CSE 463 Data Warehousing and Mining

Date: 30-01-2025
Name - Debangan Ghosh
Roll - AP22110010984
Class - CSE - Q

Implementation of OLAP operations using python on Sales

AIM: The aim of this project is to implement **Online Analytical Processing (OLAP) operations** using Python by creating a **3×3×3 data cube** and performing essential OLAP operations such as **Roll-up, Drill-down, Slice, Dice, and Pivot** using **NumPy and Pandas**. This implementation helps in understanding how multidimensional data can be analyzed efficiently using Python-based data manipulation techniques in Jupyter Notebook.

Code:

```
import numpy as np
import pandas as pd

# dimensions
dimensions = {
    "Student": ["Ankit", "Dev", "Biswa"],
    "City": ["Delhi", "Kolkata", "Bangalore"],
    "Degree": ["Electronics", "Data Science", "Finance"]
}

# Creating 3x3x3 data cube for random sales data
data_cube = np.random.randint(100, 1000, size=(3, 3, 3))

# Converting to pandas DataFrame
data_list = []
for i, student in enumerate(dimensions["Student"]):
    for j, city in enumerate(dimensions["City"]):
        for k, degree in enumerate(dimensions["Degree"]):
            data_list.append([student, city, degree, data_cube[i, j, k]])
```

```
df = pd.DataFrame(data_list, columns=["Student", "City", "Degree", "Sales"])
print("Original Data Cube:\n", df.head())

#after dataframe - OLAP operations on the Sales data
# roll-up: data by Student
rollup_df = df.groupby("Student")["Sales"].sum().reset_index()
print("\nRoll-up (Aggregate by Student):\n", rollup_df)

# drill-down: showing data for a specific Student (e.g., Dev)
drilldown_df = df[df["Student"] == "Dev"]
print("\nDrill-down (Data for Dev):\n", drilldown_df)

# slice: Extracting data for specific city (e.g., Kolkata)
slice_df = df[df["City"] == "Kolkata"]
print("\nSlice (Data for Kolkata):\n", slice_df)

# dice: Extracting data for specific conditions (e.g., Dev and Data Science degree)
dice_df = df[(df["Student"] == "Dev") & (df["Degree"] == "Data Science")]
print("\nDice (Data for Dev and Data Science):\n", dice_df)

# Pivot: Rearrange data for better readability
pivot_df = df.pivot_table(values="Sales", index=["City", "Degree"], columns="Student",
aggfunc=np.sum)
print("\nPivot Table:\n", pivot_df)
```

Output:

```
Original Data Cube:
   Student     City        Degree  Sales
0   Ankit    Delhi    Electronics    150
1   Ankit    Delhi   Data Science    287
2   Ankit    Delhi        Finance    170
3   Ankit  Kolkata    Electronics    663
4   Ankit  Kolkata   Data Science    785

Roll-up (Aggregate by Student):
   Student  Sales
0   Ankit   3951
1   Biswa   4872
2     Dev   6311

Drill-down (Data for Dev):
    Student       City        Degree  Sales
9      Dev      Delhi    Electronics    577
```

10 Dev Delhi Data Science 597
11 Dev Delhi Finance 753
12 Dev Kolkata Electronics 983
13 Dev Kolkata Data Science 467
14 Dev Kolkata Finance 868
15 Dev Bangalore Electronics 426
16 Dev Bangalore Data Science 915
17 Dev Bangalore Finance 725

Slice (Data for Kolkata):

	Student	City	Degree	Sales
3	Ankit	Kolkata	Electronics	663
4	Ankit	Kolkata	Data Science	785
5	Ankit	Kolkata	Finance	720
12	Dev	Kolkata	Electronics	983
13	Dev	Kolkata	Data Science	467
14	Dev	Kolkata	Finance	868
21	Biswa	Kolkata	Electronics	109
22	Biswa	Kolkata	Data Science	215
23	Biswa	Kolkata	Finance	630

Dice (Data for Dev and Data Science):

	Student	City	Degree	Sales
10	Dev	Delhi	Data Science	597
13	Dev	Kolkata	Data Science	467
16	Dev	Bangalore	Data Science	915

Pivot Table:

Student		Ankit	Biswa	Dev
City	Degree			
Bangalore	Data Science	162	932	915
	Electronics	599	846	426
	Finance	415	289	725
Delhi	Data Science	287	856	597
	Electronics	150	594	577
	Finance	170	401	753
Kolkata	Data Science	785	215	467
	Electronics	663	109	983
	Finance	720	630	868