Lab 5

Predictive Analysis ||

Karim Mahmoud Kamal	Sec: 2	BN: 12
Mustafa Mahmoud Hamada	Sec: 2	BN: 25

Eng. Omar Samir

Q1) Write the variable pairs that are not correlated at all to each other.

(Price, Age)

(Price, Income)

A matrix: 3 × 3 of type dbl

	Price	Income	Age
Price	1	0.00000000	0.00000000
Income	0	1.00000000	0.09612083
Age	0	0.09612083	1.00000000

Q2) Are there any highly correlated variables in this dataset?

No, as all pair have a low correlation value

Q3) How many categories are there for the Price variable?

There are 3 categories for price (10, 20, 30)

Price	MYDEPV 0	1
10	115	135
20	137	113
30	174	76

Q4) Why it is divided into two entries only in model?

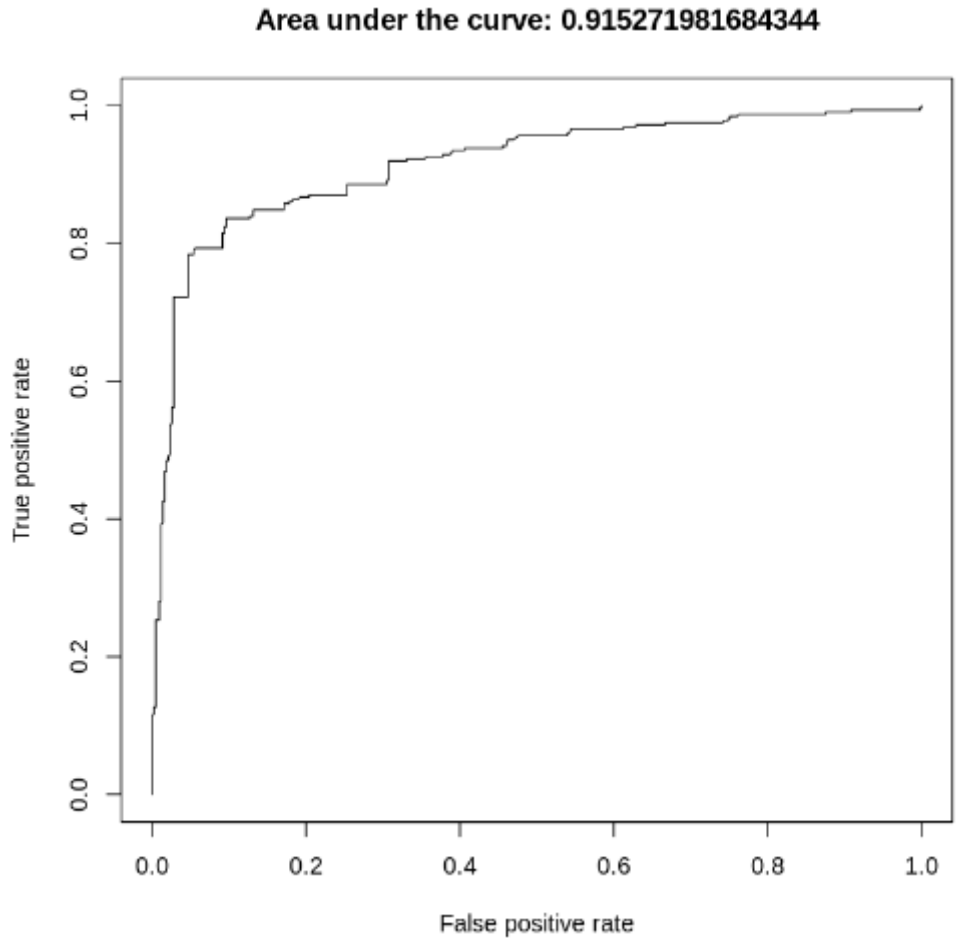
Because if we have n categories (which is 3 in this example) so it is divided into n-1 variables (which is 2).

