Case 1

Objective - To identify the impact of independent variables

Independent Variables - (PromotionalIndex, Feature Advertising)

Dependent variables - Sales

Justification - Since dependent variables and all the independent variables are quantitative in nature so we will use regression analysis.

Data Analysis -

H0 (Null Hypothesis) - This model is not statistically significant

H1 (Alternative Hypothesis) - This model is statistically significant

| ANOVA |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | SS | MS | F | Significance F |
| Regression | 2 | 7.74176E+11 | 3.87088E+11 | 11.56429488 | 0.00 |
| Residual | 97 | 3.24685E+12 | 33472681844 |  |  |
| Total | 99 | 4.02103E+12 |  |  |  |

If p(significance) = 0< alpha

=> Reject H0

H1

Here, p < alpha so we reject null hypothesis and accept H1, Therefore we can say model is significant

Interepretation

Of R2 (coefficiant of determination ), Here R2 value is 0.19

# Case 2 - Experiential Retailing: Influence on Young Indian Consumer’s Response

Multivariate linear regression model

## Objective -

To identify the impact of all independent variables i.e. sound, light, layout, music, fragrance, etc on the customers retail experience

## Justification -

Since all variables are quantitative(numerical) in nature therefore, to check the above said objective we will use multivariate regression model

## Data Analysis

Hypothesis for multivariate linear regression model

Null hypothesis (H0) : the overall model is not statistically significant

Hypothesis (H1) : Overall model is statistically significant

### Step 1 -

If , p<α(alpha) - we reject H0

Here from output P = 0.0001341 less than Alpha (0.05)

We reject , Null hypothesis (H0) and accept H1

Therefore we can say that model is statistically is significant

### Step 2 -

Hypothesis for beta ß coefficient

H0i = All beta coefficient are not statistically significant

H1i = At least one of the beta coefficient is statistically significant

| Coefficients: |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t | P Value |  |
| (Intercept) | 0.818982 | 0.707732 | 1.157 | 0.250096 |  |
| shoping.when.bored | 0.009209 | 0.089835 | 0.103 | 0.918567 |  |
| waste.of.time | -0.102435 | 0.110548 | -0.927 | 0.356477 | 0 |
| wall.colour | 0.175777 | 0.099336 | 1.77 | 0.080015 | . |
| fragrance | 0.084468 | 0.110294 | 0.766 | 0.445669 |  |
| emp.knowledge | -0.367028 | 0.1491 | -2.462 | 0.015635 | \* |
| layout.flooring | -0.078535 | 0.105669 | -0.743 | 0.459183 |  |
| recommend | 0.044022 | 0.124292 | 0.354 | 0.723986 |  |
| emp.concerned | 0.176481 | 0.122625 | 1.439 | 0.153383 |  |
| layout.spacious | -0.151162 | 0.099258 | -1.523 | 0.131102 |  |
| emp.trustworthy | 0.150259 | 0.116543 | 1.289 | 0.200424 |  |
| layout.design.display | 0.030934 | 0.122993 | 0.252 | 0.801963 |  |
| entertain | -0.076202 | 0.162124 | -0.47 | 0.639416 |  |
| enthusiam | 0.172139 | 0.134744 | 1.278 | 0.204527 |  |
| moretime.spent | 0.451503 | 0.116039 | 3.891 | 0.000185 | \*\*\* |
| buy.more | 0.107271 | 0.098752 | 1.086 | 0.28011 |  |
| light.dull | 0.108277 | 0.117792 | 0.919 | 0.360307 |  |
| design.good | -0.053493 | 0.160039 | -0.334 | 0.738926 |  |
| music.bothersome | -0.208864 | 0.104116 | -2.006 | 0.047693 | \* |
| emp.not.assist | 0.12882 | 0.101218 | 1.273 | 0.20623 |  |

