Predicting Average Basket Value
BADM – Group 9A

INDIAN SCHOOL OF BUSINESS, HYDERABAD

DECEMBER 2012

AASHISH SHARMA

ABHISHEK AGRAWAL

MADHAV PATHAK

MADHUR CHADHA

SAURABH MALHOTRA

Predicting the Average Basket Value

Business Objective

- As the hypermarket, we are interested in increasing the store sales and customer loyalty by offering DISCOUNTS to those customers whose predicted basket value (next purchase) is above the threshold level (Threshold value Rs.1,500 per basket)

Forecasting Objective

As the hypermarket, we are interested in predicting the average Basket value of the next customer who walks in based on his/her demographic data as well as his previous purchase pattern prior to this visit

- Since we have customer data for the past 1 year, we will use that to predict the basket value
- We will use basket level data for each customer at individual visit level as well as at an aggregate data level where we will club the information of his/her baskets purchased in all prior visits

Data Preparation

Used BASE SAS (9.1) for data preparation

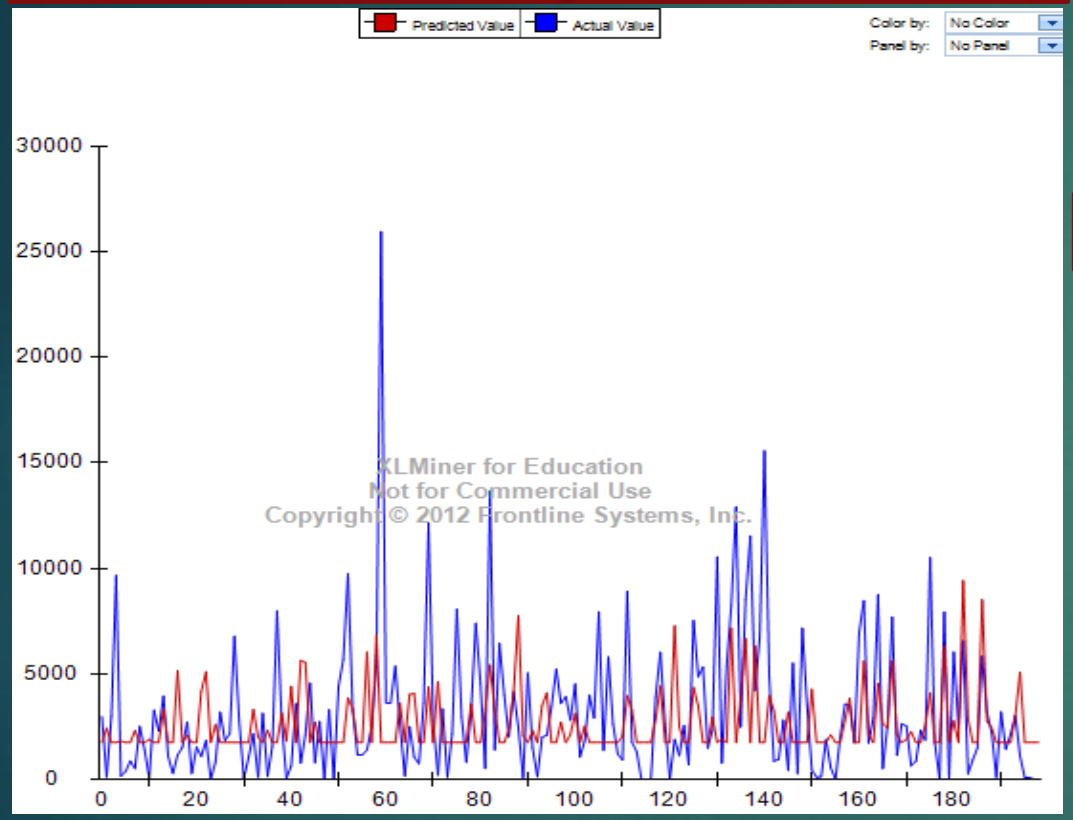
- Compiled data at the Basket level for each customer – Merged individual rows to compute basket price, quantity, etc
- Computed the Average Basket Price & Quantity for each customer at transaction level using all purchases prior to current transaction date
- Included 4 dummy variables in data set - Sex, Married, Mobile # and Email address
- Used the following 3 methods to analyze data and compute a predictive model
  - Multi Linear Regression
  - K-Nearest Neighbors
  - Classification and Regression Trees (CART)

