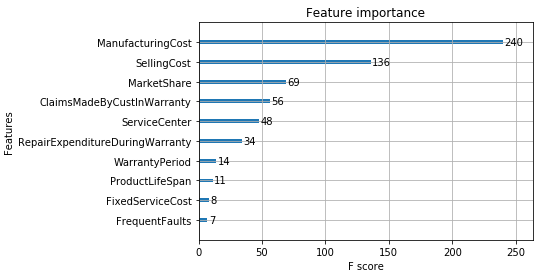
**Product Analysis**

1. **Develop a solution architecture for identifying the product variant which will be launched in the market for enhanced profit margin(X%), customer segmentation etc? Also detail out the estimated profit margin of the new-variant product ?**

Ans:- First we developed the sample dataset for past 9 years including all the features. After that we calculated the mean of every products for past 9 years, then we applied feature selection on the dataset to select which feature are important for predicting the profit margin. Using XGBoost we got the following feature importance graph.



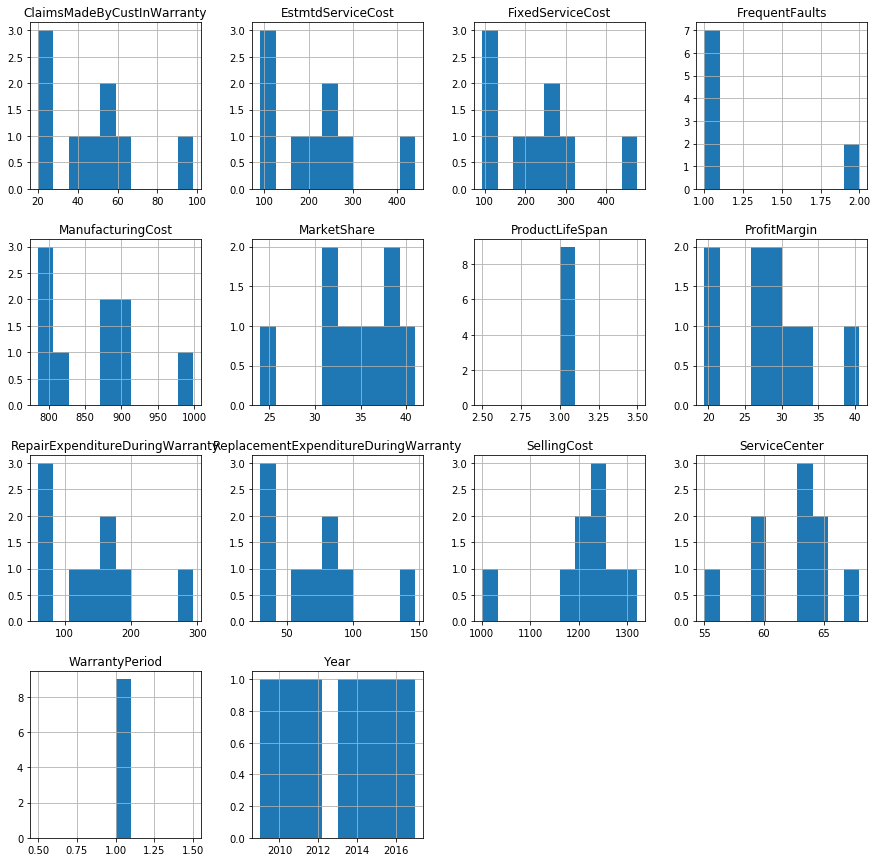
Following major features for predicting the profit margin are :

1. Manufacturing cost
2. Selling Cost
3. Market Share
4. Claims Made by Customer in warranty
5. Service Centre
6. Repair Expenditure During Warranty
7. Warranty Period
8. Product Life Span
9. Fixed service Cost
10. Frequent Faults

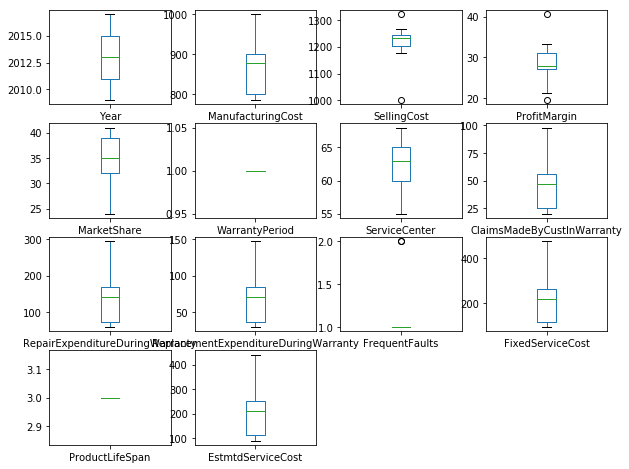
So finally we came to conclusion that product abcd\_23 will be the new variant of product for launching in the market with profit margin of 28.53.

