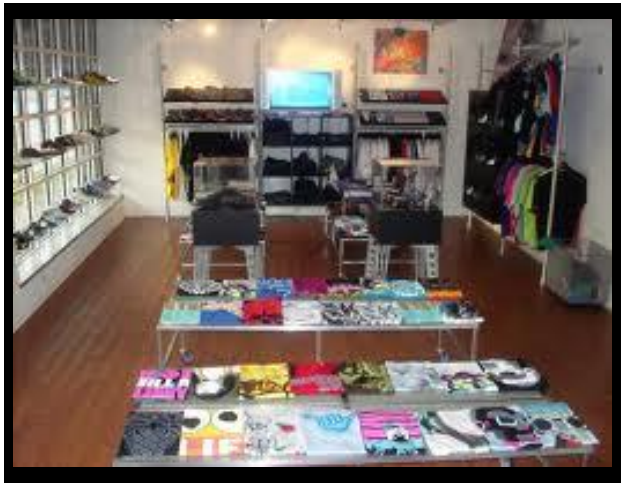
Forecasting sales for the TOP 5 selling SKUs

Term 5: Forecasting Analytics

Presented To: Prof. Galit Shmueli and Prof. Mayukh Dass

Presented By:

- Arka Sarkar
- Kushal Paliwal
- Malvika Gaur
- Shwaitang Singh



Business Goal

- **Business Driver**: To predict sales (units sold)
  - Predict volatility in earnings
  - Protect against stock outs
  - Better promotions

- Identify the **top 5 selling SKUs** at the retail store

- Forecast ***daily sales*** for the top 5 selling SKUs over the ***next 1 week*** (i.e. the first week of August 2012)

Business Goal

- **Top 5 selling (in terms of revenues) SKUs in the Year 2012**
- **Represent approximately *2% of the total revenues***
- **Total number of SKUs sold in the store: *10,493***

