**Changes: After 06/03 Meeting**

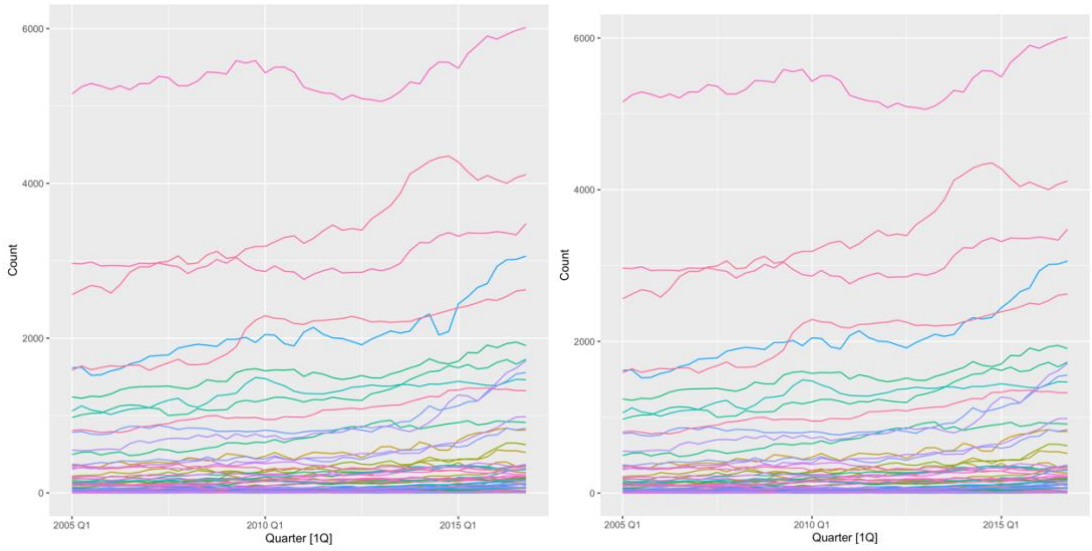
1. One step result calculated
2. Verified error calculation (overall error is calculated considering the whole hierarchy)
3. DeepAR and WaveNet cluster wise results – 20 clusters per dataset

**Changes:**

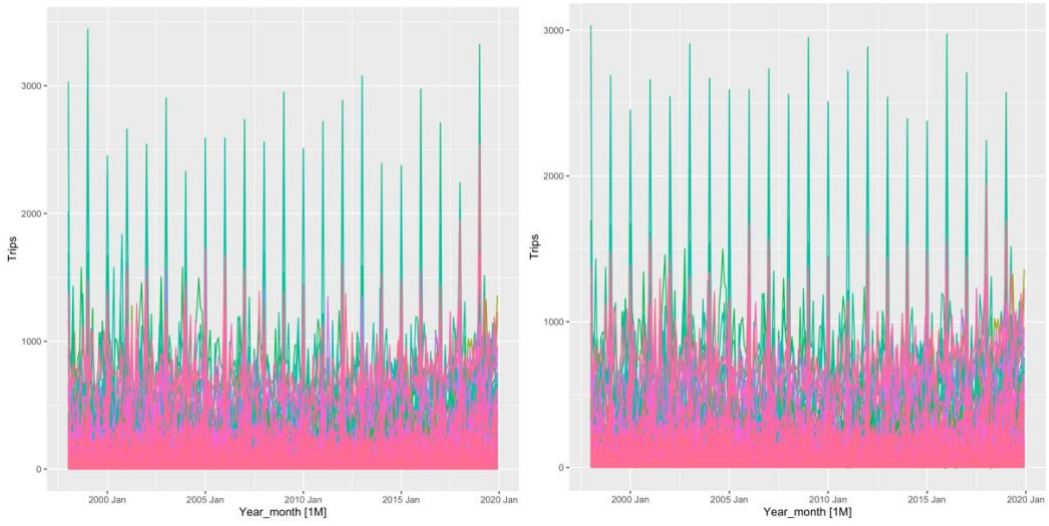
1. Data Cleaning for all datasets
2. Wikipedia dataset – hierarchy structure changed based of the paper, Total X Access X Agent X Language X Purpose X Article

