SAYANTAN RAHA

Roll # : BAI09056

IIMB - BAI09 - Assignment 2

Out[2]: Toggle on/off Code

Q 1.1

We will use the following formula to calculate the coefficient of CRIM.

$$\beta = r * \frac{SD_Y}{SD_X}$$

$$\text{where r = Correlation of X (CRIM) and Y (PRICE) \&}$$
$$SD_x = \text{Standard deviation of X}$$
$$SD_y = \text{Standard deviation of Y}$$

From table 1.1 we can find SDx = 8.60154511 & SDy = 9.197 From table 1.2 we can find r = -.388

Using the above we can find:

```
B1 -0.41485988323788486, implies as crime rate increases by 1 unit, unit price reduces by 0.41485988323788486 units (Lac IN
R)
```

Q 1.2

The range of coefficients is given by:

$$\beta \pm \text{t-crit} * SE_{beta}$$

where t-critical is the critical value of T for significance alpha.

Interpretation:

$$\beta = \text{Increase in Y as X changes by 1 Unit}$$

```
T-critical at alpha 0.05 and df 505 is 1.964672638739595
Min B1 -0.3284142871333427
Max B1 -0.5013054793424271
Price will reduce between 32K to 50K with 95% CI, hence his assumption that it reduces by at least 30K is correct
```

Q 1.3

Regression is valid for only the observed ranges of X (Predcitor). The min value of Crime rate = .0068 > 0. Hence it is incorrect to draw any conclusion about the predicted values of Y for Crim==0 as that value is unobserved.

We cannot claim the value will be 24.03

Q 1.4

Here Y predicted can be calculated from the regression equation: 24.033 - 0.414 * 1 (Value of CRIM)

For large values of n the range of Y-predicted is given by:

$$\hat{Y} \pm \text{t-crit} * SE_Y$$

where t-critical is the critical value of T for significance alpha (0.05).

```
Max Value of Price for CRIM ==1 is 40.28728266706672
```

Q 1.5

Here Y predicted (mean value of regression) can be calculated from the regression equation: 24.033 + 6.346 * 1 (Value of SEZ)

t-critical is computed as:

$$t = \frac{(t_o - t_{mean})}{SE_{estimate}}$$

We can calculate the probability using CDF of a normal Distribution. Since the value is >= 40 Lac, hence we will consider the right-tail of the t-value to compute the probability.

```
Mean Regression value 28.44
t-crit at alpha 0.05 is 1.2753751103265665
Y-pred follows a normal distribution. Probability of Price being at least 40 lac is 10.11 percent
```

Q 1.6 - a

From the residual plot, by visual inspection we can see that the spread of standardised errors are higher for lower values of standardised prediction compared to higher values.

Hence the variance of the residuals are not equal and it demonstrates heteroscedasticity

Q 1.6 - b

1. It is a right skewed distribution
2. The left tail has less proportion of data than that of a normal distribution
3. Between 40-80 % range the distribution has much less proportion of data compared to a normal distribution

From observing the P-P plot we conclude there is considerable difference between this distribution and normal distribution.

Q 1.6 - c

Based on the above we can conclude that this regression equation may not be functionally correct. It may not be correct to rely on predictions using this model.

Q 1.7

The increase in R-squared when a new variable is added to a model is the given by the **Square of the Semi-Partial (PART) Correlation**.

- From Table 1.7: R-squared @ Step 2 = 0.542
- From Table 1.8: PART Correlation for adding RES = -.153

```
R-squared in Step 3 is 0.565409
```

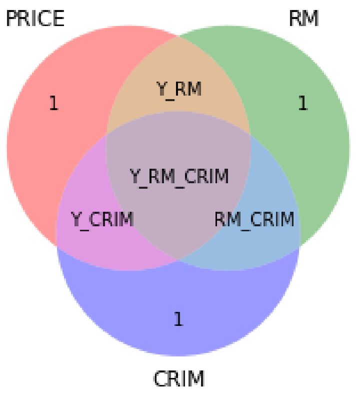
Q 1.8

It reduces as there is correlation among RM and CRIM. Part of what was explained by RM in model 1 is now being explained by CRIM in model 2 as CRIM and RM is correlated.

Technically this is call Omitted Variable Bias. The reduction can be explained by the following equation:

$$\alpha_{RM_{Model1}} = \beta_{RM_{Model2}} + \frac{\beta_{CRIM_{Model2}} * Cor(RM, CRIM)}{Var(RM)}$$

From Correlation table we see that RM and CRIM has negative Correlation, hence the overall value of coefficient for RM reduces



Q 1.9

We will use the model in step - 6 for answering this question.

- Since the variables are not standardised we cannot use the magnitude of the coefficients as a measure of impact on dependent variable (Price)
- We will use the notion of the Standardised Coefficients to measure how much 1 SD change in the variable X (Predictor) changes Y (dependant)
- From Tables 1.1 and 1.8 we can easily obtain the Standardised Coefficients for the regression model for all variables except for RM as the SD of RM is not provided in table 1.1 and the Standardised coefficient of RM is not provided in table 1.8. Standardised Coefficient is calculated using:

$$\beta_{STANDARDISED} = \hat{\beta} * \frac{S_X}{S_Y}$$

where

$$\text{Standard Deviation X} = S_X$$

&

$$\text{Standard Deviation Y} = S_Y$$

- To calculate the variance of RM we will use the Model 1
- In Model 1 the coefficient of RM is 9.102
- Standardized Coefficient of RM = .695, SD of PRICE (Y) = 9.197
- Using these values and rearranging the equation discussed above, we get SD of RM = .7022
- From the below table we can see that **RM** has the highest impact on PRICE.

Out[82]:

	_	Coefficients	Standardized Coefficients
0	INTERCEPT	-8.993	
1	RM	7.182	0.548353
2	CRIM	-0.194	-0.18144
3	RES	-0.318	-0.238
4	SEZ	4.499	0.124
5	Highway	-1.154	0.264
6	AGE	-0.077	-0.235671

Q 2.1

Correct:

1. The model explains 42.25% of variation in box office collection.

2. There are outliers in the model.

3. The residuals do not follow a normal distribution.

Incorrect:

4.The model cannot be used since R-square is low.

5.Box office collection increases as the budget increases.

Q 2.2

Here Budget (X) can never be = 0, as it may not be possible to produce a movie without money and X = 0 is unobserved i.e. X = 0 falls outside the domain of the observed values of the variable X. The relationship between the variables can change as we move outside the observed region. The Model explains the relationship between Y and X within the range of observed values only. We cannot predict for a point that is outside the range of observed values using the regression model.

Hence Mr Chellapa's observation is incorrect

Q 2.3

Since the variable is insignificant at alpha = 0.05, hence the coefficient may not be different from zero. There is is no statistical validity that the collection of movie released in Releasing_Time Normal_Season is different from Releasing_Time Holiday_Season (which is factored in the intercept / constant).

Since we do not have the data hence we cannot rerun the model without the insignificant variable. We will assume that the co-efficient is 0 and it's removal does not have any effect on the overall equation (other significant variables).

Hence the difference is **Zero**.

Q 2.4

The beta for Release Normal Time is being considered as 0 as it is statistically insignificant at alpha. Hence it will be factored in the Intercept term. Releasing_Time Long_Weekend is statistically significant and the coefficient = 1.247.

The range of values will be considered because of variability of the coefficient.

SE =0.588, tCrit @ 0.05 = 1.964 Max Value = Constant + tcrit *SE MIn Value = Constant - tcrit* SE

```
Max earning from Long weekend movie releases can be 161.87622500117592
Min earning from Long weekend movie releases can be 16.073436458805958
Movies released in normal Weekends may earn on Average 14.658201380262703
Movies released in Long Weekends may or may not earn at least 5 Cr more than movies released in normal season as the min di
fference is around 2 Cr
Mr. Chellapa's statement is incorrect.
```

Q 2.5

The increase in R-squared when a new variable is added to a model is the given by the **Square of the Semi-Partial (PART) Correlation**.

The assumption here is the variable "Director_CAT C" was the last variable added to model at Step 6. We have to make this assumption as variables added in prior stages are not available.

- From Table 2.5 : R-squared @ Step 5 = 0.810 ** 2 = .6561
- From Table 2.6: PART Correlation for adding Director_CAT C = -.104

```
R-squared in Step 3 is 0.6669160000000001
```

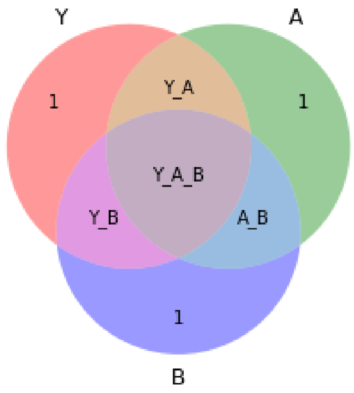
Q2.6

- Budget_35_Cr is the highest impact on the performance of the movie. On average a move with budget exceeding 35 Cr adds 1.53 Cr extra than a move with lesser budget.
- Recommendation: Use high enough budget to:

```
- Hire Category A Production House
- Do not hire Category C Director
- Do not hire Category C Music Director
- Produce a Comedy movie
```

Q 2.7

- We cannot say that the variables have no relationship to Y (BOX Office Collection)
- We can conclude that in presence of the other variables the variables in Model 2 are not explaining additional information about Y



From chart above we can see that as we add new variables (A, B) it explains variations in Y. The explained variation in Y due to addition of a new variable should be significant enough. This is measured by:

1. t-test for individual variable
2. Partial F-test for the models generated consecutively

We may conclude that the variables of Model 2 may not be explaining significant variations in Y in presence of the additional variables added later on and hence was dropped.

Q 2.8

We are making the assumption that the variable Youtube views imply views of the actual movie and not the trailers before movie release dates. The following explanation will not be valid in that case. Also, we are assuming that revenue collected from advertisements during Youtube views do not fall under the Box Office Collection.

