


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

Consider the following output obtained by using multiple regression technique in python in order to predict the Employed people in a country as a percentage based on the Gross National product (GNP). Assuming alpha is 0.05 what would be the employed people as a percentage in a country with GNP is 258.

	coef	std err	t	P> t	[95.0% Conf. Int.]
const	51.8436	0.681	76.087	0.000	50.382 53.305
GNP	0.0348	0.002	20.374	0.000	0.031 0.038

Select one:

- ☐ a. 60%
- ☐ b. 76 %
- ☐ c. 51 %
- ☐ d. 80 %
- ☐ e. 0.03 %

The relationship between price and sales

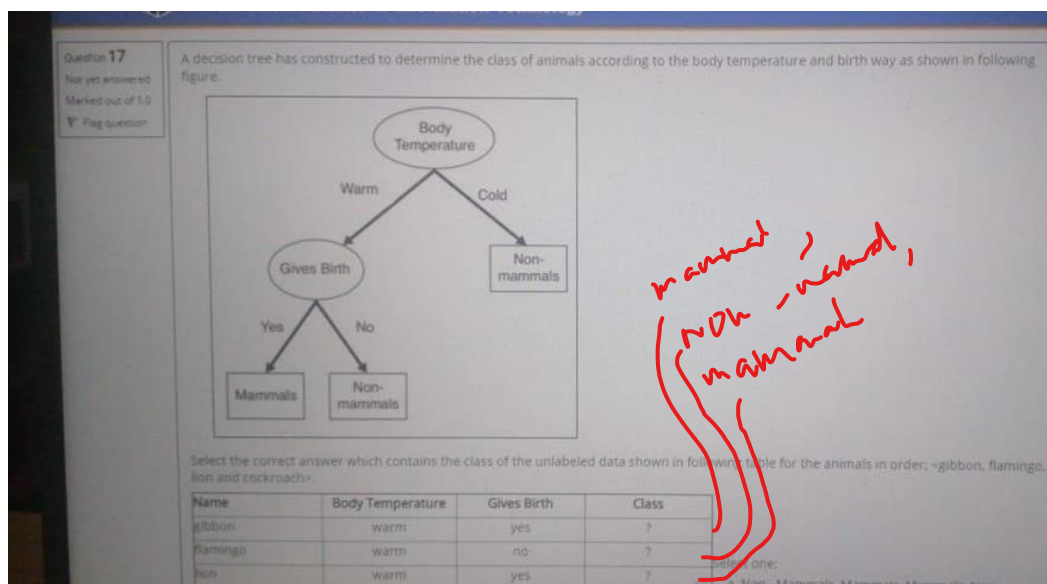
The relationship between price and sales of an item is analyzed using simple linear regression with the following data. What would be the value for β_0 in the resulted regression equation?

Price (k)	Sales (k)
3	20
6	9
11	0

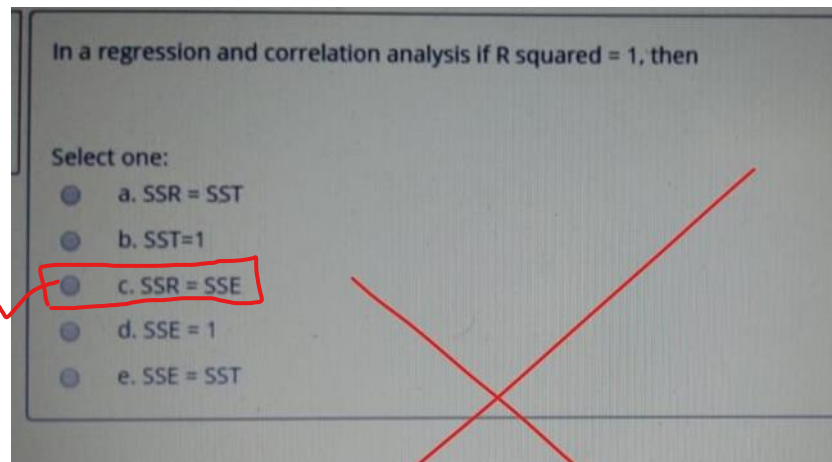
Select one:

- ☐ a. 9.6
- ☒ b. 25
- ☐ c. 1.06
- ☐ d. 2.5
- ☐ e. -2.4

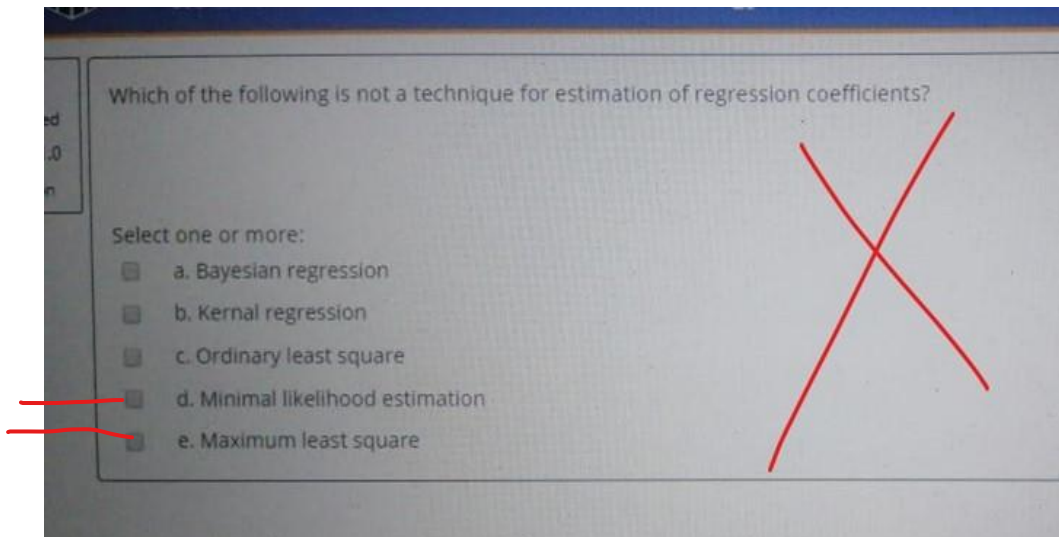
A decision tree has constructed to determine



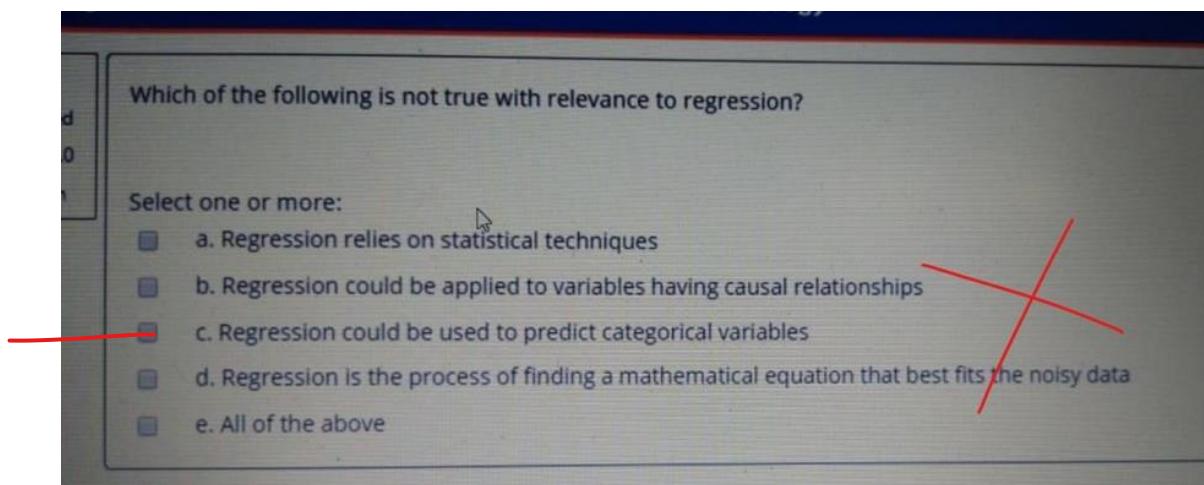
In a regression and correlation analysis



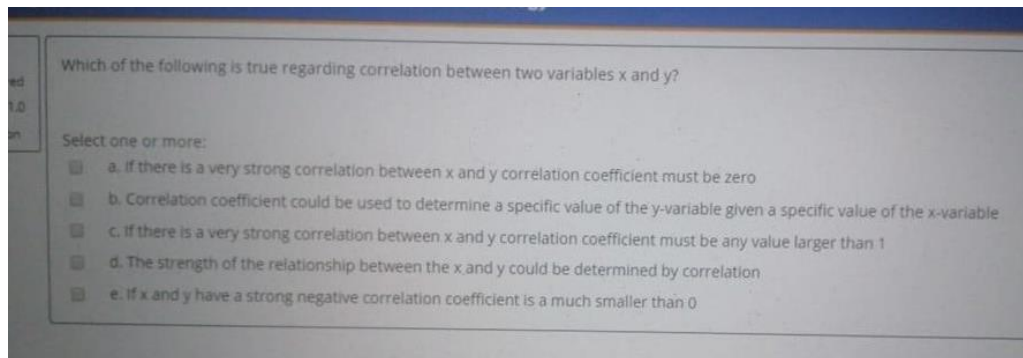
Which of the following is not a technique



Which of the following is true with relevance



Which of the following is true regarding correlation



Consider the following data frame

SRM Lanka Institute of Information Technology

Consider the following data frame named data.

