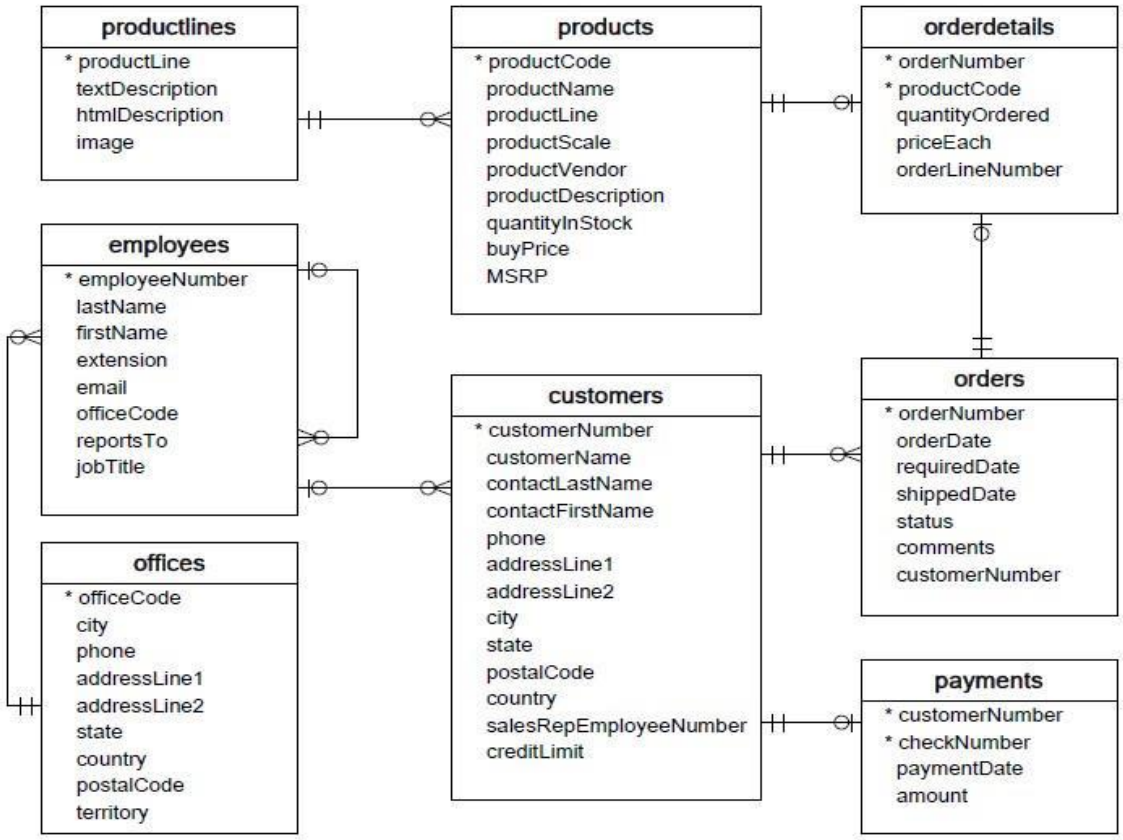
WORKSHEET 3 SQL

Refer the following ERD and answer all the questions in this worksheet. You have to write the queries using my sql for the required Operation.



1) Write SQL query to create table **Customers.**
Answer:

create table Customers (customer Number int primary key,

customerName varchar(30),

contactLastName varchar(20),

contactFirstName varchar(20),

phone int,

addressLine1 varchar(20),

addressLine2 varchar(20),

city varchar(20),

state varchar(20),

postalCode int,

country varchar(20),

salesRepEmployeeNumber int,

creditLimit float);

2)create table Orders(orderNumber int primary key,

Answer:

orderDate int,

requiredDate int,

shippedDate int,

status varchar(20),

comments varchar(30),

customerNumber int);

3)Write SQL query to show all the columns data from the **Orders** Table

Answer : select * from Orders;

4. Write SQL query to show all the comments from the **Orders** Table

Answer : select comments from Orders;

5) Write a SQL query to show orderDate and Total number of orders placed on that date, from **Orders** table

Answer: select orderDate, count(orderNumber) from orders group by orderDate;

6) Write a SQL query to show employeeNumber, lastName, firstName of all the employees from **employees** table.

Answer: select employeeNumber, lastName,firstName from employees;

7) Write a SQL query to show all orderNumber, customerName of the person who placed the respective order.

Answer: select orderNumber,customerName from orders,customers where order.customerNumber=customers.customerNumber;

8) Write a SQL query to show name of all the customers in one column and salerepemployee name in another column.

Answer : select customerName, salesRepEmployeeNumber from customers;

9) Write a SQL query to show Date in one column and total payment amount of the payments made on that date from the **payments** table.