Youtube_Views = Will not contribute anything meaningful functionally to the Box Office collection as the movie has been created and released in theaters and all possible collection is completed. The main essence of the prediction here is to understand before making a movie, what all factors may lead to better revenue collection for a movie

Q 3.1

Table 3.1

- **Observations** (N) = 543
- **Standard Error**
  - $$SE = \sqrt{\frac{\sum_{k=1}^{N} (Y_k - \hat{Y}_k)^2}{N-2}}$$

$$(Y_k - \hat{Y}_k)^2 = \epsilon_k^2 = \text{Residual SS (SSE)} = 17104.06 \text{ (Table 3.2)}$$

- **R-Squared** = 1 - SSE / SST
  - SSE = 17104.06 (Table 3.2)
  - SST = 36481.89 (Table 3.2)

- **Adjuated R-Squared** = 1 - (SSE / N-k-1) / (SST/N-1)
  - N = 543
  - K = 3

- **Multiple R** =

$$\sqrt{R}_{Squared}$$

Out[15]:

	Regression Statistics	_
0	Multiple R	0.728809
1	R Square	0.531163
2	Adjusted R Squared	0.528554
3	Standard Error	5.622778
4	Observations	543.000000

Table 3.2

- **DF Calculation**
  - DF for Regression (K) = Number of variables = 3
  - DF for Residual = N - K - 1 = 539

- **SS Calculation**
  - Residual SS (SSE) = 17104.06 (given)
  - Total SS (TSS)= 36481.89 (given)
  - Regression SS (SSR) = TSS - SSE = 19377.83

- **MS Calculation**
  - MSR (Regression) = SSR / DF for SSR (=3)
  - MSE (Error) = SSE / DF for SSE (= 539)

- **F Claculation**
  - F = MSR / MSE

Out[16]:

	_	DF	SS	MS	F	SignificanceF
0	Regression	3	19377.83	6459.28	203.552	1.11022e-16
1	Residual	539	17104.06	31.7328		
2	Total	542	36481.89			

Table 3.3 - Coefficients

- MLR T-Test
  - $$t_i = \frac{\beta_i - 0}{Se(\beta_i)}$$

    where i denotes the different variables (here i = 3)

Out[17]:

	_	Coefficeints	Standard Error	t Stat	P-Value	Lower 95%	Upper 95%
0	Intercept	38.592350	0.937225	41.177252		36.751290	40.433411
1	Margin	0.000053	0.000002	24.403670		0.000049	0.000057
2	Gender	1.551306	0.777806	1.994464		0.023404	3.079208
3	College	-1.475060	0.586995	-2.512900		-2.628140	-0.321978

Q 3.2

From the table above we see that for all the variables the t-value > 1.964. hence all the variables are significant. 1.964 = Critical value of t @ significance 0.05

Q 3.3

F-distribution with DF = 3, 539 at significance = 95% is 2.621. Hence the model is significant.

Out[18]:

```
2.621440036586561
```

Q 3.4

The increase in R-squared when a new variable is added to a model is the given by the **Square of the Semi-Partial (PART) Correlation**.

- R-squared for Model 2 = 0.52567 (R1)
- R-squared for Model 3 = 0.531163 (R2)

Part Correlation of College & % Votes =

$$\sqrt{R_2 - R_1}$$

```
Increase in R-Squared due to adding College = 0.005493000000000081
Part Correlation of College & % Votes = 0.07411477585475167
```

Q 3.5

We will conduct Partial F-test between models to test for significance of each model. We make the assumption that the variables added are significant at each step (model) at alpha 0.05

$$F_{PARTIAL} = \frac{\frac{R^2_{FULL} - R^2_{PARTIAL}}{k-r}}{\frac{1-R^2_{FULL}}{N-k-1}}$$

where k = variables in full model, r = variables in reduced model, N = Total number of records

```
Model 3 Partial F 4.358744633992214
Model 3 Critical F at Df = (1, 539) 0.03726112108923041
Model 4 Partial F 25.131765753275783
Model 4 Critical F at Df = (1, 539) 7.281767351319246e-07
Model 5 Partial F 19.291406535763336
Model 5 Critical F at Df = (1, 539) 1.3522586193692732e-05
```

```
Hence we can see that all the models are significant. The number of features (5) are not very high, hence we conclude it's
justified to add the additional variables
```

Q 3.6

- Since the variables are not standardised we cannot use the magnitude of the coefficients as a measure of impact on dependent variable (Vote %)
- We will use the notion of the Standardised Coefficients to measure how much 1 SD change in the variable X (Predictor) changes Y (dependant)
- Using Table 3.5 and equations below we will compute Standardised Coefficient:

$$\beta_{STANDARDISED} = \hat{\beta} * \frac{S_X}{S_Y}$$

where

$$\text{Standard Deviation X} = S_X$$

&

$$\text{Standard Deviation Y} = S_Y$$

- From the below table we can see that **MARGIN** has the highest impact on Vote %. 1 SD change in Margin changes .75 SD in Vote %

Out[21]:

	_	Coefficients	Standard deviation	Standardized Coefficients
0	INTERCEPT	38.569930		
1	MARGIN	0.000056	111366	0.757437
2	Gender	1.498308	0.311494	0.0568868
3	College	-1.537740	0.412796	-0.0773712
4	UP	-3.714390	0.354761	-0.160614
5	AP	5.715821	0.209766	0.146142

Q 4.1

```
Total Positives: 1045  ::  Total Negatives: 955  ::  Total Records: 2000
P(Y=1) = positives / N = 0.5225  ::  P(Y=0) = negatives /N = 0.4775
-2LL0 = 2768.5373542564103
```

- -2LLo is called the "Null Deviance" of a model. It is -2 Log Likelihood of a model which had no predictor variables. Hence we obtain the probabilities of positive and negative in the dataset using the frequencies for such model.
- After adding "Premium" 2LL reduces to 2629.318 (Table 4.2). Hence reduction is equal to (-2LLo -(-2LLm)):

```
139.2189999999996
```

Q 4.2

```
True Positive :Actually Positive and Predicted Positive = 692
False Positive :Actually Negative and Predicted Positive = 204
Precision = True Positive /  (True Positive + False Positive)  = 0.7723214285714286
```

Q 4.3

exp(B) = change in odds ratio. The odds ratio can be interpreted as the multiplicative adjustment to the odds of the outcome, given a **unit** change in the independent variable. In this case the unit of measurement for Premium (1 INR) which is very small compared to the actual Premium (1000s INR), hence a unit change does not lead to a meaningful change in odds ratio, subsequently the odds ratio will be very close to one.

Q 4.4

Assumptions: Actual Data was not available. Decision would be made based on outcome of Model results

```
The model predicts 751 + 353 = 1104 customers have a probability less than 0.5 of paying premium
They will call 1104 customers through Call Center
```

Q 4.5

Total points we are getting is 1960.

total = tp + fp + fn + tn

Formula :

sensitivity = tp/ (tp + fn)

specificity = tn / (tn + fp)

recall = sensitivity precision = tp / (tp + fp)

f-score = 2 * precision * recall / (precision + recall)

```
Number of records ::
Precision 0.75 ::
Recall 0.05555555555555555 ::
sensitivity 0.05555555555555555 ::
specificity 0.9772727272727273 ::
f-score 0.10344827586206895
```

Q 4.6

Probability of Y==1 can be calculated using the following formula:

$$P(Y=1) = \frac{\exp^{z}}{1+\exp^{z}}$$

$$\text{where z} = \beta_0 + \beta_1 * Salaried + \beta_2 * HouseWife + \beta_3 * others$$

However in this case the variable Housewife is not a significant variable. Hence using this equation to calculate probability for the variable house wife may not be appropriate. We will procced to compute the probability using the equation but will consider the coefficient of Housewife as 0 (B is not significantly different from 0 for insignificant variables). Ideally we need to rerun the Model removing the insignificant variable, but since we do not have the data we will use the same equation and assume the coefficients for the other variables will not change if we had removed Housewife.

```
Probability of House wife paying the Premium is (beta ==22.061): 0.9999999993810521
Probability of House wife paying the Premium is (beta = 0): 0.29775737226938176
Since Beta is insignificant B == 0, hence .298 is the probability for housewife paying renewal
```

Q 4.7

The Constant / Intercept measures for people with the following occupations **Professionals, Business and Agriculture** and they have a lower probability of renewal payment. From Model 3 - Coefficient of intercept is negative, hence our conclusion

Q 4.8

Probability can be calculated using the following formula:

$$P(Y=1) = \frac{\exp^{z}}{1+\exp^{z}}$$

$$\text{where z} = constant + \beta_1 * PolicyTerm$$

The regression equations reduces to the simple term as shown above because SSC Education, Agriculturist Profession & Marital Status Single will be factored in the term constant of the given equation and the remainder of the variable will be Zero.

```
Probability : 0.824190402911071
```

Q 4.9

The coefficients tell about the relationship between the independent variables and the dependent variable, where the dependent variable is on the logit scale. These estimates tell the amount of increase in the predicted log odds that would be predicted by a 1 unit increase in the predictor, holding all other predictors constant.

Findings:

- Married People have higher possibility of renewals (log odds ratio increases)
- As payment term increases it leads to slightly reduced log odds of renewals
- Professionals, Business men have much higher chance of defaulting on log odds of renewals
- Being a graduate does increase the chance of payment of renewals (log odds)
- Annual / Half yearly / Quarterly policy renewal schemes see reduced payment of renewals (log odds)
- Model Change - Premuim : Variable scale should be changed for better understanding of Premium's contribution to affinity to renew policy (may be reduce unit to 1000s)

Recommendations :

- For new customers target Married people and graduates
- For existing customers send more reminders (via Call centers / messgaes etc) to Business men, Professionals for renewal
- For people paying premiums in yearly / quarterly / halfyearly terms, send reminders to them before renewal dates
- For people with long payment terms keep sending them payment reminders as the tenure of their engagement advances

Q 4.10

The bins are computes as following:

- Decile=1 = 0 -.1 (both inclusive)
- Decile=.9 = 1.00001 - .2 (both incusive and so on)
- upto Decile1

We arrange the table in descending order of probabilities, i.e. Decile=.1 contains .90001 till 1 probability values, Decile=.2 contain .800001 till 0.9 pronbability values.

