



Top 20 Data Analytics Interview Questions and Answers

Introduction

You want a job in analytics. You must prepare well. Interviewers test your basics and your practical knowledge. They ask about SQL, Excel, data cleaning, dashboards, and statistics. They also check how you solve problems with data.

This list of Data Analytics Interview Questions and Answers helps you revise important topics step by step. It keeps your preparation simple and focused. If you practice these Data Analytics Interview Questions and Answers, you can answer clearly and attend your interview with more confidence.

List of Data Analytics Interview Questions for Freshers

1. Define data validation
2. What are the primary steps in the data validation process?
3. Describe the process of data cleansing.
4. Describe the outlier.
5. What methods exist for identifying outliers? Describe various approaches to addressing it.
6. What is visualization of data?
7. Describe a hash table.
8. Describe KNN Imputation.
9. Explain collaborative filtering.
10. Define data wrangling
11. Explain Time Series Analysis
12. What distinguishes time-series forecasting from time-series analysis?

Data Analytics Technical Interview Questions and Answers for Freshers

Freshers must show clear basics in analytics. Interviewers test SQL, Excel, data cleaning, and simple statistics. This section helps you prepare common technical questions and answer them in a direct way.

1. Define data validation

The process of ensuring the data is high-quality and correct is known as data validation. This is accomplished by including checks in a report to ensure the consistency of the statistics throughout storage.

2. What are the primary steps in the data validation process?

When validating data, data analysts generally take two steps:

- Evaluating the data for incorrect values and screening it using various techniques.
- Assessing and validating the data to determine whether the values need to be used in the data.

3. Describe the process of data cleansing.

Data cleansing is essentially the act of detecting and then changing, replacing, or removing the wrong, incomplete, inaccurate, irrelevant, or missing sections of the data as needed.

It is also referred to as data cleaning, data scrubbing, or data wrangling. This essential component of data science guarantees accurate, consistent, and usable data.

4. Describe the outlier.

Values that deviate noticeably from the dataset's mean of defining characteristics are known as outliers. We can identify measurement variability or experimental error with the help of an outlier. Outliers can be classified as either univariate or multivariate.

5. What methods exist for identifying outliers? Describe various approaches to addressing it.

There are two ways to find outliers:

- *The box plot approach* defines an outlier as a number that is either above or below the $1.5 \times \text{IQR}$ (interquartile range), or above the top quartile (Q3) or below the bottom quartile (Q1).
- By using the *standard deviation approach*, a number that deviates from the mean $\pm (3 \times \text{standard deviation})$ is considered an outlier.

Take your Knowledge
Test Report

Check your Score



6. What is visualization of data?

Data and information represented graphically are referred to as data visualization. By using visual components like charts, graphs, and maps, data visualization tools make it simple for users to identify trends, outliers, and patterns in data.

With the help of this technology, data may be examined, processed more intelligently, and transformed into charts and graphs.

Check out: [Data Analytics Syllabus PDF](#)

7. Describe a hash table.

Generally speaking, hash tables are data structures that use associative storage of data. In this way, data is typically kept in an array format, enabling a unique index value for each data value.

A hash table creates an index into an array of slots using the hash technique, allowing us to obtain the desired data from there.

8. Describe KNN Imputation.

A dataset's missing values can be replaced with some believable values using an algorithm called K-Nearest Neighbors (KNN). KNN assumes that you can find a missing value by comparing it to other values that are closest to it.

Using libraries like Scikit-Learn is easier to execute and more accurate and efficient than utilizing mean, median, and mode.

9. Explain collaborative filtering.

One type of recommendation system that uses group behavior data to generate recommendations is called collaborative filtering. It is predicated on the idea that user groups that have exhibited a particular behavior in the past will continue to exhibit similar behavior going forward.

For example, give a movie a 5-star rating. The algorithm uses this information to suggest the same products to those groups.

10. Define data wrangling

The process of taking raw data and enhancing and cleaning it to make it easier to analyze and produce patterns and trends is known as data wrangling. This strategy dramatically increases the efficiency of all data used downstream.

11. Explain Time Series Analysis

A data analysis method called time series analysis examines a dataset across predetermined time intervals. It can be particularly helpful in situations where analyzing data over time can provide insightful information.

For instance, a time-series analysis of COVID-19 can reveal patterns in the manner in which the illness has proliferated.

12. What distinguishes time-series forecasting from time-series analysis?

Time series analysis is the process of examining data points gathered over an extended period to extract meaningful insights. On the other hand, time series forecasting entails making forecasts based on data analyzed over an extended period.

List of Data Analytics Interview Questions for Experienced

13. Explain multivariate, bivariate, and univariate analysis.

14. What Part Does Linear Regression Play in the Analysis of Statistical Data?

15. How Should Missing Data in a Dataset Be Handled?

16. What stages are there in every analytics project?
17. Which technical tools have you employed for purposes of analysis and presentation?
18. Which techniques work best for cleansing data?
19. What does exploratory data analysis, or EDA, mean?
20. What does a SQL subquery mean?

