Machine Learning

1. D
2. D
3. C
4. B
5. D
6. C
7. D
8. A
9. A
10. B
11. A
12. B
13. The importance of clustering is to extract value from large sets of structured and unstructured data. If we are working with huge volume of unstructured data, it only makes sense to try to partition the data into some logical groups before attempting to analyze it.
14. K-means [clustering algorithm](https://www.sciencedirect.com/topics/computer-science/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.

SQL

1. create table Customers(customerNumber integer primary key,

customerName varchar(100) not null,

contactLastName varchar(50) not null ,

contactFirstName varchar(50) not null,

phone integer not null,

addressLine1 varchar(100) not null,

addressLine2 varchar(100),

city varchar(20) not null,

state varchar(20) not null,

postalCode varchar(10) not null,

country varchar(20) not null,

salesRepEmployeeNumber integer,

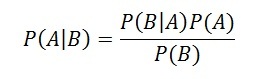
creditLimit integer

);

1. create table orders (orderNumber integer primary key, orderDate date not null , requiredDate date not null , shippedDate date not null, status varchar(50) not null, comments varchar(100), customerNumber integer, foreign key(customerNumber) references Customers(customerNumber));
2. select \* from orders;
3. select comments from orders;
4. select orderDate, count(orderDate) as number of orders from orders group by orderDate;
5. select employeeNumber, lastName, firstName from employees;
6. select o.orderNumber , c.customerName from orders o inner join customers c on o.customerNumber=c.customerNumber;
7. select customerName , salesRepEmployeeNumber from customers;
8. select paymentDate, sum(amount) from payments group by amount;
9. select productName, MSRP , productDescription from products;
10. select p.prodcutName, p.productDescription, count(o.productCode) from products p inner join orderDetails o on p.productCode=o.productCode group by productCode order by productCode desc;
11. select city from (select c.city, count(o.orders) num\_of\_orders from Customers c inner join orders o on c.customerNumber = o.customerNumber group by c.city order by num\_of\_orders desc) LIMIT 1;
12. select state from (select state , count(customerNumber) as customer\_count from Customers group by state order by customer\_count desc);
13. select employeeNumber , CONCAT(firstName, “ ”,lastName) as Fullname from employees;
14. select o.customerNumber , c.customerName, (od.quantityOrdered\*od.priceEach) as Total amount from Customers c inner join orders o on c.customerNumber=o.customerNumber inner join orderdetails od inner join o.orderNumber= od.orderNumber ;

Statistics

1. B
2. C
3. A
4. A
5. B
6. B
7. B
8. D
9. A
10. Bayes theorem is a mathematical formula for determining conditional probability. Conditional probability is the probability of an event happening, given that it has some relationship to one or more other events. For example, your probability of getting a parking space is connected to the time of day you park, where you park, and what conventions are going on at any time.



A,B = events

P(A|B)= probability of A given B is true

P(B|A)= probability of B given A is true

P(A),P(B) = independent probabilities of A and B

1. In simple terms Z-score is the number of standard deviation a point is away from the mean.

**z = (x – μ) / σ**

**x= observed value**

**μ=mean**

**σ= standard deviation**

1. A t-test is a type of inferential [statistic](https://www.investopedia.com/terms/s/statistics.asp) used to determine if there is a significant difference between the means of two groups, which may be related in certain features.T-test are performed when sample size is very small (less than 30)
2. In statistics, a percentile is a score below which a given percentage of scores in its frequency distribution falls or a score at or below which a given percentage falls. For example, the 50th percentile is the score below which or at or below which 50% of the scores in the distribution may be found.
3. An **ANOVA**test is a way to find out if survey or experiment results are [significant](https://www.statisticshowto.com/what-is-statistical-significance/). In other words, they help you to figure out if you need to [reject the null hypothesis](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/support-or-reject-null-hypothesis/) or accept the [alternate hypothesis](https://www.statisticshowto.com/what-is-an-alternate-hypothesis/).They are of two types one way and two way.

**One way:** One-way has one independent variable (with 2 [levels](https://www.statisticshowto.com/levels-in-statistics/)). For example: *brand of cereal*,

**Two way:** Two-way has two independent variables (it can have multiple levels). For example: brand of cereal, calories.

1. Like the [t-test](https://www.qualtrics.com/experience-management/research/t-test-analysis/), ANOVA helps you find out whether the differences between groups of data are statistically significant. It works by analyzing the levels of variance within the groups through [samples](https://www.qualtrics.com/experience-management/research/sampling-methods/) taken from each of them. If there is a lot of variance (spread of data away from the mean) within the data groups, then there is more chance that the mean of a sample selected from the data will be different due to chance.