Technical Mock Interview

Questions and Answers

Topics: Excel, Power BI, MySQL, Python, Data Analysis, Machine Learning

---

Entri Coding Tamil

Excel for Data Analysis (Set 1 - Beginner)

Q1: What is a cell reference in Excel?
A: A cell reference identifies a cell's location in the worksheet (e.g., A1). It can be relative, absolute, or a combination of both.

Q2: How do you apply a filter to data in Excel?
A: Select the header row and click on "Filter" under the Data tab to enable drop-down filters on each column.

Q3: What is conditional formatting?
A: It highlights cells based on specified rules like values greater than a threshold or duplicates using colors and styles.

Q4: What does the IF function do in Excel?
A: It returns one value if a condition is TRUE and another if it is FALSE. Syntax: =IF(A1>10, "Yes", "No")

Q5: How do you create a chart in Excel?
A: Select the data and choose a chart type from the "Insert" tab (e.g., column, pie, line).

Power BI (Set 1 - Beginner)

Q6: What is Power BI used for?
A: Power BI is a business intelligence tool used for data visualization, reporting, and dashboard creation.

Q7: What are visualizations in Power BI?
A: Visualizations are graphical representations like bar charts, line graphs, and pie charts used to display data insights.

Q8: How do you import data into Power BI?
A: Use the "Get Data" option to import from Excel, SQL Server, web, or other sources.

Q9: What is a data model in Power BI?
A: A data model organizes related data tables and defines relationships between them using keys.

Q10: What is the use of the Fields pane in Power BI?
A: It lists all the tables and fields available for building visuals, filters, and calculations.

MySQL (Set 1 - Beginner)

Q11: What is SQL used for?
A: SQL (Structured Query Language) is used to store, retrieve, and manipulate data in relational databases.

Q12: How do you retrieve all data from a table?
A: Use the command: SELECT * FROM table_name;

Q13: What is a primary key?
A: A primary key uniquely identifies each record in a table and cannot have NULL values.

Q14: How do you filter records in SQL?
A: Use the WHERE clause. Example: SELECT * FROM employees WHERE department = 'HR';

Q15: What does the ORDER BY clause do?
A: It sorts the result set by one or more columns in ascending (ASC) or descending (DESC) order.

Core Python (Set 1 - Beginner)

Q16: What is a variable in Python?
A: A variable is a name that stores a value. Example: x = 5

Q17: What is the difference between a list and a tuple?
A: A list is mutable (can be changed), whereas a tuple is immutable (cannot be changed).

Q18: How do you write a comment in Python?
A: Use the # symbol. Example: # This is a comment

Q19: What is a function in Python?
A: A function is a reusable block of code defined using def keyword. Example:

```
def greet():
    print("Hello")
```

Q20: What is indentation in Python and why is it important?
A: Indentation defines blocks of code and is required in Python for loops, functions, and conditionals.

Python for Data Analysis (Set 1 - Beginner)

Q21: What is Pandas in Python?
A: Pandas is a data analysis library used to work with structured data using DataFrames and Series.

Q22: How do you create a DataFrame in Pandas?
A: Using pd.DataFrame() with a dictionary or 2D array. Example:

```
import pandas as pd
df = pd.DataFrame({'Name': ['A', 'B'], 'Age': [22, 25]})
```

Q23: What does df.head() do?
A: It returns the first five rows of the DataFrame by default.

Q24: How do you read a CSV file in Pandas?
A: Use pd.read_csv('filename.csv')

Q25: What function is used to get basic statistics in Pandas?
A: Use df.describe() to get count, mean, std, min, max, etc.

Machine Learning Algorithms (Set 1 - Beginner)

Q26: What is machine learning?
A: Machine Learning is a branch of AI that allows systems to learn from data and make predictions without being explicitly programmed.

Q27: What is the difference between training and testing data?
A: Training data is used to build the model, while testing data is used to evaluate its performance.

Q28: What is a feature in ML?
A: A feature is an input variable used by the model to make predictions.

Q29: Name two types of supervised learning algorithms.
A: Examples include Linear Regression and Decision Trees.

Q30: What is classification in ML?
A: Classification is a task where the model assigns data to categories (e.g., spam or not spam).

Excel for Data Analysis (Set 2 - Intermediate)

Q31: How can you remove duplicate entries in Excel?
A: Select the range, go to Data tab, and click “Remove Duplicates.”

Q32: What is the difference between VLOOKUP and INDEX-MATCH?
A: VLOOKUP searches in the leftmost column, while INDEX-MATCH allows lookup in any direction and is more flexible.