Single star, double star, or triple star means it is significant

Dot mean significance is less than 10%

| Coefficients: |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | std error | t | P Value |  |
| (Intercept) | 1.13789 | 0.31771 | 3.582 | 0.000508 | \*\*\* |
| wall.colour | 0.1865 | 0.08737 | 2.135 | 0.034985 | \* |
| emp.knowledge | -0.22314 | 0.10351 | -2.156 | 0.033259 | \* |
| moretime.spent | 0.52414 | 0.09432 | 5.557 | 1.91E-07 | \*\*\* |

### Step 3

y=α+ß1.x1+ß2.x2+..++E

Frequency of Visit=

1.1378+0.1865(wall.color) - 0.223(emp.knowledge)+0.52414(moretime.spent)+E

#### Interpretation:

If employment knowledge and more time spent is constant and also they increase the wall color by 1 unit, then frequency of visit will increase by 18.65%

If again wc and more time spent constant and if we increase employee knowledge by 1 unit then frequency of visit will decrease by 22.3%

If again wc and employee knowledge constant and if we increase moretime spent by 1 unit then frequency of visit will increase by 52.4%

After comparing beta coefficient we conclude that more time spent in retail store is more important and most influencing variable followed by employee knowledge and wall color

#### Step 4 -

Multi-familiarity (multicollinearity )

If there is high positive correlation between variables then we can say that multicollinearity is present between the correlation

wall.colour emp.knowledge moretime.spent

1.067438 1.313283 1.298247

Since the vif value for all independent variables is below 5 , therefore no multicollinearity is present between the variables

# Multivariate Linear regression model

# Objective

To identify the impact of the all independent variables i.e Price of egg, Price of cookies on the dependent variable that is sales

# Justification

Since all the dependent and independent variables are numerical in nature therefore we use multivariate linear regression model

# Data Analysis

1. Hypothesis for model

H0 : The model is not statically significant

H1 : Model is statistically significant

α : significance level

α should be less than 0.05 for variable to be significant

From output we observe that p=0 i.e less that α (0.05)

We Reject null hypothesis and accept H1

We can say our model is statistically significant

1. Hypothesis for ß significant

H0i : All ß coefficient are not statistically significant

H1i : At least one of the ß coefficient is statistically significant

| Coefficients: |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | t value | Pr(>|t|) | Significance |
| (Intercept) | 151.318 | 12.776 | 11.844 | 3.34E-12 | \*\*\* |
| Price.Eggs | -18.727 | 1.882 | -9.953 | 1.57E-10 | \*\*\* |
| Price.Cookies | -8.786 | 2.369 | -3.709 | 0.00095 | \*\*\* |

Interpretation

For both the variables the p value is 0 thet is less than α so we reject the null hypothesis i.e. H0 and we conclude that the ß coefficients for both variables are statistically significant

1. Estimation

If we increase the price of the egg by 1 unit, sales will decrease by 18.727 units

Similarly, if we increase the price of the cookies by 1 unit sales will decrease by 8.786

1. R squared - Coefficient of determination

Here R squared value id 0.8157 i.e. my model will predict the variation only 81.57 percent for the dependent variable with respect to the changes in the independent variables

And remaining 18.5 percent variation is due to external factors

1. Multicollinearity

The VIF(variance inflation factor) value for all the variables is <5 for all the independent variables therefore there is no multicollinearity between the variables

