PRACTICAL NO 5

Aim: Implementing OLAP using Python

Step 1: Create 4 types of below data, which can be converted to the data frame using Pandas:
1. Person
2. Sales
3. Quarter
4. Country

Step 2:
1. Get the country wise total sales, 2. Find Sales by both the person and the country, 3. Print Maximum individual sale by country

Code :-

```
import pandas as pd
data={
'persons':['manoj','sonali','manish','dhruv','ankit','amruta'],
'sales':[2000,3000,4000,5000,6000,7000],

'quarter':['Q1','Q2','Q3','Q4','Q5','Q6'],
'country':['India','UK','japan','korea','london','UK']
    }
print(data)
df=pd.DataFrame(data)
print(df)

country_wise_sales=df.groupby('country')['sales'].sum().reset_index()
```

```
print('/n country_wise sales:/n')
print(country_wise_sales)

person_country=df.groupby(['persons','country'])['sales'].sum().reset_index()
print('/n sales by both person and country :/n')
print(person_country)

pivot_data=df.pivot_table(index='country',values='sales',aggfunc='max')
print('/n maximum sales of each country')
print(pivot_data)
```

Output:-

{'persons': ['manoj', 'sonali', 'manish', 'dhruv', 'ankit', 'amruta'], 'sales': [2000, 3000, 4000, 5000, 6000, 7000], 'quarter': ['Q1', 'Q2', 'Q3', 'Q4', 'Q5', 'Q6'], 'country': ['India', 'UK', 'japan', 'korea', 'london', 'UK']}
  persons sales quarter country

```
0   manoj  2000    Q1   India
1  sonali  3000    Q2      UK
2  manish  4000    Q3   japan
3   dhruv  5000    Q4   korea
4   ankit  6000    Q5  london
5  amruta  7000    Q6      Uk
```

/n country_wise sales:/n
  country sales

```
0   India   2000
1      UK  10000
2   japan   4000
3   korea   5000
4  london   6000
```

/n sales by both person and country :/n
  persons country sales

```
0  amruta      UK   7000
1   ankit  london   6000
2   dhruv   korea   5000
3  manish   japan   4000
4   manoj   India   2000
5  sonali      UK    300
```

/n maximum sales of each country
        sales
country

```
India    2000
UK       7000
japan    4000
korea    5000
london   6000
```

Step 3: create a data frame using the following columns-
1.stud_names,
2. category(online/offline),
3. Gender,
4. Fees

Code:-
```
import pandas as pd
data={                          'Stud_names':['amruta','shubham','saniya','manoj','ankit','dhruv'],
```

```python
'Category':['Online','Offline','Online','Offline','Online','Offline'],

'Fees':[3400,5000,2500,5500,2000,1500],

'Gender':['female','Male','Female','Male','male','male']
}
#Preparing Dataframe
df=pd.DataFrame(data)
print("Dataframe: \n")
print(df)

#pivot table
pivot_data=df.pivot_table(index='Category',columns='Gender',values="Stud_names",aggfunc="count")
print("\n")

pivot_data['Male/Female Ratio']=pivot_data['Male']/pivot_data['Female']
print(pivot_data[['Male','Female','Male/Female Ratio']]
```

Step 4: find the male/female ratio under OLAP using Python
```python
print(pivot_data[['Male','Female','Male/Female Ratio']])
```
Output:-
Dataframe:

```
  Stud_names Category  Fees  Gender
0    amruta   Online   3400  female
1   shubham   Offline  5000    Male
2    saniya   Online   2500  Female
3     manoj   Offline  5500    Male
4     ankit   Online   2000    male
5     dhruv   Offline  1500    male
```

```
Gender    Male  Female  Male/Female Ratio
Category
Offline   2.0    NaN            NaN
Online    NaN    1.0            NaN
```