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443998.53	California	191792.06
2	153441.51	101145.55	407934.54	California	191050.39
3	144372.41	110671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	California	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12
6	134615.46	147190.87	127716.82	California	156122.51
7	130298.13	145530.96	323876.68	New York	155752.60
8	120542.52	148718.95	311613.29	New York	152211.77
9	123334.88	108679.17	304981.62	California	149759.96

which of the following is true?

Select one or more:

- ☐ a. `data.iloc[:, 2: 4]` returns the content of Administration and Marketing Spend columns
- ☐ b. `data[Administration].describe()` generates descriptive statistics for the data in column Administration.
- ☐ c. `data[:, 4]` command provide the same output as `data[Profit]`
- ☐ d. `data.shape[1]` returns 4
- ☐ e. `data.head()` returns all column values for the first five rows in the data frame.

Consider the following distance matrix.

SRM Lanka Institute of Information Technology

Consider the following distance matrix. Assuming that C and D are initial centroids for K-means algorithm, what would be the clusters resulted after the first epoch?

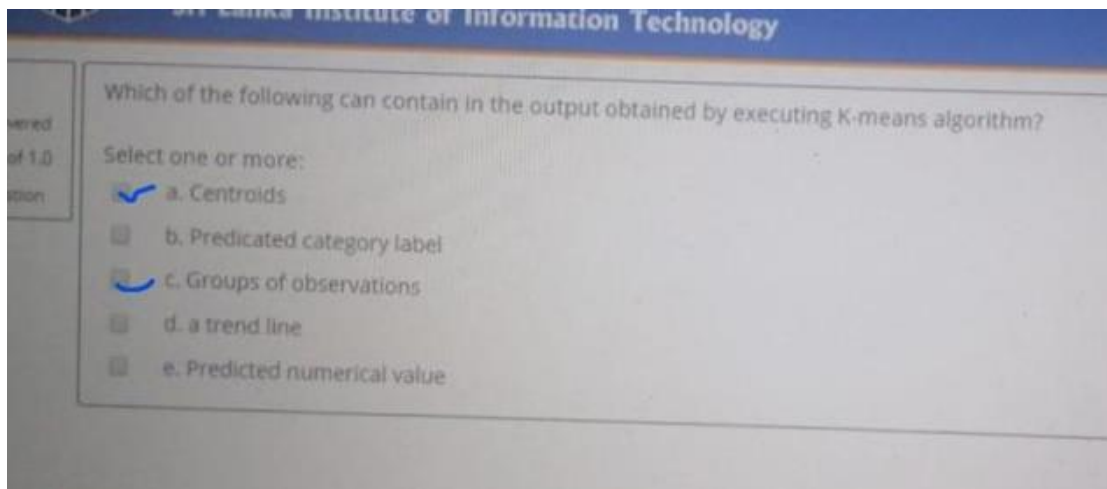
	A	B	C	D	E	F	G
A	0						
B	45	0					
C	13	6	0				
D	21	21	9	0			
E	32	6	40	1	0		
F	15	16	9	28	11	0	
G	19	25	32	9	5	10	0

Select one:

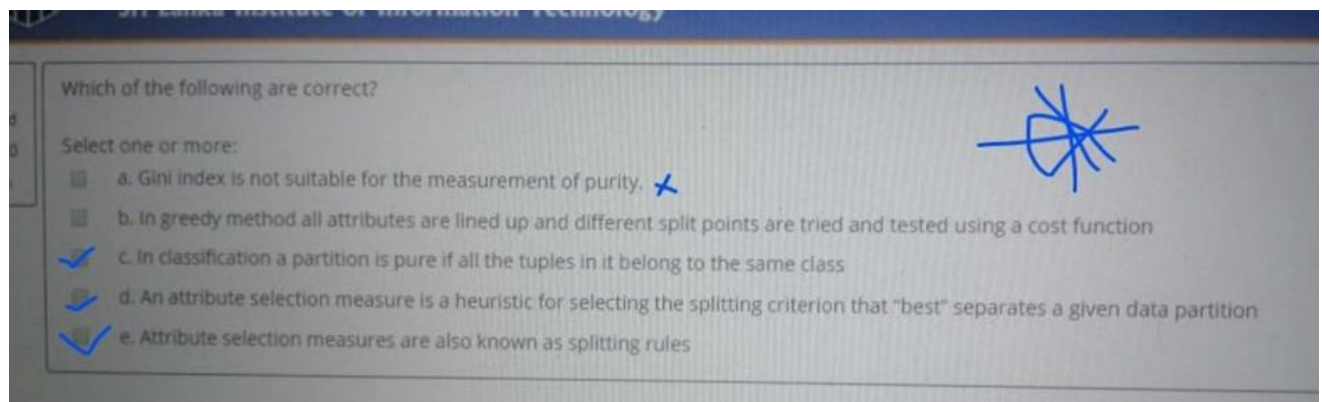
- ☐ a. A, D, C and B, E, F, G
- ☐ b. B, C, G and A, D, E, F
- ☐ c. A, B, C, E and D, F, G
- ☐ d. B, C, F, G and A, D, E
- ☒ e. A, B, C, F and D, E, G

Handwritten notes: Cluster 1 (C): A, B, F. Cluster 2 (D): D, E, G.

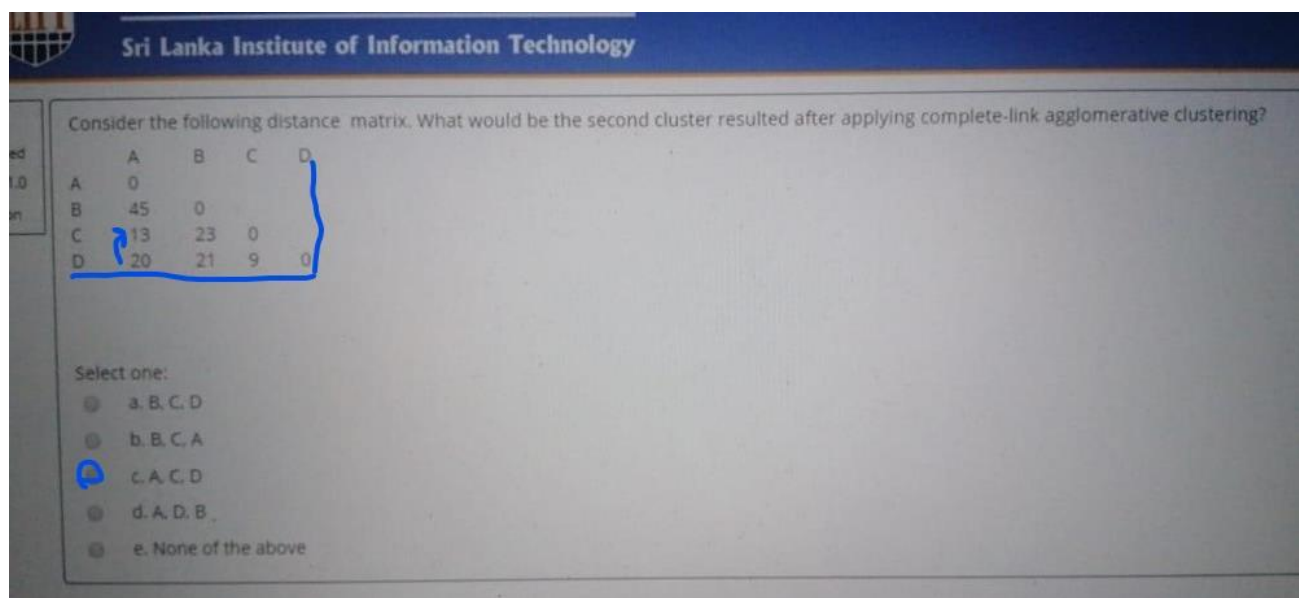
Which of the following can contain in the output obtained



Which of the following are correct



Consider the following distance matrix.



Which of the following is not an application of clustering

Sri Lanka Institute of Information Technology

Which of the following is not an application of clustering?