Example Data Set

Customer_No	Transaction_Date	Day of Week	Quantity_Sold	Extended_Price	Avg_Quantity_Lag	Avg_Ext_Price_Lag	Count_Footfall	Total_qty_sold_tilldate	Total_cart_tilldate	AGE	SEX_M	ENROLLMENT_AGE	ENROLLMENT_STORE	MARRIED	MOBILE	EMAIL
3000001804	5/8/2012	3	9	1500.03	0	0	0	9	1500.03	36.69	1	3.71	1001	1	1	1
3000002778	8/11/2012	7	30	1478.85	0	0	0	30	1478.85	29.40	1	3.71	1001	0	1	1
3000008858	10/17/2011	2	80.82	4703.67	0	0	0	80.82	4703.67	30.27	0	3.70	1001	0	1	1
3000008858	3/12/2012	2	90	6558.51	80.82	4703.67	1	170.82	11262.18	30.27	0	3.70	1001	0	1	1
3000008858	3/29/2012	5	12	897.00	85.41	5631.09	2	182.82	12159.18	30.27	0	3.70	1001	0	1	1
3000010854	1/27/2012	6	45	1047.00	0	0	0	45	1047	45.69	1	3.69	1001	0	1	1
3000011100	7/26/2012	5	12	1782.03	0	0	0	12	1782.03	74.49	1	3.70	1001	1	1	0
3000016604	1/16/2012	2	97.938	6300.72	0	0	0	97.938	6300.72	54.04	1	3.68	1001	0	1	0
3000017362	8/17/2012	6	57	3671.91	0	0	0	57	3671.91	45.18	1	3.68	1001	1	1	1
3000027247	12/19/2011	2	246	10761.51	0	0	0	246	10761.51	29.83	1	3.67	1001	0	1	1

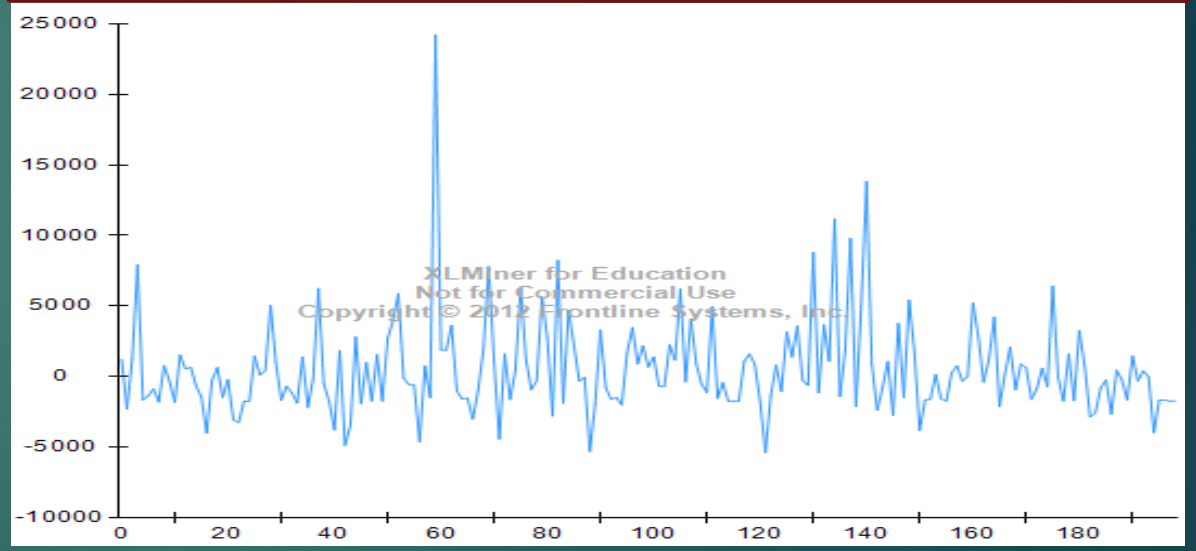
Naïve Forecast

Actual vs. Predicted plot



RMSE	RMSE	R^2
Training Data	Validation Data	Training Data
4101.0903	4170.9186	0.2411

