

Some Frequently asked SQL Interview Questions for Data Analysts:

1. Write a SQL query to find the second highest salary from the table emp. (Column name – id, salary)
2. Write a SQL query to find the numbers which consecutively occurs 3 times. (Column name – id, numbers)
3. Write a SQL query to find the days when temperature was higher than its previous dates. (Column name – Days, Temp)
4. Write a SQL query to delete Duplicate rows in a table.
5. Write a SQL query for cumulative sum of salary of each employee from Jan to July. (Column name – Emp_id, Month, Salary).
6. Write a SQL query to display year on year growth for each product. (Column name – transaction_id, Product_id, transaction_date, spend). Output will have year, product_id & yoy_growth.
7. Write a SQL query to find rolling average of posts on daily basis for each user_id. (Column_name – user_id, date, post_count). Round up the average upto two decimal places.
8. Write a SQL query to get emp id and department for each department who recently joined the organization and still in working. (column - emp id, first name, last name, date of join, date of exit , department.)

9. How many rows will come in outputs of Left, Right, Inner and outer join from two tables having duplicate rows.

For Example:-

Left table A-

Column

1

1

1

2

2

3

4

5

Right table B-

Column

1

1

2

2

2

3

3

3

4

10. Write a query to get mean, median and mode for earning? (Column – Emp_id, salary)

Question 11.

Given:

Table X:

Columns: ids with values 1, 1, 1, 1

Table Y:

Columns: ids with values 1, 1, 1, 1, 1, 1, 1, 1

Task: Determine the count of rows in the output of the following queries:

Select * From X join Y on X.ids != Y.ids

Select * From X left join Y on X.ids != Y.ids

Select * From X right join Y on X.ids != Y.ids

Select * From X full outer join Y on X.ids != Y.ids

Verbally asked SQL interview questions -

Explain order of execution of SQL.

What is difference between where and having?

What is the use of group by?

Explain all types of joins in SQL?

What are triggers in SQL?

What is stored procedure in SQL

Explain all types of window functions? (Mainly rank, row_num, dense_rank, lead & lag)

What is difference between Delete and Truncate?

What is difference between DML, DDL and DCL?

What are aggregate function and when do we use them? explain with few example.

Which is faster between CTE and Subquery?

What are constraints and types of Constraints?

Types of Keys?

Different types of Operators ?

Difference between Group By and Where?

What are Views?

What are different types of constraints?

What is difference between varchar and nvarchar?

Similar for char and nchar?

What are index and their types?

What is an index? Explain its different types.

List the different types of relationships in SQL.

Differentiate between UNION and UNION ALL.

How many types of clauses in SQL ?

What is the difference between UNION and UNION ALL in SQL?

What are the various types of relationships in SQL?

Difference between Primary Key and Secondary Key?

What is the difference between where and having?

Find the second highest salary of an employee?

Write retention query in SQL?

Write year-on-year growth in SQL?

Write a query for cumulative sum in SQL?

Difference between Function and Store procedure ?

Do we use variable in views?

What are the limitations of views?

Some Frequently asked Excel Interview Questions for Data Analysts:

Data Handling & Import

- How do you import data from external sources into Excel?
- How do you use the Data Form feature in Excel for data entry and editing?
- How do you use the Consolidate feature in Excel to combine data from multiple worksheets or workbooks?

Pivot Tables & Analysis Tools

- How do you create a pivot table in Excel and what is its purpose?
- How do you use the GETPIVOTDATA function in Excel to extract data from a pivot table?
- How do you use the GROUP BY feature in Excel to summarize data?
- How do you use the Data Analysis Toolpak in Excel for statistical analysis?

Visualization, Formatting & Validation

- What are the different types of charts available in Excel and how do you create them?
- How do you filter and sort data in Excel?
- How do you use conditional formatting in Excel?
- What are Excel tables and how do you use them for data analysis?
- What is the purpose of using data validation in Excel?
- How do you use the Data Validation feature in Excel to create drop-down lists and input restrictions?

