hierarchical_forecasting_report

Janice Hsu

2023-09-30

Contents

1 Preliminary Analysis **3**

1.1 Data Introduction . . . 3

2 Exploratory Data Analysis **4**

2.1 Number of patients entering ED under different hospital hierarchy . . . 4

2.2 Group the changed Local Health Board together . . . 5

2.3 Number of patients who enter ED under 6 different local health boards . . . 5

2.4 Seasonality of number of attendances . . . 7

2.5 Change the Age_Code structure into different groups for simplicity and interpretability . . . 9

2.6 Plot Number of Patients in different age groups . . . 9

3 Forecast with ARIMA **10**

3.1 Plotting using ARIMA . . . 10

4 Forecasting with ETS **14**

4.1 Plotting with ets . . . 14

5 Forecasting with ETS **18**

5.1 Plotting with ets . . . 18

6 Accuracy assessment for ARIMA and ETS **22**

6.1 Calculating Accuracy Metrics for ARIMA and ETS . . . 22

7 Reconciliation **23**

7.1 Aggregate data . . . 23

7.2 Time series cross validation . . . 23

7.3 stretch_tsibble(.init = 135-36, .step = 1) . . . 23

7.4 Remove tscount . . . 24

8 Running t-tests and extracting p-value and t-value **24**

8.1 Reconcile 25

8.2 Include Age Group 26

- Change the accuracy test (the one with 0)
- probably remove the seasonal effect and see if there's a pattern or sth

1 Preliminary Analysis

1.1 Data Introduction

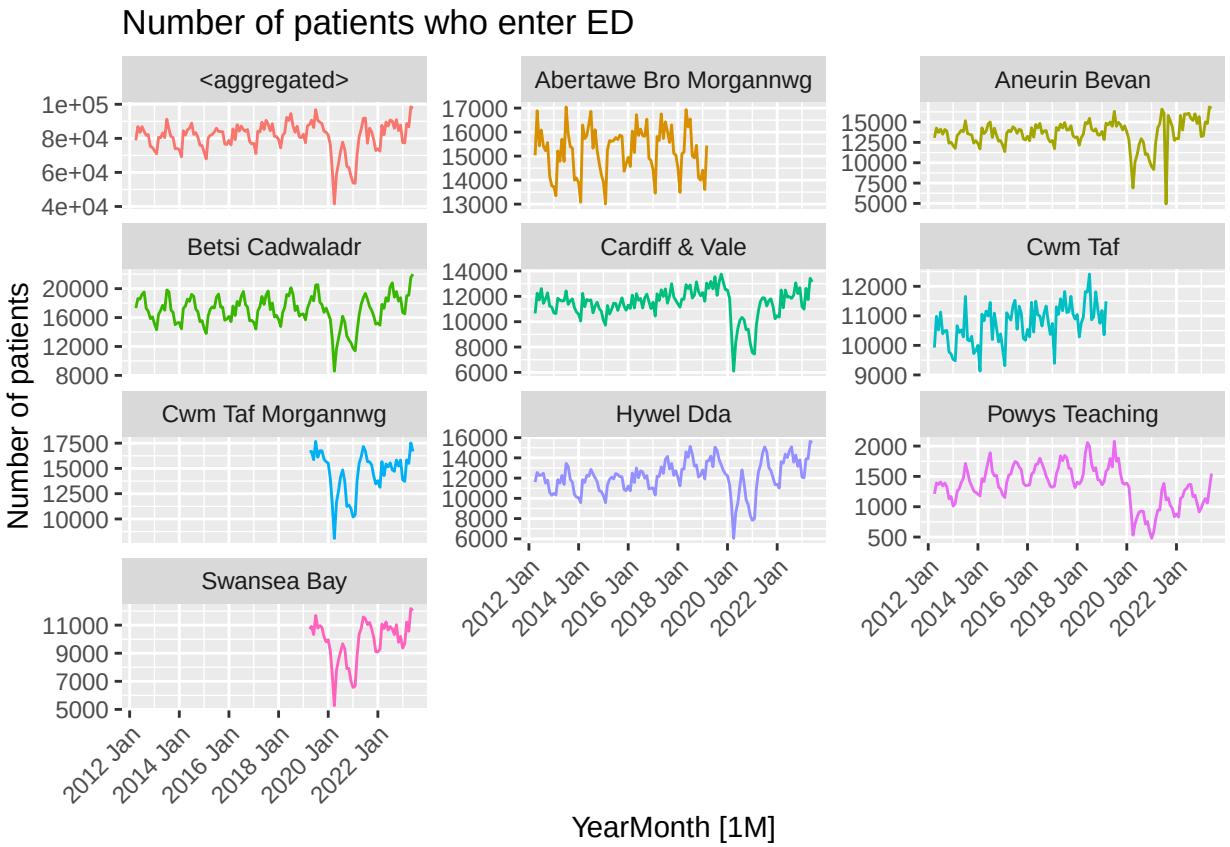
The dataset contains variables related to monthly Emergency Attendances in hospitals in Wales, UK. The data was publicly available and retrieved from StatsWales.

- Data: This column represents the aggregated number of attendances in each emergency department. The data was aggregated according to the other columns.
- YearMonth: This column represents data from 2012 April to 2023 June (135 time points).
- Age_Code: This column provides the patient's age group. There are 17 different age groups: "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 patient's gender using 3 categories: "Female", "Male" and "Unknown".
- 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 Local Health Board (LHB), which are responsible for planning and delivering NHS Wales services in their areas, including Emergency Care. In Wales there are 7 distinct LHBs, however two of them, Cwm Taf Morgannwg and Swansea Bay, were previously (before 1 st of April 2019) known as Cwm Taf and Abertawe Bro Morgannwg (therefore this column contains 9 distinct values).
- Organisation_Code: A code for the organisation as well as the LHB.

There are three hierarchies in this dataset: - On the top level, there is all the hospitals in Wales. - On the second level, there are the different LHB. - At the bottom level, there are 42 hospitals.

2 Exploratory Data Analysis

2.1 Number of patients entering ED under different hospital hierarchy



As already mentioned, two LHBs were redefined from the 1st of April 2019 onwards: Cwm Taf ((# hospitals = 27)–> Cwm Taf Morgannwg ((# hospitals = 30)// Abertawe Bro Morgannwg ((# hospitals = 26) –> Swansea Bay ((# hospitals = 31).

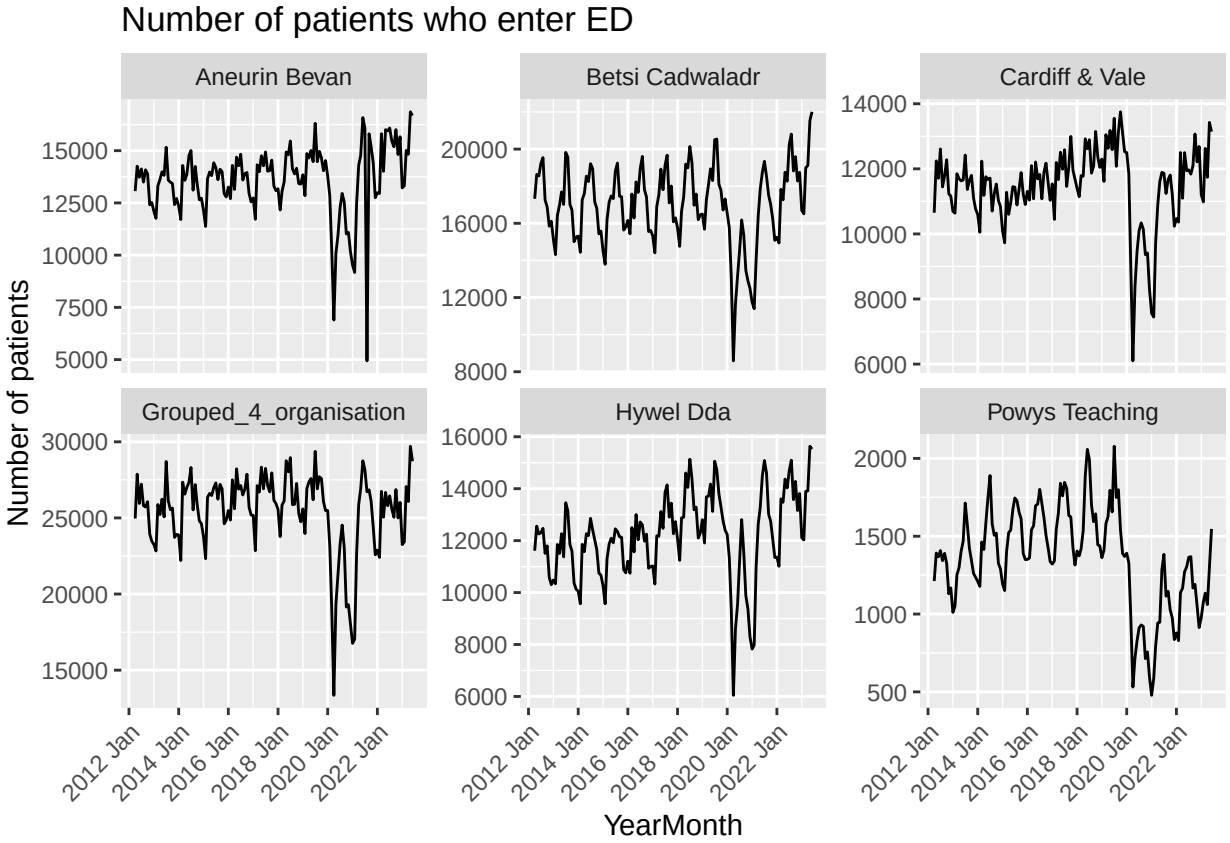
Since the number of hospitals included before and after the change in each LHB was different, these 4 organisations were aggregated into one unique group to enable their inclusion in the forecasting analysis (defined as "Grouped_4_organisation"in the code snippets).

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:

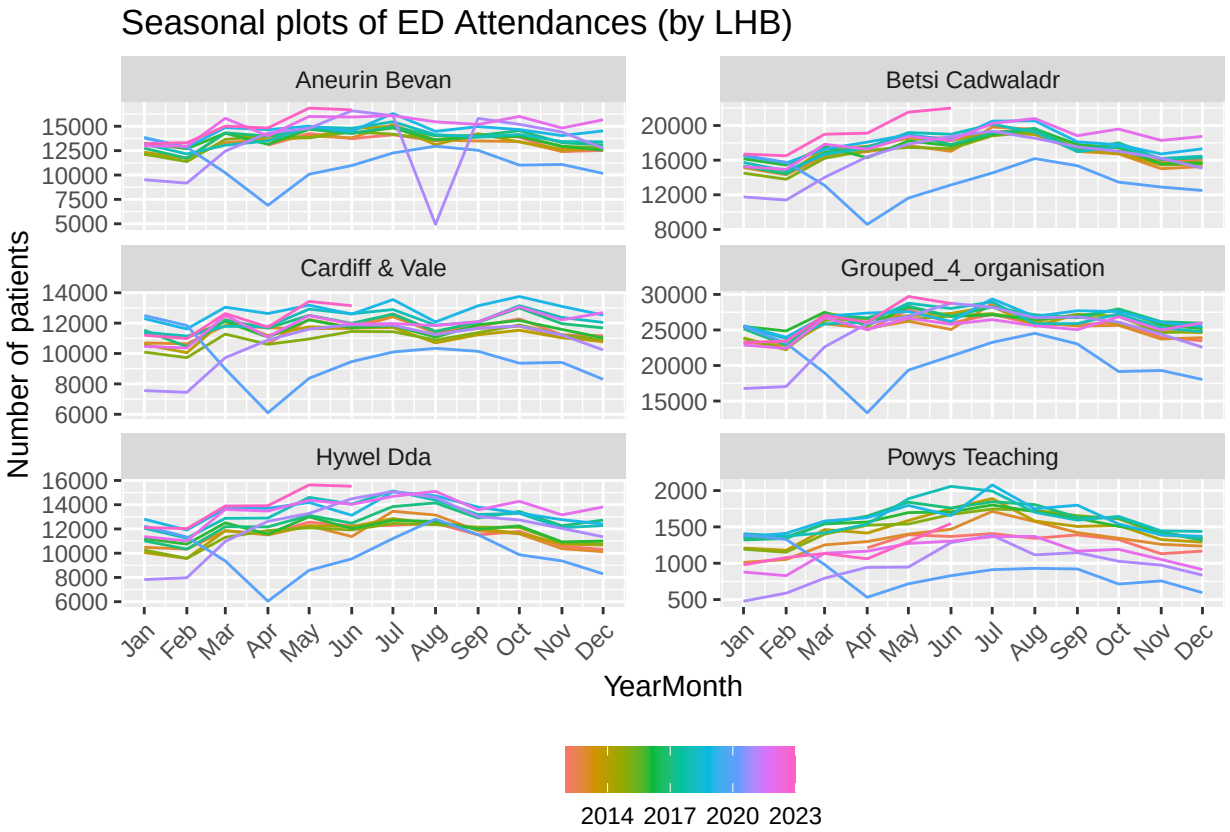
Identify Underlying Causes: Researching seasonality's probable triggers, such as public health initiatives, flu seasons, or holiday seasons, in order to comprehend the factors influencing these patterns.

