**WORKSHEET WORKSHEET 3 SQL**

1. Write SQL query to create table Customers.

**Answer.** create table Customers (customerNumber 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. Write SQL query to create table Orders.

**Answer.** create table Orders(orderNumber int primary key,

orderDate date,

requiredDate date,

shippedDate date,

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 OrdersTable.

**Answer.** select comments from Orders;

5. Write a SQL query to show orderDate and Total number of orders placed on that date, from Orderstable.

**Answer.** select orderDate, count(orderNumber) as count from Orders group by orderDate;

6. Write a SQL query to show employeNumber, 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 Orders.customerNumber=Customer.customerNumber;

8. Write a SQL query to show name of all the customers in one column and salerepemployee name inanother 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, orderdetails.quantityOrdered\*orderdetails\*priceEach as totalamount

from ((Orders

inner join Customerson Orders.customerNumber=Customers.customerNumber)

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

**WORKSHEET 3 MACHINE LEARNING**

**Answer 1.** D. All of the above

**Answer 2.** D. None

**Answer 3.** C. Reinforcement learning and Unsupervised learning

**Answer 4.** D. All of the above

**Answer 5.** D. None

**Answer 6.** C. k-nearest neighbour is same as k-means

**Answer 7.** D. 1, 2 and 3

**Answer 8.** A. 1 only

**Answer 9.** A. 2

**Answer 10.** B. Given a database of information about your users, automatically group them into different market segments.

**Answer 11.** A

**Answer 12.** B

**Answer 13.** Clustering is a method to classify the raw data and find hidden patterns in the dataset. The purpose of clustering is to make sense out of unstructured data and extract some informative values out of the data. It tries to partition the data into groups of similar objects and analyse the groups. Clustering allows us to learn about the unlabelled data.

**Answer 14.** Performance of clustering can be improved by choosing a better initialization technique, by choosing lesser number of clusters as more the number of clusters more the error, by repeating the algorithm many times and also by avoiding unbalanced cluster size.

**WORKSHEET 3 STATISTICS**

**Answer 1**. b) Total Variation = Residual Variation + Regression Variation

**Answer 2.** c) binomial

**Answer 3.** a) 2

**Answer 4.** a) Type-I error

**Answer 5.** c) level of confidence

**Answer 6.** a) decreases

**Answer 7.** b) hypothesis

**Answer 8.** d) All of the mentioned

**Answer 9.** a) 0

**Answer 10.** Baye’s theorem in statistics is a mathematical formula to determine the conditional probability of events. In simple words Baye’s theorem describes the probability of occurance of an event based on the conditions that might be related to the event. Formula for baye’s theorem is P(A|B)=P(B|A)P(A)/P(B).

**Answer 11.** A z score is the score that gives us an idea of how far the data points are from the mean. Technically, a z-score tells how many standard deviations below or above the raw data point is from the mean. Formula for z-score is z**= (x – μ) / σ.**

**Answer 12.** T-test is one of the tests used for hypothesis testing in statistics. A t-test allows us to compare the average values of two datasets and determine if they belong to the same population. A t-test is a type if inferential statistic which is used to determine if there is a significant difference between the mean of two groups, which may be related in a certain way.

**Answer 13.** A percentile is a comparison score between a particular score and the rest of the group. It shows the percentage of scores that a score surpassed. For example, if a score of 87 is ranked 90th percentile it means that 87 is higher than 90 percent of the scores.

**Answer 14.** ANOVA also called analysis of variance (is an expansion of t-test) is the method of comparing if the means of two or more samples are distinct from each other. ANOVA checks the impact of one or more factors by comparing the means of different groups.

**Answer 15.** ANOVA helps us to figure out if there is significant difference between the means of the samples on the basis of which we can either accept the null hypothesis or reject it.