Answer : select date(paymentdate), sum(amount) as total from payments group by date(paymentdate);

10) Write a SQL query to show all the products productName, MSRP, productDescription from the **products** table.

Answer : select productName, MSRP, productDescription from products;

11) Write a SQL query to print the productName, productDescription of the most ordered product

Answer : select productName, productDescription from products group by productName order by count(productName) desc limit 1;

12) Write a SQL query to print the city name where maximum number of orders were placed.

Answer : select city from customers inner join orders on customers.customerNumber=orders.customerNumber

group by city order by city desc limit 1;

13) Write a SQL query to get the name of the state having maximum number of customers.

Answer : select state from customers group by state order by count(customerNumber) desc limit 1;

14) Write a SQL query to print the employee number in one column and Full name of the employee in the second column for all the employees

Answer : select employeeNumber, concat((firstName,'',lastName) as fullname from employees;

15) Write a SQL query to print the orderNumber, customer Name and total amount paid by the customer for that order (quantityOrdered × priceEach).

Answer :

select orders.orderNumber, customers.customerName,

oderdetails.quantityOrdered*orderdetails*priceEach as totalamount

from ((orders

inner join customerson orders.customerNumber=customer.customerNumber)

inner join orderdetails on orders.orderNumber=orderdetails.orderNumber)

STATISTICS WORKSHEET-3

1. Which of the following is the correct formula for total variation?
Answer :B) Total Variation = Residual Variation + Regression Variation

2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.
Answer : C) binomial

3) How many outcomes are possible with Bernoulli trial?
Answer : A) 2

4) If Ho is true and we reject it is called

Answer : A) Type-I error

5) Level of significance is also called:
Answer:C) Level of confidence

6) The chance of rejecting a true hypothesis decreases when sample size is:
Answer :A) Decrease

7)Which of the following testing is concerned with making decisions using data?
Answer : B) Hypothesis

8) What is the purpose of multiple testing in statistical inference?
Answer:D) All of the mentioned (Minimize errors, Minimize false positives ,Minimize false negatives)

9)Normalized data are centred at and have units equal to standard deviations of the original data
Answer :A)0

10) What Is Bayes' Theorem?

In statistics the Bayes' theorem is a mathematical formula used to determine the conditional probability of events. Essentially, the Bayes' theorem describes the probability of an event based on prior knowledge of the conditions that might be relevant to the event.

$$P(A|B) = P(A)\, P(B|A)/\mathbf{P(B)}$$

11) What is z-score?

Z-scores are expressed in terms of standard deviations from their means. These z-scores have a distribution with a mean of 0 and a standard deviation of 1.

$$Z= x-\mu)/\sigma$$

12) What is t-test?

A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process actually has an effect on the population of interest, or whether two groups are different from one another.

13) What is percentile?

In statistics percentiles are used to understand and interpret data. Percentiles indicate the percentage of scores that fall below a particular value. They tell you where a score stands relative to other scores.

14) What is ANOVA?

ANNOVA is one way Analysis of Variance. One-Way Analysis of Variance tells you if there are any statistical differences between the means of three or more independent groups.

15) How can ANOVA help?

ANOVA can help you know whether or not there are significant differences between the means of your independent variables. When you get to know how each independent variable mean is different from the others then you can begin to understand which of them has a connection to your dependent variable, and we begin to learn what is driving that behaviour.

MACHINE LEARNING

1. Which of the following is an application of clustering?
Answer : D) All of the above(Biological network analysis ,Market trend prediction , Topic modeling)

2. On which data type, we cannot perform cluster analysis?
Answer : D) None

3. Netflix's movie recommendation system uses-
Answer :C) Reinforcement learning and Unsupervised learning

4. The final output of Hierarchical clustering is-
Answer : D) All of the above(The number of cluster centroids , The tree representing how close the data points are to each other ,A map defining the similar data points into individual groups)

5. Which of the step is not required for K-means clustering?
Answer :D) None

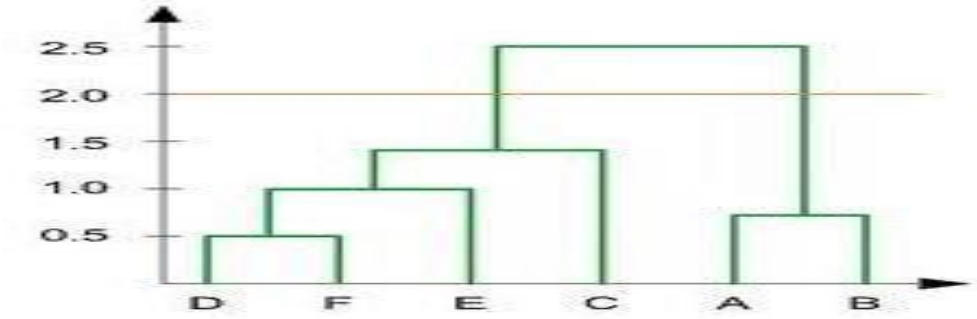
6.Which is the following is wrong?
Answer : C) k-nearest neighbour is same as k-means

