1. Which of the following are assumption of simple linear regression?

   - Linearity
   - Homoscedasticity
   - Independence
   - Normality

2. Which of the statements are true on classical time series models?

   - The additive model can be used when the seasonal variation is constant over the time
   - The multiplicative model can be used when the seasonal variation increases over the time

3. What is the standard query technique that can be used to aggregate the information within the OLAP?

   - Roll Up

4. Identify the situations cannot use as big data related problems?

5. What does veracity in big data mean?
   measuring accuracy or truthfulness/trustworthiness of high volumes of data is hard

6. In OLAP operations. slicing is the technique of_____

   - Selects one particular dimension from a given cube and provides a new sub-cube

7. Which among the following are the duties of the NameNodes

   - manage file system namespace
   - (keep of the following are the duties of the namenodes)

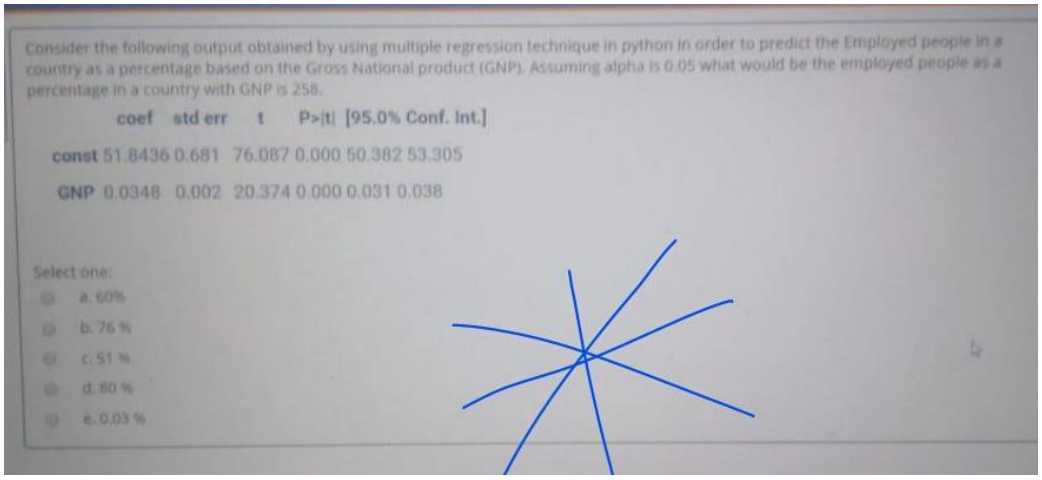
8. Which of the following incident/s can be apply the time series analysis?

9. Which one of these is not a tree based learner?

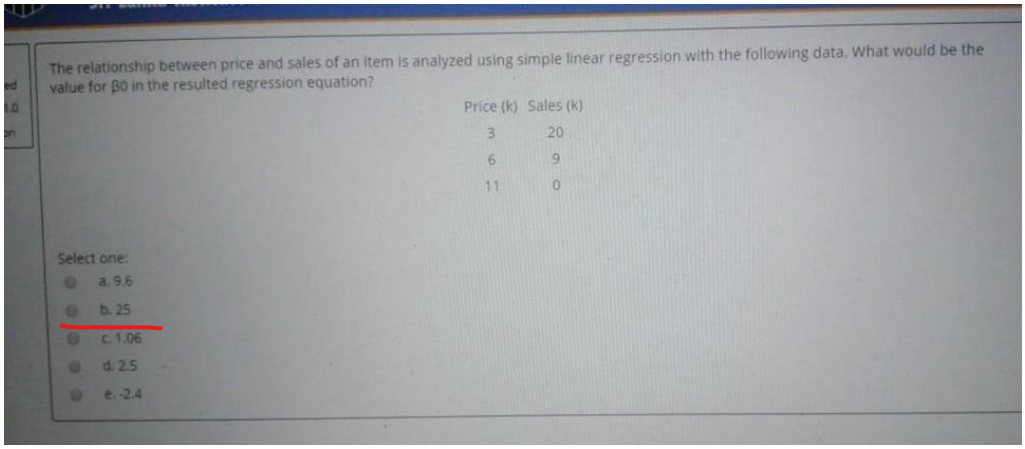
   - bayesian classifier
   - neural network

10. which of the following is/are not a classification technique?
    - DBscan

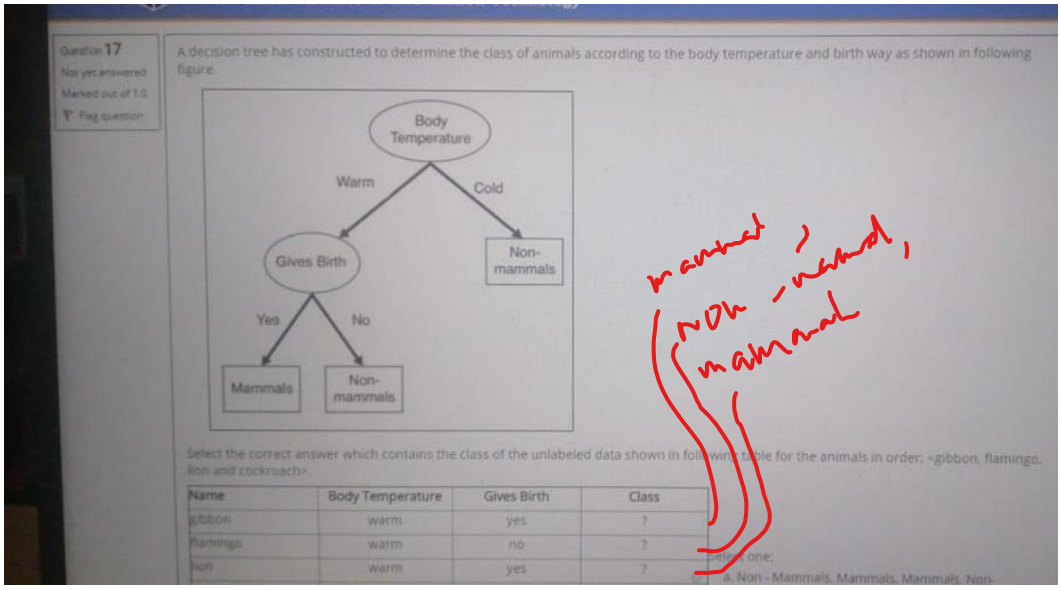
Consider the following output obtained by using multiple regression



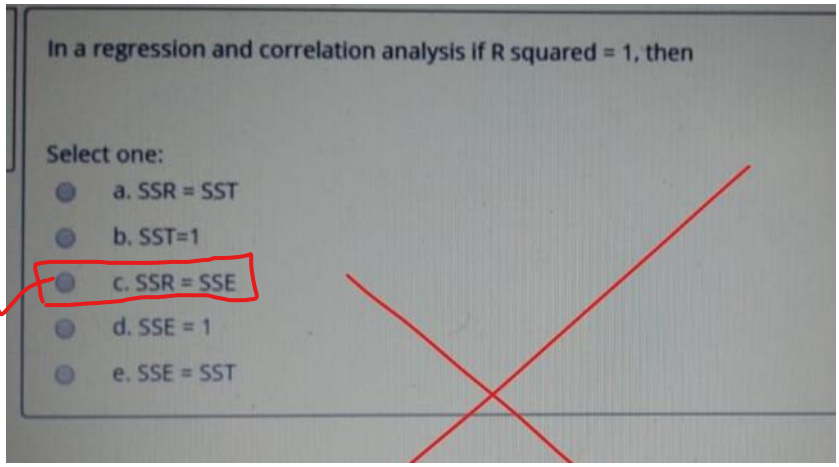
The relationship between price and sales



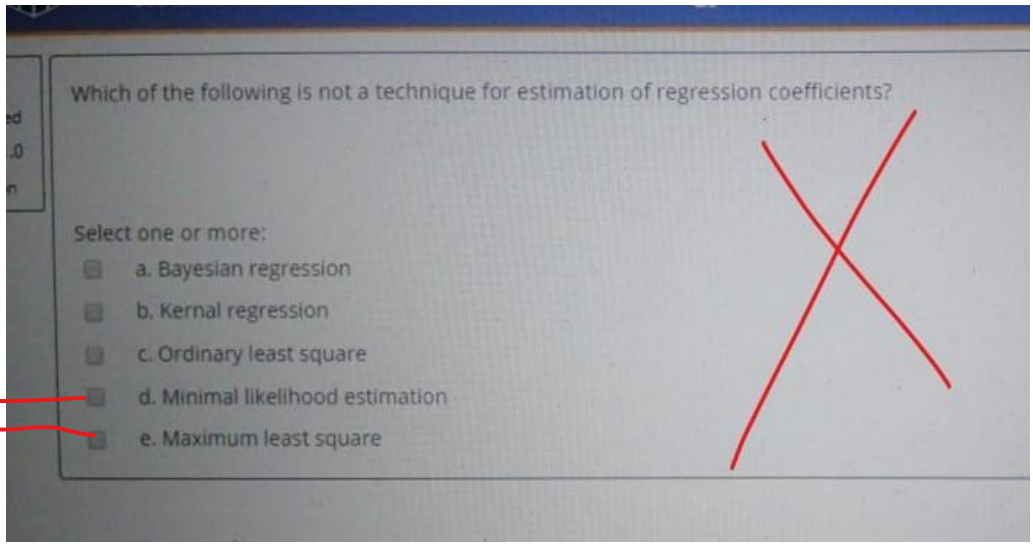
A decision tree has constructed to determine



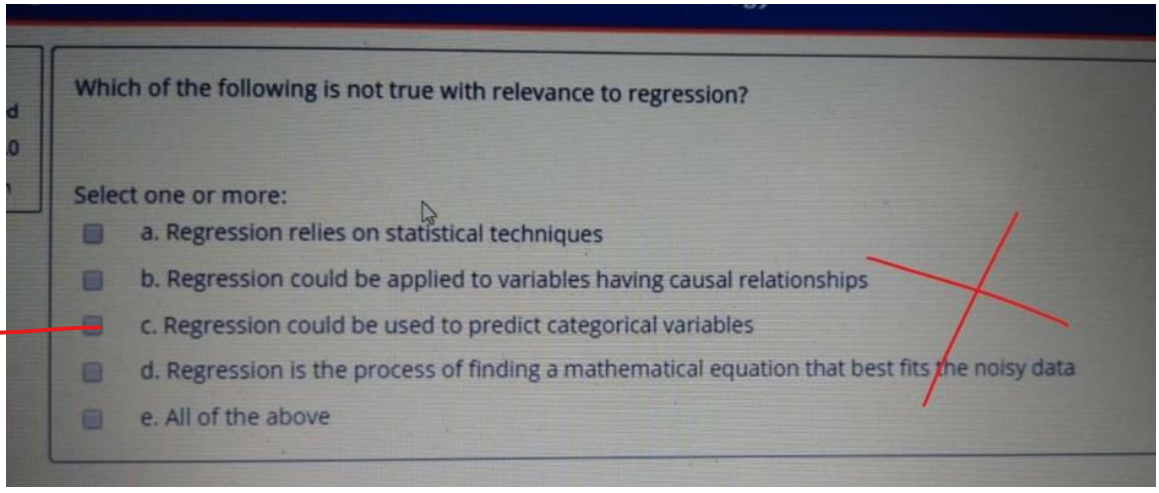
In a regression and correlation analysis



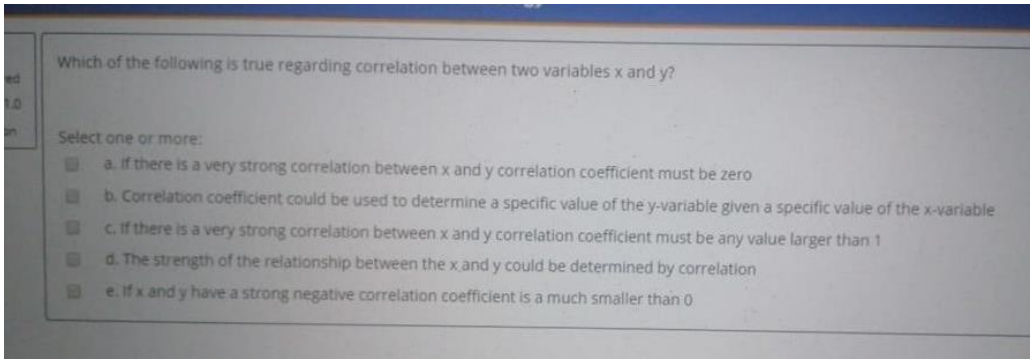
Which of the following is not a technique



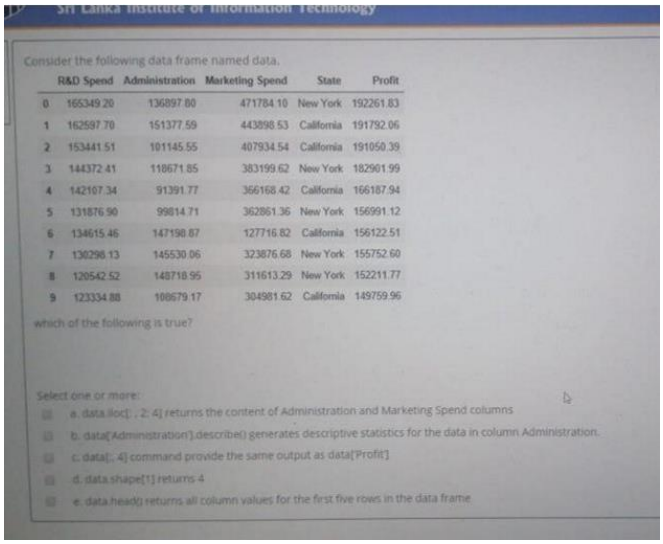
Which of the following is true with relevance



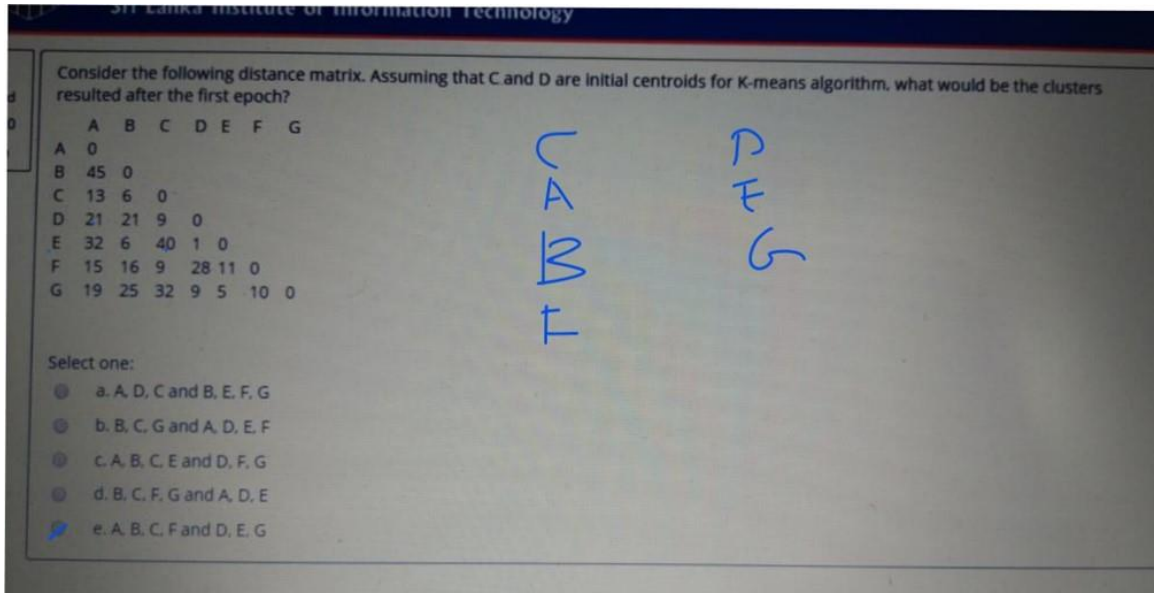
Which of the following is true regarding correlation



