Adelaide Airport – Forecasting Airline Passengers for Improving Passenger Load Factor

Group – A9

61710410	Dinesh Lulla
61710546	Simarjit Singh
61710545	Shubhankar Bivalkar
61710621	Lata Jha
61710528	Vibin Thomas
61710226	Lakshika Kothari

Contents

Executive Summary

Business Objective

Improve the Passenger Load Factor of Australian Domestic Airline to 82.5% by arriving at an optimum number of seats from Adelaide to six cities - Brisbane, Canberra, Gold Coast, Melbourne, Perth and Sydney for time period September 2016- September 2017.

Passenger Load Factor – Passenger load factor measures the capacity utilization for airlines. It signifies the efficiency with which an airline fills seats and generates revenues. 80% of passenger load factor is considered as standard in the domestic airline industry.

Forecasting Objective

Build a model **to forecast the aggregate passenger traffic on Australian Domestic Airlines for each month from September 2016 – September 2017** based on prior values. Evaluate multiple models and identify the optimum model based on MAPE (Mean Absolute Percentage Error), goodness of fit, distribution of residuals and auto correlation of residuals.

Data Set Overview

Monthly data with the respective passenger load factor for each of the 6 cities is available.

- **Data Set**- Airline passengers to 6 cities
- **Source**- Kaggle (https://www.kaggle.com/alphajuliet/au-dom-traffic)
- **Training data** – 42- 66 months of data
- **Validation data** - 18 months of data set
- **Forecasting horizon** - forecast passenger trips for 12 months Sep 2016- Sep 2017

Forecasting Model Overview

The data set was split into 6 different time series based on various cities – Canberra, Sydney, Melbourne, Perth, Gold Coast, and Brisbane. For each of the time series various forecasting models were based on the specified criteria and the optimum model was identified. (refer section Forecast Methods for more information)

Recommendations

- Based on the forecasted seats calculate the maximum capacity needed and the number of flights based on below equations
    - Max Seats = Forecasted Traffic / 82.5%

- Number of Flights = Max Seats/ 50 (1 domestic flight = 50 seats (appx))

- Based on the forecasted number determine the fleet optimization plan
  - Flight Planning and Schedule Development
  - Maintenance schedule
  - Flight lease plan

As the forecasted period is 6 months ahead the airline has enough time to optimize the number / type of planes on the route. Determining optimal number of flights would help Australian airlines in better scheduling of its flights and effective capacity utilization. Decisions such as leasing new flights or scheduling maintenance plans for individual aircrafts can be done effectively if an estimated demand is available with the airlines.

The estimated demand will hold true given underlying conditions remain unchanged. This model would work well if the forecasts are updated over time based on the actual values.

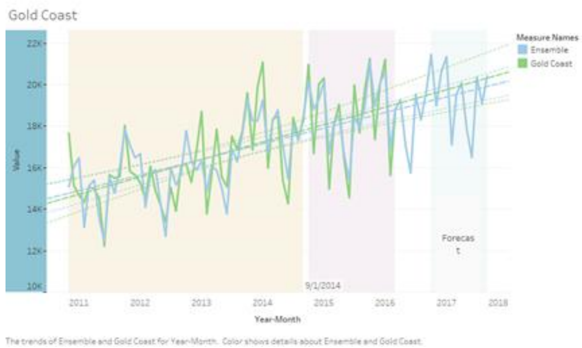
Technical Summary

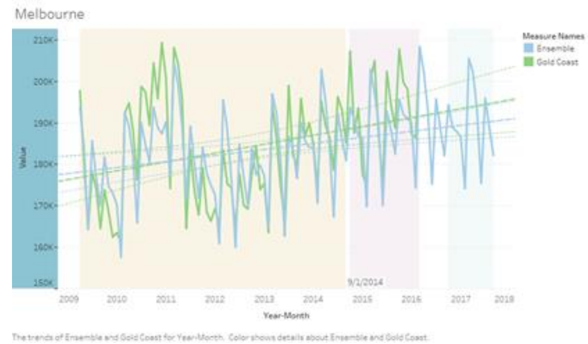
Data Preparation

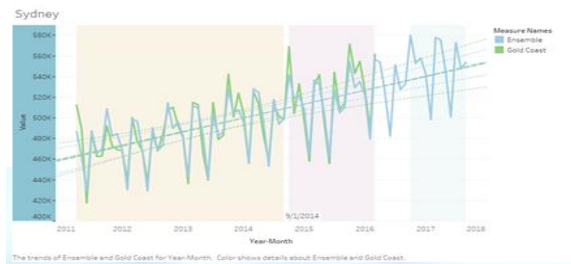
Data set from Kaggle was cleansed and filtered to restrict passenger trips from Adelaide to six cities, and for the duration from 2009 to 2016.

