Forecast sales of Walmart departments to allocate marketing budget

Team A3	
Gandharv Paliwal	61710328
Asmita Sonik	61710866
Aayam Ankan	61710631
Abhinav Mani	61710499
Swagnik Chatterjee	61710076
Samta Jhingan	61710109

Business Problem: Due to intense competition from E-commerce companies like Amazon, sales from Walmart's brick and mortar stores have been declining steadily. There is a need for Walmart to increase sales through its stores, and also boost repeat customer purchases.

Forecasting Problem: Walmart operates through multiple departments like Furniture, Electronics, Clothing etc. Aim is to forecast sales of different departments in order to correctly allocate marketing budget to different departments for the coming month.

Underlying Assumption: Departments expected to have greater sales in the coming month will perform better with higher marketing budget (sales promotions & advertisements)

Forecasting Horizon: Next 2 months (roll forward forecasting)

Available Data: Feb' 2014-Oct'2016

Data: Total 99 departments, 45 stores

Store	Dept.	Date	Weekly_Sales	IsHoliday
1	1	2/5/2014	24924.5	FALSE
1	1	2/12/2014	46039.49	TRUE
1	1	2/19/2014	41595.55	FALSE
1	1	2/26/2014	19403.54	FALSE
1	1	3/5/2014	21827.9	FALSE

Total : 33 data points
Data partition : Training (24), Validation (9)

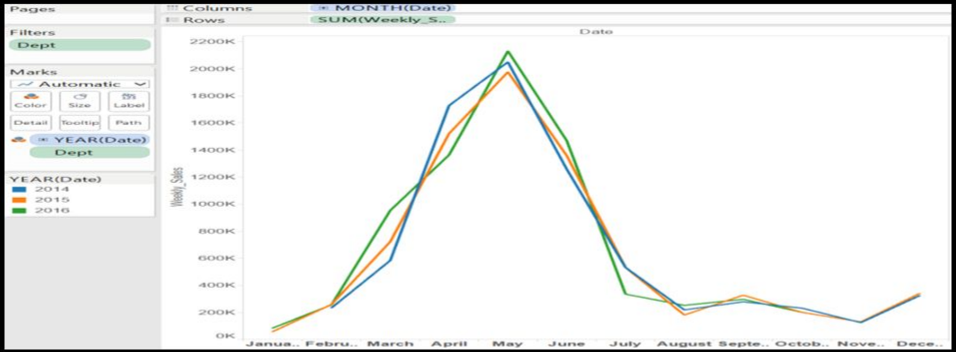
1. Data Plotting
2. Random Walk
3. Non-Seasonal & Seasonal Naïve Forecasting
4. Moving Average of Seasonalized Data
5. Holt's/ Holt's Winter/ Exponential Smoothening ARIMA
6. Multiple Linear Regression (Additive , Multiplicative Seasonality) ARIMA

General Strategy Pursued for Different Datasets

Department #56

Compnents

Trend	Constant
Seasonality	Annual
Level	2M-300K
Noise	Always there

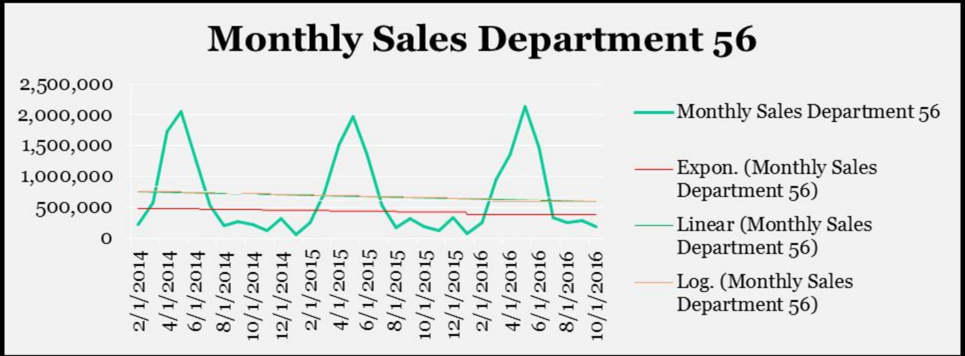


RANDOM WALK

In the ARIMA model AR1 coefficient is 3xStErr away from 1

ARIMA Model

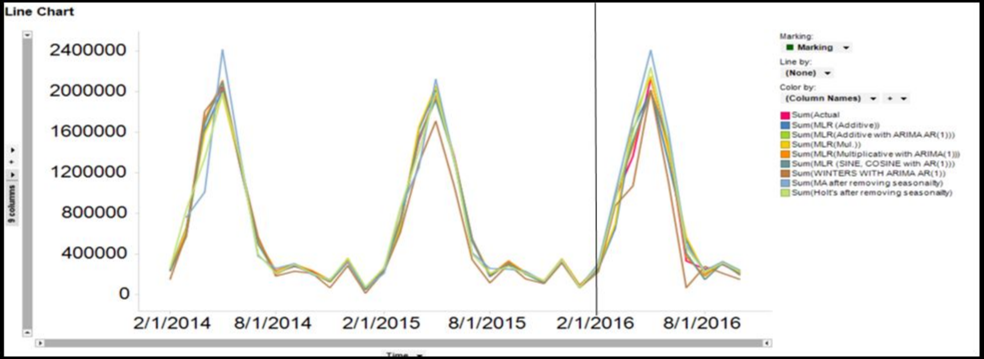
ARIMA	Coeff	StErr	p-value
Const. tern	221442.8	36157.36	9.1016E-10
AR1	0.675383	0.124015	5.15258E-08