Gain is calculated as:

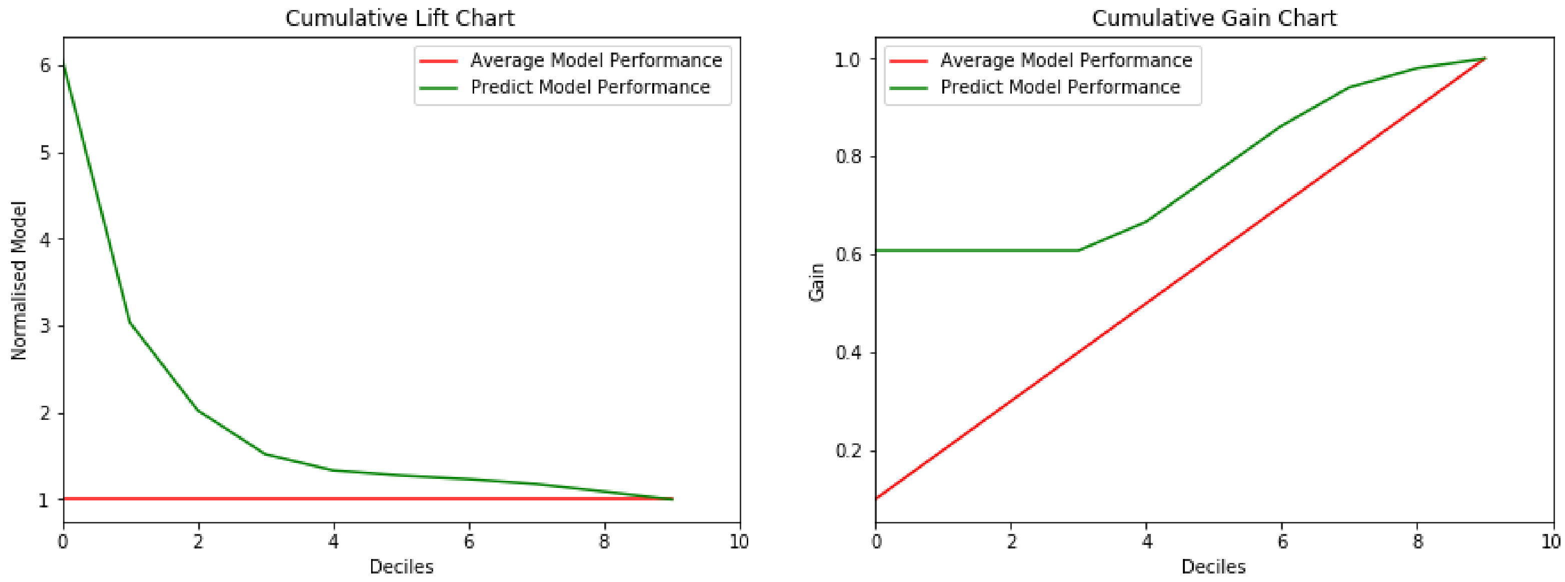
$$gain = \frac{\text{cumulative number of positive obs upto decile i}}{\text{Total number of positive observations}}$$

Lift is calculated as:

$$lift = \frac{\text{cumulative number of positive obs upto decile i}}{\text{Total number of positive observations upto decile i from random model}}$$

Out[29]:

	Decile	posunits	negunits	posCountunits	negCountunits	avgCountunits	cumPosCountunits	cumAvgCountunits	lift	gain	avgLift
0	0.1	31	0	620	0	102.0	620	102.0	6.078431	0.607843	1
1	0.2	0	0	0	0	102.0	620	204.0	3.039216	0.607843	1
2	0.3	0	0	0	0	102.0	620	306.0	2.026144	0.607843	1
3	0.4	0	0	0	0	102.0	620	408.0	1.519608	0.607843	1
4	0.5	3	0	60	0	102.0	680	510.0	1.333333	0.666667	1
5	0.6	5	5	100	100	102.0	780	612.0	1.274510	0.764706	1
6	0.7	5	11	100	220	102.0	880	714.0	1.232493	0.862745	1
7	0.8	4	17	80	340	102.0	960	816.0	1.176471	0.941176	1
8	0.9	2	12	40	240	102.0	1000	918.0	1.089325	0.980392	1
9	1.0	1	2	20	40	102.0	1020	1020.0	1.000000	1.000000	1



Observaions

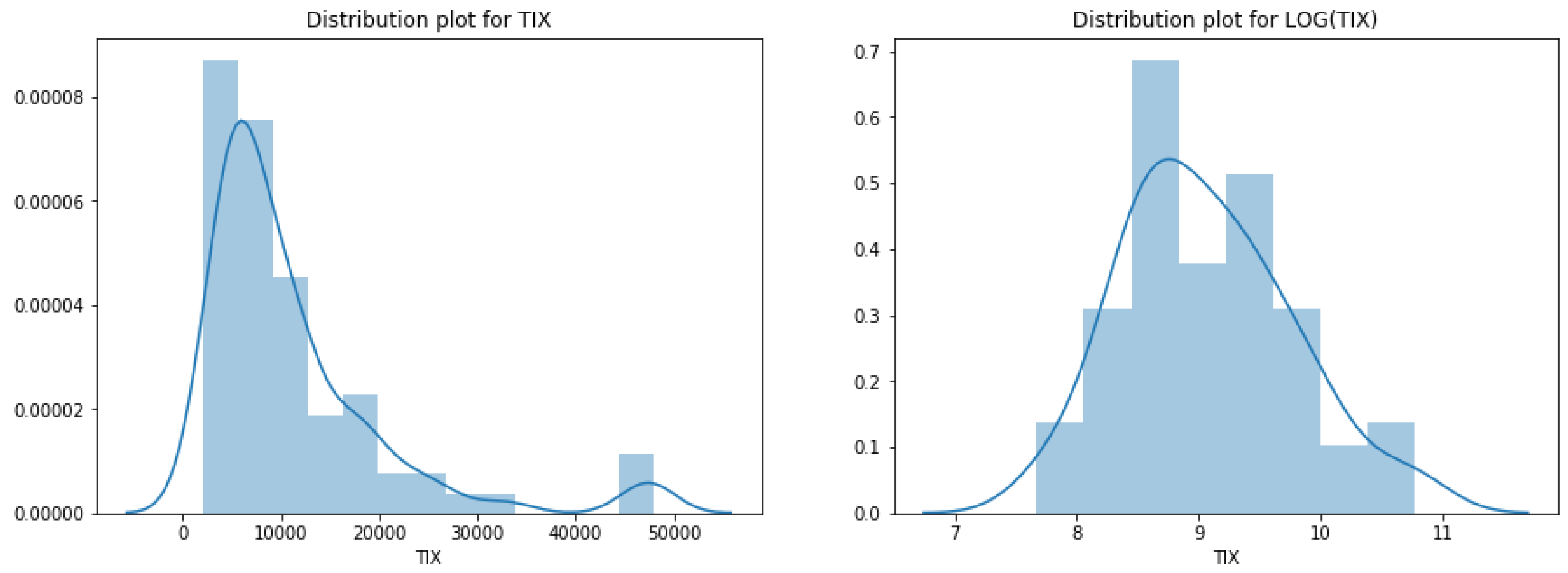
- From gain we see that the model captures 76% positives by the fifth decile
- From Lift we see for the 1st decile model captures 6 times more positives than an ordinary model, 3 times for second decile, 2 times for 3rd decile, 1.5 times for 4th decile and 1.27 times for the 5th decile

Q 5

```
There are no Missing Values in Data
```

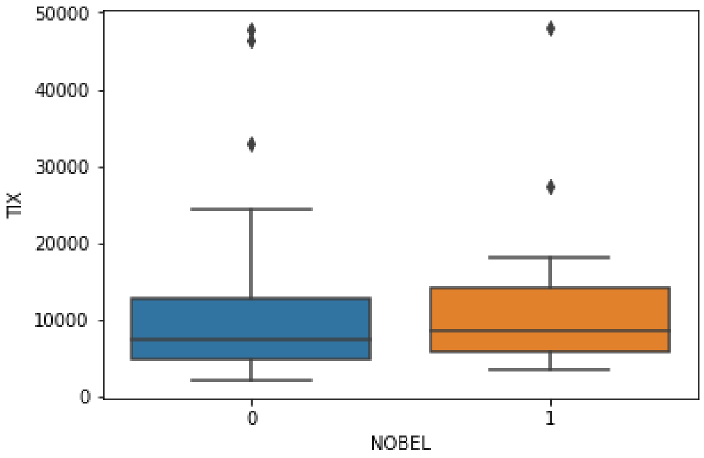
Out[32]:

	NUMBER	TIX	OPP	POS	GB	DOW	TEMP	PREC	TOG	TV	PROMO	NOBEL	YANKS	WKEND	OD
count	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.00000	75.000000	75.000000	75.000000	75.000000	75.
mean	38.000000	11244.253333	7.053333	2.853333	8.760000	4.240000	62.026667	0.040000	1.480000	0.12000	0.173333	0.213333	0.066667	0.520000	0.
std	21.794495	9729.863870	3.834034	1.342581	6.064607	2.058864	3.324655	0.197279	0.502964	0.32715	0.381084	0.412420	0.251124	0.502964	0.
min	1.000000	2140.000000	1.000000	1.000000	0.000000	1.000000	55.000000	0.000000	1.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.
25%	19.500000	5151.500000	4.000000	2.000000	2.000000	2.000000	60.000000	0.000000	1.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.
50%	38.000000	7696.000000	7.000000	3.000000	11.000000	5.000000	62.000000	0.000000	1.000000	0.00000	0.000000	0.000000	0.000000	1.000000	0.
75%	56.500000	13026.000000	10.000000	3.000000	13.000000	6.000000	64.000000	0.000000	2.000000	0.00000	0.000000	0.000000	0.000000	1.000000	0.
max	75.000000	47946.000000	13.000000	7.000000	19.000000	7.000000	70.000000	1.000000	2.000000	1.00000	1.000000	1.000000	1.000000	1.000000	1.



```
TIX is right skewed distribution. The log Transformed TIX is more of an approximate normal distribution.
```

