

Considering the 3 datasets below, write PySpark statements to perform the required operations.

Datasets:

1.

```
employeeColumn = ["emp_no", "birth_date", "first_name", "last_name", "gender", "hire_date"]
employeeData = [
    ["10001", "1953-09-02", "Georgi", "Facello", "M", "1986-06-26"],
    ["10002", "1964-06-02", "Bezael", "Simmel", "F", "1985-11-21"],
    ["10003", "1959-12-03", "Parto", "Bamford", "M", "1986-08-28"],
    ["10004", "1954-05-01", "Chirstian", "Koblick", "M", "1986-12-01"],
    ["10005", "1955-01-21", "Kyoichi", "Maliniak", "M", "1989-09-12"]
]
```

2.

```
jobColumn = ["emp_no", "title", "from_date", "to_date"] jobData
= [
    ["10001", "Senior Engineer", "1986-06-26", "9999-01-01"],
    ["10002", "Staff", "1996-08-03", "9999-01-01"],
    ["10003", "Senior Engineer", "1995-12-03", "9999-01-01"],
    ["10004", "Senior Engineer", "1995-12-01", "9999-01-01"],
    ["10005", "Senior Staff", "1996-09-12", "9999-01-01"]
]
```

3.

```
salaryColumn = ["emp_no", "title", "from_date", "to_date"] salaryData
= [
```

```
[
  ["10001", "66074", "1988-06-25", "1989-06-25"], ["10001", "62102", "1987-06-26", "1988-06-25"],
  ["10001", "60117", "1986-06-26", "1987-06-26"], ["10002", "72527", "2001-08-02", "9999-01-01"],
  ["10002", "71963", "2000-08-02", "2001-08-02"], ["10002", "69366", "1999-08-03", "2000-08-02"],
  ["10003", "43311", "2001-12-01", "9999-01-01"], ["10003", "43699", "2000-12-01", "2001-12-01"],
  ["10003", "43478", "1999-12-02", "2000-12-01"],
  ["10004", "74057", "2001-11-27", "9999-01-01"], ["10004", "70698", "2000-11-27", "2001-11-27"],
  ["10004", "69722", "1999-11-28", "2000-11-27"],
  ["10005", "94692", "2001-09-09", "9999-01-01"], ["10005", "91453", "2000-09-09", "2001-09-09"],
  ["10005", "90531", "1999-09-10", "2000-09-09"]
]
```

Requested operations:

1. Create 3 data frames with the above data
2. Rename the columns by using using capital letters and replace '_' with space
3. Format birth_date as 01.Jan.2021
4. Add a new column in employeeData where you compute the company email address by the following rule: [first 2 letter of first_name][last_name]@company.com
5. Calculate the average salary for each job role
6. Add a flag (set value to True) in salaryData if the average salary of the person is lower than the average salary for their job role