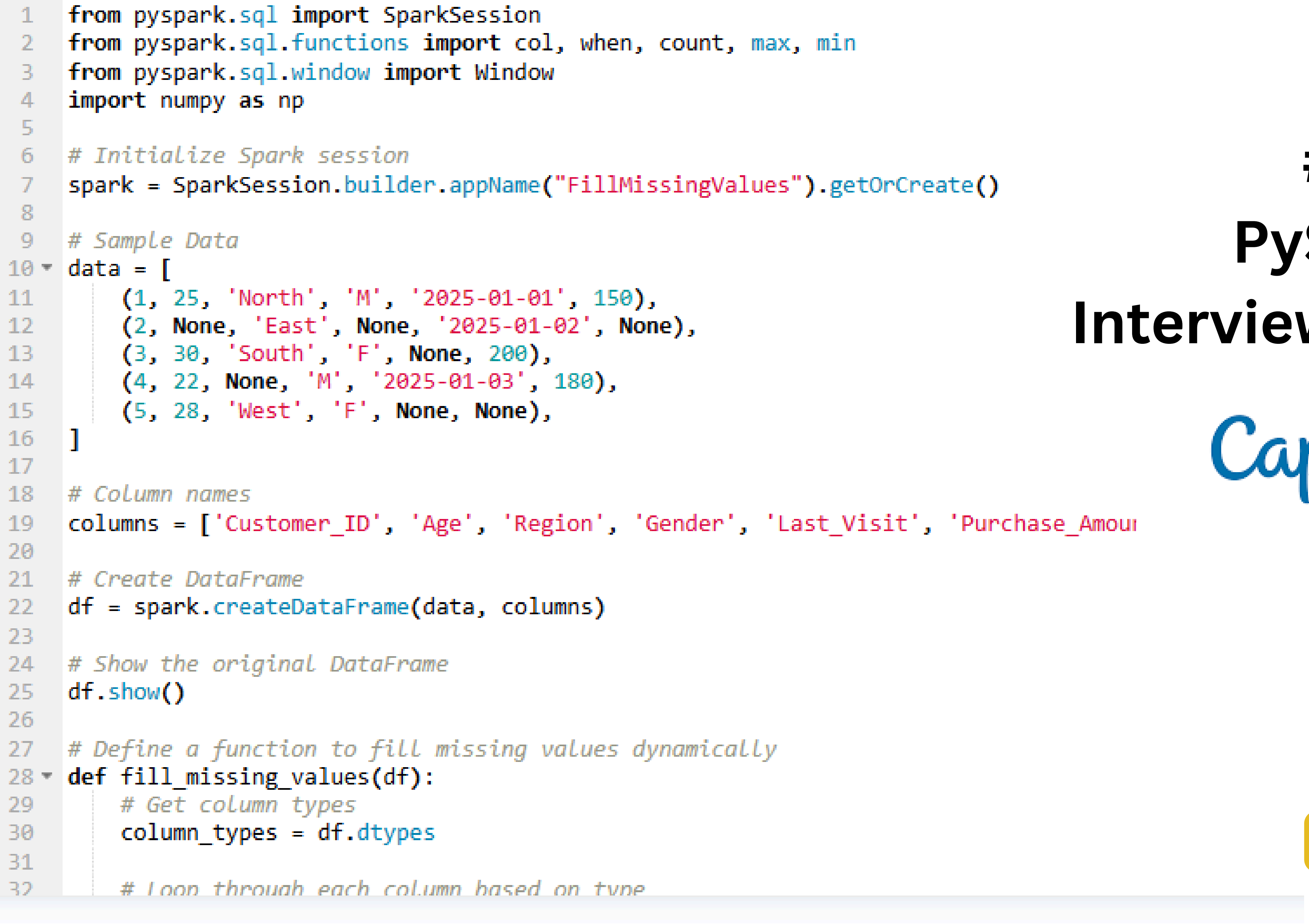
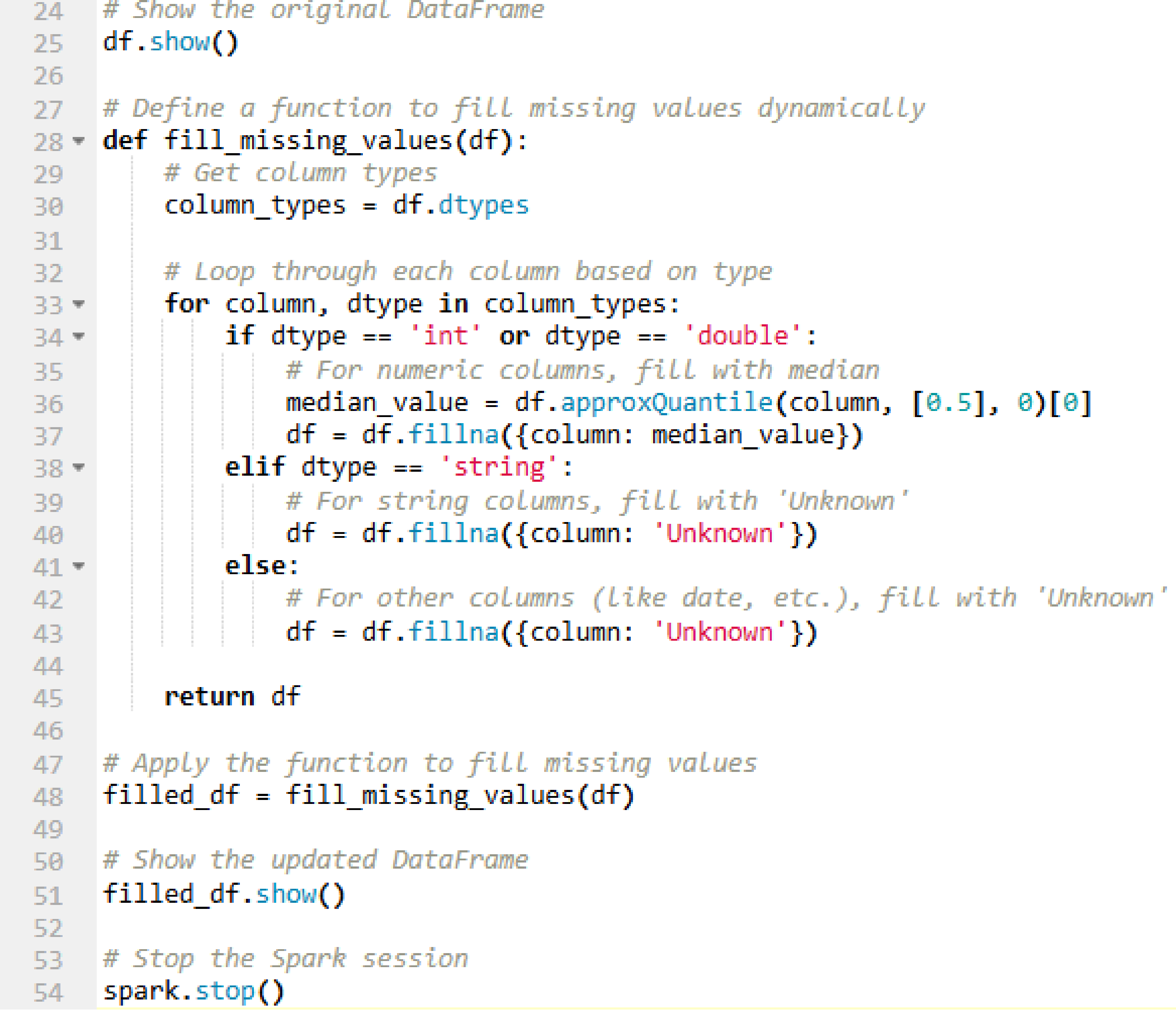


```python
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, when, count, max, min
from pyspark.sql.window import Window
import numpy as np

# Initialize Spark session
spark = SparkSession.builder.appName("FillMissingValues").getOrCreate()

# Sample Data
data = [
    (1, 25, 'North', 'M', '2025-01-01', 150),
    (2, None, 'East', None, '2025-01-02', None),
    (3, 30, 'South', 'F', None, 200),
    (4, 22, None, 'M', '2025-01-03', 180),
    (5, 28, 'West', 'F', None, None),
]

# Column names
columns = ['Customer_ID', 'Age', 'Region', 'Gender', 'Last_Visit', 'Purchase_Amou

# Create DataFrame
df = spark.createDataFrame(data, columns)

# Show the original DataFrame
df.show()

# Define a function to fill missing values dynamically
def fill_missing_values(df):
    # Get column types
    column_types = df.dtypes

    # Loop through each column based on type
```