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 further explore whether there was any trend and/or seasonality in the data by LHB, the time series was decomposed using the STL method and the extracted components (trend, seasonal and residuals) plotted:



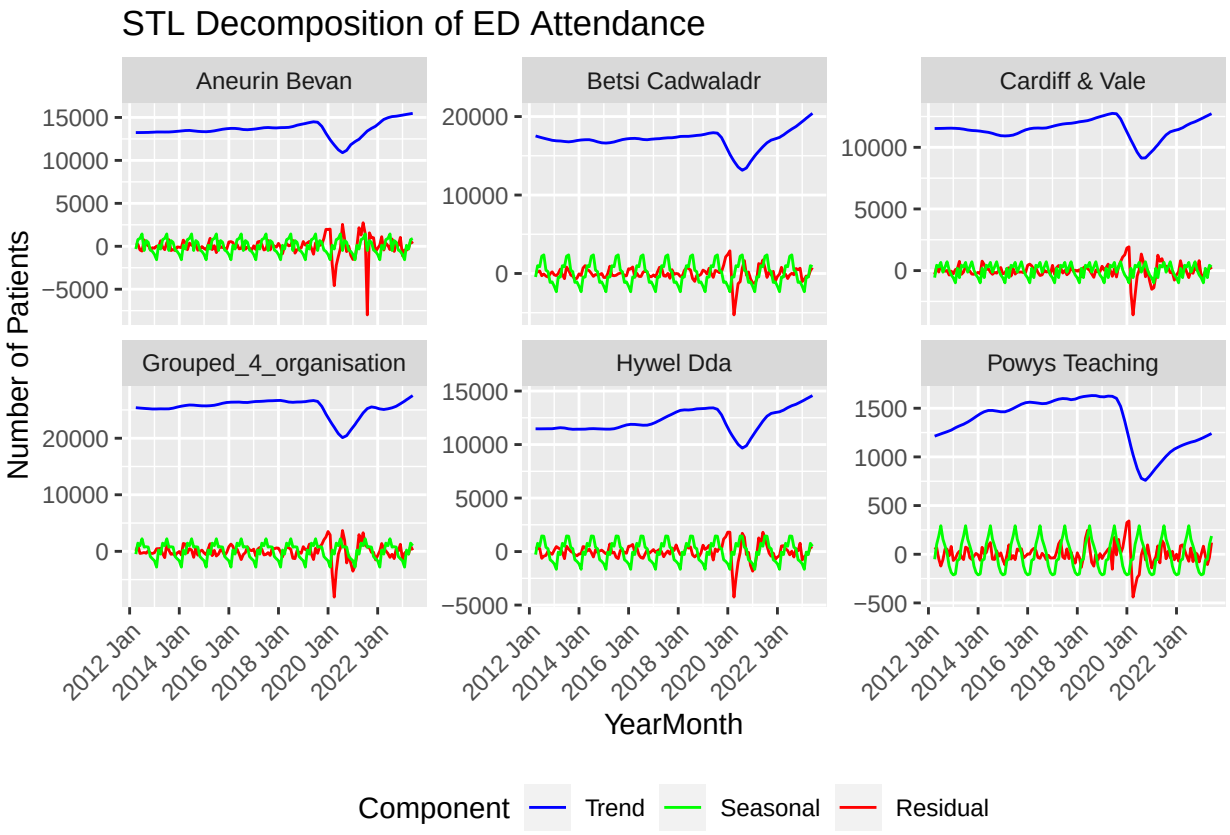
- Findings:

1. Most LHBs have a trend whereby attendance is lower in some months, like April and November, and higher in others, like June and July.
2. In many of the LHBs, the attendance in 2020 differs noticeably from other years. In comparison to similar months in past years, the majority of LHBs appear to have a fall or reduction in ED attendances during several months of 2020. This could be a sign of outside influences influencing access to or use of healthcare because of the pandemic.
3.

- Aneurin Bevan: The middle of the year for 2020 shows a clear decline.
- Cardiff & Vale: Attendances in 2020 appear to be continuously lower than in previous years, particularly in the early months.
- Hywel Dda: Around mid-year, there is a noticeable decline in attendance.
- Betsi Cadwaladr: While the trend for 2020 appears to be largely constant, several months still show a decline in attendance.
- Grouped_4_organisation: Attendance clearly drops off during particular months.
- Powys Teaching: The variations are more pronounced, with obvious dips during particular months.

2.4.1 Decompose Time Series

2.4.2 Plotting



Findings:

1. Trend: There appears to be an upward trend in the number of ED visits over time for all LHBs, but there are outliers. The most obvious decline occurs around 2020, which may be a sign of the COVID-19 pandemic's effects.

2. Seasonality: The seasonal component illustrates the periodicity in ED attendances by showing predictable patterns that repeat annually. 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.

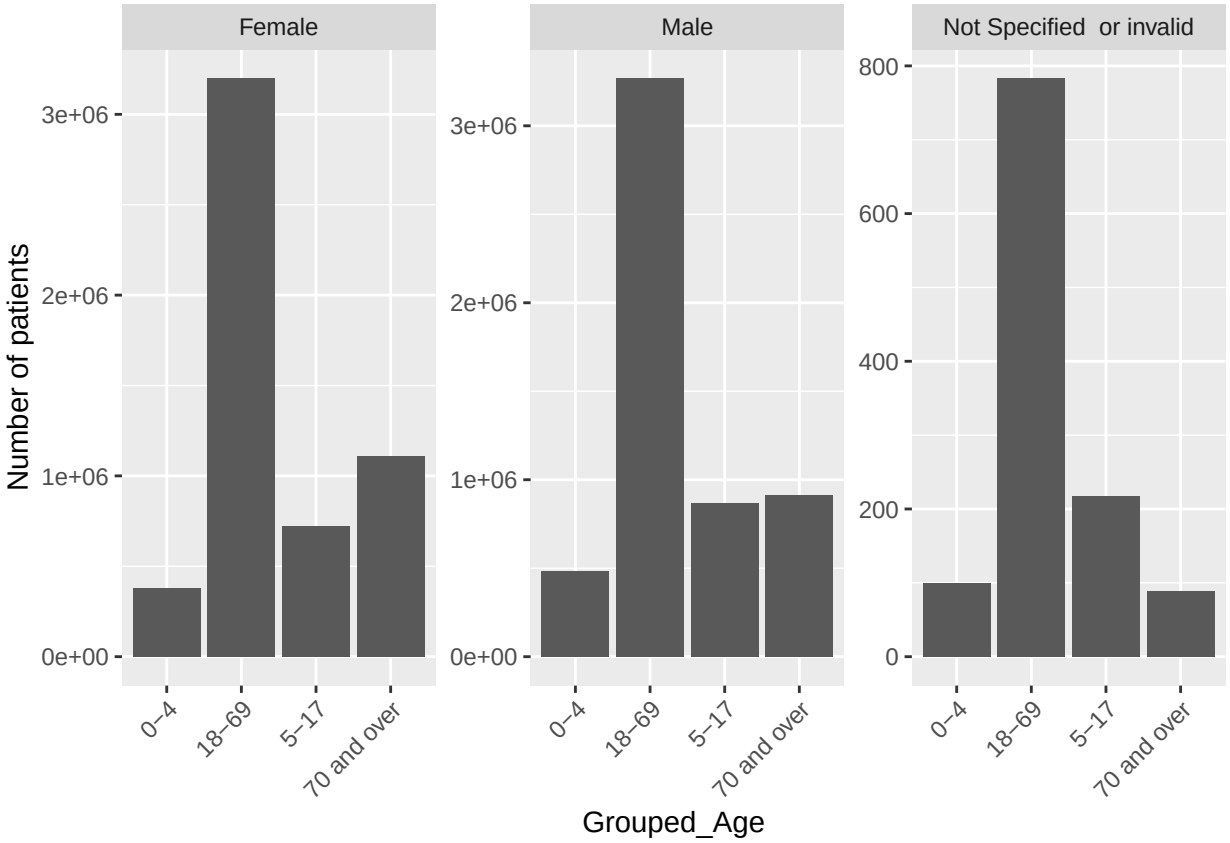
3. Residuals: The residuals are the data's unexplained variation after the trend and seasonal components have been removed. The majority of LHBs have residuals that oscillate near zero, which shows that the STL decomposition has effectively caught the main trends in the data.