Data Analytics Technical Interview Questions and Answers for Experienced

Experienced candidates must prove practical knowledge. Interviewers check complex queries, reporting logic, and problem solving skills. This section covers technical questions that test real work experience and decision making ability.

13. Explain multivariate, bivariate, and univariate analysis.

When there is just one variable, it is called *univariate* analysis. This is the most basic type of analysis, similar to trends; causal or relationship analysis cannot be done in this manner.

For instance, the population growth of a particular city over the previous 50 years.

When there are two variables, the analysis is called *bivariate*. Both causal and connectional analyses are possible for you. This might be a gender-specific analysis of a city's population growth.

When there are three or more variables, the analysis is called *multivariate*. Here, you examine multiple variables at once to identify trends in multidimensional data. This could be the distribution of population growth by gender, income, kind of occupation, etc. in a particular city.

**Take your Knowledge
Test Report**

Check your Score



14. What Part Does Linear Regression Play in the Analysis of Statistical Data?

One of the most effective methods for statistical data analysis is linear regression. Establishing correlations between various variables is a useful tool for assessing company outcomes.

Take the following scenario: a credit card firm wants to determine what causes its customers to miss payments. By utilizing linear regression, the business can better target the traits of defaulters and enhance the clientele that it serves.

15. How Should Missing Data in a Dataset Be Handled?

In data analysis, there are two primary approaches to handling missing data.

- Making a well-informed estimate regarding the possible value of the missing data point is known as imputation.
- When there is little to no missing data and there seems to be some natural variation in the data that is available, it is employed.

- Removing the data is the alternative. This is typically carried out when random data is missing and it is impossible to infer plausible values for the missing values.

16. What stages are there in every analytics project?

Among the most fundamental interview questions for data analysts is this one. Any typical analytics project will involve the following steps:

- **Understanding the Problem:** Recognize the business issue, specify the objectives of the company, and make a profitable solution plan.
- **Data Collection:** Obtain pertinent information from a range of sources and additional data, following your priorities.
- **Data Cleansing:** To prepare the data for analysis, clean it up by removing unnecessary, redundant, and missing variables.
- **Data Exploring:** To analyze data, employ data mining strategies, predictive modeling, data visualization tools, and **business intelligence** tools.
- **Interpreting the Results:** Analyze the data to obtain insights, uncover hidden patterns, and predict future trends.

17. Which technical tools have you employed for purposes of analysis and presentation?

It is required of you as a data analyst to be familiar with the following tools for analysis and presentation. Here are a few well-known tools you should be aware of:

MySQL and MS SQL Server: To manipulate information kept in relational databases

Tableau, Microsoft Excel: For building dashboards and reporting

R, SPSS, and Python: For data modeling, exploratory analysis, and statistical analysis

Using Microsoft PowerPoint for the presentation, show the essential conclusions and end outcomes.

Check out: [Python Full Stack Training in Chennai](#)

18. Which techniques work best for cleansing data?

- Make a plan for cleaning up the data by *identifying the common error locations* and maintaining open lines of communication.
- Determine and *eliminate duplicates from the data* before modifying it. This will result in a simple and efficient procedure for data analysis.
- *Pay attention to how accurate the data is.* Establish required constraints, preserve data value types, and enable cross-field validation.
- To make the data less disorganized, *normalize it at the entry point.* You'll be able to guarantee that all data is consistent, which will reduce entry errors.

19. What does exploratory data analysis, or EDA, mean?

EDA, or exploratory data analysis, aids in improving data comprehension.

- It assists you in building your data's confidence to the extent that you're prepared to use a machine learning method.
- You can utilize it to further hone the feature variables you've chosen to use in the model development process.
- The data contains insights and hidden trends that you can find.

20. What does a SQL subquery mean?

A subquery in **SQL** is a query enclosed in another query. It is also known as a nested query or an inner query at times. The purpose of subqueries is to improve the data that the main query will be querying.

There are two varieties: linked and unrelated queries.

An example of a subquery that yields the name, email address, and phone number of a Chennai City employee is shown below.

```
SELECT name, email, phone
```

```
FROM employee
```

```
WHERE emp_id IN (
```

```
SELECT emp_id
```

```
FROM employee
```

```
WHERE city = 'Chennai');
```

Conclusion

These Top 20 **Data Analytics Interview Questions and Answers** give you a clear idea of what companies expect. Strong basics matter. Clear thinking matters more. Practice with real datasets. Focus on solving problems, not just giving definitions.

If you want serious preparation, choose [Data Analytics Training in Chennai](#).
Learn concepts in a simple way. Work on practical tasks. Get interview
guidance. Build skills that companies look for. Start your preparation today.