Python Interview Q & A

Data Analyst (Fresher Role)

# Technical Questions (Basics to Intermediate)

1. What is Python and why is it popular in data analysis?

→ High-level, easy-to-learn, supports data libraries (Pandas, NumPy).

2. Difference between list, tuple, and set?

→ List = mutable, Tuple = immutable, Set = unique items.

3. How do you declare a list in Python?

→ my\_list = [1,2,3]

4. What is a dictionary in Python?

→ Key-value pairs, e.g., {"name":"Ali", "age":25}.

5. How to get length of a list?

→ len(my\_list)

6. How do you slice a list?

→ my\_list[1:4]

7. Difference between append() and extend() in lists?

→ Append = adds one item, Extend = adds multiple.

8. What is indentation in Python?

→ Space/tab that defines code blocks.

9. Explain mutable vs immutable with examples.

→ List = mutable, String = immutable.

10. How to reverse a string?

→ "hello"[::-1]

11. Difference between == and is?

→ == checks value, is checks memory location.

12. How to handle errors in Python?

→ Using try-except block.

13. Explain list comprehension.

→ Short way to create lists: [x\*2 for x in range(5)].

14. How to remove duplicates from a list?

→ list(set(my\_list))

15. Explain for loop with example.

→ for i in range(5): print(i)

16. What is a function in Python?

→ def add(x,y): return x+y

17. Explain lambda function.

→ lambda x: x\*2

18. What are Python modules?

→ Pre-written code files you can import, e.g., import math.

19. How to read a CSV file in Python?

→ import pandas as pd; df = pd.read\_csv("file.csv")

20. Difference between NumPy and Pandas.

→ NumPy = numeric arrays, Pandas = tabular data.

21. How to check null values in Pandas?

→ df.isnull().sum()

22. How to filter rows in Pandas?

→ df[df["age"] > 25]

23. How to get unique values in a column?

→ df["col"].unique()

24. How to merge two DataFrames?

→ pd.merge(df1, df2, on="id")

25. Explain difference between loc and iloc.

→ loc = label-based, iloc = index-based.

26. What is groupby in Pandas?

→ Aggregates data by categories.

27. How to sort DataFrame by column?

→ df.sort\_values("col")

28. What is a Series in Pandas?

→ One-dimensional labeled data structure.

29. Explain Python’s with open() usage.

→ Context manager for file handling.

30. How to install external packages?

→ pip install package\_name.

31. What is the difference between shallow copy and deep copy?

→ Shallow = references same objects, Deep = full clone.

32. Explain the use of map() in Python.

→ Applies a function to iterable: map(str.upper, names).

33. Difference between any() and all().

→ any = at least one True, all = all True.

34. How do you drop a column in Pandas?

→ df.drop("col", axis=1)

35. What is typecasting in Python?

→ Converting data types, e.g., int("5").

36. How to generate random numbers in Python?

→ import random; random.randint(1,10)

37. How to check Python version?

→ python --version or sys.version.

38. Difference between None and NaN.

→ None = absence of value, NaN = Not a Number (numeric missing).

39. Explain try-except-finally.

→ try = test, except = handle error, finally = always runs.

40. What is a DataFrame index?

→ Row identifier in Pandas.

# Situational Questions (Data Analyst Context)

1. You have a dataset with 1M rows. How would you optimize analysis?

→ Use Pandas chunksize, vectorized operations, drop unused cols.

2. A CSV has missing values. What do you do?

→ Drop or fill using mean/median depending on context.

3. How to find top 5 customers by revenue in a DataFrame?

→ df.sort\_values("revenue", ascending=False).head(5)

4. How do you combine multiple CSV files into one DataFrame?

→ Use glob + loop or pd.concat.

5. You need to calculate profit margin from Revenue & Cost columns.

→ df["profit\_margin"] = (df["Revenue"]-df["Cost"])/df["Revenue"]

6. How do you check duplicates in dataset?

→ df.duplicated().sum()

7. If dataset is slow to process, what libraries would you consider?

→ Dask, Vaex, or SQL pushdown.

8. How to find correlation between columns?

→ df.corr()

9. You need only numeric columns from DataFrame.

→ df.select\_dtypes(include='number')

10. A manager asks for "data summary" — how do you provide it?

→ df.describe() + visuals in Power BI.

# Behavioural Questions (Mindset & Approach)

1. How do you approach learning a new Python library?

→ Start with docs, small project, then apply in real dataset.

2. Tell me about a time you debugged a Python error.

→ Checked error log, used print/logging, Googled error message.

3. How do you explain Python concepts to non-technical stakeholders?

→ Use real-life analogies, avoid jargon, show simple examples.

4. If stuck in a coding problem, what do you do?

→ Break into small steps, use docs, ask peers, test iteratively.

5. Why do you prefer Python over Excel for analysis?

→ Python handles large datasets, automation, reproducibility.