Functions & Formulas: Basics to Advanced

- How do you use VLOOKUP, HLOOKUP, and XLOOKUP functions in Excel?
- How do you use the IF, COUNTIF, SUMIF, AVERAGE, MEDIAN, and MODE functions in Excel?
- How do you use CONCATENATE, TEXTJOIN, LEFT, RIGHT, MID, DATE, TIME, NOW, ROUND, ROUNDUP, ROUNDDOWN, INDEX, MATCH, CHOOSE, SWITCH, IFS, RAND, RANDBETWEEN, TRANSPOSE, SUMPRODUCT, TEXT, LEN, TRIM, SUBSTITUTE, NETWORKDAYS, WORKDAY, DATEDIF, WEEKDAY, MONTH, PROPER, UPPER, LOWER, PIVOTDATA, CUBESET, and CUBEMEMBER functions in Excel?

Optimization & Add-ons

- How do you create and use macros in Excel?
- How do you use the Solver add-in in Excel for optimization problems?

Some Frequently asked Pandas(Python) Interview Questions for Data Analysts:

1. Find all unique employee names who work in more than one department.

Sample DataFrame:

```
df = pd.DataFrame({'EmployeeName': ['John Doe', 'Jane Smith', 'Alice Johnson', 'John Doe'], 'Department': ['Sales', 'Marketing', 'Sales', 'Marketing']})
```

2. Calculate the monthly average sales for each product. Assume sales data is daily.

Sample DataFrame:

```
df = pd.DataFrame({'Date': pd.date_range(start='2023-01-01',  
end='2023-03-31', freq='D'), 'Product': np.random.choice(['ProductA',  
'ProductB'], 90), 'Sales': np.random.randint(100, 500, 90)})
```

3. Identify the top 3 employees with the highest sales in each quarter.

Sample DataFrame:

```
df = pd.DataFrame({'Employee': ['John', 'Jane', 'Doe', 'Smith', 'Alice'],  
'Quarter': ['Q1', 'Q1', 'Q2', 'Q2', 'Q3'], 'Sales': [200, 150, 300, 250, 400]})
```

4. Analyze the attendance records to find employees with more than 95% attendance throughout the year.

Sample DataFrame:

```
df = pd.DataFrame({'Employee': ['John', 'Jane', 'Doe'], 'TotalDays': [365,  
365, 365], 'DaysAttended': [365, 350, 360]})
```

5. Calculate the monthly customer retention rate based on the transaction logs.

Sample DataFrame:

```
df = pd.DataFrame({'Month': ['Jan', 'Feb', 'Mar', 'Jan', 'Feb', 'Mar'],  
'CustomerID': [1, 1, 1, 2, 2, 3], 'TransactionCount': [1, 2, 1, 3, 2, 1]})
```

6. Determine the average time employees spent on projects, assuming you have start and end dates for each project participation.

Sample DataFrame:

```
df = pd.DataFrame({'Employee': ['John', 'Jane', 'Doe'], 'ProjectStart':  
pd.to_datetime(['2023-01-01', '2023-02-15', '2023-03-01']), 'ProjectEnd':  
pd.to_datetime(['2023-01-31', '2023-03-15', '2023-04-01'])})
```

7. Compute the month-on-month growth rate in sales for each product, highlighting products with more than 10% growth for consecutive months.

Sample DataFrame:

```
df = pd.DataFrame({'Month': ['Jan', 'Feb', 'Mar', 'Jan', 'Feb', 'Mar'],  
'Product': ['A', 'A', 'A', 'B', 'B', 'B'], 'Sales': [200, 220, 240, 150, 165, 180]})
```

8. Identify the time of day (morning, afternoon, evening) when sales peak for each category of products.

Sample DataFrame:

```
df = pd.DataFrame({'Category': ['Electronics', 'Clothing', 'Electronics',  
'Clothing'], 'TimeOfDay': ['Morning', 'Afternoon', 'Evening', 'Morning'],  
'Sales': [300, 150, 500, 200]})
```

9. Evaluate the distribution of workload among employees based on the number of tasks assigned in the last six months.

Sample DataFrame:

```
df = pd.DataFrame({'Employee': ['John', 'Jane', 'Doe'], 'TasksAssigned':  
[20, 25, 15]})
```