Residuals plot

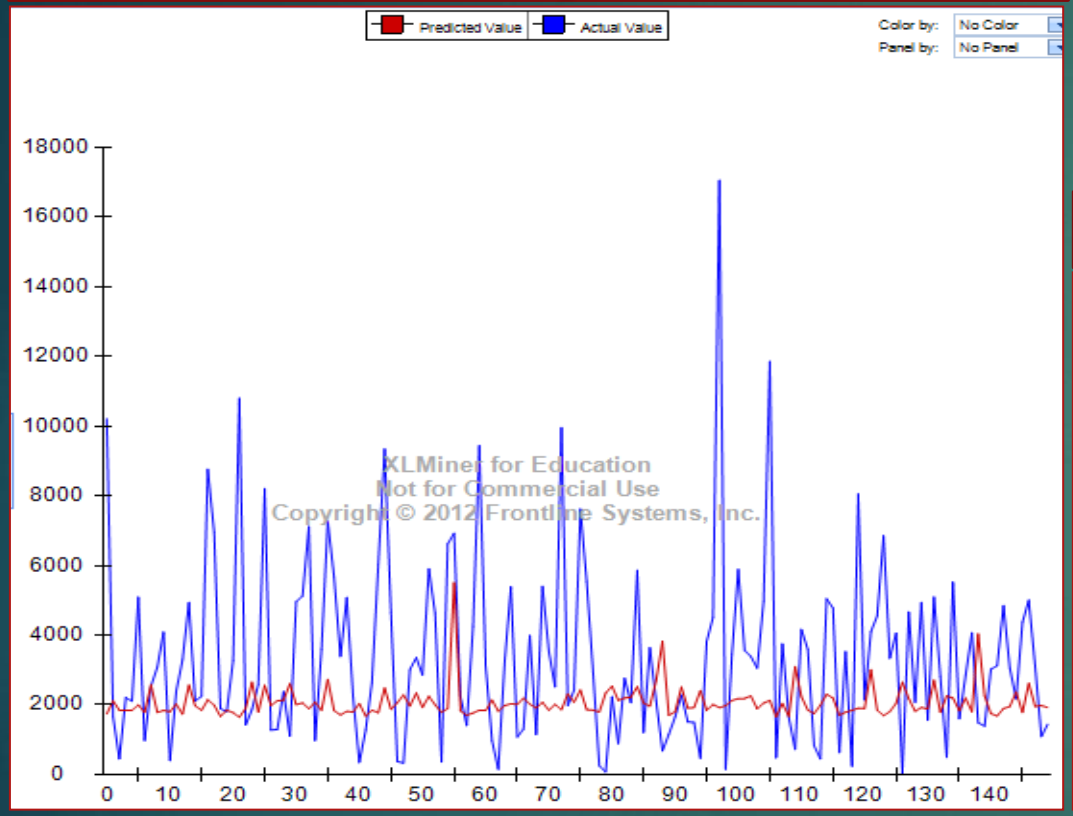


Model

Y = 1736.7644 + 0.6333 * Avg. Ext Price_Lag

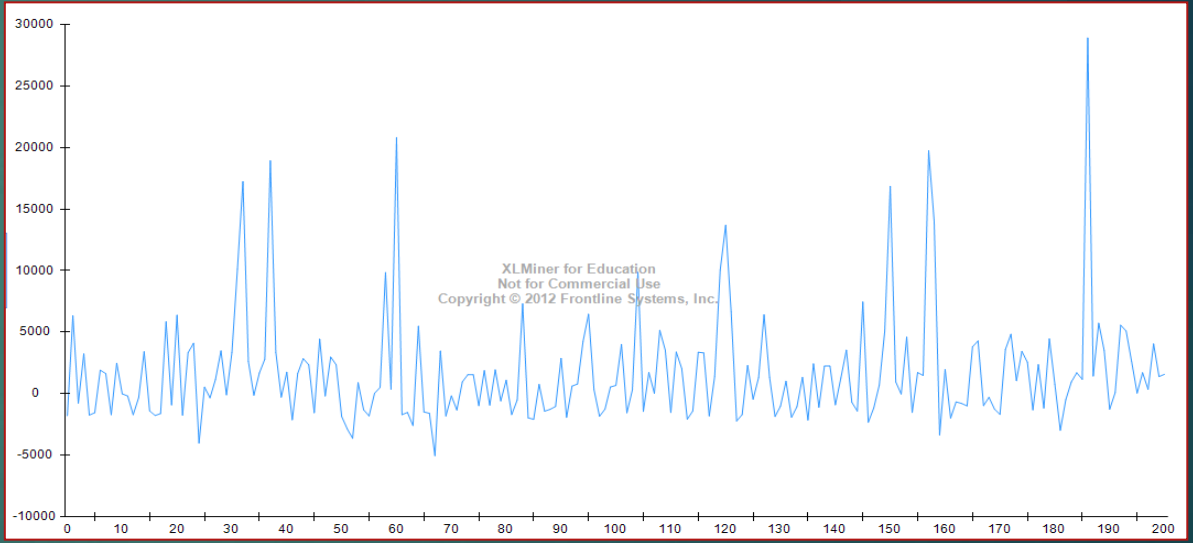
Multiple Linear Regression

Actual vs. Predicted plot



RMSE	RMSE	R^2
Training Data	Validation Data	Training Data
3980.4571	4019.9053	0.2371