INR 0.9 million

INR 0.8 million

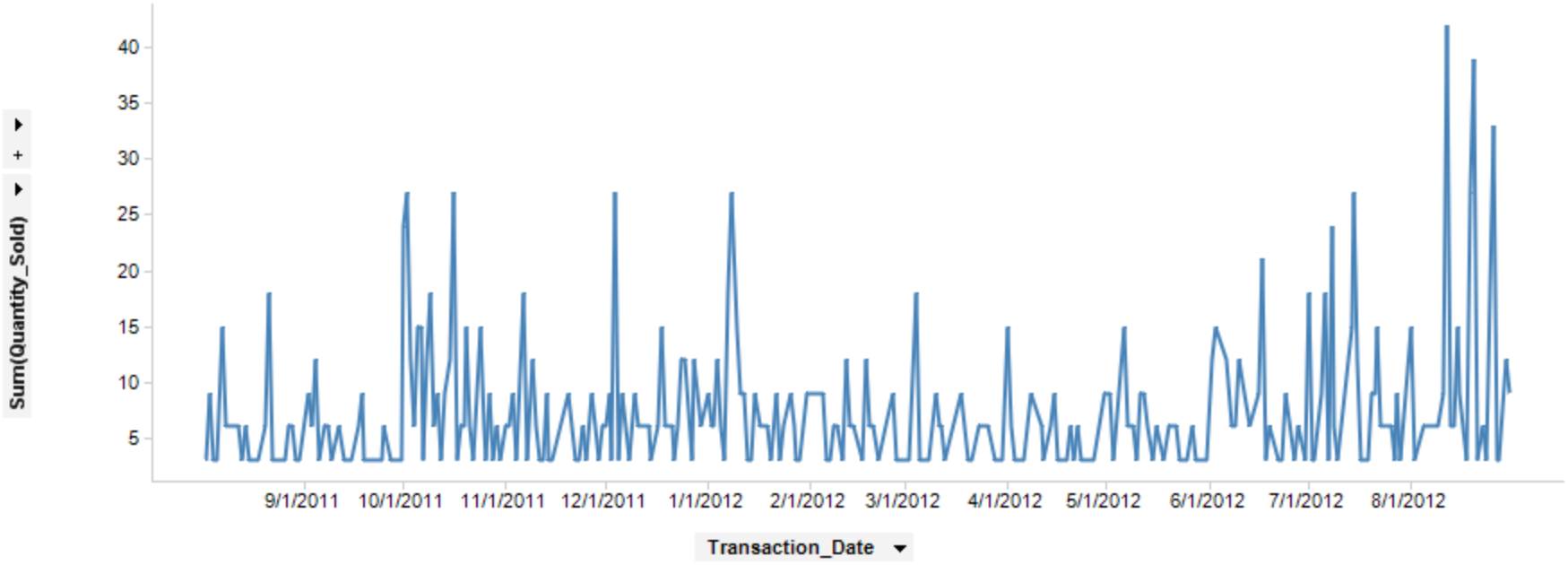
INR 0.78 million

INR 0.58 million

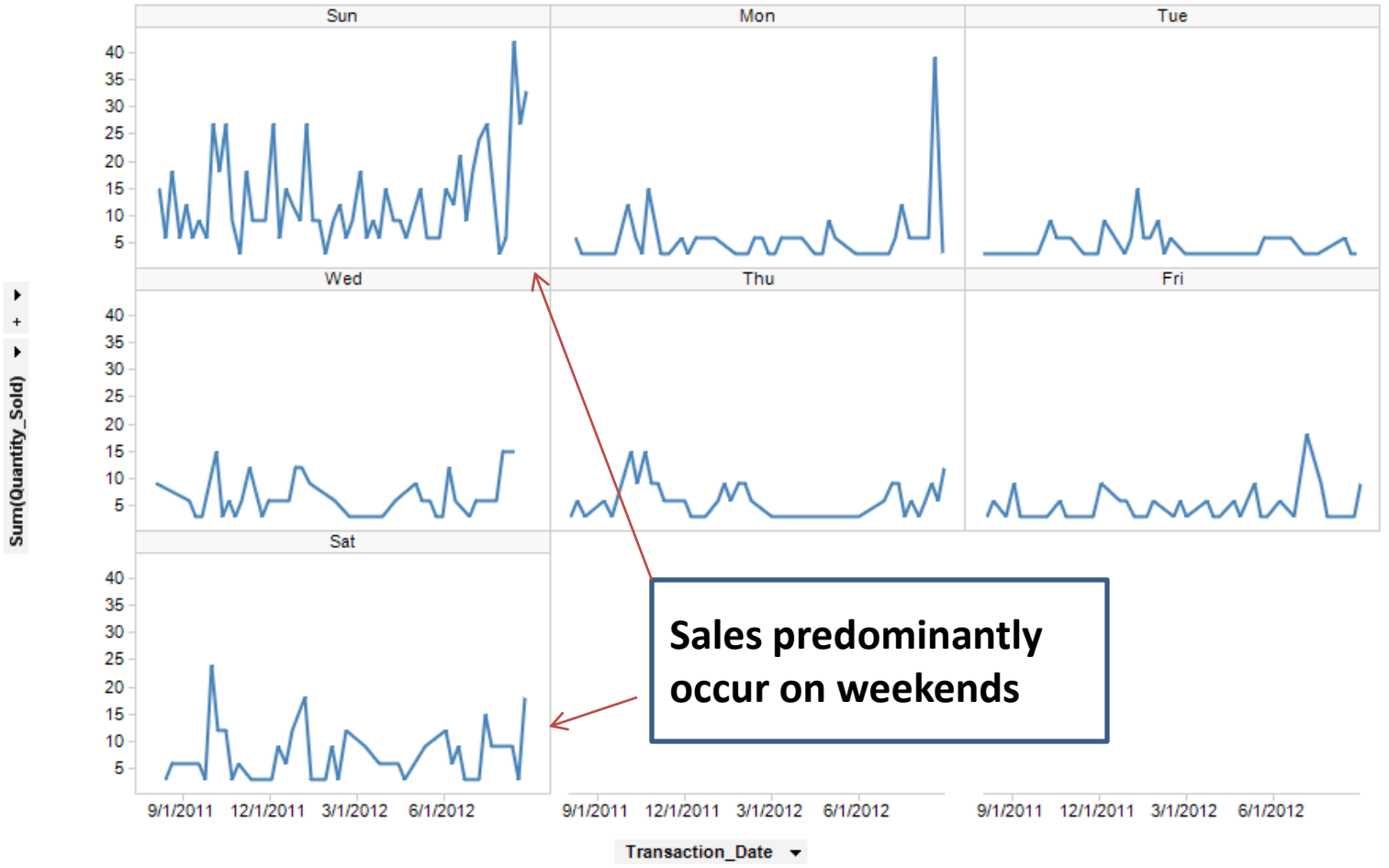
INR 0.53 million

Visualizing the data

SKU: 100004925
Total Sales (2011 & 2012): INR 1.4 Million
Sells 10 times more than 2 ltr. Jar, and 5 times more than 1 ltr. Pouch



Visualizing the data



Initial Analysis: Peaks and Outliers?

Preprocessing: Possible Explanation

Average Quantity bought per day: 7.35
Most occurring purchase size: 3 units
Although not in this case, we've found a 'bulk buyer' who shops sporadically for other SKUs

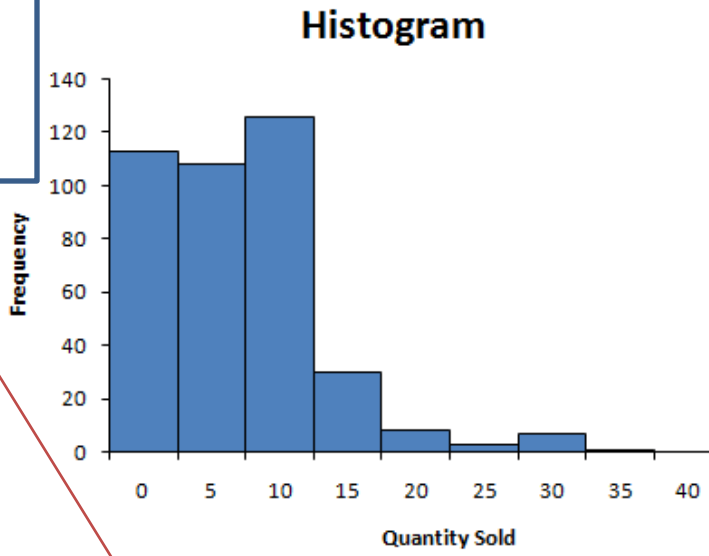
Quantity bought per transaction	Number of such transactions
3	664
6	11
9	3
15	1
Grand Total	**679**

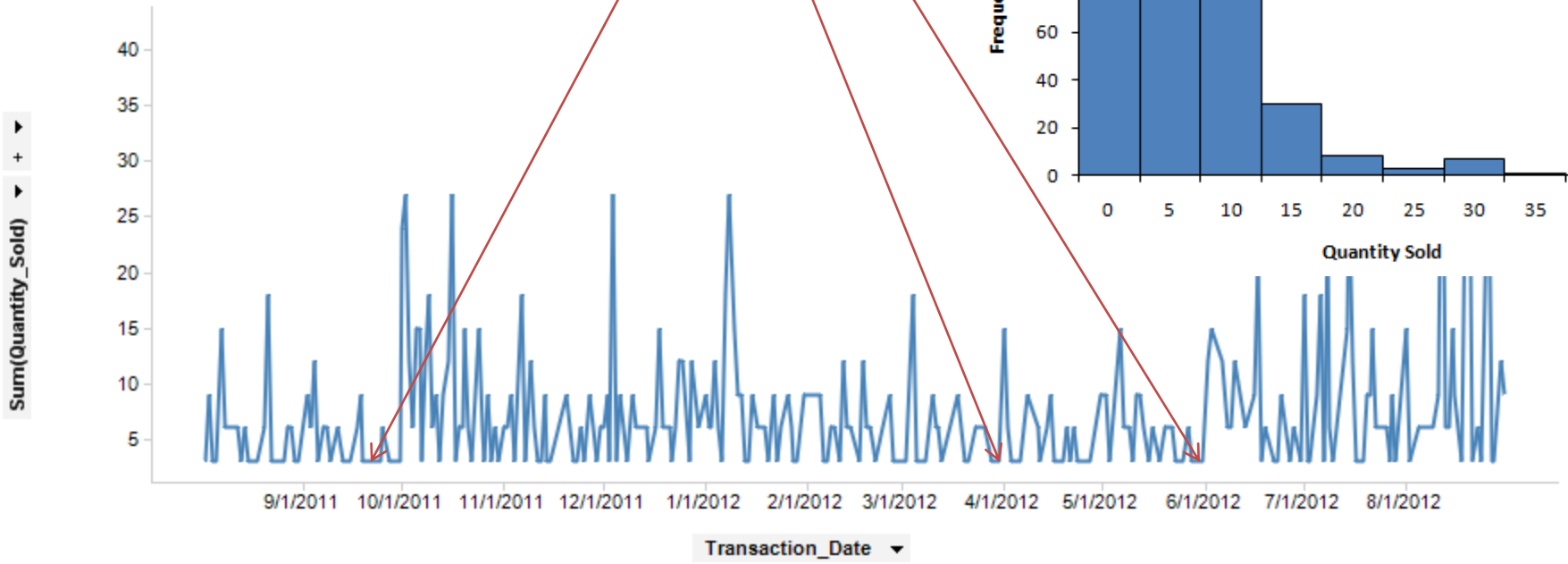
Replaced with most occurring ticket size