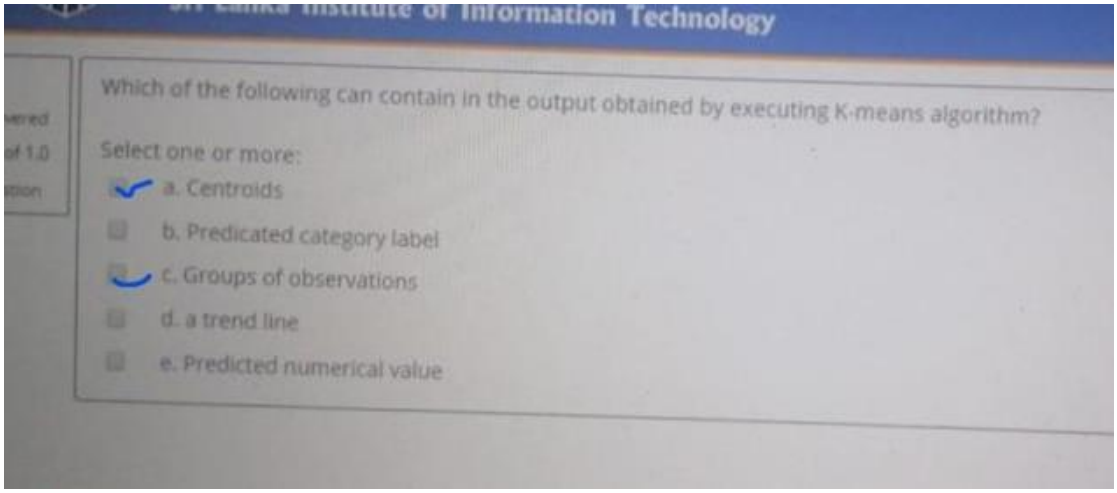
Consider the following data frame



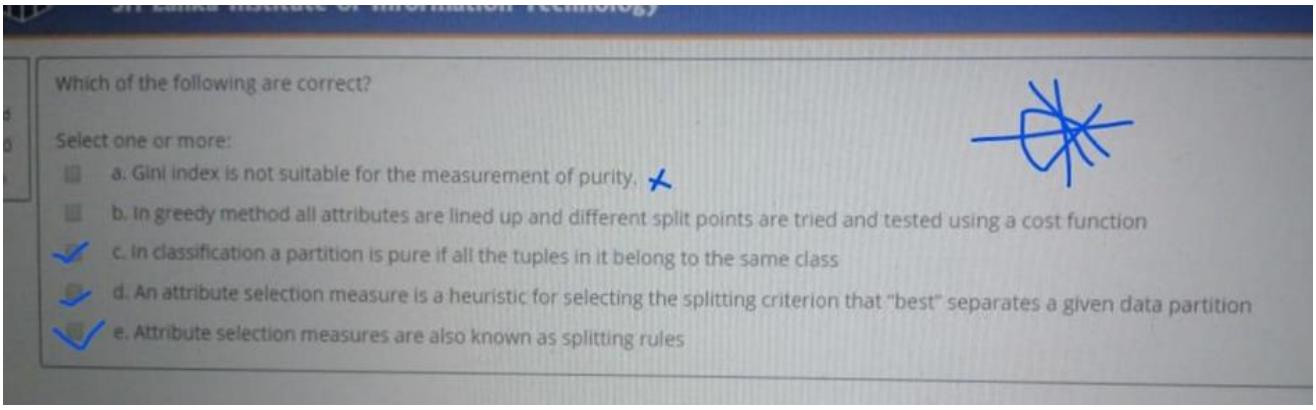
Consider the following distance matrix.



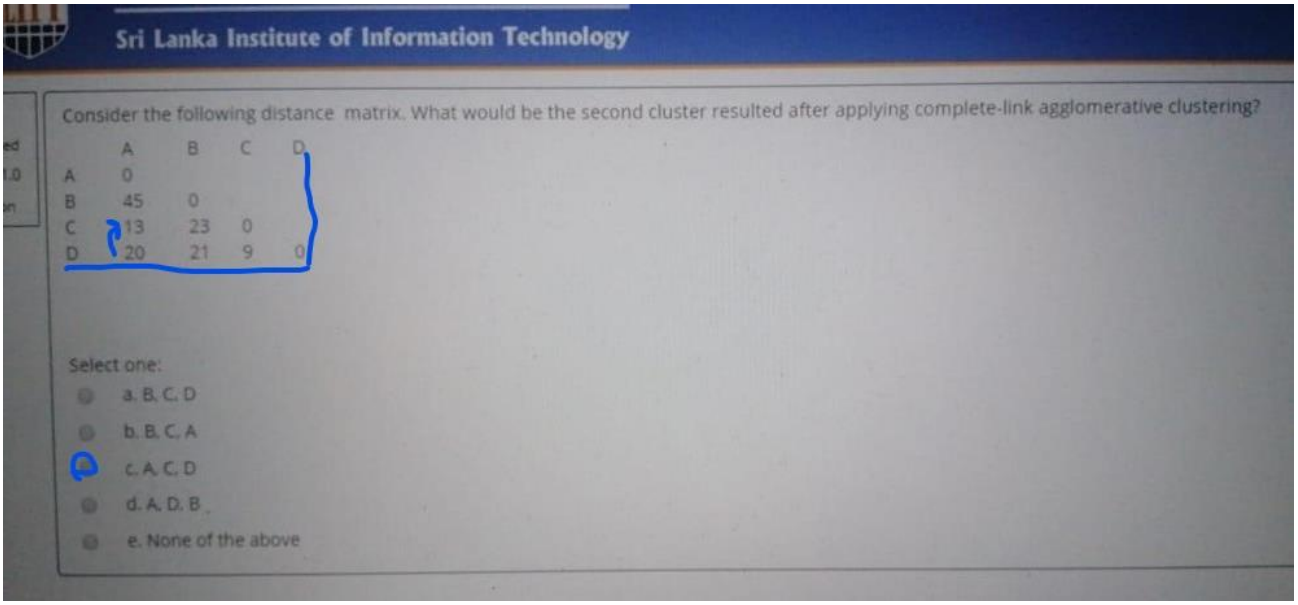
Which of the following can contain in the output obtained



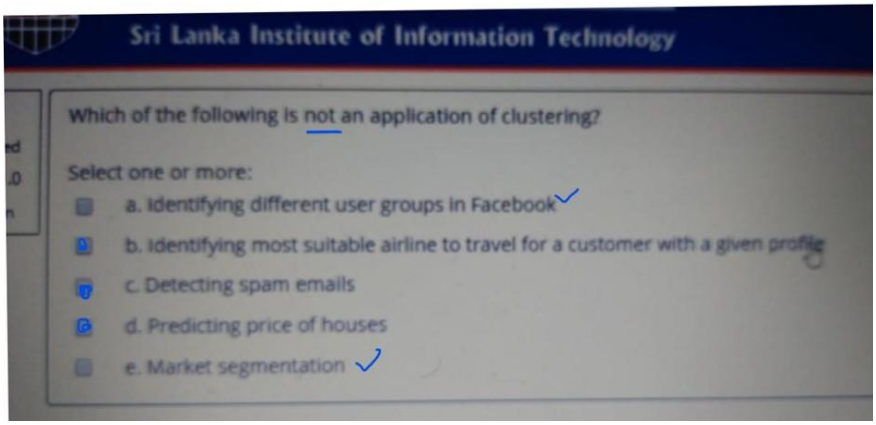
Which of the following are correct



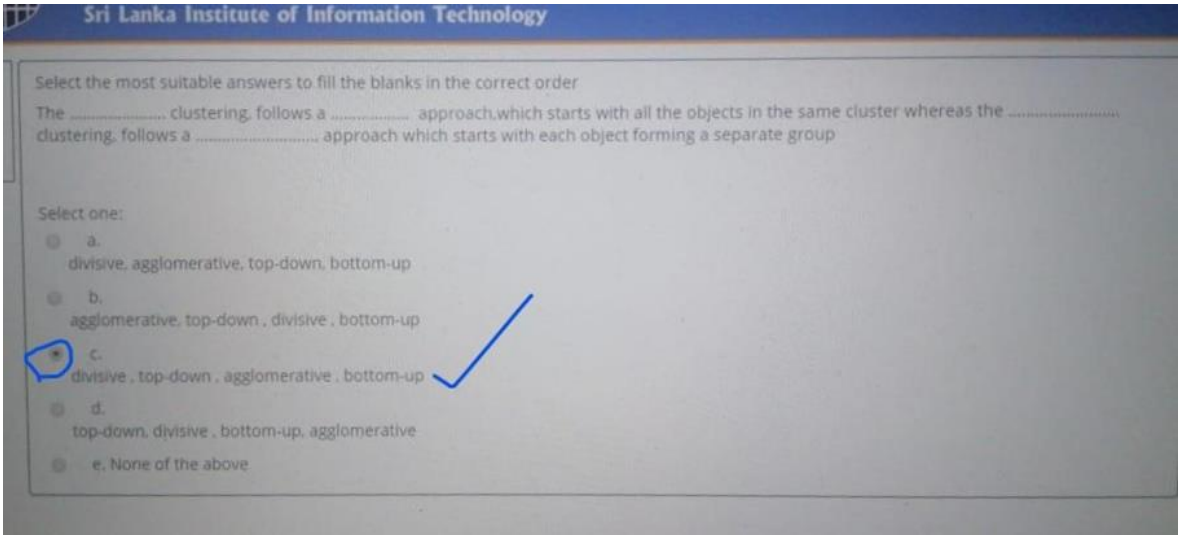
Consider the following distance matrix.



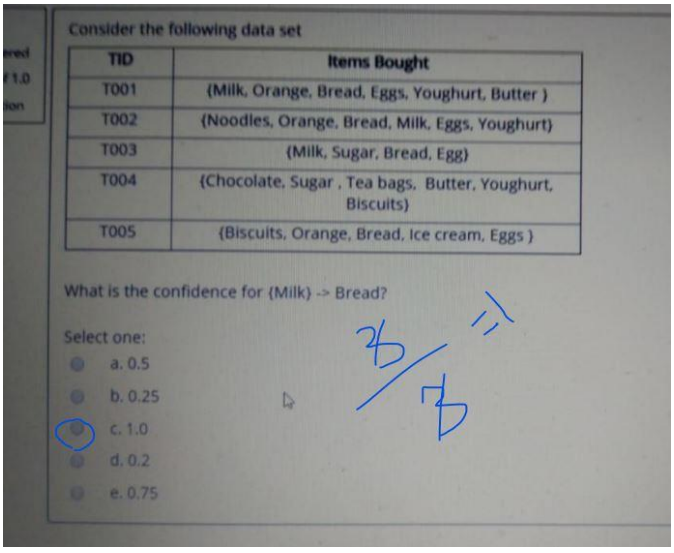
Which of the following is not an application of clustering



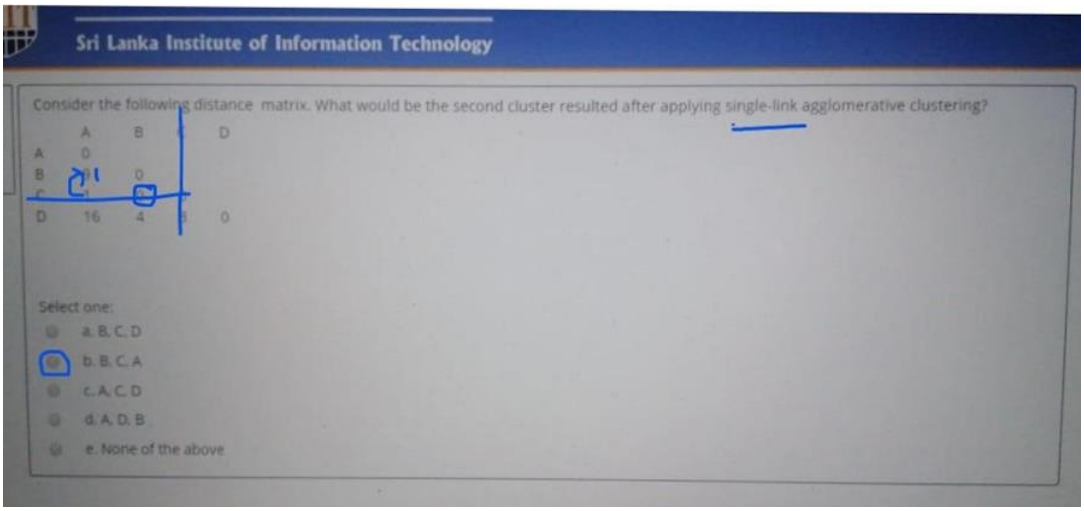
Select the most suitable answers to fill the blanks



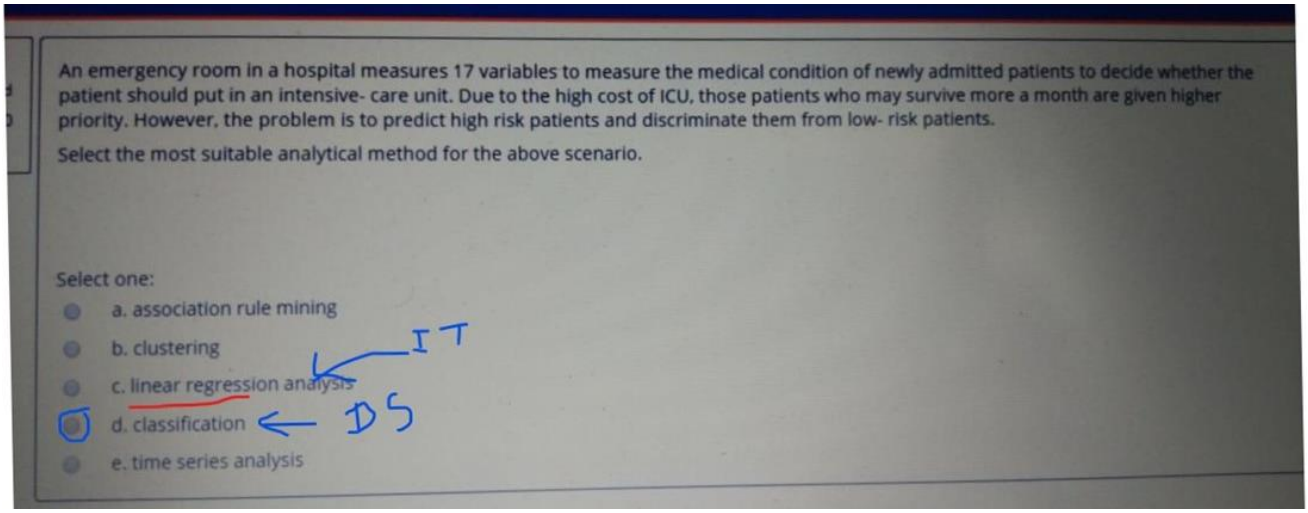
Consider the following data set



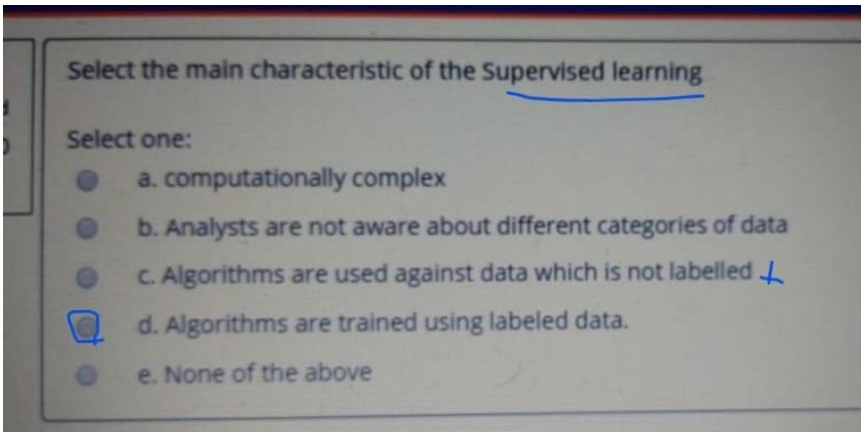
Consider the following distance matrix.