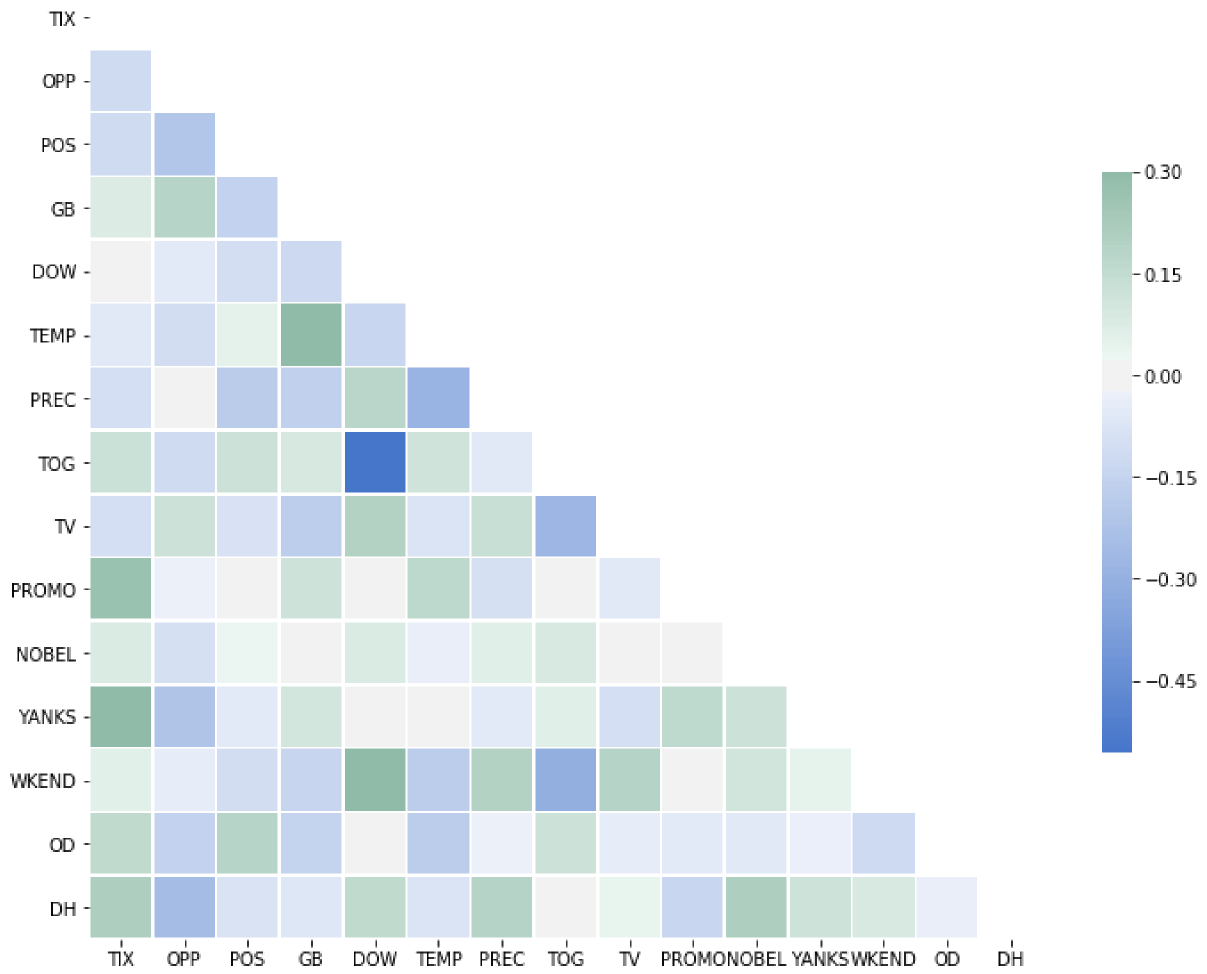
- Mark Nobel has played for 21.33% games for Oakland A during the period when the data was captured
- We will perform a Two Sample T-test between the mean of TIX when Nobel Played vs When Nobel did not play to check wthether ther was any significant difference of mean between the two categories



```
Alpha: 0.05
Mean TIX (x1) = 12663.5625, STD TIX = 11211.620411987733 and number of games = 16 with Nobel
Mean TIX (x1) = 10859.35593220339, STD TIX = 9357.940067245701 and number of games = 59 without Nobel
NUll Hypothesis: x1 - x2 <= 0
Alternate Hypothesis: x1 - x2 >0
This is 2 Sample T test, with unknown population SD and the SD of the two are unequal
SE 3056.228389809929
DF 21
T-stat 0.5903376114861676
This is a two sided T-Test
Significant t-value at alpha - 0.05 is : 1.7207429028118777
p-value:0.2806319971052741 is greater than alpha(0.05)
Hence we can retain the NULL Hypothesis (ho)
```

- In general we see that there is not statistical evidence that a single factor, presence of Nobel has any effect on increasing ticket sales
- We will check whether this factor become important in presence of other factors before drawing any final conclusions

```
Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2c477afe80>
```

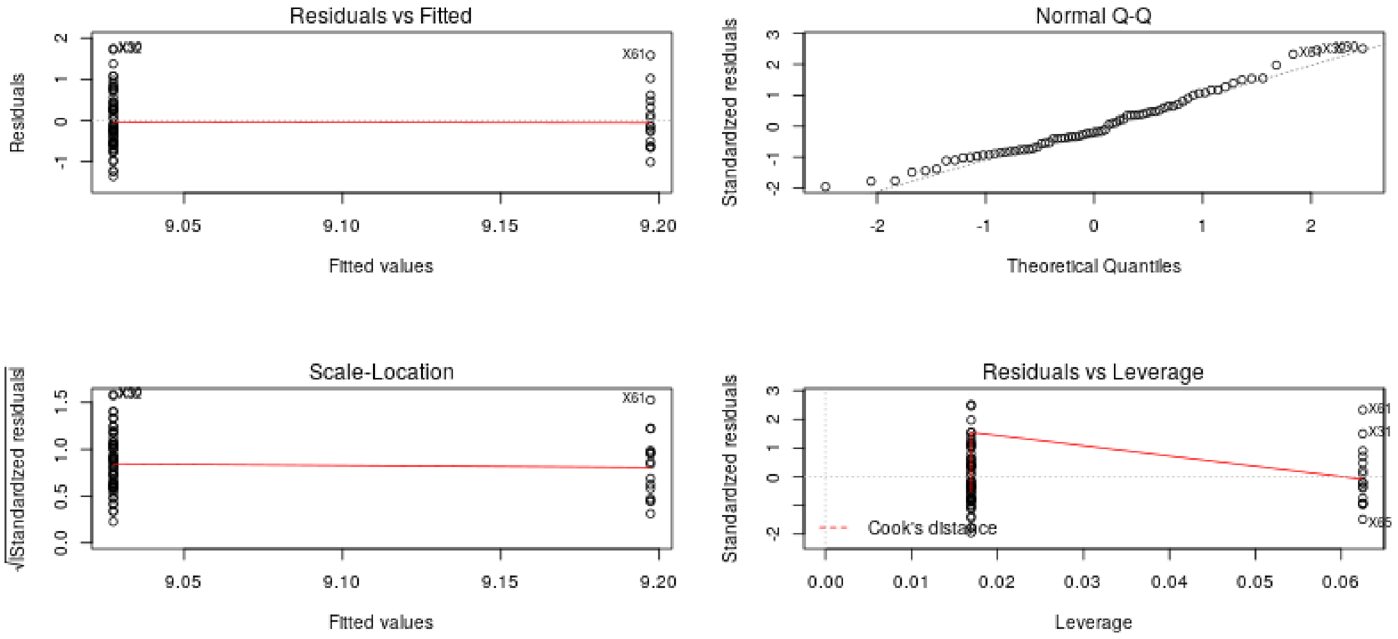


- From the correlation plot above we see that "Game with YANKS" and PROMO along with whether the match is a "DOUBLE HEADER" has high correlation to TIX sales

We will now create a series of Regression Models to check the validity of the claim that MARK NOBEL's presence increase the TIX and revenue generation for OAKLAND A

- From the plots of TIX we noticed that TIX is not normally distributed. The Regression Model developed with TIX may end up with Error terms which are not Normally distributed
- To address this issue we will build the models using the Log transformed values of TIX, as from the plot it is clear that the log transformed variable is closer to a Normal Distribution.