Select one or more:

- ☒ a. Identifying different user groups in Facebook ✓
- ☐ b. Identifying most suitable airline to travel for a customer with a given profile
- ☐ c. Detecting spam emails
- ☐ d. Predicting price of houses
- ☒ e. Market segmentation ✓

Select the most suitable answers to fill the blanks

Sri Lanka Institute of Information Technology

Select the most suitable answers to fill the blanks in the correct order.

The clustering, follows a approach, which starts with all the objects in the same cluster whereas the clustering, follows a approach which starts with each object forming a separate group

Select one:

- ☐ a. divisive, agglomerative, top-down, bottom-up
- ☐ b. agglomerative, top-down, divisive, bottom-up
- ☒ c. divisive, top-down, agglomerative, bottom-up ✓
- ☐ d. top-down, divisive, bottom-up, agglomerative
- ☐ e. None of the above

Consider the following data set

Consider the following data set

TID	Items Bought
T001	{Milk, Orange, Bread, Eggs, Youghurt, Butter }
T002	{Noodles, Orange, Bread, Milk, Eggs, Youghurt}
T003	{Milk, Sugar, Bread, Egg}
T004	{Chocolate, Sugar, Tea bags, Butter, Youghurt, Biscuits}
T005	{Biscuits, Orange, Bread, Ice cream, Eggs }

What is the confidence for {Milk} → Bread?

Select one:

- ☐ a. 0.5
- ☐ b. 0.25
- ☒ c. 1.0
- ☐ d. 0.2
- ☐ e. 0.75

Handwritten calculation: 2/2 = 1

Consider the following distance matrix.

Sri Lanka Institute of Information Technology

Consider the following distance matrix. What would be the second cluster resulted after applying single-link agglomerative clustering?

	A	B	C	D
A	0	7	1	16
B	7	0	4	3
C	1	4	0	5
D	16	3	5	0

Select one:

- ☐ a. B, C, D
- ☒ b. B, C, A
- ☐ c. A, C, D
- ☐ d. A, D, B
- ☐ e. None of the above

An emergency room in a hospital measures 17

An emergency room in a hospital measures 17 variables to measure the medical condition of newly admitted patients to decide whether the patient should put in an intensive- care unit. Due to the high cost of ICU, those patients who may survive more a month are given higher priority. However, the problem is to predict high risk patients and discriminate them from low- risk patients.

Select the most suitable analytical method for the above scenario.

Select one:

- ☐ a. association rule mining
- ☐ b. clustering
- ☐ c. linear regression analysis
- ☒ d. classification
- ☐ e. time series analysis

IT → c
DS → d

Select the main characteristic of the Supervised learning

Select the main characteristic of the Supervised learning

Select one:

- ☐ a. computationally complex
- ☐ b. Analysts are not aware about different categories of data
- ☐ c. Algorithms are used against data which is not labelled
- ☒ d. Algorithms are trained using labeled data.
- ☐ e. None of the above

A credit card company typically

Sri Lanka Institute of Information Technology

A credit card company typically receives hundreds of thousands of applications for new cards with several different attributes of applicant details. The company needs to identify the applicants who can be placed into good credit, bad credit or fall into a gray area. Select the most suitable analytical technique for the above scenario.

Select one:

- ☐ a. K-Means analysis
- ☒ b. decision tree analysis ✓
- ☐ c. linear regression analysis
- ☐ d. Time series analysis
- ☐ e. Apriori algorithm

Which of the following is not true

Sri Lanka Institute of Information Technology

Which of the following is not true about classification?

Select one or more:

- ☒ a. The only input for the classifier are tuples ✓
- ☒ b. Second step of classification involves testing the classification model and predicting ✓
- ☒ c. Classification have two main steps namely learning and predicting ✓
- ☒ d. Same set of rows are usually used in first and second phase of classification
- ☒ e. The class label for of the data set is a continuous variable

Consider the following data set

Sri Lanka Institute of Information Technology

Consider the following data set. Assuming that A, B and C are initial centroids for K-means algorithm, what would be the resulted clusters after the first epoch?

	A	B	C	D	E	F	G
A	0						
B	9	0					
C	1	0	0				
D	16	4	8	0			
E	3	12	0	20	0		
F	11	26	13	12	1	0	
G	19	25	23	2	8	10	0

Select one:

- ☐ a. {A, D} {B, E} and {C, F, G}
- ☐ b. {A, E} {B, D, F} {C, G}
- ☒ c. {A, F, G}, {B, D} and {C, E}
- ☐ d. {A, F, G} {B, D} and {C, E}
- ☐ e. {A, D, E} {B, F} {C, G}

Handwritten red text:
A B C
G D E
F

Select the correct answer which contains the class

Select the correct answer which contains the class of the unlabeled data shown in following table for the animals in order: <gibbon, flamingo, lion and cockroach>.

Name	Body Temperature	Gives Birth	Class
gibbon	warm	yes	?
flamingo	warm	no	?
lion	warm	yes	?
cockroach	cold	no	?

Select one:

☐ a. Non - Mammals, Mammals, Mammals, Non-mammals

☒ b. Mammals, Non-mammals, Non - Mammals, Non-mammals

☐ c. Mammals, Non-mammals, Mammals, Non-mammals

☐ d. Non - Mammals, Non- Mammals, Mammals, Non-mammals

☐ e. Mammals, Mammals, Mammals, Non-mammals

Consider the following distance matrix

Consider the following distance matrix. What would be the second cluster resulted after applying complete-link agglomerative clustering?

	A	B	C	D
A	0			
B	45	0		
C	13	23	0	
D	20	21	9	0

Select one:

☐ a. B, C, D

☐ b. B, C, A

☒ c. A, C, D

☐ d. A, D, B

☐ e. None of the above

Consider the following distance matrix developed

Consider the following count matrix developed for the prediction of students' grade using specialty.

Specialty	Grade	
	Yes	No
IT	4	1
Medicine	1	3
Engineering	0	2
Sociology	1	1

Which of the following is Gini(Specialty)?

Select one:

☐ a. 0.0

☒ b. 0.315

☐ c. 0.500

☐ d. 0.037

☐ e. 0.375

Which of the following are imitations

Which of the following are imitations of the K-means algorithm

Select one or more:

- ☐ a. Not Sensitive to outliers ~~X~~
- ☒ b. Issues in clustering data of various sizes and density
- ☒ c. cannot find clusters with arbitrary shapes
- ☒ d. K- should be selected
- ☐ e. Initial centroids are selected systematically ~~X~~

Consider the following distance matrix

Consider the following distance matrix. Assuming A and E are initial centroids for K-means algorithm, what would be the resulted clusters after the first epoch?

	A	B	C	D	E	F	G
A	0						
B	9	0					
C	1	0	0				
D	16	4	8	0			
E	3	12	0	20	0		
F	11	26	13	12	1	0	
G	19	25	23	2	8	10	0

Select one:

- ☐ a. A, C, D, G and B, E, F
- ☐ b. A, D and B, C, E, F, G
- ☐ c. A, B, G and C, D, E, F
- ☐ d. A, B, C and D, E, F, G
- ☒ e. A, B, D and C, E, F, G

Handwritten notes: A, E, B, C, D, F, G

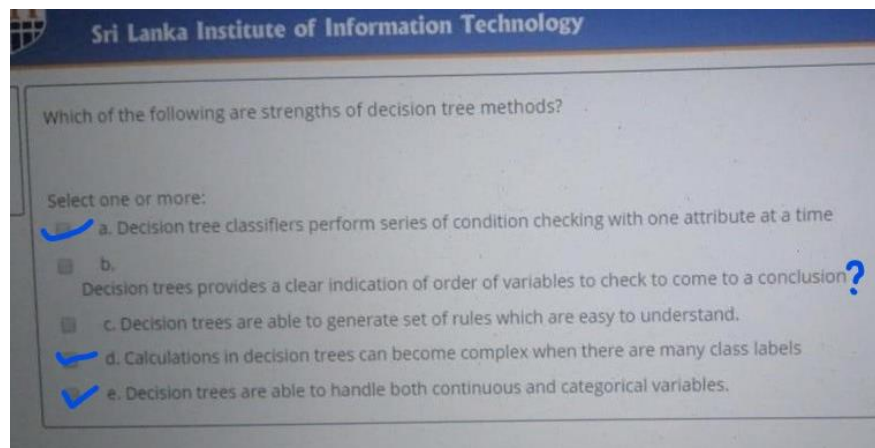
Which of the following is not a classification

Which of the following is not a classification technique?

