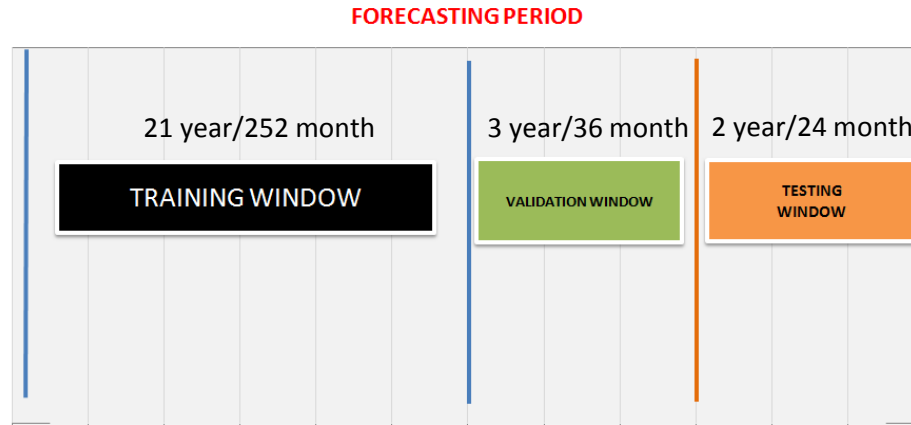**Forecasting overnight stay in country1 by Indian Tourist**
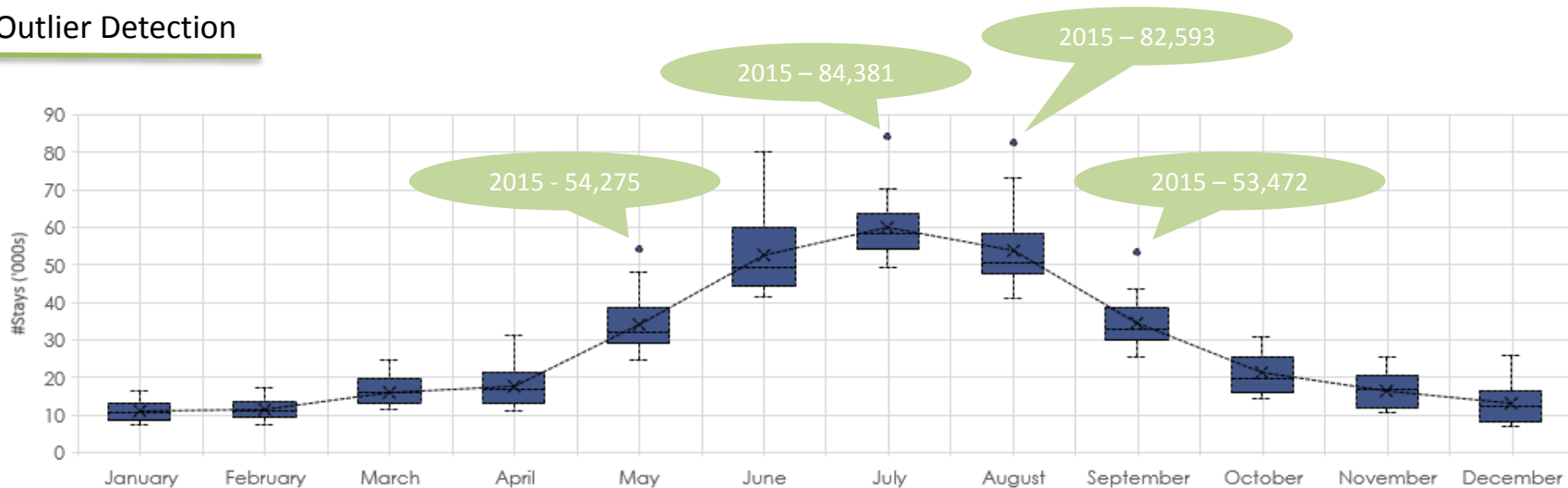
Data Summary

o 24 years of monthly data from 1992 to 2015 were considered

o No missing value.

o The 21 years of monthly data was taken as training set, next 3 years of monthly data as validation set, final 2 years as test period.

o 5 different methods were used and the best result was picked up for each instance.

o For deciding accuracy, Weighted Absolute Percentage Error (WAPE) was used as the metric. WAPE is defined as $\sum(|A-F|)/\sum A$, where A are actual values and F are the forecasted values.