Initial Analysis: Missing Values

Preprocessing: Missing Values

Replaced with zero in the dataset.
Represent no sale on that particular day





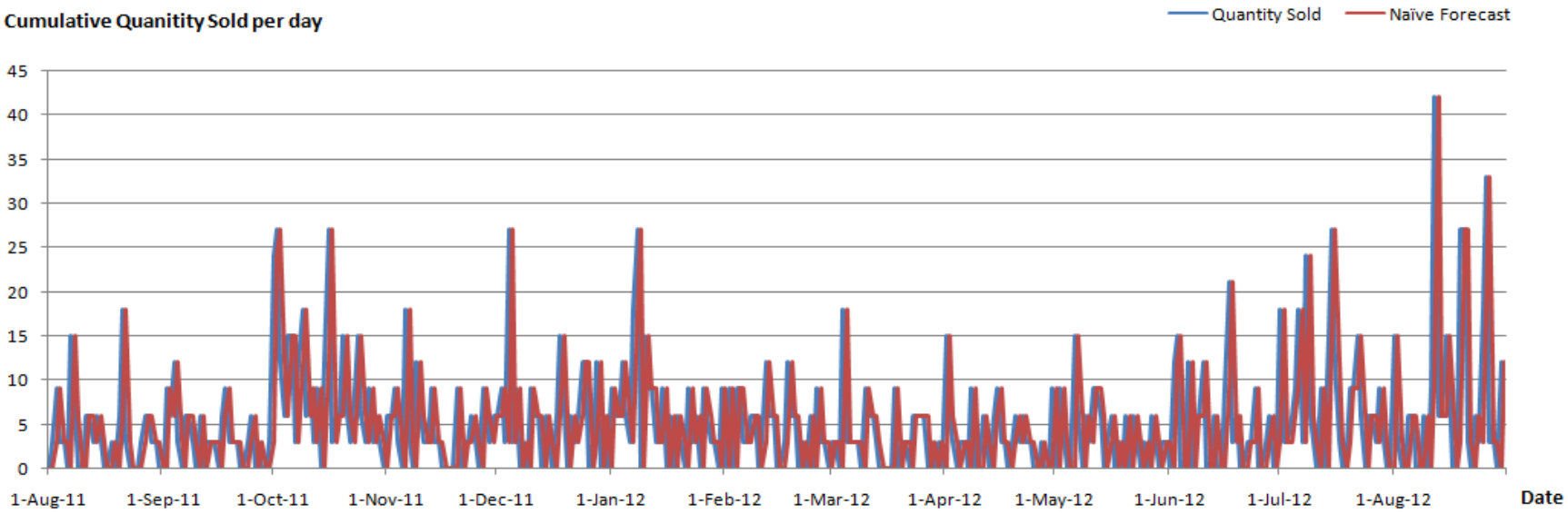
Is this a Random Walk?

Check to see if the data can actually be predicted
Tested for all SKUs
Results: Slope coefficient of AR(1) models significantly (more than 3 standard deviations away) different from 1 – hence they do not follow a random walk

ARIMA Model

ARIMA	Coeff	StErr	p-value
Const. term	4.48121786	0.38925543	0
AR1	0.1479677	0.02359452	0

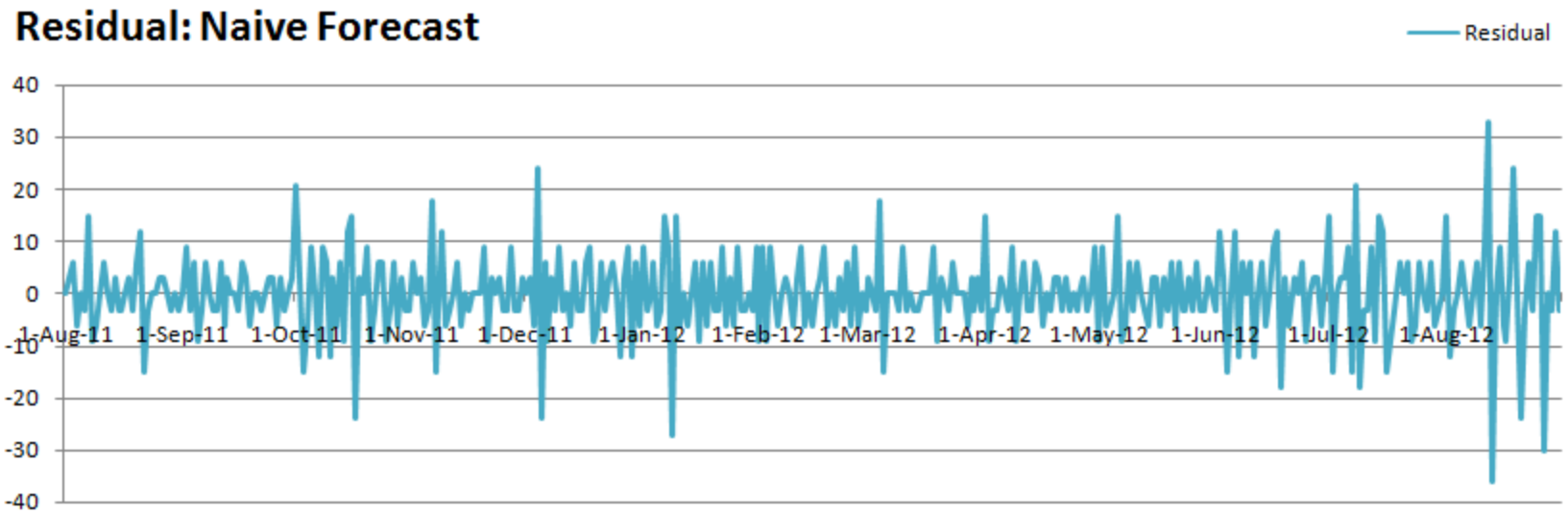
Performance: Naïve Forecast



Performance: Naïve Forecast

Used the Naïve Forecast as a **performance benchmark**

RMSE	13.81094
MAPE	150.20%



Choice of the model

- Does the data exhibit level, trend and seasonality?
- Can seasonality be captured by dummy variables?
- Model Choices:
  - Multi layered model
  - Smoothing model
- Did not consider so far:
  - Neural Network approach

Does data exhibit seasonality?