We also applied the descriptive analysis in our dataset.

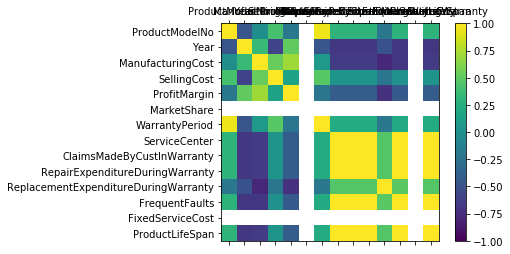
We plotted the histogram for every variable to get the distribution of each attribute.



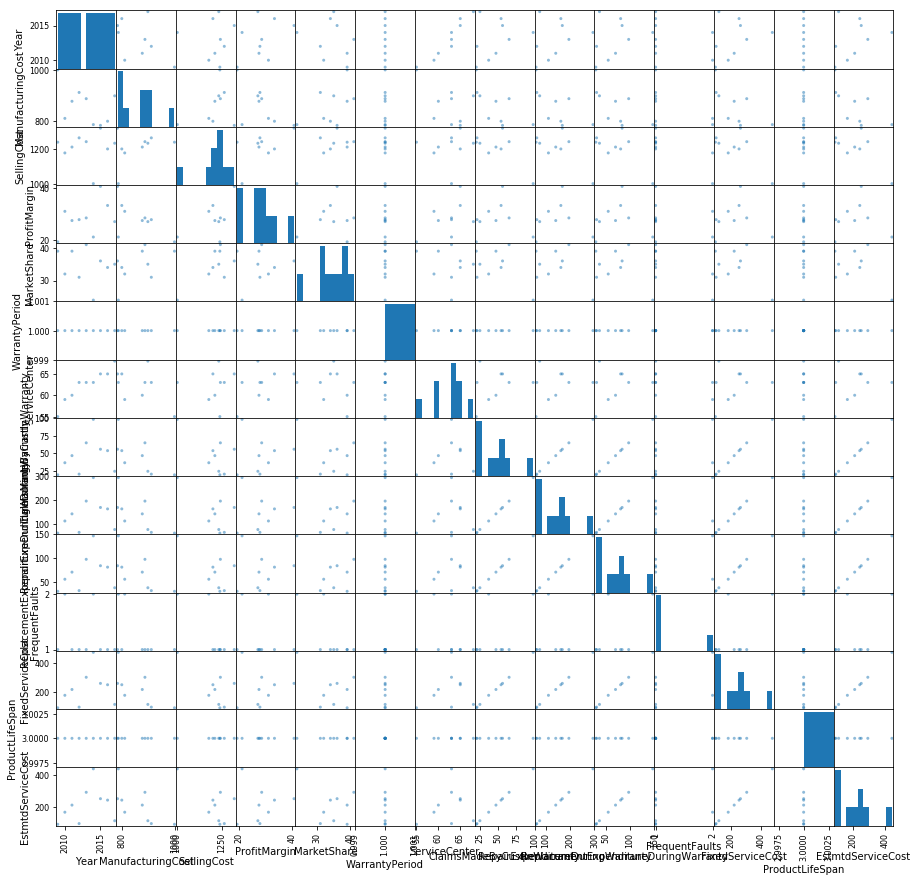
Then we plotted the boxplot for every variable to get the distribution of each attribute.Boxplots summarize the distribution of each attribute, drawing a line for the median (middle value) and a box around the 25th and 75th percentiles (the middle 50% of the data). The whiskers give an idea of the spread of the data and dots outside of the whiskers show candidate outlier values



Then we plotted the correlation matrix plot for every variable.Correlation gives an indication of how related the changes are between two variables. If two variables change in the same direction they are positively correlated. If the change in opposite directions together, then they are negatively correlated.You can calculate the correlation between each pair of attributes. This is called a correlation matrix. You can then plot the correlation matrix and get an idea of which variables have a high correlation with each other.



