CS 101 - Foundation of Data Science and Engineering

PSET-7 - Object Oriented Programming

Objective:

In this assignment you will be working with Class, Objects and Inheritance that was discussed in the class.

This is an individual assignment. No collaboration is allowed!

DataAnalyzer for Student Projects

As a data science student, you are tasked with analyzing project data for a research lab. The lab wants to track the progress and results of various student projects to optimize resource allocation and promote effective study habits. Your program will need to manage project data and calculate statistics based on this data.

Question 1: (5 pts)

Create a class called Student with the following attributes

1.student_id: a unique identifier for the student as an integer

2.major: the student's major as a string (e.g., "Computer Science", "Data Science")

3.university: the name of the university (e.g., "Harvard", "MIT", "Other")

Please be sure to include a constructor that can initialize the class with the above attributes. Assume inputs passed to your class will always be correct i.e. you are not required to check types, and verify uppercase or lower case.

In [2]:
```
class Student:
    def __init__(self, student_id, major, university):
        self.student_id = student_id
        self.major = major
        self.university = university
```

Question 2: (50 pts)

Create a class called Project which is a subclass of Student with the following attributes and methods.

Attributes (5 points) :

Public attributes :

1. project_id: a unique identifier for the project as an integer

2. data_points: a list of numerical values collected from the project

Private attributes : (In python private members are stored by adding 2 underscores in front of the name of the variable. Example: analysis_results should be declared as __analysis_results)

3. analysis_results : a dictionary to store results of various analysis

4. active : a boolean indicating if the project is currently active (default is True)

Methods 1-4(15 points) :

1. add_data(self, new_data): Appends a new data point to the data_points list.

2. get_results(self): Returns the analysis_results dictionary

3. is_active(self): Returns the status of the active attribute.

4. set_active(self,status): Sets the status of the active attribute (True/False).

5. perform_analysis(self): Analyzes data_points to calculate and store/update analysis_results with statistical results (mean, median, variance). (20 pts)

Notes for perform_analysis:

1. The method should only store the mean, median, and variance in analysis_results dictionary.It should not return the analysis_results.

2. Use the sample formula to calculate variance.

3. Use of inbuilt methods is not allowed for calculation of mean, median and variance

4. Please do not omit the decimal points for each of the statistic calculated.

Exception Handling :(10 points)

1. Implement a mechanism to handle incorrect data type. At the minimum you should check all the data in the list are numerical in the constructor, and check if new data point being added is of numerical type. Raise an error if you encounter list that have non-numerical values and if non-numerical values are being added to the list. https://docs.python.org/3/library/exceptions.html#TypeError (5 pts)

2. In your perform_analysis method use try-except to catch errors resulting from ZeroDivisionError if the list is empty. You should print a friendly text if you encounter such error in perform_analysis method, i.e. do not raise an error. (5 points)

Note: Please see Question 3 for order of parameters being passed to the constructor in the example

In [207…

```python
class Project(Student):
    def __init__(self, student_id, major, university, project_id, data_point
        super().__init__(student_id, major, university)
        self.project_id = project_id

        if not all(isinstance(x, (int, float)) for x in data_points):
            raise TypeError("All data points must be numerical values (int c

        self.data_points = data_points
        self.__analysis_results = {}
        self.active = True

    def perform_analysis(self):
        try:
            avg = sum(self.data_points) / len(self.data_points)
            self.__analysis_results = {
                "min": min(self.data_points),
                "max": max(self.data_points),
                "avg": avg
            }
        except ZeroDivisionError:
            print("Cannot perform analysis: No data points available.")

    def get_results(self):
        return self.__analysis_results

    def is_active(self):
        return self.active and len(self.data_points) > 0

    def set_active(self, value):
        if not isinstance(value, bool):
            raise TypeError("Active status must be a boolean value.")
        self.active = value

    def add_data(self, value):
        if not isinstance(value, (int, float)):
            raise TypeError("New data point must be a numerical value.")
        self.data_points.append(value)
```

In [209…

```
#TEST PROJECT
# Create a Project instance
project1 = Project(
    student_id=1234,
    major="Data Science",
    university="Harvard",
    project_id=5678,
    data_points=[85, 90, 95, 100]
)

print(f"Project ID: {project1.project_id}")
print(f"Data Points: {project1.data_points}")
print(f"Student Major: {project1.major}")
```

```
Project ID: 5678
Data Points: [85, 90, 95, 100]
Student Major: Data Science
```

Question-3 - Testing your code (5 pts)

Write at least 5 test cases to test different parts of your implementation. We have provided an example. You will get full points as long as you have 5 test cases and a description of what is being tested. Please be sure to include at lease 2 test cases for perform_analyis method which should include at least one test case for checking exceptions. You are welcome to write your test cases using python unit test library. However, it is not required.

