Data Analytics Engineer

Python Coding Interview Preparation Guide

Generated on: December 01, 2025

What to Expect

1. Python Fundamentals

- Data structures: lists, dictionaries, sets, tuples
- List comprehensions and generator expressions
- Functions, decorators, lambda functions
- Object-oriented programming basics
- Error handling (try/except)
- Working with modules and packages

2. Data Manipulation (High Priority)

- Pandas: filtering, grouping, aggregations, merges, pivots
- NumPy: array operations, broadcasting
- Data cleaning: handling missing values, duplicates, outliers
- Data transformation: reshaping, pivoting, melting

3. Problem-Solving Approach

- Breaking down problems into smaller parts
- Writing clean, readable code
- Explaining your thought process
- Edge cases and error handling
- Code efficiency (time/space complexity)

4. Common Interview Patterns

- Data processing tasks (ETL-like)
- Aggregations and statistics
- Data validation and quality checks
- Working with nested data structures
- String/text processing

What an ML Engineer Will Focus On

Code Quality

- Clean, readable code
- Proper naming conventions
- DRY (Don't Repeat Yourself) principles
- Comments and docstrings
- Modular functions

Data Handling

- Efficient data processing
- Memory considerations
- Handling large datasets conceptually
- Data validation

Problem-Solving

- Logical thinking
- Breaking complex problems into steps
- Testing edge cases

Common Interview Question Types

1. Data Transformation

Example: Given a list of dictionaries, group by category and calculate averages. Clean and transform a messy dataset.

2. Aggregations and Statistics

Example: Find the top N items by some metric. Calculate rolling averages or percentiles.

3. Data Validation

Example: Write a function to validate data quality. Check for missing values, duplicates, outliers.

4. String/Text Processing

Example: Parse and clean text data. Extract patterns from strings.

5. Algorithmic Thinking

Example: Find duplicates efficiently. Merge or join datasets.

Preparation Tips

1. Practice Coding

- LeetCode Easy/Medium (focus on array/string problems)
- HackerRank Python challenges
- Pandas exercises (Kaggle, DataCamp)

2. Review These Topics

- List comprehensions
- Dictionary operations
- Pandas: groupby, merge, pivot, apply
- Lambda functions and map/filter/reduce
- Generators and iterators
- Error handling

3. Practice Explaining

- Think out loud
- Explain your approach before coding
- Discuss trade-offs
- Mention edge cases

4. Code Style

- Use meaningful variable names
- Write small, focused functions
- Add comments for complex logic
- Handle edge cases

Sample Questions You Might Get

1. Given a CSV-like structure, find the average value grouped by category
2. Write a function to clean phone numbers in various formats
3. Find duplicate records in a dataset efficiently
4. Transform nested JSON data into a flat structure
5. Calculate running statistics (mean, median) on a time series

What to Emphasize

- Your data engineering experience (ETL, pipelines)
- Understanding of data quality and validation
- Ability to write production-ready code
- Problem-solving approach
- Willingness to learn ML concepts if needed

Red Flags to Avoid

- Writing code without explaining
- Not handling edge cases
- Inefficient solutions (nested loops when not needed)
- Poor variable naming
- No error handling

Day-of Interview Tips

- Ask clarifying questions
- Start with a simple solution, then optimize
- Test your code with examples
- Discuss time/space complexity if relevant
- Show enthusiasm for learning

Quick Review Checklist

- ■ Python data structures (list, dict, set, tuple)
- ■ List/dict comprehensions
- ■ Pandas basics (groupby, merge, filter)
- ■ Error handling (try/except)
- ■ Functions and lambda functions
- ■ String manipulation
- ■ Working with dictionaries and nested structures

Good luck with your interview!