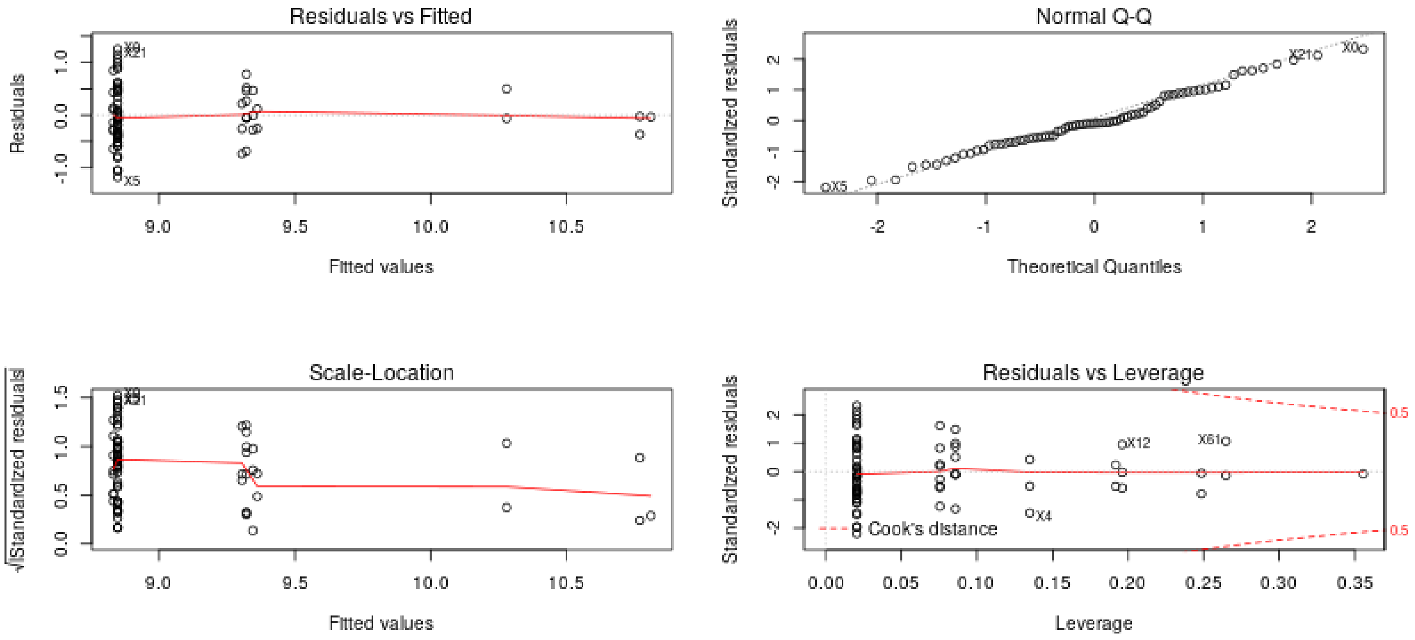
```
Call:
lm(formula = .outcome ~ ., data = dat, verbose = FALSE)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.3590 -0.5277 -0.1325  0.4293  1.7465 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.02758    0.09137  98.807   <2e-16 ***
x1           0.16970    0.19781   0.858    0.394    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7018 on 73 degrees of freedom
Multiple R-squared:  0.009981,	Adjusted R-squared:  -0.003581 
F-statistic: 0.7359 on 1 and 73 DF,  p-value: 0.3938

NULL
```



- As noticed with the Hypothesis test, from the model above we can see that on its own the variable checking for the presence of Nobel is not Significant in predicting for TIX

We will build a Model with NOBEL, YANKS, DH and PROMO

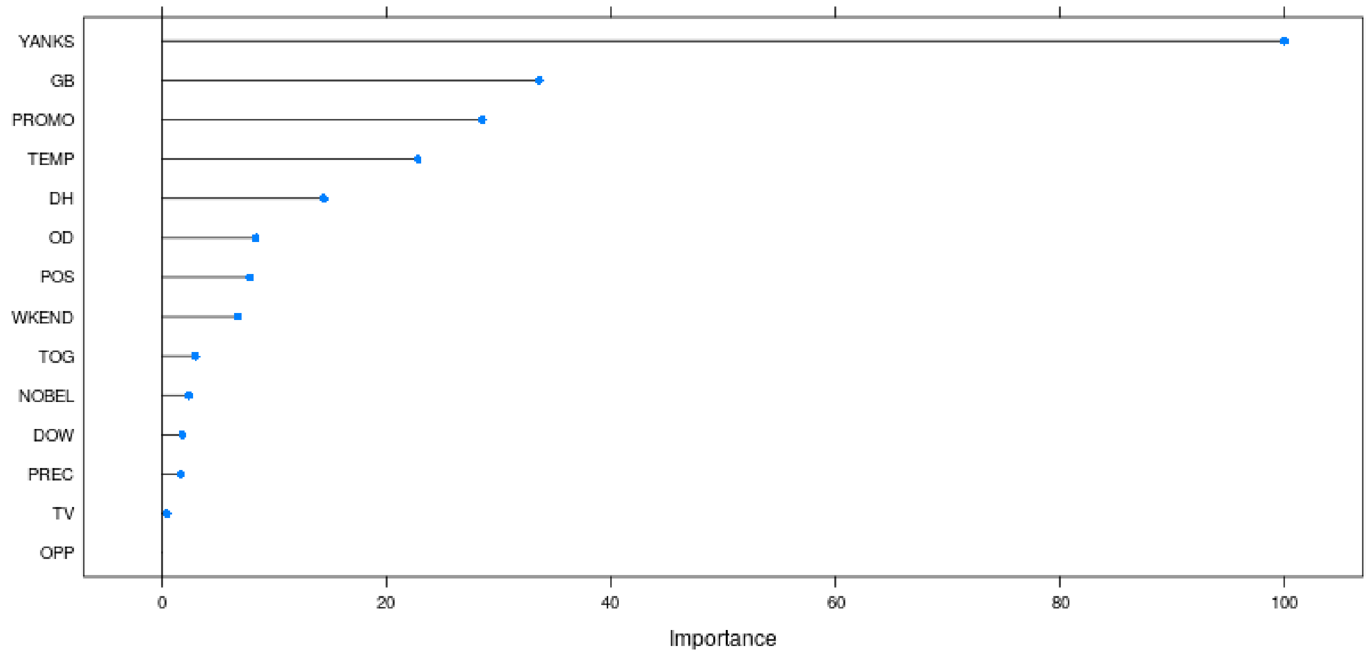
```
Call:
lm(formula = .outcome ~ ., data = dat, verbose = FALSE)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.17886 -0.34759 -0.05221  0.43435  1.25553 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.84742    0.07830 112.993  < 2e-16 ***
NOBEL1      -0.01655    0.15781  -0.105   0.9168    
YANKS1       1.44901    0.25925   5.589 4.09e-07 ***
DH1          0.51414    0.24118   2.132   0.0365 *  
PROMO1       0.47397    0.17055   2.779   0.0070 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5448 on 70 degrees of freedom
Multiple R-squared:  0.4279,	Adjusted R-squared:  0.3953 
F-statistic: 13.09 on 4 and 70 DF,  p-value: 5.174e-08

NULL
```

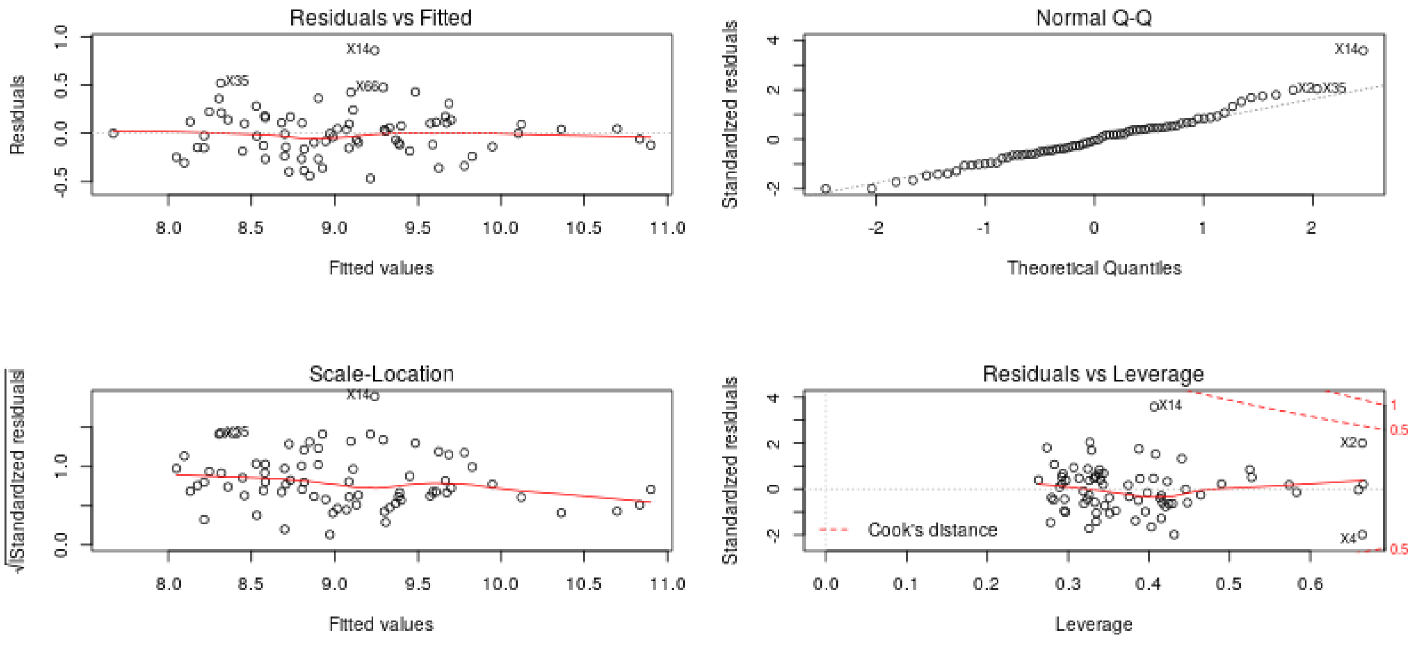


- As noticed with the Hypothesis test, from the model above we can see that the variable checking for the presence of Nobel is not Significant in predicting for TIX

