**Data Mining, Big Data and Analytics.**

Lab 6 – Logistic Regression

|  |  |  |
| --- | --- | --- |
| Name | Section | BN |
| Donia Gameel | 1 | 24 |
| Heba Ashraf Raslan | 2 | 32 |

**1. Write the variable pairs that are not correlated at all to each other.**

1. Price and Income
2. Price and Age

**2. Are there any highly correlated variables in this dataset?**

No

**3. How many categories are there for the Price variable?**

3 categories (10,20,30)

**4. Why is it divided into two entries only in the model?**

because it is treated as a categorical variable with three levels. When using a categorical variable in a regression model, one level is used as the reference category, and the model estimates coefficients for the other levels relative to this reference category.

**5. 1. Write the value of AUC.**

**2. What is the maximum value of AUC (ideal case)?**

The AUC score: 0.915272

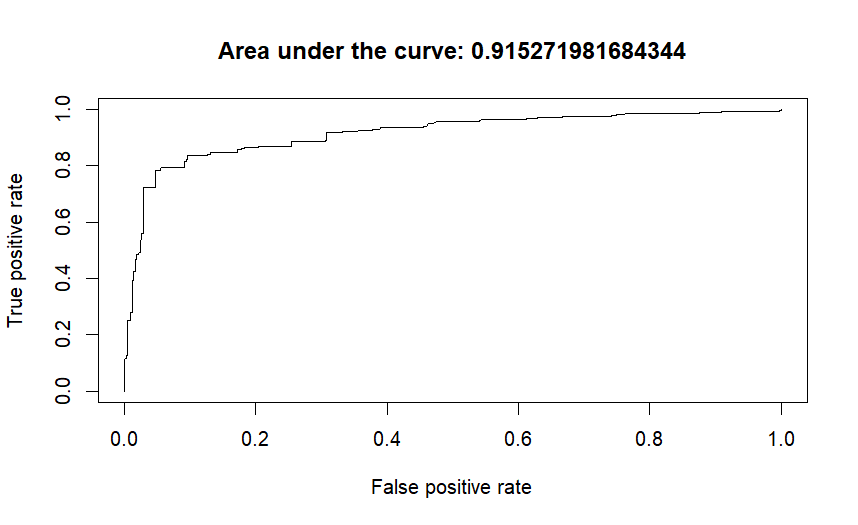
The maximum value of AUC=1

**6. What does each point in the ROC graph represent?**

**In other words, what is the value that changes and drives TPR and FPR to change too from one point to another in the graph?**

Each point on the ROC (Receiver Operating Characteristic) curve represents a different threshold for classifying the positive and negative classes. The ROC curve is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings.

As the threshold for classifying a positive outcome is increased, the model becomes more conservative, leading to fewer false positives but potentially more false negatives. This change in threshold affects the TPR and FPR, as the true positive rate increases (more actual positives correctly classified) and the false positive rate also increases (more negatives incorrectly classified as positives).

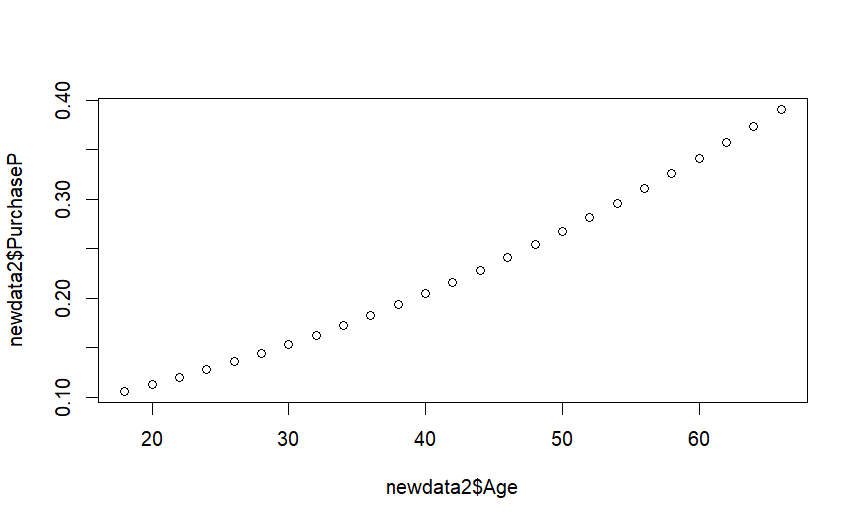
****

**7. How is the predicted probability affected by changing only Price holding all other variables constant?**

As the price increases from 10 to 30, the predicted probability of purchase decreases. This suggests that higher prices are associated with a lower likelihood of purchase, assuming all other factors remain the same.

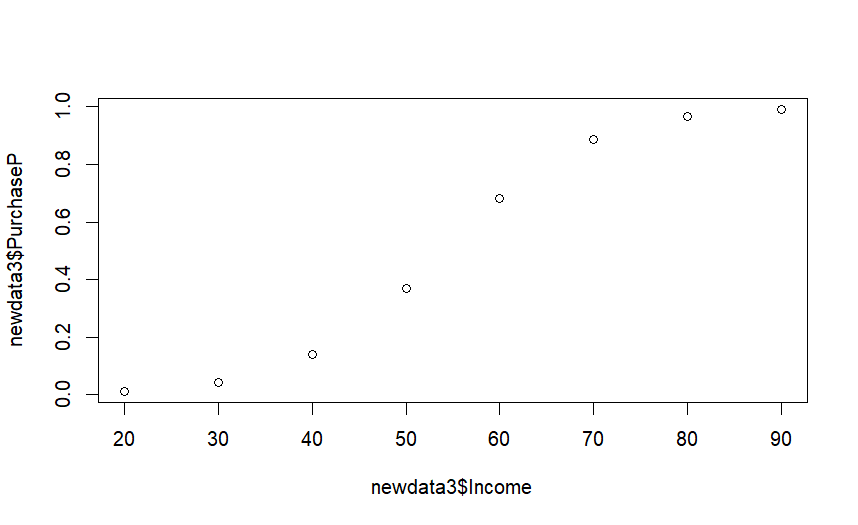
**8. How is the predicted probability affected by changing only Age holding all other variables constant?**

As Age increases, the predicted probability of purchase also increases. This suggests that older individuals are more likely to make a purchase, assuming all other factors remain the same.

****

**9. How is the predicted probability affected by changing only Income holding all other variables constant?**

As Income increases, the predicted probability of purchase also increases. This suggests that individuals with higher incomes are more likely to make a purchase, assuming all other factors remain the same.

****