# Initialize a Spark session

spark = SparkSession.builder \

.appName("Employee Data Analysis") \

.getOrCreate()

# Sample employee data

data = [

(1, 'Arjun', 'IT', 75000),

(2, 'Vijay', 'Finance', 85000),

(3, 'Shalini', 'IT', 90000),

(4, 'Sneha', 'HR', 50000),

(5, 'Rahul', 'Finance', 60000),

(6, 'Amit', 'IT', 55000)

]

# Define schema (columns)

columns = ['EmployeeID', 'EmployeeName', 'Department', 'Salary']

# Create DataFrame

employee\_df = spark.createDataFrame(data, columns)

# Show the DataFrame

employee\_df.show()

1. \*\*Task 1: Filter Employees by Salary\*\*

Filter the employees who have a salary greater than 60,000 and display the result.

employee\_df.filter(employee\_df["Salary"]>60000).show()

\*\*Hint\*\*: Use the `filter` method to filter based on the salary column.

2. \*\*Task 2: Calculate the Average Salary by Department\*\*

Group the employees by department and calculate the average salary for each department.

employee\_df.groupBy("Department").avg("Salary").show()

\*\*Hint\*\*: Use `groupBy` and `avg` functions.

3. \*\*Task 3: Sort Employees by Salary\*\*

Sort the employees in descending order of their salary.

\*\*Hint\*\*: Use the `orderBy` function and sort by the `Salary` column.

employee\_df.orderBy(col("Salary").desc()).show()

4. \*\*Task 4: Add a Bonus Column\*\*

Add a new column called `Bonus` which should be 10% of the employee's salary.

Bonus=employee\_df.withColumn("Bonus",col("Salary")\*0.1)

Bonus.show()

\*\*Hint\*\*: Use `withColumn` to add a new column.