**Data Before and After Cleaning**

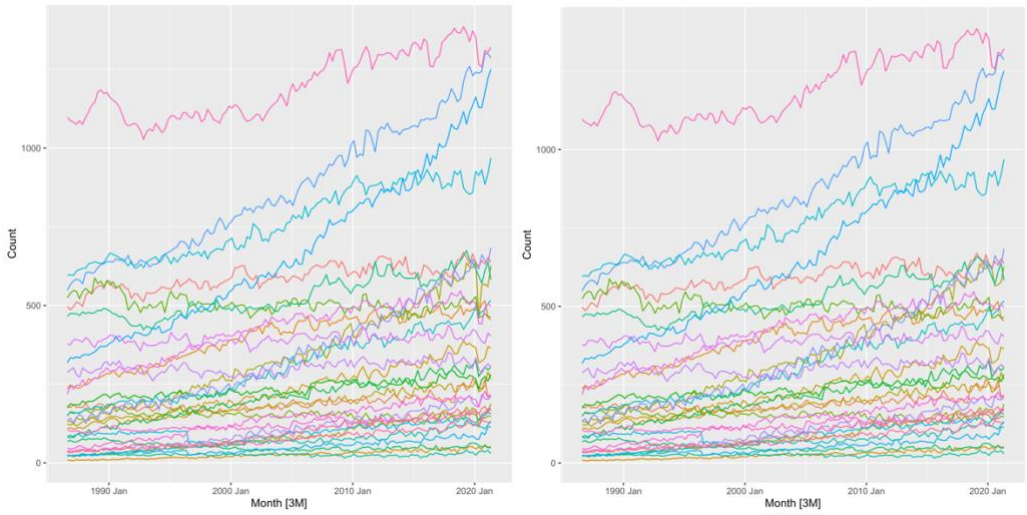
**Prison**



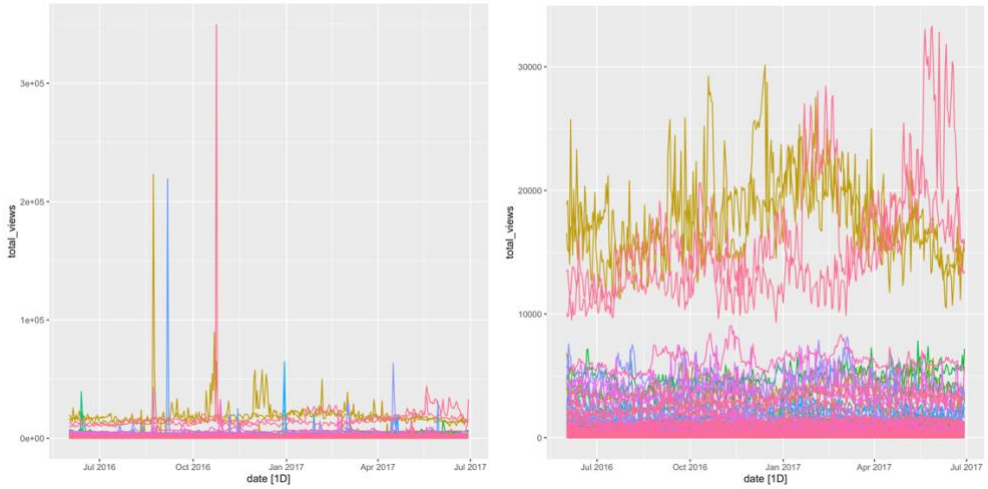

**Tourism**



**Labour**




**Wikipedia**




**Summary Data and Results**

| Dataset | Frequency | Horizon | Number of Samples | Minimum training sample length | Number of levels | Number of total time series |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| Prison | 4 (quarterly) | 8 | 3 | 24 | 5 | 121 |
| Tourism | 12 (monthly) | 12 | 10 | 144 | 3 | 85 |
| Wikipedia | 7 (weekly) | 7 | 10 | 324 | 6 | 1095 |
| Labour | 4 (quarterly) | 12 | 5 | 68 | 4 | 57 |

**Error calculation**

- Step 1 – For each time series in the hierarchy calculate the error metric (MSE, MAPE)
- Step 2 – Calculate the mean error across the time series in each level (this gives the mean error for each level)
- Step 3 – Calculate the overall error by getting the mean error across all time series errors we have in Step 1
- Step 4 – Repeat step 1-3 for all samples
- Step 5 – Calculate the mean error for each level and overall, across the samples
- Step 6 – Calculate the percentage improvement for each level and overall

**Results Across Samples – Latest**

**Full Horizon**

Base Errors (MSE) - Overall

| Dataset | ARIMA | ETS | DeepAR | DeepAR Cluster | WaveNet | WaveNet Cluster |
|---|---|---|---|---|---|---|
| Prison | 52548.6 | 49629.380 | 121514 | 150438 | 1.836182e+08 | 1.278420e+06 |
| Tourism | 70573.5 | 65806.981 | 75410 | 81104 | 117250.810 | 126934.165 |
| Labour | 1231.7 | 2313.894 | 3221.896 | 2587 | 1922.840 | 3053.138 |
| Wikipedia | 3.477905e+05 | 2.862994e+05 | 4.503426e+05 | 5.041211e+05 | 8.904260e+05 | 8.794240e+05 |

