



TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS

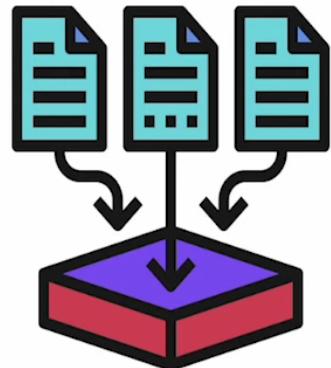


1

What are the steps involved in Data Analysis?



Understand
Problem



Collect
Data



Clean
Data



Explore
Data



Interpret
Data

ChatGPT

chatgpt.com/c/320eb461-be13-48d1-ba73-507633b53779

Gmail YouTube Maps All Bookmarks

New chat

ChatGPT 3.5

Today

Data Analysis Steps ...

Previous 7 Days

Data Analysis and Visualization

Previous 30 Days

Data Analyst Portfolio Advice

Improving SQL Table Schema

Cumulative Salary SQL Query.

Gaussian Features Assessment

April

Data Setup: Import, Load

LogReg Explained Succinctly

Upgrade plan
Get GPT-4, DALL-E, and more

Payel Roy

Message ChatGPT

ChatGPT can make mistakes. Check important info.

29°C Haze

Search

10:58 PM 12-May-24 ENG IN

ChatGPT 3.5

- Define the Problem or Research Question:** Clearly outline what you're trying to achieve through your analysis. Understand the purpose and context of the data analysis.
- Data Collection:** Gather the relevant data that you'll be analyzing. This can involve sourcing data from databases, surveys, experiments, or other sources. Ensuring data quality is crucial at this stage.
- Data Cleaning and Preparation:** This step involves cleaning the data to remove any errors, inconsistencies, or missing values. It may also involve transforming the data into a format suitable for analysis, such as aggregating, normalizing, or encoding categorical variables.
- Exploratory Data Analysis (EDA):** Explore the data to understand its underlying patterns, distributions, and relationships. This can involve techniques such as summary statistics, data visualization, and correlation analysis.
- Hypothesis Testing or Modeling:** Depending on the nature of the analysis, you may formulate hypotheses to test or develop predictive models to make forecasts or classifications. This step involves selecting appropriate statistical or machine learning techniques and applying them to the data.



TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



2

State the difference between Data Analysis and Data Mining.

Data Analysis

Data Analysis involves the process of cleaning, organising and using that data to produce meaningful insights.



Data Mining

Data Mining is used to search for hidden patterns in the data.





TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



3

What is Data Validation? Explain different types of data validation techniques.

Data validation is a process of ensuring that data is accurate, consistent, and meets the required quality standards.

A word cloud diagram centered around the words "DATA VALIDATION". Other words include "SYSTEM", "VERIFICATION", "CLEAN", "DATA", "RULES", "PROCESS", "CORRECT", "CHECK", "CONSISTENCY", "USEFUL", "ACCURACY", "SECURITY", "METHODS", and "APPLICATION".



TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



3

What is Data Validation? Explain different types of data validation techniques.

Types of Data Validation

FIELD LEVEL VALIDATION

FORM LEVEL VALIDATION

DATA SAVING VALIDATION

SEARCH CRITERIA VALIDATION



ChatGPT

chatgpt.com/c/321f7cb9-3b1b-4a0a-a53c-3d76d95f2bb6

Gmail YouTube Maps All Bookmarks

New chat

Today

Validation Techniques Overview

Previous 7 Days

Data Analysis Steps

Previous 30 Days

Data Analysis and Visualization Skills

Data Analyst Portfolio Advice

Improving SQL Table Schema

Cumulative Salary SQL Query

April

Gaussian Features Assessment

Data Setup: Import, Load

Upgrade plan
Get GPT-4, DALL-E, and more

Payel Roy

94°F Partly sunny

Search

PC

12:09 PM 14-May-24 PRE

ChatGPT 3.5

1. Field-Level Validation:

- Field-level validation occurs when the data entered into individual fields of a form or database record is checked for accuracy and consistency.
- Examples include:
 - Checking for the correct format of an email address.
 - Verifying that a phone number contains only numbers and follows a specific format.
 - Ensuring that a date is entered in a valid format and falls within a certain range.
- Field-level validation helps maintain data integrity by ensuring that only valid data is entered into each field.