In [212…

```
#Example 1- Test Student Class constructor with student_id 1001, major "Data

john = Student(1001, "Data Science", "Harvard")
print(f"Student id is : {john.student_id}\nMajor: {john.major}\nUniversity:{
```

```
Student id is : 1001
Major: Data Science
University:Harvard
```

In [214…

```
#Example 2- Test Project Class constructor with student_id 1001, major "Data
#data_points [65,75,95,99]"

my_project = Project(1001, "Data Science", "Harvard", 1, [65,75,95,99])
print(f"Student id is : {my_project.student_id}\nMajor: {my_project.major}\n
project_id:{my_project.project_id}\ndata_points : {my_project.data_points}\n
```

```
Student id is : 1001
Major: Data Science
University:Harvard
project_id:1
data_points : [65, 75, 95, 99]
is_active:True
```

In [216…

```
#Test 1: Student Class Constructor
print("Test 1: Student Class Constructor")
```

```
john = Student(1001, "Data Science", "Harvard")
print(f"Student id is : {john.student_id}\nMajor: {john.major}\nUniversity:
```

```
Test 1: Student Class Constructor
Student id is : 1001
Major: Data Science
University: Harvard
```

In [218…

```
#Test 2: Project Class Constructor
print("\nTest 2: Project Class Constructor")
my_project = Project(1001, "Data Science", "Harvard", 1, [65, 75, 95, 99])
print(f"Student id is : {my_project.student_id}\nMajor: {my_project.major}\r
      f"project_id:{my_project.project_id}\ndata_points : {my_project.data_p
```

```
Test 2: Project Class Constructor
Student id is : 1001
Major: Data Science
University:Harvard
project_id:1
data_points : [65, 75, 95, 99]
is_active:True
```

In [220…

```
#Test 3: perform_analysis with valid data
print("\nTest 3: perform_analysis with valid data")
result = my_project.perform_analysis()
print("Analysis Results:", result)
```

```
Test 3: perform_analysis with valid data
Analysis Results: None
```

In [222…

```
#Test 4: perform_analysis with empty data (Exception Handling)
print("\nTest 4: perform_analysis with empty data")
empty_project = Project(1002, "Computer Science", "MIT", 2, [])
try:
    empty_project.perform_analysis()
except ValueError as e:
    print("Caught expected exception:", e)
```

```
Test 4: perform_analysis with empty data
Cannot perform analysis: No data points available.
```

In [224…

```
#Test 5: perform_analysis with single data point
print("\nTest 5: perform_analysis with one data point")
single_data_project = Project(1003, "AI", "Stanford", 3, [100])
result = single_data_project.perform_analysis()
print("Analysis Results (single point):", result)
```

```
Test 5: perform_analysis with one data point
Analysis Results (single point): None
```

Question 4-9 (30 points)

Absolute Tests - For Question 4-9 simply run the cell below. We will verify your output with expected outputs. You implementation should be based on the requirement and not these test cases. If you have implemented the requirement

correctly, your output should be as expected. Your implementation should be correct to get full credit here. Please do not modify the code for question 4-9

Question 4: Tests Class and Objects are created, and perform_analysis yields results as expected. (10 pts)

In [228…

```
test1 = [(1000, 'Computer Science', 'HARVARD', 1, [65,75,95,99]),(1001, 'Com
         (1003,'Data Science', 'Cornell', 3, [75,85,95,99,33])]


for test in test1:
    project = Project(test[0],test[1],test[2],test[3],test[4])
    project.perform_analysis()
    project_info = f"Student id : {project.student_id}\nMajor: {project.majc
project_id:{project.project_id}\ndata_points : {project.data_points}\nis_act
    print(project_info)
    print(f"analysis results :{project.get_results()}\n")
```

```
Student id : 1000
Major: Computer Science
University:HARVARD
project_id:1
data_points : [65, 75, 95, 99]
is_active:True
analysis results :{'min': 65, 'max': 99, 'avg': 83.5}

Student id : 1001
Major: Computer Science
University:MIT
project_id:2
data_points : [95, 35, 75, 90, 91]
is_active:True
analysis results :{'min': 35, 'max': 95, 'avg': 77.2}

Student id : 1003
Major: Data Science
University:Cornell
project_id:3
data_points : [75, 85, 95, 99, 33]
is_active:True
analysis results :{'min': 33, 'max': 99, 'avg': 77.4}
```