We will build a Stepwise Model with all variables and select the best model. If the variable NOBEL is significant it will be added by the STEPWISE Selection Algorithm

```
Start:  AIC=-137.62
.outcome ~ OPP + POS + GB + DOW + TEMP + PREC + TOG + TV + PROMO + 
    NOBEL + YANKS + WKEND + OD + DH


Step:  AIC=-137.62
.outcome ~ OPP + POS + GB + DOW + TEMP + PREC + TOG + TV + PROMO + 
    NOBEL + YANKS + OD + DH


Step:  AIC=-137.62
.outcome ~ OPP + POS + GB + DOW + TEMP + PREC + TOG + TV + PROMO + 
    NOBEL + OD + DH

        Df Sum of Sq    RSS      AIC
- GB    16    1.0683 4.3093 -148.253
- POS    5    0.2947 3.5357 -141.094
- NOBEL  1    0.0015 3.2425 -139.586
- TEMP   1    0.0102 3.2512 -139.384
- TOG    1    0.0302 3.2712 -138.925
- TV     1    0.0854 3.3264 -137.670
<none>               3.2410 -137.621
- PREC   1    0.2414 3.4824 -134.232
- PROMO  1    0.8455 4.0865 -122.236
- OD     1    0.9411 4.1821 -120.501
- DH     1    0.9740 4.2150 -119.913
- OPP   12    6.0107 9.2517  -82.951
- DOW    6    4.8968 8.1378  -80.572

Step:  AIC=-148.25
.outcome ~ OPP + POS + DOW + TEMP + PREC + TOG + TV + PROMO + 
    NOBEL + OD + DH

        Df Sum of Sq     RSS      AIC
- TOG    1    0.0005  4.3098 -150.244
- NOBEL  1    0.0053  4.3146 -150.161
- TEMP   1    0.0125  4.3218 -150.036
<none>                4.3093 -148.253
- TV     1    0.1677  4.4771 -147.389
- PREC   1    0.2532  4.5625 -145.971
- POS    6    1.3336  5.6429 -140.031
- OD     1    0.9311  5.2404 -135.581
- PROMO  1    1.5890  5.8983 -126.711
- DH     1    1.6242  5.9335 -126.265
- DOW    6    6.6753 10.9846  -90.075
- OPP   12   12.6881 16.9974  -69.332

Step:  AIC=-150.24
.outcome ~ OPP + POS + DOW + TEMP + PREC + TV + PROMO + NOBEL + 
    OD + DH

        Df Sum of Sq     RSS      AIC
- NOBEL  1    0.0051  4.3149 -152.156
- TEMP   1    0.0133  4.3231 -152.014
<none>                4.3098 -150.244
- TV     1    0.1683  4.4781 -149.371
- PREC   1    0.2527  4.5625 -147.971
- POS    6    1.3522  5.6621 -141.777
- OD     1    1.0705  5.3803 -135.605
- PROMO  1    1.5921  5.9020 -128.665
- DH     1    1.6913  6.0012 -127.415
- DOW    6    6.9278 11.2376  -90.367
- OPP   12   12.9542 17.2641  -70.165

Step:  AIC=-152.16
.outcome ~ OPP + POS + DOW + TEMP + PREC + TV + PROMO + OD + 
    DH

        Df Sum of Sq     RSS      AIC
- TEMP   1    0.0145  4.3294 -153.905
<none>                4.3149 -152.156
- TV     1    0.1684  4.4833 -151.285
- PREC   1    0.2522  4.5671 -149.896
- POS    6    1.3484  5.6633 -143.761
- OD     1    1.0704  5.3853 -137.537
- PROMO  1    1.5871  5.9020 -130.665
- DH     1    1.7216  6.0365 -128.975
- DOW    6    6.9499 11.2648  -92.185
- OPP   12   12.9817 17.2966  -72.023

Step:  AIC=-153.91
.outcome ~ OPP + POS + DOW + PREC + TV + PROMO + OD + DH

        Df Sum of Sq     RSS      AIC
<none>                4.3294 -153.905
- TV     1    0.1751  4.5044 -152.932
- PREC   1    0.2801  4.6095 -151.203
- POS    6    1.3866  5.7159 -145.067
- OD     1    1.0859  5.4153 -139.119
- PROMO  1    1.7217  6.0511 -130.794
- DH     1    1.8443  6.1737 -129.289
- DOW    6    6.9391 11.2684  -94.161
- OPP   12   13.0650 17.3944  -73.601

Call:
lm(formula = .outcome ~ OPP + POS + DOW + PREC + TV + PROMO + 
    OD + DH, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.46633 -0.14812 -0.00399  0.11824  0.85710 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.09003    0.26331  34.522  < 2e-16 ***
OPP2         0.44416    0.20678   2.148 0.037133 *  
OPP3         0.30877    0.25620   1.205 0.234437    
OPP4         1.73429    0.21481   8.073 2.70e-10 ***
OPP5         0.62308    0.26591   2.343 0.023598 *  
OPP6         0.21812    0.20011   1.090 0.281516    
OPP7         0.37260    0.23989   1.553 0.127375    
OPP8         0.04801    0.20888   0.230 0.819247    
OPP9         1.20131    0.22159   5.421 2.24e-06 ***
OPP10        0.52861    0.20772   2.545 0.014435 *  
OPP11        0.68907    0.25195   2.735 0.008893 ** 
OPP12        0.13104    0.21979   0.596 0.554029    
OPP13        0.70907    0.20678   3.429 0.001307 ** 
POS2         0.07531    0.19319   0.390 0.698512    
POS3        -0.13506    0.16212  -0.833 0.409191    
POS4         0.08729    0.22849   0.382 0.704247    
POS5         0.17222    0.21214   0.812 0.421152    
POS6        -0.38361    0.41983  -0.914 0.365735    
POS7        -0.66226    0.31119  -2.128 0.038839 *  
DOW2        -1.03787    0.15275  -6.794 2.06e-08 ***
DOW3        -1.08154    0.14853  -7.282 3.90e-09 ***
DOW4        -0.99885    0.22516  -4.436 5.85e-05 ***
DOW5        -0.58277    0.15017  -3.881 0.000337 ***
DOW6        -0.56788    0.14304  -3.970 0.000256 ***
DOW7        -0.49177    0.14823  -3.318 0.001804 ** 
PREC1       -0.36312    0.21281  -1.706 0.094841 .  
TV1         -0.20025    0.14845  -1.349 0.184096    
PROMO1       0.49874    0.11789   4.230 0.000113 ***
OD1          1.39539    0.41533   3.360 0.001598 ** 
DH1          0.72698    0.16604   4.378 7.05e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3102 on 45 degrees of freedom
Multiple R-squared:  0.8808,	Adjusted R-squared:  0.804 
F-statistic: 11.46 on 29 and 45 DF,  p-value: 9.674e-13

NULL
```





From the models created above including building the stepwise regression model and the analysis done above we can see that presence of Nobel is not Significant in increasing Ticket Sales and Revenue collected from Ticket sales. He Does not have any contribution to increased revenue colection due to ticket Sales.

Q6

Q6-1

- NPS is a KPI which is used by many organizations to understand and measure customer satisfaction
- Organizations also believe that it is important for every organization to know what their customers tell their friends about the organization. NPS is considered by many organizations as a measurement of whether a customer will recommend the company or product/service to a friend or colleague

Business Problem

- Managment at Manipal Hospitals belived that loyalty in healthcare depends on technical and emotional aspects
- Happy customer may lead to new business, unhapy customers may lead to lack of new business / erosion of exising business
- Through NPS forms they wanted to collect customer feedback and sentiments
- By analysing the NPS data they also wanted to understand the reasons that led to the customer giving such a NPS score
- They wanted to analyse the reasons that would help to resolve the issues and then keeping the customers informed about the corrective actions; they believed they could improve the customer satisfaction and hence the NPS by such action

How Analytics can help with the Problem

- The historical paper based feedback when conevrted into digital data and the digital data captured post March 2014 can be analysed using analytics to derive insights
- By analysing past data, analytics can help unearth patterns in data that may be related to high or low customer statisfaction and NPS
- These patterns can be formualted into prescriptive actions which can help improve the process for the future there by improving the overall customer satisfaction and better NPS
- If analytics can help link customer demographics / behaviour to NPS then hospital can devise different startegies for different customer profiles, which also can lead to better NPS and satisfied customer

Q6-2

Sensitivity, Specificity for a multinomial / 3-class problem can be calculated in the following manner. We will elaborate the method using the following tables and derive the formula for the metrics.

total records = tp + fp + fn + tn

For 2-class the following are the definition for sensitivity and specificity:

sensitivity = tp/ (tp + fn)

specificity = tn / (tn + fp)

where tp = True positive fp = False Postive tn = True Negative fn = False Negative

The definition for Specificity / sensitivity does not change from the above in 3-class scenario. The way we compute the tp, tn, fp and fn changes. We will demonstrate the same below.

Lets say we have 3 classes A, B, C.