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

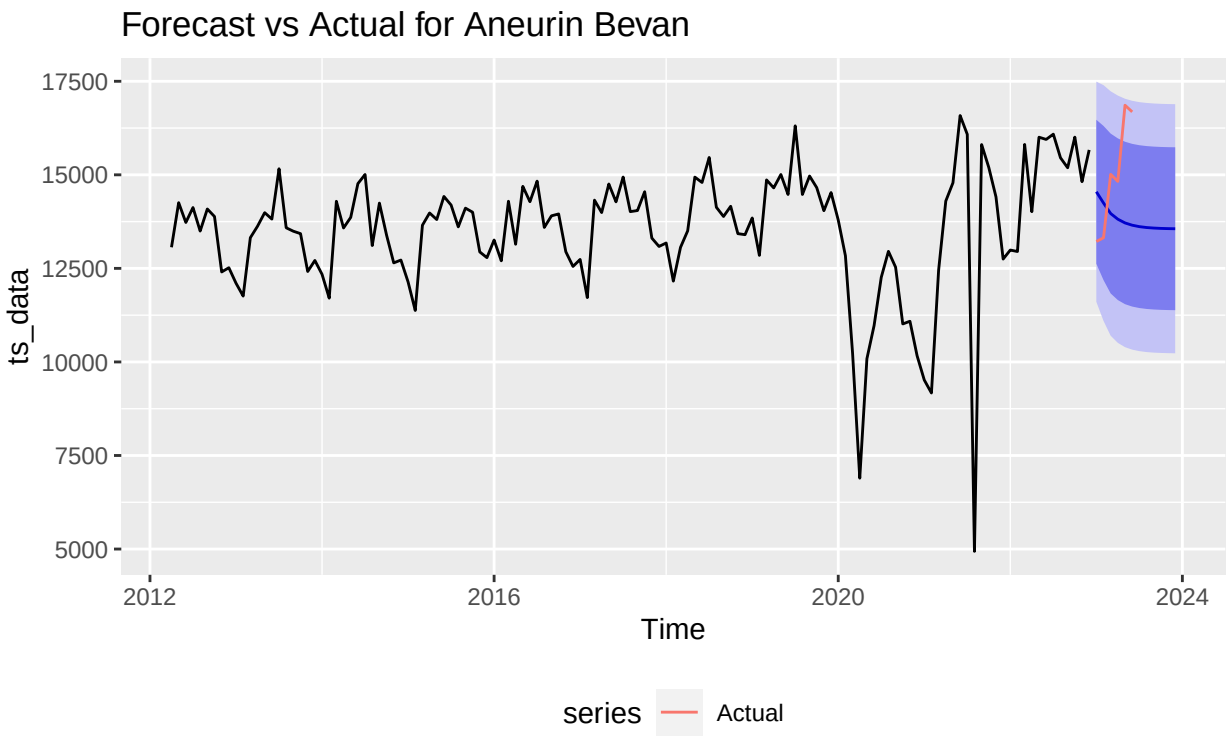


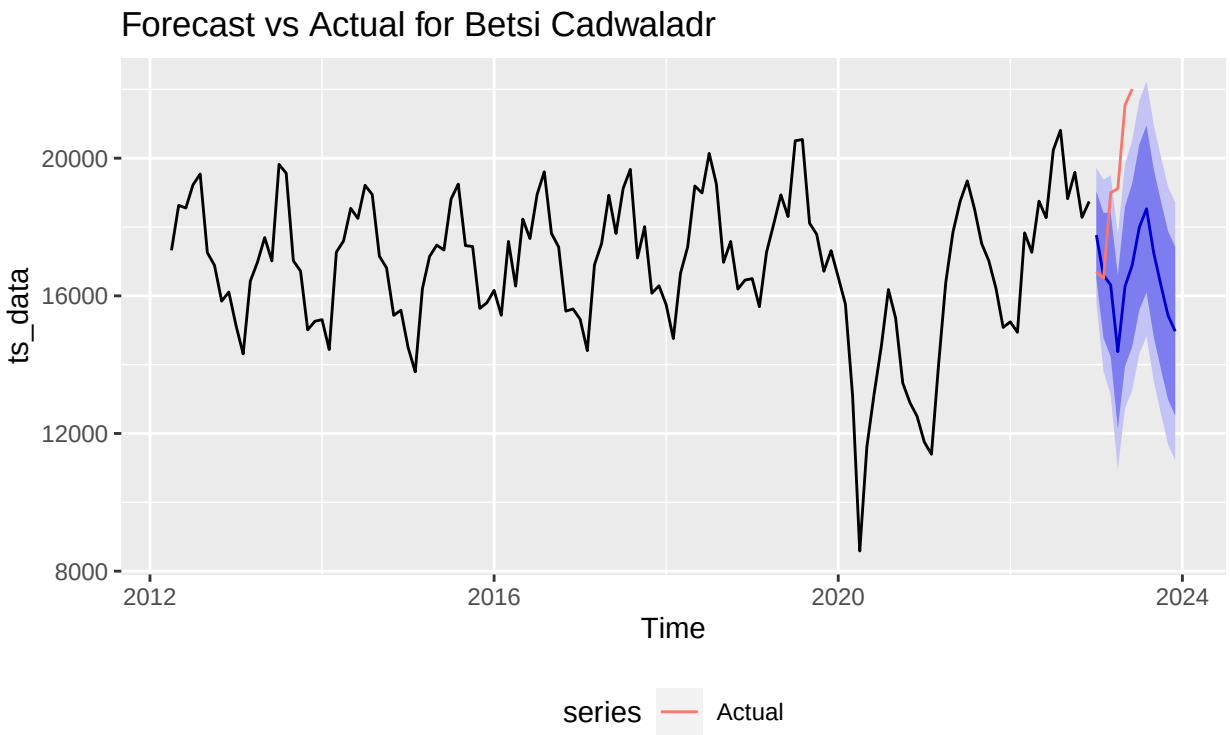
Findings:

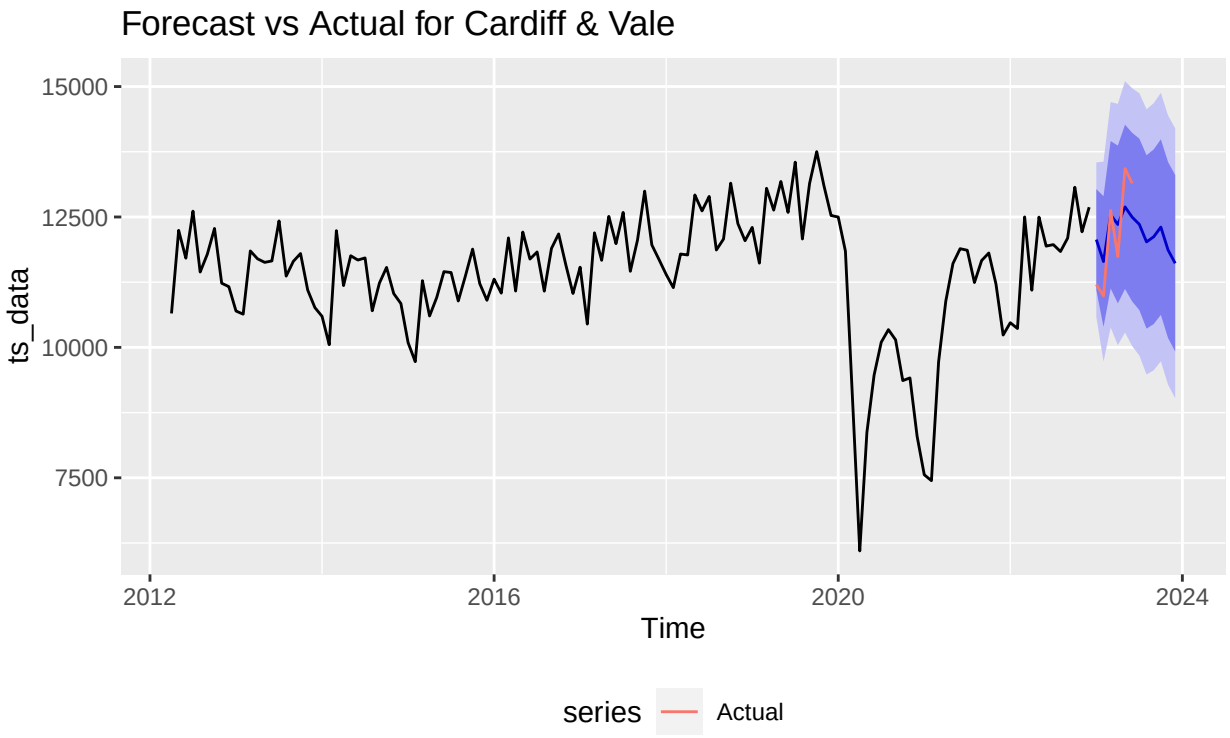
- The observation that the age group 18-69 has the most amount of patient attendance is expected, as it is the biggest group among all. However, it is noteworthy that the second biggest group are from the oldest age bracket, aligning with the general understanding of the health care need for the elders.

