

Part 1: Multiple-Choice Questions (15 total)

#	Category	Example Question
1	Data Fundamentals	Which of the following best describes <i>structured data</i> ? A) Data stored in fixed fields ✓ B) Free text and multimedia C) Sensor logs D) Web scraping results
2	Data Cleaning	Which method is commonly used to handle missing values? A) Deleting the entire dataset B) Replacing with mean/median ✓ C) Sorting values D) Ignoring errors
3	Excel/Spreadsheets	Which Excel function combines text from multiple cells? A) CONCATENATE ✓ B) SUMIF C) INDEX D) TEXTJOIN
4	Statistics Basics	What does the <i>median</i> represent in a dataset? A) Average B) Middle value ✓ C) Highest value D) Range
5	Data Visualization	Which chart type is best for showing <i>part-to-whole</i> relationships? A) Line chart B) Pie chart ✓ C) Scatter plot D) Box plot
6	SQL Basics	What does the SQL command SELECT DISTINCT City FROM Customers; do? A) Selects all cities B) Removes duplicate city names ✓ C) Deletes duplicates D) Counts city records
7	Joins & Relationships	Which SQL join returns all records from both tables, matching where possible? A) INNER JOIN B) LEFT JOIN C) FULL OUTER JOIN ✓ D) CROSS JOIN

8	Data Types (SQL/Python)	Which of these is a <i>categorical</i> variable? A) Age B) Temperature C) Gender <input checked="" type="checkbox"/> D) Salary
9	Python Basics	What will <code>len([1,2,3,4])</code> return in Python? A) 3 B) 4 <input checked="" type="checkbox"/> C) 5 D) Error
10	Libraries & Tools	Which Python library is most commonly used for data manipulation? A) Matplotlib B) NumPy C) Pandas <input checked="" type="checkbox"/> D) Seaborn
11	Logical Thinking	Which of the following best describes a correlation of -0.9? A) Strong negative relationship <input checked="" type="checkbox"/> B) Weak positive relationship C) No relationship D) Strong positive relationship
12	Data Ethics	Why is anonymization important in data analysis? A) Reduces file size B) Protects personal data <input checked="" type="checkbox"/> C) Improves performance D) Avoids duplication
13	Error Checking	What does a <i>data outlier</i> typically indicate? A) Missing value B) Extreme or unusual observation <input checked="" type="checkbox"/> C) Typing mistake D) Average data point
14	Data Interpretation	If sales increase but profit margins decrease, what could that mean? A) Costs increased <input checked="" type="checkbox"/> B) Demand dropped C) Prices rose D) Data is incorrect
15	Reporting & Storytelling	What makes a data dashboard effective? A) It's colorful B) It uses fewer visuals C) It communicates key insights quickly <input checked="" type="checkbox"/> D) It has many charts

Part 2: Coding / Simulator Tasks (5 total)

#	Category	Example Practical Task
1	Python (Data Cleaning)	Given a CSV file with missing values in the “Age” column, write a Python script using Pandas to fill missing values with the column mean.
2	SQL Querying	Using the Orders table, write an SQL query to find the total revenue per customer , sorted from highest to lowest.
3	Data Analysis & Aggregation	Using Python (Pandas), load a dataset of employee records and calculate the average salary by department .
4	Data Visualization	Write a short Python script using Matplotlib or Seaborn to create a bar chart of product sales by category.
5	Logical Problem / Insight Generation	Given a dataset showing monthly expenses, identify the month with the highest variance and print an insight such as “Expense volatility peaked in Month X.”

Let's Get To Work.

This model is challenging but immensely rewarding. **You will be building a new ecosystem for talent in South Africa.**