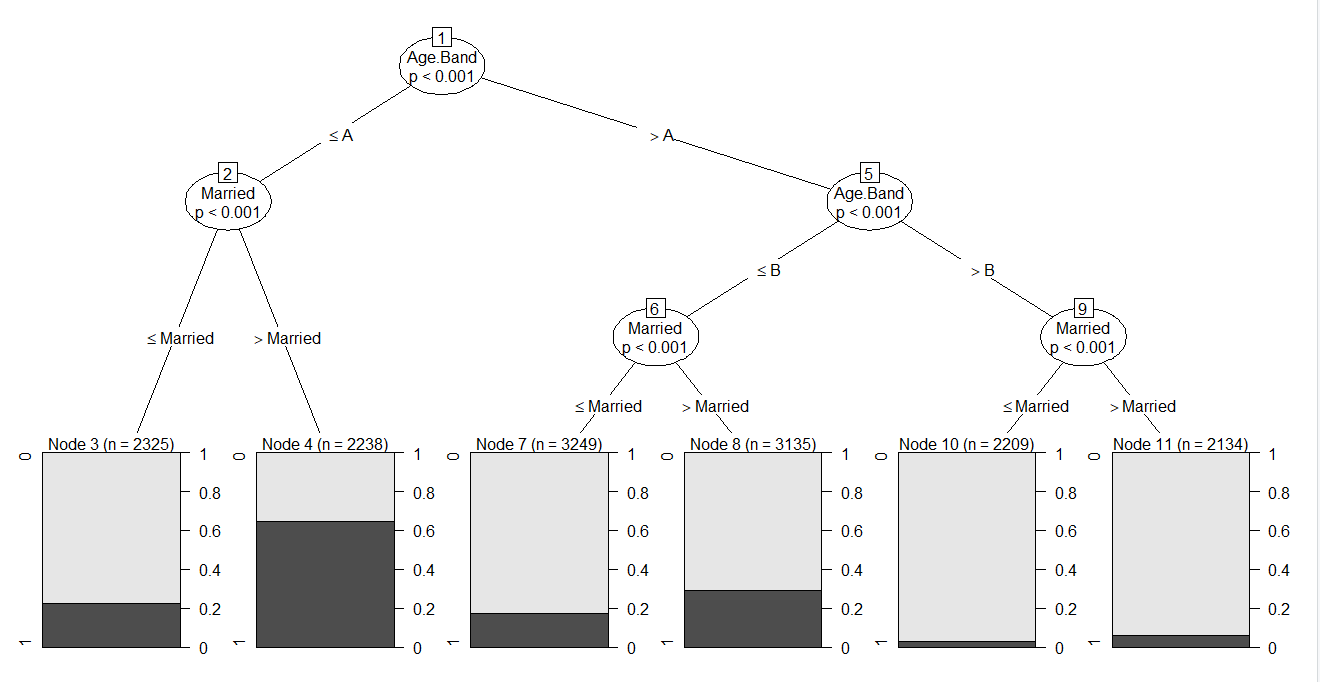
Auto Insurance Policy Records

Topic - Decision tree(Chad Analysis

Objective :- to identify the probability of non defaulter customers in a group who have paid the premium on time.

Justification :- Since, the dependent variable is qualitative in nature so we will use the chad analysis

Data Analysis :-



Node3 - Customers belongs to age group A at they are married, their probability of non defaulter is 0.2

Node 4 - Customers belongs to the age group A and they are single, their probability of non defaulters is more than 0.6

Node 7 - Customer belongs to age group B and they are married, their probability of non defaulter is less than 0.2

Node 8 - Customer belongs to age group B and they are Single, their probability of non defaulter is more than 0.2

Node 10 - Customer belongs to age group C and they are married, their probability of non defaulter is less than 0.1

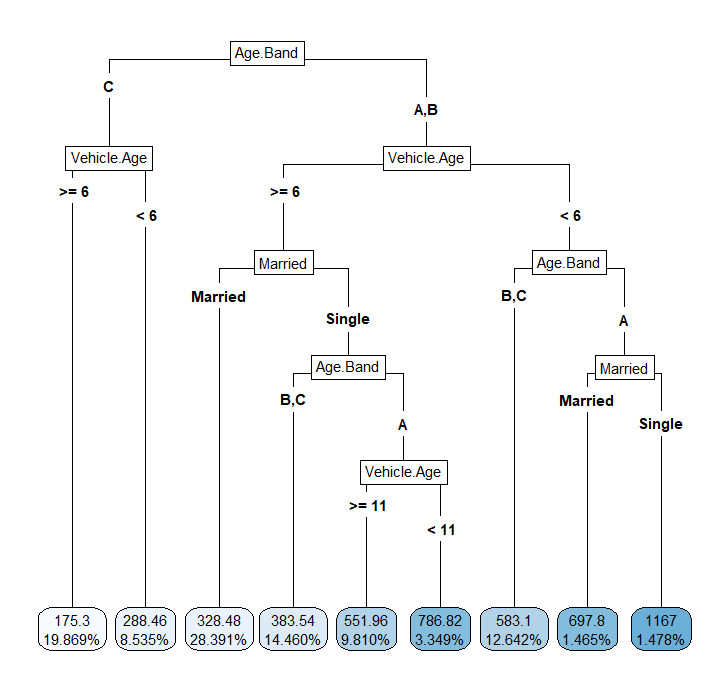
Node 11 - Customer belongs to age group C and they are Single, their probability of non defaulter is more than Node 11