Residuals plot

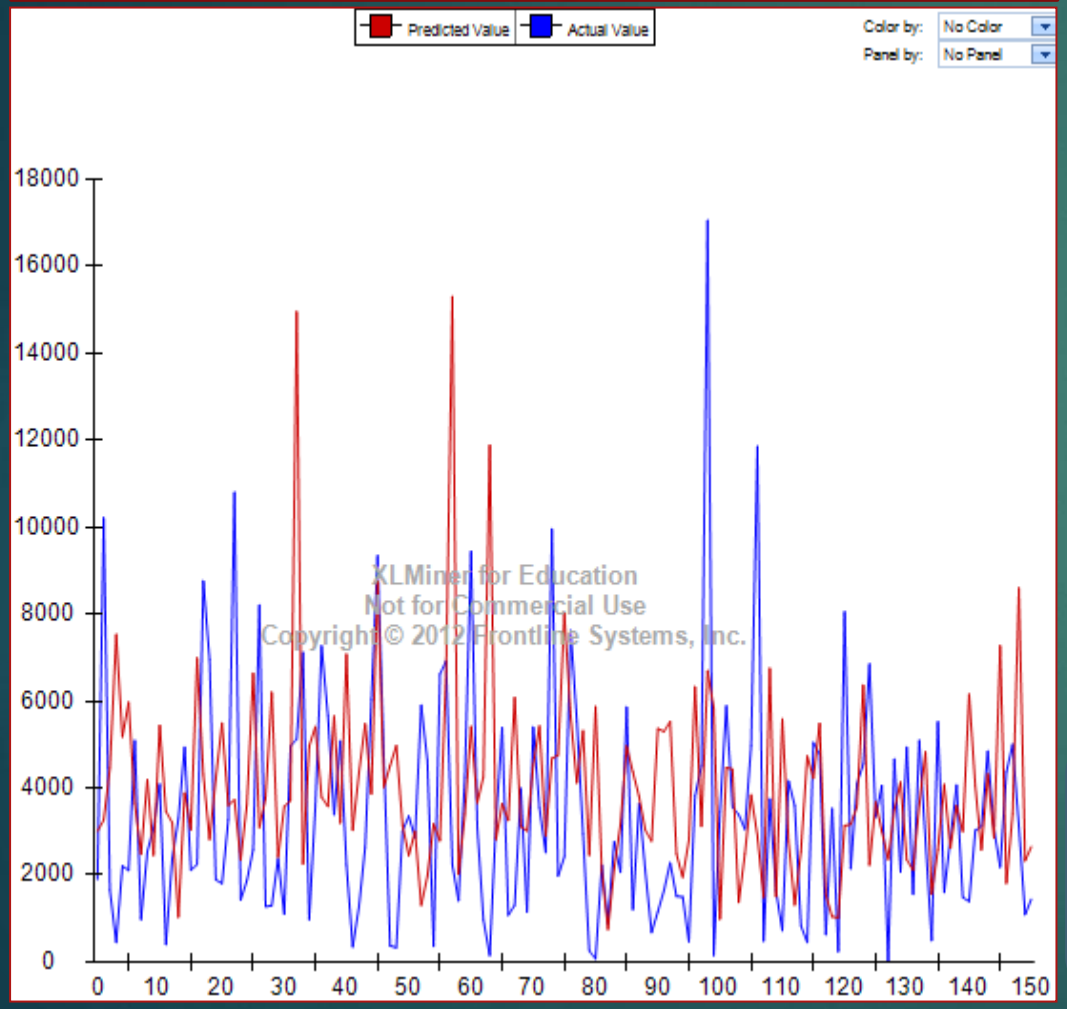


Model

Y = 2066.58 + -76.29*Day of week + 0.5505*Avg Ext Price (Lag) + -23.30*Count (Footfall) + 16.93*AGE + -494.82*SEX_M + 100.08*ENROLLMENT AGE + 150.20*MARRIED + -154.86*CLEAN_EMAIL_FLAG

K-NN

Actual vs. Predicted plot



RMSE			
Training Data	Validation Data	Test Data	Nearest K
24.4435	3957.3642	4085.73	20

