SQL WORKSHEET 3

Answer No. 1: 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);

Answer No. 2: create table Orders(orderNumber int primary key,

orderDate int,

requiredDate int,

shippedDate int,

status varchar(20),

comments varchar(30),

customerNumber int);

Answer No. 3: select * from Orders;

Answer No. 4: select comments from Orders;

Answer No. 5: select orderDate, count(orderNumber) from orders group by orderDate;

Answer No. 6: select employeeNumber, lastName,firstName from employees;

Answer No. 7: select orderNumber,customerName from orders,customers where order.customerNumber=customers.customerNumber;

Answer No. 8: select customerName, salesRepEmployeeNumber from customers;

Answer No. 9: select date(paymentdate), sum(amount) as total from payments group by date(paymentdate);

Answer No. 10: select productName, MSRP, productDescription from products;

Answer No. 11: select productName, productDescription from products group by productName order by count(productName) desc limit 1;

Answer No. 12: select city

from customers

inner join orders on customers.customerNumber=orders.customerNumber

group by city order by city desc limit 1;

Answer No. 13: select state from customers group by state order by count(customerNumber) desc limit 1;

Answer No. 14: select employeeNumber, concat((firstName,'',lastName) as fullname from employees;

Answer No. 15: 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

Answer No. 1: B

Answer No. 2: C

Answer No. 3: A

Answer No. 4: A

Answer No. 5: C

Answer No. 6: A

Answer No. 7: B

Answer No. 8: D

Answer No. 9: A

Answer No. 10: 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.

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

Answer No. 11: 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.

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

Answer No. 12: 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.

Answer No. 13: 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.

Answer No. 14: 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.

Answer No. 15: 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 WORKSHEET-3

Answer No. 1: D

Answer No. 2: D

Answer No. 3: C

Answer No. 4: D

Answer No. 5: D

Answer No. 6: C

Answer No. 7: D

Answer No. 8: A

Answer No. 9: A

Answer No. 10: B

Answer No. 11: A

Answer No. 12: B

Answer No. 13: 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.

Answer No. 14: 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.