Table under consideration is Sales with the sample format given below:

Sale\_date, Country, Category, product, sales

Jan-2021, USA , Retail, UUID1, 10000

Jan-2021, USA, Healthcare, UUID2, 2000

Jan-2021, Canada, Retail, UUID3, 30000

Jan-2021, Canada, Hospitality, UUID4, 5000

Feb-2021, USA , Retail, UUID5, 10000

Feb-2021, USA, Healthcare, UUID6, 2000

Feb-2021, Canada, Retail, UUID7, 30000

Feb-2021, Canada, Hospitality, UUID8, 5000

**Questions:**

1.Top 3 categories within every country based on total sales in the current year.

The output should be in this format (Country, Category, Total sales)

2. Total sales of products sold in both Feb & Jan , Total sales of products sold ONLY in Jan, Total sales of products sold ONLY in Feb.

Output should be in this format (jan\_feb\_sales, jan\_sales, feb\_sales)

3. In the query written in question #1 what are the partitions and indexes you would create for best performance?

Table: Employee

+--------------+---------+

| Column Name | Type |

+--------------+---------+

| id | int |

| name | varchar |

| departmentId | int |

+--------------+---------+

id is the primary key column for this table.

departmentId is a foreign key of the ID from the Department table.

Each row of this table indicates the ID, name, and salary of an employee. It also contains the ID of their department.

Table: Salary

+--------------+---------+

| Column Name | Type |

+--------------+---------+

| employeeid | int |

| salary | varchar |

+--------------+---------+

Table: Department

+-------------+---------+

| Column Name | Type |

+-------------+---------+

| id | int |

| name | varchar |

+-------------+---------+

id is the primary key column for this table.

Each row of this table indicates the ID of a department and its name.

Q1 :: Employees with same last name and different departments.

output format :: (employee1\_name,department1\_name, employee2\_name, department2\_name)

Q2 :: Top 3 salaries department wise

output format :: (department, employee\_name, salary)

Q3 :: Salaries that are lower than average salary of department.

output format :: (department, employee\_name, salary)

**Python**::

Q1 : Average Price per SKU.

Q2 :: Country wise Number of unique products being sold, descending order of unique products

Q3 :: Convert The given CSV to parquet, with two new columns

column 1 :: 'currency' , populate the currency column with an appropriate value.

column 2 :: 'unit of measure' , populate column with an appropriate value.

Dtypes for Price column should be Float, rest all string.