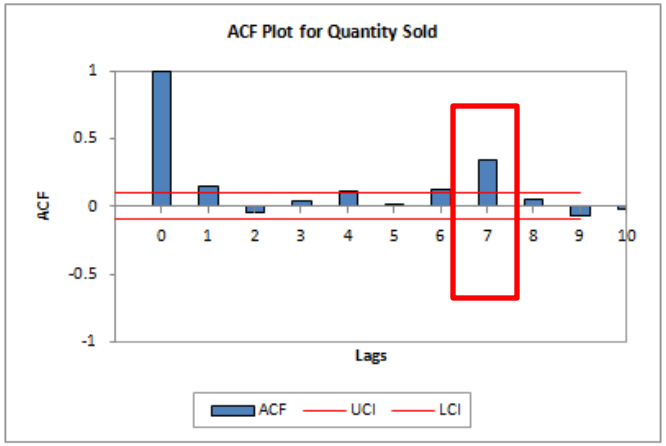
Yes, a weekly seasonality is exhibited as demonstrated by the ACF plot

Seasonal Naïve Forecast is an improvement over the naïve forecast

RMSE	11.31513
MAPE	89.70%

ACF Values

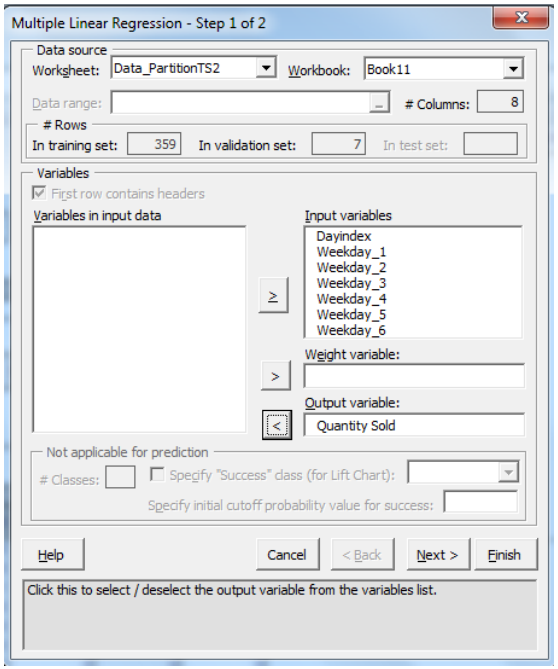
Lags	ACF
0	1
1	0.14790192
2	-0.04758157
3	0.03641458
4	0.10584228
5	0.01007687
6	0.12162134
7	0.34265426
8	0.05656664
9	-0.0686518
10	-0.01619564



Model: Multiple Linear Regression

Created using 6 dummy variables to account for weekly seasonality
Training: August 1st, 2011 to July 24th, 2012
Validation (1 week): July 25th, 2012 to July 31st, 2012
Test (1 month): August 1st, 2012 to August 31st, 2012

Data							
Data source	CategoryVar1!$D$11:$O$376						
Time variable	Dayindex						
Selected variables	Quantity Sold	Weekday_1	Weekday_2	Weekday_3	Weekday_4	Weekday_5	Weekday_6
Partitioning Method	Sequential						
# training rows	359						
# validation rows	7						



Model: Multiple Linear Regression

The Regression Model

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	12.23248482	0.75717109	0	8935.043945
Dayindex	-0.00127739	0.00229641	0.57839102	4.24783516
Weekday_1	-8.42627048	0.88860762	0	121.3146515
Weekday_2	-8.8288393	0.88859576	0	264.2489014
Weekday_3	-8.12275696	0.89293945	0	199.0139618
Weekday_4	-8.06265545	0.89291877	0	313.9673767
Weekday_5	-8.70843697	0.89290398	0	943.0252686
Weekday_6	-6.88363028	0.89289516	0	1208.293335

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
7135.844104	4.458363273	1.74485E-07

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
124.3544094	4.214844674	-0.90699376