CART Analysis

Objective :- To calculate the average losses (Dependent variables) for different different groups with the help of decision tree

Justification :- Since the dependent variables is losses is quantitative in nature therefore we will use the cart analysis. Cart analysis is based on regression method therefore, we are able to calculate average losses for different different groups

Data analysis :-



* Firstly the variable is going to be split in two branches, in one branch we have the customers oh age group C and in another group the customers belong to age group A&B
* FUrther Age group C customers going to be classified with respect to the vehicle age in two groups, one group has the vehicle more than = 6 and other has vehicle less than 6
* The average losses for the group (age band = C, vehicle age>= 6) is $175.3
* The average losses for the age group C and vehicle <6 is $288.45

Date: 10-10-2022

Objective to identify the impact of all independent variables on our dependent variables

independent variable(Promotion index,FeatureAdvertising Index)

dependent variable(sales)

Justification

Since DV and all the IV’s are quantitative in nature, so we wil use regression analysis

Hypothesis:

H0 (Null Hypothesis)

The model is not Statistically Significant

H1(Alternative Hypothesis) : The model is Statistically Significant

| ANOVA |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | SS | MS | F | Significance F |
| Regression | 2 | 7.34E+11 | 3.67E+11 | 10.82866 | 0.00005690 |
| Residual | 97 | 3.29E+12 | 3.39E+10 |  |  |
| Total | 99 | 4.02E+12 |  |  |  |

Here p < alpha so we reject null hypothesis and accept H1 therefore we can say model is significant

Interpretation of R2 (Coefficient of determination):here R2 value is 0.18 that is below 0.50 that means model is not very good . this value explains only 18% variation in the change of dependent variable with respect to changes is the independent variables

Data analysis:

1 represent when walmart was open post and 0 represent when walmart was closed pre.

1 represent holiday is thre and 0 means there is no holiday.

Null hypothesis H0 :

Model is significant

Case 2:

**07 november 2022**

Multivariate linear regression

Likert scale - scales consist of responses from strongly disagree to strongly agree

**Experiential Retailing: Influence on Young Indian Consumer’s Response**

Dependent variable - frequency of visit

All other variables are independent variable

Objective:

To identify the impact of all the independent variables that is sound,light,layout,music,etc on the customer’s retail experience.

Justification

Since all the variables are quantitative(numeric values) in nature therefore to achieve the above said objective we will use multivariate regression model

Hypothesis for multivariate regression model

Null hypothesis H0: Overall model is not statistically significant.

From the output p= 0.0001341 < alpha α (0.5) i.e we reject the null hypothesis and accept H1