Best Approach from Reconciliation

| Dataset | ARIMA | ETS | DeepAR | DeepAR Cluster | WaveNet | WaveNet Cluster |
|---|---|---|---|---|---|---|
| Prison | OLS<br><br>Best ML Rank – 4 (Case 2 Lambda 0.1-0.9) | Case 1 Lambda 0. 1-0. 9 No skip /<br><br>Case 1 Lambda 1 | Bottom-UP<br><br>Best ML Rank – 5 (Case2 Lambda 0.1, 0.9) | Bottom-UP<br><br>Best ML Rank – 2 (Case1 Lambda 0.01, 5) | BottomUp<br><br>Best ML Rank – 2 (Case 1 Lambda 1-4) | Case 2 Lambda 0.01-5 |
| Tourism | OLS<br><br>Best ML Rank – 3 Case 1 Lambda 0.01-5 | OLS<br><br>Best ML Rank- 2 Case 2 Lambda 1 | MintShrink<br><br>Best ML Rank -4 (Case 1 Lambda 0.1-0.9) | OLS<br><br>Best ML Rank- 2 Case 1 Lambda 1-4 | Case 2 Lambda 0.01-5 | Case 2 Lambda 0.01-5 |
| Labour | OLS Best ML Rank – 4 (Case 2 Lambda 0.1-0.9) | Case 2 Lambda 0.01-0.09 | Case 1/2 Lambda 0.01-0.09 | Case 2 Lambda 1 | MintShrink<br><br>Best ML Case – 4 (Case 2 Lambda 0.01-0.09) | Case 1 Lambda 0.01-5 |
| Wikipedia | Case 1 Lambda 0.01-0.09 / Case 2 Lambda 1-4 | Bottom-Up<br><br>Best ML Rank – 4 (Case 1 Lambda 0.01-5) | WLS<br><br>Best ML Case – 5 (Case 2 Lambda 1-4) | Bottom Up Best ML Rank – 4 (Case 2 Lambda 1 | Case 2 Lambda 1-4 | MintShrink<br><br>Best ML Rank – 2 Case 1 Lambda 0.01-5 |

**Short Horizon**

Base Errors (MSE)

| Dataset | ARIMA | ETS | DeepAR | DeepAR Cluster | WaveNet | WaveNet Cluster |
|---|---|---|---|---|---|---|
| Prison | 16240.217 | 15644.543 | 56893.103 | 62184.950 | 1.861848e+08 | 1.442180e+06 |
| Tourism | 67303.917 | 60432.873 | 77717.717 | 82379.230 | 129727.399 | 118862.489 |
| Labour | 603.014 | 689.654 | 1285.324 | 1365.070 | 1154.330 | 1886.196 |
| Wikipedia | 2.835496e+05 | 2.140000e+05 | 3.452034e+05 | 4.231807e+05 | 9.779535e+05 | 8.054878e+05 |

Best Approach from Reconciliation

| Dataset | ARIMA | ETS | DeepAR | DeepAR Cluster | WaveNet | WaveNet Cluster |
|---|---|---|---|---|---|---|
| Prison | Case 1 Lambda 0.1-0.9 | Case 2 Lambda 0.01-0.09 | Bottom-UP<br><br>Best ML Rank – 5 (Case2 Lambda 0.1, 0.9) | Bottom-UP<br><br>Best ML Rank – 3 (Case1 Lambda 0.01-5) | Bottom-UP<br><br>Best ML Rank 2 (Case 1 Lambda 1-4) | Case 2 Lambda 0.01-5 |
| Tourism | OLS<br><br>Best ML Rank – 3 Case 2 Lambda 0.01-5 | Case 2 Lambda 1 | Case 1 Lambda 0.01-0.09 | Case 1 Lambda 1-4 | Case 2 Lambda 0.01-5 | Case 2 Lambda 0.01-5 |
| Labour | OLS Best ML Rank – 3 (Case 2 Lambda 0.01-5) | Case 1 Lambda 0.01-5 | Case 1 Lambda 0.01-0.09 | MinTSample<br><br>Best ML Rank -2 Case 2 Lambda 1 | Case 1 Lambda 0.1-0.9 | Case 2 Lambda 1 |
| Wikipedia | Case 1 Lambda 0.1-0.9 | Case 2 Lambda 1-4 | OLS<br><br>Best ML Rank – 5 (Case 2 Lambda 1-4) | BottomUp<br><br>Best ML Rank – 3 (Case 1 Lambda 0.01-5) | MinTShrink<br><br>Best ML Rank 2 (Case 2 Lambda 1-4) | Case 1 Lambda 0.01-5 |