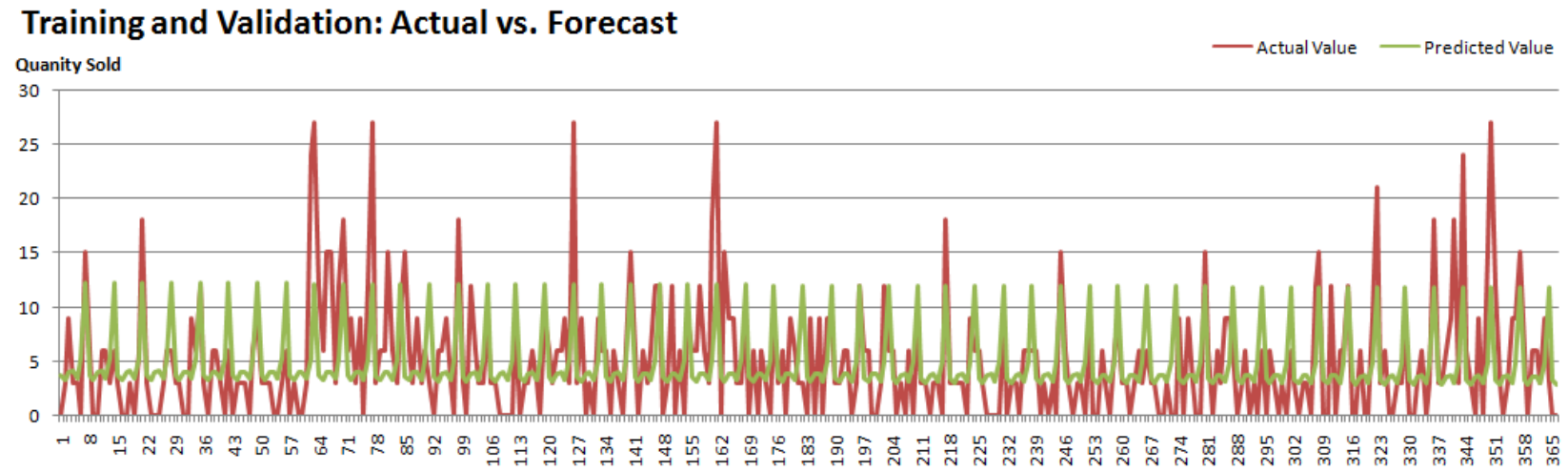
Seasonal Naïve Forecast

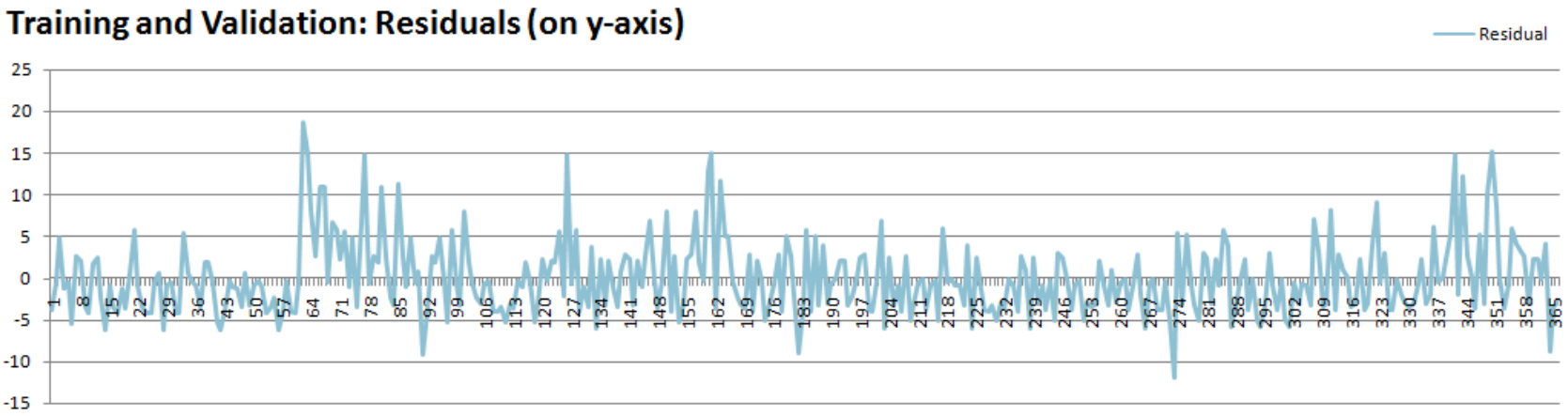
RMSE	11.31513
MAPE	89.70%

Multiple Linear Regression Forecast

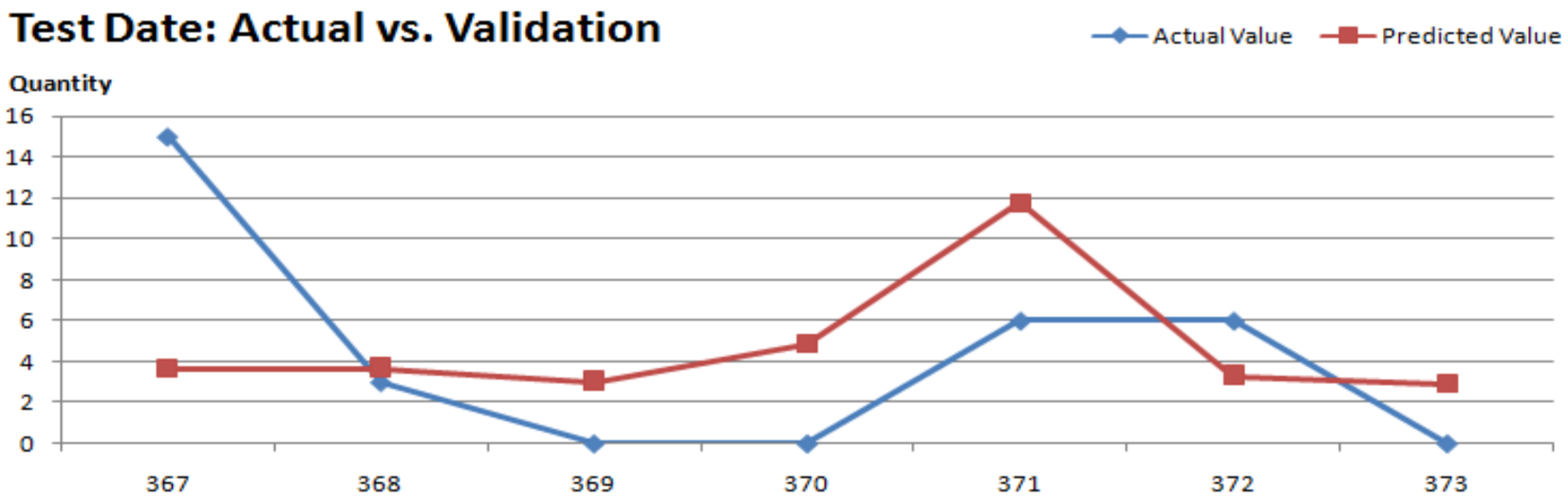
RMSE	5.49633
MAPE	77.07%

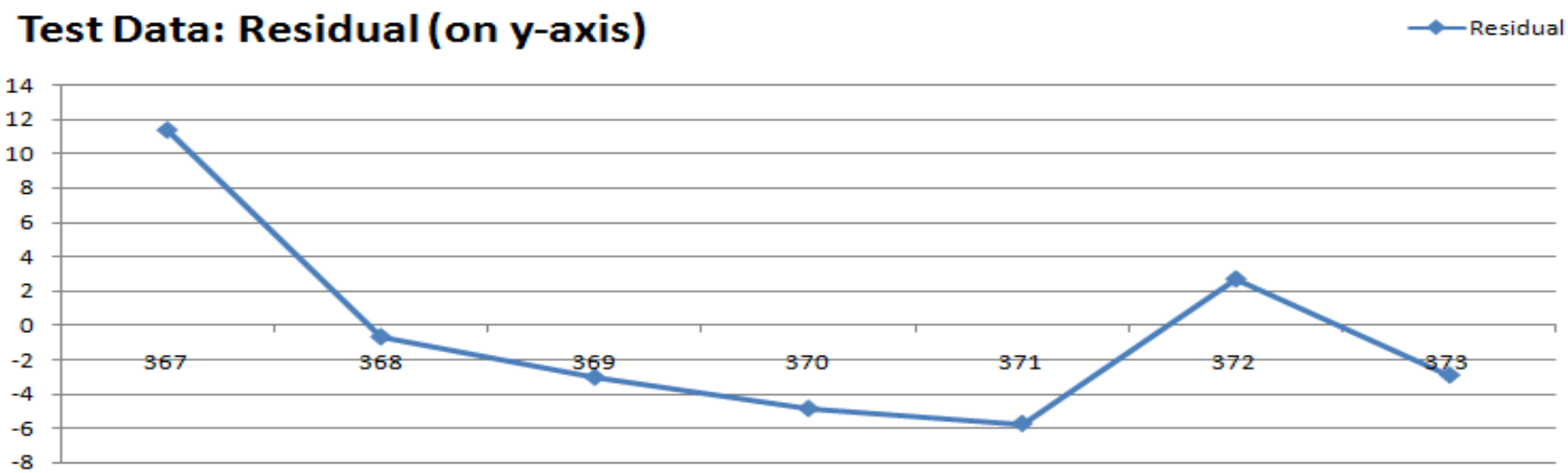
Model: Plots (for SKU: 100004925)