FORECASTING PERIOD

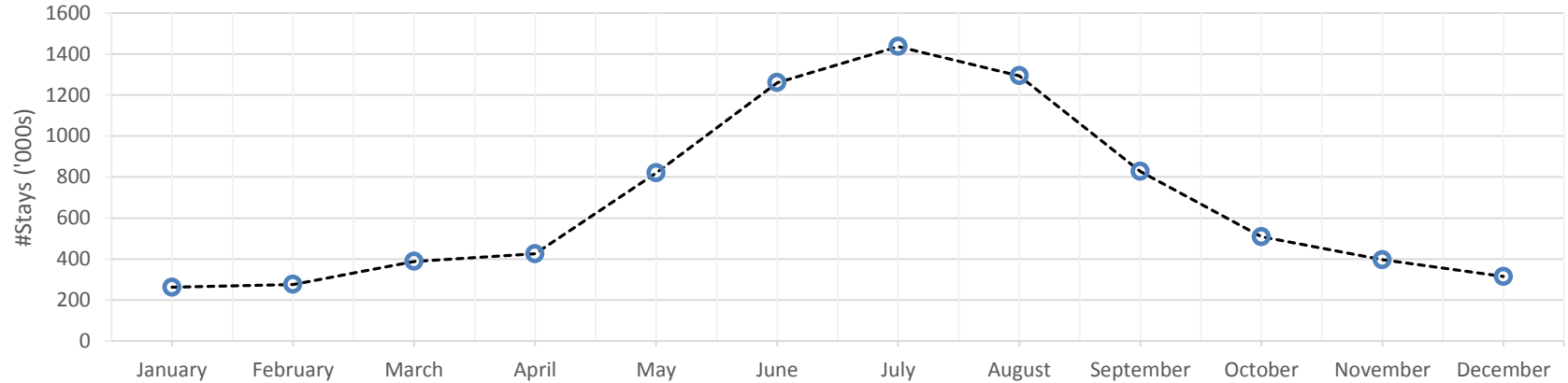| 21 year/252 month	3 year/36 month	2 year/24 month
TRAINING WINDOW	VALIDATION WINDOW	TESTING WINDOW

Data Pre-Processing

Outlier Detection

o Outliers of overnight stays is seen in 2015, reason could be high overnight stays due to some world sports event, conference, etc.
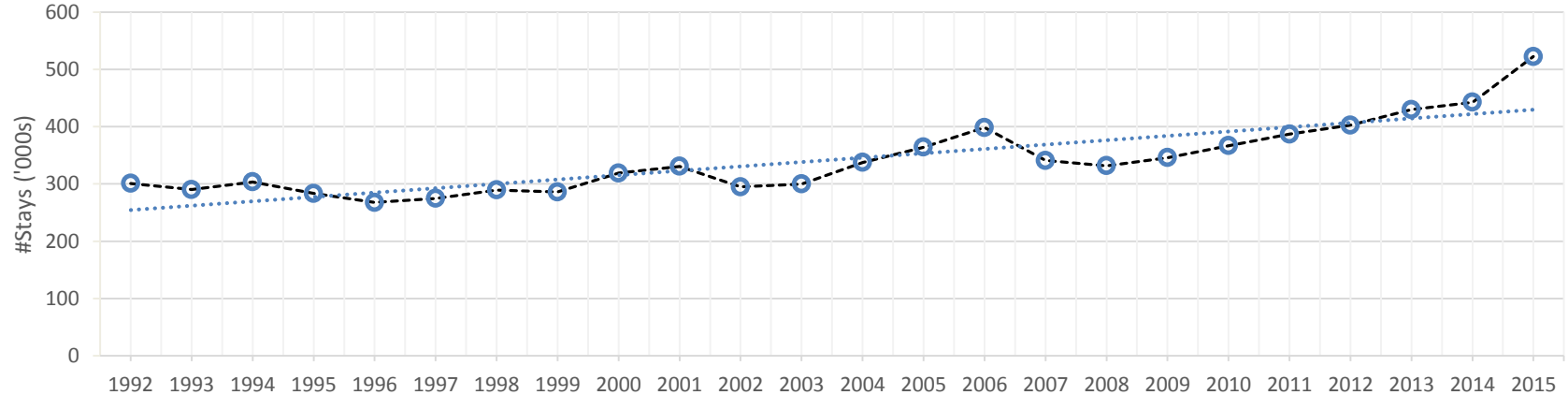
High Level Trends

Monthly Overnight Stay Trend

o   Looking at Monthly trend, #stay increase in June, July and August, reason could be in India, summer holidays starts from May to August.

