- □ SparkSession
- □ DataFrames
- □ Column operations
- □ Filtering
- □ Computed fields

□ Dataset (Inline): `employee_data`

Prepare this in your notebook manually:

```
data = [
    ("Ananya", "HR", 52000),
    ("Rahul", "Engineering", 65000),
    ("Priya", "Engineering", 60000),
    ("Zoya", "Marketing", 48000),
    ("Karan", "HR", 53000),
    ("Naveen", "Engineering", 70000),
    ("Fatima", "Marketing", 45000)
]
columns = ["Name", "Department", "Salary"]

df = spark.createDataFrame(data, columns)
```

□ PySpark Exercises – No Solutions

□ Exercise Set 1: Basics

1. Display all records in the DataFrame.
2. Print the schema of the DataFrame.
3. Count total number of employees.

□ Exercise Set 2: Column Operations

4. Add a new column `Bonus` which is 15% of Salary.
5. Add a new column `NetPay` = Salary + Bonus.

□ Exercise Set 3: Filtering and Conditions

6. Display only employees from the “Engineering” department.
7. Display employees whose salary is greater than 60000.
8. Display employees who are not in the “Marketing” department.

□ Exercise Set 4: Sorting and Limiting

9. Show top 3 highest paid employees.
10. Sort the data by Department ascending and Salary descending.

□ Exercise Set 5: String and Case Logic

11. Add a new column `Level`:
    - “Senior” if salary > 60000

- “Mid” if salary between 50000 and 60000
- “Junior” otherwise

12. Convert all names to uppercase.