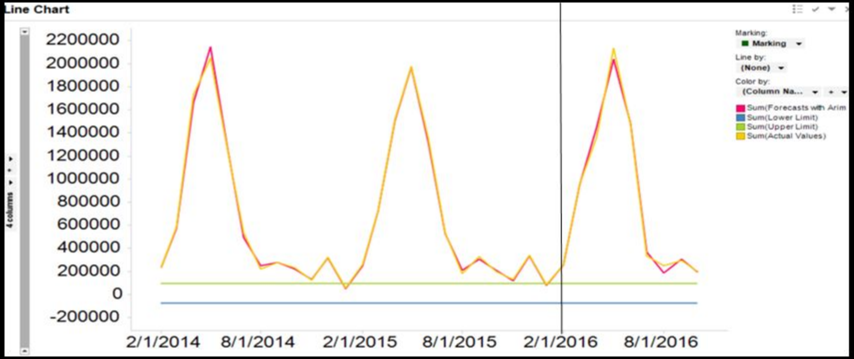
Data Plotting → **Random Walk** → **Non-Seasonal & Seasonal Naïve Forecasting** → **Moving Average of Seasonalized Data** → **Holt's/ Holt's Winter/ Exponential Smoothening ARIMA** → **Multiple Linear Regression (Additive , Multiplicative Seasonality) ARIMA**

MODELING OUTPUTS

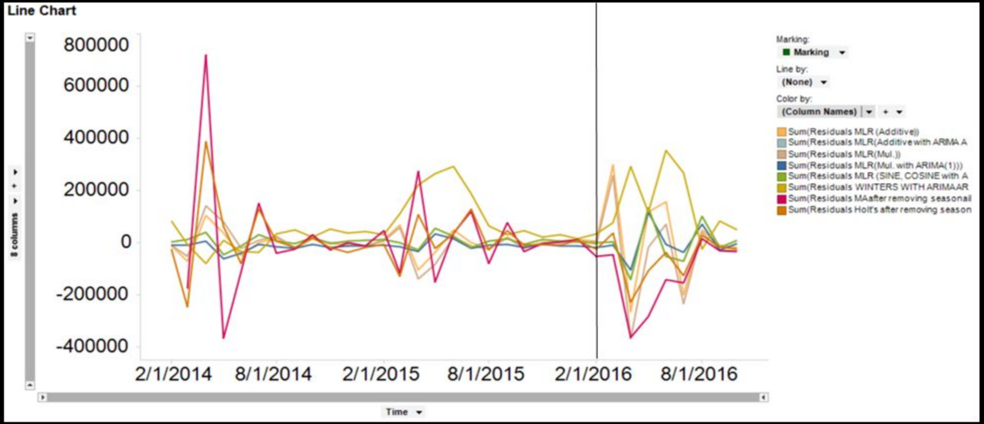
METHODS	Training Error	Validation Error	
NAÏVE SEASONAL	12.08%	16.59%	
NAÏVE NON- SEASONAL	99.13%	76.24%	**Overfitting**
MOVING AVERAGE OF DE-SEASONALIZED DATA	17.11%	17.04%	Seasonality index to de seasonalize data, which was then fed to the model. Forecasts were re seasonalized for result analysis
HOLT'S ON DE-SEASONALIZED DATA	12.16%	11.37%	Default Alpha(level)=0.2, Beta (Trend) =0.15 (Best Output)
WINTERS METHOD WITH ARIMA	18.9%	23.955%	Default Alpha(level)=0.2, Beta (Trend) =0.15, Gamma(Seasonality)=0.05 (Best output values)
MULTIPLE LINEAR REGRESSION WITH ADDITIVE SEASONALITY	6.29%	17.83%	Input variables: 11 Monthly dummy variables, no. of holiday weeks in the month, t Output variables: Monthly Sales
MULTIPLE LINEAR REGRESSION WITH ADDITIVE SEASONALITY & ARIMA	3.78%	10.47% (ARIMA 1)	
MULTIPLE LINEAR REGRESSION WITH SINE AND COSINE FUNCTIONS & ARIMA	3.78%	10.47%	Input variables: 11 Monthly dummy variables, no. of holiday weeks , sin(2∏t/12), cosine(2∏t/12) Output variables: Monthly Sales
MULTIPLE LINEAR REGRESSION WITH MULTIPLICATIVE SEASONALITY	5.988%	19.042%	Input variables: 11 Monthly dummy variables, no. of holiday weeks in the month, t Output variables: ln(Monthly Sales)
MULTIPLE LINEAR REGRESSION WITH MULTIPLICATIVE SEASONALITY & ARIMA	4.96%	7.68% (ARIMA 1)	



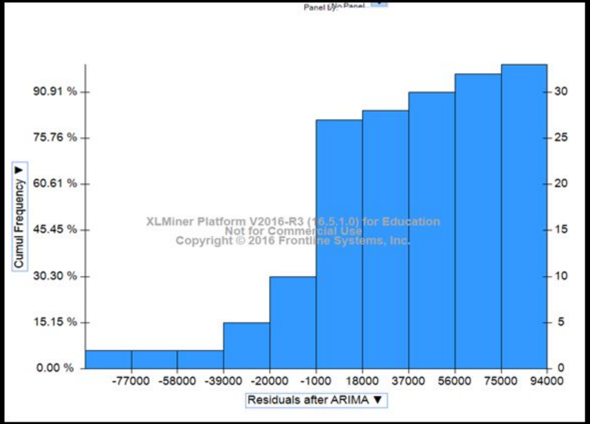
Actual & Forecasts of Different Models Vs. Time



