**Day 5 - Python Refresher & Advanced Python for Data Engineering**

🗓️ Date: **10 June 2025**

* **Core Python Refresher**

**Python Basics**

* **Features:** Easy syntax, interpreted, dynamic typing, huge libraries.
* **First Program:**
* print("Hello, Data Engineers!")
* **Variables & Data Types**
* name = "Sakshi" # String
* age = 23 # Integer
* score = 89.5 # Float
* passed = True # Boolean

**Control Structures**

for i in range(3):

print("Looping", i)

if score > 90:

print("Excellent!")

elif score >= 75:

print("Good job")

else:

print("Keep learning")

**Collections Overview**

| **Type** | **Example** |
| --- | --- |
| List | marks = [85, 78, 92] |
| Dictionary | {"name": "Sakshi", "score": 95} |
| Set | {1, 2, 3} |
| Tuple | ("Alice", [85, 78, 92]) |

**Functions & Lambda**

def greet(name="Friend"):

return f"Hello, {name}!"

add = lambda x, y: x + y

print(add(10, 5)) # Output: 15

**OOP in Python**

class Student:

def \_\_init\_\_(self, name, marks):

self.name = name

self.marks = marks

def average(self):

return sum(self.marks) / len(self.marks)

s = Student("Sakshi", [85, 90, 95])

print(s.average()) # Output: 90.0

**File Handling & Exception**

try:

with open("data.csv", "r") as f:

content = f.read()

except FileNotFoundError:

print("File not found!")

**Modules and Packages**

# mymodule.py

def add(a, b):

return a + b

# main.py

import mymodule

print(mymodule.add(2, 3)) # Output: 5

**Data Engineering Essentials**

**CSV / JSON Handling with Pandas**

import pandas as pd

df = pd.read\_csv("students.csv")

print(df.head())

df.to\_json("students.json")

**MySQL with Python**

import mysql.connector

conn = mysql.connector.connect(

host="localhost", user="root", password="1234", database="school"

)

cursor = conn.cursor()

cursor.execute("SELECT \* FROM students")

for row in cursor.fetchall():

print(row)

**Data Cleaning with NumPy & Pandas**

import numpy as np

import pandas as pd

arr = np.array([[1, np.nan, 3], [4, 5, np.nan]])

df = pd.DataFrame(arr)

df.fillna(0, inplace=True)

**Task 1 – Students Marks Management System**

students = [

("Sakshi",[88,89,90]),

("Jennifer",[89,88,90]),

("Jack",[90,91,85])

]

print("Average Marks of Each Student: ")

topper\_name = ""

highest\_avg = 0

for name, marks in students:

avg = sum(marks)/len(marks)

print(f"{name} - Average: {avg:.2f}")

if avg > highest\_avg:

highest\_avg = avg

topper\_name = name

print("\nTopper:",topper\_name)

students.append(("Rock", [80,89,95]))

for i in range(len(students)):

if students[i][0] == "Jennifer":

students[i][1][1] = 89

break

print("\n Updated Student's List: ")

for name, marks in students:

print(f"{name}: {marks}")

student\_stats = []

for name, marks in students:

total = sum(marks)

avg = total/len(marks)

student\_stats.append((name, total, round(avg, 2)))

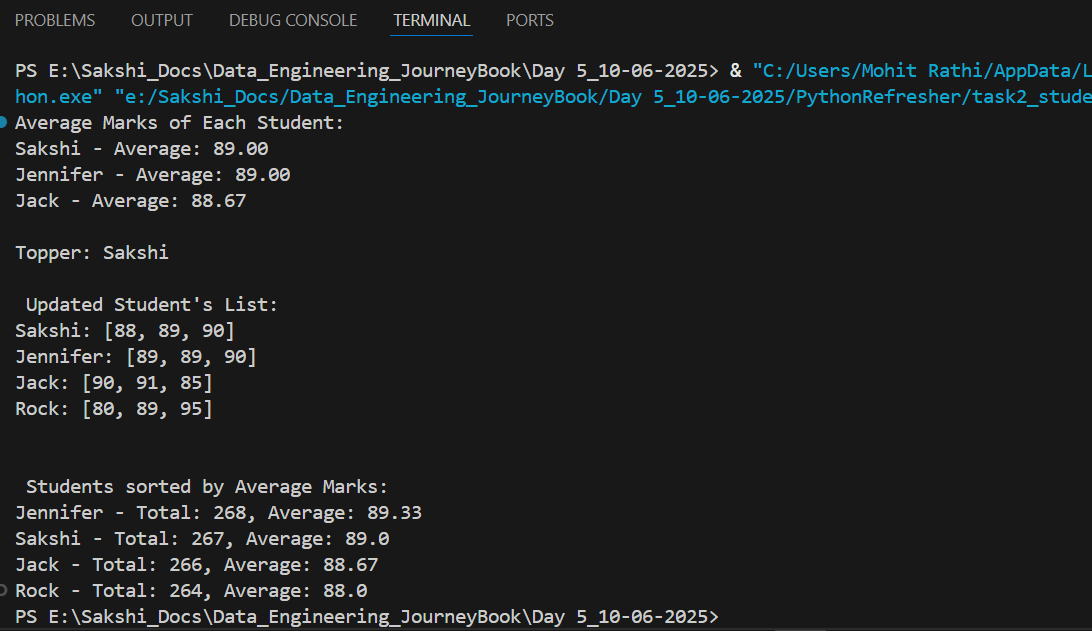
student\_stats.sort(key=lambda x: x[2], reverse=True)