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?
Answer : D) (Single-link ,Complete-link , Average-link)

8. Which of the following are true?

Answer : A) Clustering analysis is negatively affected by multicollinearity of features

9. In the figure above, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed?



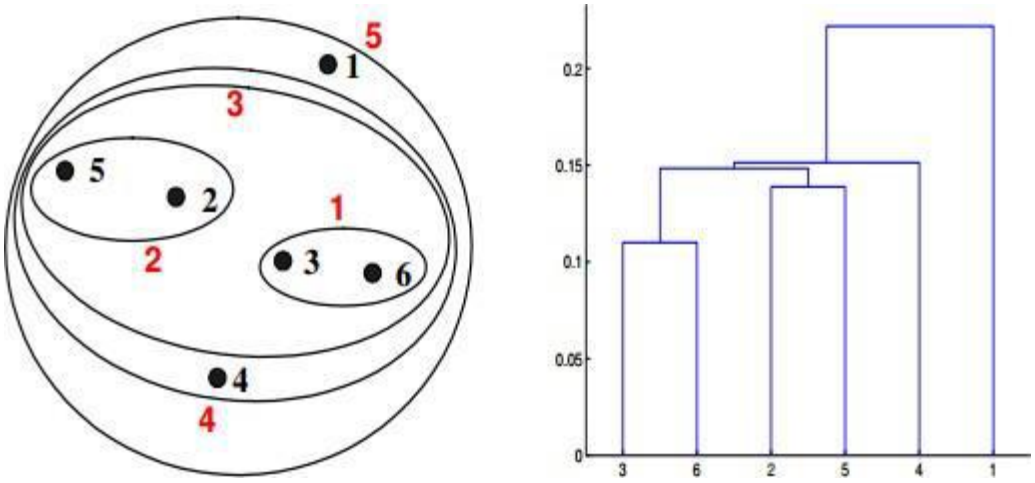
Answer : A) 2

10. For which of the following tasks might clustering be a suitable approach?

Answer: B) Given a database of information about your users, automatically group them into different market segments.

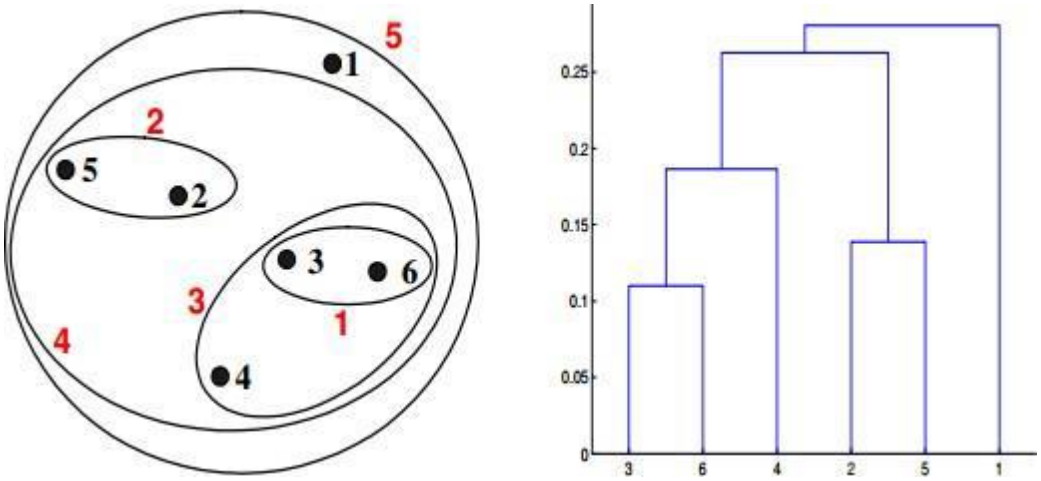
11. Given, six points with the following attributes:

Answer : A



12.Given, six points with the following attributes:

Answer: B



13. What is the importance of clustering?

Answer: The purpose of clustering and classification algorithms is to make sense of and extract value from large sets of structured and unstructured data. If you're working with huge volumes of unstructured data, it only makes sense to try to partition the data into some sort of logical groupings before attempting to analyze it.

14. How can I improve my clustering performance?

Answer :K-means clustering algorithm can be significantly improved by using a better initialization technique, and by re starting the algorithm, or avoiding unbalanced cluster size.