Forecast from selected Model

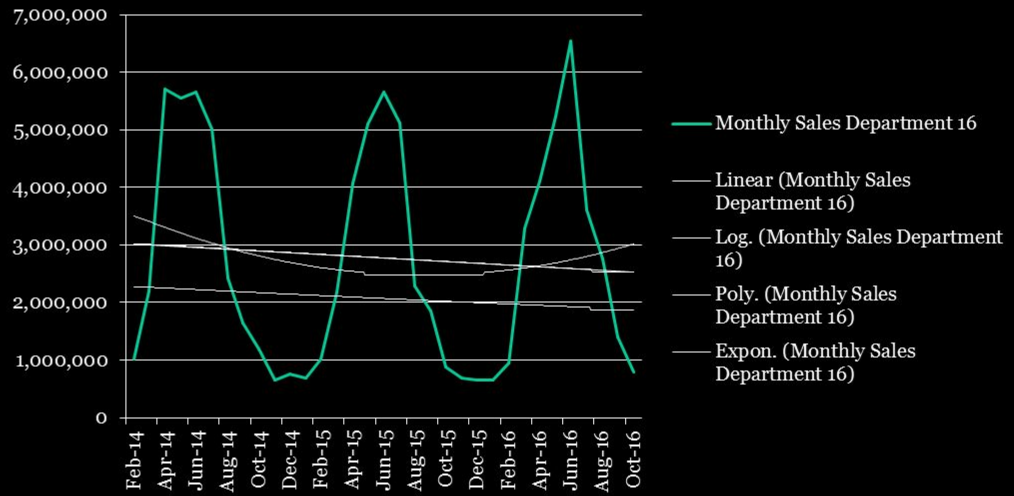


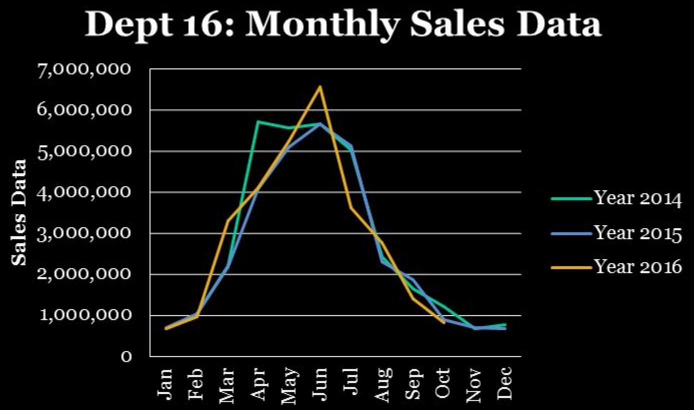
Residuals of Different Models Vs. Time



Histogram of Residuals

Department #16





Random walk

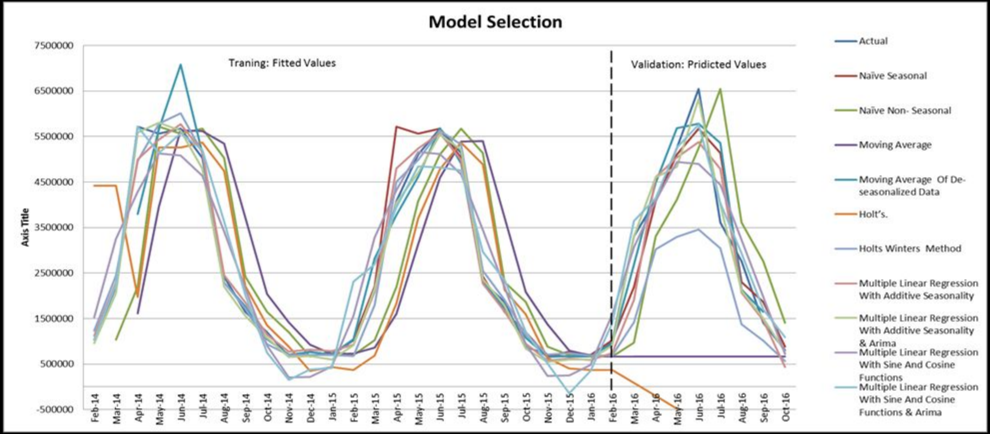
ARIMA	Coeff	StErr	p-value
Const. term	679526.3	82701.39	2.09309E-16
AR1	0.755381	0.110888	9.61858E-12

Naive -seasonal and non seasonal

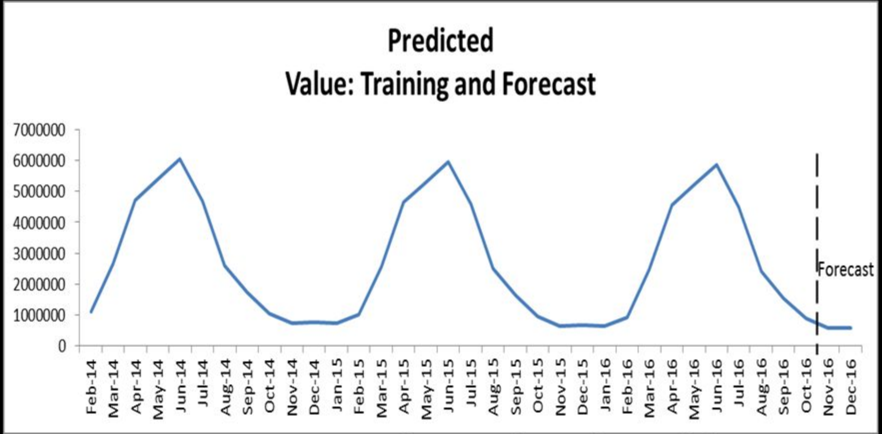
	MAPE	RMSE
Training	10.93%	507603.7
Validation	17.63%	730825.6