Visualization

Data sets were inspected for seasonality, trend and levels. Most of the city wise data showed monthly seasonality with Gold coast, Sydney Melbourne and Brisbane showing an increasing trend, Canberra showing a declining trend and no trend in Perth. After the visualization process data from 2009 was removed on finding an irregularity in the data for all cities except Melbourne and Perth.







Forecast Methods

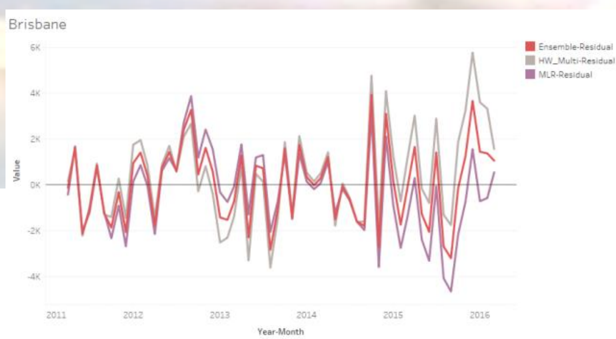
The final forecasting models used combinations of smoothing methods like the Holt Winters Model (No trend, additive and multiplicative) , Multiple Linear Regression and Ensemble Forecast. Each city used different sets of models based on the trend, seasonality and nature of the data set. To arrive at the final method the following models were evaluated and performance was compared against Naive (Lag1).

The forecasting models were evaluated on four parameters

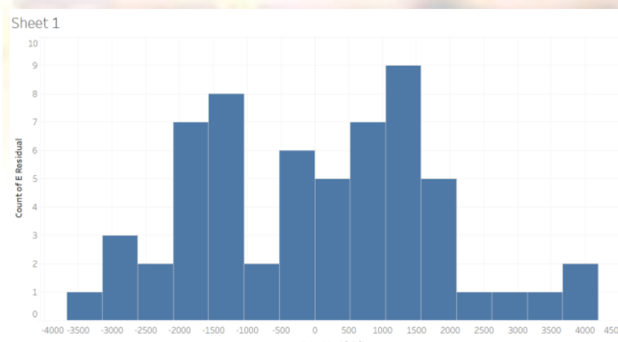
1. **MAPE**- The mean absolute error of the forecasting model, tested against the actual values, indicates the overall accuracy and variance of the prediction against the forecast.
2. **Time series of Residuals**- The time series of residuals were compared to identify the nature of the fit and outliers. If the MAPE of the validation data set is high due to the presence of a few outliers the seasons were separated from the data set.
3. **Histogram of Residuals**- Since the business objective is to improve the load factor negative residual values (under predicting) does not impact the final business objective. Thus if the model residuals are skewed to negative values the model performance is rated higher.
4. **Autocorrelations** - autocorrelation test (lag12) is done to check for presence of correlations. The absence of correlation indicates the model captured trend and seasonality satisfactorily.

Below table provides statistics for various forecast models across each city. (**Ensemble for each of the method is the average of other forecasting methods**)