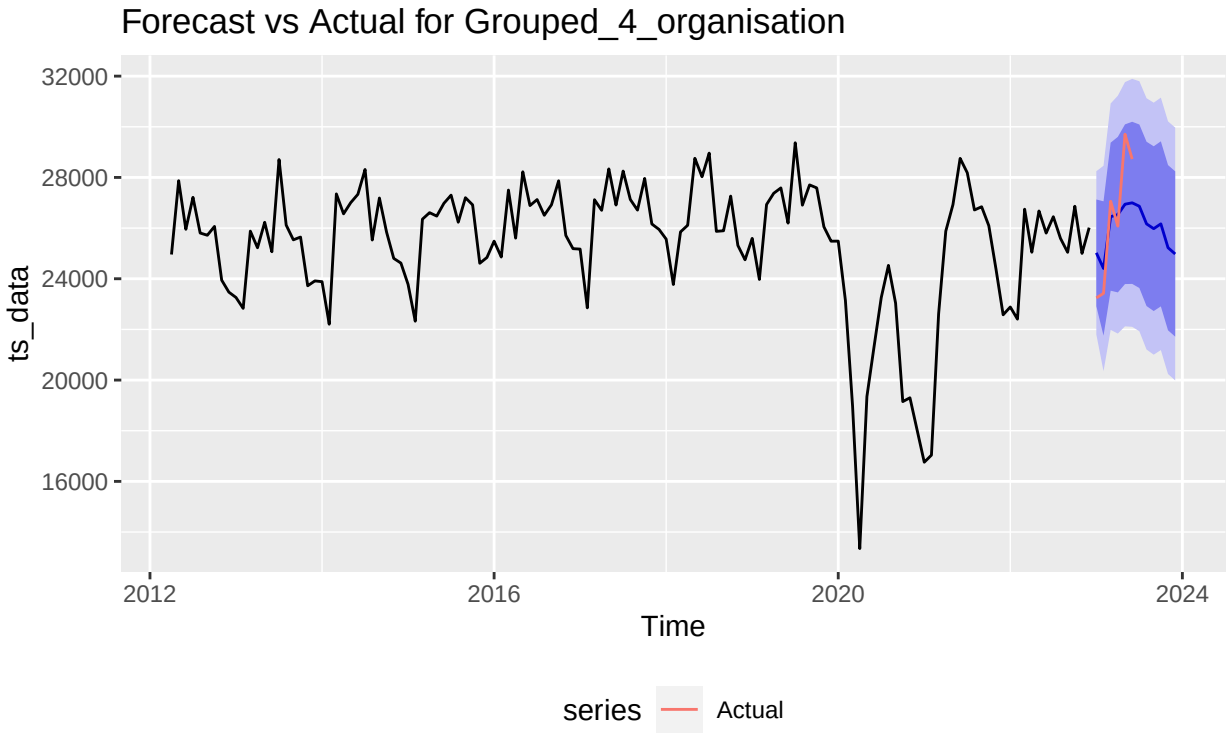
3 Forecast with ARIMA

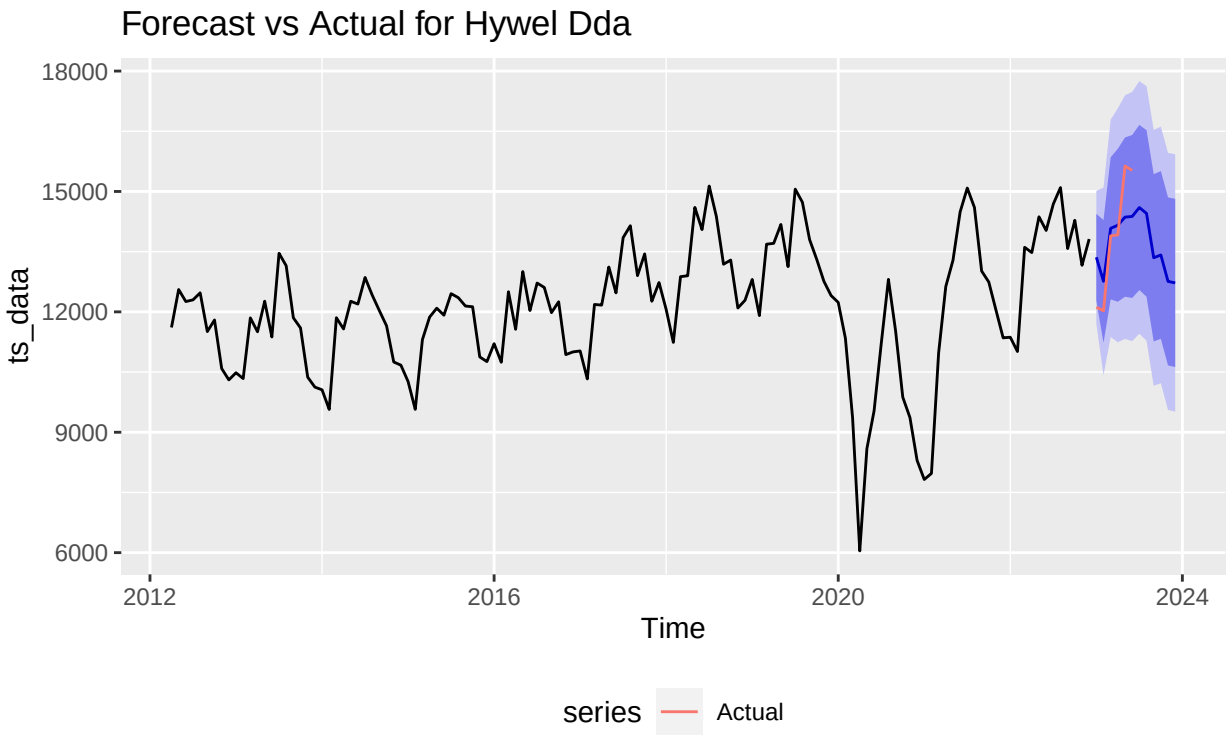
3.1 Plotting using ARIMA

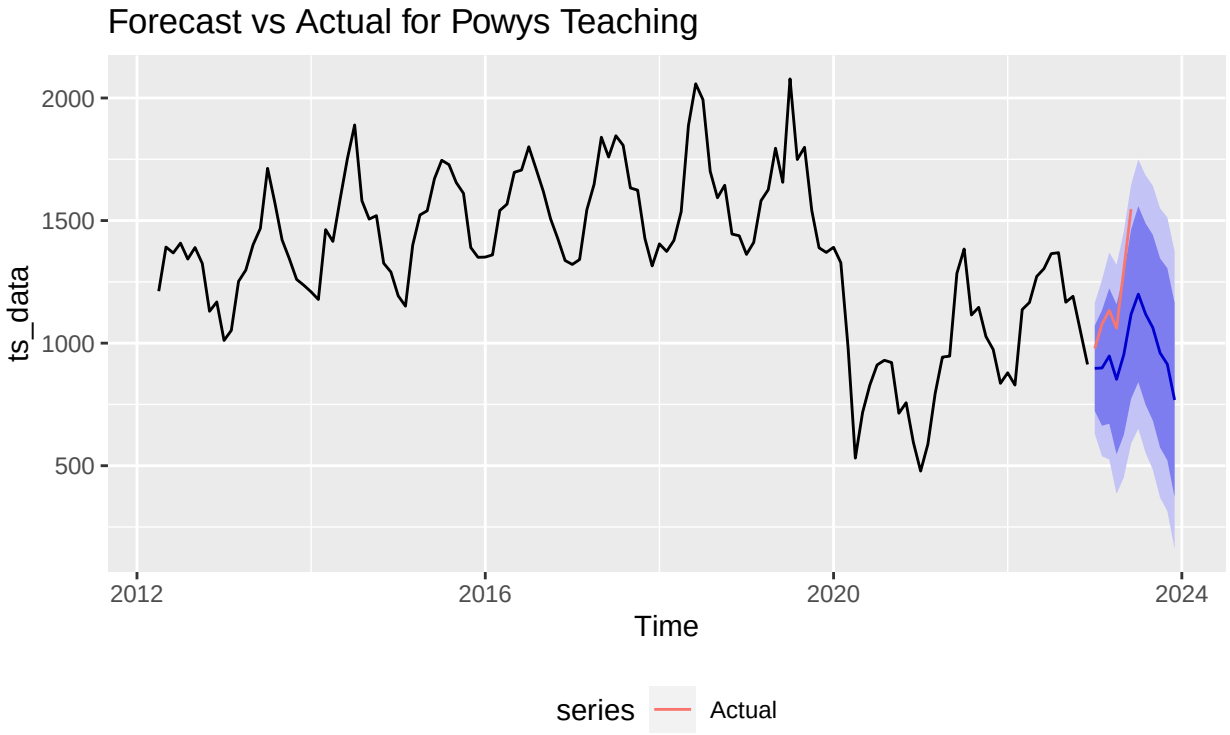


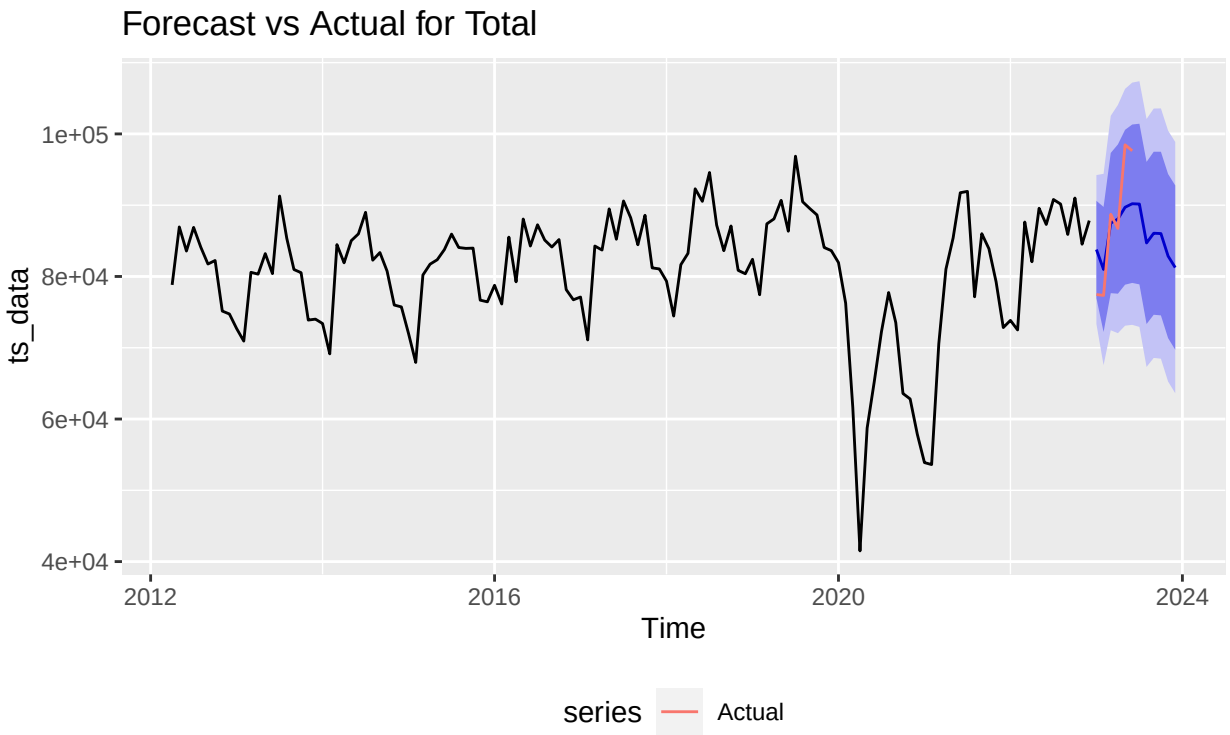



Forecast vs Actual for Cardiff & Vale
15000
12500
10000
7500
ts_data
2012
2016
2020
2024
Time
series
Actual


Forecast vs Actual for Grouped_4_organisation
32000
28000
24000
20000
16000
ts_data
2012
2016
2020
2024
Time
series
Actual

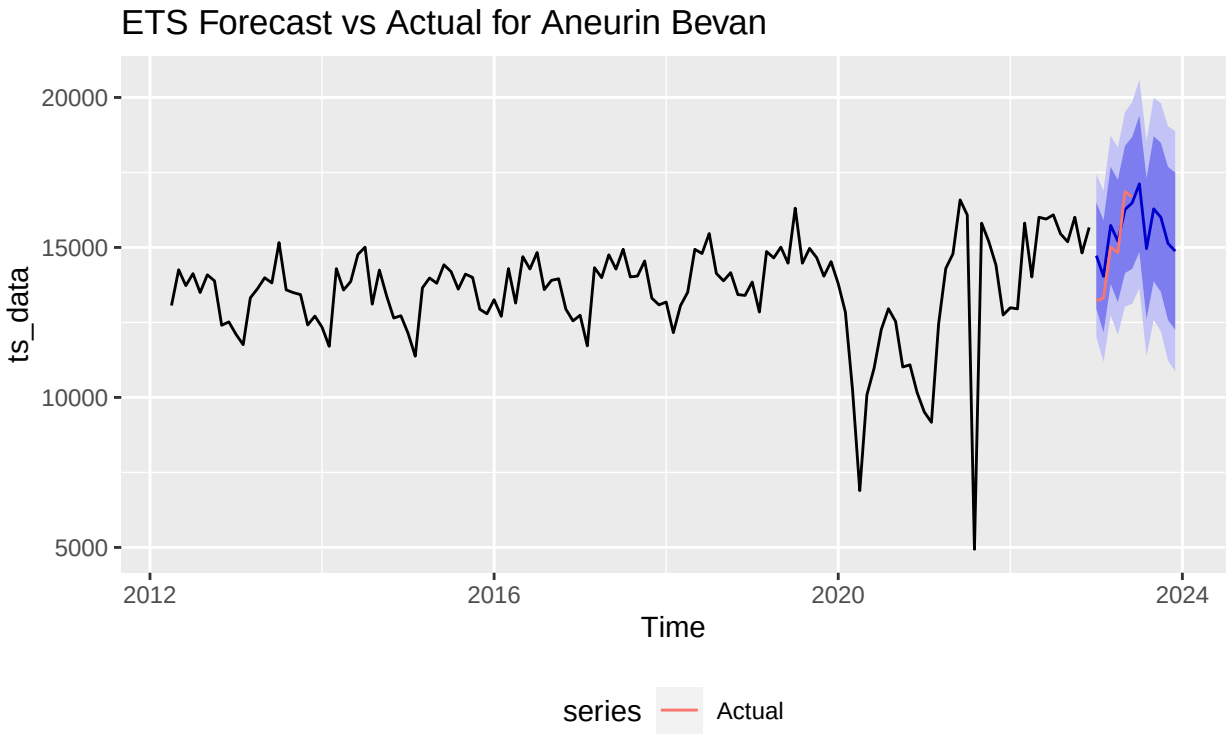

Forecast vs Actual for Hywel Dda
18000
15000
12000
9000
6000
ts_data
2012
2016
2020
2024
Time
series
Actual

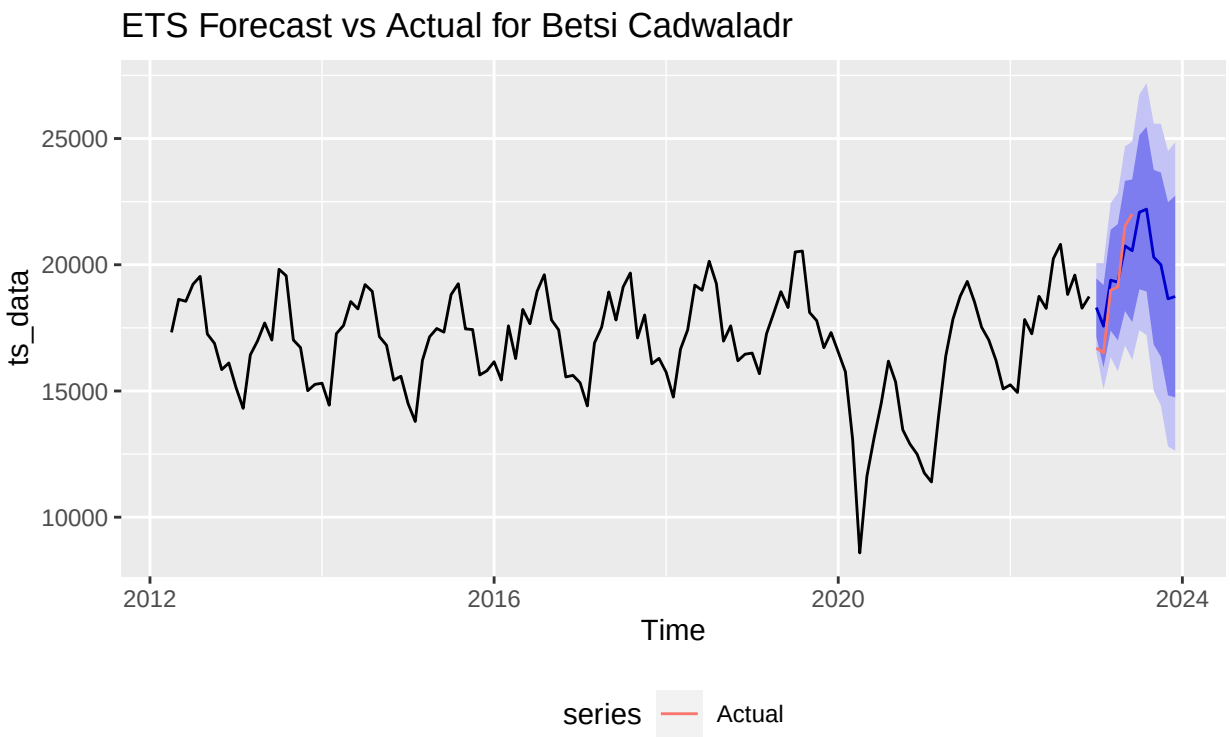

Forecast vs Actual for Powys Teaching
2000
1500
1000
500
ts_data
2012
2016
2020
2024
Time
series
Actual