Modeling Outputs

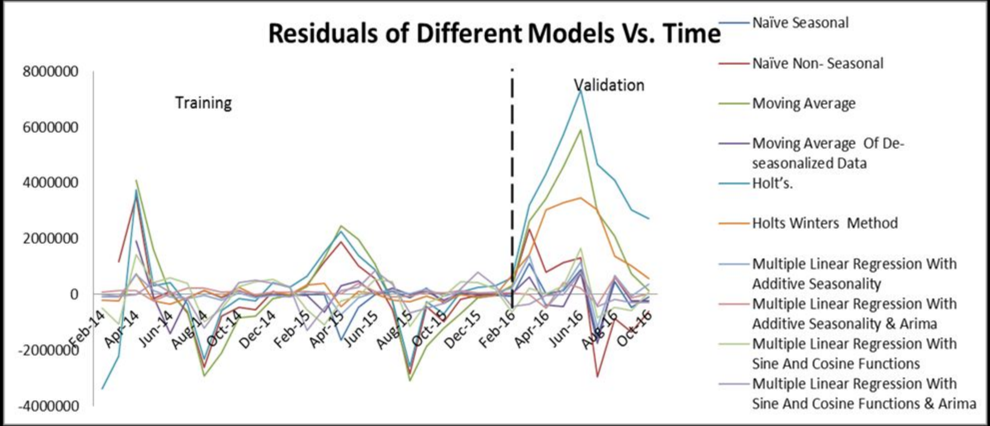
METHODS	Training Error	Validation Error	Comments
Naïve Seasonal	10.93%	17.63%	
Naïve Non- Seasonal	38.80%	49.54%	Overfitting
Moving Average	57.60%	66.57%	
Moving Average Of De-seasonalized Data	8.65%	18.64%	Seasonality index to de seasonalize data, which was then fed to the model. Forecasts were re seasonalized for result analysis
Holt's.	54.98%	146.06%	Default Alpha
Holts Winters Method	9.62%	35.56%	Default Alpha, Beta, gamma
Multiple Linear Regression With Additive Seasonality	7.48%	22.98%	Input variables: 11 Monthly dummy variables, no. of holiday weeks in the month, Output variables: Monthly Sales
Multiple Linear Regression With Additive Seasonality & Arima	6.51%	8.30%	
Multiple Linear Regression With Sine And Cosine Functions	32.41%	21.75%	Input variables: 11 Monthly dummy variables, no. of holiday weeks , sin(2∏t/12), cosine(2∏t/12)
Multiple Linear Regression With Sine And Cosine Functions & Arima	31.65%	15.83%	



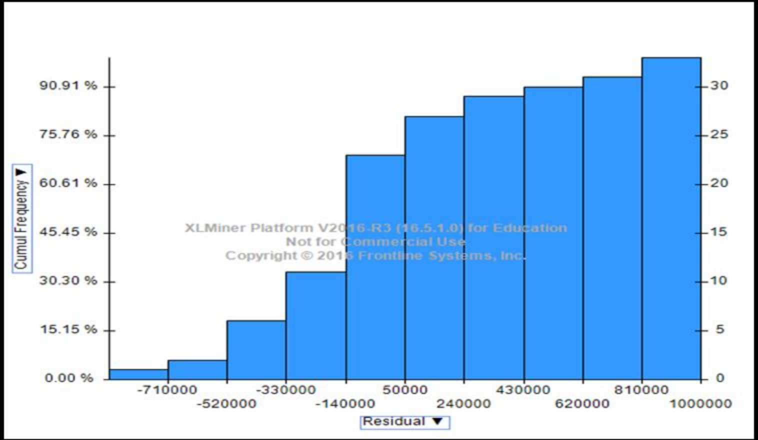
Actual & Forecasts of Different Models Vs. Time



Forecast from selected Model

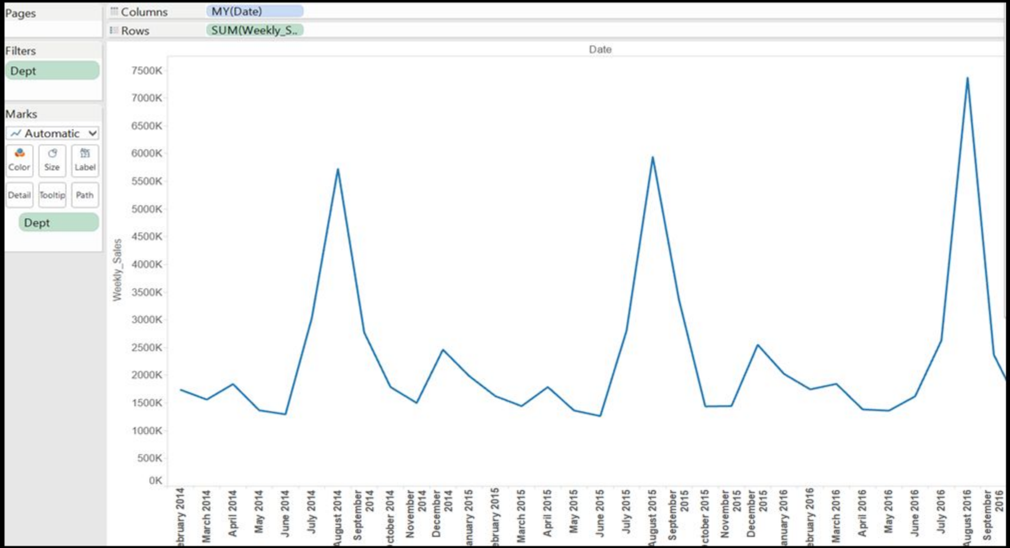


Residuals of Different Models Vs. Time



Histogram of Residuals

Department 3



Compnents	
Trend	Constant
Seasonality	Annual
Level	7.5M-1.5K
Noise	Always there

RANDOM WALK CHECK

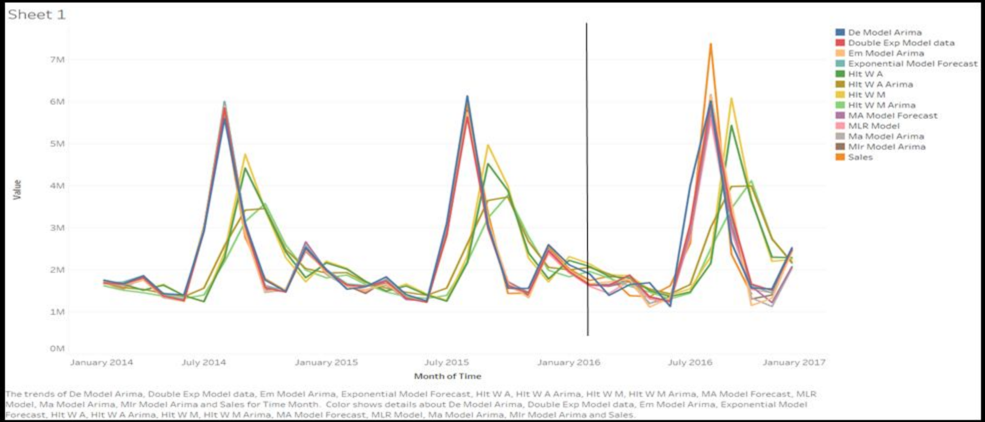
In the ARIMA model AR1 coefficient is approx. 5xStErr away from 1. Hence not a random walk.