**One Step Horizon**

Base Errors (MSE)

| Dataset | ARIMA | ETS | DeepAR | DeepAR Cluster | WaveNet | WaveNet Cluster |
|---|---|---|---|---|---|---|
| Prison | 3766.827 | 3805.240 | 33142.810 | 24172.620 | 1.863123e+08 | 1.412184e+06 |
| Tourism | 84939.040 | 83091.175 | 96057.683 | 103622.035 | 3.298383e+05 | 200148.687 |
| Labour | 320.678 | 346.358 | 713.982 | 823.582 | 475.626 | 822.036 |
| Wikipedia | 1.447196e+05 | 9.047025e+04 | 1.697056e+05 | 2.116968e+05 | 4.565305e+05 | 3.783312e+05 |

Best Approach from Reconciliation

| Dataset | ARIMA | ETS | DeepAR | DeepAR Cluster | WaveNet | WaveNet Cluster |
|---|---|---|---|---|---|---|
| Prison | Case 1 Lambda 0.1-0.9 | Case 2 Lambda 0.01-0.09 | Bottom - Up<br><br>Best ML Rank – 5 (Case1 Lambda 0.01, 0.09) | Bottom -Up<br><br>Best ML Rank – 3 (Case1 Lambda 0.01, 5) | Bottom-UP<br><br>Best ML Rank 2 (Case 1 Lambda 1-4) | Case 2 Lambda 0.01-5 |
| Tourism | Case 1 Lambda 1 | MinT Shrink<br><br>Best ML Rank – 4 Case 2 Lambda 0.01-0.09 | WLS<br><br>Best ML Rank – 5<br><br>Case 2 Lambda 0.01-0.09 | Case 2 Lambda 0.1-0.9 | Case 2 Lambda 0.01-5 | Case 2 Lambda 0.01-0.09 |
| Labour | OLS<br><br>Best ML Rank – 4 (Case 2 Lambda 0.1-0.9) | OLS<br><br>Best ML Rank - 2 Case 1 Lambda 1 | Bottom Up<br><br>Best ML Rank – 6 Case 1 Lambda 0.01-0.09 | MinT Shrink<br><br>Best ML Rank – 6 Case 1 Lambda 0.1-0.9 | WLS<br><br>Best ML Rank – 6 Case 1 lambda 0.1-0.9 | OLS<br><br>Best ML Rank – 3 Case 1 lambda 0.01-5 |
| Wikipedia | Case 1 Lambda 0.1-0.9 | Bottom-UP<br><br>Best ML Rank - 2 Case 2 Lambda 0.01-0.09 | WLS<br><br>Best ML Rank – 5 (Case 2 Lambda 1-4) | BottomUp<br><br>Best ML Rank – 5 (Case 1 Lambda 0.01-5) | MinTShrink<br><br>Best ML Rank 5 (Case 2 Lambda 1-4) | Case 1 Lambda 0.01-5 |

OLD EXPERIMENTS BEFORE CLEANING THE DATASETS

**Results Across Samples**

Full Horizon (*In the full horizon ML reconciliation performs best for most 3 out of datasets when the base model is ETS. However, when the base forecasts are from ARIMA the ML reconciliation performs best only for the Tourism dataset*)

| **Dataset** | **ARIMA** | **ETS** |
|---|---|---|
| Prison | OLS<br>Best ML Rank – 4 (Case 1 Lambda 1-4) | Case 1 Lambda 0.01-0.09 |
| Tourism | Case 1 Lambda 0.01-5 | Case 2 Lambda 1 |
| Labour | OLS<br>Best ML Rank – 4 (Case 2 Lambda 1) | Case 1 Lambda 0.01-0.09 |
| Wikipedia | OLS<br>Best ML Rank – 5 (Case 2 Lambda 0.1-0.9) | OLS<br>Best ML Rank – 5 (Case 2 Lambda 1) |