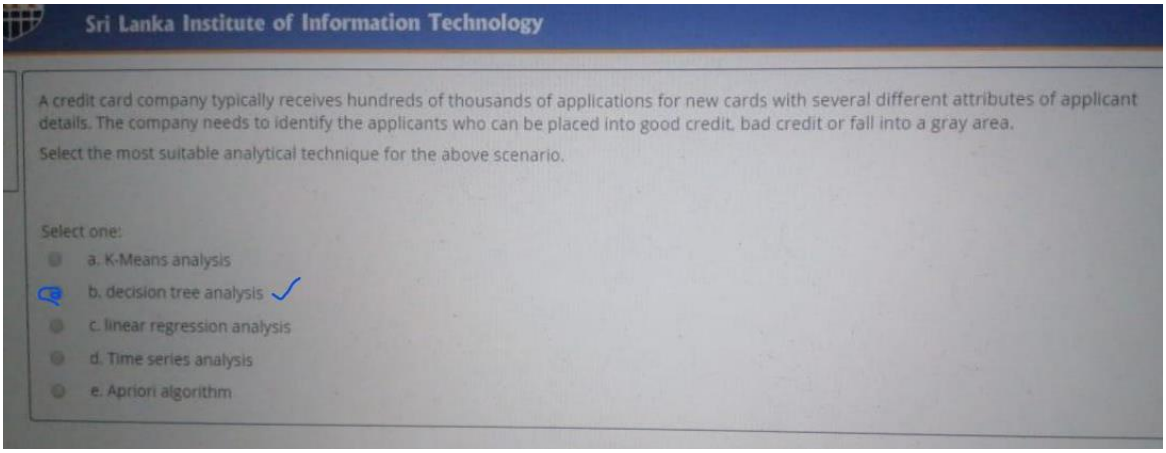
An emergency room in a hospital measures 17



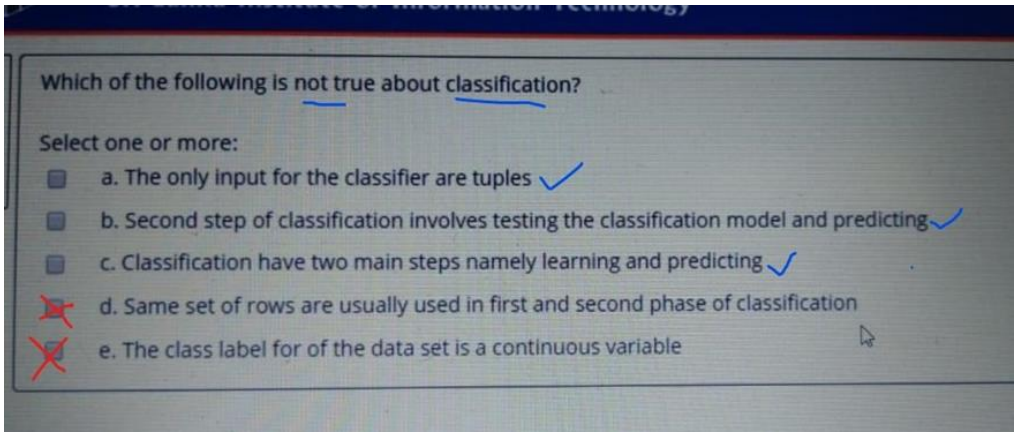
Select the main characteristic of the Supervised learning



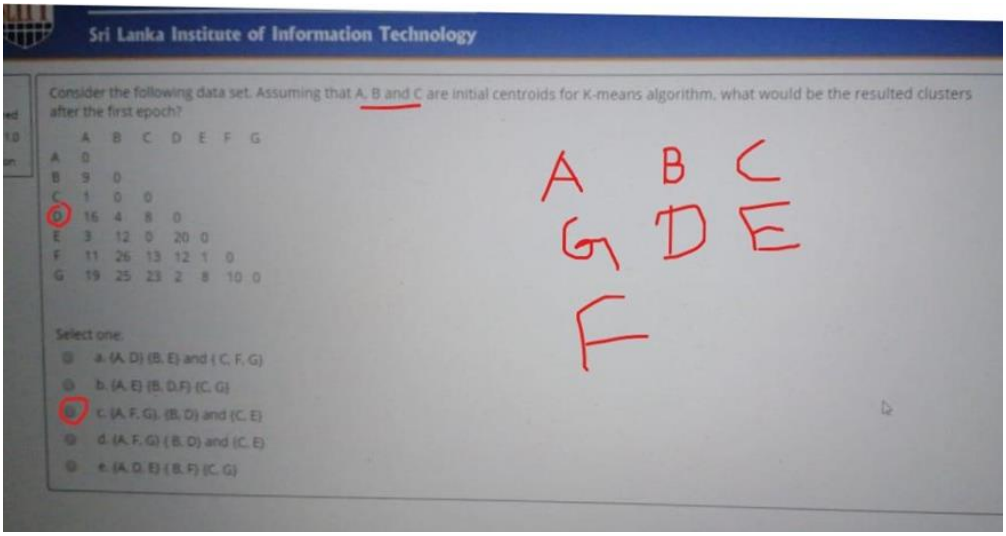
A credit card company typically



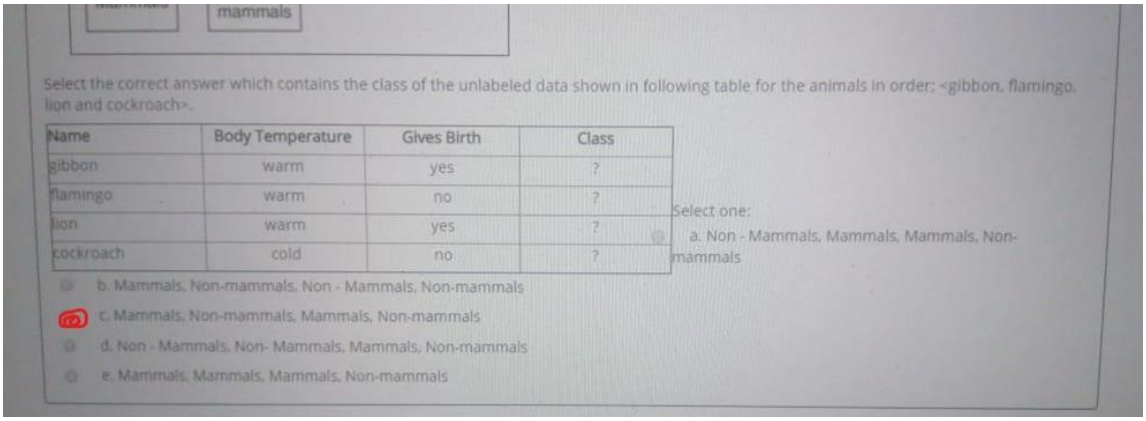
Which of the following is not true



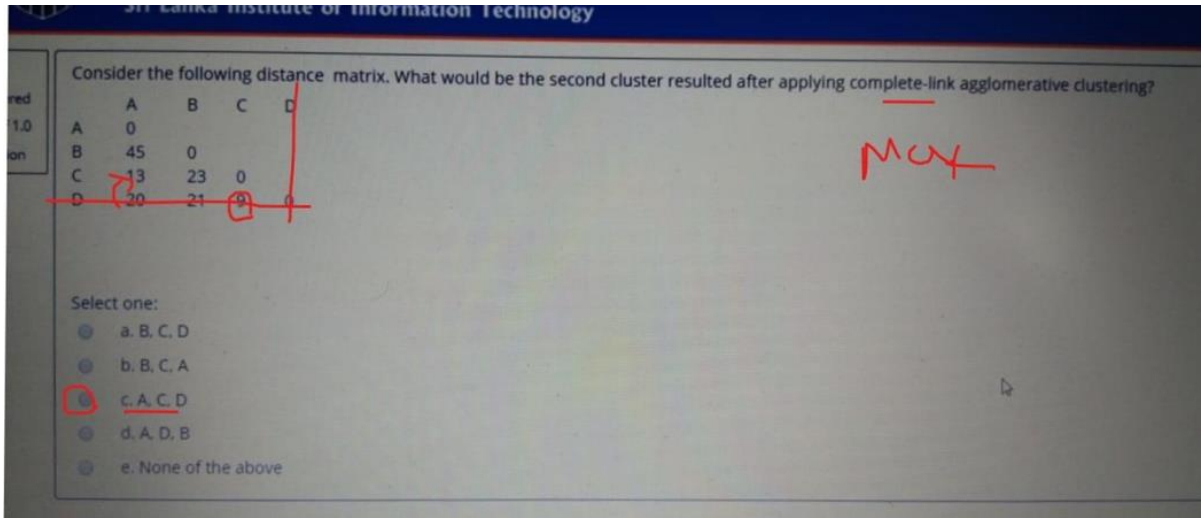
Consider the following data set



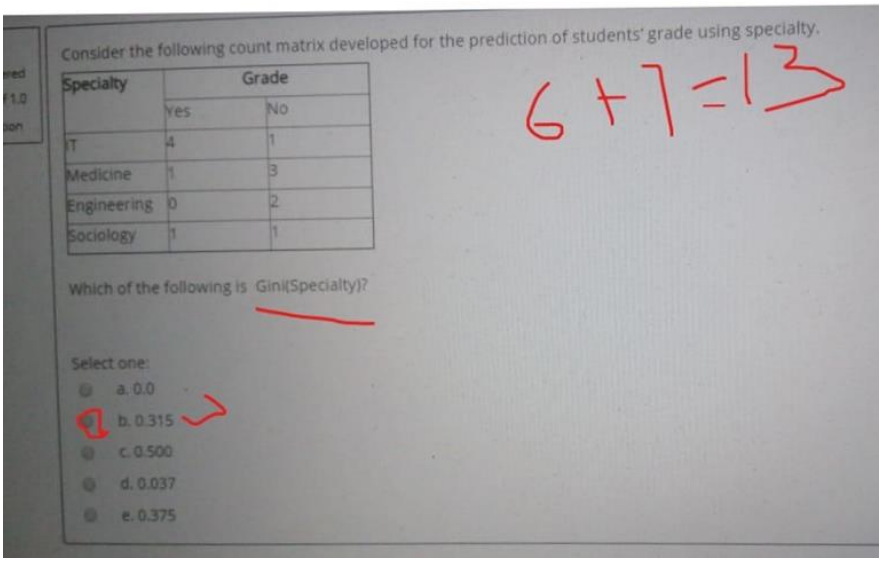
Select the correct answer which contains the class



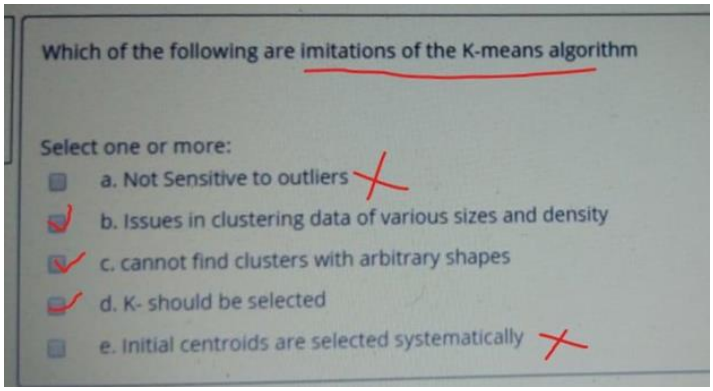
Consider the following distance matrix



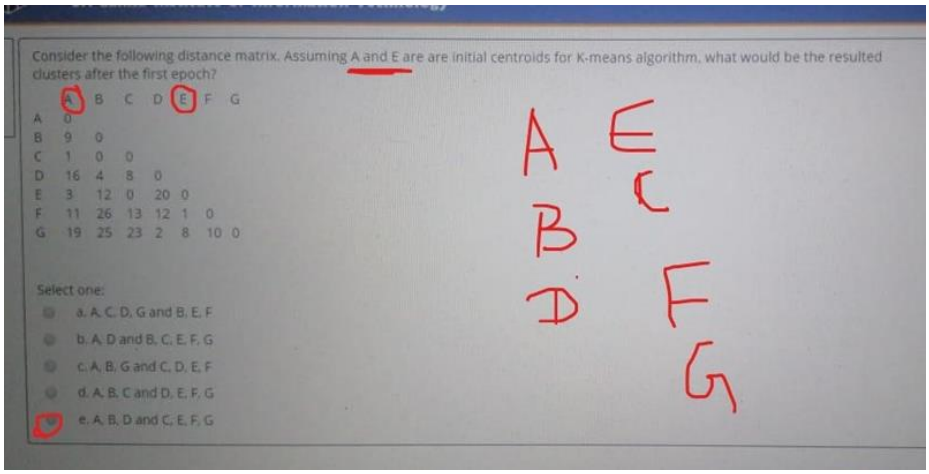
Consider the following distance matrix developed



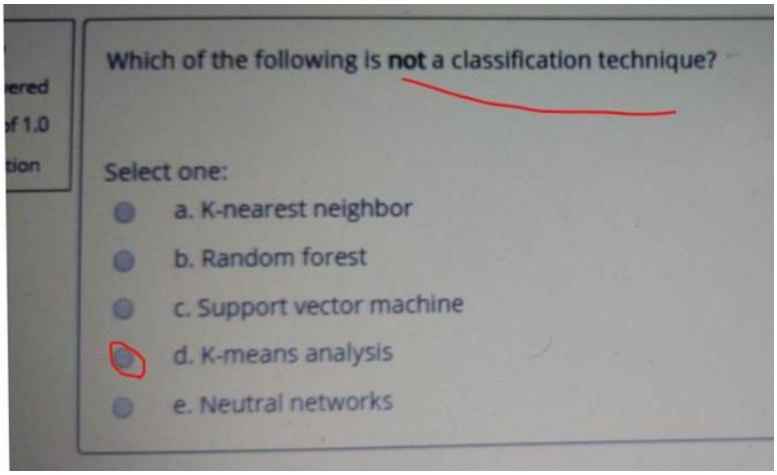
Which of the following are imitations



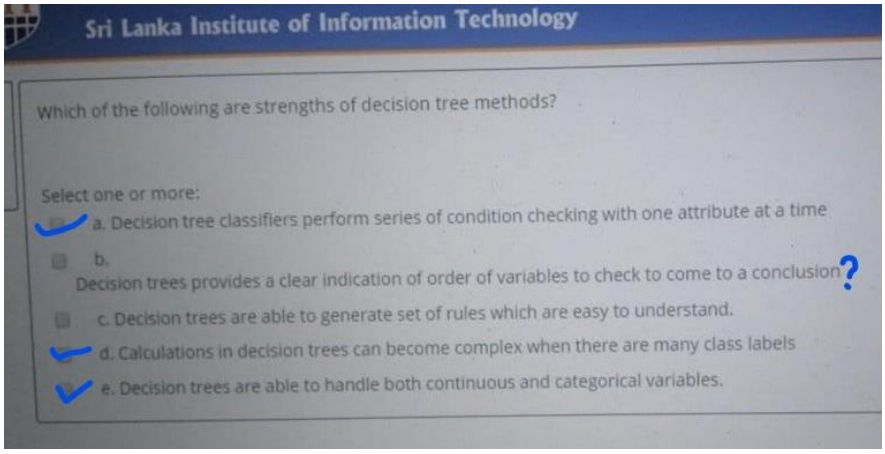
Consider the following distance matrix



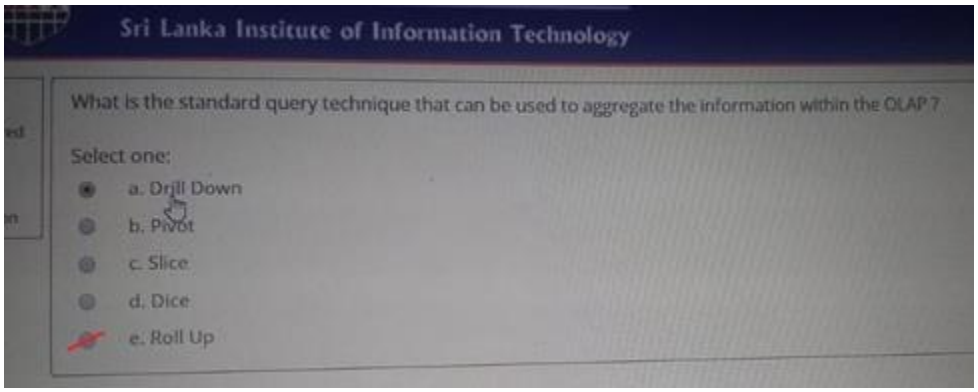
Which of the following is not a classification