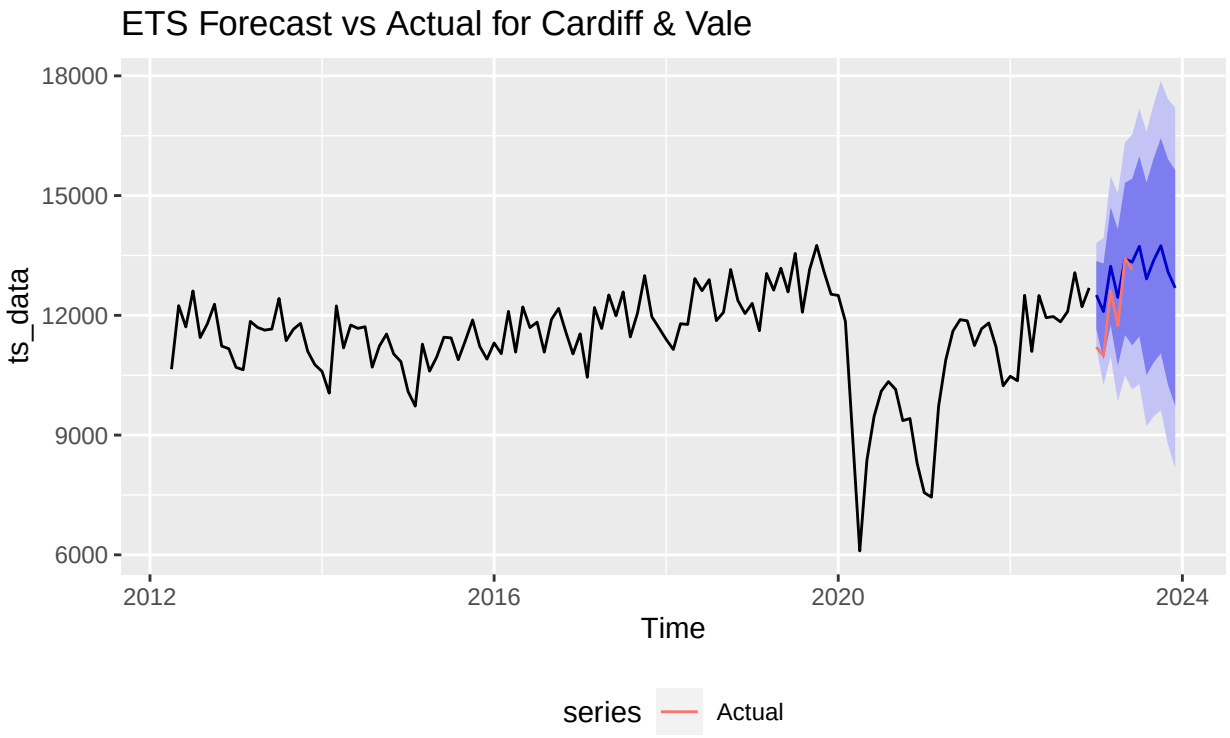

Forecast vs Actual for Total
1e+05
8e+04
6e+04
4e+04
ts_data
2012
2016
2020
2024
Time
series
Actual