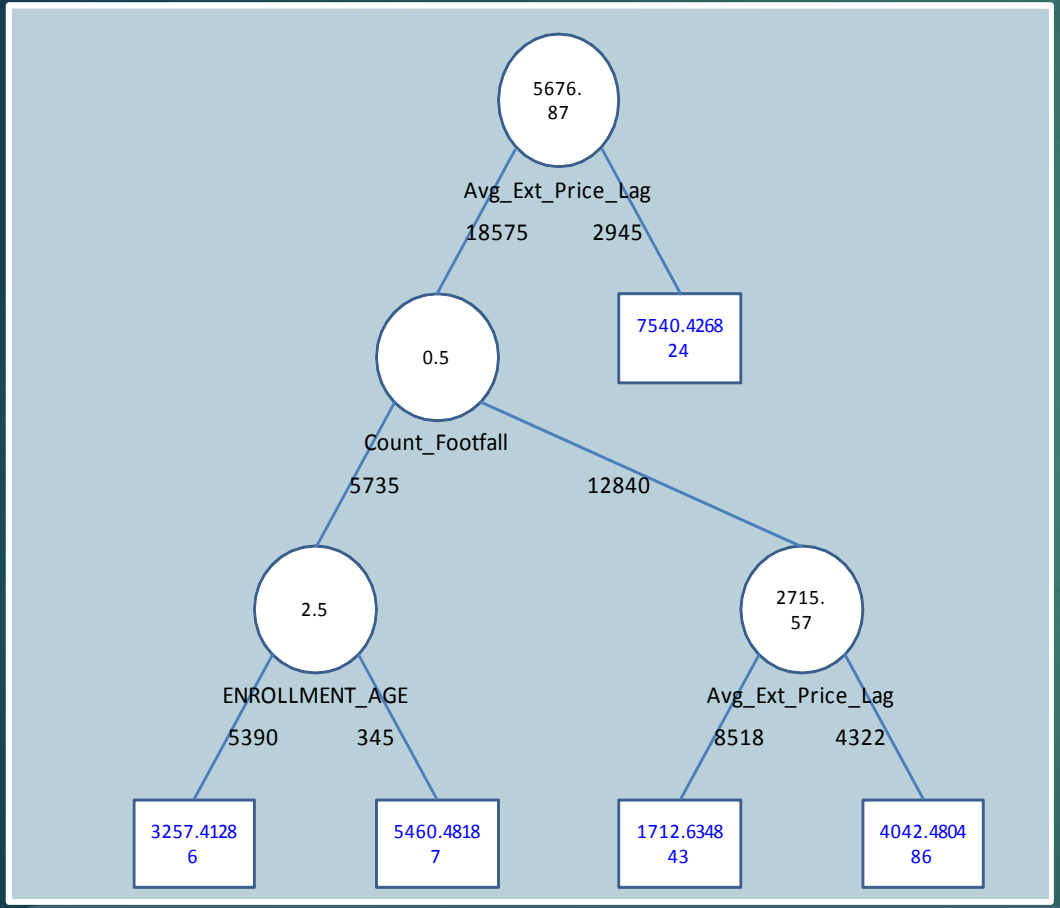
Residuals plot

Model

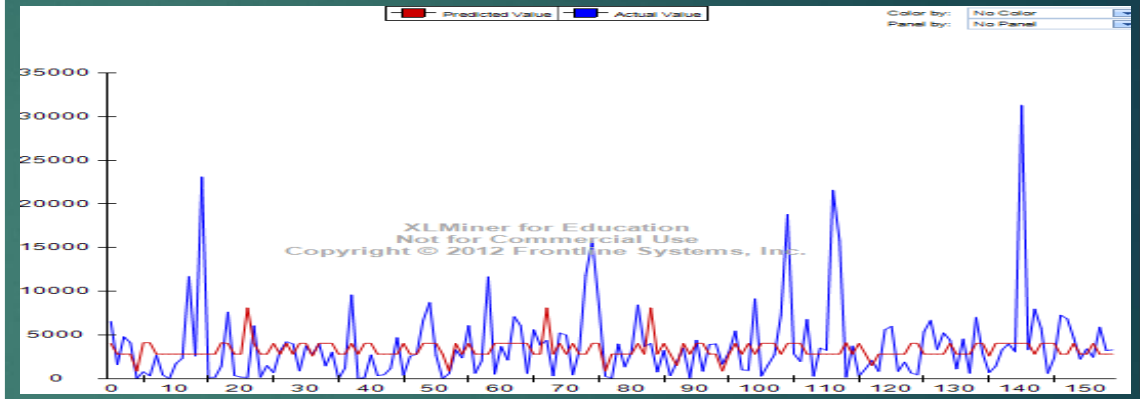
- **Input Variables: Day of