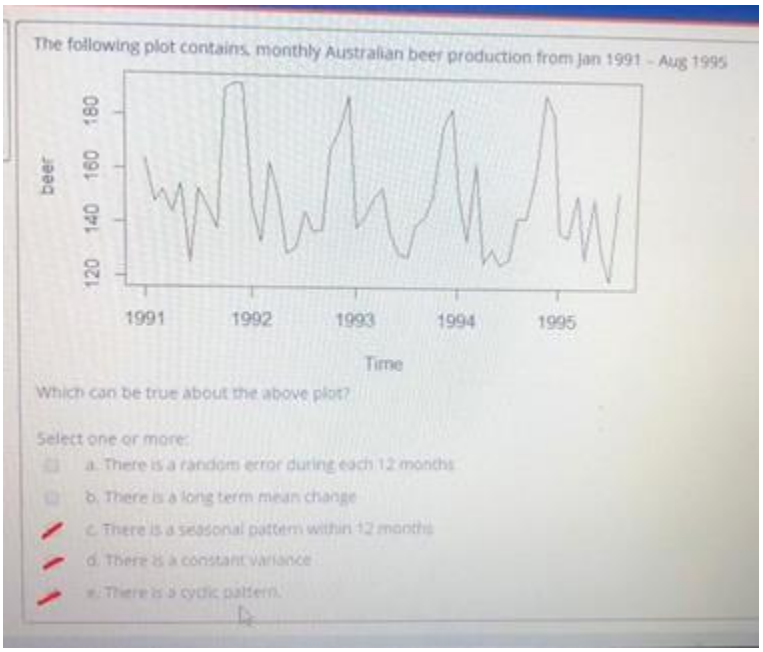
Which of the following are strengths



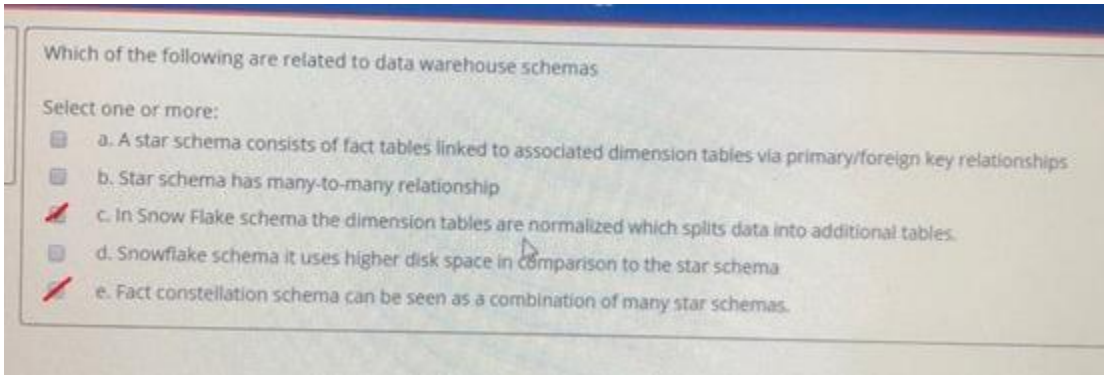
What is the standard query technique that can be used



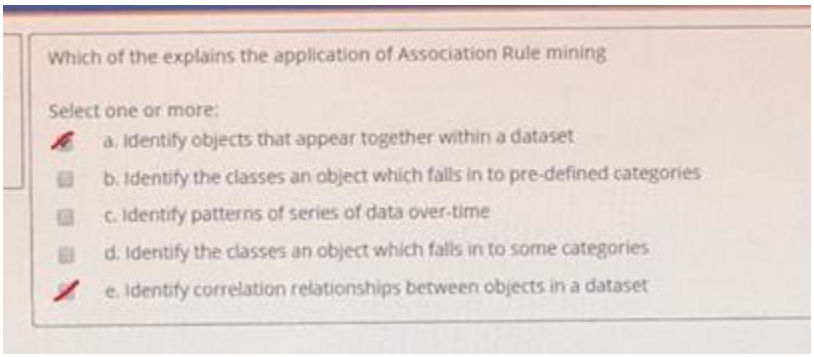
The following plot contains



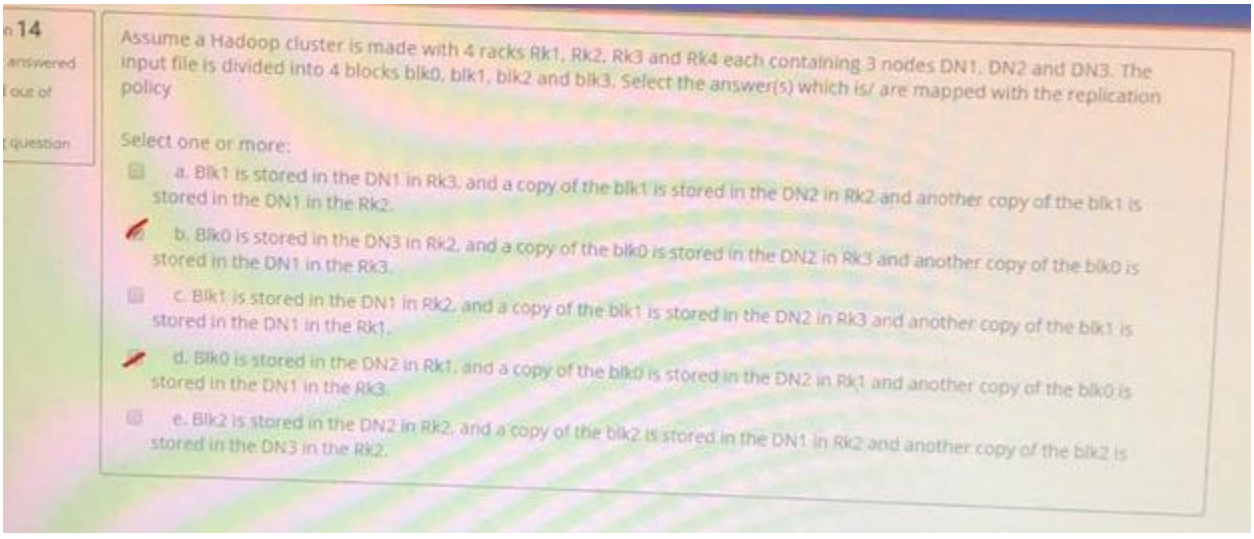
Which of the following are related



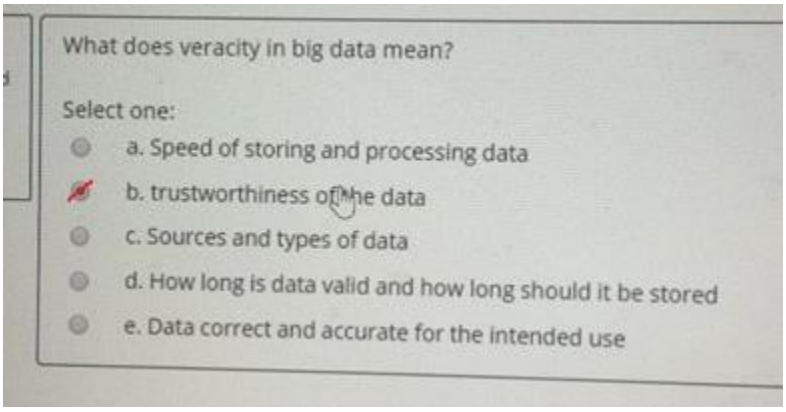
Which of the explains the application



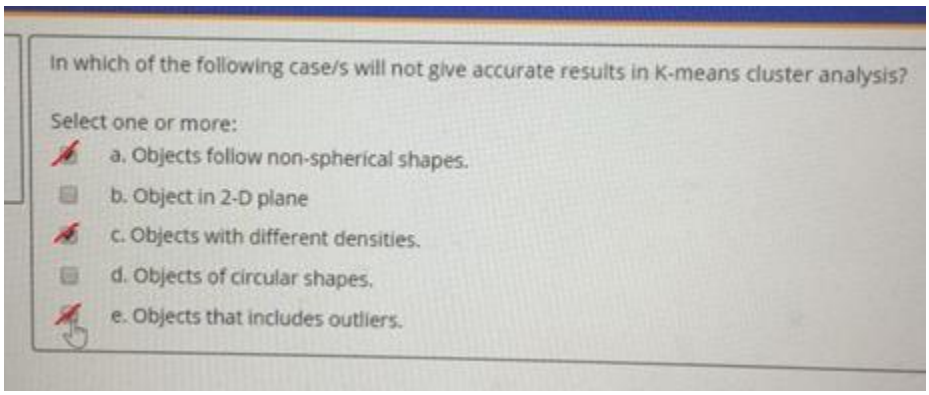
Assume a Hadoop cluster is made with 4 racks



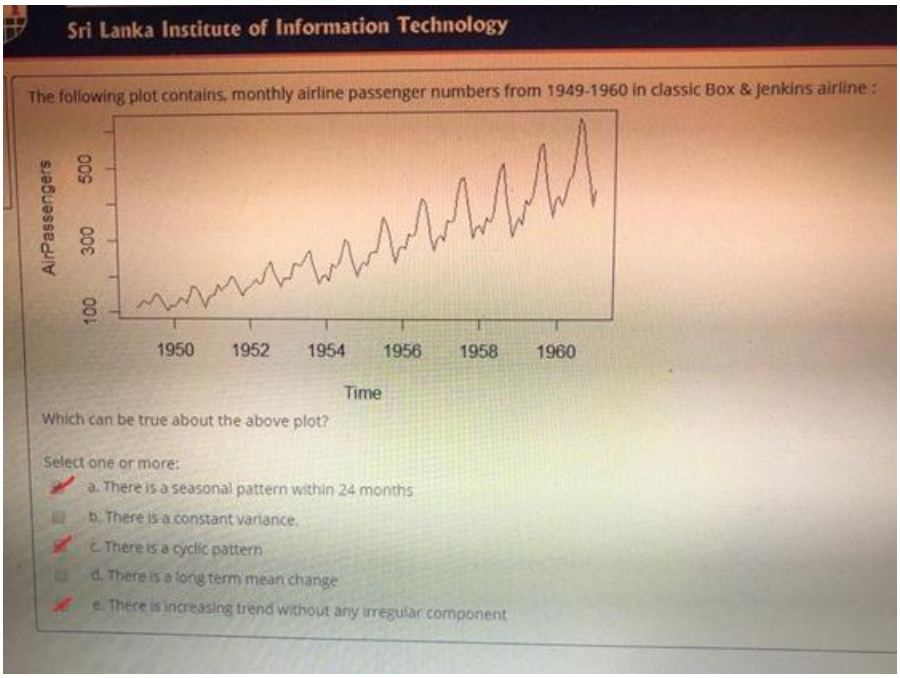
What does veracity in big data mean



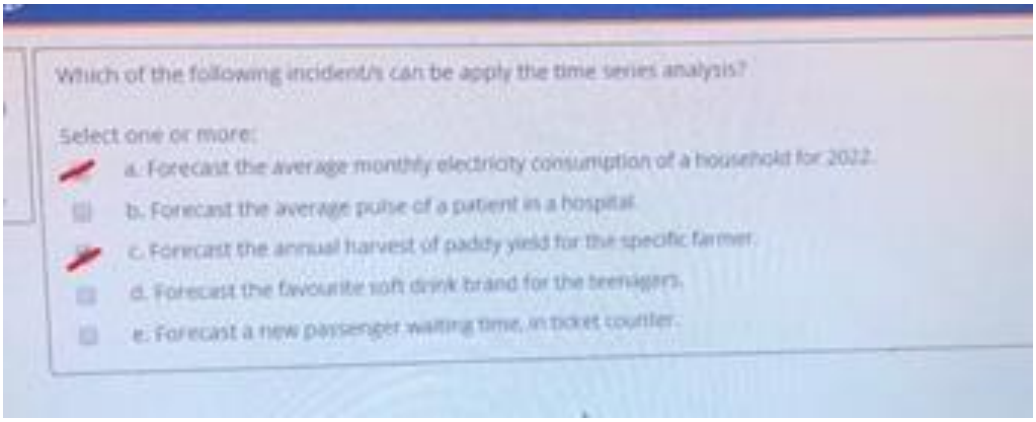
In which of the following case/s will not give accurate results