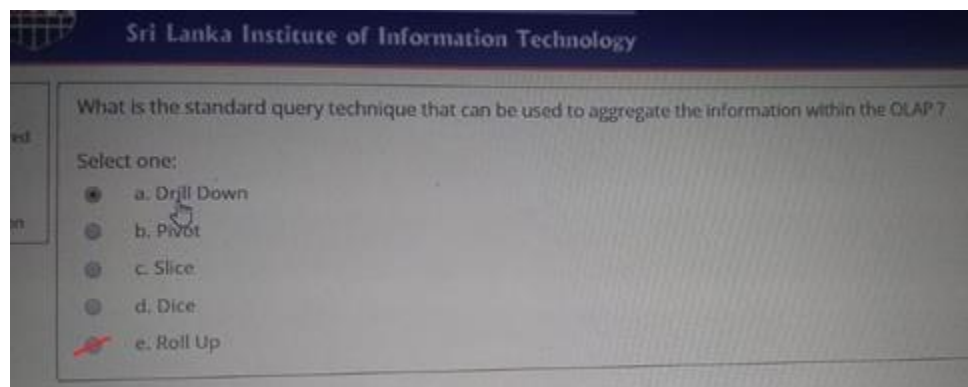
Select one:

- ☐ a. K-nearest neighbor
- ☐ b. Random forest
- ☐ c. Support vector machine
- ☒ d. K-means analysis
- ☐ e. Neural networks

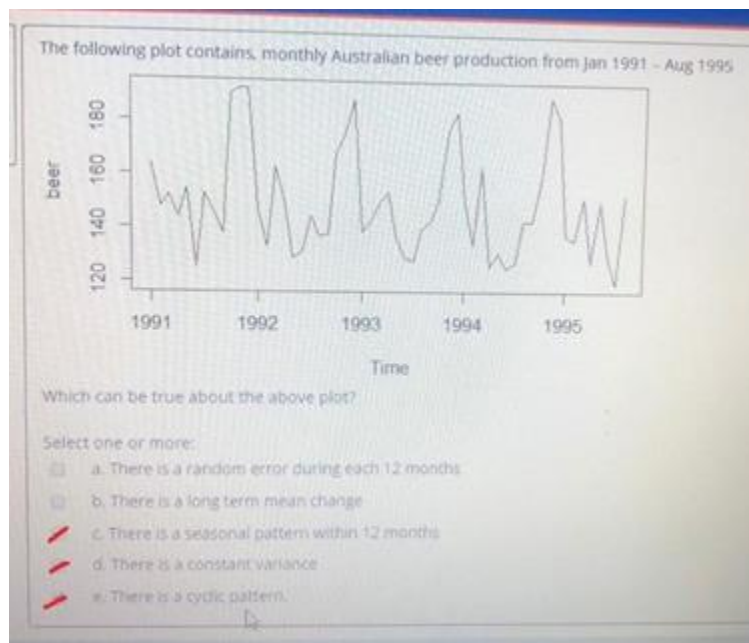
Which of the following are strengths



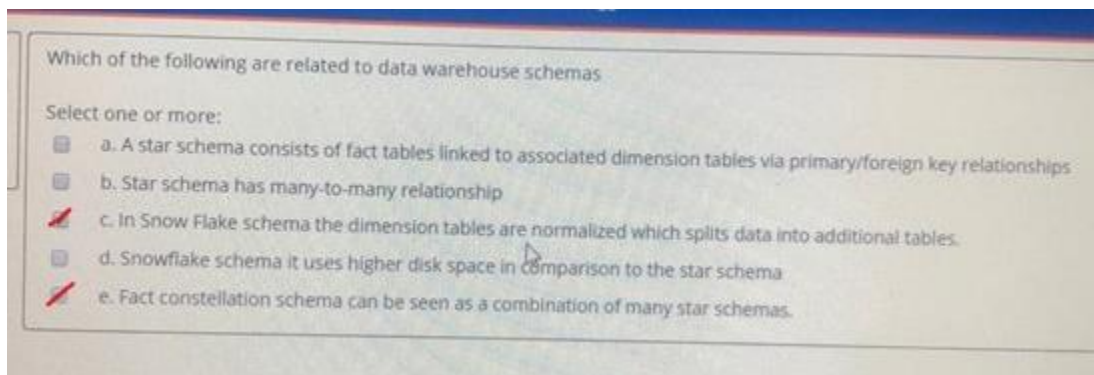
What is the standard query technique that can be used



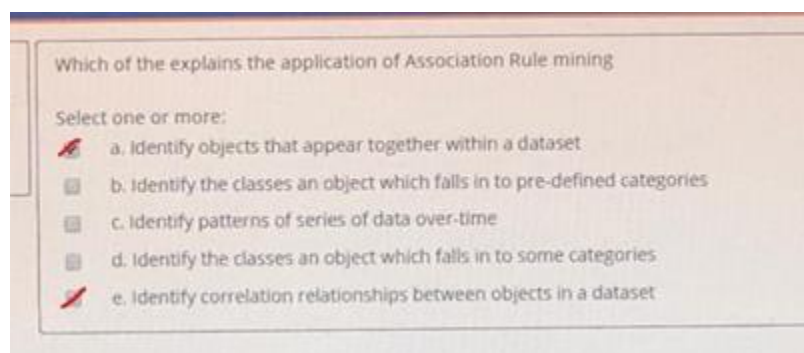
The following plot contains



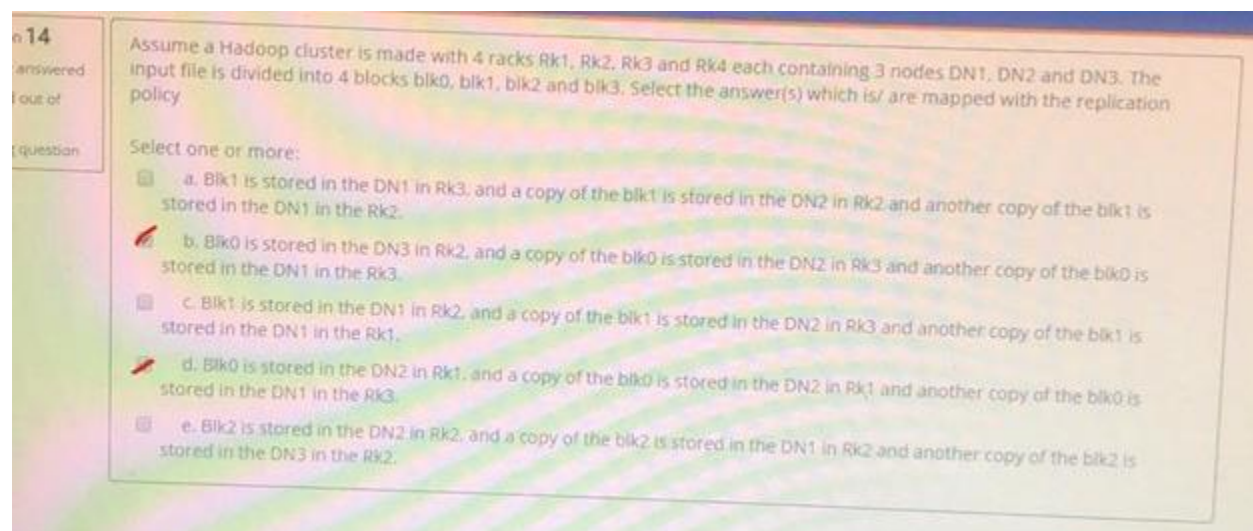
Which of the following are related



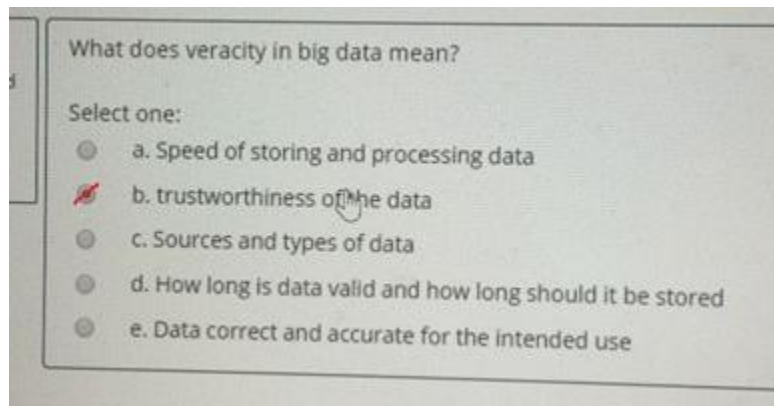
Which of the explains the application



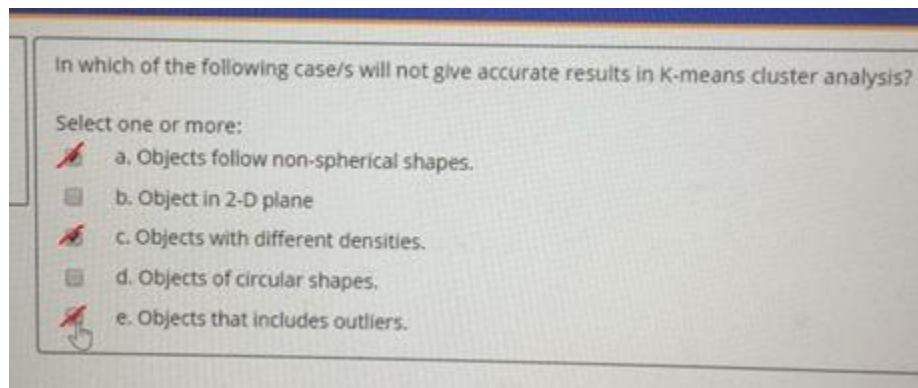
Assume a Hadoop cluster is made with 4 racks



