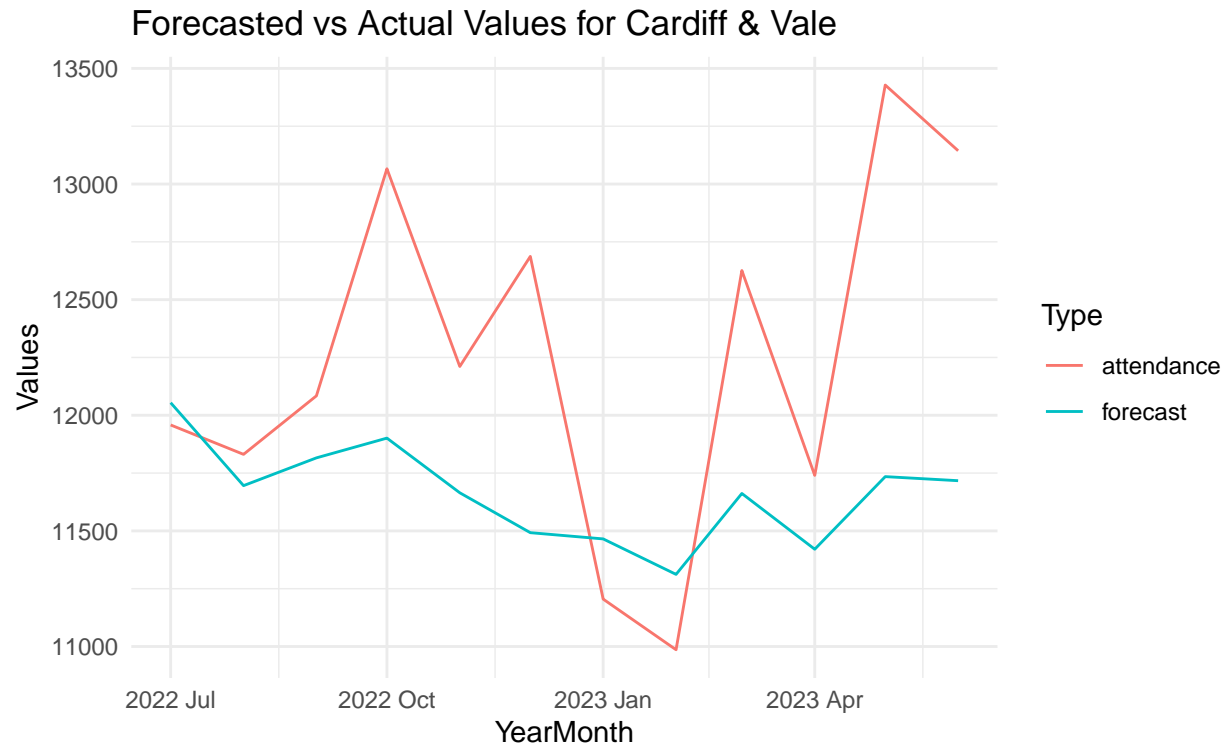



```
##  
## [[3]]
```

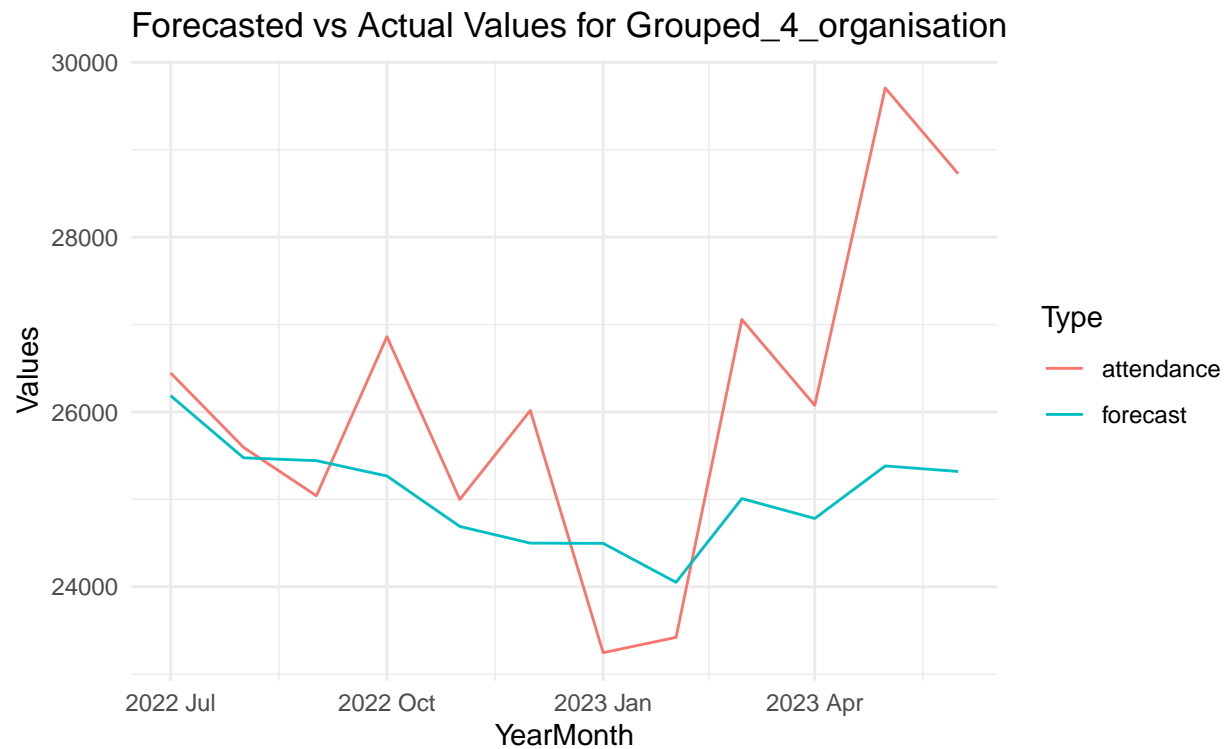