print("\n Students sorted by Average Marks: ")

for name, total, avg in student\_stats:

print(f"{name} - Total: {total}, Average: {avg}")

**Output –**

****

**Task 2 – Smart City Project**

import numpy as np

temps = np.array([

[32, 33, 31, 30, 29, 28, 27],

[35, 36, 34, 32, 33, 31, 30],

[25, 26, 27, 28, 29, 30, 31],

[20, 22, 24, 23, 25, 26, 27],

[38, 39, 37, 36, 35, 34, 33]

])

days = ["Mon", "Tue", "Wed", "Thu", "Fri", "Sat", "Sun"]

#1. Function: Average temperature of a city for the week

def average\_temperature(city\_index):

return np.mean(temps[city\_index])

# 2. Function: Hottest day index for a city

def hottest\_day(city\_index):

return np.argmax(temps[city\_index])

# 3. Loop through all cities and print average temperature and hottest day

print("City-wise Average Temperature and Hottest Day:")

for i in range(5):

avg\_temp = average\_temperature(i)

hot\_day\_index = hottest\_day(i)

hot\_day\_name = days[hot\_day\_index]

print(f"City {i+1}: Avg Temp = {avg\_temp:.2f}°C, Hottest Day = {hot\_day\_name} (Index: {hot\_day\_index})")

print()

#4. Function: Day-wise average temperature across all cities

def day\_wise\_avg():

return np.mean(temps, axis=0)

# Print day-wise average

day\_avg = day\_wise\_avg()

print("Day-wise Average Temperatures Across All Cities:")

for i in range(7):

print(f"{days[i]}: {day\_avg[i]:.2f}°C")

print()

#5. Loop to find:

# City with highest weekly average

city\_avgs = [average\_temperature(i) for i in range(5)]

highest\_city\_index = np.argmax(city\_avgs)

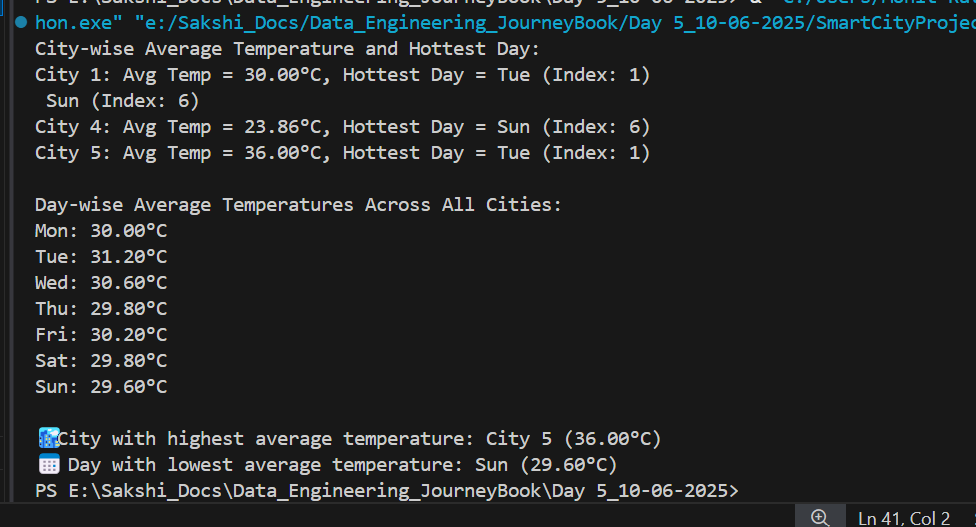
# Day with lowest average temperature

lowest\_day\_index = np.argmin(day\_avg)

print(f"🏙️ City with highest average temperature: City {highest\_city\_index + 1} ({city\_avgs[highest\_city\_index]:.2f}°C)")

print(f"📅 Day with lowest average temperature: {days[lowest\_day\_index]} ({day\_avg[lowest\_day\_index]:.2f}°C)")

**Output –**

****