**WEEK 2:- GRADED ASSIGNMENT**

**1.** A test for “driving while intoxicated” was given 100 times. 20 people tested were actually intoxicated, and 10 people were mis-classified as intoxicated. What would the False Positive rate be?

***False Positive Rate : When it's actually no, how often does it predict yes?***

***False positive = 10; Actual no = Total(100)-Actual yes(20) = 80***

***= FP/actual no = 10/80 = 0.125 = 12.5%***

**ANSWER:- 12.5%**

**2.** If a fire alarm malfunctions and fails to go off when there actually is a fire, that is a:

**ANSWER:- False Negative(FN)**

**3.** Use the Binary Classification Metrics Spreadsheet Definitions to answer the following:

If the “classification incidence/test incidence” is 10% for the whole population, and the true “condition incidence” is 12% for the whole population, the True Positive rate:

***The condition actually present is given as: 12%; The condition predicted as present is given as: 10%***

***True Positive Rate = TP/ actual yes;***

***There are more Condition "+" than test "Positives." So not all "+" can be correctly identified by the test (hence, cannot be 100%)***

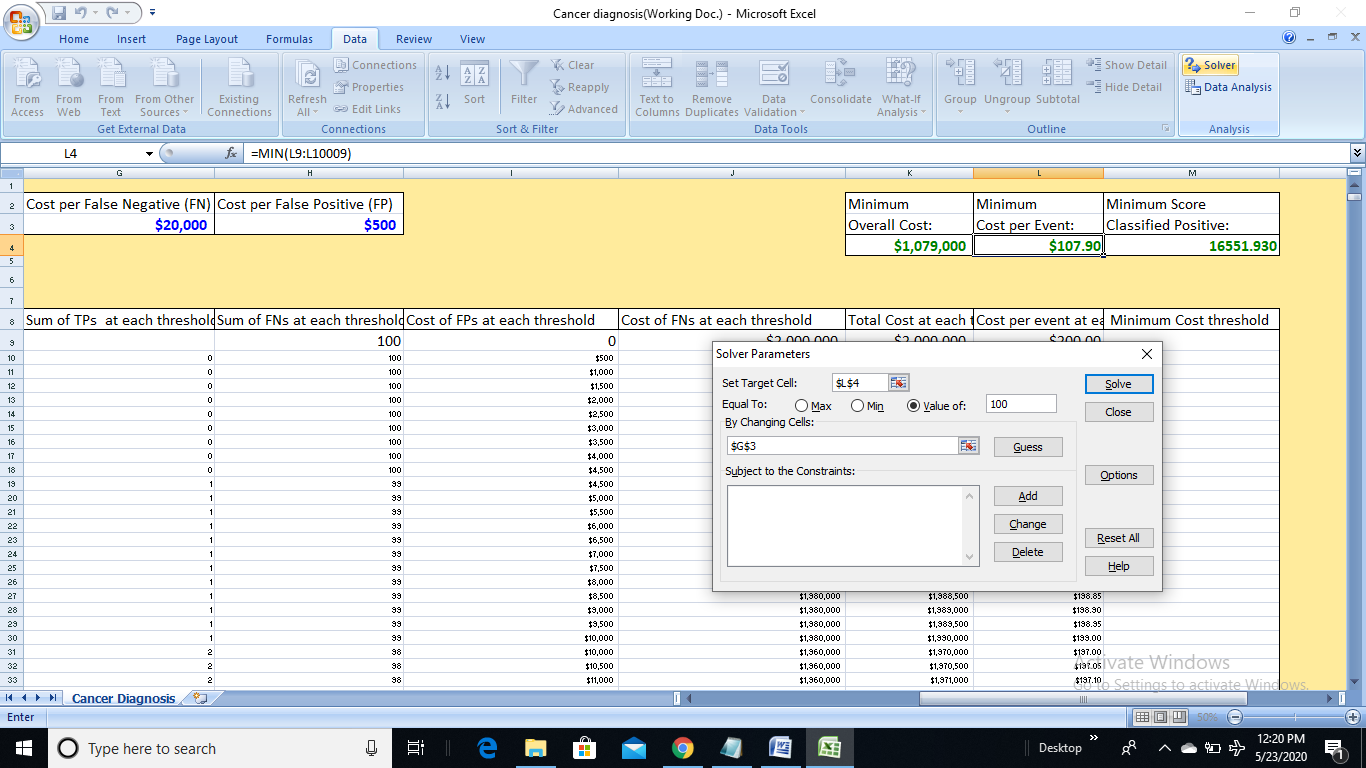
**ANSWER:- cannot be 100%**

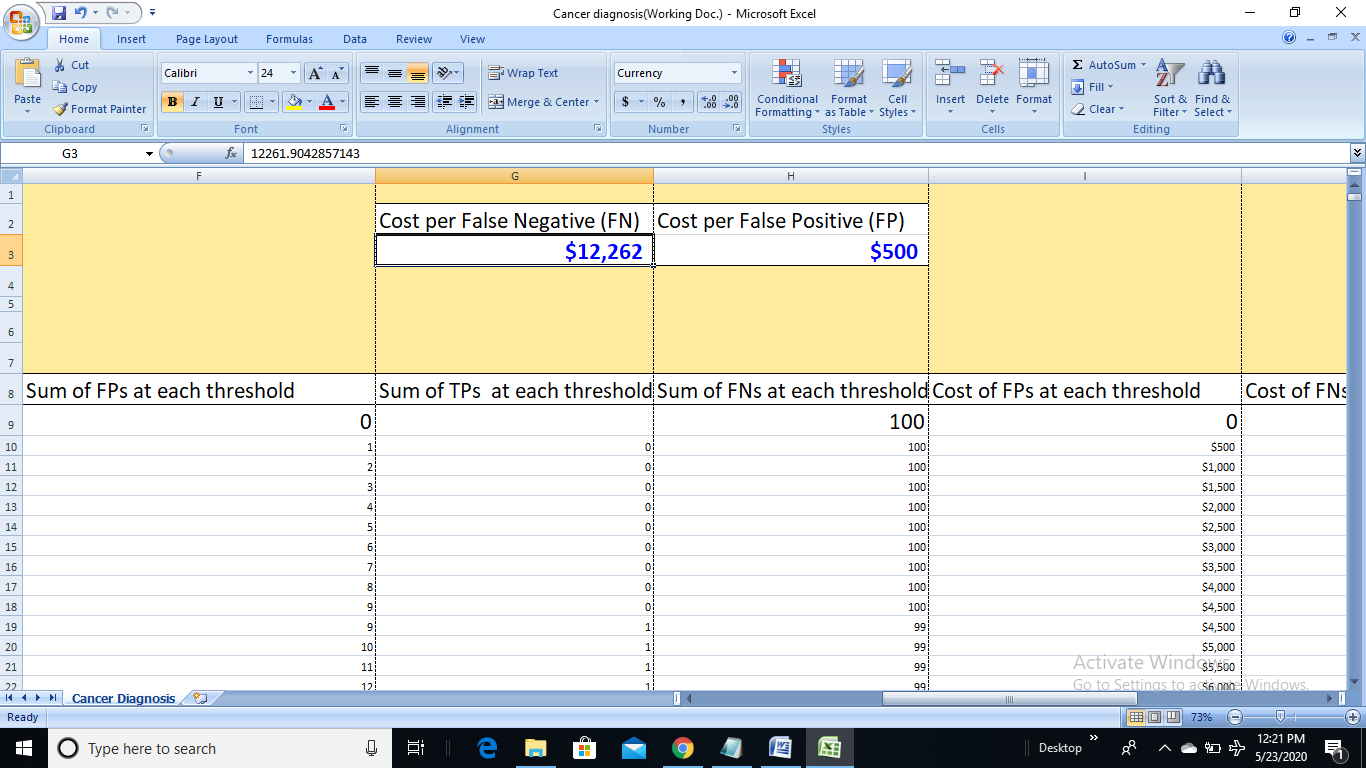
**4.** Use the *Cancer Diagnosis* Spreadsheet to answer Questions 4 to 6.