Model: Forecasts (for SKU: 100004925)





Signal in the Residual?

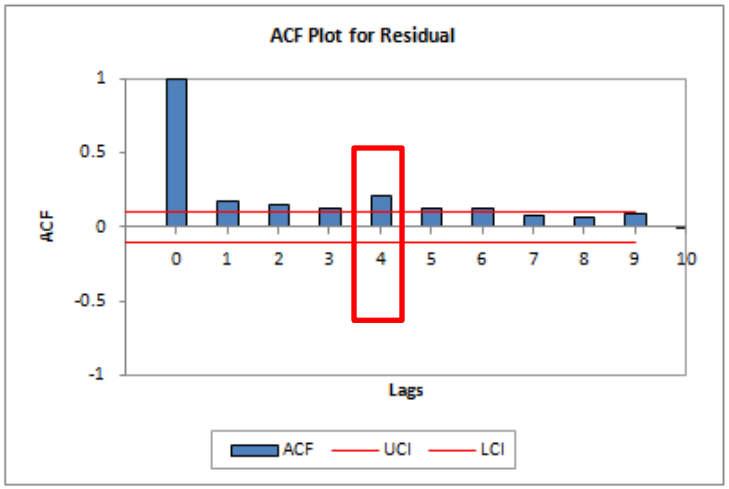
There appears to be some signal (lag(4)) in the residuals; we remodel using an AR model

ARIMA Model

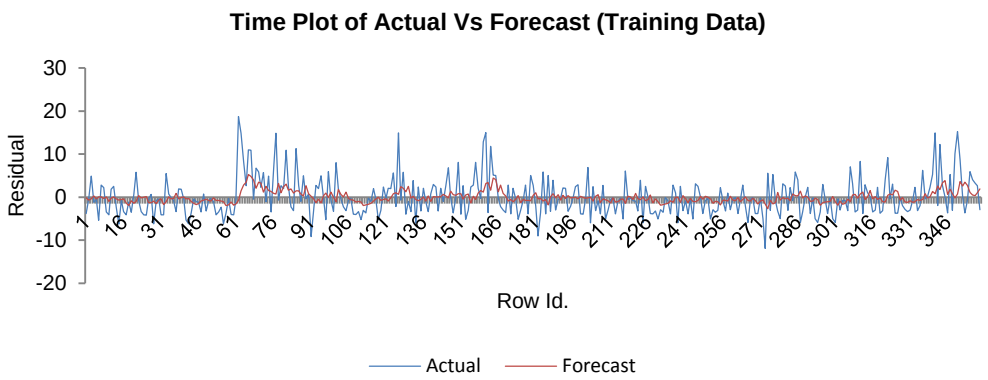
ARIMA	Coeff	StErr	p-value
Const. term	0.00000009	0.25780311	0.9999997
AR1	0.13082664	0.05197655	0.01183482
AR2	0.08756854	0.0522197	0.09355704
AR3	0.06450702	0.05234129	0.21778819
AR4	0.17353147	0.05194433	0.00083561

ACF Values

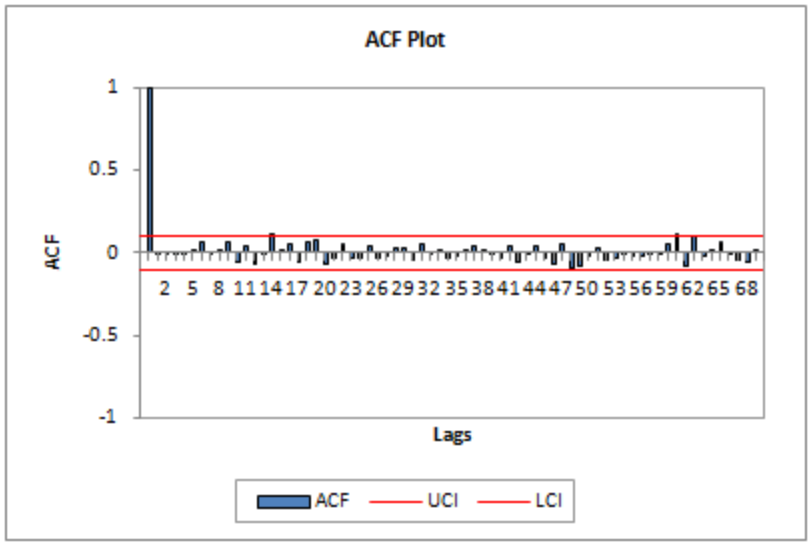
Lags	ACF
0	1
1	0.1770826
2	0.15019101
3	0.12846524
4	0.20902474
5	0.11898037
6	0.12250761
7	0.07009474
8	0.06833635
9	0.08505475
10	0.00809798