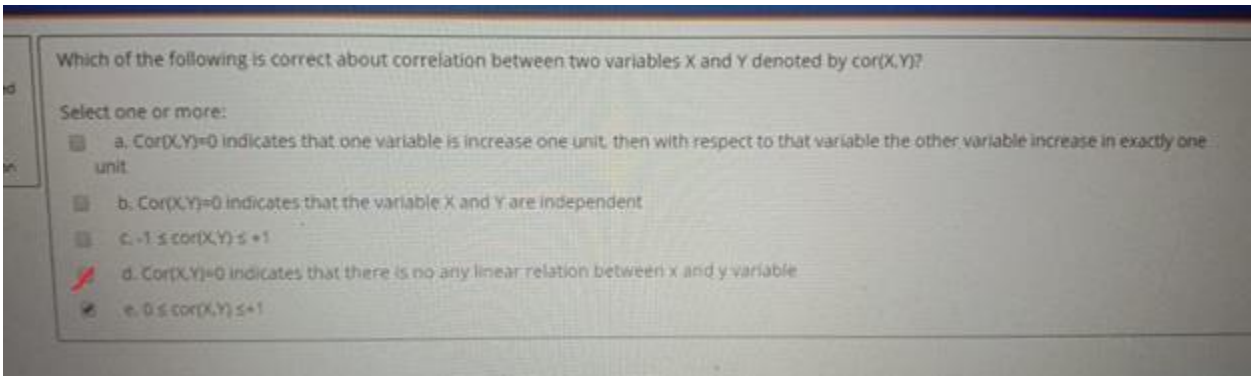
The following plot contains



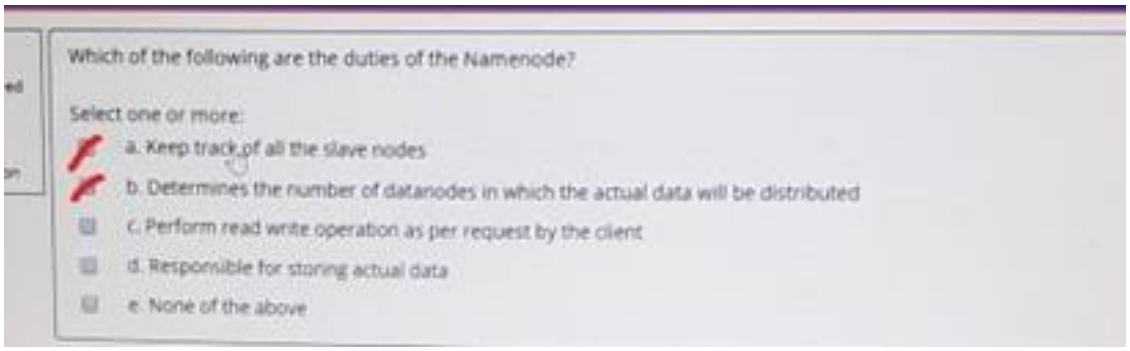
Which of the following incident



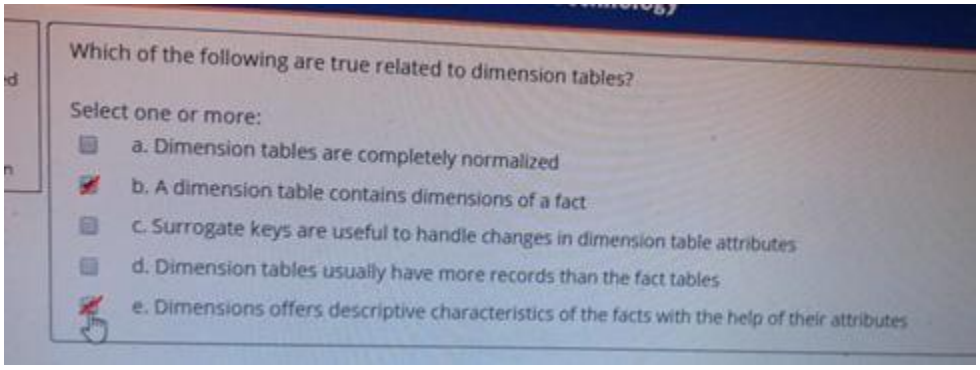
Which of the following is correct about



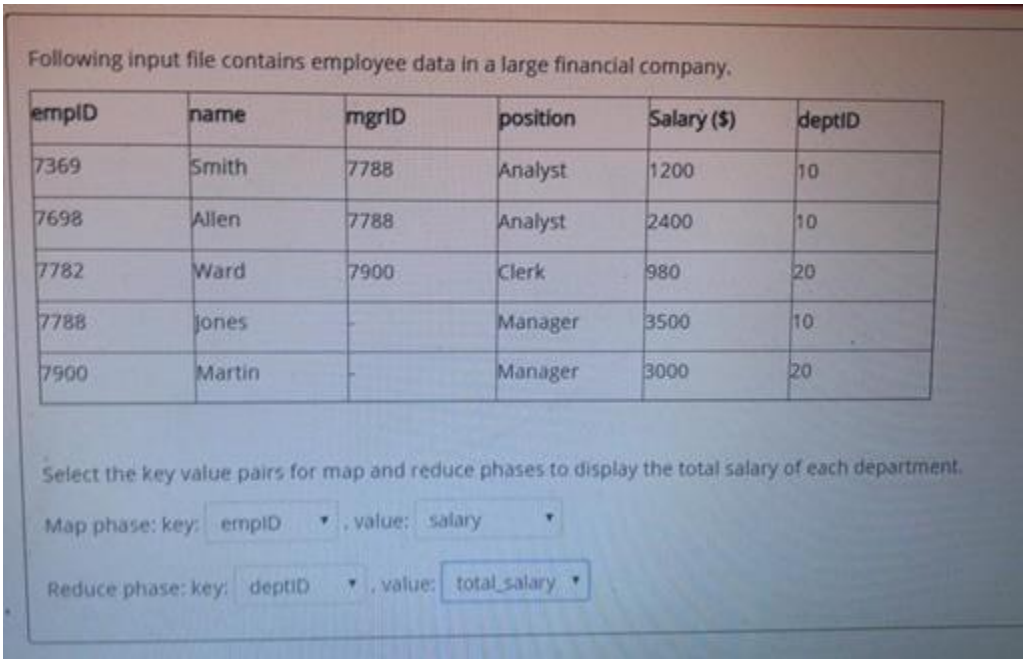
Which of the following are the duties



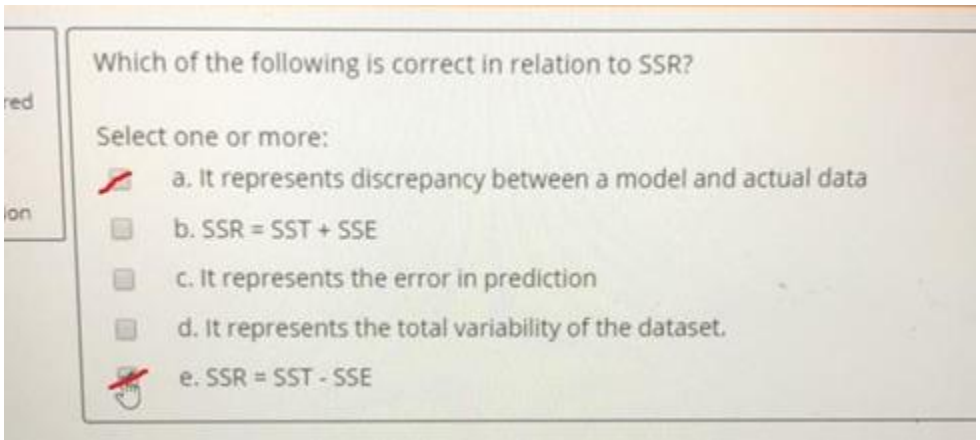
Which of the following are true



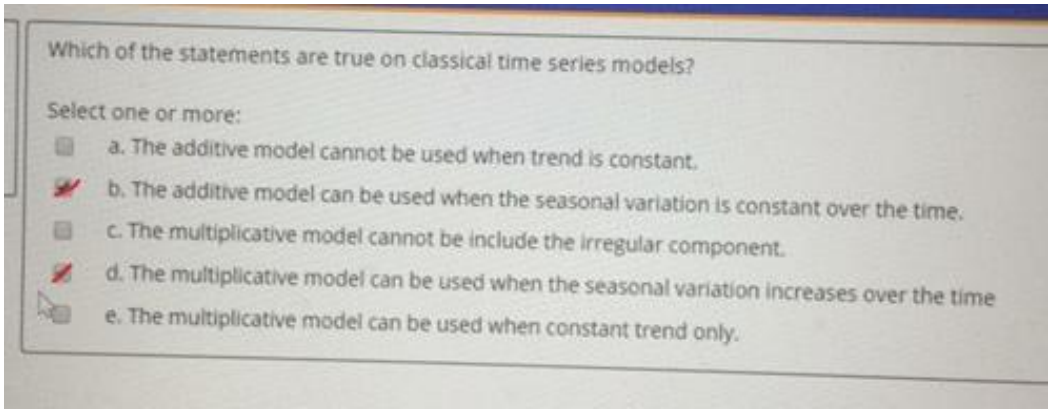
Following input file contains employee data



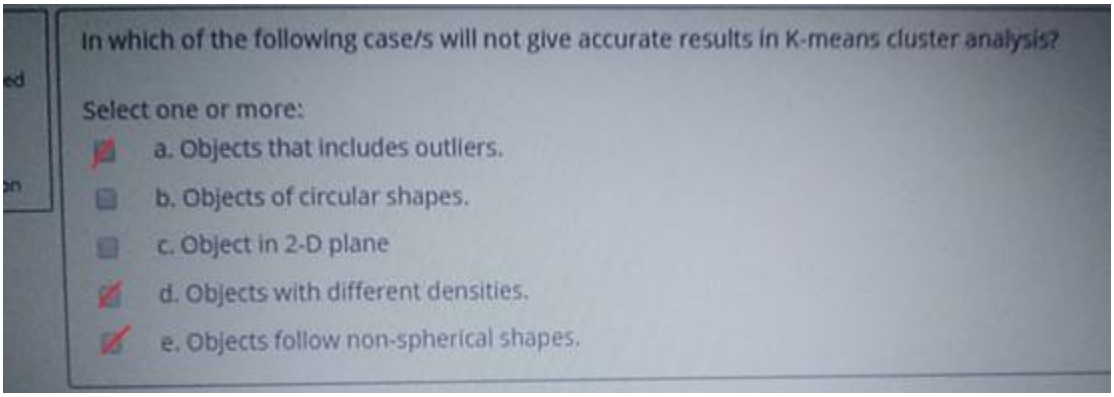
Which of the following is correct



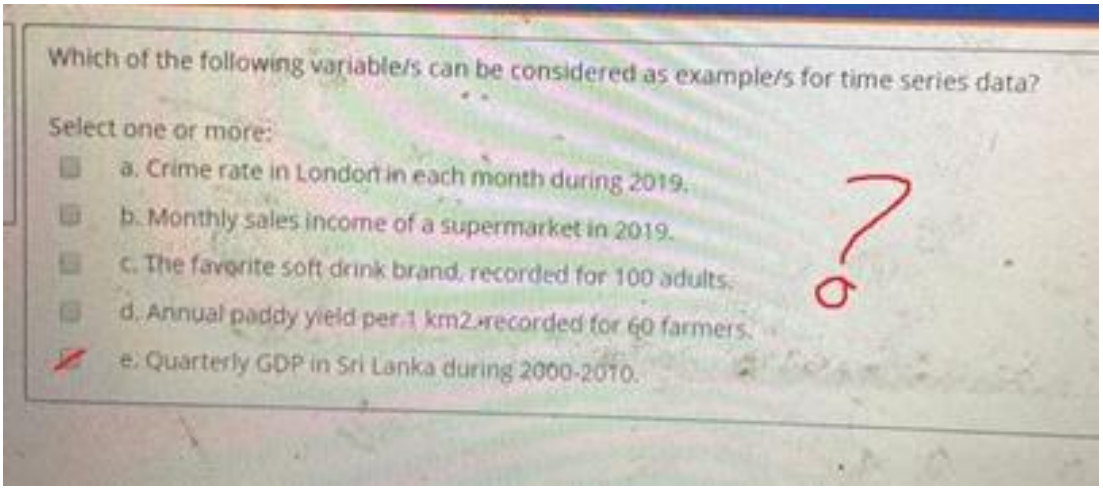
Which of the statement are true on classical time