ARIMA Model

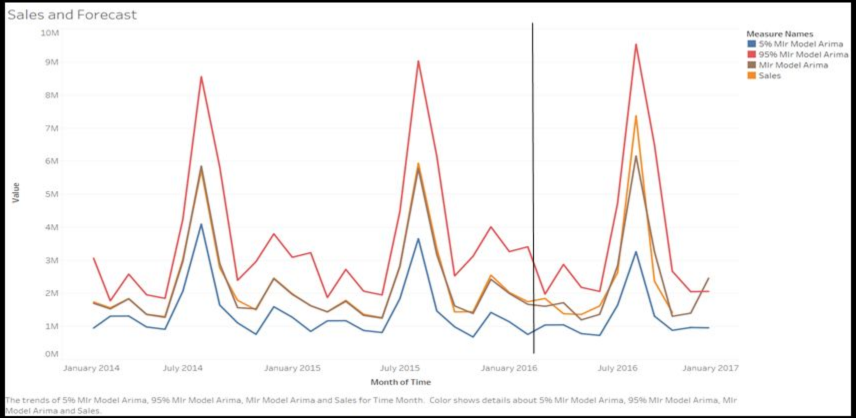
ARIMA	Coeff	StErr	p-value
Const. terr	1712824	182529.4	6.36E-21
AR1	0.25522	0.166617	0.125577

MODELING OUTPUTS – Department 3

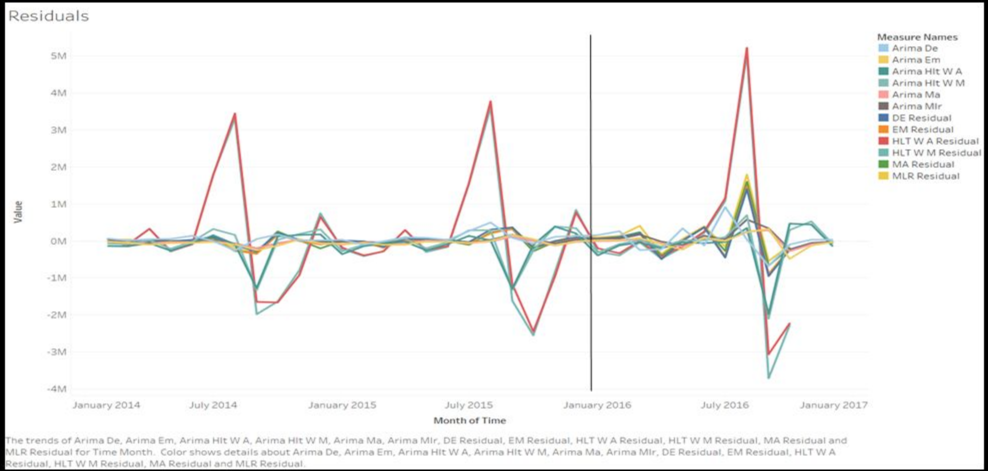
METHODS	Training Error	Validation Error	
NAÏVE SEASONAL	8%	13%	
NAÏVE NON- SEASONAL	39%	44%	**Overfitting**
MOVING AVERAGE OF DE-SEASONALIZED DATA	5.446%	12.489%	Seasonality index to de seasonalize data, which was then fed to the model. Forecasts were re seasonalized for result analysis
EXPONENTIAL METHOD ON DE-SEASONALIZED DATA	4.713%	14.190%	
HOLT'S ON DE-SEASONALIZED DATA	4.627%	14.676%	Default Alpha(level)=0.2, Beta (Trend) =0.15 (Best Output)
WINTERS METHOD WITH ADDITIVE SEASONALITY AND ARIMA	31.995%	43.318%	Default Alpha(level)=0.2, Beta (Trend) =0.15, Gamma(Seasonality)=0.5 (Best output values)
WINTERS METHOD WITH MULTIPLICATIVE SEASONALITY AND ARIMA	33.053%	44.037%	Default Alpha(level)=0.2, Beta (Trend) =0.15, Gamma(Seasonality)=0.5 (Best output values)
MULTIPLE LINEAR REGRESSION WITH MULTIPLICATIVE SEASONALITY	3.062%	13.095%	Input variables: 11 Monthly dummy variables, no. of holiday weeks in the month, t Output variables: ln(Monthly Sales)
MULTIPLE LINEAR REGRESSION WITH MULTIPLICATIVE SEASONALITY & ARIMA	2.824%	12.438%	



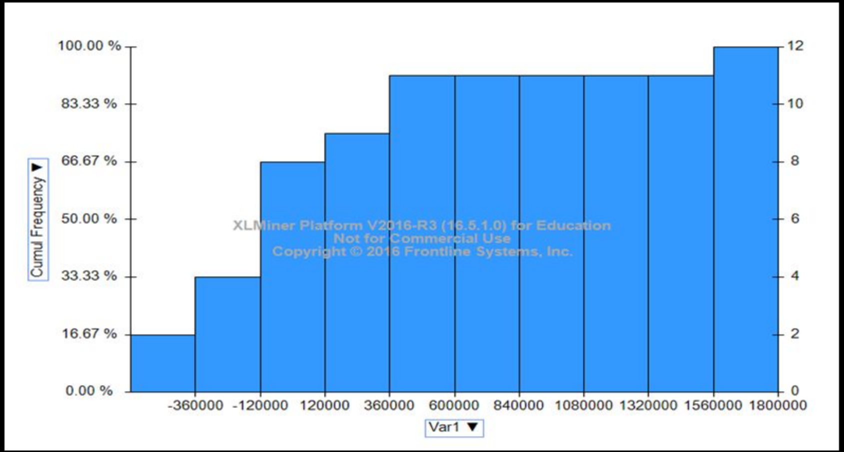
Actual & Forecasts of Different Models Vs. Time