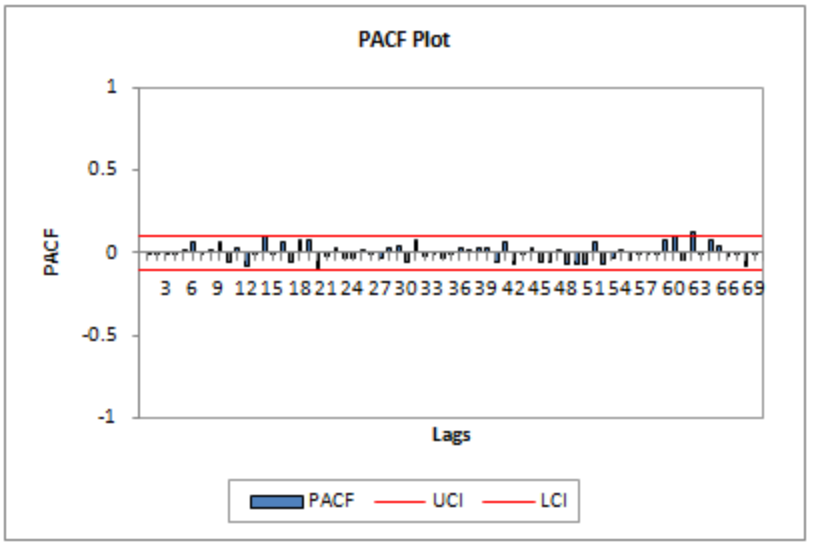


Signal in the Residual? Yes!!



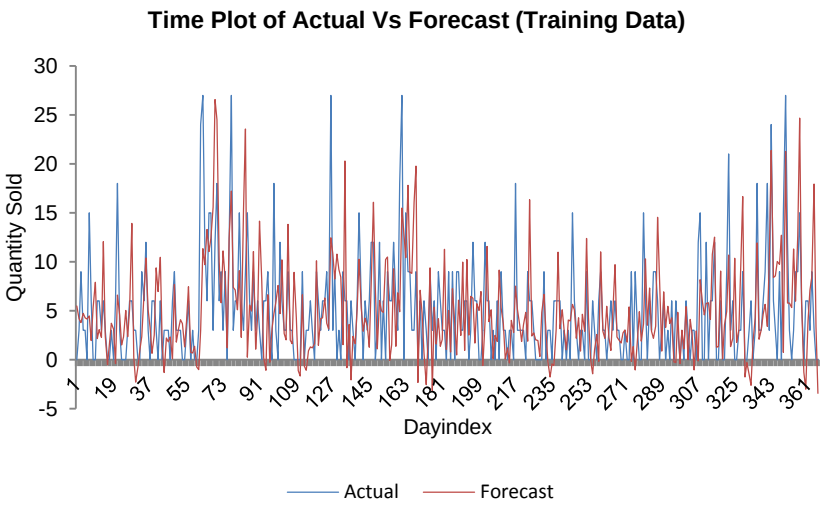
RMSE	5.262024111
MAPE	71.66%





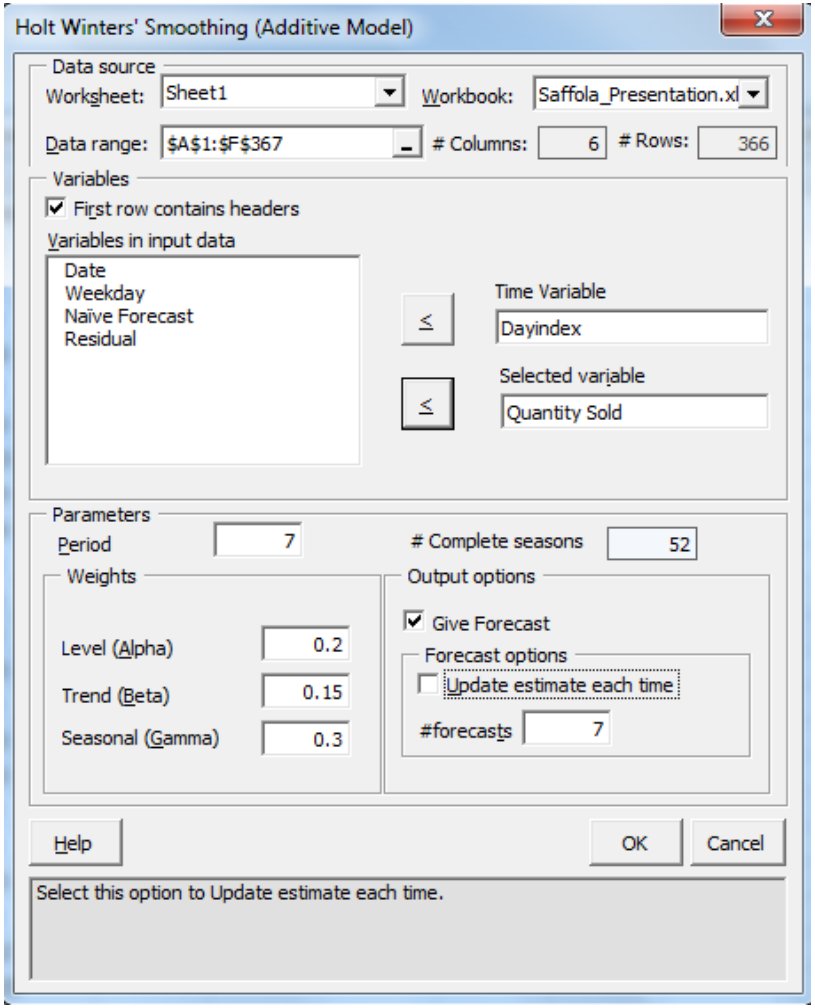
Model: Holt-Winters (Additive)

Actual vs. Forecast for Holt Winters



Various scenarios tried and the results:

alpha	beta	gamma	RMSE	MAPE
0.2	0.15	0.3	6.666152	87.56%
0.2	0.15	0.5	6.692811	81.70%
0.2	0.15	0.7	6.360498	78.15%
0.2	0.15	0.9	5.827601	80.04%