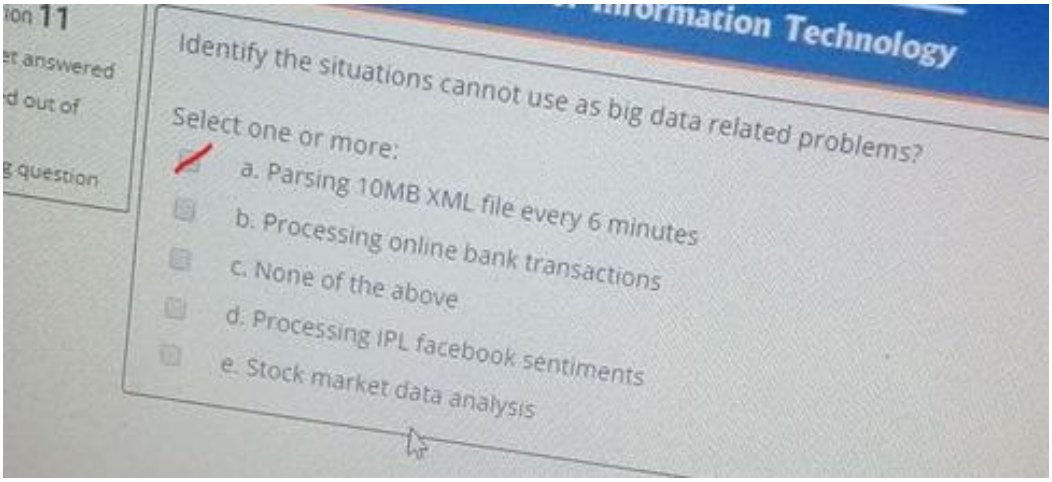
In which of the following case



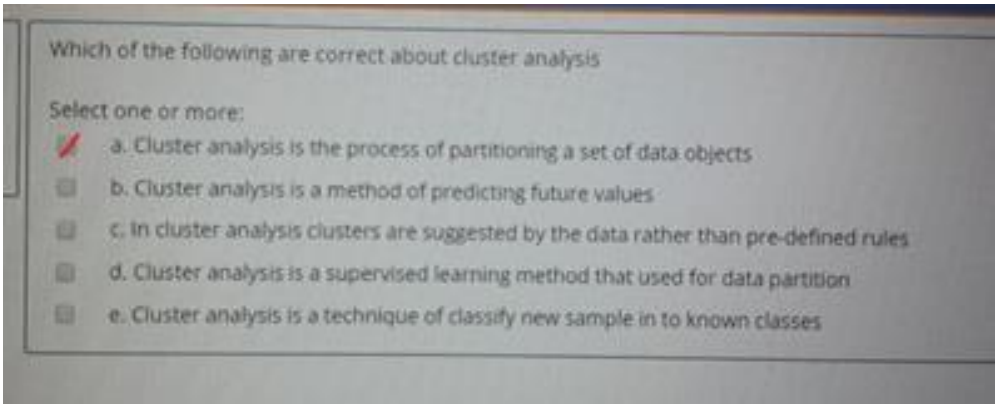
Which of the following variable



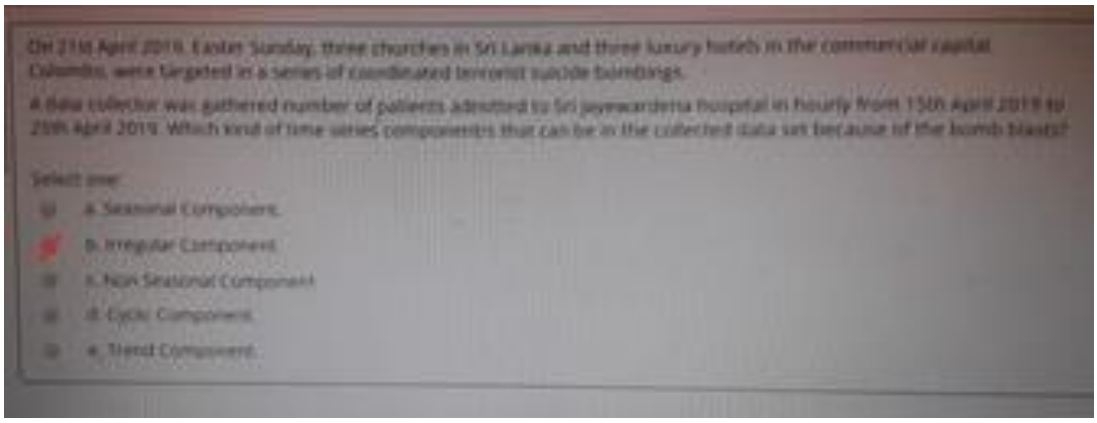
Identify the situations



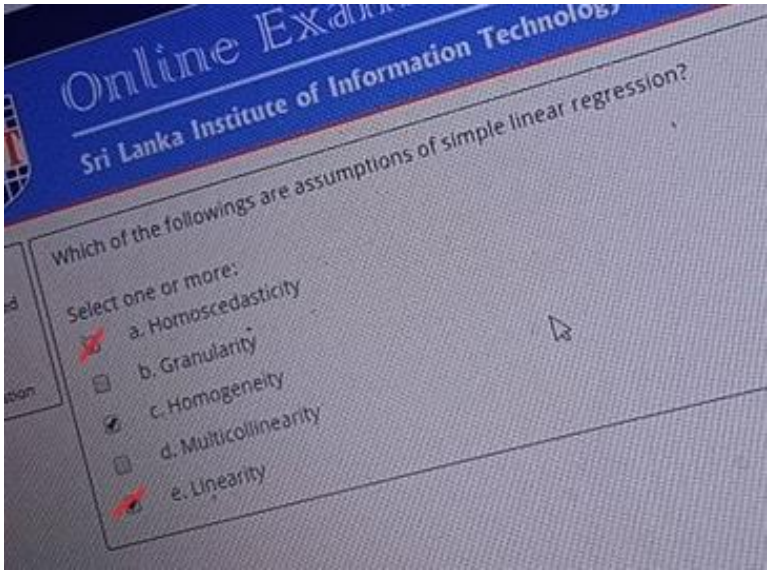
Which of the following are correct



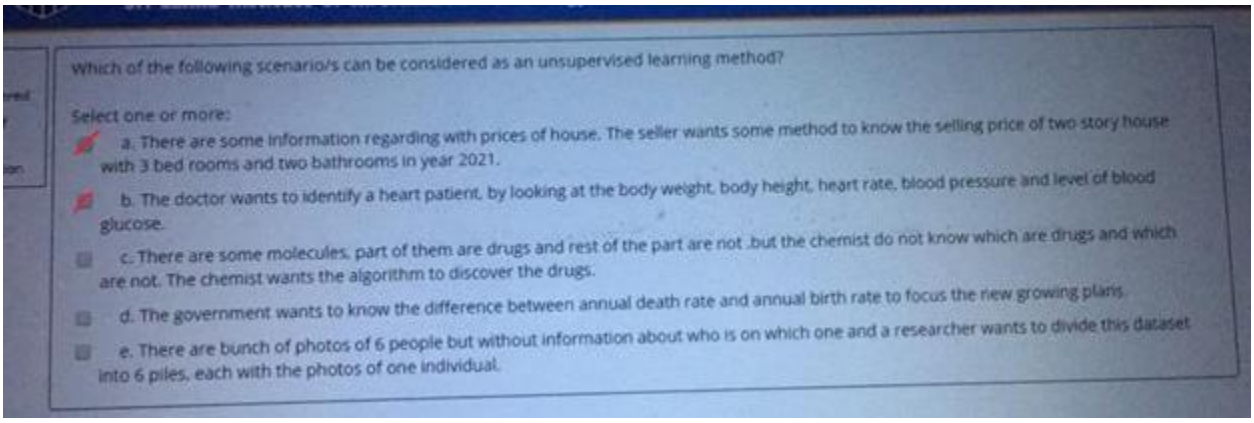
On 21st April 2019



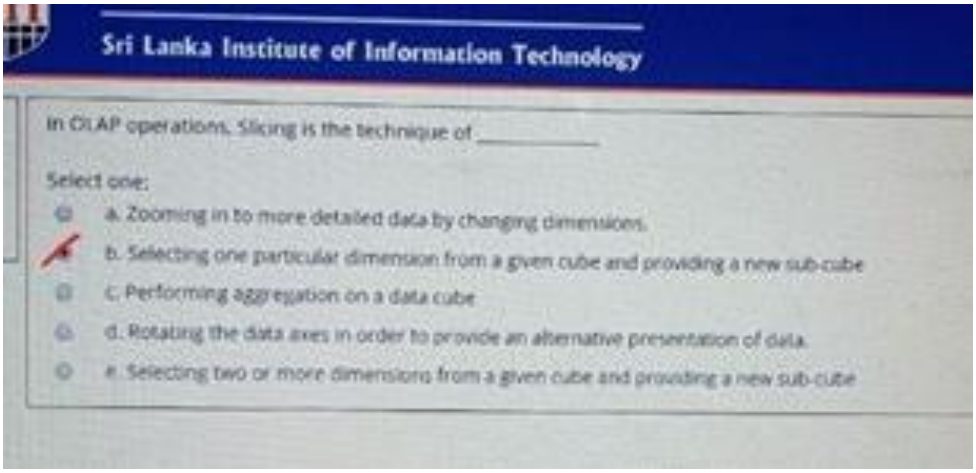
Which of the followings are assumptions



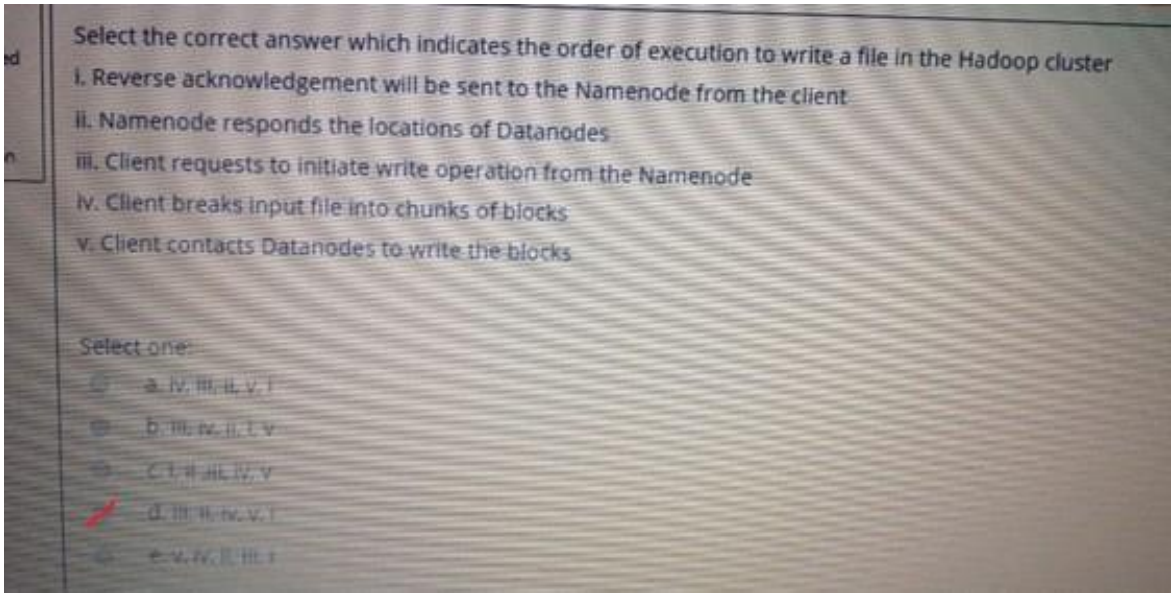
Which of the following scenario/s



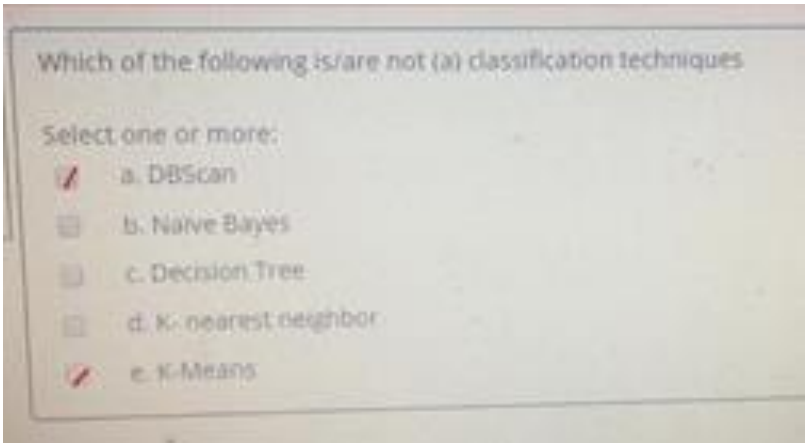
In OLAP operations



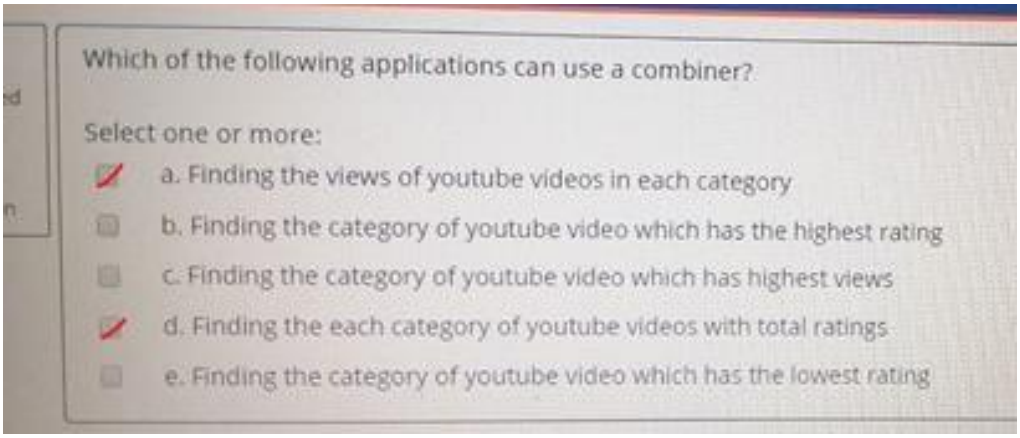
Select the correct answer which indicates



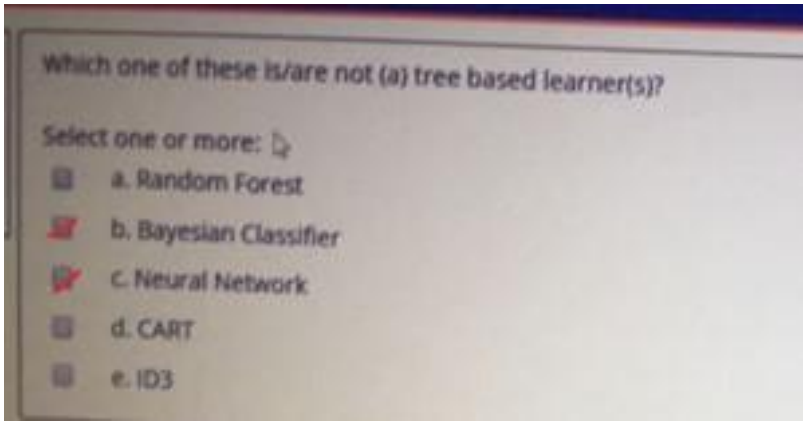
Which of the following is/are not (a)



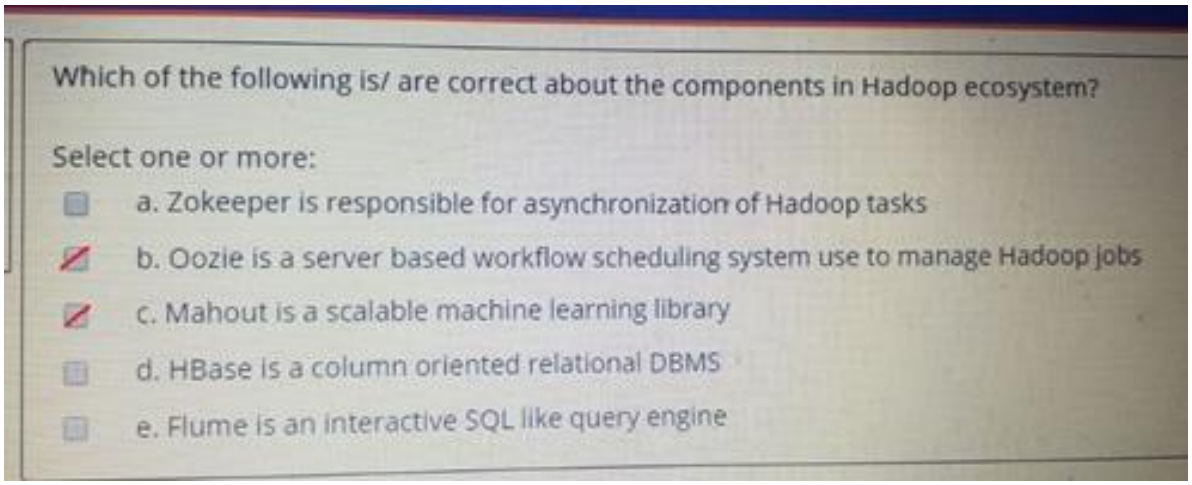
Which of the following applications



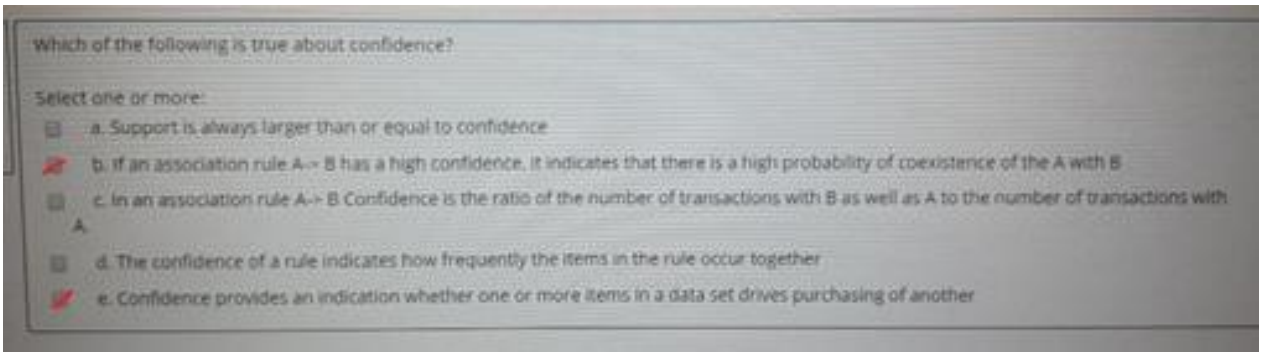
Which one of these is/are not



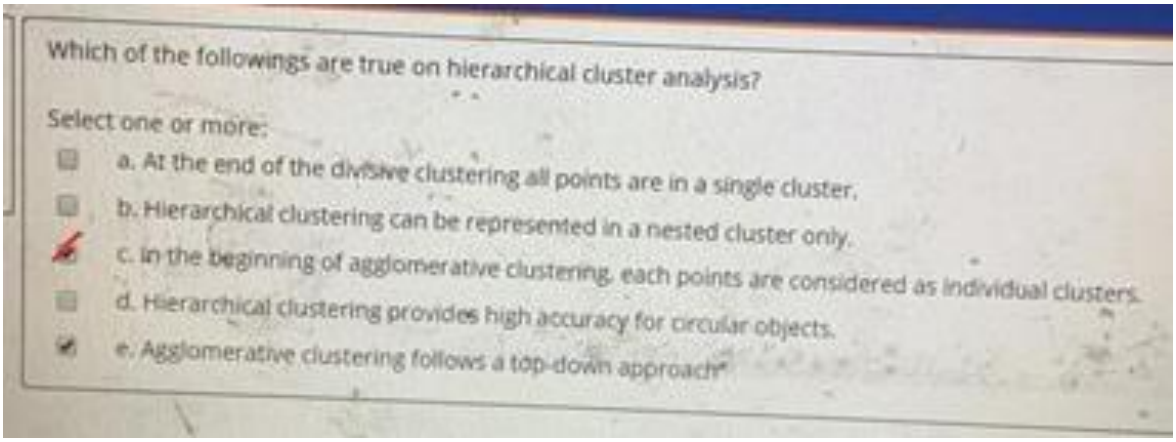
Which of the following is



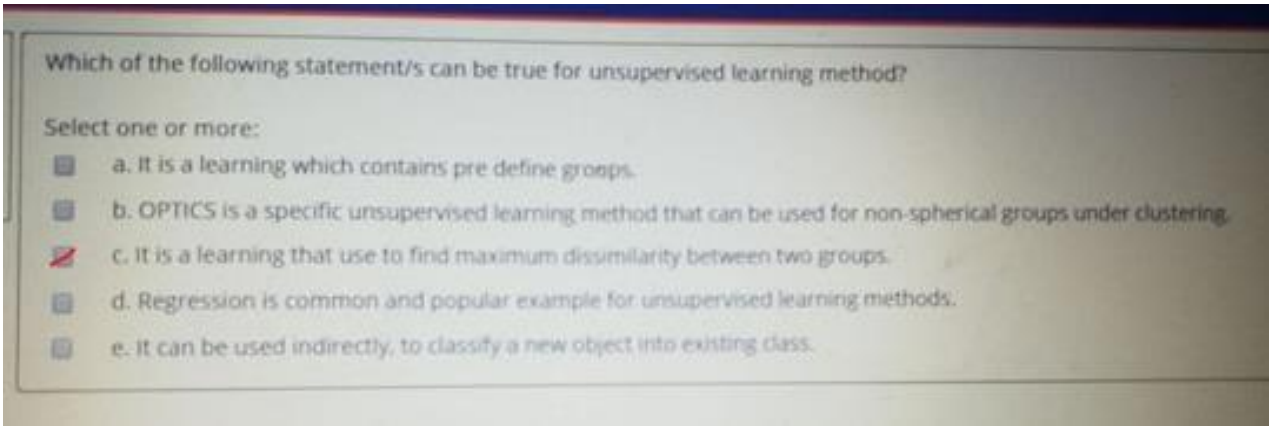
Which of the following is true



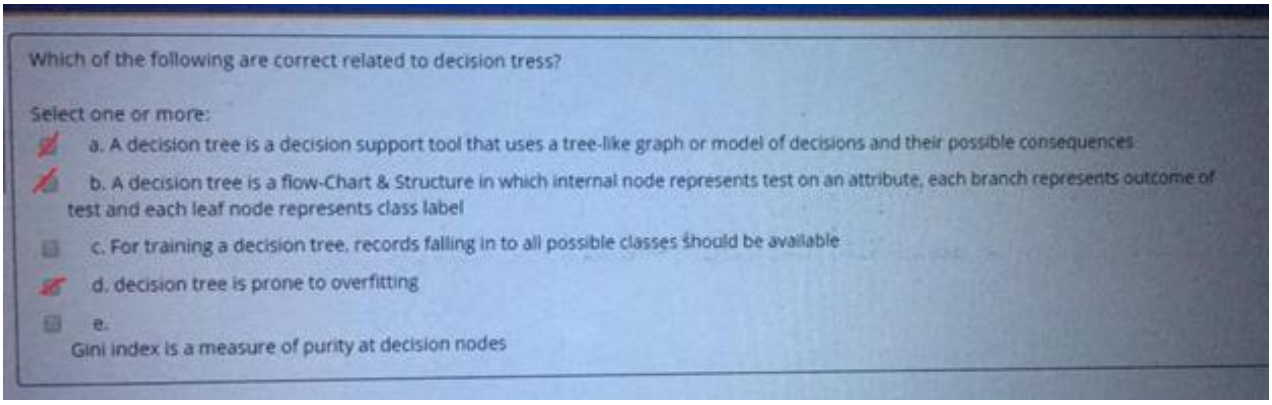
Which of the followings are true on hierarchical



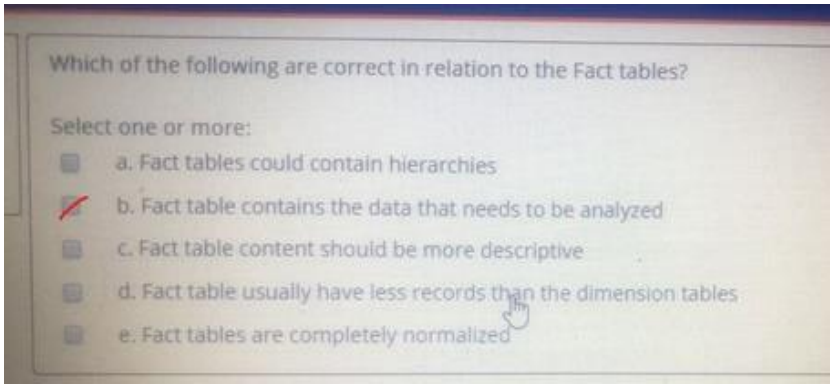
Which of the following statement/s



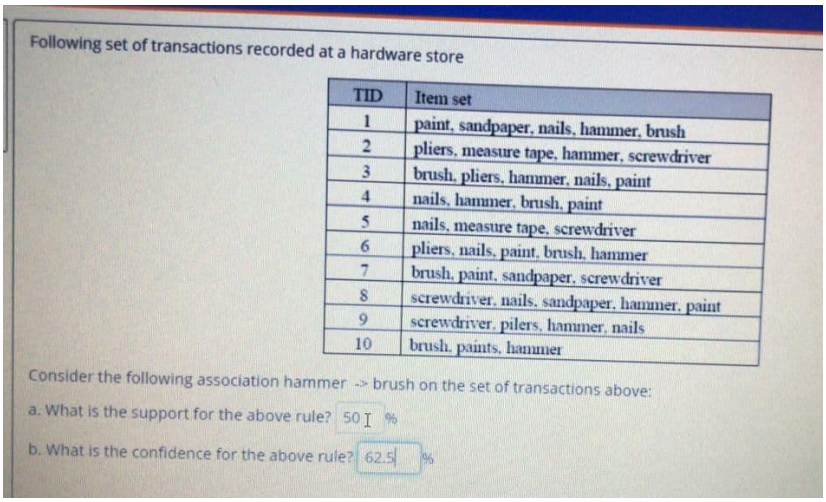
Which of the following are correct related