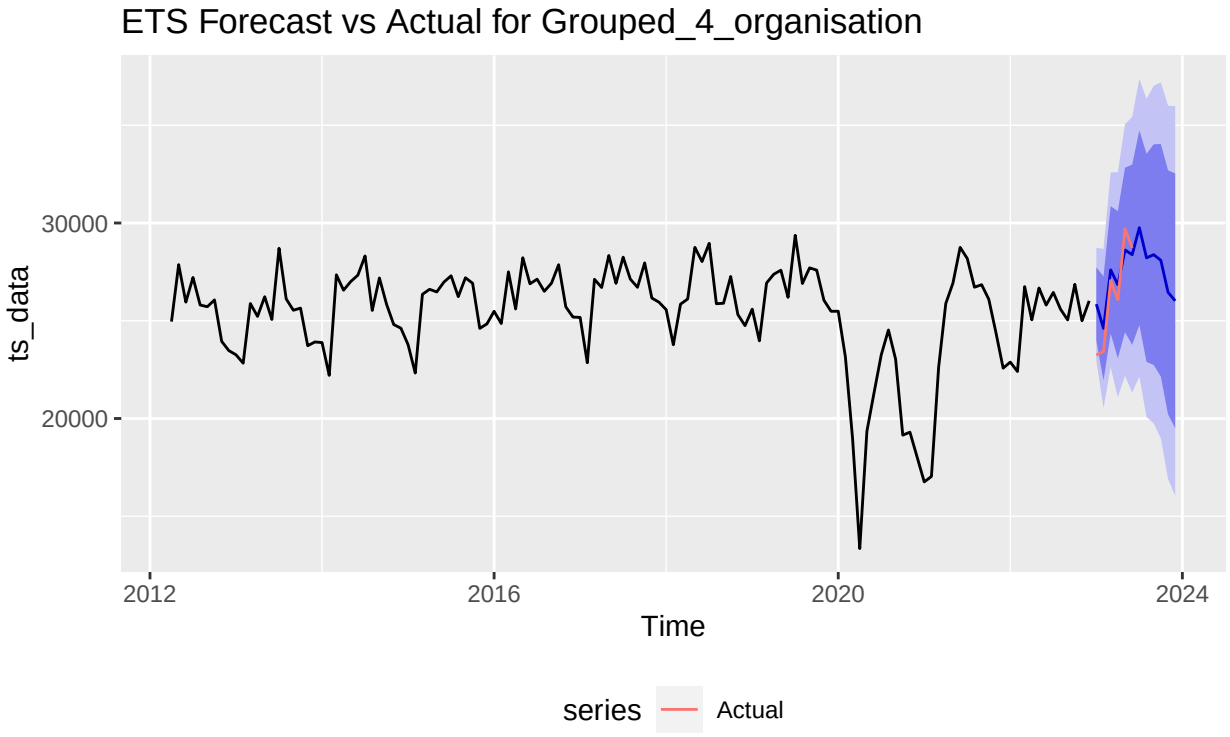
4 Forecasting with ETS

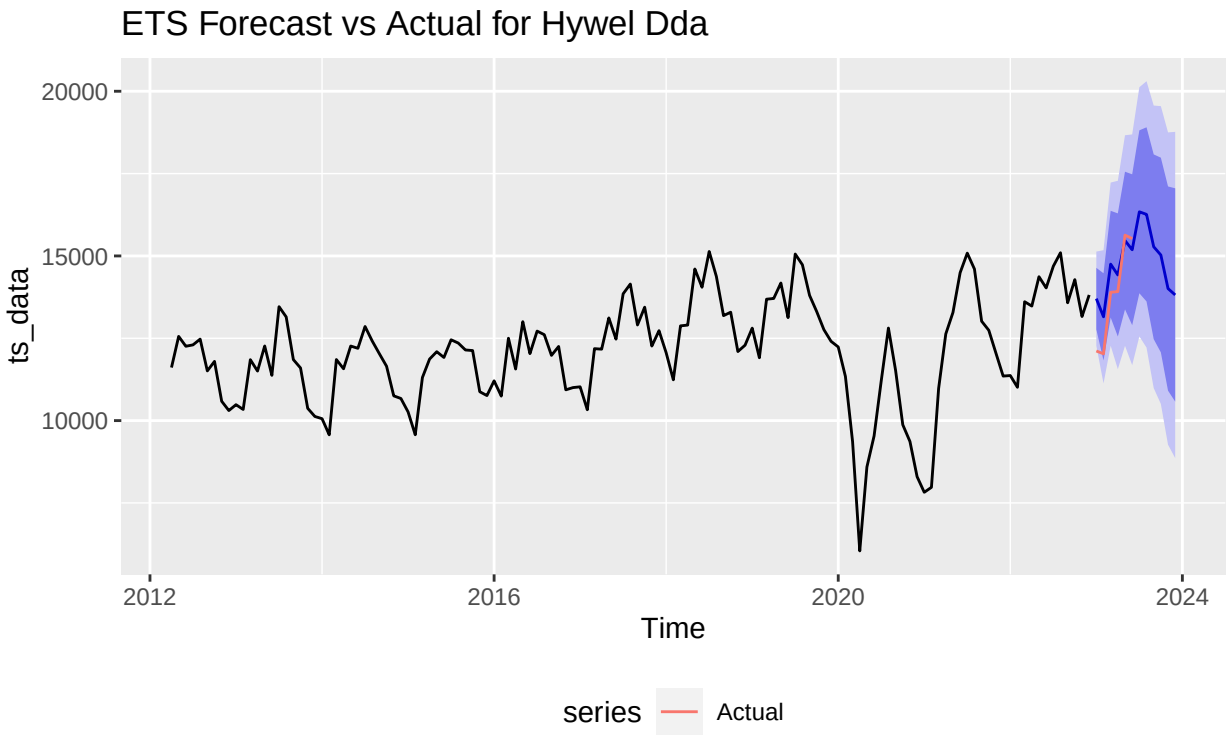
4.1 Plotting with ets

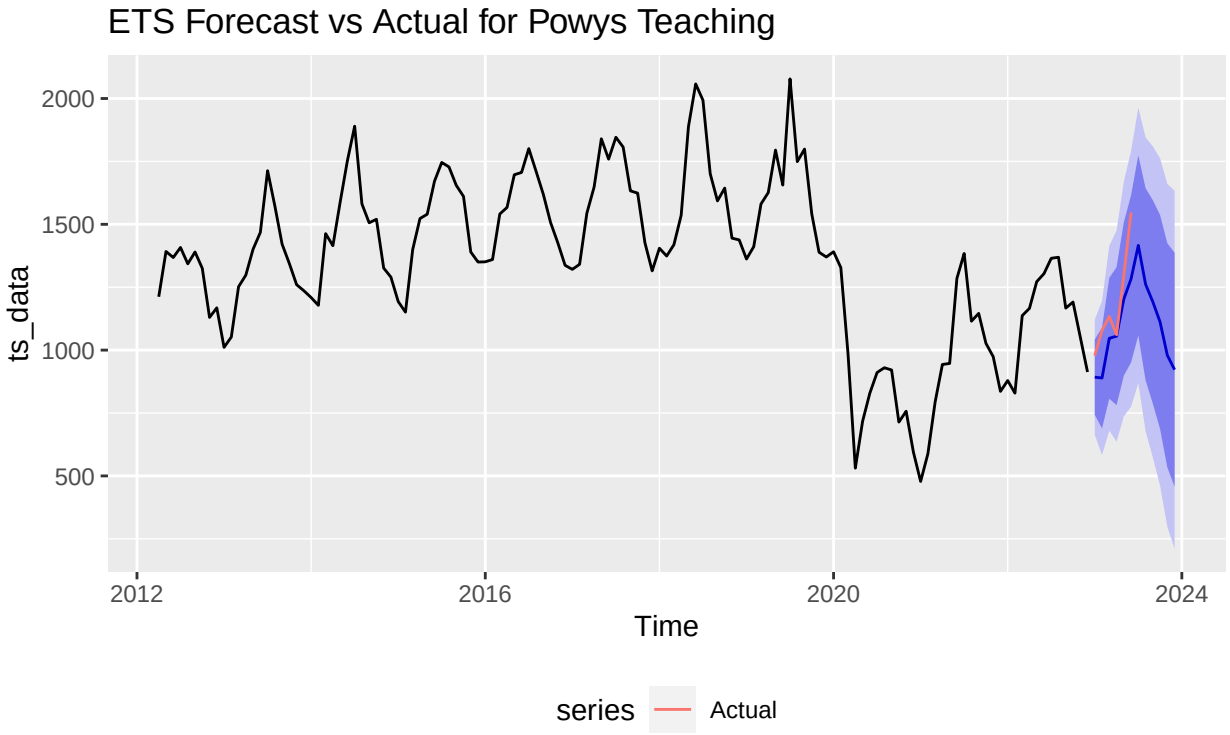


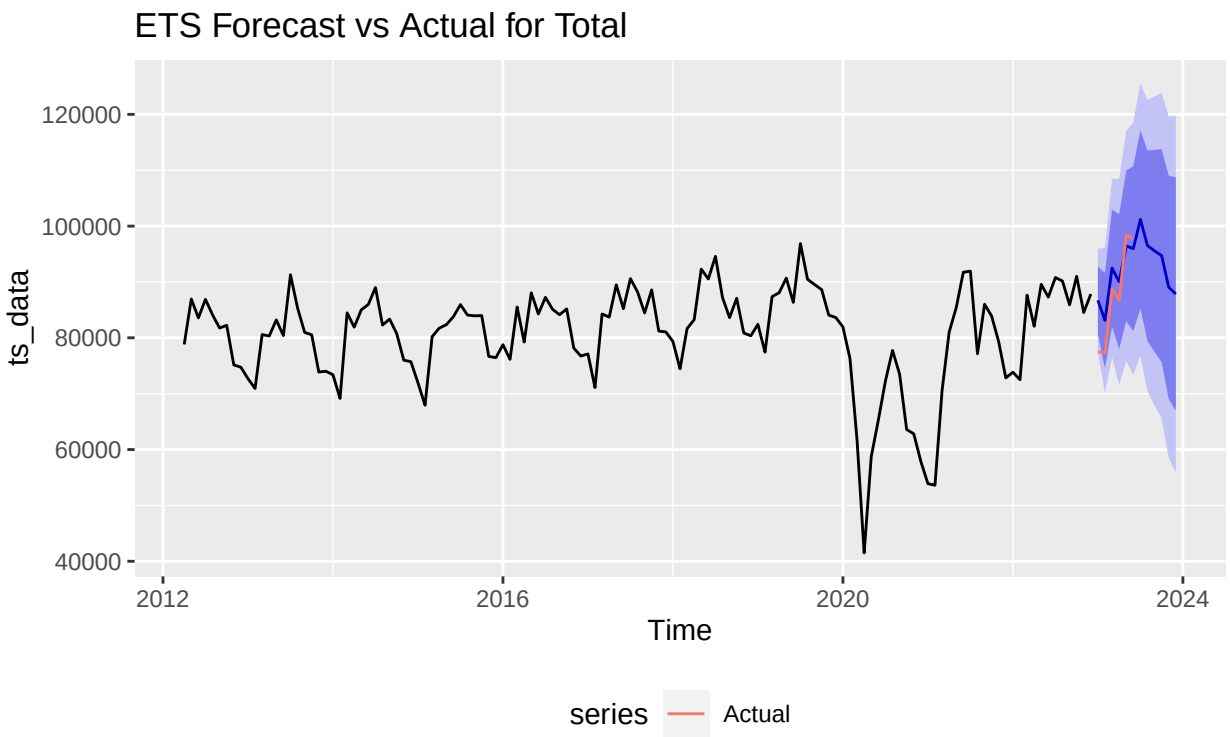



ETS Forecast vs Actual for Cardiff & Vale
18000
15000
12000
9000
6000
ts_data
2012
2016
2020
2024
Time
series
Actual


ETS Forecast vs Actual for Grouped_4_organisation
30000
20000
ts_data
2012
2016
2020
2024
Time
series
Actual


ETS Forecast vs Actual for Hywel Dda
20000
15000
10000
ts_data
2012
2016
2020
2024
Time
series
Actual


ETS Forecast vs Actual for Powys Teaching
2000
1500
1000
500
ts_data
2012
2016
2020
2024
Time
series
Actual



- Conformity:

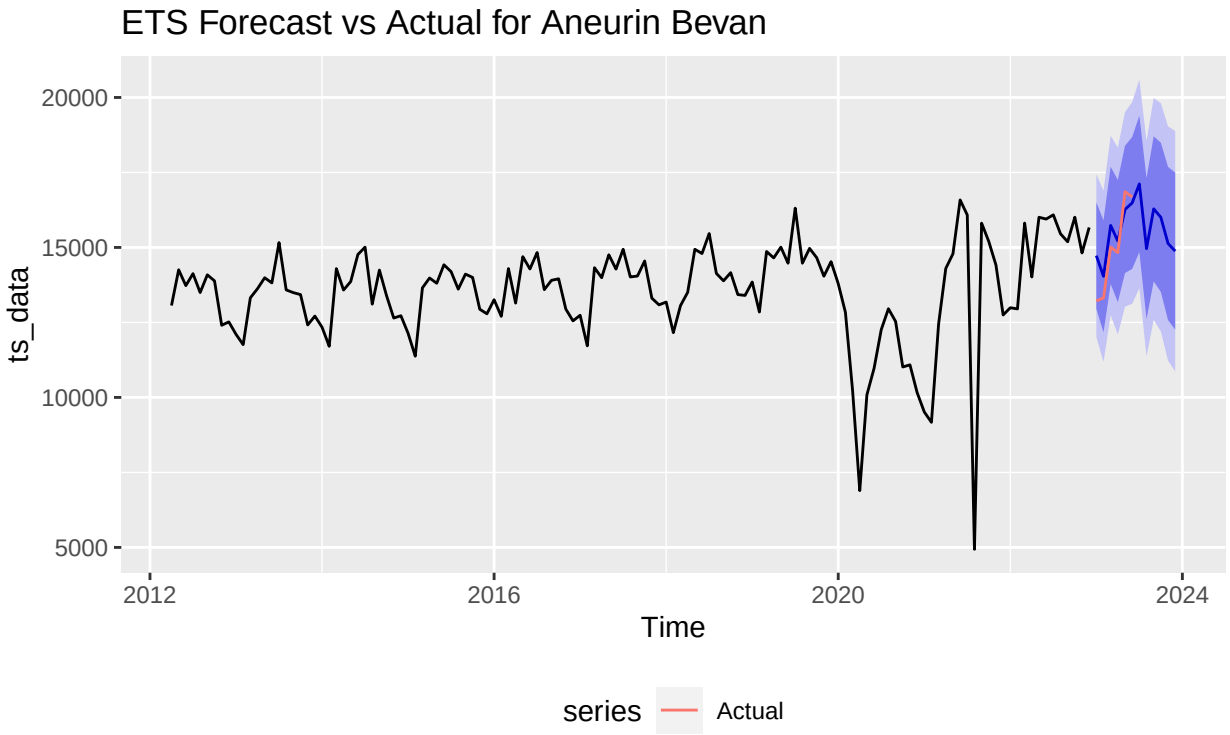
Hywel Dda, Powys Teaching, Cardiff & Vale, Grouped_4_Organisation, and Total (All-Wales): For these health boards, the ARIMA model effectively predicted the downturn that was observed in the latter portions of 2023 and successfully reflected the actual data, indicating a cogent direction. This demonstrates 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.

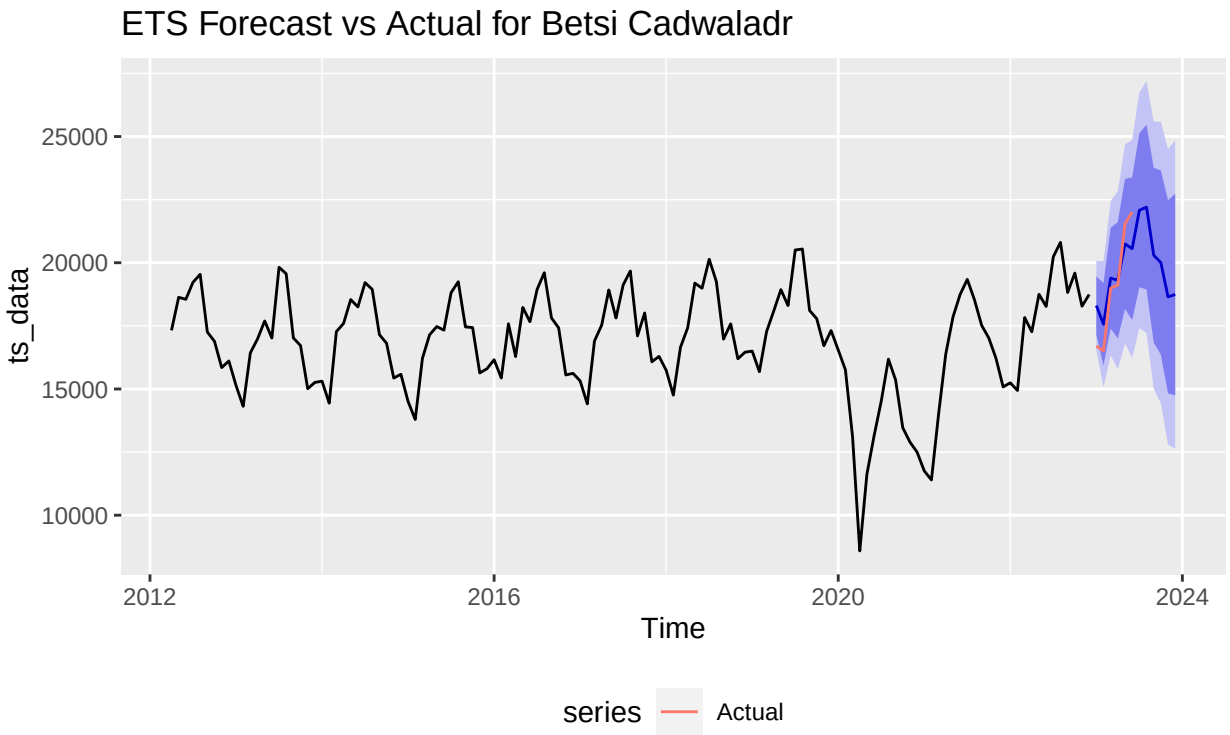
- Contradictions:

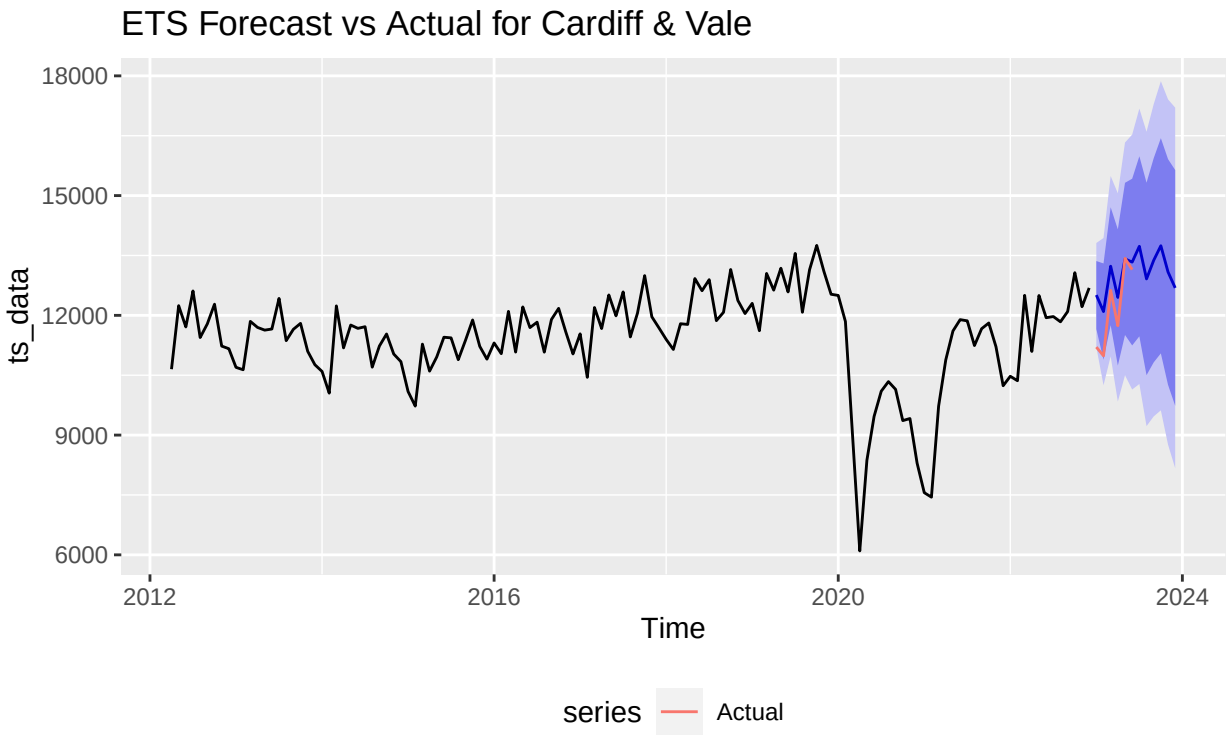
Aneurin Bevan and Betsi Cadwaladr: By contrast, the prediction statistics for Aneurin Bevan and Betsi Cadwaladr show a clear departure from the actual numbers. It is necessary to assess the ARIMA model's applicability and accuracy with regard to these specific datasets because there is a discrepancy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