Forecast from selected Model



Residuals of Different Models Vs. Time



Histogram of Residuals

Department No. 67

February 2014 to October 2016

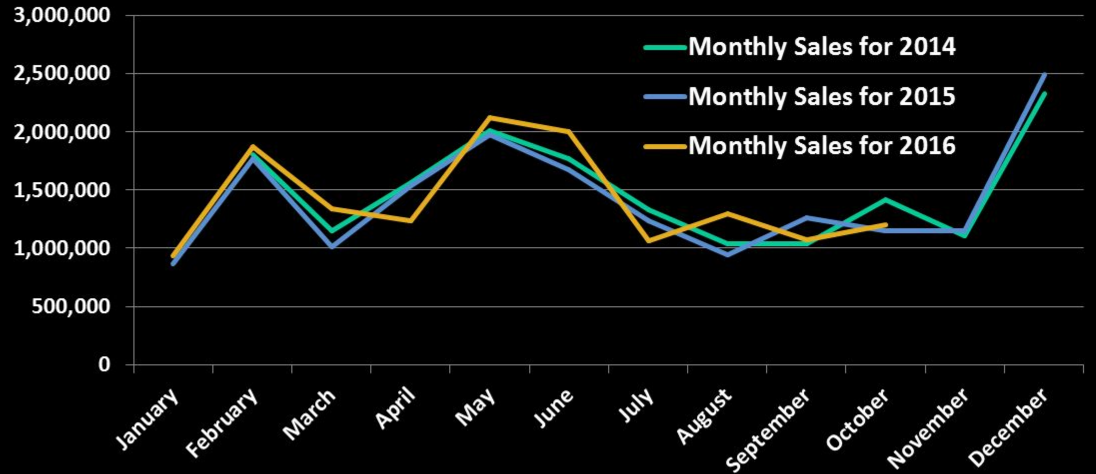
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Monthly Sales for 2014		1,810,771	1,151,855	1,562,939	2,016,670	1,770,089	1,331,383	1,042,685	1,042,464	1,419,963	1,110,677	2,329,182
Monthly Sales for 2015	871,254	1,771,180	1,020,878	1,539,813	1,976,636	1,677,350	1,237,082	951,704	1,264,811	1,152,051	1,156,677	2,495,602
Monthly Sales for 2016	941,779	1,878,468	1,344,016	1,239,628	2,124,512	2,001,654	1,071,933	1,296,984	1,075,257	1,202,830		

Random Walk?

Coefficient of AR1 = 0.029. This is 7*StdErr away from 1.

ARIMA Model

ARIMA	Coeff	StErr	p-value
Const. term	1751587.651	90131.38	4E-84
AR1	-0.207214989	0.170402	0.22397