[Cancer Diagnosis.xlsx](https://d3c33hcgiwev3.cloudfront.net/_80a72b1369dfd15319274bec0f5f5f92_Cancer-Diagnosis.xlsx?Expires=1590278400&Signature=NcBhLv4ldf2zh0bzbE1-FSTRXN1Flj-72z2j70nojLSonPaY6I0iJqawgaSstJDvpMbvNYc3WtAvKQ0QO53W4mvxF5VwOe2zOXH9BlfkYV6YCSmiPnHgWQUc07G18spkUBYkfwty-JY~bGzOzjF2KSGLpN975jYGfo7TMZST5WE_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

Keep the cost per False Positive test set at $500. Use MS Solver to determine the maximum cost per False Negative test that permits an average cost per test of $100.

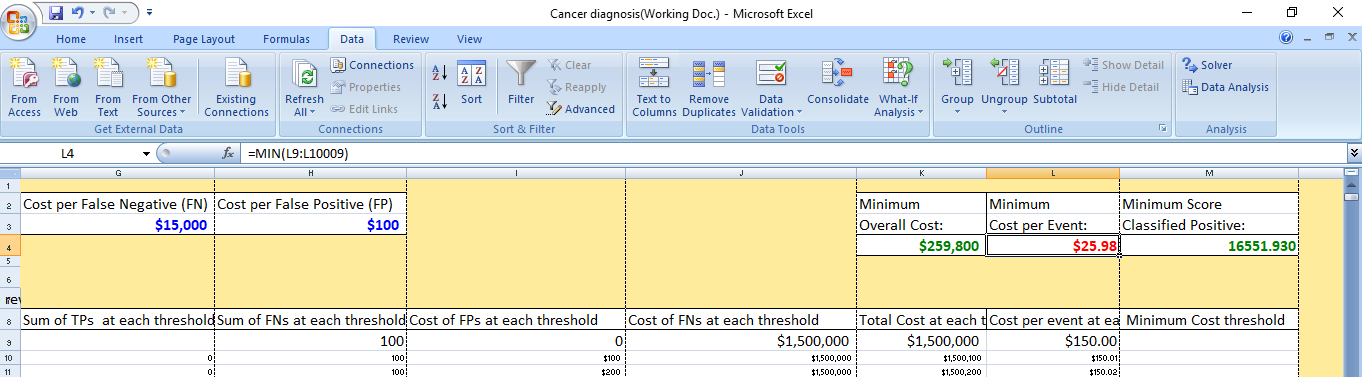
***Use Solve option available in Data Tab under Analysis Group, and follow the below procedure.***





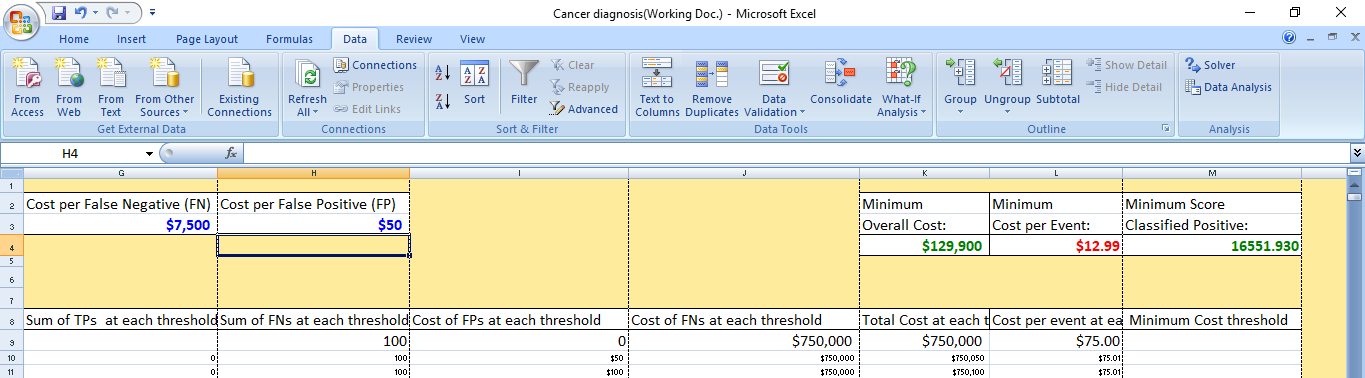
**ANSWER:- $12,262**

**5.** Assume a cost of $15,000 per False Negative (FN) and $100 per False Positive (FP). What is the minimum average cost per test?