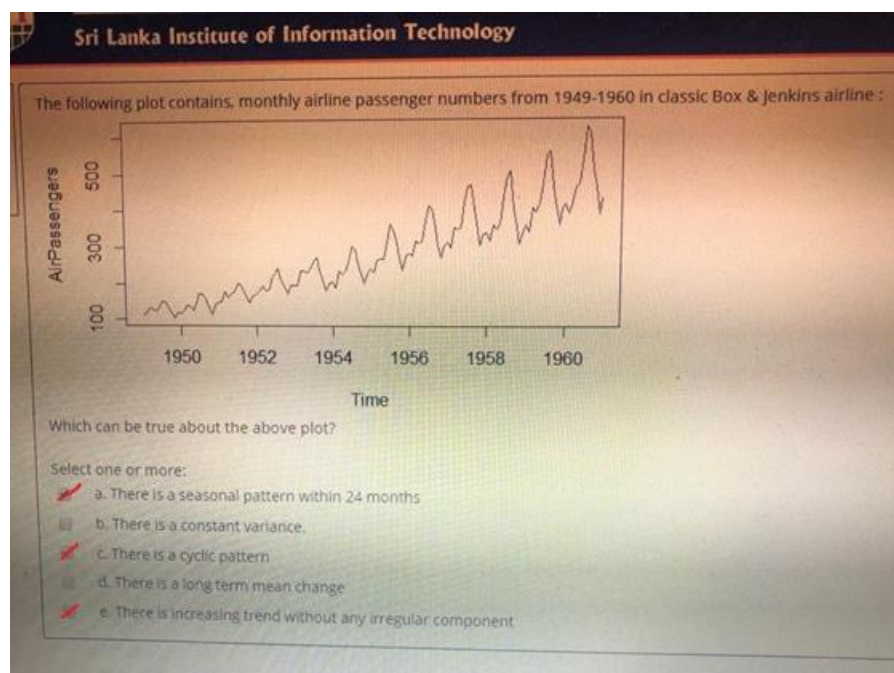
What does veracity in big data mean



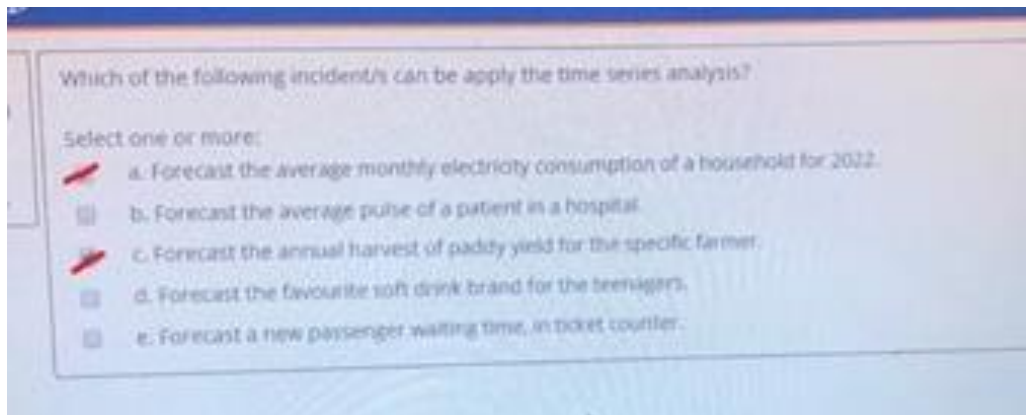
In which of the following case/s will not give accurate results



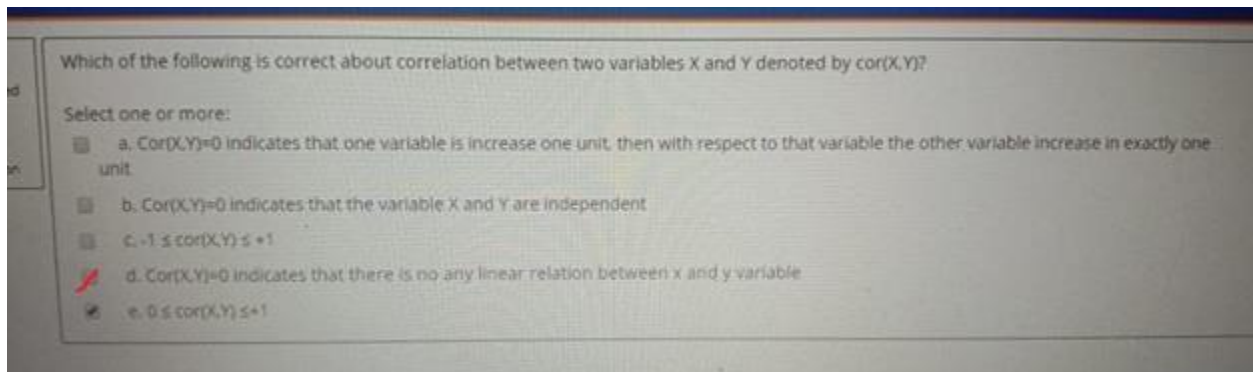
The following plot contains



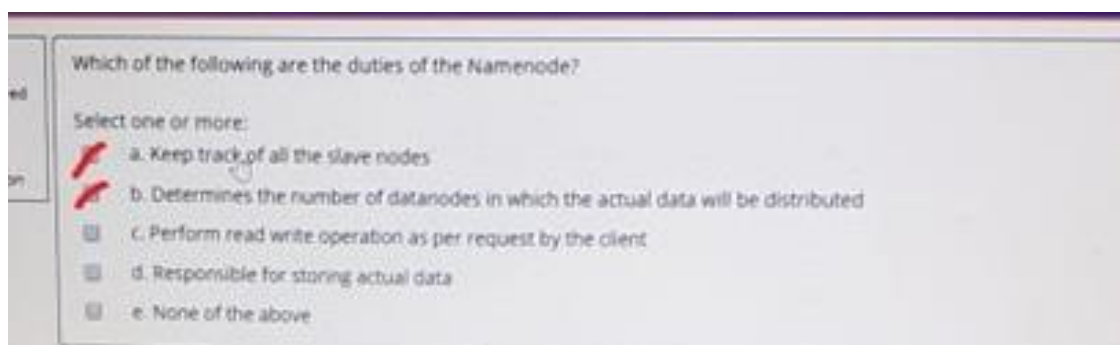
Which of the following incident



Which of the following is correct about



Which of the following are the duties



Which of the following are true

Which of the following are true related to dimension tables?

Select one or more:

- ☐ a. Dimension tables are completely normalized
- ☒ b. A dimension table contains dimensions of a fact
- ☐ c. Surrogate keys are useful to handle changes in dimension table attributes
- ☐ d. Dimension tables usually have more records than the fact tables
- ☒ e. Dimensions offers descriptive characteristics of the facts with the help of their attributes

Following input file contains employee data

Following input file contains employee data in a large financial company.

empID	name	mgrID	position	Salary (\$)	deptID
7369	Smith	7788	Analyst	1200	10
7698	Allen	7788	Analyst	2400	10
7782	Ward	7900	Clerk	980	20
7788	Jones	-	Manager	3500	10
7900	Martin	-	Manager	3000	20

Select the key value pairs for map and reduce phases to display the total salary of each department.

Map phase: key: empID , value: salary

Reduce phase: key: deptID , value: total_salary

Which of the following is correct

Which of the following is correct in relation to SSR?

Select one or more:

- ☒ a. It represents discrepancy between a model and actual data
- ☐ b. $SSR = SST + SSE$
- ☐ c. It represents the error in prediction
- ☐ d. It represents the total variability of the dataset.
- ☒ e. $SSR = SST - SSE$

Which of the statements are true on classical time

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

Select one or more:

- ☐ a. The additive model cannot be used when trend is constant.
- ☒ b. The additive model can be used when the seasonal variation is constant over the time.
- ☐ c. The multiplicative model cannot include the irregular component.
- ☒ d. The multiplicative model can be used when the seasonal variation increases over the time
- ☐ e. The multiplicative model can be used when constant trend only.

In which of the following case

In which of the following case/s will not give accurate results in K-means cluster analysis?

Select one or more:

- ☒ a. Objects that includes outliers.
- ☐ b. Objects of circular shapes.
- ☐ c. Object in 2-D plane
- ☒ d. Objects with different densities.
- ☒ e. Objects follow non-spherical shapes.