Q33: What does the TEXT function do in Excel?
A: It formats numbers and dates as text using a specified format. Example: TEXT(A1, "dd-mm-yyyy")

Q34: What are pivot tables used for?
A: To summarize, group, and analyze large datasets dynamically.

Q35: What is a named range?
A: It assigns a name to a cell or range of cells for easier reference in formulas.

Power BI (Set 2 - Intermediate)

Q36: What is DAX in Power BI?
A: DAX (Data Analysis Expressions) is a formula language used to create custom calculations in Power BI.

Q37: What is the difference between calculated column and measure in Power BI?
A: Calculated columns are stored in tables and recalculated row-wise; measures are dynamic and calculated at runtime.

Q38: How do you create a relationship between tables?
A: Go to “Model” view and drag a field from one table to a matching field in another.

Q39: What are slicers in Power BI?
A: Slicers are visual filters that allow users to filter data interactively.

Q40: What is a Power BI dashboard?
A: A dashboard is a single-page summary that combines visuals from multiple reports.

MySQL (Set 2 - Intermediate)

Q41: What is a JOIN in SQL?
A: A JOIN is used to combine rows from two or more tables based on a related column.

Q42: What is the difference between INNER JOIN and LEFT JOIN?
A: INNER JOIN returns only matching rows; LEFT JOIN returns all from the left table and matched from the right.

Q43: How do you count distinct values in a column?
A: Use: SELECT COUNT(DISTINCT column_name) FROM table_name;

Q44: What is a subquery?
A: A query within another query, used for complex filters or aggregations.

Q45: What are indexes in SQL?
A: Indexes improve the speed of data retrieval operations on a table.

Core Python (Set 2 - Intermediate)

Q46: What is list comprehension in Python?
A: It is a concise way to create lists. Example: [x for x in range(5)]

Q47: What is a lambda function?
A: An anonymous function defined using lambda keyword. Example: lambda x: x + 1

Q48: How does exception handling work in Python?
A: Use try, except, finally blocks to handle runtime errors gracefully.

Q49: What are Python dictionaries used for?
A: To store key-value pairs. Example: {'name': 'Alice', 'age': 25}

Q50: What is the difference between is and == in Python?
A: is checks object identity, == checks value equality.

Python for Data Analysis (Set 2 - Intermediate)

Q51: What is the difference between loc[] and iloc[] in Pandas?
A: loc[] uses labels/index names; iloc[] uses integer-based positions.

Q52: How do you group data in Pandas?
A: Use df.groupby('column_name') for aggregation.

Q53: What is broadcasting in NumPy?
A: A technique that allows NumPy to perform operations on arrays of different shapes.

Q54: How do you handle missing values in Pandas?
A: Use df.isnull(), df.dropna(), or df.fillna().

Q55: How do you merge two DataFrames in Pandas?
A: Use pd.merge(df1, df2, on='key')

Machine Learning Algorithms (Set 2 - Intermediate)

Q56: What is overfitting in ML?
A: Overfitting is when a model performs well on training data but poorly on unseen data.

Q57: What is cross-validation?
A: A technique to evaluate model performance by splitting data into training and validation sets multiple times.

Q58: What is the difference between classification and regression?
A: Classification predicts categories, regression predicts continuous values.

Q59: What are hyperparameters in ML?
A: Settings used to control the learning process (e.g., learning rate, depth of tree).

Q60: What is a confusion matrix?
A: A table used to evaluate the performance of a classification model by comparing predicted vs actual values.

Technical Mock Interview Questions and Answers

Program: Data Science and Machine Learning
Topics: Excel, Power BI, MySQL, Python, Data Analysis, Machine Learning
Platform: AI-Powered Mock Interview Tool - ZappyVue

Advanced Questions (Set 3)

Excel for Data Analysis (Set 3 – Advanced)

Q61: How can you perform dynamic data consolidation in Excel across multiple sheets?
A: Use the INDIRECT() function combined with SUMPRODUCT() or INDEX() to reference dynamic sheet names and ranges.

Q62: What is Power Query and how is it used in Excel?
A: Power Query is a data transformation tool that helps in importing, cleaning, reshaping, and loading data without manual steps.

Q63: How do you create a Pivot Table with slicers connected to multiple tables?
A: Use a data model in Excel by adding relationships between tables and inserting slicers linked via the model.

Q64: How do you use array formulas in Excel?
A: Array formulas perform multiple calculations in a single cell or across a range using functions like SEQUENCE(), MMULT(), or INDEX(MATCH()) combinations.

Q65: What is the difference between Excel Tables and Named Ranges?
A: Tables are structured references with dynamic behavior (like autofill and filters); named ranges are static references to cells.