10. Calculate the profit margin for each product category based on revenue and cost data.

Sample DataFrame:

```
df = pd.DataFrame({'Category': ['Electronics', 'Clothing'], 'Revenue': [1000, 500], 'Cost': [700, 300]})
```

Some Verbally asked Pandas(Python) Interview Questions for Data Analysts:

1. What are the differences between lists and tuples in Python, and how does this distinction relate to Pandas operations?
2. What is a DataFrame in Pandas, and how does it differ from a Series?
3. Can you explain how to handle missing data in Pandas, including the difference between 'fillna()' and 'dropna()'?
4. Describe the process of renaming a column in a Pandas DataFrame.
5. What is the purpose of the 'groupby' function in Pandas, and provide an example of its usage?
6. How can you merge two DataFrames in Pandas, and what are the different types of joins available?
7. Explain the purpose of the 'apply' function in Pandas, and give an example of when you might use it.
8. What is the difference between 'loc' and 'iloc' in Pandas, and when would you use each?
9. Explain the difference between a join and a merge in Pandas with examples.
10. How do you remove duplicates from a DataFrame in Pandas?
11. How do you join two DataFrames on multiple columns in Pandas?
12. Discuss the use of the 'pivot_table' method in Pandas and provide an example scenario where it is useful.

13. Explain the difference between the 'agg' and 'transform' methods in groupby operations.
14. Describe a method to handle large datasets in Pandas that do not fit into memory.
15. How can you convert categorical data into 'dummy' or 'indicator' variables in Pandas?
16. What is the difference between 'concat' and 'append' methods in Pandas?
17. How would you use the 'melt' function in Pandas, and what is its purpose?
18. Describe how you would perform a vectorized operation on DataFrame columns.
19. How can you set a column as the index of a DataFrame, and why would you want to do this?
20. Explain how to sort a DataFrame by multiple columns in Pandas.
21. How do you deal with time series data in Pandas, and what functionalities support its manipulation?
22. What are some ways to optimize a Pandas DataFrame for better performance?
23. Explain the purpose of the 'crosstab' function in Pandas and provide a use case.
24. How can you reshape a DataFrame in Pandas using the 'stack' and 'unstack' methods?
25. Describe how to use the 'query' method in Pandas and why it might be more efficient than other methods.
26. Discuss the importance of vectorization in Pandas and provide an example of a non-vectorized operation versus a vectorized one.
27. How would you export a DataFrame to a CSV file, and what are some common parameters you might adjust?
28. Explain the use of multi-indexing in Pandas and provide a scenario where it's beneficial.

29. How can you handle different timezones in Pandas?

Some Business Case Study & Guesstimate Questions:

1. Picture this: Netflix has seen a decline in the number of new subscribers from urban areas but an uptick from rural regions. Frame a strategy to leverage this trend and ensure consistent growth across both demographics.

2. Uber is contemplating launching a new service named "Uber Tours" in major tourist cities. This would be a guided tour using an Uber. Define the critical metrics you'd consider to gauge the initial success of this service.

3. A global coffee chain, "CoffeeHub," notices peak sales during winters but a significant dip during summers in tropical countries. Outline a structured methodology on how you would delve into and address this seasonality issue.

4. Estimate the number of cups of coffee consumed daily in Bangalore City.

5. Picture a scenario where a leading online music streaming platform, "BeatStream," grapples with competition from piracy websites. Craft a business model or a feature aimed at curtailing the impact of these unauthorized sources on BeatStream's revenue.

6. Zoom, the video conferencing software, is mulling over diversifying into the fitness sector. Which key indicators would be on your radar to judge the triumph or slip-up of this venture?

7. The government is in the blueprint phase of a new express highway between two pivotal cities. With access to pertinent data, how would you project the potential daily traffic on this route two years after its inauguration?

8. Electronic books (e-books) are on an uptrend, yet physical bookstores chart a downswing. Propose an avant-garde strategy for a conventional bookstore chain to adapt to this digital wave and yet ensure sustained growth.