Which of the following variable

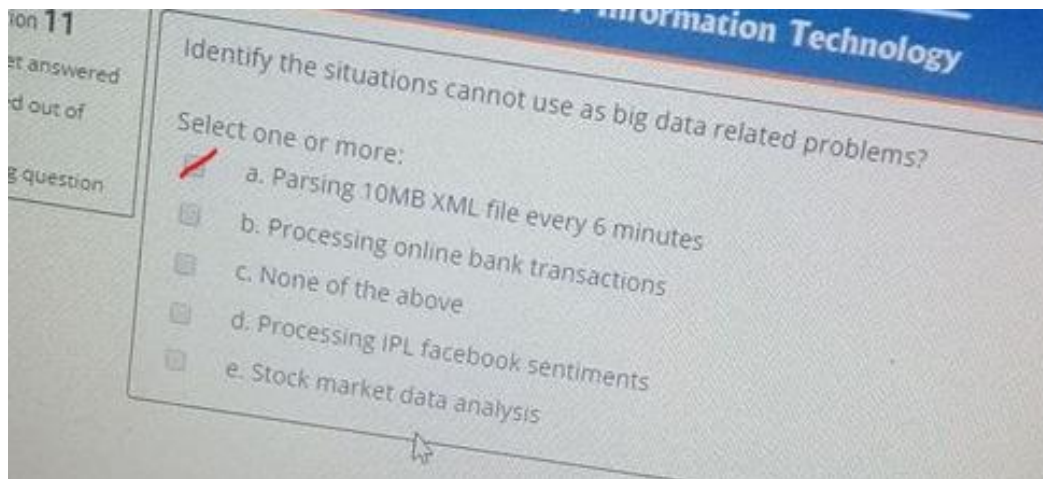
Which of the following variable/s can be considered as example/s for time series data?

Select one or more:

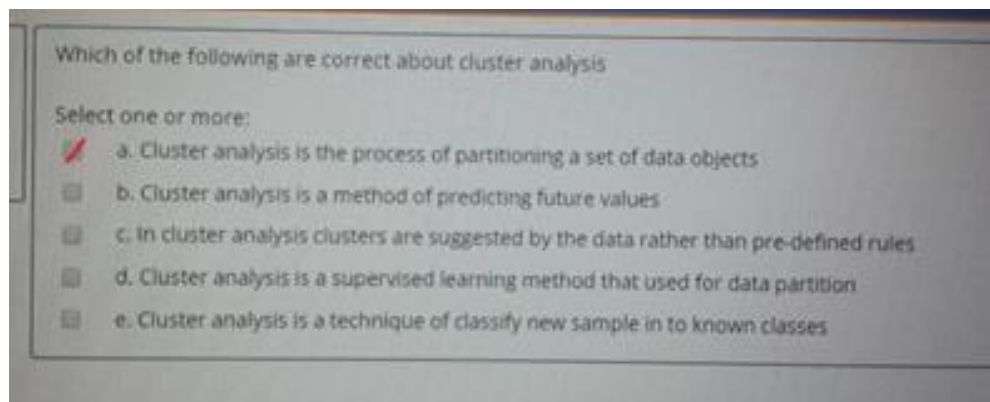
- ☐ a. Crime rate in London in each month during 2019.
- ☐ b. Monthly sales income of a supermarket in 2019.
- ☐ c. The favorite soft drink brand, recorded for 100 adults.
- ☐ d. Annual paddy yield per 1 km², recorded for 60 farmers.
- ☒ e. Quarterly GDP in Sri Lanka during 2000-2010.

?