Therefore we can say that model is statistically significant

Step2: Hypothesis for Beta ß coefficient

Null Hypothesis H0:All the ß coefficients are not statistically Significant

H1:At least one of the ß coefficients is Statistically significant

| Coefficients: |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | t value | P value :Pr(>|t|) |  | significant/insignificant(alpha=0.05) |  |
| (Intercept) | 0.818982 | 0.707732 | 1.157 | 0.250096 |  |  |  |
| shoping.when.bored | 0.009209 | 0.089835 | 0.103 | 0.918567 |  | p>α >> p does not less than alpha so we donot reject null hypothesis | not significant |
| waste.of.time | -0.102435 | 0.110548 | -0.927 | 0.356477 |  | p>α >> p does not less than alpha so we donot reject null hypothesis | not significant |
| wall.colour | 0.175777 | 0.099336 | 1.77 | 0.080015 | . | p>α >> p does not less than alpha so we donot reject null hypothesis | \*significant at 10% |
| fragrance | 0.084468 | 0.110294 | 0.766 | 0.445669 |  | p>α >> p does not less than alpha so we donot reject null hypothesis |  |
| emp.knowledge | -0.367028 | 0.1491 | -2.462 | 0.015635 | \* |  | significant |
| layout.flooring | -0.078535 | 0.105669 | -0.743 | 0.459183 |  |  | insignificant |
| recommend | 0.044022 | 0.124292 | 0.354 | 0.723986 |  |  | insignificant |
| emp.concerned | 0.176481 | 0.122625 | 1.439 | 0.153383 |  |  | insignificant |
| layout.spacious | -0.151162 | 0.099258 | -1.523 | 0.131102 |  |  | insignificant |
| emp.trustworthy | 0.150259 | 0.116543 | 1.289 | 0.200424 |  |  | insignificant |
| layout.design.display | 0.030934 | 0.122993 | 0.252 | 0.801963 |  |  | insignificant |
| entertain | -0.076202 | 0.162124 | -0.47 | 0.639416 |  |  | insignificant |
| enthusiam | 0.172139 | 0.134744 | 1.278 | 0.204527 |  |  | insignificant |
| moretime.spent | 0.451503 | 0.116039 | 3.891 | 0.000185 | \*\*\* |  | significant |
| buy.more | 0.107271 | 0.098752 | 1.086 | 0.28011 |  |  | insignificant |
| light.dull | 0.108277 | 0.117792 | 0.919 | 0.360307 |  |  | insignificant |
| design.good | -0.053493 | 0.160039 | -0.334 | 0.738926 |  |  | insignificant |
| music.bothersome | -0.208864 | 0.104116 | -2.006 | 0.047693 | \* |  | significant |
| emp.not.assist | 0.12882 | 0.101218 | 1.273 | 0.2062 |  |  | not significant |

single,double and triple star means it is significant

Single dot means it is significant at 10% but not 5%

Main Model

| Coefficients: |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | t value | P value :Pr(>|t|) |  |
| (Intercept) | 1.13789 | 0.31771 | 3.582 | 0.000508 | \*\*\* |
| wall.colour | 0.1865 | 0.08737 | 2.135 | 0.034985 | \* |
| emp.knowledge | -0.22314 | 0.10351 | -2.156 | 0.033259 | \* |
| moretime.spent | 0.52414 | 0.09432 | 5.557 | 1.91E-07 | \*\*\* |

Step3: Regression Model

y=α+ß1.x1+ß2.x2+..++E

Frequency of Visit=

1.1378(intercept)+0.1865(wall.color) - 0.223(emp.knowledge)+0.52414(more time.spent)+E

Interpretation:

If employment knowledge and more time spent is constant and also they increase the wall color by 1 unit, then frequency of visit will increase by 18.65%

If again wc and more time spent constant and if we increase employee knowledge by 1 unit then frequency of visit will decrease by 22.3%

If again wc and employee knowledge constant and if we increase moretime spent by 1 unit then frequency of visit will increase by 52.4%

After comparing beta coefficient we conclude that more time spent in retail store is more important and most influencing variable followed by employee knowledge and wall color

Step4: multicollinearity

Means if there is high degree +ve correlation between the independent variables then we cna say that multicollinearity is present between the variables.

VIF- **V**ariance **I**nflation **F**actor

| wall.colour | emp.knowledge | moretime.spent |
| --- | --- | --- |
| 1.067438 | 1.313283 | 1.298247 |

Since VIF value for all Independent variables is below 5 that means no multicollinearity is present between the variables.

14 NOV 2022

Model: Multivariate Regression Model

Objective

To identify the impact of all independent variables i.e price of egg and price of cookies and dependent variables on the dependent variable that is sales.

Justification:

Since all the dependent variables and independent variables numerical in nature therefore we will use multivariate regression model.