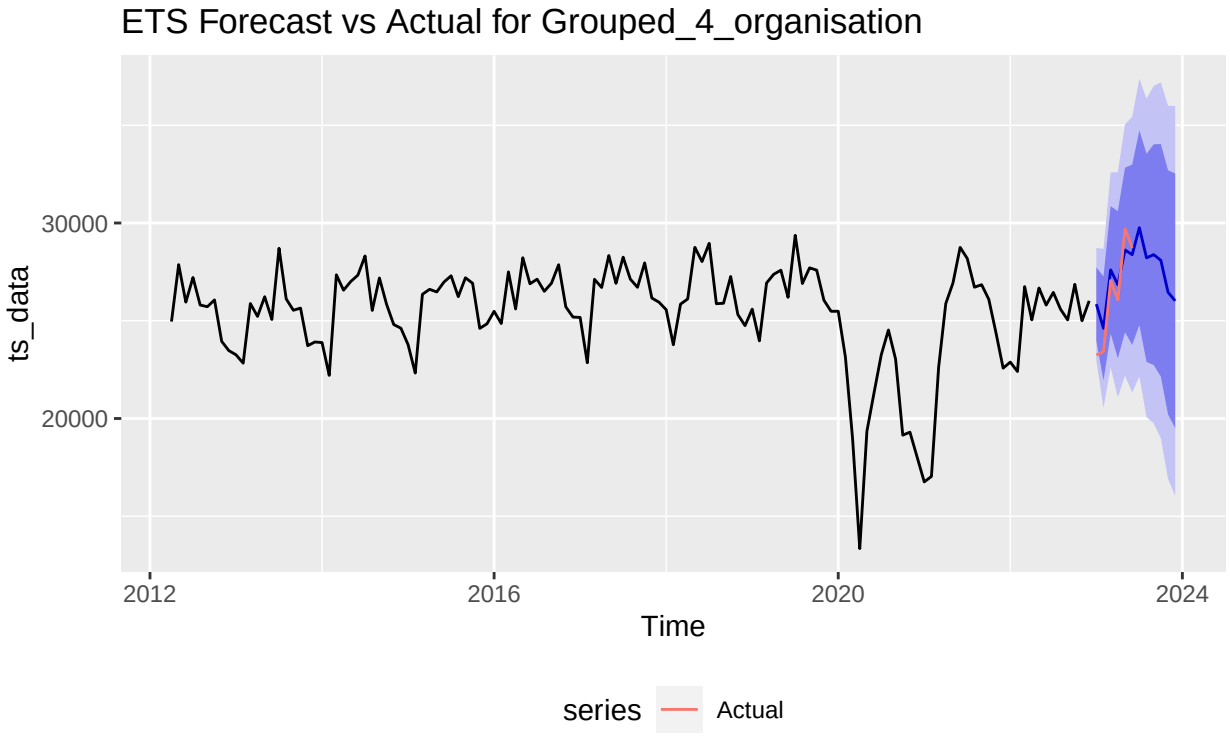
5 Forecasting with ETS

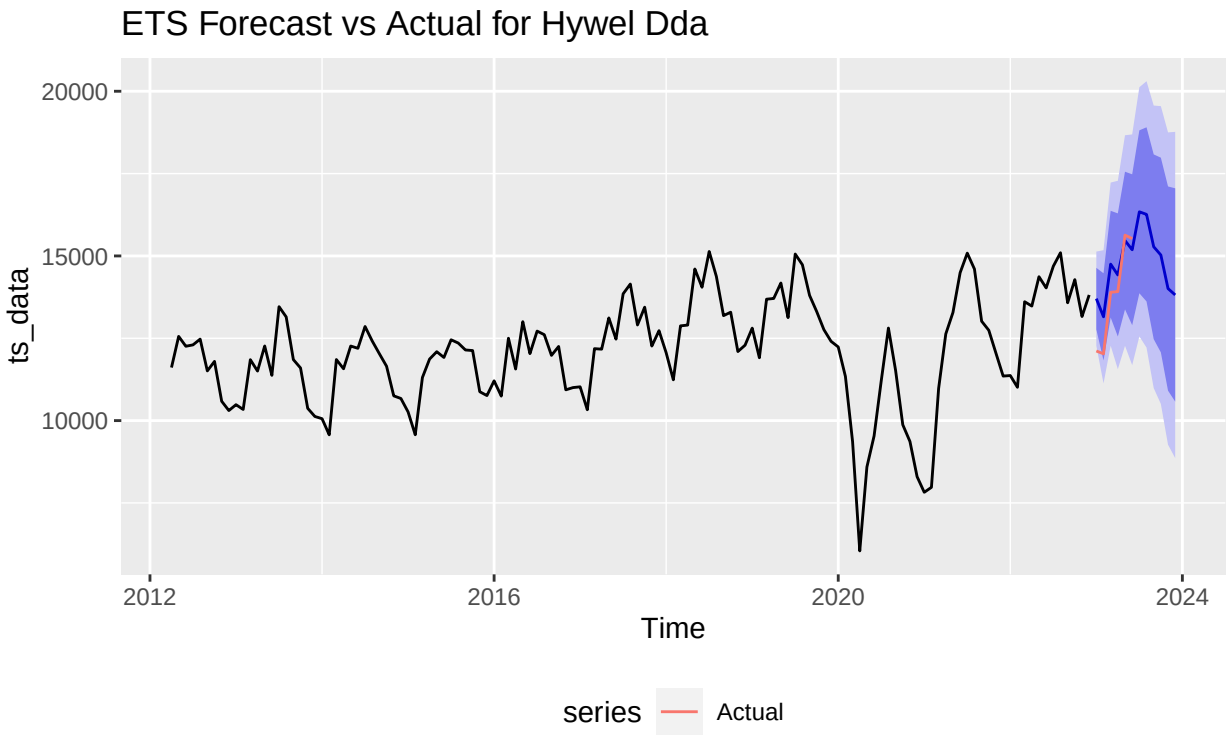
5.1 Plotting with ets

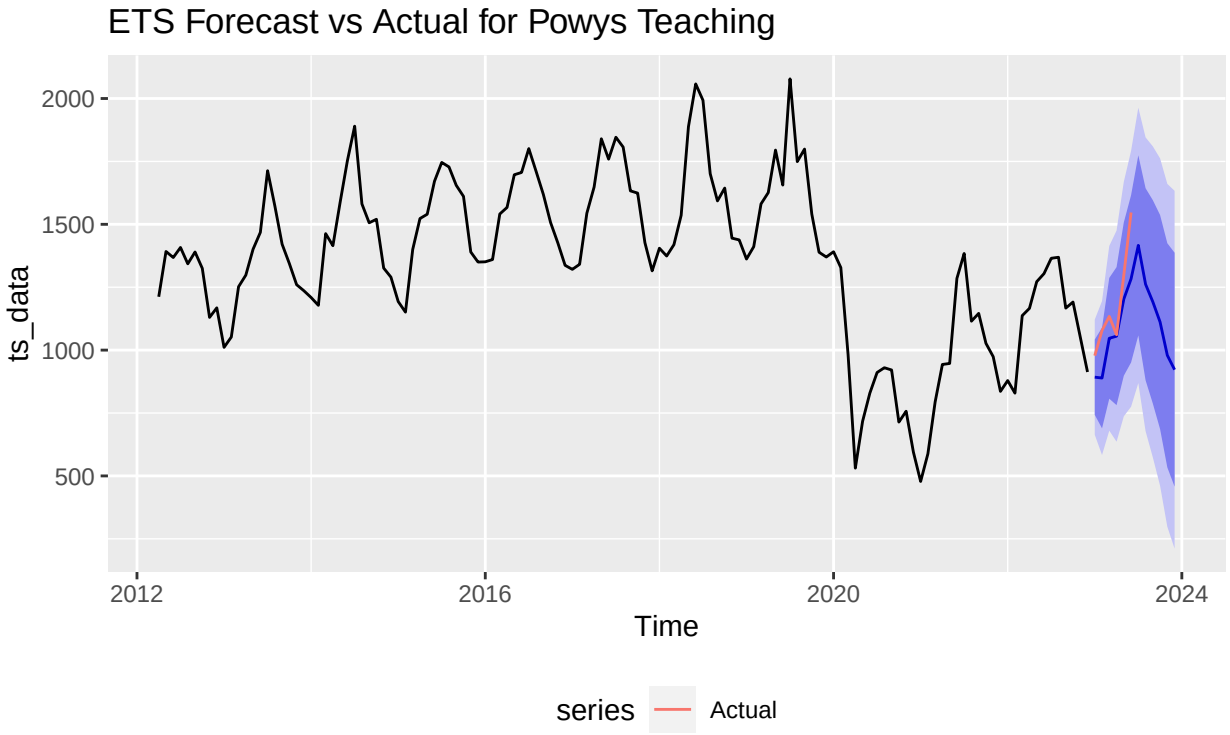


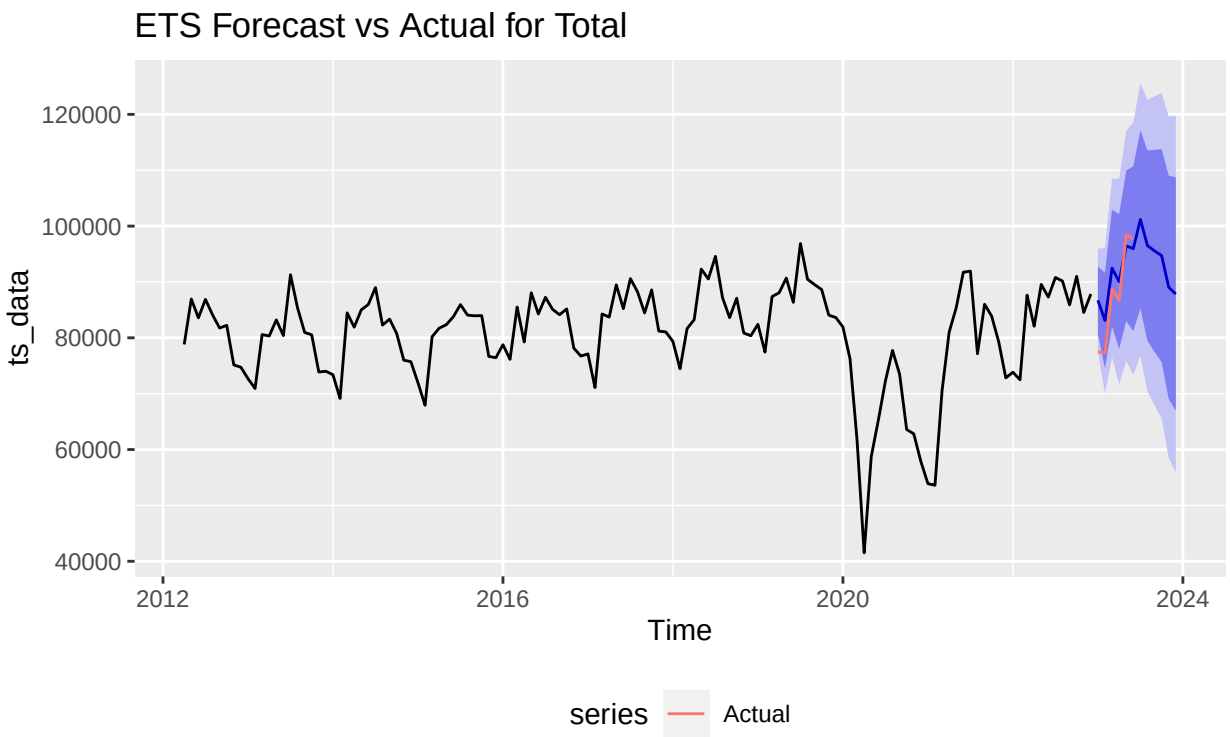



ETS Forecast vs Actual for Cardiff & Vale
18000
15000
12000
9000
6000
ts_data
2012
2016
2020
2024
Time
series
Actual


ETS Forecast vs Actual for Grouped_4_organisation
30000
20000
ts_data
2012
2016
2020
2024
Time
series
Actual


ETS Forecast vs Actual for Hywel Dda
20000
15000
10000
ts_data
2012
2016
2020
2024
Time
series
Actual


ETS Forecast vs Actual for Powys Teaching
2000
1500
1000
500
ts_data
2012
2016
2020
2024
Time
series
Actual



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 recurrent pattern 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.

An accuracy analysis was then carried out upon for both ARIMA and ETS models to further assess the quality of their predictions.