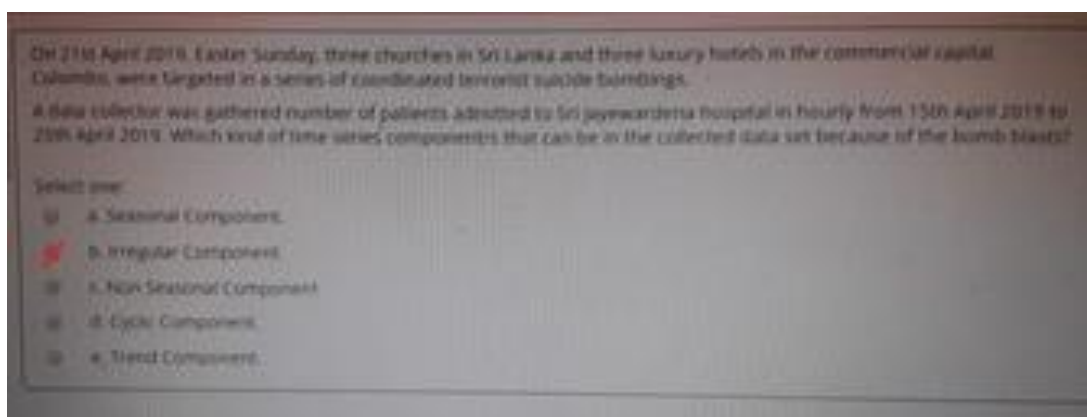
Identify the situations



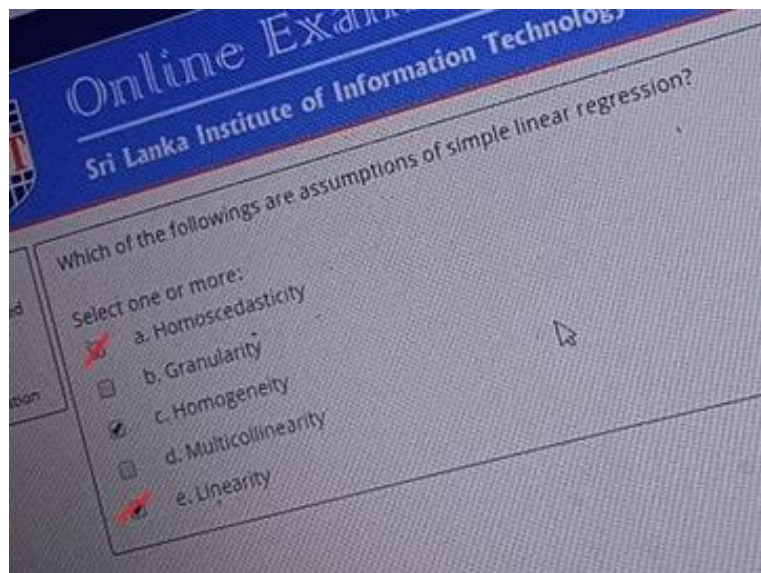
Which of the following are correct



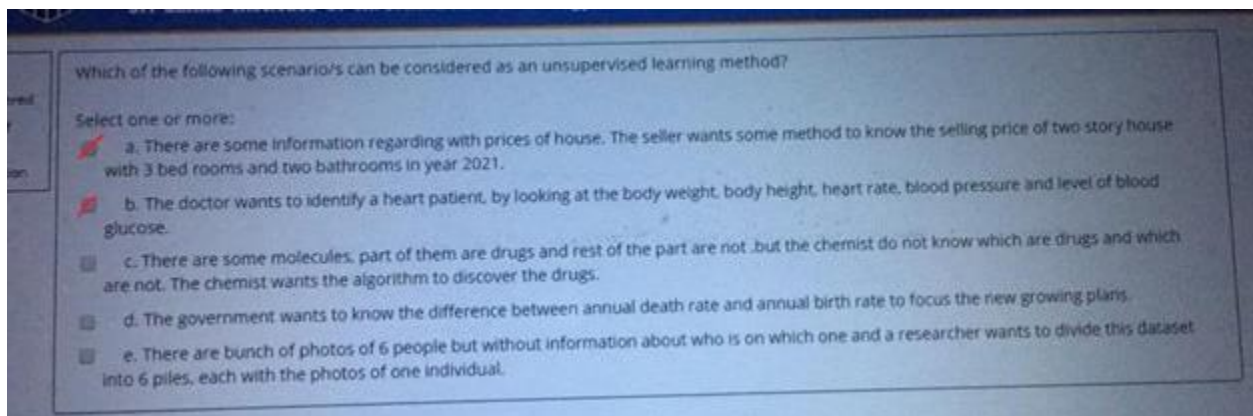
On 21st April 2019



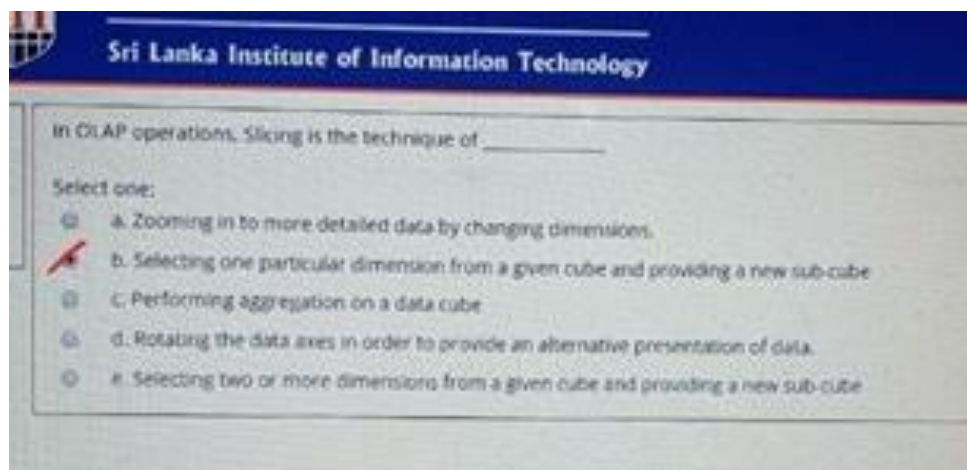
Which of the followings are assumptions



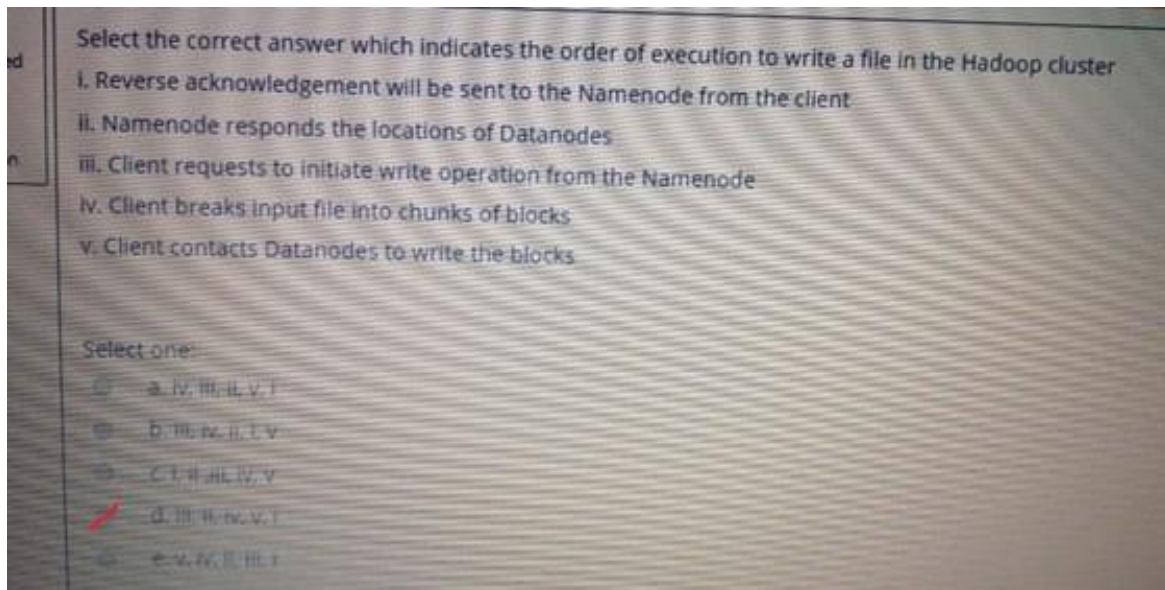
Which of the following scenario/s



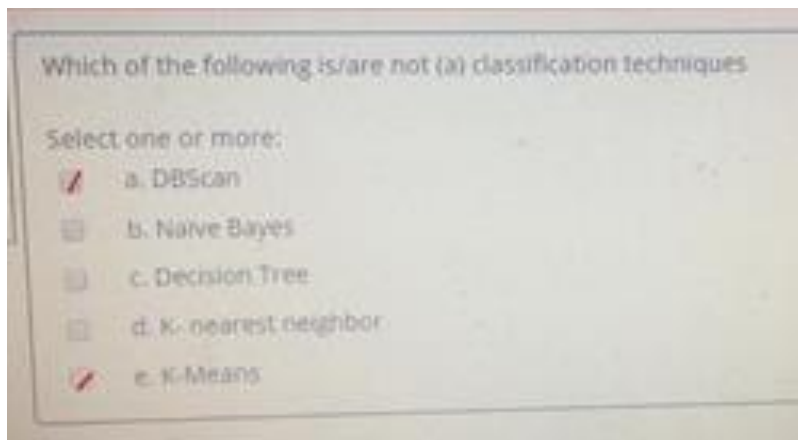
In OLAP operations



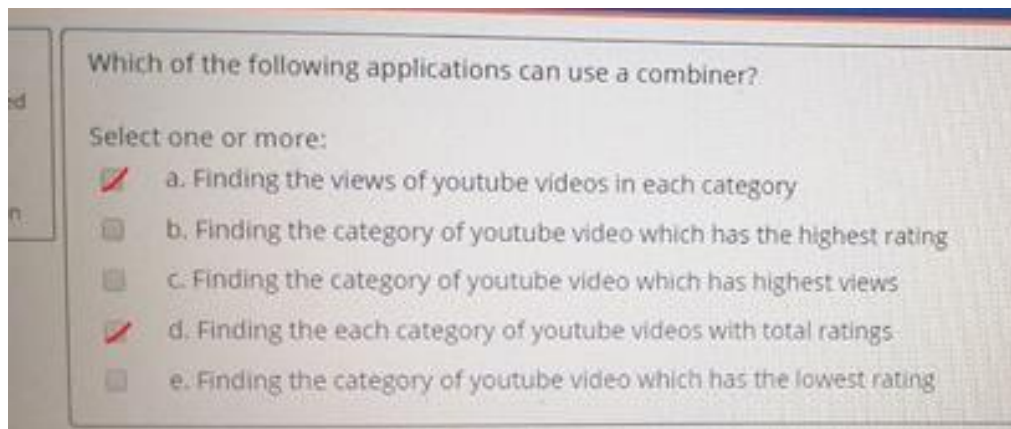
Select the correct answer which indicates



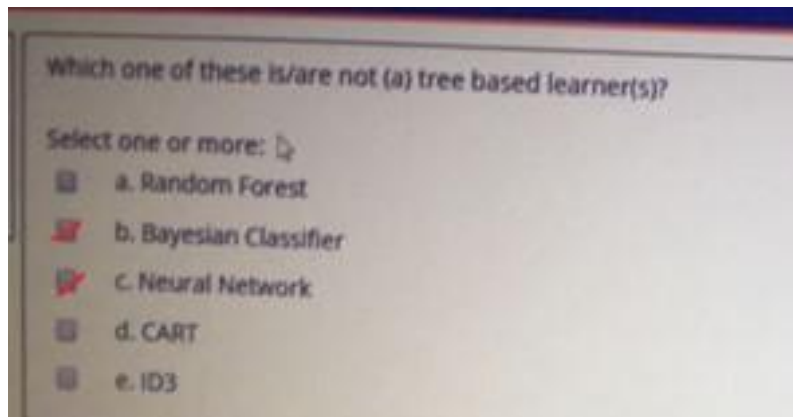
Which of the following is/are not (a)