Step 1: We will construct the Confusion Matrix for "A". Table belows shows FP1 and FP2 etc. information. Here :

```
fp = FP1 + FP2

fn = FN1 + FN2

tn = Sum(X)
```

The formula for the metrics changes to:

sensitivity = tp/ (tp + fn1 + fn2)

specificity = tn / (tn + fp1 + fp2)

Out[43]:

		Predcited		
		A	B	C
Actual	A	TP	FN1	FN2
	B	FP1	X	X
	C	FP2	X	X

Step 2: We will construct the Confusion Matrix for "B". Table belows shows FP1 and FP2 etc. information. Here:

```
fp = FP1 + FP2

fn = FN1 + FN2

tn = sum(X)
```

The formula for the metrics changes to:

sensitivity = tp/ (tp + fn1 + fn2)

specificity = tn / (tn + fp1 + fp2)

Out[44]:

		Predcited		
		A	B	C
Actual	A	X	FP1	X
	B	FN1	TP	FN2
	C	X	FP2	X

Step 3: We will construct the Confusion Matrix for "C". Table belows shows FP1 and FP2 etc. information. Here :

```
fp = FP1 + FP2

fn = FN1 + FN2

tn = sum(X)
```

The formula for the metrics changes to:

sensitivity = tp/ (tp + fn1 + fn2)

specificity = tn / (tn + fp1 + fp2)

Out[45]:

		Predicted		
		A	B	C
Actual	A	X	X	FP1
	B	X	X	FP2
	C	FN1	FN2	TP

Q6-3

Binary Classification Model

Train Data Source: Training Data or Binary Class - tab

Test Data Source: Test Data for Binary Class - tab

```
There are no Nulls in data, hence missing value treatment is not required.
```

- There are 5000 records approximately
- To reduce initial complexity and to improve the ability of the model to generalise, we will not encode any variable which has less than 100 rows per category into seperate encoded variables, but merge all such variables into one bucket (constant / others / intercept)
- Please note 100 is not a magic number, and its not deduced by any statistical / mathematical way; more complex testing can be performed for optimality of such number, but we will keep things simple for now
- Also the count is based on training set and not testing set
- For Dep column: "GEN" is the base category
- Estimated cost is at a whole different range, hence we will take a Log transform of estimated cost
- Promoter is encoded as 0 and Passive & Detractors are encoded as 1

Out[59]:

```
11543
```

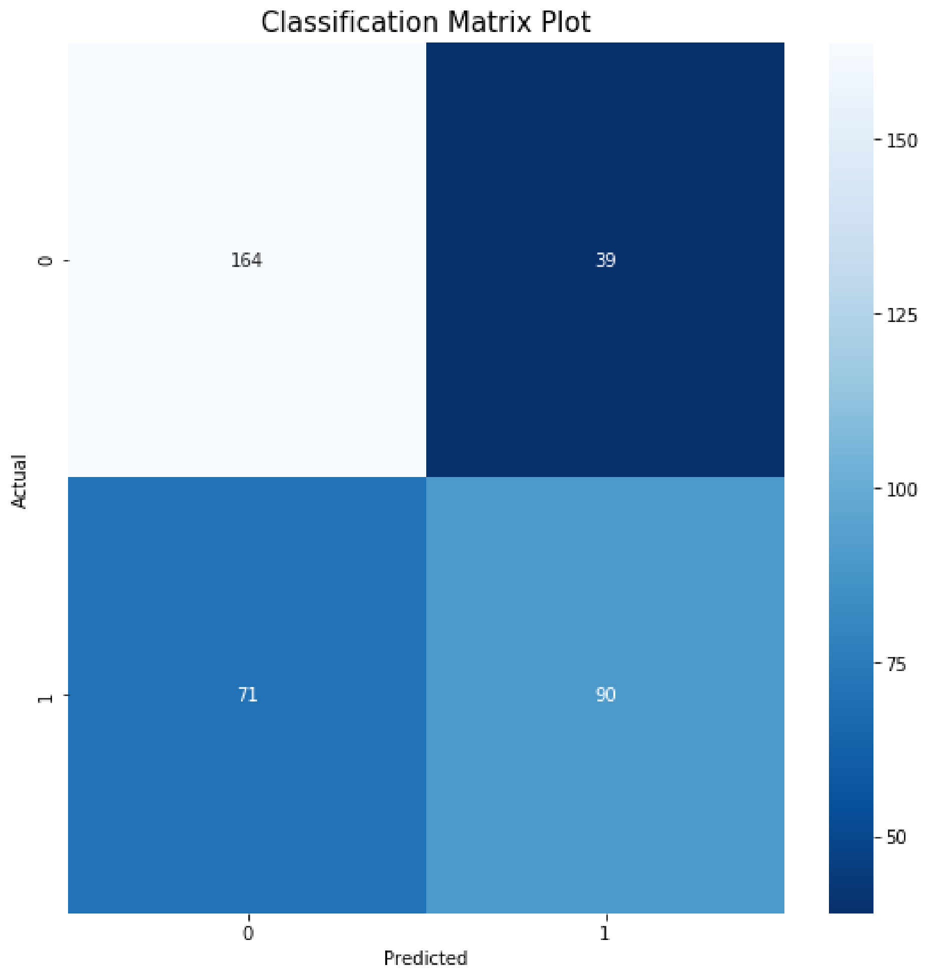
We run a stepwise model and select important varibales at significance of 0.1

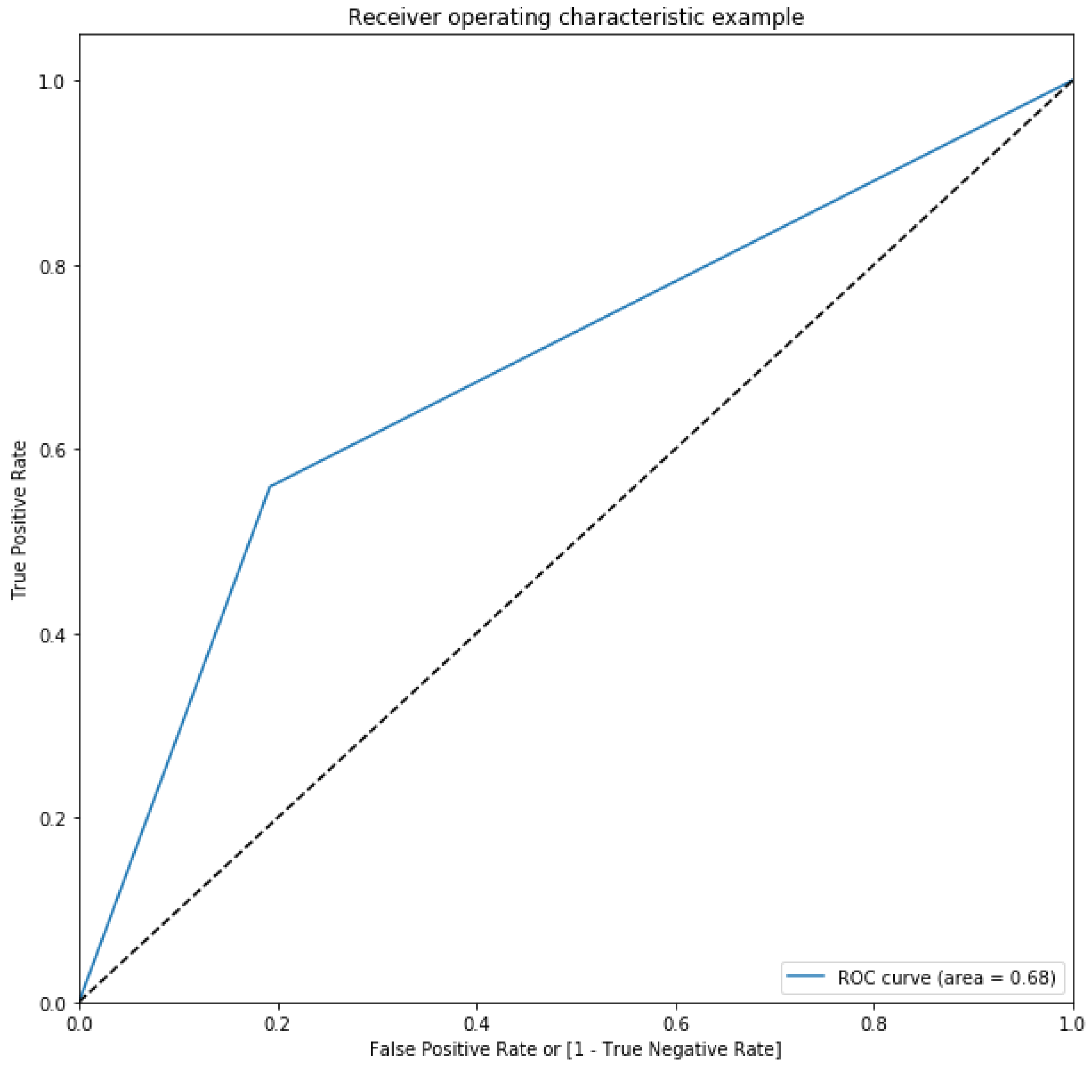
We rebuild the model with just the signifact factors

- Details of the models is as below

Out[62]:

```
Generalized Linear Model Regression Results

      Dep. Variable:          NPS_bin   No. Observations:       4989
             Model:               GLM       Df Residuals:       4963
      Model Family:          Binomial           Df Model:         25
     Link Function:             logit              Scale:        1.0
            Method:              IRLS     Log-Likelihood:    -2581.6
              Date:  Sun, 30 Sep 2018          Deviance:     5163.2
              Time:          19:27:09      Pearson chi2:   5.03e+03
    No. Iterations:                 5

                                        coef   std err        z   P>|z|   [0.025   0.975]
                              const   7.8345    0.450   17.425   0.000    6.953    8.716
                            CE_CSAT  -0.7503    0.080   -9.425   0.000   -0.906   -0.594
                    CE_VALUEFORMONEY  -0.5809    0.072   -8.092   0.000   -0.722   -0.440
                          EM_NURSING  -0.3909    0.097   -4.033   0.000   -0.581   -0.201
      AD_TARRIFFPACKAGESEXPLAINATION  -0.3898    0.078   -5.026   0.000   -0.542   -0.238
                   AD_STAFFATTITUDE   0.3214    0.084    3.821   0.000    0.157    0.486
               INR_ROOMCLEANLINESS   0.1721    0.080    2.161   0.031    0.016    0.328
                 INR_ROOMAMBIENCE  -0.1611    0.090   -1.781   0.075   -0.338    0.016
                   FNB_FOODQUALITY  -0.1302    0.067   -1.952   0.051   -0.261    0.001
              FNB_FOODDELIVERYTIME  -0.2209    0.077   -2.858   0.004   -0.372   -0.069
                FNB_STAFFATTITUDE   0.1523    0.089    1.708   0.088   -0.022    0.327
              AE_PATIENTSTATUSINFO  -0.3111    0.090   -3.440   0.001   -0.488   -0.134
                  AE_ATTENDEEFOOD  -0.1338    0.069   -1.943   0.052   -0.269    0.001
         DOC_TREATMENTEXPLAINATION   0.3036    0.111    2.745   0.006    0.087    0.520
                         DOC_VISITS  -0.4253    0.103   -4.145   0.000   -0.626   -0.224
               NS_NURSESATTITUDE   0.3832    0.107    3.587   0.000    0.174    0.593
     OVS_OVERALLSTAFFPROMPTNESS  -0.4096    0.109   -3.750   0.000   -0.624   -0.196
             OVS_SECURITYATTITUDE   0.2888    0.095    3.040   0.002    0.103    0.475
              DP_DISCHARGEQUERIES  -0.2463    0.070   -3.506   0.000   -0.384   -0.109
                          PEDIATRIC   0.2196    0.102    2.145   0.032    0.019    0.420
                            GENERAL  -0.3748    0.083   -4.514   0.000   -0.537   -0.212
                         ULTRA SPL  -0.3097    0.156   -1.982   0.048   -0.616   -0.003
                              RENAL  -0.3643    0.179   -2.038   0.042   -0.715   -0.014
                         CORPORATE  -0.2695    0.134   -2.011   0.044   -0.532   -0.007
                         Karnataka   0.3325    0.089    3.726   0.000    0.158    0.507
                         EXEMPTION  -0.1948    0.117   -1.666   0.096   -0.424    0.034
```