Week, Avg_Ext_Price_Lag, Count_Footfall, AGE, SEX_M, ENROLLMENT_AGE, MARRIED**

CART

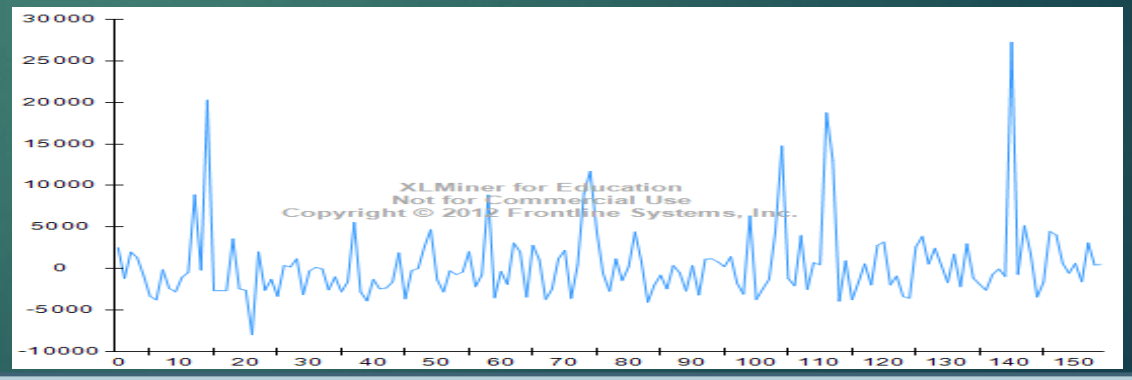
Full Regression Tree (Training Data)