Data analysis

Hypothesis

H0:The model is not statistically significant

H1:the model is statistically significant

α=Significance Level

Significance Level is 0.5% and Confidence Level is 95%

If p<α we reject H0.

From Ouptut we observe that p=0 .ie we reject null hypothesis

Hypothesis for ß coefficient

H0i:All ß coefficients are not statistically Significant

H1i:At least on of the ß is statistically Significant

| Coefficients: |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | t value | Pr(>|t|) | Significance |
| (Intercept) | 151.318 | 12.776 | 11.844 | 3.34E-12 | \*\*\* |
| Price.Eggs | -18.727 | 1.882 | -9.953 | 1.57E-10 | \*\*\* |
| Price.Cookies | -8.786 | 2.369 | -3.709 | 0.00095 | \*\*\* |

Interpretation:

For Both the Variables The P value is 0 ie P<α so we reject Null Hypothesis and conclude that the ß coefficients for both variables are statistically significant

Step 3:Model

If we increase tha price of the egg by 1 unit then sales will decrease by 18.727 units

similarly,If we increase tha price of the cookies by 1 unit then sales will decrease by 8.786 units

Step 4: R Square

R Square means Coefficient of Determination

Here R Square Value is 0.8157 i.e 81.57%.

Which means the Model will Predict the Variation only 81.57% for the dependent variable with respect to the changes in the Independent variable.

And remaining 18.53 variation is due to the external factors.

Step 5: Multicollinearity between the variables

The VIF (**V**ariance **I**nflation **F**actor) value for all the independent variables is less than 5 therefore there is no multicollinearity between the variables.

16 Nov 2022

Intercept+sumProduct(X1:x2,ß1+ß2)

28 NOV 2022

Auto Insurance Policy Record

Decision Tree Case Study:CHAID analysis

Objective:

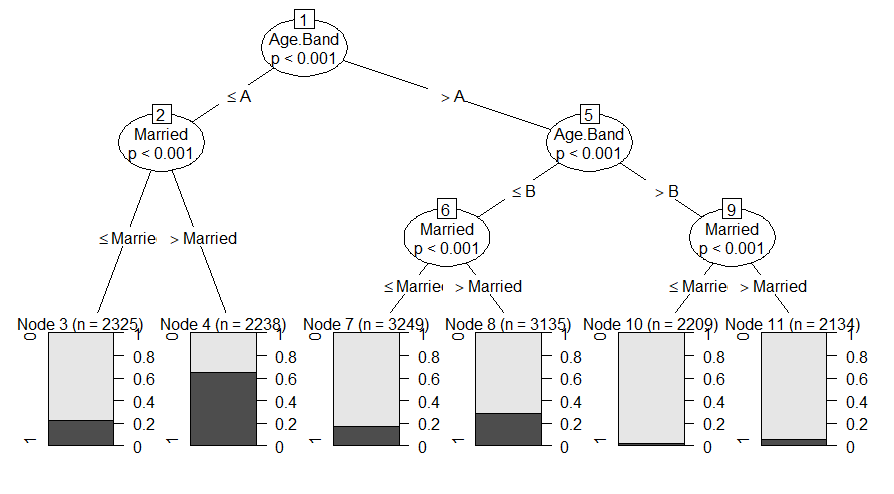
To identify the probability of non defaulter customers in a group who had paid the premium on time

Justification:

SInce the dependent variable is Qualitative in Nature so we will use the CHAID Analysis

CHI Square Automatic

Data Analysis



Interpretation:

Node 3: ON Node 3 the customers belongs to age group A and they are married Their probability of non defaulter is 0.2 and their sample size is 2325

On Node 4 customers belongs to Age Group A and they are single and their probability of non Defaulter is Greater than 0.6 and their sample size is 2238

On Node 7 customers belongs to Age Group B and they are married and their probability of non Defaulter is lesser than 0.2 and their sample size is 3249

On Node 8 customers belongs to Age Group B and they are single and their probability of non Defaulter is Greater than 0.2 and their sample size is 3135

On Node 10 customers belongs to Age Group C and they are married and their probability of non Defaulter is equivalent to 0 and their sample size is 2209

On Node 11 customers belongs to Age Group C and they are single and their probability of non Defaulter is Greater than 0 and less than 0.1 and their sample size is 2134

28 Nov 2022

CART Analysis

Objective :

To calculate the average losses(dependent variable) for different groups with the help of decision tree.