- The Regression has been set up to identify detractors and understand reasons that may lead to a not such a good score
- This is not a model to understand on day 1 when a customer comes in whether he will turn out to be a detractor or not. Such a model will be based on Customer Demograhics and other customer attributes vs NPS_Score. This model includes NPS_Scores provided by customers for individual departments which will be not available for a new customer. Hence using this model for such analysis may not be prudent

Observations

- Areas to improve: As these are coming out as key features leading to higher Detractors / Passive responders
  - Admission Staff Attitude
  - Cleanliness and Hygiene of the Room and Bath Room
  - Karnataka residents are more dis-satisfied
  - Helpfulness or lack of it of security staff
  - Nursing Attitude
  - Food and Beverage Staff Attitude
- Some areas that are working well for them:
  - Prompt response to concerns or complaints made
  - Regular process updates and visits by Doctors
  - Emergency Nursing
  - Explanation of tariff & packages available
  - Guidance and Information on Patient Health Status

Recommendations

- Focus on Staff and Nurse behavioural training
- Improve room and bathroom hygiene
- Given a large number of patients are from Karnataka, and given these people have a higher chance of giving poor NPS_Scores, it is advisable to understand the need of patients from this geographic region and if possible cater to those needs. A follow up study can be conducted to understand the need of people from these regions to further improve their scores.

Q6-4

Ordinal Logistic Classification Model

Train data source : Training Data for Multi-Class M - tab

Test data source : Test Data for Multi-Class Model

```
There are no Nulls in data, hence missing value treatment is not required.
```

```
Out[66]: 11162
```

```
Optimization terminated successfully.
         Current function value: 0.679890
         Iterations 8
```

```
Optimization terminated successfully.
         Current function value: 0.695003
         Iterations 8
```

Out[69]:

MNLogit Regression Results

Dep. Variable:	NPS_Status	**No. Observations:**	4989
Model:	MNLogit	**Df Residuals:**	4947
Method:	MLE	**Df Model:**	40
Date:	Sun, 30 Sep 2018	**Pseudo R-squ.:**	0.2066
Time:	19:27:18	**Log-Likelihood:**	-3467.4
converged:	True	**LL-Null:**	-4370.3
		LLR p-value:	0.000

NPS_Status=Detractor	**coef**	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	6.7859	0.589	11.519	0.000	5.631	7.941
CE_CSAT	-0.5216	0.134	-3.904	0.000	-0.783	-0.260
CE_VALUEFORMONEY	-0.9356	0.112	-8.329	0.000	-1.156	-0.715
FNB_FOODDELIVERYTIME	-0.1319	0.104	-1.264	0.206	-0.337	0.073
RENAL	-0.5748	0.352	-1.634	0.102	-1.265	0.115
AE_ATTENDEEFOOD	0.1583	0.103	1.531	0.126	-0.044	0.361
OVS_SECURITYATTITUDE	0.3137	0.141	2.230	0.026	0.038	0.589
OVS_OVERALLSTAFFPROMPTNESS	0.0809	0.141	0.575	0.565	-0.195	0.357
AgeYrs	0.0057	0.003	2.112	0.035	0.000	0.011
CE_ACCESSIBILITY	-0.5430	0.110	-4.924	0.000	-0.759	-0.327
AE_PATIENTSTATUSINFO	0.0210	0.132	0.159	0.873	-0.237	0.279
SPECIAL	-0.2964	0.177	-1.670	0.095	-0.644	0.051
INR_ROOMCLEANLINESS	-0.3039	0.099	-3.077	0.002	-0.497	-0.110
NS_NURSESATTITUDE	0.1285	0.142	0.905	0.365	-0.150	0.407
AD_TARRIFFPACKAGESEXPLAINATION	-0.1041	0.122	-0.852	0.394	-0.344	0.135
DP_DISCHARGEQUERIES	-0.4682	0.151	-3.099	0.002	-0.764	-0.172
DP_DISCHARGEPROCESS	0.1844	0.146	1.259	0.208	-0.103	0.472
DOC_VISITS	-0.1929	0.133	-1.454	0.146	-0.453	0.067
DOC_TREATMENTEXPLAINATION	-0.2527	0.146	-1.735	0.083	-0.538	0.033
AD_STAFFATTITUDE	-0.2766	0.128	-2.163	0.031	-0.527	-0.026
AD_TIME	0.2802	0.113	2.479	0.013	0.059	0.502
NPS_Status=Promotor	**coef**	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	-5.0880	0.390	-13.061	0.000	-5.852	-4.324
CE_CSAT	0.6408	0.089	7.202	0.000	0.466	0.815
CE_VALUEFORMONEY	0.4456	0.077	5.763	0.000	0.294	0.597
FNB_FOODDELIVERYTIME	0.2145	0.070	3.071	0.002	0.078	0.351
RENAL	0.3525	0.184	1.912	0.056	-0.009	0.714
AE_ATTENDEEFOOD	0.2712	0.065	4.188	0.000	0.144	0.398
OVS_SECURITYATTITUDE	-0.2040	0.098	-2.075	0.038	-0.397	-0.011
OVS_OVERALLSTAFFPROMPTNESS	0.4147	0.114	3.642	0.000	0.192	0.638
AgeYrs	0.0049	0.002	3.000	0.003	0.002	0.008
CE_ACCESSIBILITY	-0.0913	0.080	-1.140	0.254	-0.248	0.066
AE_PATIENTSTATUSINFO	0.2766	0.090	3.088	0.002	0.101	0.452
SPECIAL	0.0356	0.101	0.353	0.724	-0.162	0.233
INR_ROOMCLEANLINESS	-0.1579	0.069	-2.273	0.023	-0.294	-0.022
NS_NURSESATTITUDE	-0.3953	0.111	-3.560	0.000	-0.613	-0.178
AD_TARRIFFPACKAGESEXPLAINATION	0.3989	0.086	4.633	0.000	0.230	0.568
DP_DISCHARGEQUERIES	0.1450	0.100	1.444	0.149	-0.052	0.342
DP_DISCHARGEPROCESS	0.0870	0.092	0.943	0.346	-0.094	0.268
DOC_VISITS	0.3548	0.108	3.295	0.001	0.144	0.566
DOC_TREATMENTEXPLAINATION	-0.3773	0.117	-3.227	0.001	-0.606	-0.148
AD_STAFFATTITUDE	-0.3809	0.092	-4.159	0.000	-0.560	-0.201
AD_TIME	0.1043	0.075	1.391	0.164	-0.043	0.251

Out[70]:

pred	**BasePassive**	**Detractor**	**Promotor**
NPS_Status			
BasePassive	34	11	72
Detractor	13	18	13
Promotor	14	3	186

Compare with Binary Model

- In Binary Model focus was identifying who were non-Promoters and trying to find reasons for why they gave non-positive rating. In Ordinal Logistic Model, the base class was considered as the people who gave passive scores and we are trying to find the reasons which led to negative scores and also identify the areas/reasons working well by studying the positive scores, so that, better practices can be used in the areas not doing so well
- What is working well / contributing to good NPS score
  - Ateendee Food
  - Food Delivery time
  - Age (Increase in age of Patients leads to improved NPS score)
  - Discharge Queries
  - Overall Staff Promptness
  - AE_PATIENTSTATUSINFO
  - AD_TARRIFFPACKAGESEXPLAINATION
  - CE_VALUEFORMONEY
- What is contributing to Detractors
  - OVS_SECURITYATTITUDE
  - Admission Time
- What is needed to push Passive customers to Promoters
  - Improve Room cleanliness
  - Better Explanation of Doctor Treatment
  - Improvement of Security Attitude
  - Improvement of Staff Attitude
  - Value for Money - Improve people's perception of the treatment value / may be better explained with explanation of Doctor Treatment

The results are in line with the findings from the binary classification model. However this model is more powerful as it provides complete segregation of the Passive and Detractor communities. It is easier to identify the reasons for huge dissatisfaction among some patients.

Passive responders at time are more difficult to understand and react to as they are not completely open in coming out with their observations. Where as Promoters and Detractors (though not desirable) voice clear cut opinions about what is working well and what is not. This clear preferences / feedback helps in taking corrective action / continuing with what is working well

Q6-5

Conclusions

- Better Explanation of Doctor Treatment is needed, notion of Value for Money - Improve people's perception of the treatment value / may also improve
- Improvement of Security Attitude via trainings
- Improvement of Staff Attitude via trainings
- Focus on Staff and Nurse behavioural training
- Improve room and bathroom hygiene
- Given a large number of patients are from Karnataka, and given these people have a higher chance of giving poor NPS_Scores, it is advisable to understand the need of patients from this geographic region and if possible cater to those needs. A follow up study can be conducted to understand the need of people from these regions to further improve their scores.