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

LHB	ARIMA	ETS
Aneurin Bevan.MAE	1748.634	688.417
Aneurin Bevan.RMSE	1992.543	801.644
Aneurin Bevan.MAPE	0.113	0.048
Betsi Cadwaladr.MAE	3157.257	915.735
Betsi Cadwaladr.RMSE	3756.403	1050.122
Betsi Cadwaladr.MAPE	0.156	0.049
Cardiff & Vale.MAE	600.429	654.811
Cardiff & Vale.RMSE	648.376	799.030
Cardiff & Vale.MAPE	0.050	0.057
Grouped_4_organisation.MAE	1386.277	1089.509
Grouped_4_organisation.RMSE	1598.559	1315.071
Grouped_4_organisation.MAPE	0.052	0.043
Hywel Dda.MAE	804.597	763.085
Hywel Dda.RMSE	923.853	905.503
Hywel Dda.MAPE	0.058	0.059
Powys Teaching.MAE	238.265	120.999
Powys Teaching.RMSE	264.667	146.796
Powys Teaching.MAPE	0.192	0.098
Total.MAE	4781.817	4297.460
Total.RMSE	5601.875	5014.551
Total.MAPE	0.054	0.052

- 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

7.3 stretch_tsibble(.init = 135-36, .step = 1)

```
## [1] 1.831502
```

```
## [1] 1.767515
```

```
## # A tibble: 8 x 5
##   model_name  model        organisation  MASE RMSSE
##   <chr>       <chr>        <chr>        <dbl> <dbl>
## 1 naive_model <aggregated> MASE          1.60  1.55
## 2 naive_model <aggregated> RMSSE         1.60  1.55
## 3 ets_model   <aggregated> MASE          1.59  1.55
## 4 ets_model   <aggregated> RMSSE         1.59  1.55
## 5 tscount     <aggregated> MASE          1.93  1.94
## 6 tscount     <aggregated> RMSSE         1.93  1.94
## 7 comb        <aggregated> MASE          1.69  1.65
## 8 comb        <aggregated> RMSSE         1.69  1.65
```

7.3.1 tscount seems to be performing worse than the naive model, therefore, I would like to remove that from the models included.

7.4 Remove tscount

```
## # A tibble: 8 x 5
##   model_name  model        organisation   MASE  RMSSE
##   <chr>       <chr>        <chr>         <dbl>  <dbl>
## 1 naive_model <aggregated> MASE           1.60   1.55
## 2 naive_model <aggregated> RMSSE          1.60   1.55
## 3 ets_model   <aggregated> MASE           1.59   1.55
## 4 ets_model   <aggregated> RMSSE          1.59   1.55
## 5 tscount     <aggregated> MASE         NaN    NaN
## 6 tscount     <aggregated> RMSSE        NaN    NaN
## 7 comb        <aggregated> MASE           1.59   1.55
## 8 comb        <aggregated> RMSSE          1.59   1.55
```

joined_data <- as_tibble(test_data) %>% left_join(as_tibble(fc_cv), by = c("YearMonth", "Aggregated_Organisation")) %>% select(-attendance.y) %>% rename(attendance = attendance.x) %>% rename(forecast = .mean)

joined_data_long<- joined_data %>% pivot_longer(cols = c(attendance, forecast), names_to = "Type", values_to = "Value")

unique_organisations <- unique(joined_data_long$Aggregated_Organisation)

plot_list <- map(unique_organisations, ~{ ggplot(joined_data_long[joined_data_long$Aggregated_Organisation == .x,], aes(x = YearMonth, y = Value, color = Type)) + geom_line() + labs(title = paste("Forecasted vs Actual Values for", .x), y = "Values") + theme_minimal() }) print(plot_list)

8 Running t-tests and extracting p-value and t-value

t_test_summary <- joined_data %>% summarise(t_value = list(t.test(attendance, forecast)$statistic), p_value = $list(t.test(attendance, forecast)$p.value)) %>% unnest(cols = c(t_value, p_value))

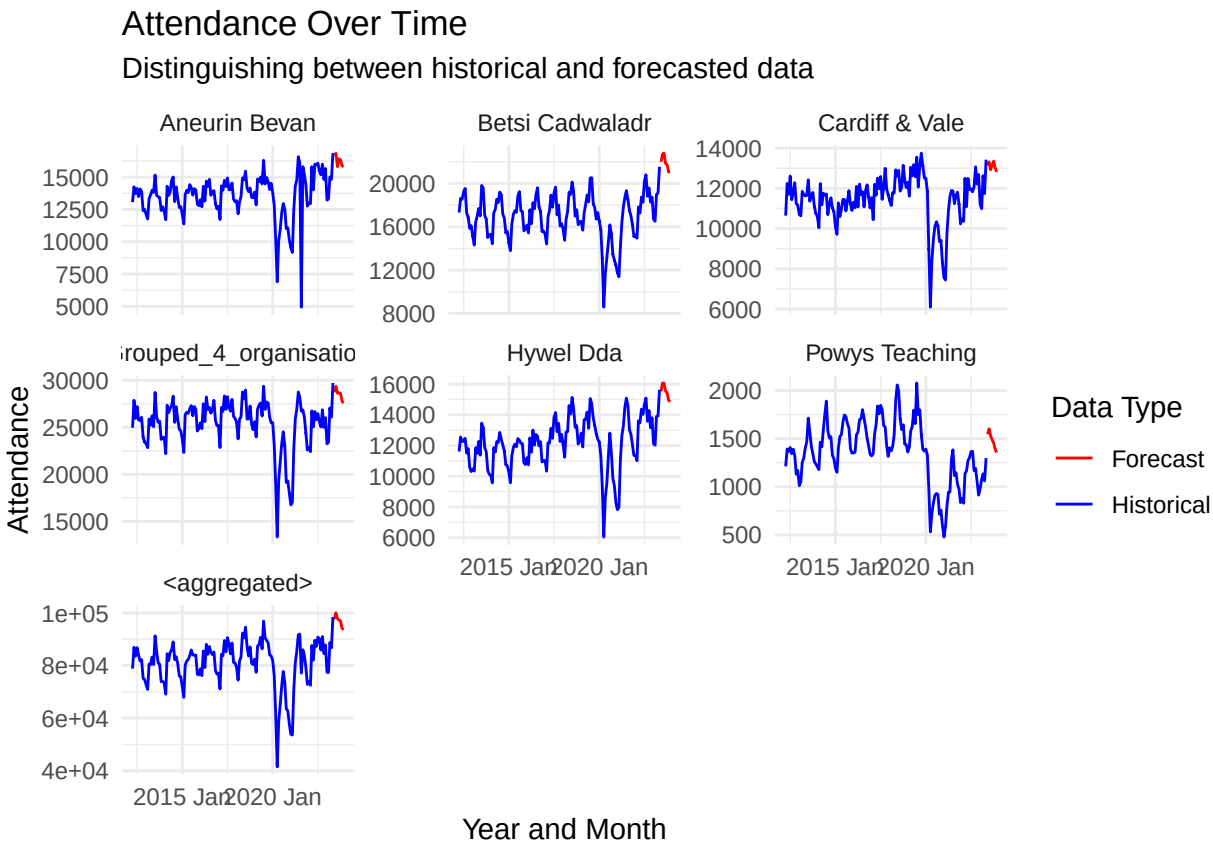
print(t_test_summary)

The p value is relatively high. It suggests that we fail to reject the null hypothesis, meaning there is no 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.

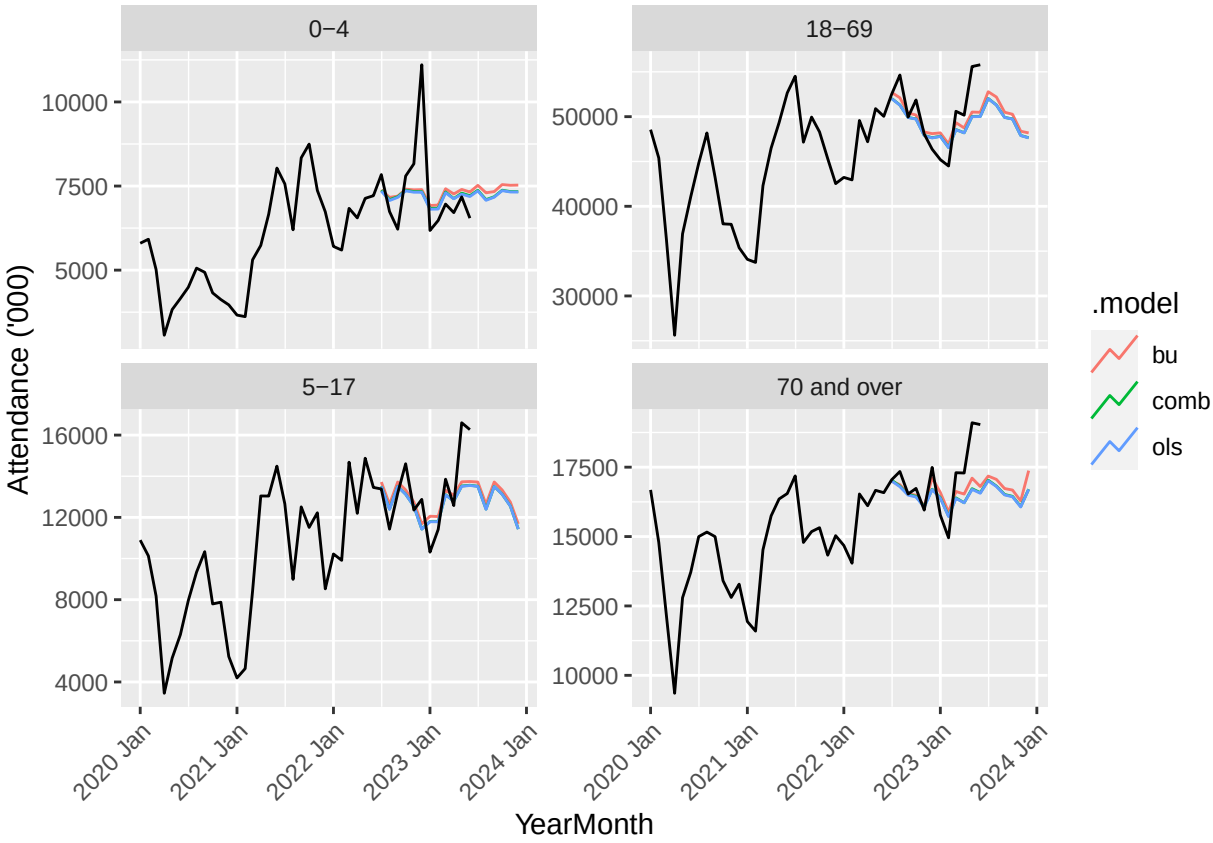
- Base Forecasting: This technique is used to forecast straightforward time series data. Using statistical or machine learning models to extrapolate future data points based on observed past data is the main focus of this technique. Base forecasting, which is frequently applied when dealing with a single time series, tries to find patterns, trends, and seasonality in the historical data in order to produce precise and trustworthy future projections.
- Reconciliation Forecasting (Hierarchical Forecasting): Reconciliation or Hierarchical forecasting is typically used when managing hierarchical or grouped time series data. It involves generating forecasts on several levels of aggregation. This process not only creates projections for each individual time series, but also reconciles the forecasts to guarantee consistency throughout the hierarchy's many levels. The forecasts may be distributed across hierarchical levels using a variety of strategies, such as top-down, bottom-up, to make sure that the projections at each level coincide with their corresponding aggregate levels. When the data contains a hierarchical structure, such as organisational, product, or geographic hierarchies, hierarchical forecasting is especially advantageous since it ensures consistent and consolidated projections at all levels.

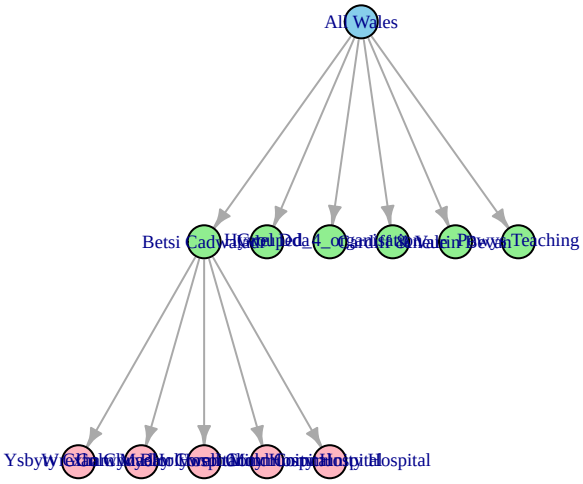
8.1 Reconcile

8.1.1 1. Fitting models and Forecasting without tscount



8.2 Include Age Group




All Wales
Betsi Cadwaladr
Teaching