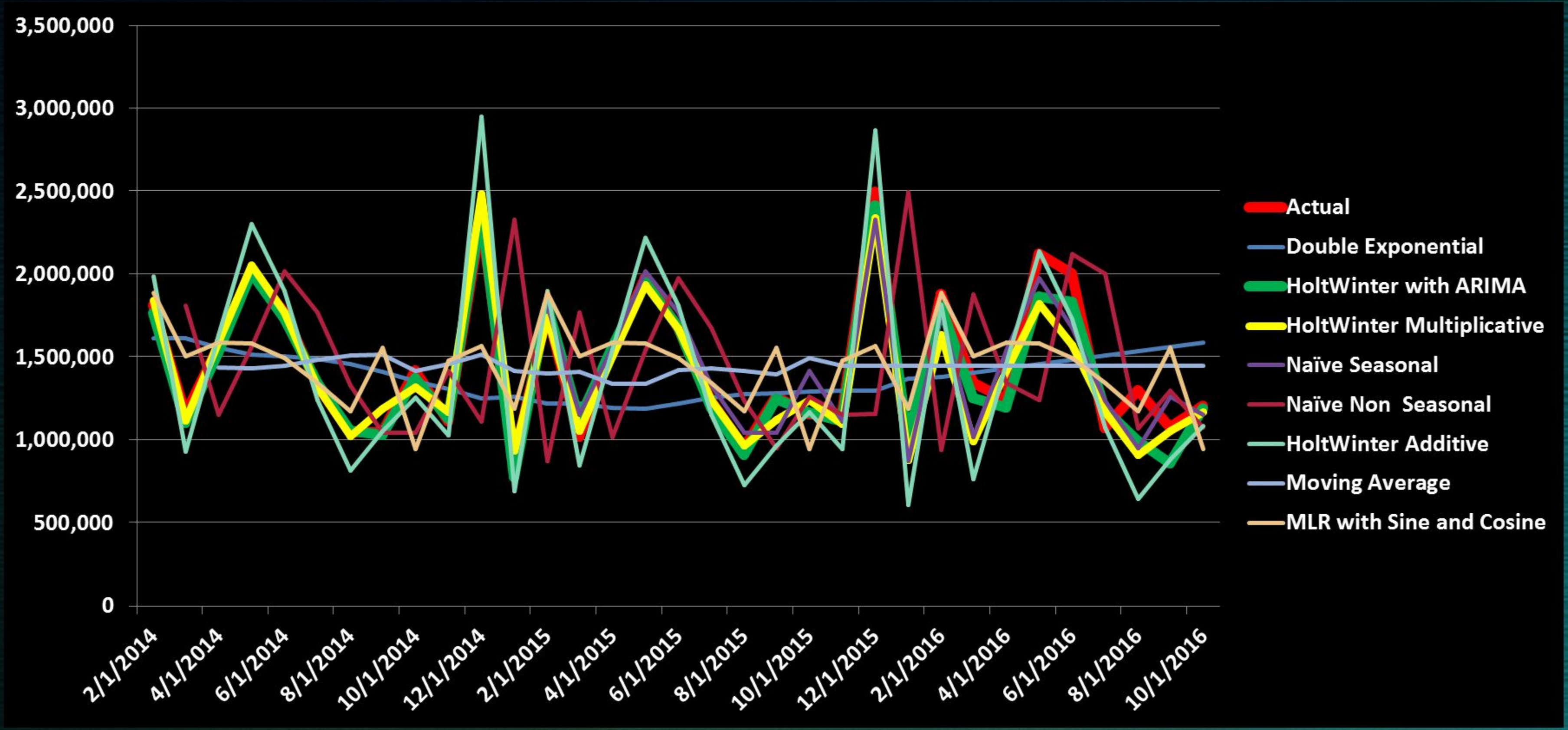
Components

Seasonality	Annual
Trend	Constant Trend
Level	1 mn- 2.5 mn
Noise	Present

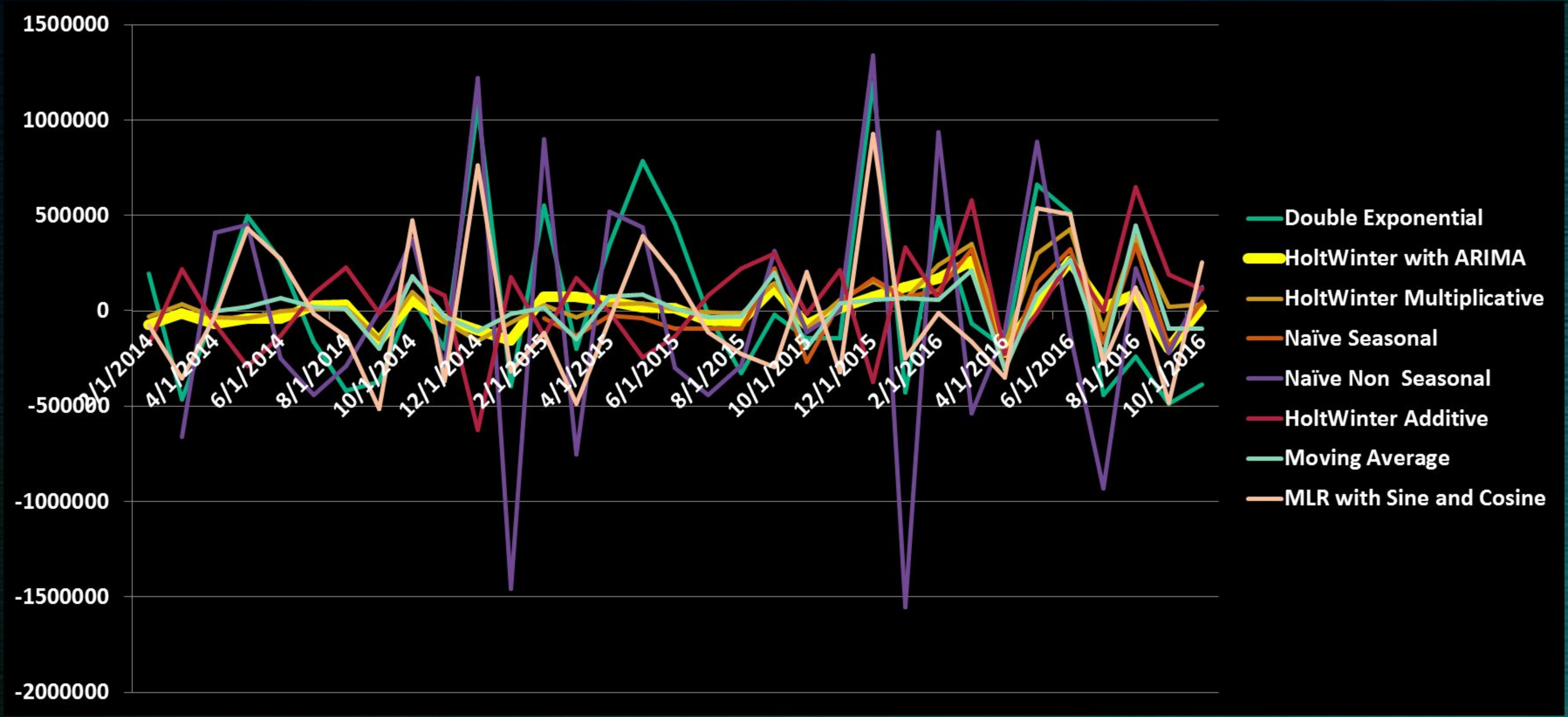
Modeling Output

Methods	Training Error	Validation Error	Other Inputs
Double Exponential	24.514	26.768	Default Alpha(level)=0.2, Beta (Trend) =0.15
HoltWinter Multiplicative	4.194	14.757	Period = 12
Naïve Seasonal	8.356	15.667	
Naïve Non-Seasonal	39.154		
HoltWinter Additive	13.312	17.354	Default values of Alpha(level)=0.2, Beta (Trend) =0.15, Gamma(Seasonality)=0.05 (Best output values)
Moving Average	5.414	11.03	Seasonality index to de seasonalize data, which was then fed to the model. Forecasts were re seasonalized for result analysis
HoltWinter Multiplicative with ARIMA	**3.016**	**10.024**	**Default Alpha(level)=0.2, Beta (Trend) =0.15, Gamma(Seasonality)=0.05 (Best output values)** **Period = 12** **ARIMA = 3**
MLR with Sine and Cosine	21.862	21.367	Input variables: 11 Monthly dummy variables, no. of holiday weeks , sin(2∏t/12), cosine(2∏t/12) Output variables: Monthly Sales