2. Form-Level Validation:

- Form-level validation occurs when data entered across multiple fields of a form is validated as a whole.
- It involves checking relationships and constraints between fields.
- Examples include:

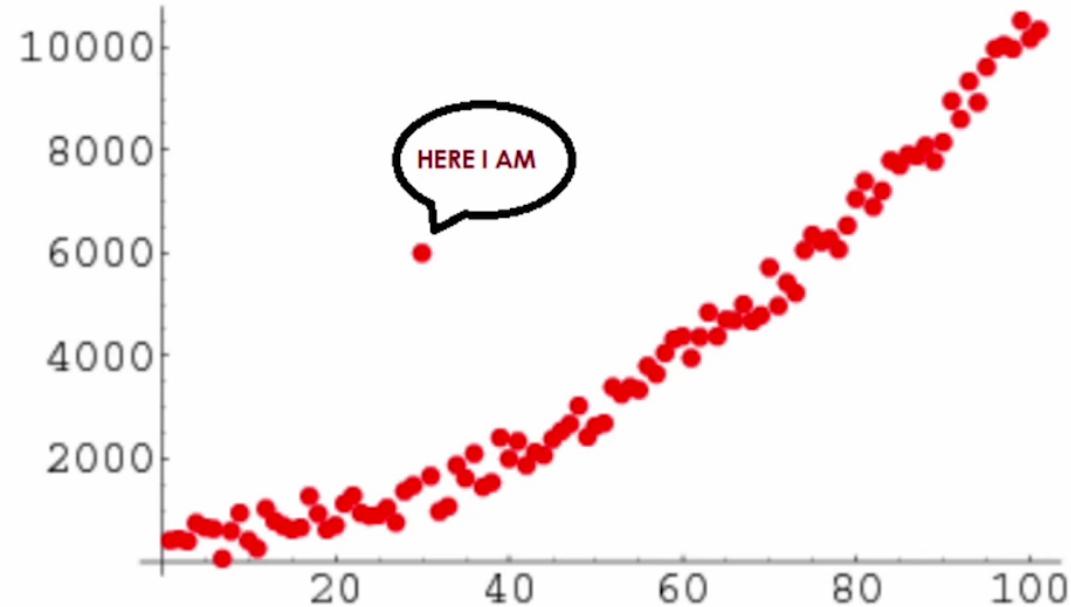


TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



4 What are outliers? How do you detect and treat outliers?

- Outlier is **an observation** in a given dataset that **lies far from the rest** of the observations.
- Outlier is **vastly larger or smaller** than the remaining values in the set.





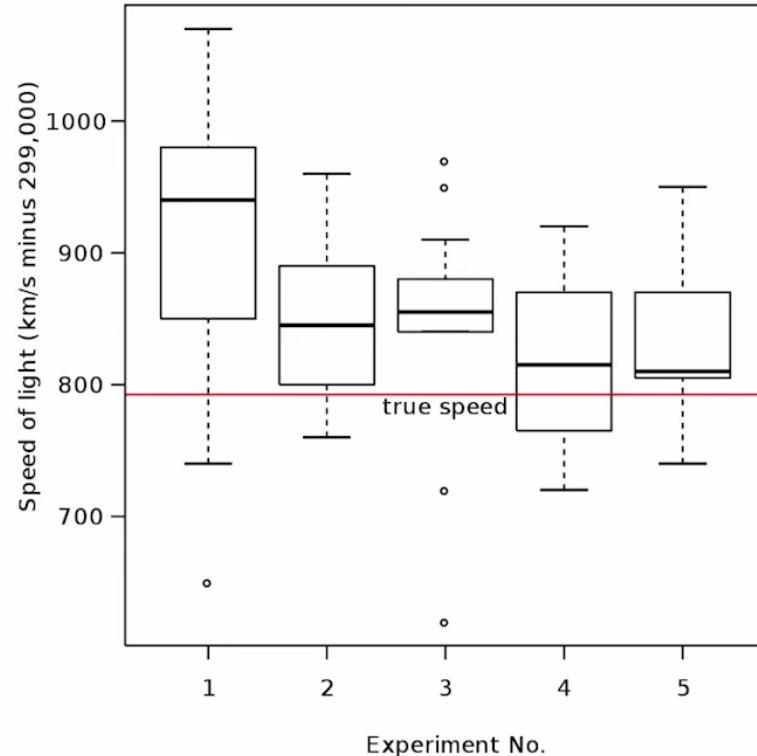
TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