Then we plotted the scatter plot for every variable.A scatterplot shows the relationship between two variables as dots in two dimensions, one axis for each attribute. Scatter plots are useful for spotting structured relationships between variables, like whether you could summarize the relationship between two variables with a line. Attributes with structured relationships may also be correlated and good candidates for removal from your dataset.This is useful to look at the pair-wise relationships from different perspectives



Linear Regressor : divided the whole 9 year data into train and testing the for test data we got the approx same value for 2 row which is having less difference

Mean squared error: 2.21

Variance score: 0.6047792305915647

KNN : it predicted the one row approx same having less difference for other two row diff was little bit more

Mean Square Error 8.555613199869791

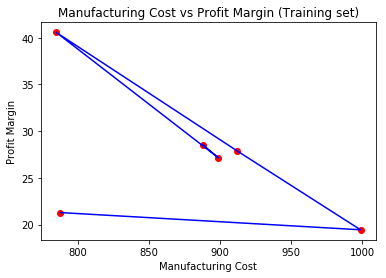
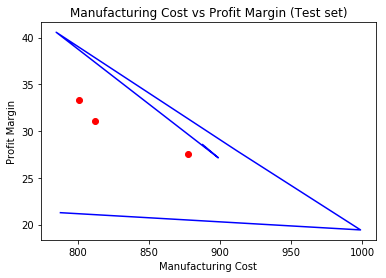
Variance score: 0.528693377088242

XGBoostRegressor : it successfully able to predict the one row correctly out of 3 row

Mean Square Error 32.809834798177086

Variance score: 4.862371208735267

Linear regression gave the best accuracy for profit margin prediction

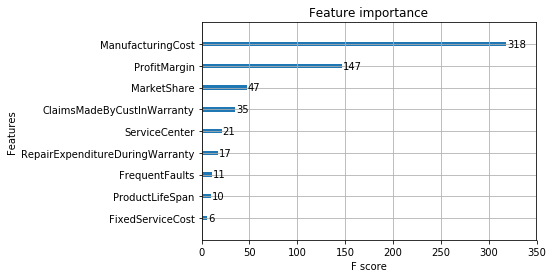


**2. What should be the selling cost of this new variant of a product considering a profit margin of X%?**

Ans:- For profit margin of 28.53% the selling cost of product will be 1210.55

We also created the ML model for prediction where our target variable was selling cost

For the second problem statement we calculated the selling cost of the product by using 3 model having the selling cost as the target variable. Here is the feature importance for predicting selling cost



Linear Regressor : divided the whole 9 year data into train and testing the for test data we got the approx same value for 2 row which is having less difference

Mean squared error: 1600.54 Variance score: 6.976096802139511

KNN : it predicted the one row approx same having less difference for other two row diff was little bit more

Mean squared error: 1294.15 Variance score: 5.449227672082642

XGBoostRegressor : it successfully able to predict the one row correctly out of 3 row

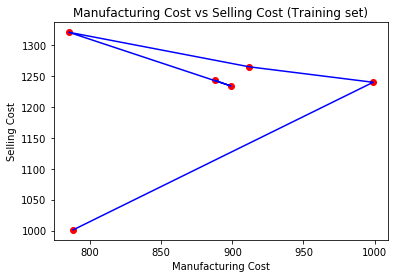
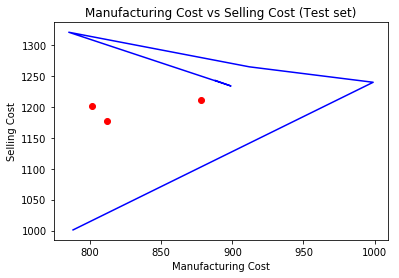
Mean squared error: 3206.33 Variance score: 14.978405315614618

If we want to predict the selling cost for new variant of product than we have to provide the some details for test, we can take random data for predicting the selling cost.

KNN gave the good accuracy for prediction.

Visualising the Training and Test Result for Manufacturing Cost vs Profit Margin using Linear Regresion , where Manufacturing Cost is contributing the most.

Visualising the Training and Test Result for Manufacturing Cost vs Selling Cost using Linear Regresion , where Manufacturing Cost is contributing the most.