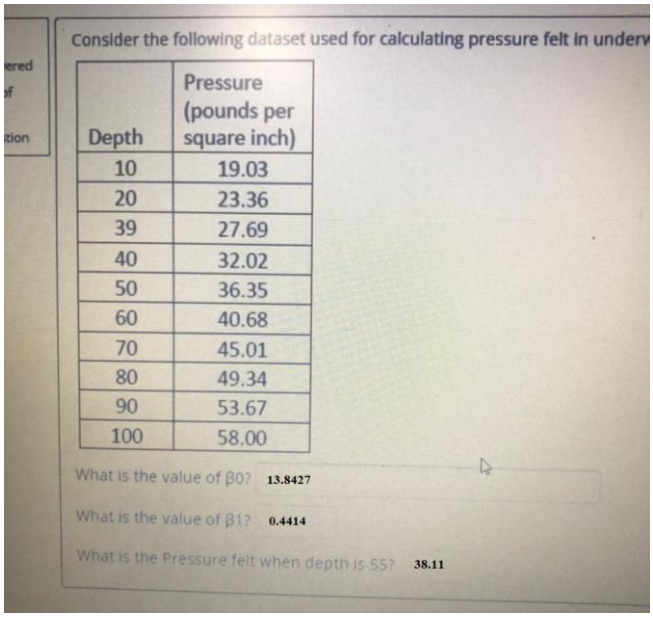
Which of the following are correct in relation to the



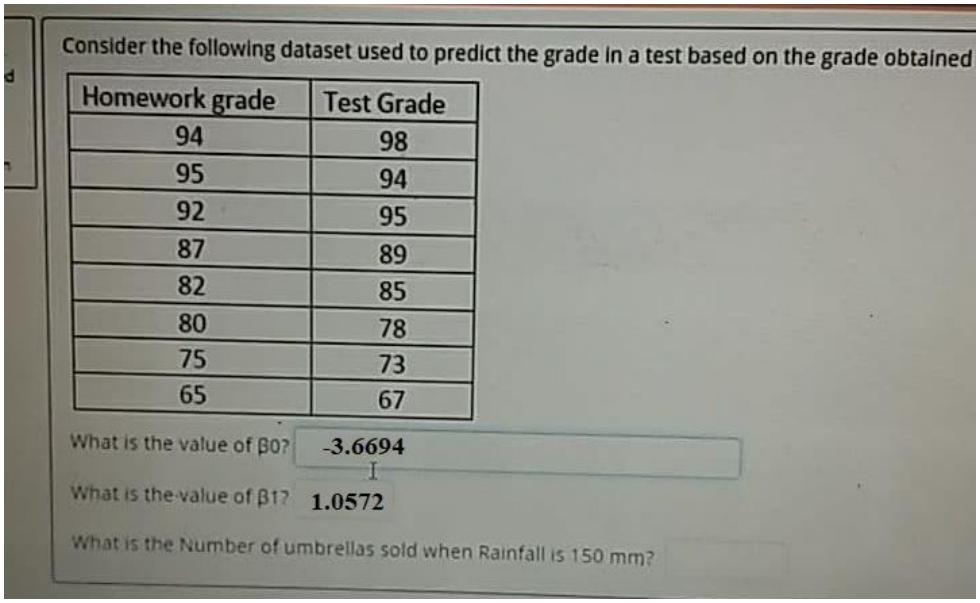
Following set of transactions recorded



Consider the following dataset



Consider the following dataset used



Consider the following dataset used to predict the grade in a test based on the grade obtained

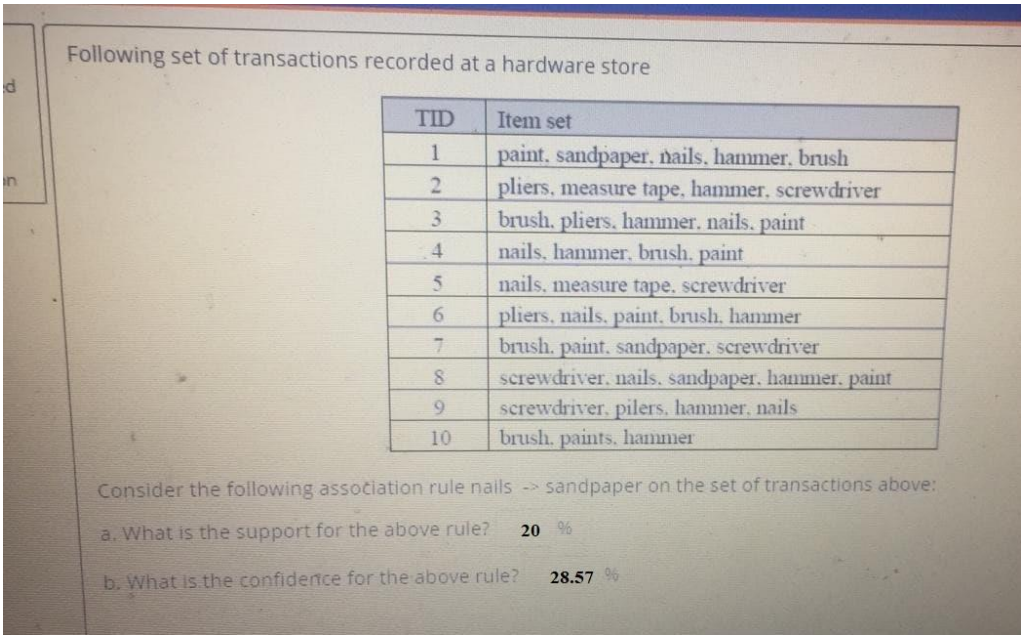
Homework grade	Test Grade
94	98
95	94
92	95
87	89
82	85
80	78
75	73
65	67

What is the value of β0? -3.6694

What is the value of β1? 1.0572

What is the Number of umbrellas sold when Rainfall is 150 mm?

Following set of transactions



Following set of transactions recorded at a hardware store

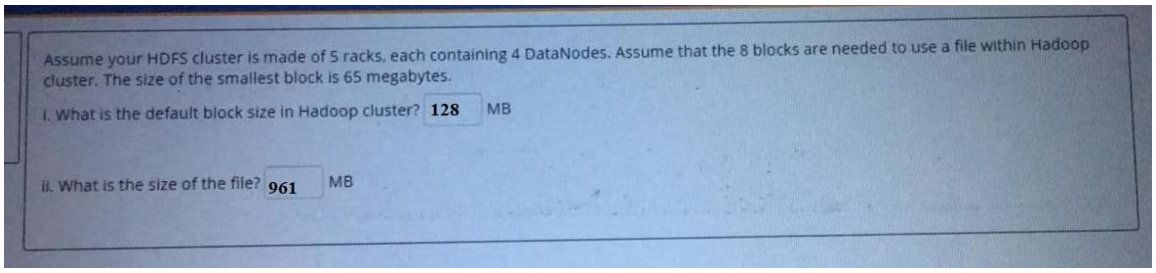
TID	Item set
1	paint, sandpaper, nails, hammer, brush
2	pliers, measure tape, hammer, screwdriver
3	brush, pliers, hammer, nails, paint
4	nails, hammer, brush, paint
5	nails, measure tape, screwdriver
6	pliers, nails, paint, brush, hammer
7	brush, paint, sandpaper, screwdriver
8	screwdriver, nails, sandpaper, hammer, paint
9	screwdriver, pilers, hammer, nails
10	brush, paints, hammer

Consider the following association rule nails --> sandpaper on the set of transactions above:

a. What is the support for the above rule? 20 %

b. What is the confidence for the above rule? 28.57 %

Assume your HDFS cluster

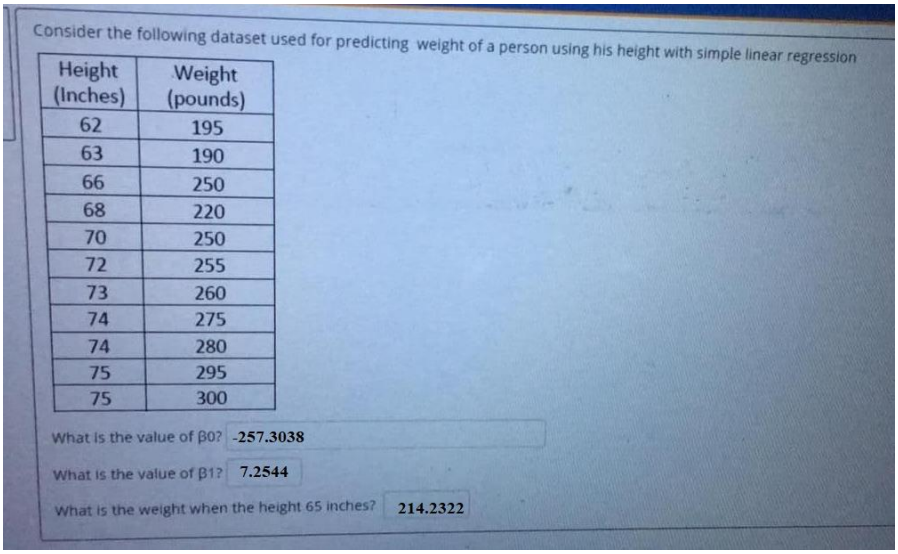


Assume your HDFS cluster is made of 5 racks, each containing 4 DataNodes. Assume that the 8 blocks are needed to use a file within Hadoop cluster. The size of the smallest block is 65 megabytes.

i. What is the default block size in Hadoop cluster? 128 MB

ii. What is the size of the file? 961 MB

Consider the following dataset



Consider the following dataset used for predicting weight of a person using his height with simple linear regression

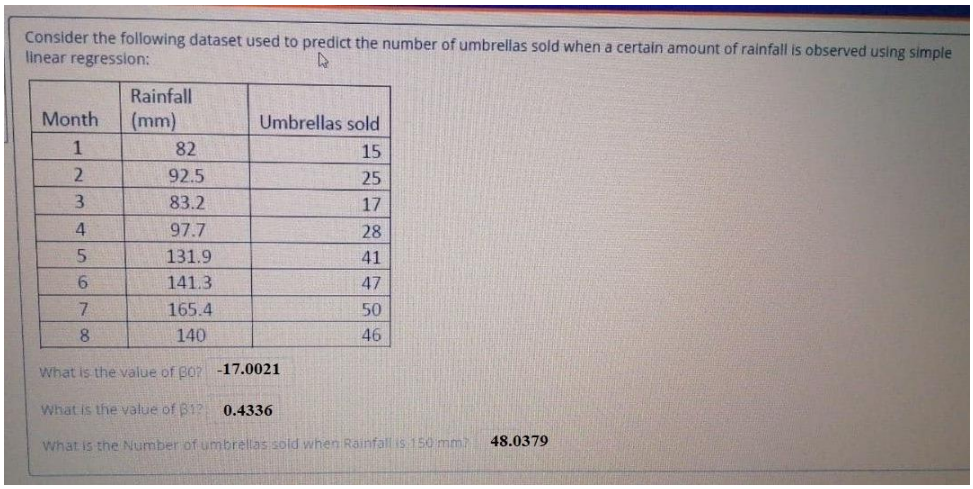
Height (Inches)	Weight (pounds)
62	195
63	190
66	250
68	220
70	250
72	255
73	260
74	275
74	280
75	295
75	300

What is the value of β0? -257.3038

What is the value of β1? 7.2544

What is the weight when the height 65 inches? 214.2322

Consider the following dataset used



Consider the following dataset used to predict the number of umbrellas sold when a certain amount of rainfall is observed using simple linear regression:

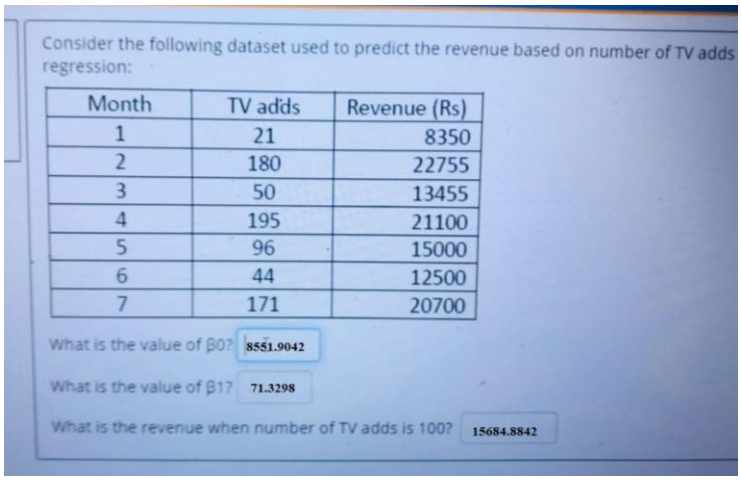
Month	Rainfall (mm)	Umbrellas sold
1	82	15
2	92.5	25
3	83.2	17
4	97.7	28
5	131.9	41
6	141.3	47
7	165.4	50
8	140	46

What is the value of β0? -17.0021

What is the value of β1? 0.4336

What is the Number of umbrellas sold when Rainfall is 150 mm? 48.0379

Consider the following dataset



Consider the following dataset used to predict the revenue based on number of TV adds regression:

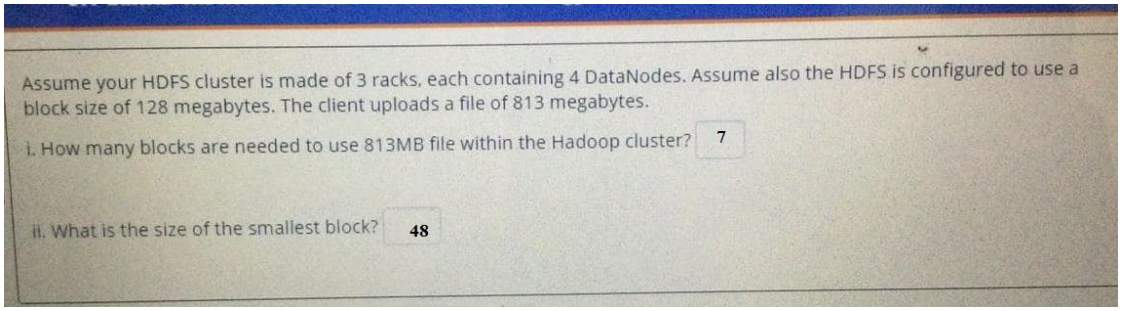
Month	TV adds	Revenue (Rs)
1	21	8350
2	180	22755
3	50	13455
4	195	21100
5	96	15000
6	44	12500
7	171	20700

What is the value of β0? 8551.9042

What is the value of β1? 71.3298

What is the revenue when number of TV adds is 100? 15684.8842

Assume your HDFS cluster is made of 3 racks

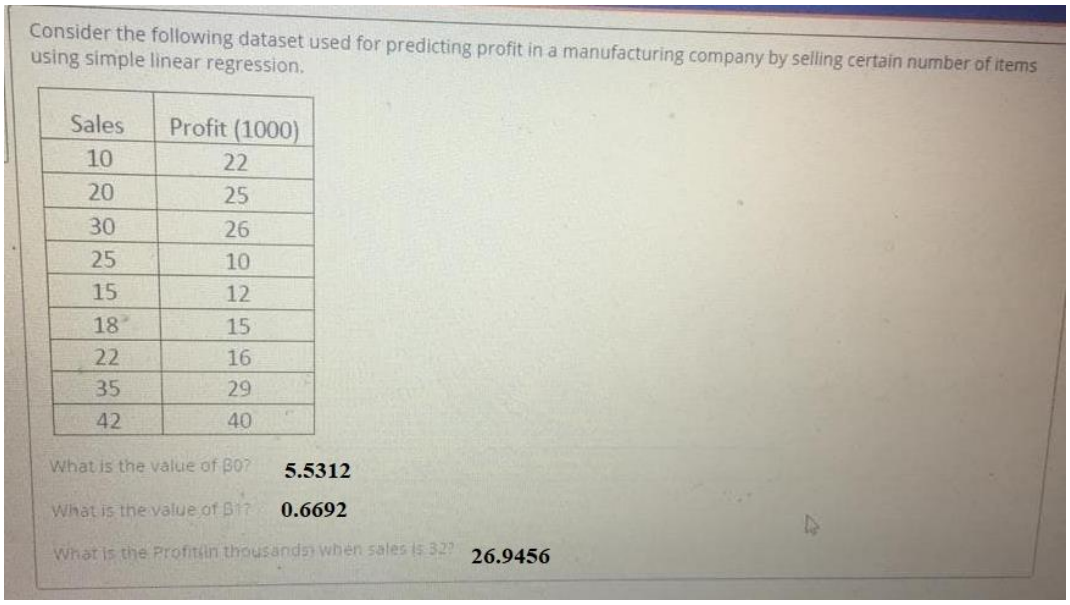


Assume your HDFS cluster is made of 3 racks, each containing 4 DataNodes. Assume also the HDFS is configured to use a block size of 128 megabytes. The client uploads a file of 813 megabytes.

i. How many blocks are needed to use 813MB file within the Hadoop cluster? 7

ii. What is the size of the smallest block? 48

Consider the following dataset



Consider the following dataset used for predicting profit in a manufacturing company by selling certain number of items using simple linear regression.