4 What are outliers? How do you detect and treat outliers?

Techniques to Detect Outlier

Box Plot





TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS

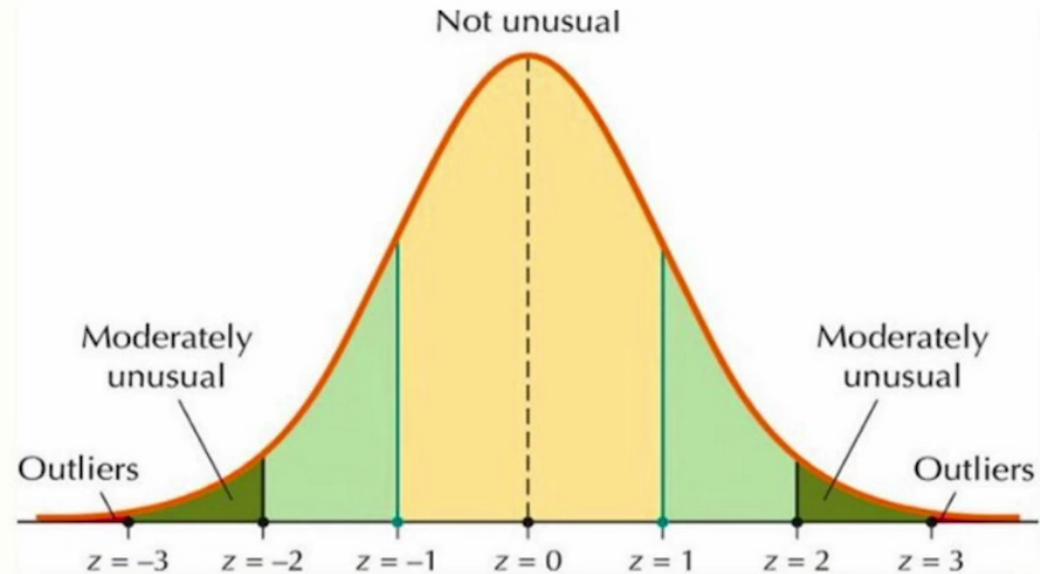


4 What are outliers? How do you detect and treat outliers?

Techniques to Detect Outlier

Z-Score

Detecting Outliers with z-Scores





TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



4 What are outliers? How do you detect and treat outliers?

Techniques to Treat Outlier

DROP OUTLIERS

ASSIGN A NEW VALUE

CAP OUTLIERS DATA

TRY NEW TRANSFORMATIONS



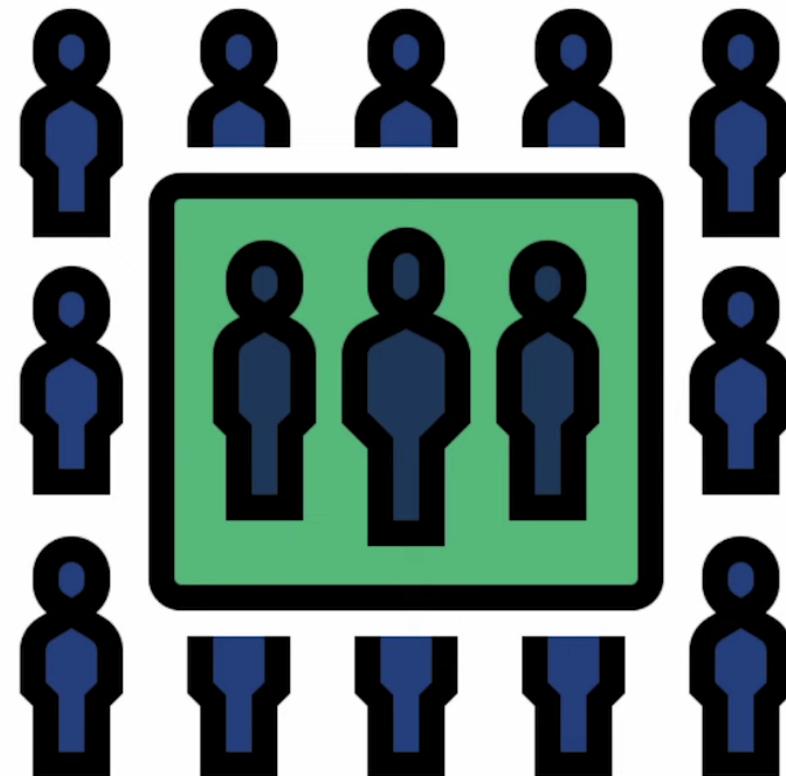
TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