****

**ANSWER:- $25.98**

**6.** If, instead of assuming a cost $15,000 per FN and $100 per FP, the costs are assumed to be $7,500 per FN and $50 per FP, what changes?

****

***On comparing above two questions, on changing the values; (Minimum cost per event changed from $25.98 to $12.99)***

**ANSWER:- The Minimum Cost per Test**

**7.** Use logic and the definition in the Binary Performance Metrics Spreadsheet to answer the following question.

[Binary Performance Metrics.xlsx](https://d3c33hcgiwev3.cloudfront.net/_6f81298f196a69e651cd11d36ee96650_Binary-Performance-Metrics.xlsx?Expires=1590278400&Signature=FqK5cKGtl8eWDjjxqHNu-mrkuPy8o0a-m0PFZK-ipZ1jFJhcrHCkUIPmfosxhPd~BG6yYUbaXxwNltvFhAhAPaNm-2-UG7855ZzT3wFotCvQyCY9JDAir8rNDUVbeRCLzMEsIcqOXmoafSo0~8QzpWjd1dB6yAS0OZKK~UEQ~Gc_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

In general, increasing the cost per FN while keeping the cost per FP constant will cause the cost-minimizing threshold score to:

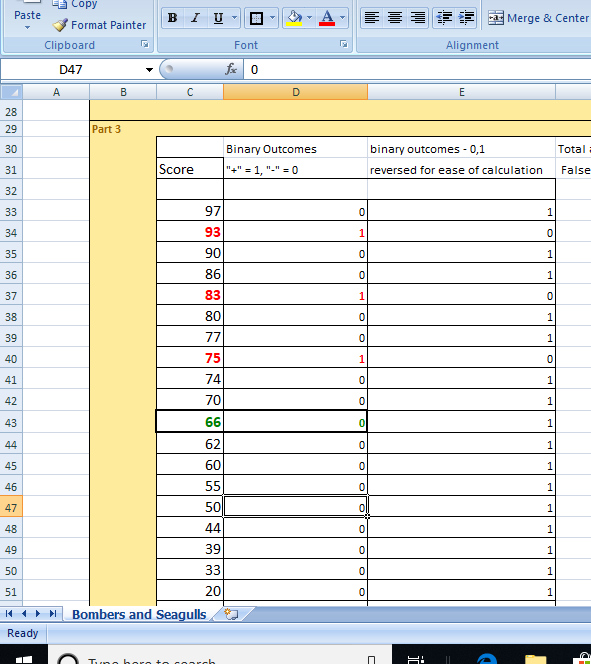
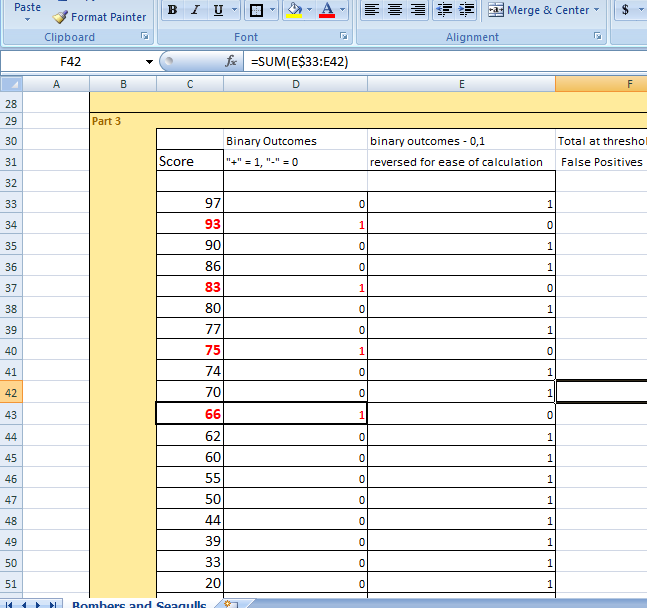
**ANSWER:- Decrease**