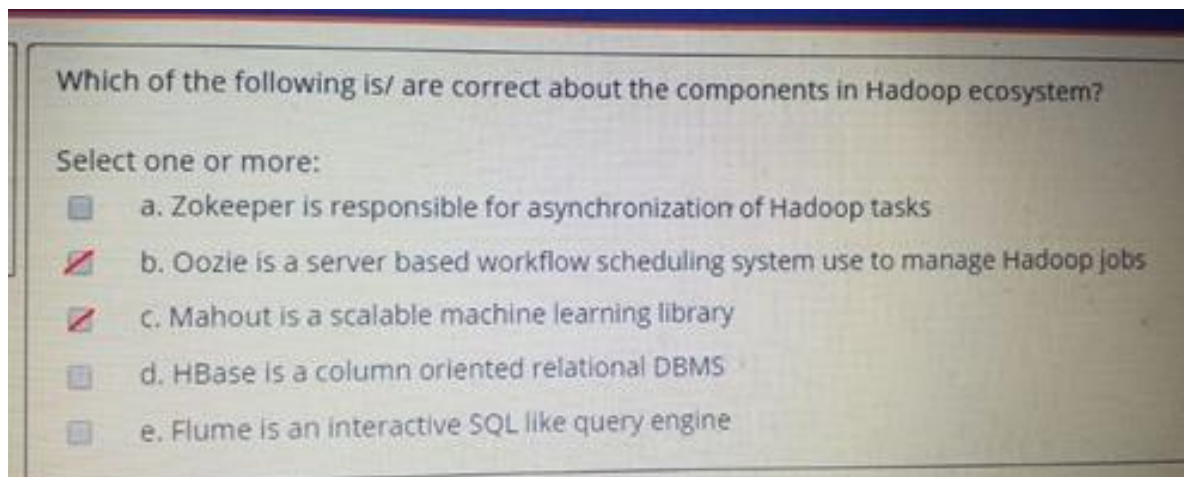
Which of the following applications



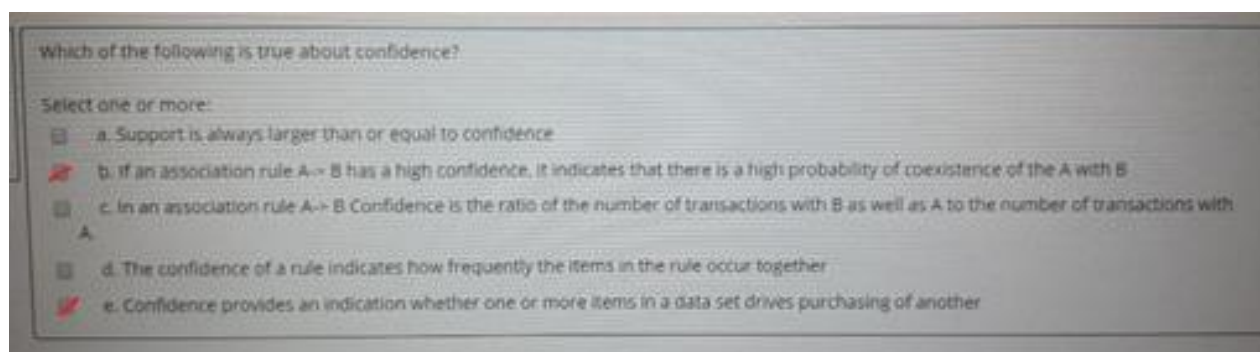
Which one of these is/are not



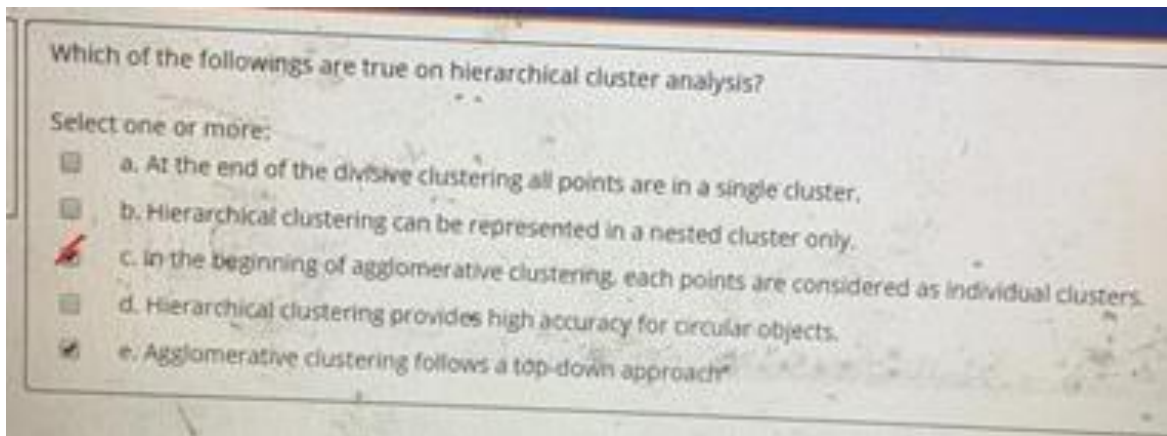
Which of the following is



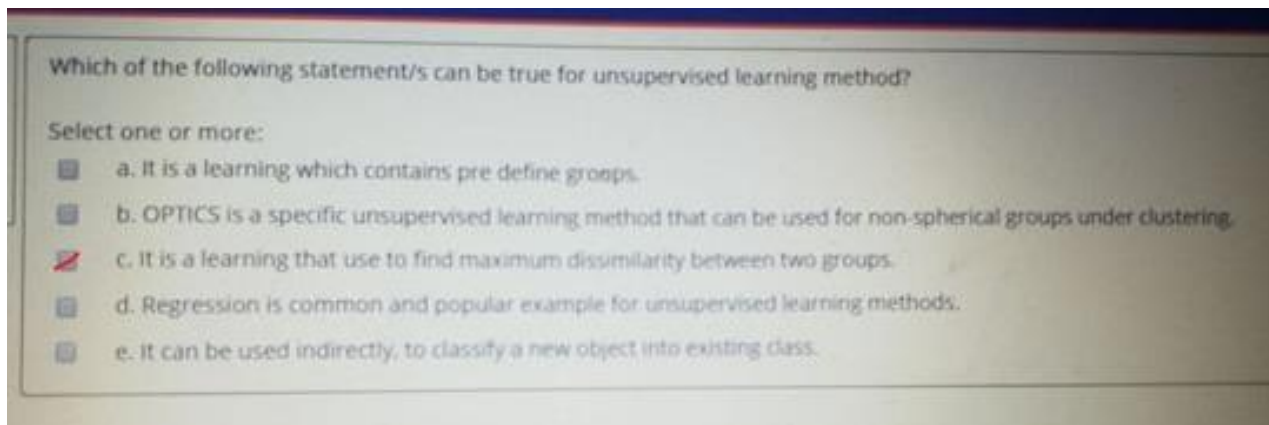
Which of the following is true



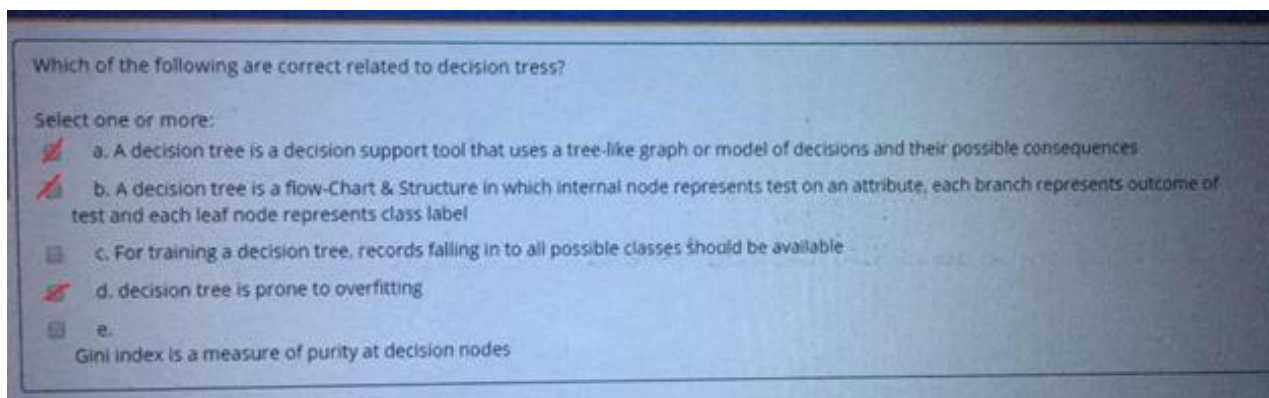
Which of the followings are true on hierarchical



Which of the following statement/s



Which of the following are correct related



Which of the following are correct in relation to the

Which of the following are correct in relation to the Fact tables?

Select one or more:

- ☐ a. Fact tables could contain hierarchies
- ☒ b. Fact table contains the data that needs to be analyzed
- ☐ c. Fact table content should be more descriptive
- ☐ d. Fact table usually have less records than the dimension tables
- ☐ e. Fact tables are completely normalized

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, pliers, hammer, nails
10	brush, paints, hammer

Consider the following association hammer \rightarrow 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

Consider the following dataset used for calculating pressure felt in underv

Depth	Pressure (pounds per square inch)
10	19.03
20	23.36
39	27.69
40	32.02
50	36.35
60	40.68
70	45.01
80	49.34
90	53.67
100	58.00

What is the value of β_0 ? 13.8427

What is the value of β_1 ? 0.4414

What is the Pressure felt when depth is 55? 38.11

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 β_0 ?

What is the value of β_1 ?

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, pliers, hammer, nails
10	brush, paints, hammer

Consider the following association rule nails \rightarrow sandpaper on the set of transactions above:

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

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

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? MB

ii. What is the size of the file? 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 β_0 ? -257.3038

What is the value of β_1 ? 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 β_0 ? -17.0021

What is the value of β_1 ? 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 β_0 ? 8551.9042

What is the value of β_1 ? 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 β_0 ? **5.5312**

What is the value of β_1 ? **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, pliers, hammer, nails
10	brush, paints, hammer

Consider the following association hammer \rightarrow 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 β_0 ? **5.5312**

What is the value of β_1 ? **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, pliers, 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 \rightarrow Sandpaper
- ☐ b. Pliers \rightarrow Nails
- ☒ c. Brush \rightarrow Paint
- ☒ d. Hammer \rightarrow Nails
- ☒ e. Paint \rightarrow Brush

Consider the following dataset used

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

Fever C°	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, pliers, hammer, nails
10	brush, paints, hammer

Consider the following association rule hammer → pliers 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 C°	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**