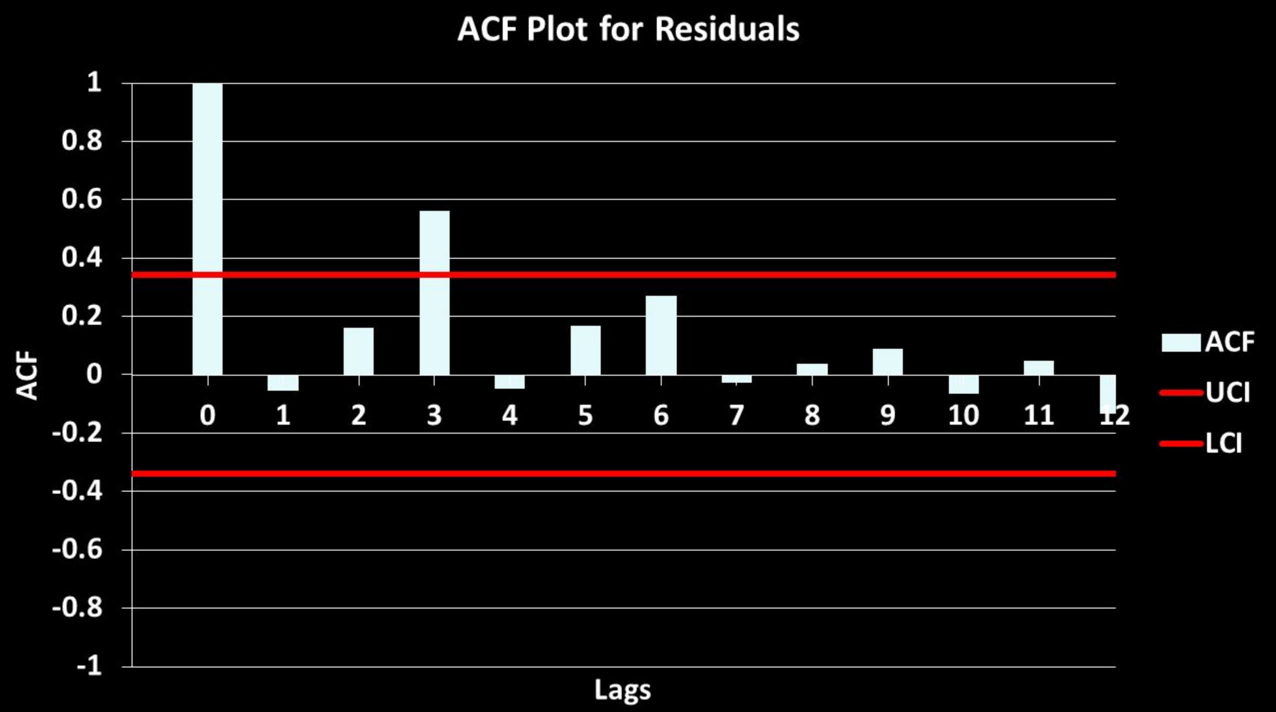
Forecasts of Different Models and Actual Values versus Time



Residuals of Different Models versus Time

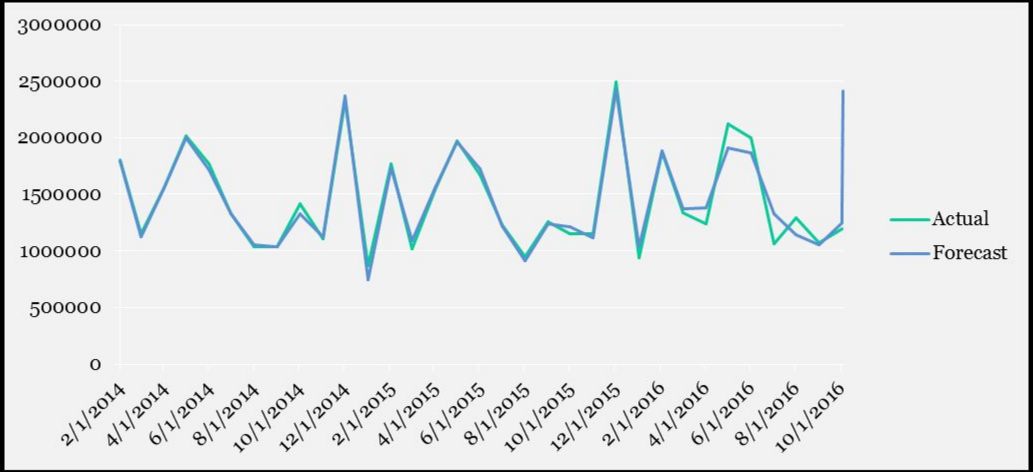


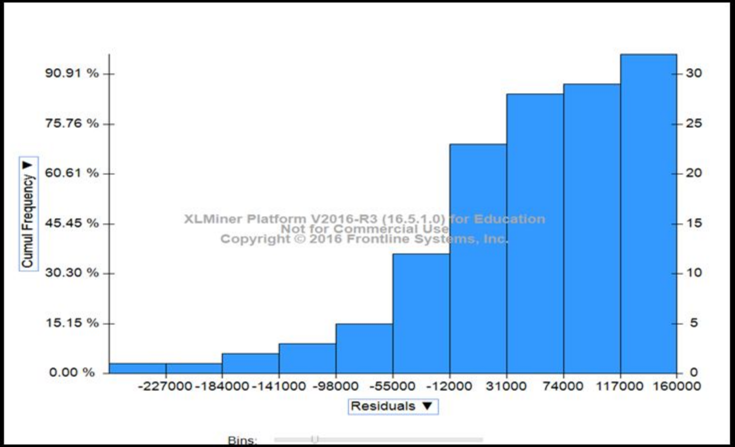
Training MAPE	3.016
Validation MAPE	10.024



Chosen Model: HoltWinter Multiplicative ARIMA output

Actual Vs. Forecasted value





Summary

- The modelling can be extended to all the departments which are not having random walk and budget can be allocated based on predicted sales

- In order to reduce complexity, modelling can be done only for the top selling departments and budget can be allocated based on predictions. For the other departments a fixed budget can be allocated.

- Possibility of developing an interface to generate forecasts for all the departments, along with optimal marketing budget allocations.

- Possibility of forecasting using econometric models especially the impact of competitive pricing from major competitors like Amazon.