Power BI (Set 3 – Advanced)

Q66: How do you create calculated tables in Power BI and where are they useful?
A: Use New Table in the Modeling tab with DAX expressions like SUMMARIZE() or ADDCOLUMNS() to create virtual tables used in slicers or advanced visuals.

Q67: What is row-level security (RLS) and how is it implemented in Power BI?
A: RLS restricts data access for users. It is implemented using roles with DAX filters in the modeling view.

Q68: Explain the context transition in DAX with an example.
A: Context transition happens when row context changes to filter context—e.g., using CALCULATE() on a row context like inside a SUMX() iterator.

Q69: How does Power BI handle large datasets?
A: Through data reduction techniques such as aggregations, composite models, incremental refresh, and DirectQuery mode.

Q70: What are bookmarks in Power BI and their use cases?
A: Bookmarks capture the current state of a report (filters, visuals, navigation) for storytelling, toggling views, and creating dynamic reports.

MySQL (Set 3 – Advanced)

Q71: How do you optimize a query in MySQL with a slow performance issue?
A: Use indexing, avoid SELECT *, use EXPLAIN, ensure WHERE clause matches index columns, and limit nested queries.

Q72: What is a subquery vs a join, and when would you use one over the other?
A: Subqueries are nested queries used in SELECT, WHERE, or FROM clauses. Joins combine rows from two tables. Use joins for performance; subqueries for filtering.

Q73: How can you perform pivoting or transposing rows to columns in MySQL?
A: Use CASE statements with GROUP BY or use PIVOT if supported, else rely on application-level pivoting.

Q74: Explain normalization and denormalization in database design.
A: Normalization organizes data to reduce redundancy; denormalization adds redundancy for read performance.

Q75: What are stored procedures and when should you use them?
A: Stored procedures are precompiled SQL code blocks used for modularity, performance, and reuse in transactions.

📚 Core Python (Set 3 – Advanced)

Q76: What is a generator and how is it different from a normal function?
A: Generators use yield to return values lazily, saving memory. Normal functions return all results at once.

Q77: Explain the use of decorators in Python.
A: Decorators are functions that modify the behavior of other functions using @decorator_name syntax.

Q78: What are metaclasses in Python?
A: Metaclasses are classes of classes that control class creation. Useful in frameworks and for modifying class behavior dynamically.

Q79: How is memory managed in Python?
A: Python uses reference counting and a cyclic garbage collector for automatic memory management.

Q80: Explain multithreading vs multiprocessing in Python.
A: Multithreading runs threads in the same process (limited by GIL); multiprocessing uses separate processes for true parallelism.

Python for Data Analysis (Set 3 – Advanced)

Q81: What is the difference between .apply() and .map() in Pandas?
A: .map() is used on Series and modifies each element; .apply() works on Series and DataFrames, enabling row/column-wise operations.

Q82: How do you handle missing data using advanced imputation techniques?
A: Use methods like KNN Imputation (sklearn.impute.KNNImputer) or regression-based imputation based on correlated features.

Q83: What is chaining in Pandas and why should it be avoided?
A: Chaining refers to accessing/setting values with multiple indexing operations; it can cause SettingWithCopyWarning and unpredictable behavior.

Q84: How can you optimize a large DataFrame for performance?
A: Use categorical data types, downcast numeric types, use chunksize while reading, and drop unused columns early.

Q85: How do you merge two DataFrames on multiple keys and handle missing matches?
A: Use pd.merge(df1, df2, on=['key1','key2'], how='outer') and use indicator=True to track matched/unmatched rows.

Machine Learning Algorithms (Set 3 – Advanced)

Q86: What is regularization in ML and name two types?
A: Regularization prevents overfitting by penalizing large coefficients. Types: L1 (Lasso) and L2 (Ridge).

Q87: Explain how decision trees are prone to overfitting.
A: Decision trees can create very deep trees capturing noise in data. Techniques like pruning, max depth, or ensemble methods help prevent this.

Q88: What is the difference between bagging and boosting?
A: Bagging builds models in parallel (e.g., Random Forest); boosting builds sequentially where each model corrects the errors of the previous (e.g., XGBoost).

Q89: How does SMOTE work for handling class imbalance?

A: SMOTE (Synthetic Minority Over-sampling Technique) creates synthetic samples for minority class by interpolating between existing examples.

Q90: What are precision-recall tradeoffs and when are they important?

A: Precision-recall tradeoffs are crucial in imbalanced datasets. High precision means few false positives, high recall means few false negatives. Important in domains like fraud detection.