5

What are different types of sampling techniques used in Data Analysis?

Sampling is a statistical method to understand a subset of data from an entire data set to estimate the characteristics of the whole population.





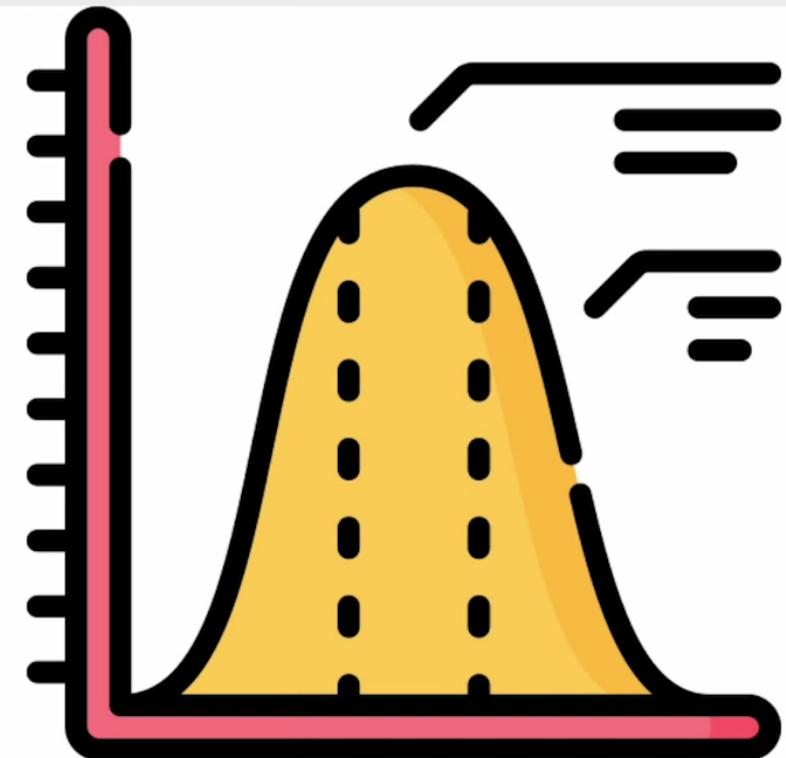
TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



7

What is Normal Distribution?

- Normal Distribution, also known as a Gaussian distribution or bell curve, **is a fundamental concept in statistics and data analysis.**
- It represents **a specific type of probability distribution** that is characterised by a symmetric, bell-shaped curve.





TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS

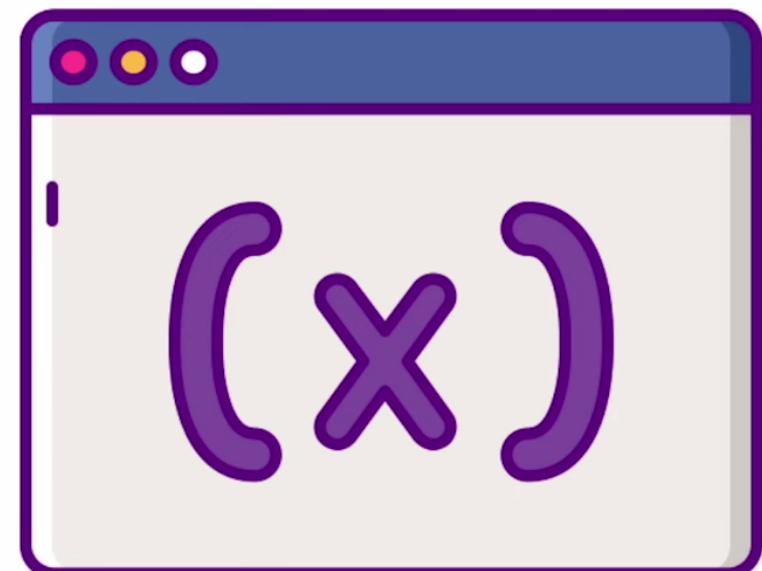


8

State the difference between Univariate, Bivariate and Multivariate Analysis.

Univariate

This type of data consists of only one variable. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes.





TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS

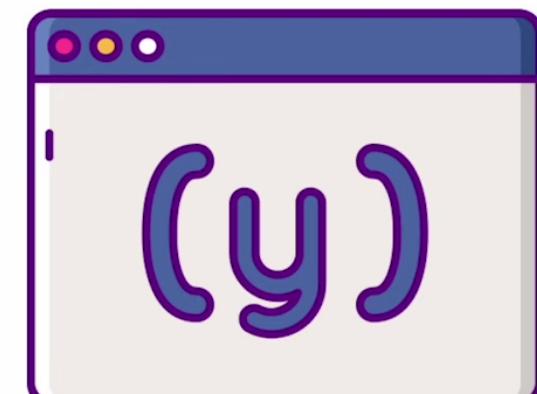
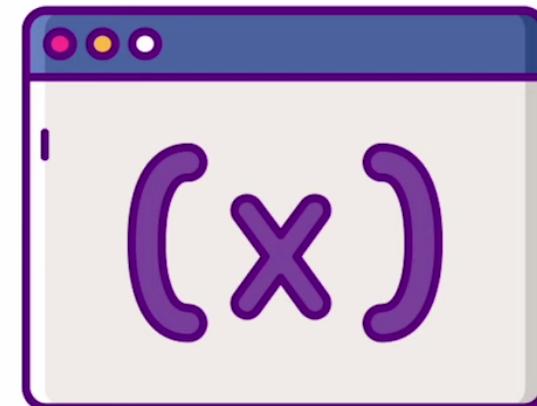


8

State the difference between Univariate, Bivariate and Multivariate Analysis.

Bivariate

This type of data involves two different variables. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables.





TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



8

State the difference between Univariate, Bivariate and Multivariate Analysis.

Multivariate

When the data involves three or more variables, it is categorised under multivariate.