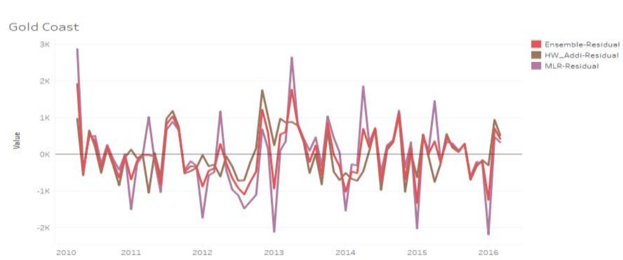
City	Forecast Model	Training Period	Training MAPE	Validation Period	Validation MAPE
Brisbane	Naïve (Lag-1)	42 Months	7.74%	18 Months	11.34%
	Multiple Linear Regression		2.00%		2.98%
	Holts Winter Multiplicative		0.20%		0.35%
	Ensemble		**2.74%**		**2.04%**
Canberra	Naïve (Lag-1)	42 Months	9.21%	18 Months	11.90%
	Multiple Linear Regression		4.80%		4.99%
	Holts Winter Multiplicative		3.52%		3.22%
	Holts Winter Additive		3.46%		3.14%
	Ensemble		**3.77%**		**3.75%**
Gold Coast	Naïve (Lag-1)	42 Months	9.47%	18 Months	16.35%
	Multiple Linear Regression		7.23%		10.35%
	Holts Winter Multiplicative		5.88%		3.84%
	Holts Winter Additive		**5.02%**		**5.17%**
	Ensemble		5.39%		4.34%
Melbourne	Naïve (Lag-1)	42 Months	5.91%	18 Months	6.53%
	Multiple Linear Regression		2.20%		2.60%
	Holts Winter Additive		2.40%		2.90%
	Ensemble		**2.12%**		**2.54%**
Perth	Naïve (Lag-1)	66 Months	7.29%	18 Months	9.37%
	Multiple Linear Regression		2.72%		3.01%
	Holts Winter No Trend		2.74%		2.88%
	Holts Winter Additive		2.98%		3.75%
	Ensemble		**2.66%**		**2.62%**
Sydney	Naïve (Lag-1)	42 Months	7.33%	18 Months	9.33%
	Multiple Linear Regression		1.58%		2.39%
	Holts Winter Multiplicative		0.15%		0.18%
	Ensemble		**2.01%**		**1.68%**



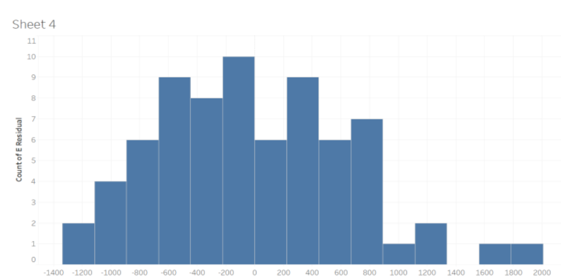
The trends of Ensemble-Residual, HW_Multi-Residual and MLR-Residual for Year-Month. Color shows details about Ensemble-Residual, HW_Multi-Residual and MLR-Residual.



The trend of count of E Residual for E Residual (bin).

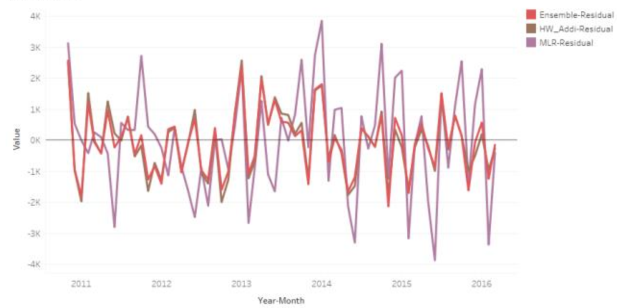


The trends of Ensemble-Residual, HW_Addi-Residual and MLR-Residual for Year-Month. Color shows details about Ensemble-Residual, HW_Addi-Residual and MLR-Residual.



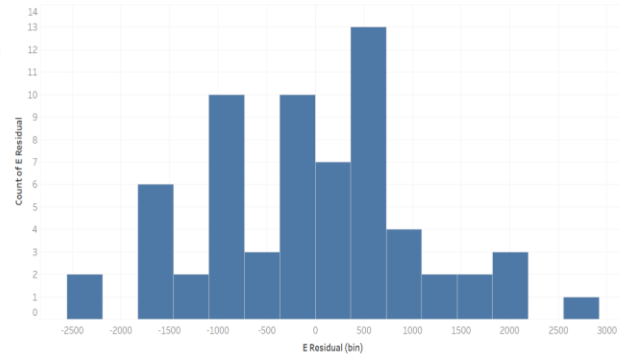
The trend of count of E Residual for E Residual (bin).

Gold Coast



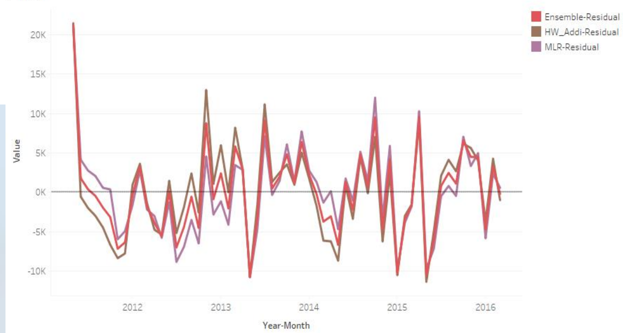
The trends of Ensemble-Residual, HW_Addi-Residual and MLR-Residual for Year-Month. Color shows details about Ensemble-Residual, HW_Addi-Residual and MLR-Residual.