#29
PySpark
Interview Questions

Capgemini

Karthik Kondpak

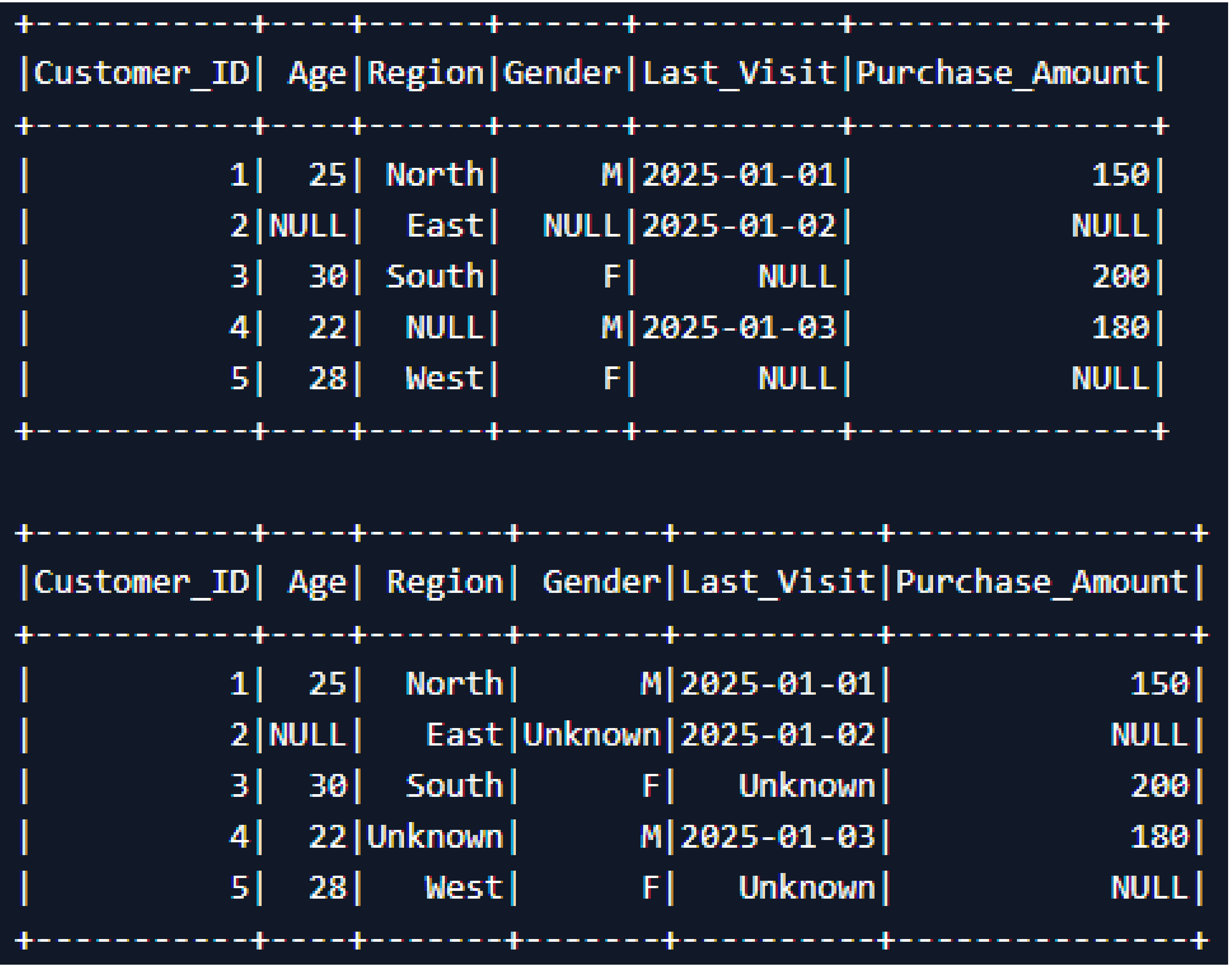


```
# Show the original DataFrame
df.show()

# Define a function to fill missing values dynamically
def fill_missing_values(df):
    # Get column types
    column_types = df.dtypes

    # Loop through each column based on type
    for column, dtype in column_types:
        if dtype == 'int' or dtype == 'double':
            # For numeric columns, fill with median
            median_value = df.approxQuantile(column, [0.5], 0)[0]
            df = df.fillna({column: median_value})
        elif dtype == 'string':
            # For string columns, fill with 'Unknown'
            df = df.fillna({column: 'Unknown'})
        else:
            # For other columns (like date, etc.), fill with 'Unknown'
            df = df.fillna({column: 'Unknown'})

    return df

# Apply the function to fill missing values
filled_df = fill_missing_values(df)

# Show the updated DataFrame
filled_df.show()

# Stop the Spark session
spark.stop()
```

#29
PySpark
Interview Questions

Capgemini

Karthik Kondpak



```
+-----------+----+------+------+----------+---------------+
|Customer_ID| Age|Region|Gender|Last_Visit|Purchase_Amount|
+-----------+----+------+------+----------+---------------+
|          1|  25| North|     M|2025-01-01|            150|
|          2|NULL|  East|  NULL|2025-01-02|           NULL|
|          3|  30| South|     F|      NULL|            200|
|          4|  22|  NULL|     M|2025-01-03|            180|
|          5|  28|  West|     F|      NULL|           NULL|
+-----------+----+------+------+----------+---------------+

+-----------+----+-------+-------+----------+---------------+
|Customer_ID| Age| Region| Gender|Last_Visit|Purchase_Amount|
+-----------+----+-------+-------+----------+---------------+
|          1|  25|  North|      M|2025-01-01|            150|
|          2|NULL|   East|Unknown|2025-01-02|           NULL|
|          3|  30|  South|      F|   Unknown|            200|
|          4|  22|Unknown|      M|2025-01-03|            180|
|          5|  28|   West|      F|   Unknown|           NULL|
+-----------+----+-------+-------+----------+---------------+
```

#29
PySpark
Interview Questions

Capgemini

Karthik Kondpak