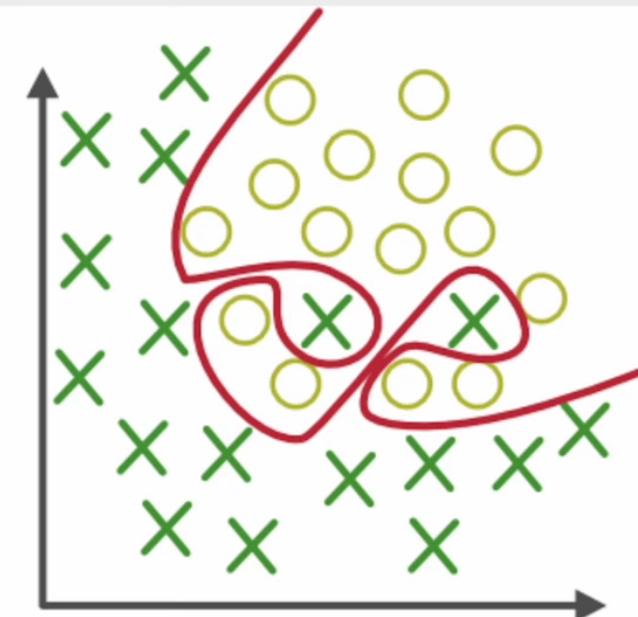
TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



9 What is the difference between overfitting and underfitting?

Underfitting A statistical model or a machine learning algorithm is said to have underfitting when **a model is too simple to capture data complexities.**

Overfitting A statistical model is said to be overfitted when the model does not make accurate predictions on testing data. When a model gets trained with so much data, it **starts learning from the noise and inaccurate data entries in our data set.**



Over-fitting
(forcefitting--too good to be true)



TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS



9

What is the difference between overfitting and underfitting?

Overfitting	Underfitting
Model trains the data too well using the training set	The model neither trains the data well nor can generalize the new data
The performance drops significantly over the test set	Performs poorly both on train and the test set
Happens when the model learns the noise and random fluctuations in the training dataset in detail	Happens when there is less data to build an accurate model and also when we try build a linear model with a non-linear data

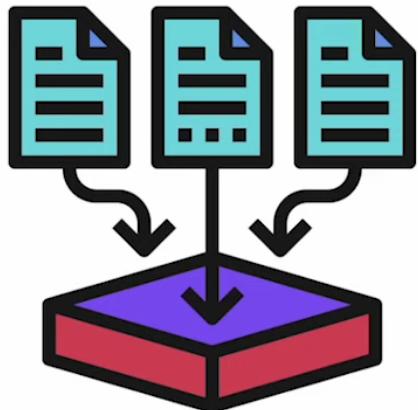


TOP 10 DATA ANALYSIS INTERVIEW QUESTIONS

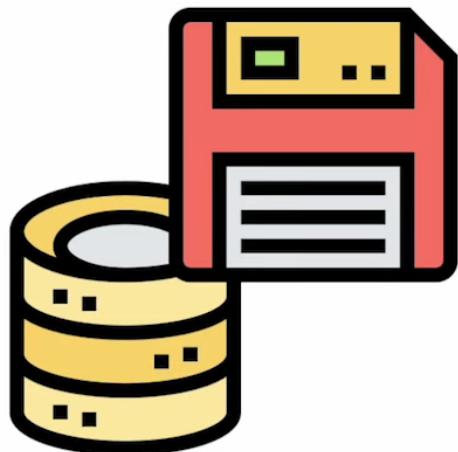


10 What are common problems encountered by data analysts?

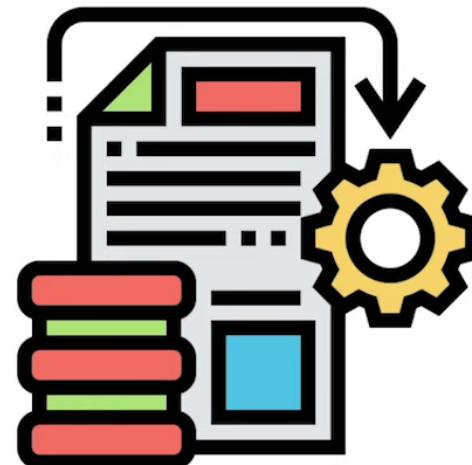
COMMON PROBLEMS FACED BY DATA ANALYSTS



Collecting Data



Storing Data



Processing Data



Data Quality and Governance