MACHINE LEARNING

1.d

2.d

3.c

4.b

5.d

6.c

7.d

8.a

9.a

10.a,b

11.a

12.b

13. **What is the importance of clustering?**

Ans: Clustering is useful for exploring data. If there are many cases and no obvious groupings, clustering algorithms can be used to find natural groupings. Clustering can also serve as a useful data-preprocessing step to identify homogeneous groups on which to build supervised models.

14. **How can I improve my clustering performance?**

Ans: K-means clustering algorithm can be significantly improved by using a better initialization technique, and by repeating (re-starting) the algorithm. When the data has overlapping clusters, k-means can improve the results of the initialization technique. When the data has well separated clusters, the performance of k-means depends completely on the goodness of the initialization.

WORKSHEET 3 SQL

1.CREATE TABLE Customers (
CustomerNumber int NOT NULL,
CustomerName varchar(255) ,
contactLastName varchar(255),
contactFirstName varchar(255),
Phone int ,
addressline1 varchar(255),
addressLine2 varchar(255),
city varchar(255),
state varchar(255),
postalcode varchar(255),
country varchar(255),
SalesRepEmpoyeeNumber varchar(255),
creditLimit float,
PRIMARY KEY(CustomerNumber));

2.CREATE TABLE Orders (orderNumber int NOT NULL,
orderDate date,
requiredDate date,
shippedDate date,
status varchar(255),
comments varchar(255),
customerNumber int ,
PRIMARY KEY(orderNumber),
FOREIGN KEY (customerNumber) REFERENCES Customers (customerNumber));

3. select * from Orders;
4. select comments from Orders;
5. select count(orderNumber) from Orders where orderDate=--;
6. select employeeNumber, lastName, firstName from employees;
7. select customerName,

8. select Cust.customerName,empl.lastName,emp.firstName from Customers cust innerjoin employees emp
on cust.
9.select paymentDate, sum(amount) from payments groupby Date;
10. select productName, MSRP, productDescription from products;
11. select pro.productName,pro. productDescription from product pro inner join orderdetails od
on pro.productCode = od.productCode
Group by(pro.productCode)

order by od.quantityOrdered asc;

12. select cust.city from Customers cust inner join Order od
on cust.customerNumber=od. customerNumber
group by cust.city
where max(count(od.orderNumber))

13. select state from Customers
Where max(count(customerNumber))

14.select employee number,cocat(firstName, “ ” ,lastName) as Name from employees

15.

STATISTICS WORKSHEET-3

1. b
2. c
3. a
4. a
5. c
6. a
7. b
8. d
9. 0
10. What Is Bayes' Theorem?

Ans: Bayes' Theorem is a mathematical formula for determining conditional probability. Conditional probability is the likelihood of an outcome occurring, based on a previous outcome occurring. Bayes' theorem provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence.

Formula:

P(A□B)=P(B)P(A□B)=P(B)P(A)□P(B□A)
where:P(A)= The probability of A occurringP(B)= The probability of B occurringP(A□B)=T he probability of A given BP(B□A)= The probability of B given AP(A□B))= The probabilit y of both A and B occurring

11. What is z-score?

Ans: Simply put, a z-score (also called a *standard score*) gives you an idea of how far from the mean a data point is. But more technically it's a measure of how many standard deviations below or above the population mean a raw score is.

Formula:

Z=X-U/std

Where X=observed value of the sample

U= mean of the sample

Std= standard deviation of the sample

12. What is t-test?

Ans: A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features. The t-test is one of many tests used for the purpose of hypothesis testing in statistics.

$t= m-u\backslash s/\sqrt{n}$

where t =t-test

m=mean
u=mean
s=standard deviation

n= variable set size

13. What is percentile?

Ans: A percentile is a comparison score between a particular score and the scores of the rest of a group. It shows the percentage of scores that a particular score surpassed. For example, if you score 75 points on a test, and are ranked in the 85 th percentile, it means that the score 75 is higher than 85% of the scores.

14. What is ANOVA?

Ans: Analysis of variance (ANOVA) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among means.

15. How can ANOVA help?
Ans:Finding Correalition between different groups of a categorical variable.