High Level Trends

Yearly Overnight Stay Trend

- o Looking at Yearly trend, #stay drop in 2002 and again in 2008, reason could be India economy facing recession in 2002 and 2008.
- o Again, #stay boost from 2013 onwards.

Forecasting Methods

METHODS	DESCRIPTION
Moving Average (maf)	Mean of last 2 month values. Continually recomputed as new data becomes available, it progresses by dropping the earliest value and adding the latest value.
Simple Exponential Smoothing (ses)	Similar to weighted moving average with exponentially decreasing weights.
Theta	Simple exponential smoothing with drift
Holt-Winters Additive (hwadd) and Holt-Winters Multiplicative (hwmult)	Forecasting seasonal time series. The additive method is preferred when the seasonal variations are roughly constant through the series. The multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

Methodology

o Training forecasting methods on 21 years of monthly data (Training Period) and then using it to predict the last 3 years of monthly data (Validation Period).

o All the methods are used to forecast the #stay in validation window, the method which gives the least error is selected as best method.

o Best method is used to forecast the year 2016 & 2017 (Testing Period)

Error at Validation Period

Prediction Period	Method	WAPE	BIAS
3 year/36 month	Simple Exponential Smoothing with drift	10.7%	-1.5%
3 year/36 month	Holt-Winters Additive	14.4%	-12.7%
3 year/36 month	2 Month Moving Average	14.8%	-14.6%
3 year/36 month	Holt-Winters Multiplicative	22.7%	-22.7%
3 year/36 month	Simple Exponential Smoothing	48.7%	-22.7%

WAPE Calculation
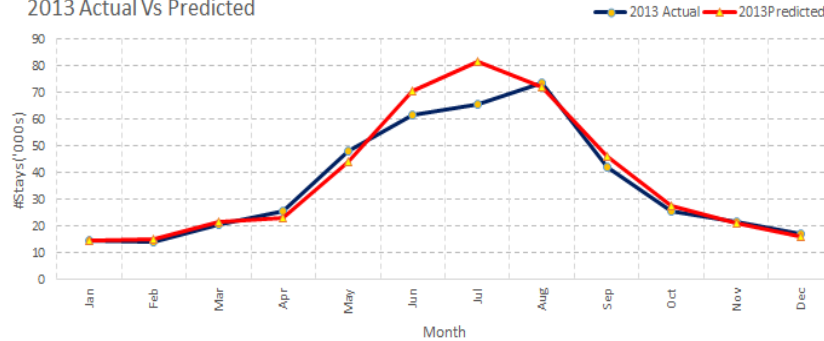
$$WAPE = \frac{\sum(ABS(Actual - Forecast))}{\sum Actual\ Value} \times 100$$

BIAS Calculation

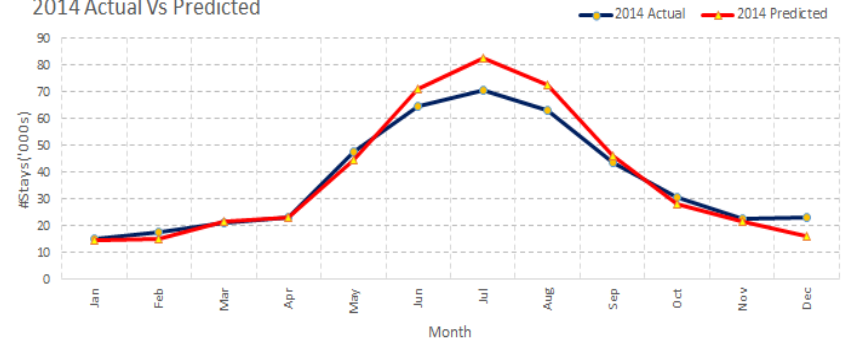$$BIAS = \frac{\sum(Actual - Forecast)}{\sum Actual\ Value} \times 100$$

Theta method (simple exponential smoothing with drift) was picked up as the best method in validation period with WAPE as 10.7% and BIAS as -1.5%.