**8.** Make a copy of the *Bombers and Seagulls*Spreadsheet to answer questions 8-10.

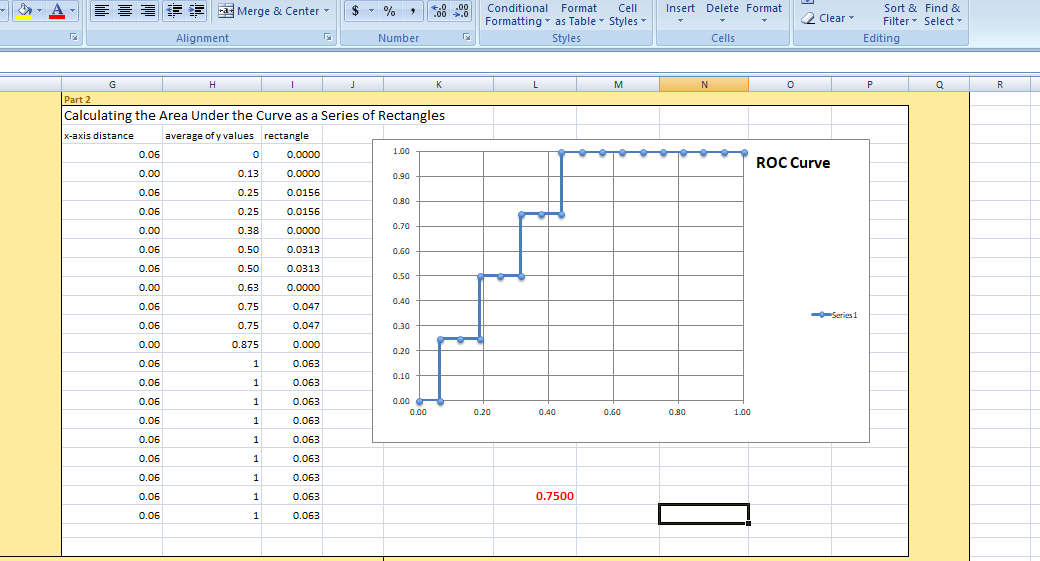
[Bombers and Seagulls.xlsx](https://d3c33hcgiwev3.cloudfront.net/_80a72b1369dfd15319274bec0f5f5f92_Bombers-and-Seagulls.xlsx?Expires=1590278400&Signature=BIvtrYVTn00zHhTRWjOZvQDM4JebwK0nvpIFKZ6N5iYmBFvR~1JV7UAZ0HQimOb2TE~pnMlQGrRNvEKSVTjwezxFc3GRUZiYUGQLrIcIv7gak8ztMjiL7tSF3X2zAWYd7HsswoOlP4YuMSrMd-04jD7HduAN2B7EL2lihnbJEOg_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

Modify the spreadsheet data so that there are 4 bombers instead of 3, and 16 seagulls instead of 17, by changing the actual condition for the radar score of 66 from a 0 to a 1 in cell D43.

**What is the new Area under the Curve:**

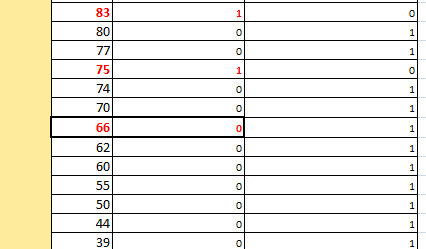
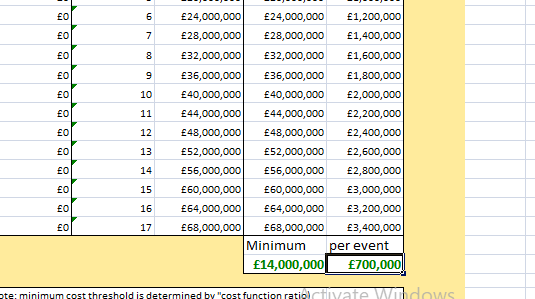
 

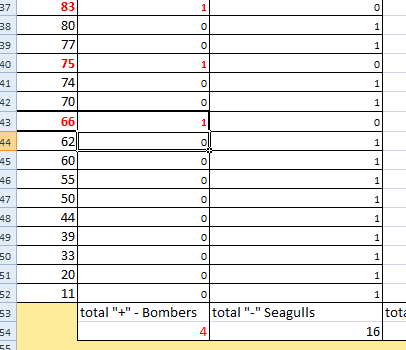
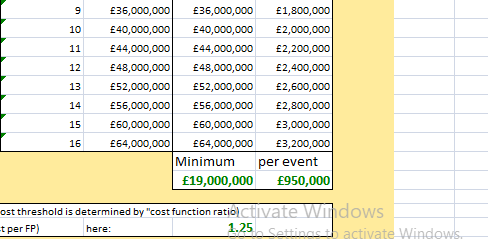
***Changing the value of 66 from black to red(seagulls to bombers); Area Under the curve value is:***

******

**ANSWER:- 0.75**

**9.** Assuming the costs for classification errors are 5 million pounds per FN and 4 million pounds per FP, how much does changing the value at Cell D43 from 0 to 1 change the minimum cost per event?

