



**16\_Select in pyspark**

Write a pyspark code perform below function

* Write a pyspark code to get all employee detail.
* Write a query to get only "FirstName" column from emp\_df
* Write a Pyspark code to get FirstName in upper case as "First Name".
* Write a pyspark code to get FirstName in lower case

**Difficult Level :** EASY

**DataFrame:**

**data = [**

**[1, "Vikas", "Ahlawat", 600000.0, "2013-02-15 11:16:28.290", "IT", "Male"],**

**[2, "nikita", "Jain", 530000.0, "2014-01-09 17:31:07.793", "HR", "Female"],**

**[3, "Ashish", "Kumar", 1000000.0, "2014-01-09 10:05:07.793", "IT", "Male"],**

**[4, "Nikhil", "Sharma", 480000.0, "2014-01-09 09:00:07.793", "HR", "Male"],**

**[5, "anish", "kadian", 500000.0, "2014-01-09 09:31:07.793", "Payroll", "Male"],**

**]**

**# Create a schema for the DataFrame**

**schema = StructType([**

**StructField("EmployeeID", IntegerType(), True),**

**StructField("First\_Name", StringType(), True),**

**StructField("Last\_Name", StringType(), True),**

**StructField("Salary", DoubleType(), True),**

**StructField("Joining\_Date", StringType(), True),**

**StructField("Department", StringType(), True),**

**StructField("Gender", StringType(), True)**

**])**



**# Creating Spark Session**

**from pyspark.sql import SparkSession**

**from pyspark.sql.types import StructType,StructField,IntegerType,StringType**

**#creating spark session**

**spark = SparkSession. \**

**builder. \**

**config('spark.shuffle.useOldFetchProtocol', 'true'). \**

**config('spark.ui.port','0'). \**

**config("spark.sql.warehouse.dir", "/user/itv008042/warehouse"). \**

**enableHiveSupport(). \**

**master('yarn'). \**

**getOrCreate()**

**# Create a list of rows from the image**

**data = [**

**[1, "Vikas", "Ahlawat", 600000.0, "2013-02-15 11:16:28.290", "IT", "Male"],**

**[2, "nikita", "Jain", 530000.0, "2014-01-09 17:31:07.793", "HR", "Female"],**

**[3, "Ashish", "Kumar", 1000000.0, "2014-01-09 10:05:07.793", "IT", "Male"],**

**[4, "Nikhil", "Sharma", 480000.0, "2014-01-09 09:00:07.793", "HR", "Male"],**

**[5, "anish", "kadian", 500000.0, "2014-01-09 09:31:07.793", "Payroll", "Male"],**

**]**

**# Create a schema for the DataFrame**

**schema = StructType([**

**StructField("EmployeeID", IntegerType(), True),**

**StructField("First\_Name", StringType(), True),**

**StructField("Last\_Name", StringType(), True),**

**StructField("Salary", DoubleType(), True),**

**StructField("Joining\_Date", StringType(), True),**

**StructField("Department", StringType(), True),**

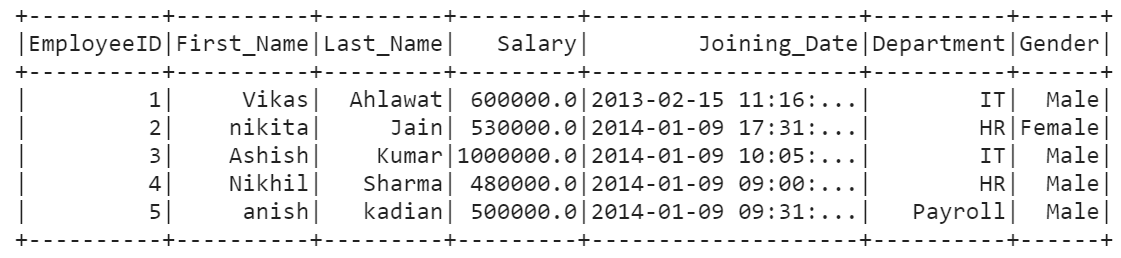
**StructField("Gender", StringType(), True)**

**])**

**emp\_df=spark.createDataFrame(data,schema)**

**#1. Write a pyspark code to get all employee detail**

**emp\_df.show()**

****

**# 2. Write a query to get only "FirstName" column from emp\_df**

**# Method 1**

**emp\_df.select("First\_Name").show()**

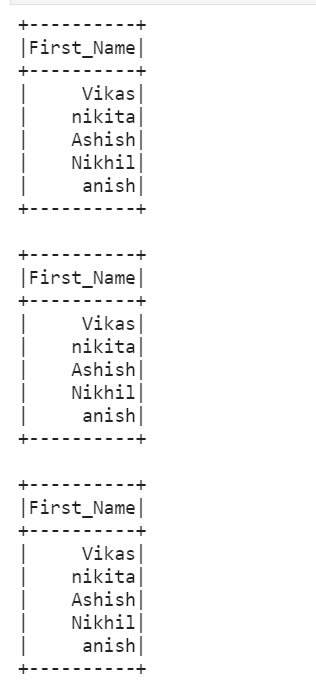
**# Method 2**

**emp\_df.select(col("First\_Name")).show()**

**# Method 3**

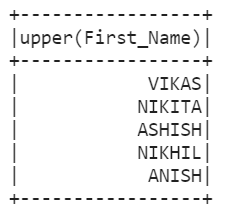
**emp\_df.createOrReplaceTempView("emp\_table")**

**spark.sql("select First\_Name from emp\_table").show()**

****

**# 3. Write a Pyspark code to get FirstName in upper case as "First Name".**

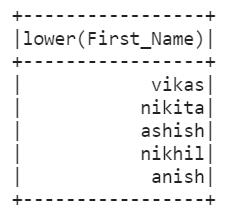
**emp\_df.select(upper("First\_Name")).show()**

****

**#4. Write a pyspark code to get FirstName in lower case**

**from pyspark.sql.functions import lower**

**emp\_df.select(lower("First\_Name")).show()**

****