RMSE		
Training Data	Validation Data	Test data
4502.3423	3923.2403	3971.0112



Actual vs. Pdredicted



Residuals plot

Model

- ❑ **Input Variables: Day of Week, Avg_Ext_Price_Lag, Count_Footfall, AGE, SEX_M, ENROLLMENT_AGE, MARRIED**

Model selection & evaluation

CART

RMSE		
Test data	Naive	% Improvement over Naïve (RMSE)
3971.01	4170.92	5%

K-NN

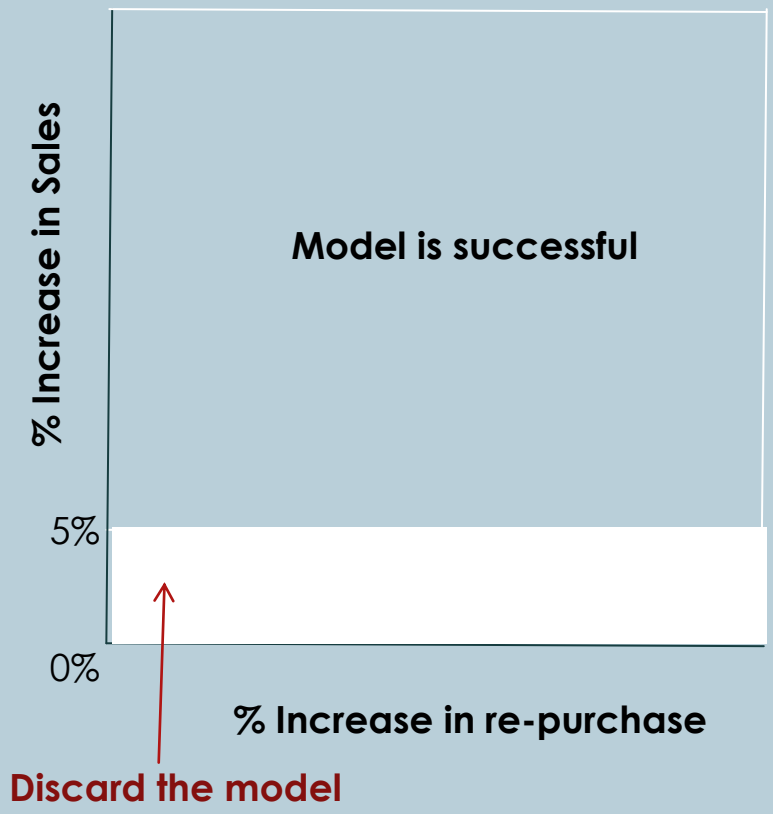
RMSE		
Test Data	Naïve	% Improvement over Naïve (RMSE)
4085.73	4170.92	2%

MLR

RMSE		
Validation Data	Naïve	% Improvement over Naïve (RMSE)
4019.91	4170.92	4%

- We have used RMSE as the evaluation criteria for the different models
- Based on the RMSE data, CART comes out to be the most successful model for prediction of the Basket value

Business Perspective - Performance evaluation



- Two ways to measure the performance of this initiative
  - % Increase in Sales and % increase in re-purchase rate for existing customers
    - Prerequisite for this benchmark: % increase in sales should be greater than 5%
  - % increase in profitability
    - Ideal measure but since data is not available we are not using this metric for evaluation

Limitations

- Macro-economic factors have been ignored for the purpose of this model
- To improve the prediction quality of the model, the hypermarket should further add more predictor variables by collecting information from the customers
- Demographic data might be misleading as many customers might have one loyalty card per family and thus the loyalty card might be used by multiple people in the family
- All customers will be given a loyalty card while he/she makes a purchase at the hypermarket

Recommendations

- We would recommend the hypermarket to adopt and deploy the CART model to predict the basket value for the next purchase
- The hypermarket should offer a discount to those customers whose predicted basket value (for the next purchase) is greater than Rs.1,500 (defined threshold)
  - Using this model, the hypermarket will be able to better satisfy its customer base, and improve its sales and profitability
  - This model should be updated on a real time basis since it extensively uses the prior purchase of these customers as an input variable

Thank You