** **

** **

***So, the cost per event changed from £700,000 to £950,000***

**ANSWER:- Increases by 250,000 pounds**

**10.** Change the cost per FN to 50 million pounds. How does changing the data in cell D43 from a 0 to a 1 change the cost-minimizing threshold?

***[Due to some error, unable to post screenshots]***

**ANSWER:- Decreases it from 75 to 66**

**11.** Use the *Binary Performance Metrics*Spreadsheet definitions to answer the following question.

[Binary Performance Metrics.xlsx](https://d3c33hcgiwev3.cloudfront.net/_6f81298f196a69e651cd11d36ee96650_Binary-Performance-Metrics.xlsx?Expires=1590278400&Signature=FqK5cKGtl8eWDjjxqHNu-mrkuPy8o0a-m0PFZK-ipZ1jFJhcrHCkUIPmfosxhPd~BG6yYUbaXxwNltvFhAhAPaNm-2-UG7855ZzT3wFotCvQyCY9JDAir8rNDUVbeRCLzMEsIcqOXmoafSo0~8QzpWjd1dB6yAS0OZKK~UEQ~Gc_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

A population tested for “driving while intoxicated” has a Condition incidence of 20%. If the test has a true positive rate of 70% and a false positive rate of 10%, what is the test’s Positive Predictive Value (PPV)?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  | Confusion Matrix |  |  |  |  |  |  |  |  |
|  |  |  |  |  | Test Classification  **Y** | | |  |  |
|  |  |  |  |  | "Positive" |  | "Negative" | |  |
|  |  |  |  |  |  | c |  | d |  |
|  |  |  |  |  |  |  |  |  |  |
|  | Condition **X** | "+" |  | a |  | e |  | f |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  | "-" |  | b |  | g |  | h |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

***Given, Condition-Incidence (a) =20%; True Positive rate (e/a) = 70%;***

* ***a = 0.2 ; && e/a = 0.7 🡺 e = 0.14;***

***False Positive rate (g/b) = 10%;***

* ***g/b = 0.1; (Since, a+b=1; 🡺 b= 0.8) 🡺 g = 0.08***

***Positive-Predictive-Value: e/c (Since, c = e+g; 🡺 c = 0.22) ; Hence, PPV = 0.14/0.22 = 0.6363 = 0.64***

**ANSWER:- 0.64**

**12.** Use the *Soldier Performance* Spreadsheet to answer question 12.

[Forecasting Soldier Performance.xlsx](https://d3c33hcgiwev3.cloudfront.net/_18267f8cf82cde5920753ca6efa3589e_Forecasting-Soldier-Performance.xlsx?Expires=1590278400&Signature=Ho9kP-wzDWhDQu-~wYhx~7kThSWFl8LizYcReSs2adaEGFPk5QHd4txkUacp~T471dD4B64txan~LpAshdHuva56XdH9542DvTE39T9CaOFG0bAqh03r1ATSp1-WFTmvcR6hsjccsklNLdTpv9crhETbAZYkITcEJZXLr6FqZFg_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

Rank the outcomes using soldier’s age as the score, with the oldest at the top. A threshold of 24 years represents what point on the ROC Curve?

***Given, Threshold = 24; (First sort the age column and select expand the selection; and sort from oldest to youngest)***

***Each (x,y) point on the ROC Curve is a False Positive rate and a True Positive rate.***

***Divide the number of 0s above the threshold (4) by 6 (condition negatives) to get the False Positive rate; and the number of 1s above the threshold (2) by 6 (condition positives) to get the True Positive rate.***

***False Positive Rate = 4/6 = 0.67; True Positive Rate = 2/6 = 0.33;***

**ANSWER:- .67, .33**