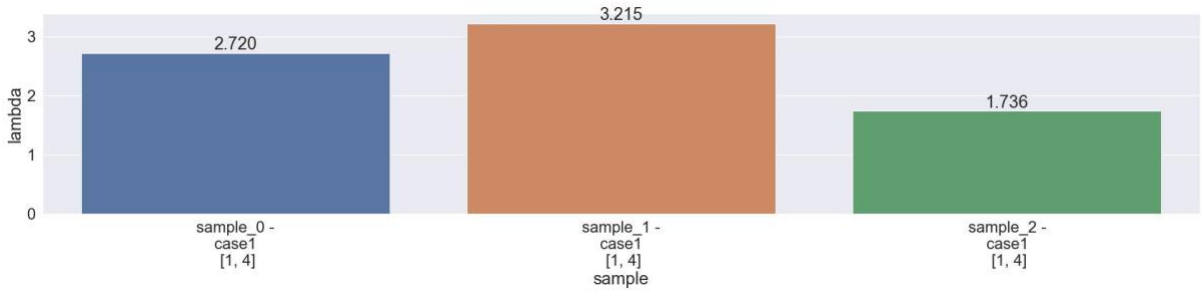
Short Horizon (*In the short horizon ML reconciliation performs best for most 3 out of datasets when the base model is ETS – similar observation as the full horizon. When the base forecasts are from ARIMA the ML reconciliation performs best for Prison and Tourism datasets*)

| **Dataset** | **ARIMA** | **ETS** |
|---|---|---|
| Prison | Case 1 Lambda 1-4 | Case 1 Lambda 0.01-0.09 |
| Tourism | Case 2 Lambda 0.01-5 | Case 1 Lambda 1-4 |
| Labour | OLS<br>Best ML Rank – 2 (Case 2 Lambda 1) | Case 1 Lambda 0.01-0.09 |
| Wikipedia | OLS<br>Best ML Rank – 5 (Case 2 Lambda 0.1-0.9) | OLS<br>Best ML Rank – 5 (Case 2 Lambda 0.01-5) |

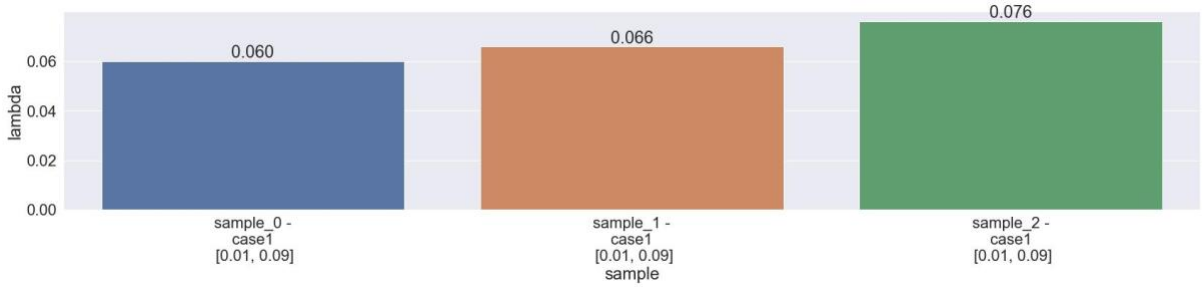
**Lambda Ranges – Lambda value for each sample from the overall best ML method**