Q5.1) Write the value of AUC.

Area under the curve = 0.915271981684344

Q5.2) What is the maximum value of AUC (ideal case)?

The maximum AUC is 1 so it maximizes the TPR and minimizes FPR



Q6) 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 ROC graph represents a different classification threshold.

The classification threshold is the point at which a predicted probability or score is converted into a binary classification (positive or negative).

A good model is characterized by a situation where, as we move from left to right along the threshold, the TPR approaches values near 1 with only a small change in the FPR.

Q7) How is the predicted probability affected by changing only Price holding all other variables constant?

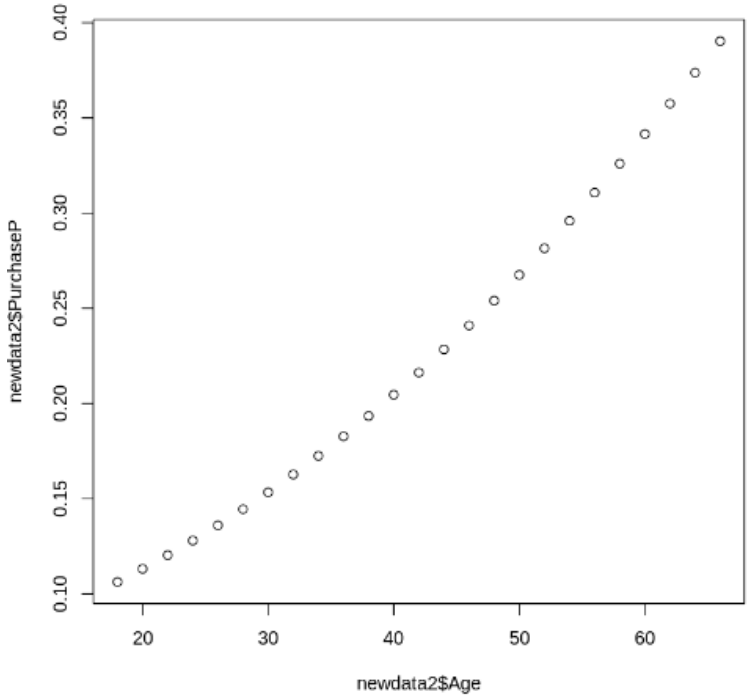
The probability of purchasing something more expensive decreases as the price increases.

A data.frame: 3 × 4

Income	Age	Price	PurchaseP
<dbl>	<dbl>	<dbl>	<dbl>
42.492	35.976	10	0.6707408
42.492	35.976	20	0.4918407
42.492	35.976	30	0.1826131

Q8) How is the predicted probability affected by changing only Age holding all other variables constant?

The predicted probability tends to increase as older clients are considered.



Age	Income	Price	PurchaseP
<dbl>	<dbl>	<dbl>	<dbl>
18	42.492	30	0.1063052
20	42.492	30	0.1131540
22	42.492	30	0.1203845
24	42.492	30	0.1280103
26	42.492	30	0.1360445
28	42.492	30	0.1444993
30	42.492	30	0.1533864
32	42.492	30	0.1627160
34	42.492	30	0.1724975
36	42.492	30	0.1827387
38	42.492	30	0.1934457
40	42.492	30	0.2046231
42	42.492	30	0.2162731
44	42.492	30	0.2283958
46	42.492	30	0.2409892
48	42.492	30	0.2540483
50	42.492	30	0.2675657
52	42.492	30	0.2815308
54	42.492	30	0.2959303
56	42.492	30	0.3107477
58	42.492	30	0.3259636
60	42.492	30	0.3415553
62	42.492	30	0.3574973
64	42.492	30	0.3737609
66	42.492	30	0.3903150

Q9) How is the predicted probability affected by changing only Income holding all other variables constant?

The predicted probability increases as the income increases.

This is logical because when income increases, clients are more likely to have more income and hence purchase with a higher probability.