Sheet 3



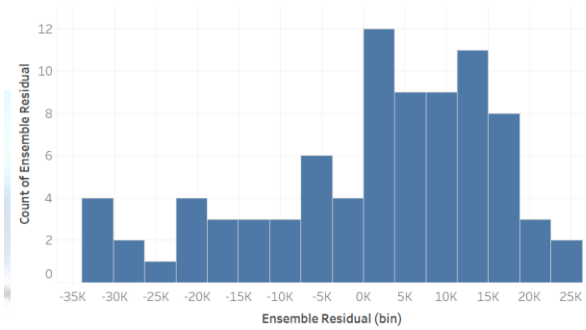
The trend of count of E Residual for E Residual (bin).

Perth



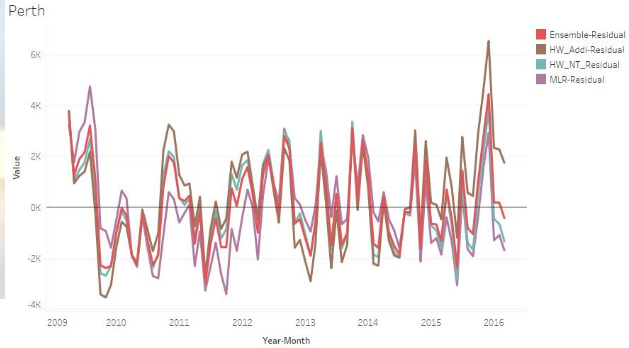
The trends of Ensemble-Residual, HW_Addi-Residual and MLR-Residual for Year-Month. Color shows details about Ensemble-Residual, HW_Addi-Residual and MLR-Residual.

Sheet 7



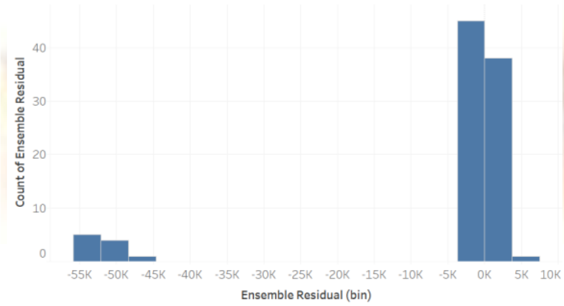
The trend of count of Ensemble Residual for Ensemble Residual (bin).

Perth



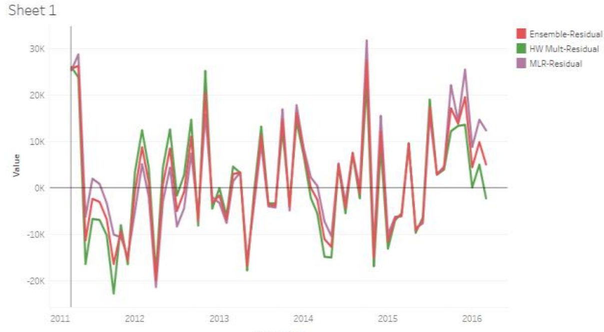
The trends of Ensemble-Residual, HW_Addi-Residual, HW_NT_Residual and MLR-Residual for Year-Month. Color shows details about Ensemble-Residual, HW_Addi-Residual, HW_NT_Residual and MLR-Residual.

Sheet 8



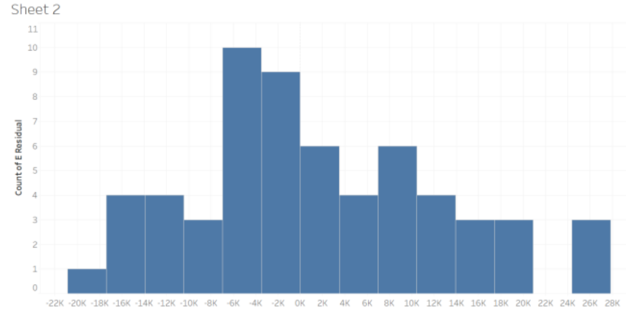
The trend of count of Ensemble Residual for Ensemble Residual (bin).

Sheet 1



The trends of Ensemble-Residual, HW Mult-Residual and MLR-Residual for Year-Month. Color shows details about Ensemble-Residual, HW Mult-Residual and MLR-Residual.

Sheet 2



The trend of count of E Residual for E Residual (bin).