**Prison**

ARIMA

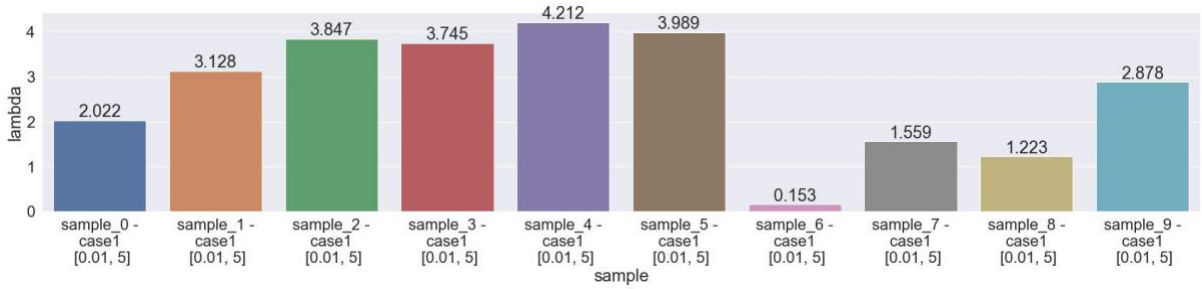



ETS

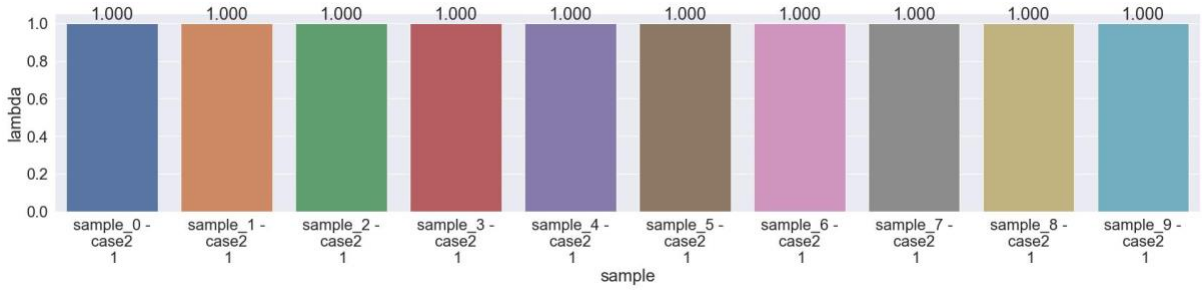



**Tourism**

ARIMA

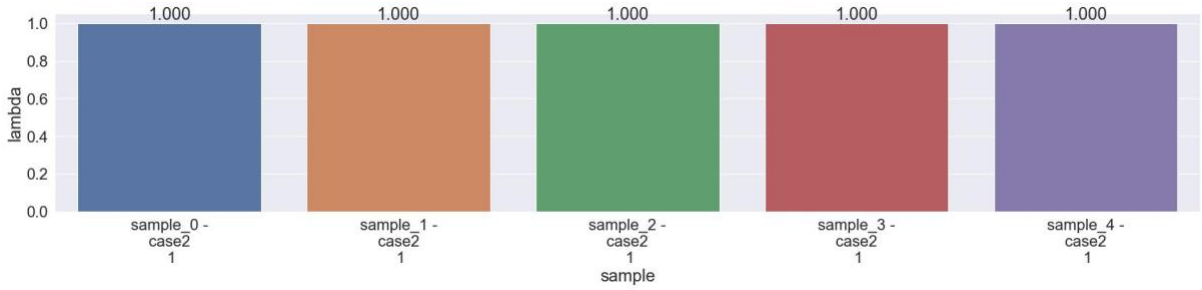



ETS

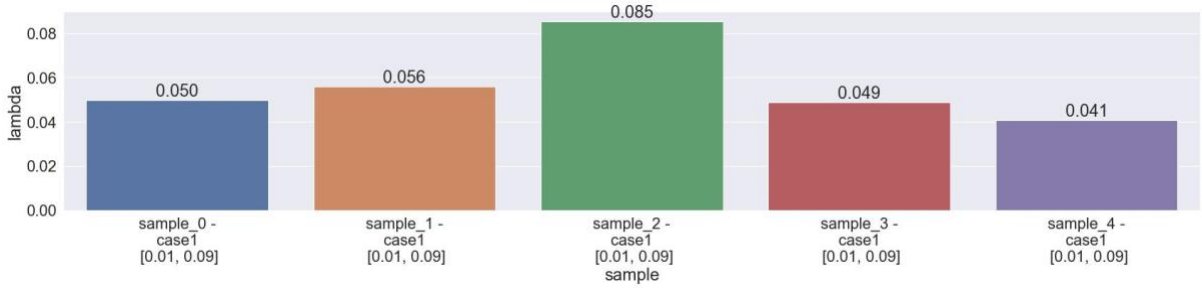


Labour

ARIMA

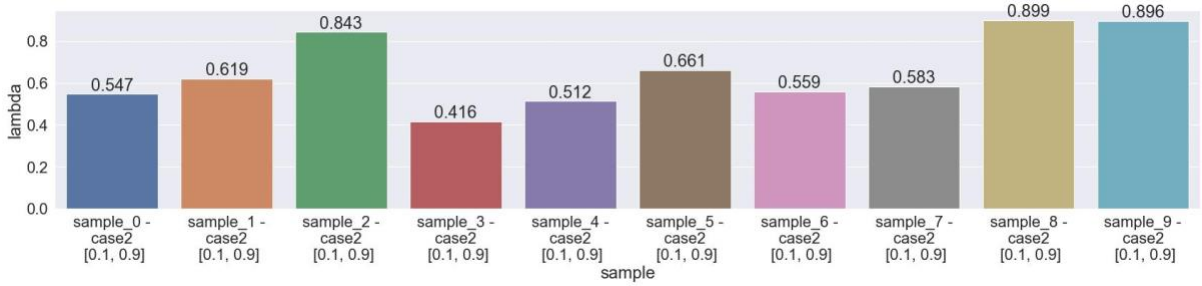



