GLS UNIVERSITY
GLS (SRP) INSTITUTE OF COMPUTER TECHNOLOGY
SUBJECT: PRACTICALS IN DATA SCIENCE II
SEMESTER: V DIVISION: B
ASSIGNMENT: 3 TOPIC: PySpark with Hadoop

FACULTY: Ms. Prerna Agrawal

1. Write a Pyspark program to perform Hadoop streaming. Take a dataset of Employee_department.txt (Format: Id, Name, Department, Age, Date_of_join, Gender,Salary) and perform the following operations:
   - Count the total number of Employees.
   - Count the females working in HR department.
   - Count the females whose date of join is in current year.
   - Count the males working in Admin and IT department.
   - Count the males whose age is more than 25 and working in Admin department.
   - Display the Employee name with minimum salary.
   - Display the Employee name with maximum salary.

2. Write a Pyspark program to perform Hadoop streaming. Take a dataset of Person.txt (Format: Name, Gender, Age) and perform the following operations:
   - Count the total number of Males and Females.
   - Find the females having age more than 30 years.
   - Find the males whose age is 50 years.
   - Find the person with minimum and maximum age.

   Person.txt

   Alka F 35
   Palak F 40
   Nikita F 25
   Ankush M 15
   Paritosh M 50
   Jay M 28
   Bharti F 35
   Nimisha F 30
   Bharat M 23
   Nishant M 32

3. Write a Pyspark program to perform Hadoop streaming. Take a dataset of WomenChessPlayers.csv and perform the following operations:
   - Find the total count of chess players.
   - Find the names of the chess players belonging to India.
   - Find the names of the Indian chess players who are born in 1995.
   - Display the names of Russian chess players whose rapid rating and Blitz rating both is more than 2500.
   - List all the China players whose title is 'GM'.

4. Write a Pyspark program to perform Hadoop streaming. Take a dataset of WorldHappinessData2019.csv and perform the following operations:
   - List the top 5 countries having maximum happiness score.
   - List the country name having minimum health life expectancy.
   - List the country name having the maximum generosity level.
   - List the country names having social support more than 1.5 and Freedom to make life choices less than 0.5.
   - List the country names having maximum and minimum GDP per capita.