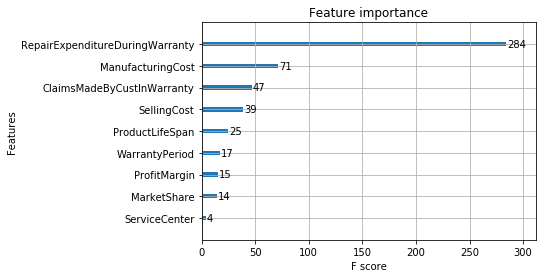


1. **What should be the fixed Service Charges (this will exclude the parts cost and labour cost ) for this new variant of a product. The company will demand this charge (cost) from the Customers.?**

Ans:- The fixed service charge for this new variant of a product will be 222.66

We also created the ML model for prediction where our target variable was fixed service charges

Below is the feature selection for fixed service charges prediction



Linear Regressor : divided the data into train and testing the for test data we got the approx same value for 3 row which is having less differences between actual and predicted.

Mean squared error: 20.32 Variance score: 0.9771372940946602

KNN : it predicted the one row approx same having less difference for other two row diff was little bit more

Mean squared error: 48.95 Variance score: 0.9449272562642429

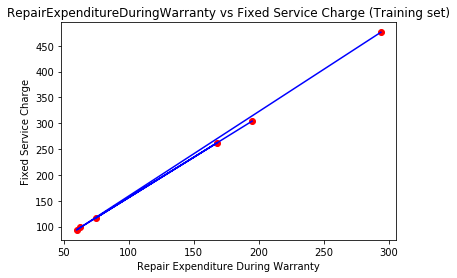
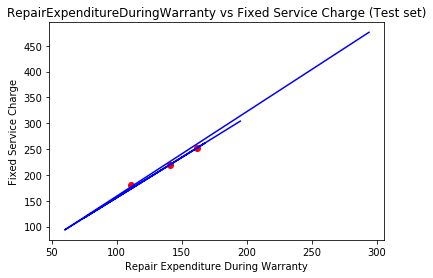
XGBoostRegressor : it successfully able to predict the one row correctly out of 3 row

Mean squared error: 3710.11 Variance score: -3.1744801400630127

If we want to predict the fixed service cost for new variant of product than we have to provide the some details for test, we can take random data for predicting the fixed service cost.

for predicting the fixed service cost linear regression gave the best accuracy.

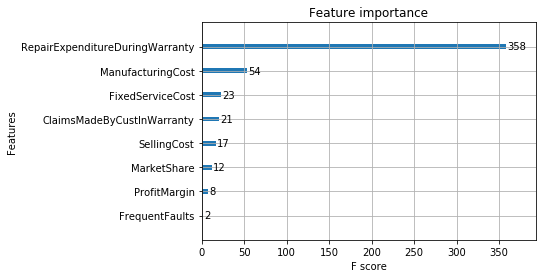
Visualising the Training and Test Result for Repair Expenditure During Warranty vs Fixed Service Charge using Linear Regresion , where Repair Expenditure During Warranty is contributing the most

**4. What would be the estimated service cost ? The company has to bear for servicing related expenditure (viz. repair, parts replacement etc.) for the expected service requests during warranty period of new-variant product? Warranty period should be assumed as 1 Year.**

Ans:- We also created the ML model for prediction where our target variable was estimated service cost

Below is the feature selection for predicting the estimated service cost.



Linear Regressor : divided the data into train and testing the for test data we got the approx same value for 3 row which is having less differences between actual and predicted.

Mean squared error: 7.00 Variance score: 0.9928926426301602

KNN : it predicted the one row approx same having less difference for other two row diff was little bit more

Mean squared error: 46.17 Variance score: 0.9531507167682296

XGBoostRegressor : it successfully able to predict the one row correctly out of 3 row

Mean squared error: 3186.00 Variance score: -2.232876712328767

If we want to predict the estimated service cost for new variant of product than we have to provide the some details for test, we can take random data for predicting the estimated service cost.

test\_model = [[789,1099,28.20746133,25,1,67,12,45,23,2,232,3]] where warranty period = 1

for this test result the estimated service cost prediction is : 154.554

For predicting the estimated service cost linear regression gave the best accuracy.

Visualising the Training and Test Result for Repair Expenditure During Warranty vs Estimated Service Cost using Linear Regresion , where Repair Expenditure During Warranty is contributing the most