ETS

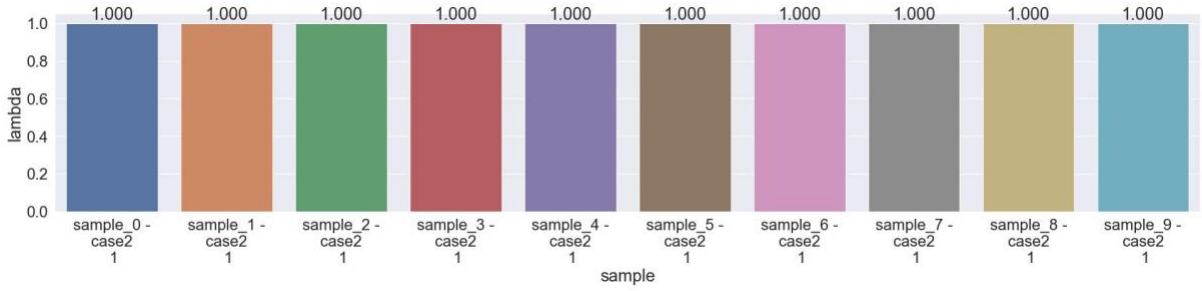



**Wikipedia**

ARIMA



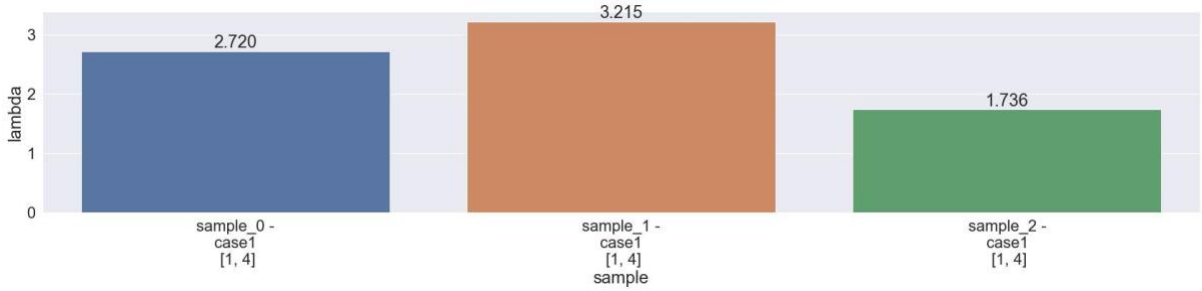

ETS



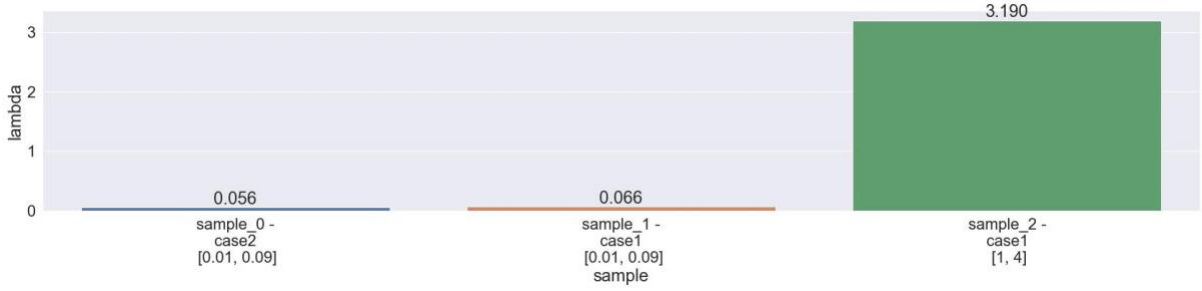

**Lambda Ranges – Lambda value and the best ML method per sample**

**Prison**

ARIMA

