DATA ANALYTICS AND VISULIZATION

LAB ASSIGNMENT WEEK-2:

HTNO:2403A52332

NAME:BOLLAM RISHIKA

OBJECTIVE:

1)As part of a university analytics project, you're analyzing student
performance across departments.
students.csv:
student_id, name, department
S001, Arjun, CSE
S002, Kavya, ECE
S003, Rahul, CSE
grades.csv:
grade_id, student_id, course, score
G01, S001, DBMS, 85
G02, S002, DSP, 78
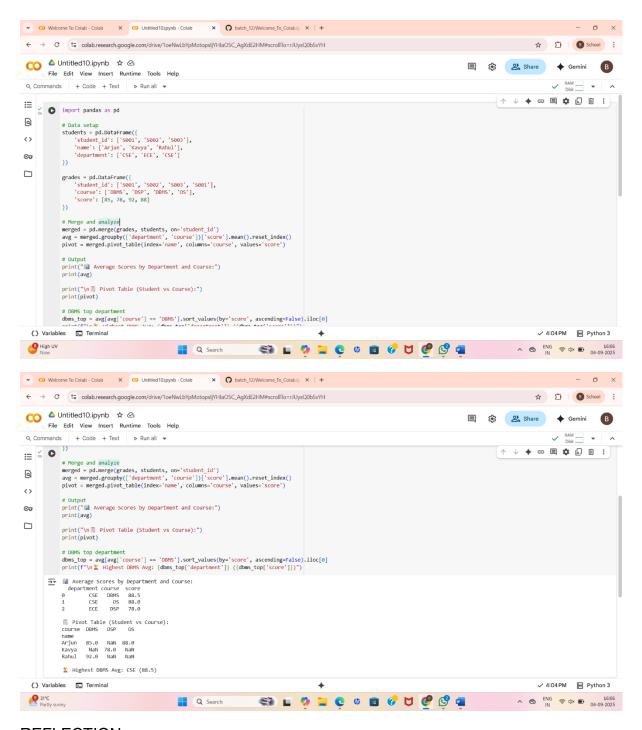G03, S003, DBMS, 92
G04, S001, OS, 88
Tasks:
• Merge the datasets to view student performance by course
and department.
• Group by department and course to calculate average
scores.
• Create a pivot table with student name as rows and courses
as columns with scores.
Questions:
• Which department has the highest average in DBMS?

CODE:

```python
import pandas as pd

# Data setup
students = pd.DataFrame({
    'student_id': ['S001', 'S002', 'S003'],
    'name': ['Arjun', 'Kavya', 'Rahul'],
    'department': ['CSE', 'ECE', 'CSE']
})

grades = pd.DataFrame({
    'student_id': ['S001', 'S002', 'S003', 'S001'],
    'course': ['DBMS', 'DSP', 'DBMS', 'OS'],
    'score': [85, 78, 92, 88]
})

# Merge and analyze
merged = pd.merge(grades, students, on='student_id')
avg = merged.groupby(['department', 'course'])['score'].mean().reset_index()
pivot = merged.pivot_table(index='name', columns='course', values='score')

# Output
print("📊 Average Scores by Department and Course:")
print(avg)

print("\n📋 Pivot Table (Student vs Course):")
print(pivot)

# DBMS top department
dbms_top = avg[avg['course'] == 'DBMS'].sort_values(by='score', ascending=False).iloc[0]
```



```python
})

# Merge and analyze
merged = pd.merge(grades, students, on='student_id')
avg = merged.groupby(['department', 'course'])['score'].mean().reset_index()
pivot = merged.pivot_table(index='name', columns='course', values='score')

# Output
print("📊 Average Scores by Department and Course:")
print(avg)

print("\n📋 Pivot Table (Student vs Course):")
print(pivot)

# DBMS top department
dbms_top = avg[avg['course'] == 'DBMS'].sort_values(by='score', ascending=False).iloc[0]
print(f"\n🏆 Highest DBMS Avg: {dbms_top['department']} ({dbms_top['score']})")
```

```
📊 Average Scores by Department and Course:
  department course  score
0        CSE   DBMS   88.5
1        CSE     OS   88.0
2        ECE    DSP   78.0

📋 Pivot Table (Student vs Course):
course  DBMS   DSP    OS
name
Arjun   85.0   NaN  88.0
Kavya    NaN  78.0   NaN
Rahul   92.0   NaN   NaN

🏆 Highest DBMS Avg: CSE (88.5)
```

REFLECTION:

This program showcases a strong grasp of data manipulation using **pandas**, one of Python's most powerful libraries for structured data. By merging two datasets—students and grades—it mirrors real-world relational database operations, such as **JOINs**, and demonstrates how to extract meaningful insights from raw tabular data.

The use of **groupby** to calculate average scores by department and course reflects an analytical mindset, turning individual records into aggregated performance metrics. This is especially valuable in academic or institutional settings where trends across departments matter more than individual scores.