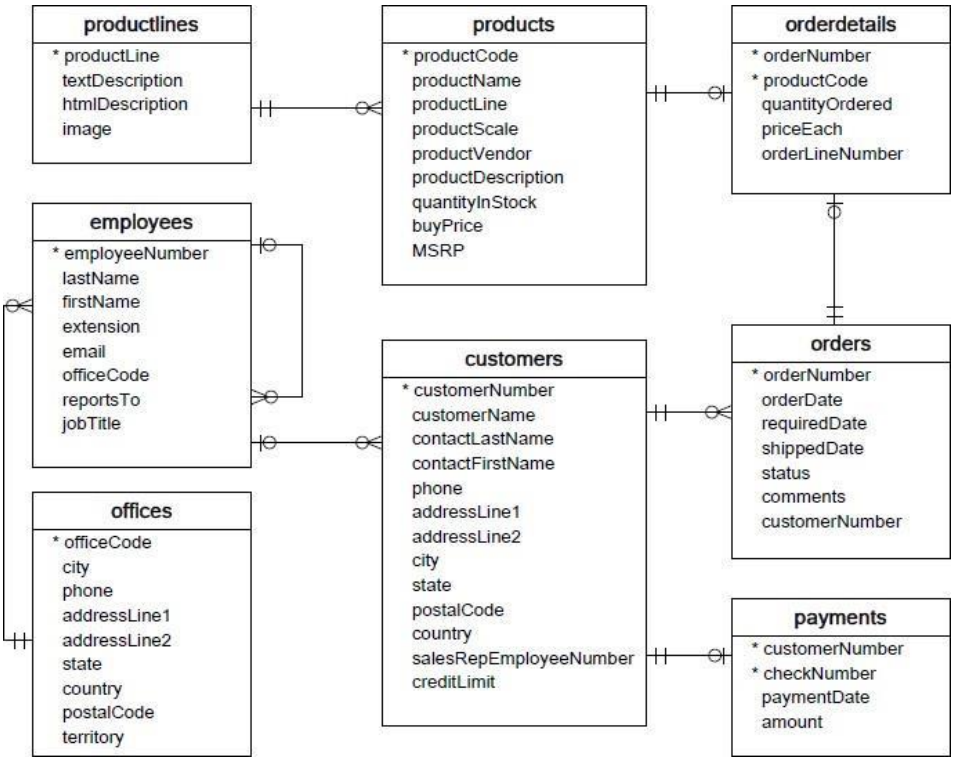


- **Customers**: stores customer's data.
- **Products**: stores a list of scale model cars.
- **Product Lines**: stores a list of product line categories.
- **Orders**: stores sales orders placed by customers.
- **Order Details**: stores sales order line items for each sales order.
- **Payments**: stores payments made by customers based on their accounts.
- **Employees**: stores all employee information as well as the organization structure such as who reports to whom.
- **Offices**: stores sales office data.

QUESTIONS:

Write a SQL query to show average number of orders shipped in a day (use Orders table).
Write a SQL query to show average number of orders placed in a day.
Write a SQL query to show the product name with minimum MSRP (use Products table).

STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

1. What is central limit theorem and why is it important?
2. What is sampling? How many sampling methods do you know?
3. What is the difference between type1 and typeII error?
4. What do you understand by the term Normal distribution?
5. What is correlation and covariance in statistics?
6. Differentiate between univariate ,Biavariate,and multivariate analysis.
7. What do you understand by sensitivity and how would you calculate it?
8. What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?
9. What is quantitative data and qualitative data?
10. How to calculate range and interquartile range?
11. What do you understand by bell curve distribution ?
12. Mention one method to find outliers.
13. What is p-value in hypothesis testing?
14. What is the Binomial Probability Formula?
15. Explain ANOVA and it's applications.

FLIP ROBO

MACHINE LEARNING

1

In a Linear Regression problem, ‘X’ is independent variable and ‘Y’ is dependent variable, where ‘X’ represents weight in pounds. If you convert the unit of ‘X’ to kilograms, then new coefficient of ‘X’ will

× coefficient of ‘X’ B) same as old coefficient of ‘X’

C) old coefficient of ‘X’ ÷

B) Identifying loan defaulters in a bank on the basis of previous years’ data of

ANSWERS;

MACHINE LEARNING

1. A
2. C

3. D
4. D
5. B
6. C
7. B
8. D
9. B,C,D
10. A,B,C

STATISTICS;
1.CENTRAL LIMIT THEROM : IT states that the distribution of sample means approximztes a normal distribution as the sample size gets larger, regardless of the population's distribution.
It is imp as it can analyse the large sample size more accurately.

2. sampling refers to sample containing analytic subset of a larger population/group.It helps researchers to conduct their studies with more manageable data and in a timely manner.