Comparison of results

Naïve Forecast

RMSE	13.81094
MAPE	150.20%

Naïve Seasonal Forecast

RMSE	11.31513
MAPE	89.70%

Multiple Linear Regression

RMSE	5.49633
MAPE	77.07%

Multiple Linear Regression with error prediction

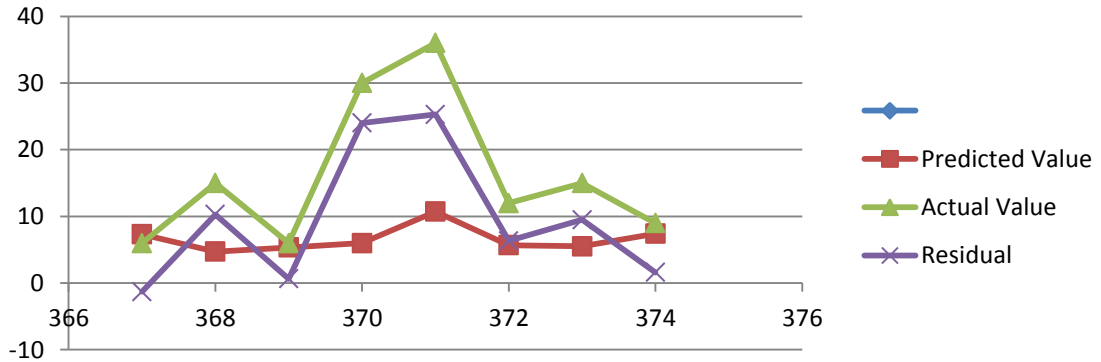
RMSE	5.2620
MAPE	71.66%

Holt Winters

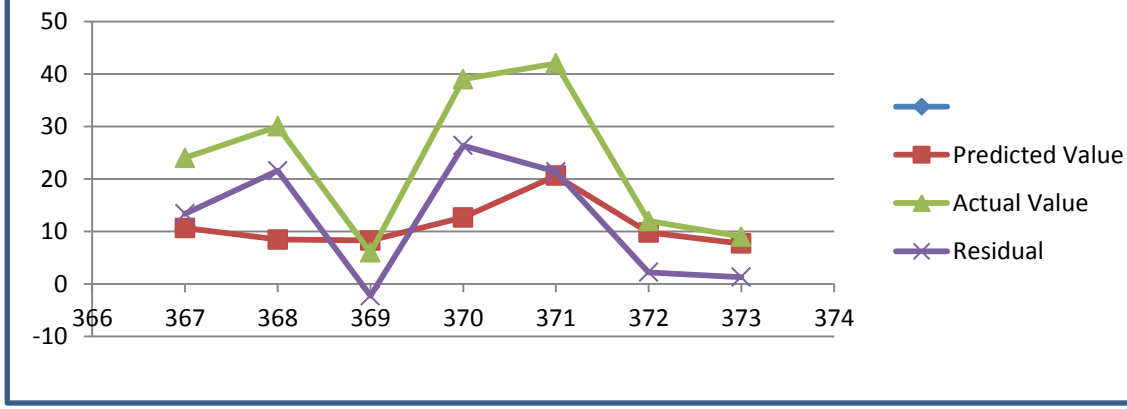
alpha	beta	gamma	RMSE	MAPE
0.2	0.15	0.3	6.666152	87.56%
0.2	0.15	0.5	6.692811	81.70%
0.2	0.15	0.7	6.360498	78.15%
0.2	0.15	0.9	5.827601	80.04%

Summary of results for other SKUs

RMSE	MAPE
14.4117	52.50%

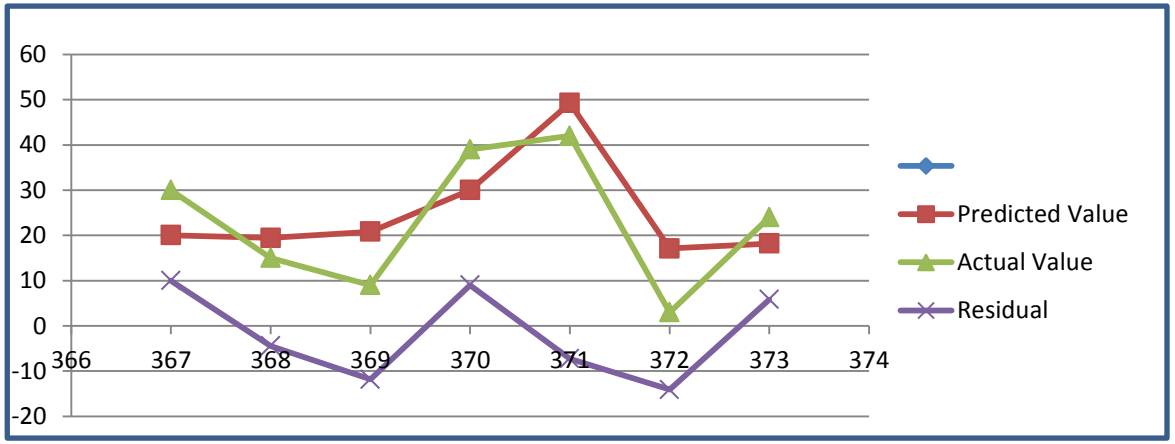


RMSE	MAPE
16.0635	45.22%

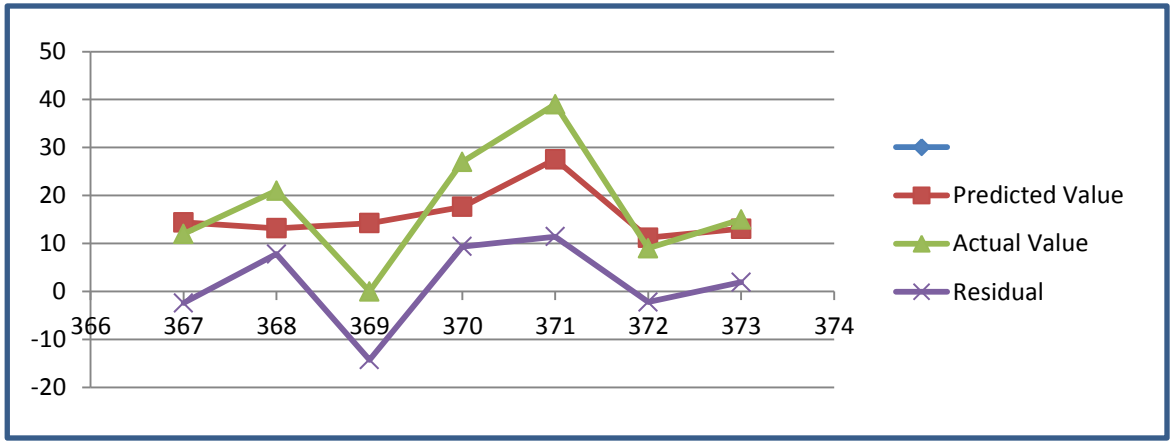


Summary of results for other SKUs

RMSE	MAPE
9.4568	104.19%



RMSE	MAPE
8.4231	36.93%



Other possible extensions

- Using the holiday calendar in sync with existing data
- Use econometric models: incorporate the effects of price changes and discounts, competitive brand pricing
- Model behavior of customers: predict/forecast repeat purchases, bulk purchases etc.