```
##
```

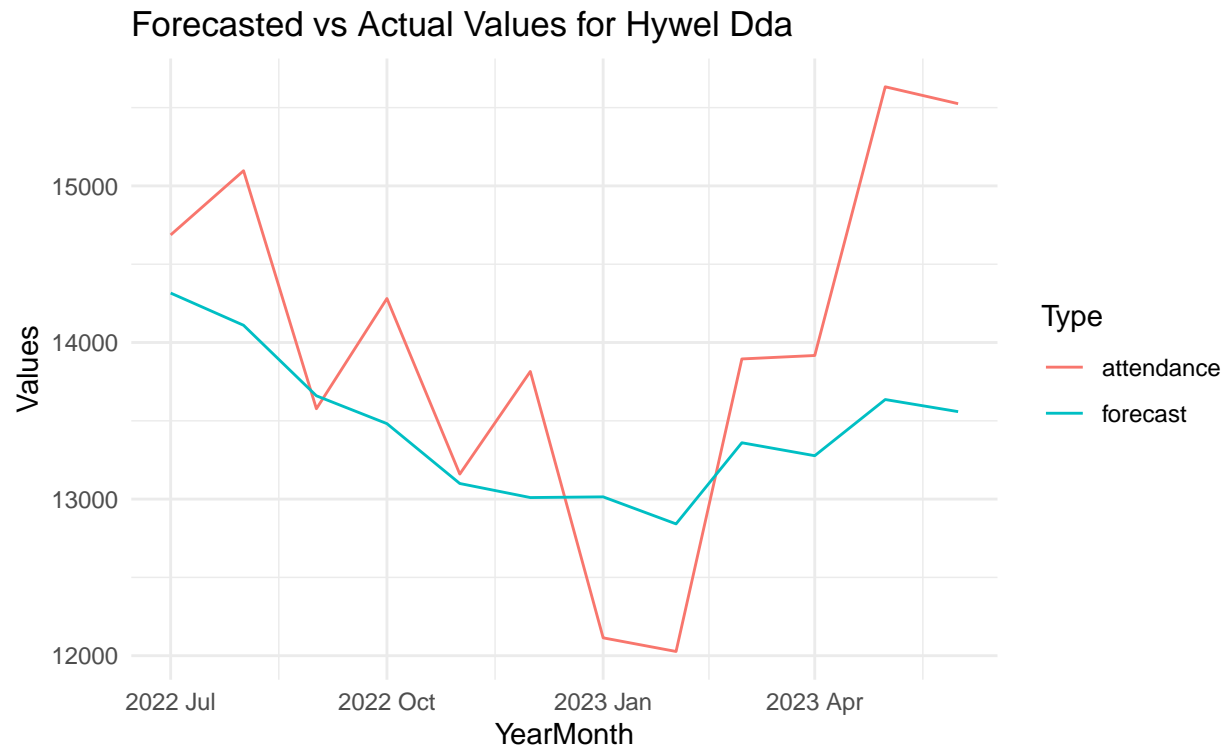
```
## [[4]]
```



```
##  
## [[5]]
```