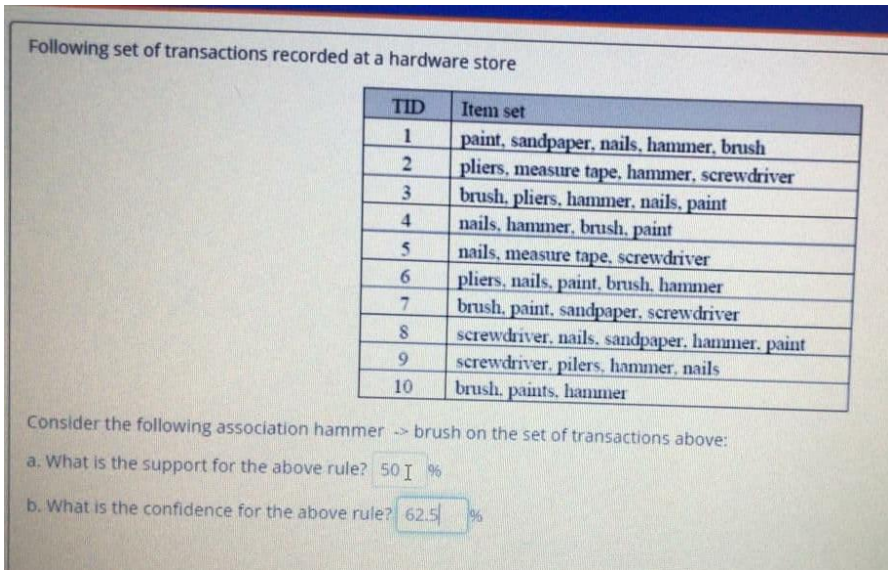
Sales	Profit (1000)
10	22
20	25
30	26
25	10
15	12
18	15
22	16
35	29
42	40

What is the value of β0? 5.5312

What is the value of β1? 0.6692

What is the Profit(in thousands) when sales is 32? 26.9456

Following set of transactions recorded



Following set of transactions recorded at a hardware store

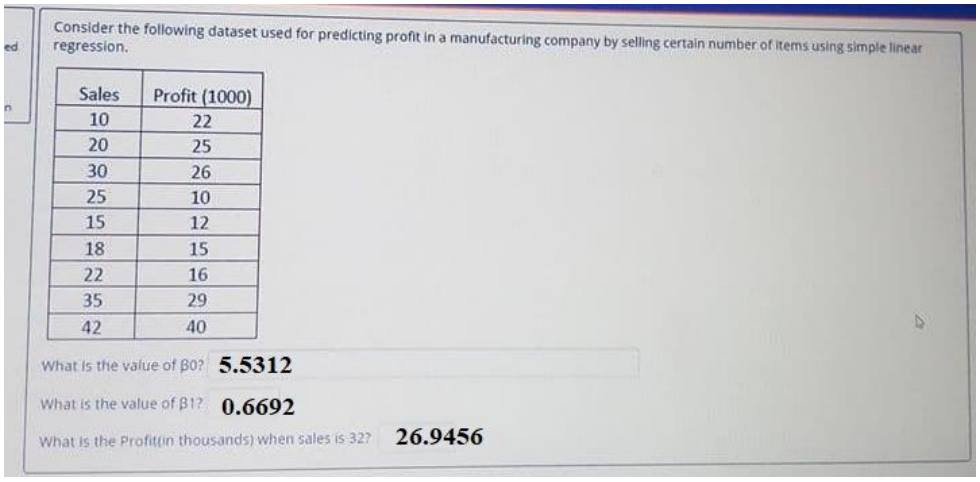
TID	Item set
1	paint, sandpaper, nails, hammer, brush
2	pliers, measure tape, hammer, screwdriver
3	brush, pliers, hammer, nails, paint
4	nails, hammer, brush, paint
5	nails, measure tape, screwdriver
6	pliers, nails, paint, brush, hammer
7	brush, paint, sandpaper, screwdriver
8	screwdriver, nails, sandpaper, hammer, paint
9	screwdriver, pilers, hammer, nails
10	brush, paints, hammer

Consider the following association hammer -> brush on the set of transactions above:

a. What is the support for the above rule? 50 %

b. What is the confidence for the above rule? 62.5 %

Consider the following dataset used



Consider the following dataset used for predicting profit in a manufacturing company by selling certain number of items using simple linear regression.

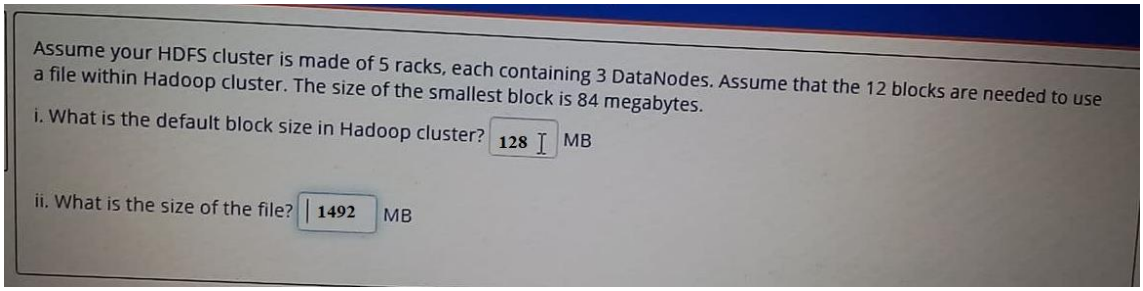
Sales	Profit (1000)
10	22
20	25
30	26
25	10
15	12
18	15
22	16
35	29
42	40

What is the value of β0? 5.5312

What is the value of β1? 0.6692

What is the Profit(in thousands) when sales is 32? 26.9456

Assume your HDFS cluster

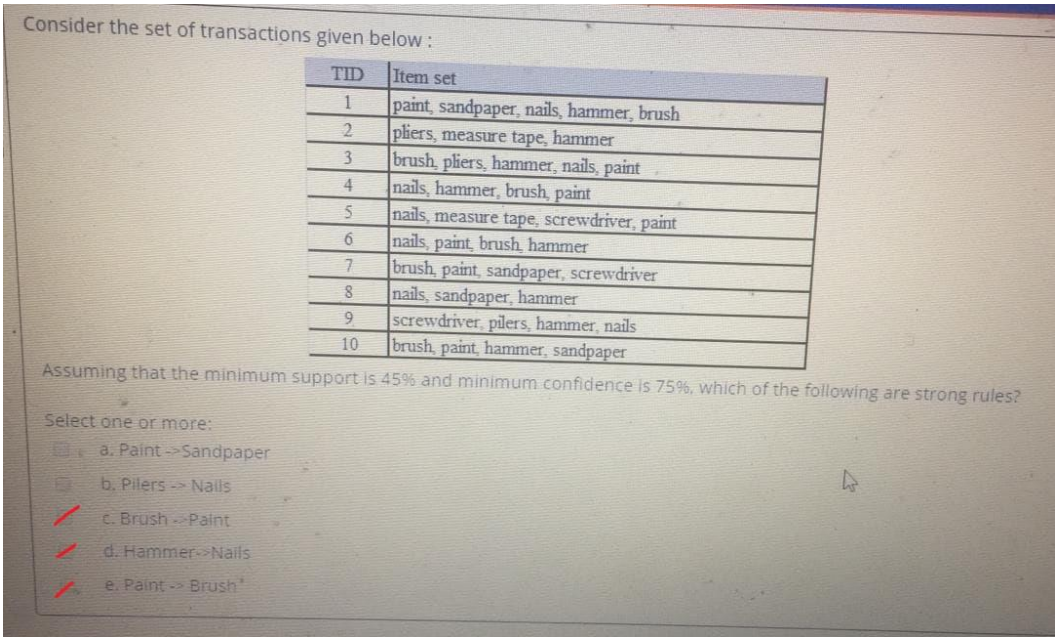


Assume your HDFS cluster is made of 5 racks, each containing 3 DataNodes. Assume that the 12 blocks are needed to use a file within Hadoop cluster. The size of the smallest block is 84 megabytes.

i. What is the default block size in Hadoop cluster? 128 MB

ii. What is the size of the file? 1492 MB

Consider the set of transactions



Consider the set of transactions given below :

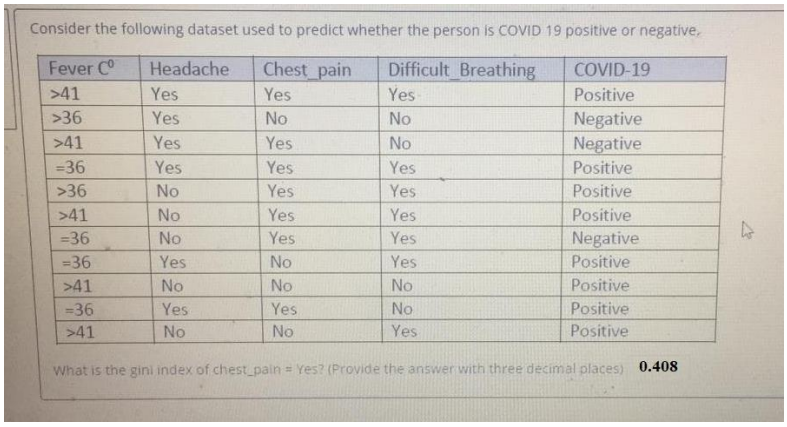
TID	Item set
1	paint, sandpaper, nails, hammer, brush
2	pliers, measure tape, hammer
3	brush, pliers, hammer, nails, paint
4	nails, hammer, brush, paint
5	nails, measure tape, screwdriver, paint
6	nails, paint, brush, hammer
7	brush, paint, sandpaper, screwdriver
8	nails, sandpaper, hammer
9	screwdriver, pilers, hammer, nails
10	brush, paint, hammer, sandpaper

Assuming that the minimum support is 45% and minimum confidence is 75%, which of the following are strong rules?

Select one or more:

- a. Paint ->Sandpaper
- b. Pilers -> Nails
- c. Brush ->Paint
- d. Hammer->Nails
- e. Paint -> Brush

Consider the following dataset used

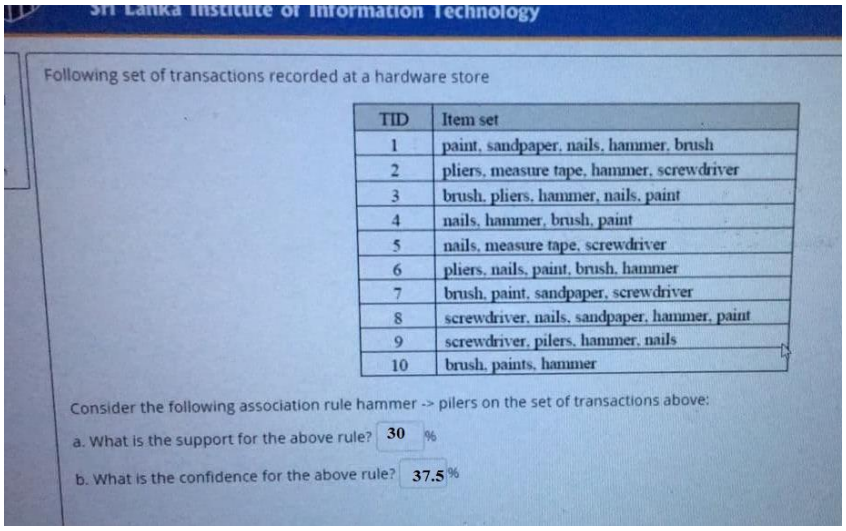


Consider the following dataset used to predict whether the person is COVID 19 positive or negative.

Fever C⁰	Headache	Chest_pain	Difficult_Breathing	COVID-19
>41	Yes	Yes	Yes	Positive
>36	Yes	No	No	Negative
>41	Yes	Yes	No	Negative
=36	Yes	Yes	Yes	Positive
>36	No	Yes	Yes	Positive
>41	No	Yes	Yes	Positive
=36	No	Yes	Yes	Negative
=36	Yes	No	Yes	Positive
>41	No	No	No	Positive
=36	Yes	Yes	No	Positive
>41	No	No	Yes	Positive

What is the gini index of chest_pain = Yes? (Provide the answer with three decimal places) 0.408

Following set of transactions



Sri Lanka Institute of Information Technology

Following set of transactions recorded at a hardware store

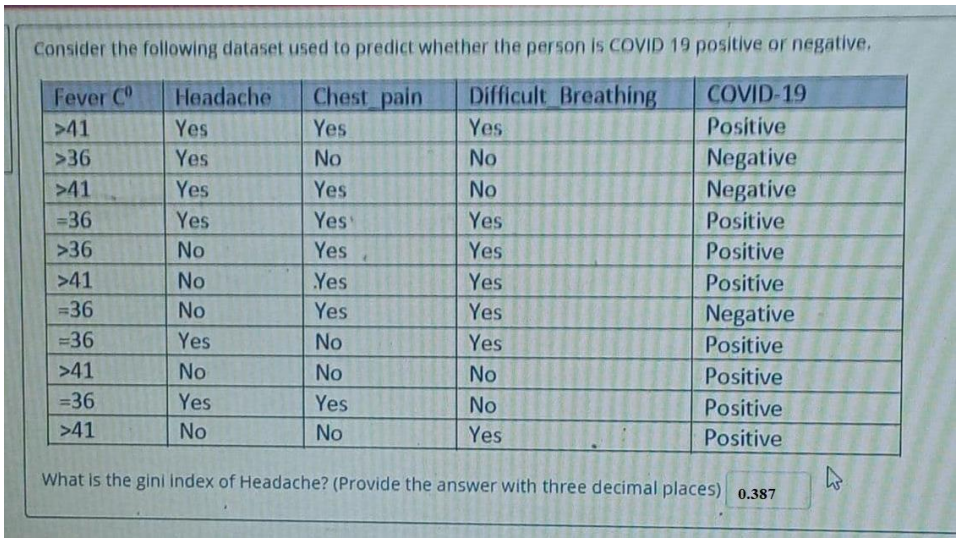
TID	Item set
1	paint, sandpaper, nails, hammer, brush
2	pliers, measure tape, hammer, screwdriver
3	brush, pliers, hammer, nails, paint
4	nails, hammer, brush, paint
5	nails, measure tape, screwdriver
6	pliers, nails, paint, brush, hammer
7	brush, paint, sandpaper, screwdriver
8	screwdriver, nails, sandpaper, hammer, paint
9	screwdriver, pilers, hammer, nails
10	brush, paints, hammer

Consider the following association rule hammer -> pilers on the set of transactions above:

a. What is the support for the above rule? 30 %

b. What is the confidence for the above rule? 37.5 %

Consider the following dataset used of predict



Consider the following dataset used to predict whether the person is COVID 19 positive or negative.

Fever C⁰	Headache	Chest_pain	Difficult_Breathing	COVID-19
>41	Yes	Yes	Yes	Positive
>36	Yes	No	No	Negative
>41	Yes	Yes	No	Negative
=36	Yes	Yes	Yes	Positive
>36	No	Yes	Yes	Positive
>41	No	Yes	Yes	Positive
=36	No	Yes	Yes	Negative
=36	Yes	No	Yes	Positive
>41	No	No	No	Positive
=36	Yes	Yes	No	Positive
>41	No	No	Yes	Positive

What is the gini index of Headache? (Provide the answer with three decimal places) 0.387