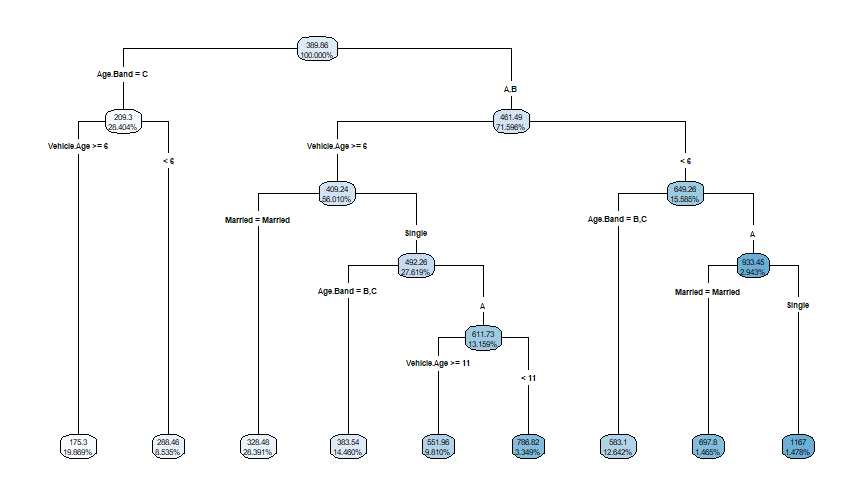
Justification:

Since the dependent variable is quantitative in nature therefore we will use the cart analysis.

Cart Analysis is Based on Regression method therefore we are able to calculate average losses for different groups

Data Analysis:

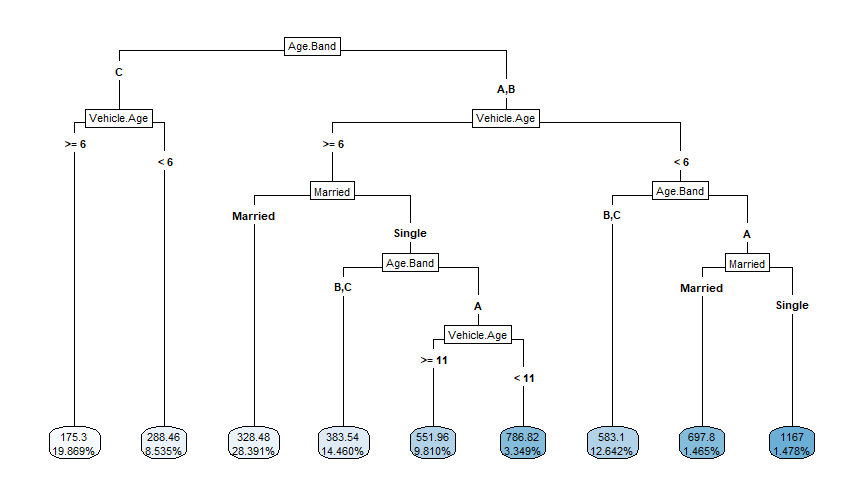
Type 4



Interpretation:

* Firstly the Variable is goint to be split in two branches , in one branch we have the customers of the age group C and in another group Customers belongs to Group A and B
* Age group C customers going to be classified with respect to the variable vehicle.age in two groups,one group has the vehicles.age more than equals to 6 and another group has the vehicles.age less than 6.
* The average Losses for the group(age.band==C and vehicle.age>=6) is $175.3
* Similarly, the average Losses for the age group C and vehicle Age < 6 is $288.45

Type 5



30 Nov 2022

Case study of Revloan.

Segmentation Cluster Analysis