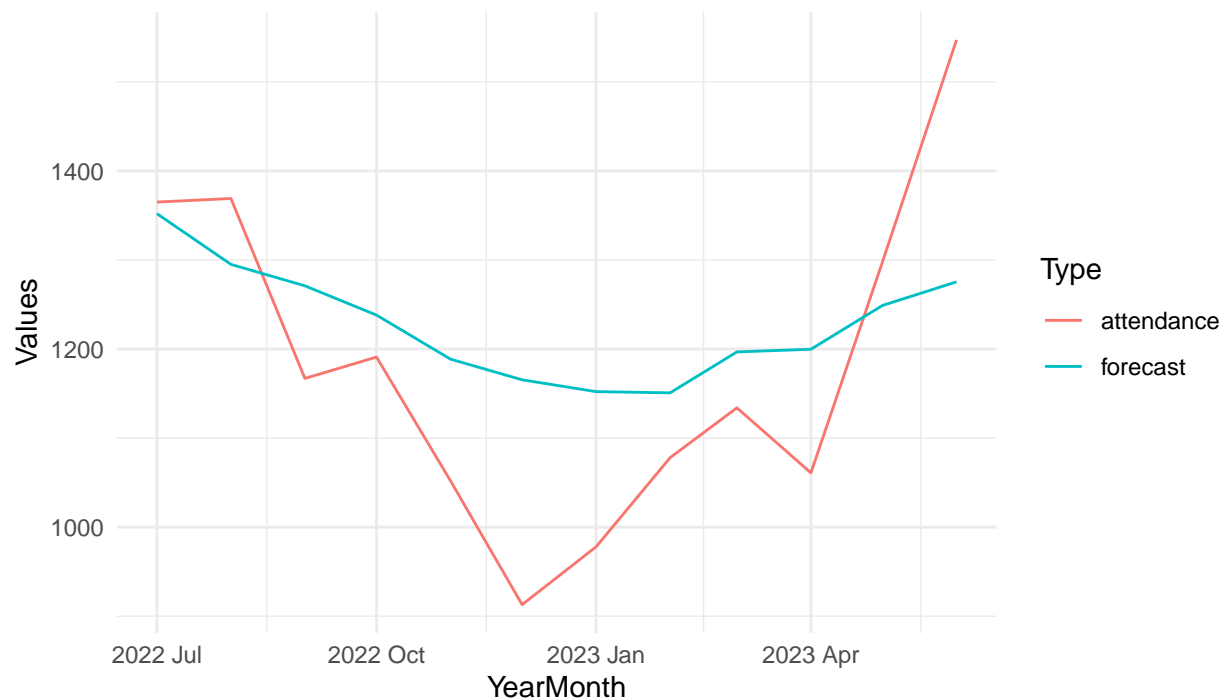


```
##  
## [[6]]
```



```
##  
## [[7]]
```