Actual vs. Predicted – Validation Window

2013 Actual Vs Predicted

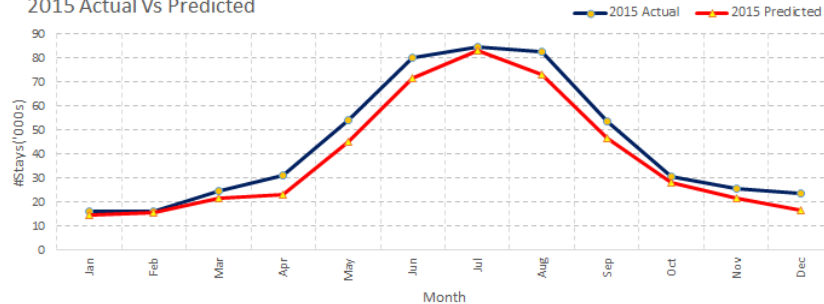2014 Actual Vs Predicted

2015 Actual Vs Predicted

Chart above depicts the last 3 years of monthly data actual vs. predicted in validation window, with a WAPE of 10.7% and BIAS of -1.5% using our best forecast method selection, Theta was picked up as the best method.

NOTE: Mentioned WAPE and BIAS is calculated over all 3 years

Insights

- #Stay increasing constantly during May to Sep, especially at Jun, Jul, Aug dominating the series.
- Smaller #stay observer during the Winter period.
- To drill down further, we can check with external parameters like weather.

APPENDIX

India Economic Growth
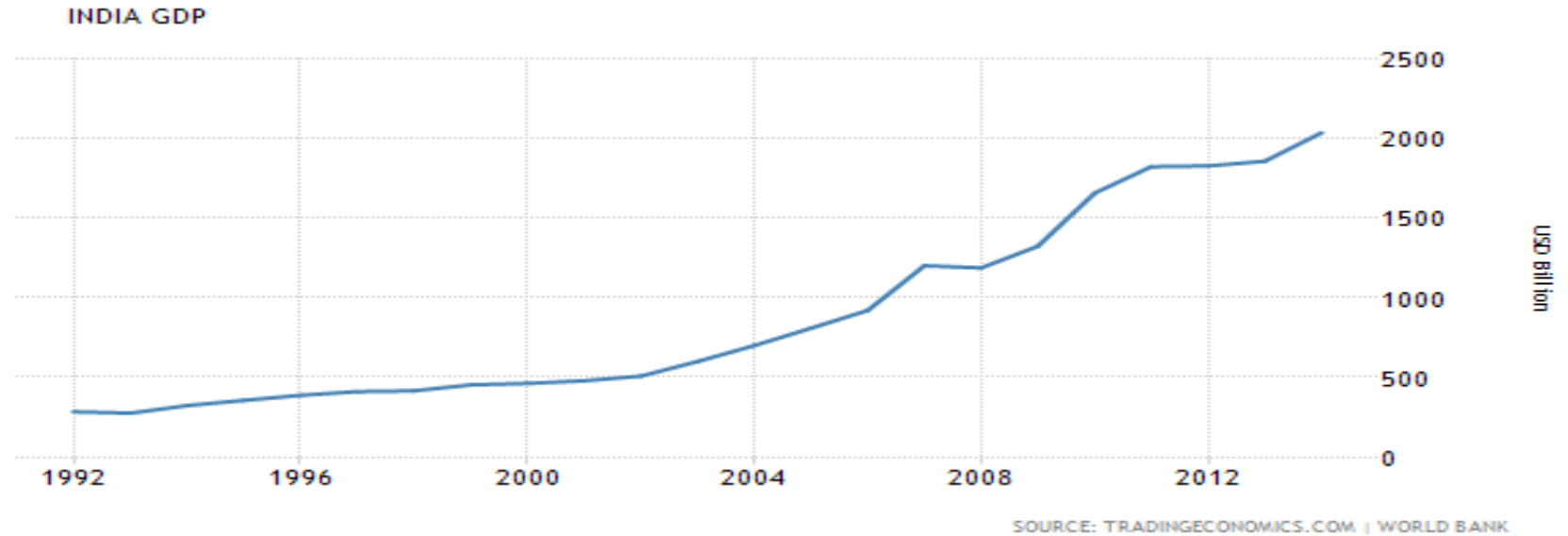
o   Above graph depict India Economic growth over the year 1992 to 2017