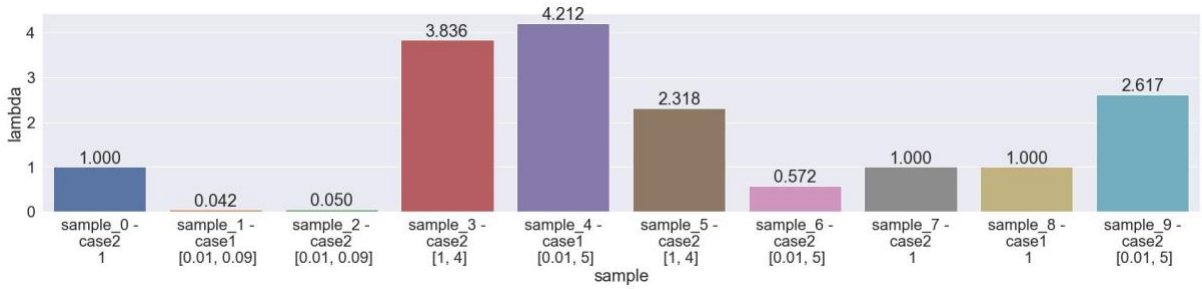



ETS

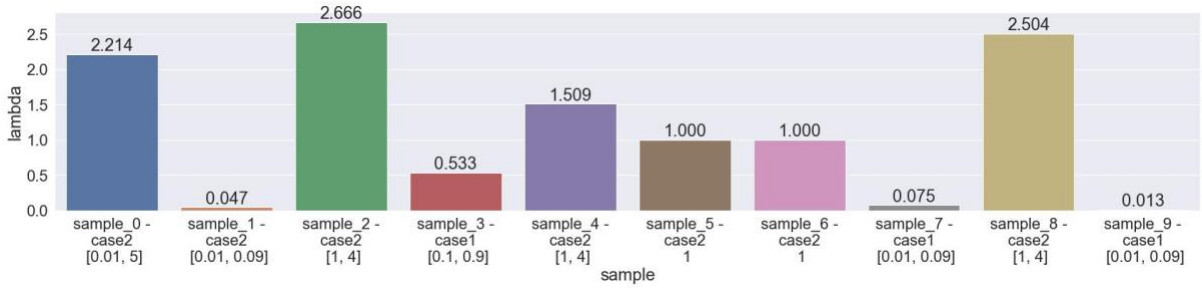



**Tourism**

ARIMA

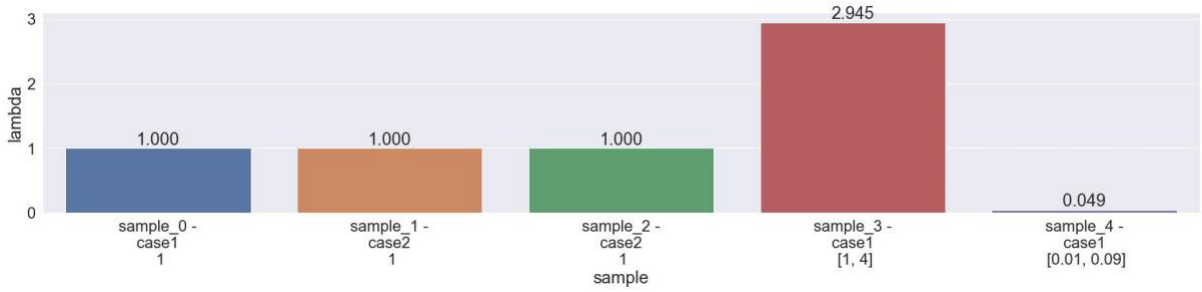



ETS

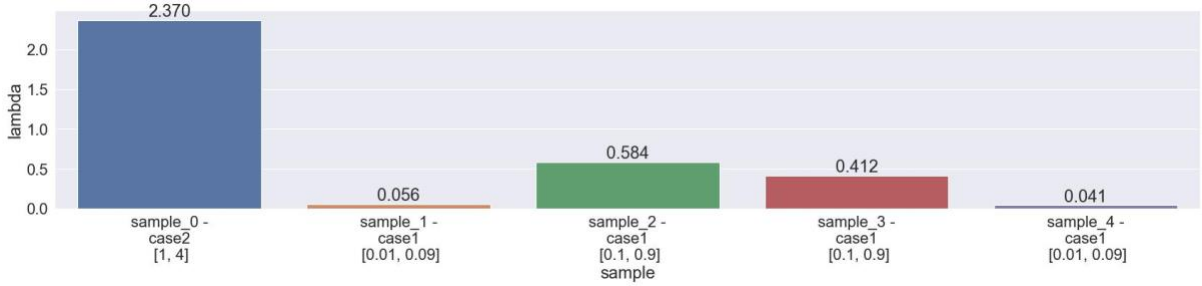



Labour

ARIMA

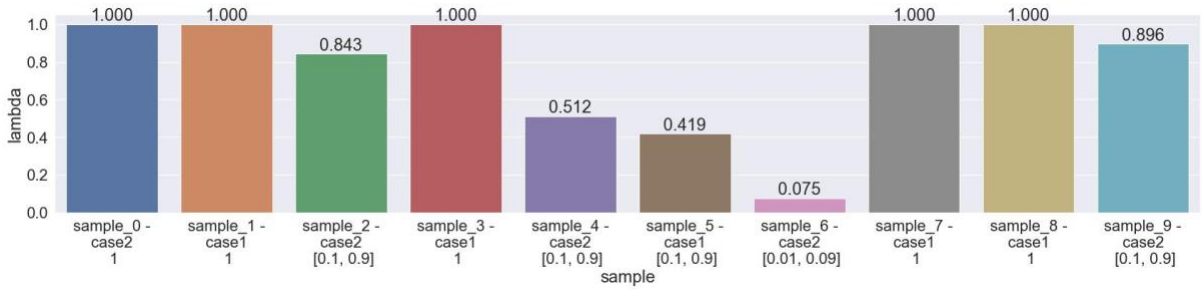



ETS



**Wikipedia**

ARIMA




ETS