Types of sampling:
1.simple random sampling:it is ideal when we have population of similar entity.
2.stratified random sampling:
It divides the overall population into smaller groups which further share similar characteristics.

3.Type 1 error means rejecting the null hypothesis,when it is actually true..(false positive)
Type 2 error means not rejecting the null hypothesis when its is actually false.(false negative).
THEY are inversely proportional to each other.

4 normal distribution-
It is a probability distribution that is symmetric about the mean, showing that data near the mean is more frequent in occurrence than data from the mean.. It appears like bell shaped curve.The mean is zero and SD is 1.

5. Correlation is a measure that determines the degree to which two or mre random variables move in sequence. It can positively or negatively correlated.

Covariance is a statistical term that refers to a systematic relationship between two random variables in which a change in the other reflects a change in one variable.

6.Univariate data :when the data consists of only one variable ..for eg height of a person.we can find pattern of the data using central tendancy,pie charts, bar charts, frequwncy polygon.

BIVARIATE DATA: WHEN the analysis deals with two different variables. This helps to know thw cause and relation ship between two variables,,,, eg temp and ice cream in summers.

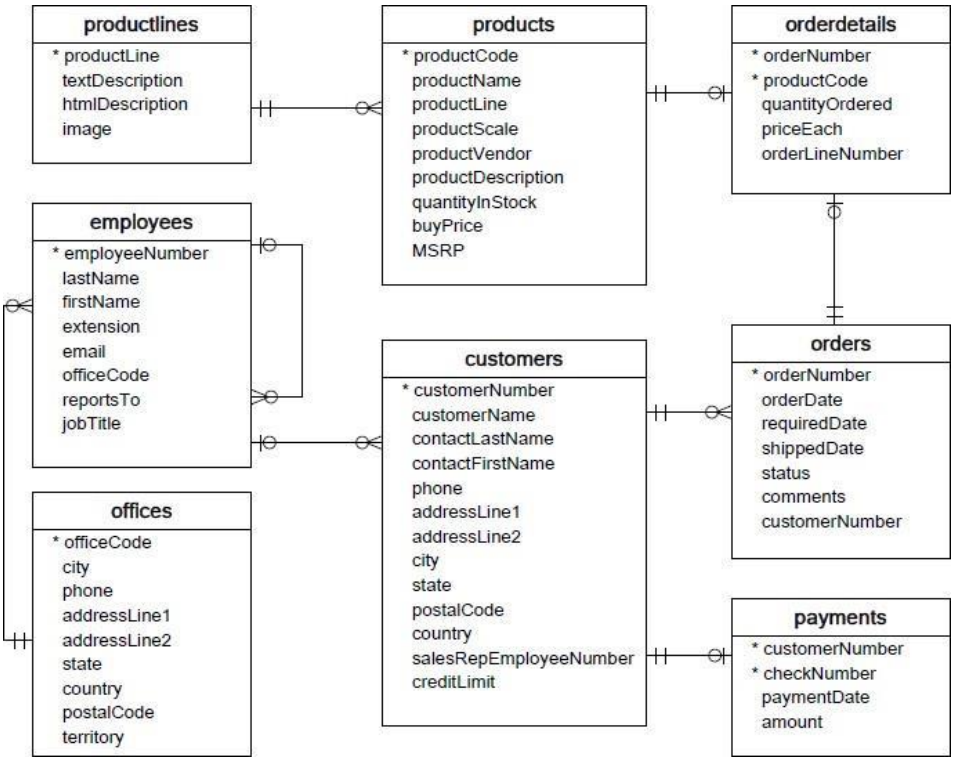
Multivariate data: when the data involves three or more variables.these variables are dependent on each other.

8. Hyopthesis testing:it is estimate the relationship between two variables,Ho -null hypothesis and alternative hypothesis H1.
In two tailed test ,the test sample is checked to be greater or less than a range of values in a two tailed test,implying that the critical distribution area is two -sided.

9.Quantitative data- are the measures of values or counts which are expressed in numbers.eg no of children in school.
Qualitative data- are the measures of type in categorical data..what is your occupation.

10 The range only takes in account the max and min observed values and ignores the data point between these extreme values.
The interquartile range is Q3-Q1…

11.Bell curve distribution
;it is visual representation of normal distribution.it is continuous probability disribution



- **Customers**: stores customer's data.
- **Products**: stores a list of scale model cars.
- **Product Lines**: stores a list of product line categories.
- **Orders**: stores sales orders placed by customers.
- **Order Details**: stores sales order line items for each sales order.
- **Payments**: stores payments made by customers based on their accounts.
- **Employees**: stores all employee information as well as the organization structure such as who reports to whom.
- **Offices**: stores sales office data.

QUESTIONS:

Write a SQL query to show average number of orders shipped in a day (use Orders table).
Write a SQL query to show average number of orders placed in a day.
Write a SQL query to show the product name with minimum MSRP (use Products table).