Forecasted vs Actual Values for Powys Teaching



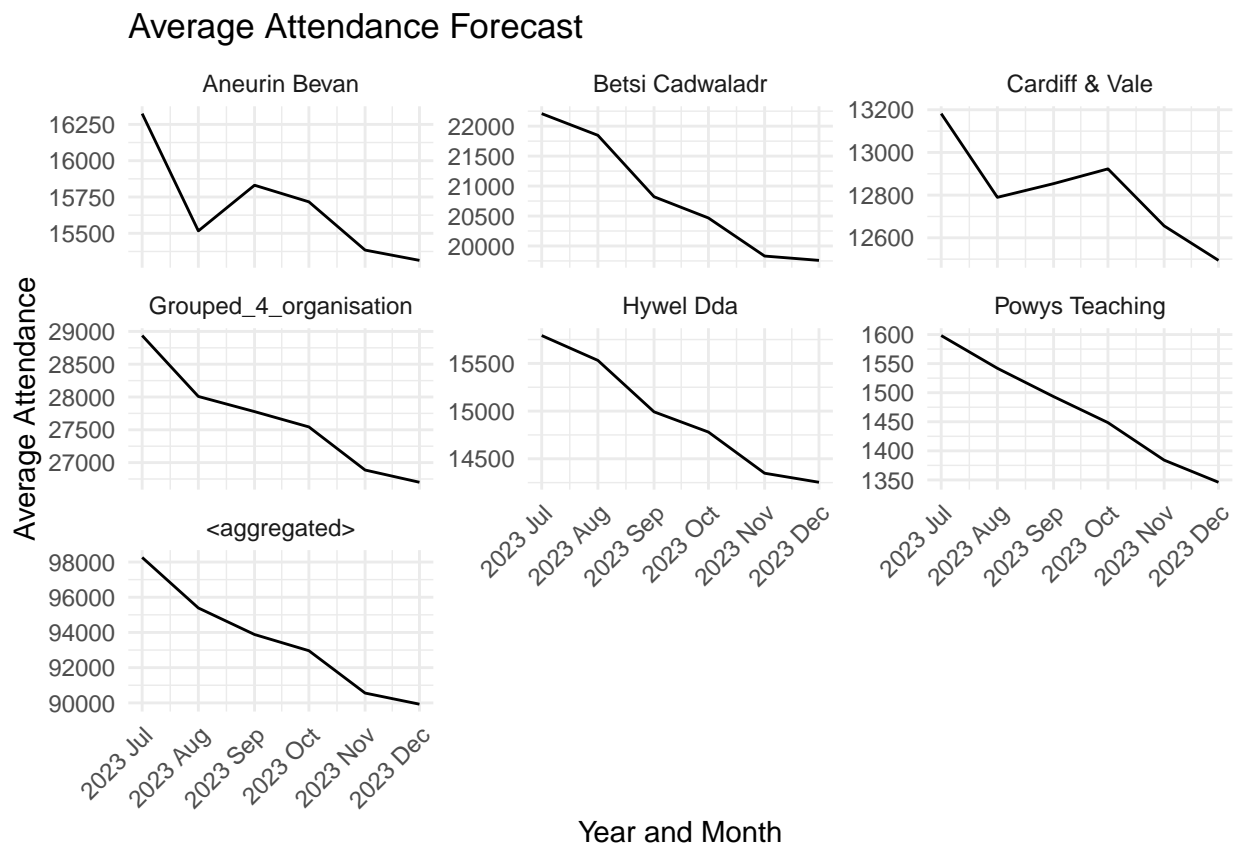
```
## # A tibble: 1 x 2
##   t_value p_value
##   <dbl>   <dbl>
## 1    0.313    0.755
```

The p value is relatively high. It suggests that we fail to reject the null hypothesis, meaning there is not enough evidence to conclude that forecast and actual values are significantly different. In other words, the t-value of 0.313 is not large enough to be considered statistically significant.

- Base forecasting: applied when forecasting simple time series.
- Reconciliation forecasting (Hierarchical forecasting): it includes forecasting on different levels of aggregation, and it should be applied when dealing with hierarchical or grouped time series.

7.3 Reconcile

7.3.1 1. Fitting models and Forecasting



7.4 Include Age Group