Question 5: Tests if data points can be added and data statistics are updated.(5 points)

In [232…

```
test = (1000, 'Computer Science', 'HARVARD', 1, [65,75,95,99])
project = Project(test[0],test[1],test[2],test[3],test[4])
project.perform_analysis()
```

```
print(f"analysis results before : {project.get_results()}")
project.add_data(22)
project.perform_analysis()
print(f"analysis results after: {project.get_results()}")
```

```
analysis results before : {'min': 65, 'max': 99, 'avg': 83.5}
analysis results after: {'min': 22, 'max': 99, 'avg': 71.2}
```

Question 6: Tests if active can be set (5 points)

In [235…

```
test = (1000, 'Computer Science', 'HARVARD', 1, [65,75,95,99])
project = Project(test[0],test[1],test[2],test[3],test[4])
print(f"value of active at initialization: {project.is_active()}")
project.set_active(False)
print(f"value of active after calling set method: {project.is_active()}")
```

```
value of active at initialization: True
value of active after calling set method: False
```

Question 7: Tests if perform_analysis can handle exceptions for empty list (4 points)

In [238…

```
test1 = (1000, 'Computer Science', 'HARVARD', 1, [])

print(f"Test #1 : Test perform analysis if the data points is an empty list.
project = Project(test1[0],test1[1],test1[2],test1[3],test1[4])
project.perform_analysis()
```

```
Test #1 : Test perform analysis if the data points is an empty list.
Cannot perform analysis: No data points available.
```

Question 8: Tests if constructor raises an error if non-numerical values are present in data_points (3 pts)

In [241…

```
test2 = (1001, 'Computer Science', 'MIT', 2, [95,35,75,90,'a'])
#print(f"\nTest #3 : Test constructor if data points has non-numerical value
project = Project(test2[0],test2[1],test2[2],test2[3],test2[4])
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[241], line 3
      1 test2 = (1001, 'Computer Science', 'MIT', 2, [95,35,75,90,'a'])
      2 #print(f"\nTest #3 : Test constructor if data points has non-numeric
al values.")
----> 3 project = Project(test2[0],test2[1],test2[2],test2[3],test2[4])

Cell In[207], line 7, in Project.__init__(self, student_id, major, universit
y, project_id, data_points)
      4 self.project_id = project_id
      6 if not all(isinstance(x, (int, float)) for x in data_points):
----> 7     raise TypeError("All data points must be numerical values (int o
r float).")
      9 self.data_points = data_points
     10 self.__analysis_results = {}

TypeError: All data points must be numerical values (int or float).
```

Question 9: Tests if adding non-numerical values to data_points raises an error (3 pts)

In [244…
```
test3 = (1000, 'Computer Science', 'HARVARD', 3, [100,100,100])
project = Project(test3[0],test3[1],test3[2],test3[3],test3[4])
project.add_data('xyz')
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[244], line 3
      1 test3 = (1000, 'Computer Science', 'HARVARD', 3, [100,100,100])
      2 project = Project(test3[0],test3[1],test3[2],test3[3],test3[4])
----> 3 project.add_data('xyz')

Cell In[207], line 37, in Project.add_data(self, value)
     35 def add_data(self, value):
     36     if not isinstance(value, (int, float)):
---> 37         raise TypeError("New data point must be a numerical value.")
     38     self.data_points.append(value)

TypeError: New data point must be a numerical value.
```

Question 10 (5 pts)

Working with Class and objects is very important in data science You will be working with a lot of python libraries which are all created using class and objects. You have already worked with pandas. Let's tie what you have learned in this homework with what you have already learned. Run the cell below, and answer the question that follow

In [247…
```
import pandas as pd
midterm_scores = [96,90,85]
df = pd.DataFrame(midterm_scores, columns=['midterm_scores'])
```

```
shape = df.shape
print(shape)
df.sort_values(by='midterm_scores', ascending=False)
```

```
(3, 1)
```

Out[247…

	midterm_scores
0	96
1	90
2	85

1. Identify the class name and the parameters which is being used to create the object df. Please refer to pandas documentation to answer this question. https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html (2pt)
2. Identify the attribute being called on the object df. (1pt)
3. Identify the method, and the method parameters being called by our object df. (2 pt)

your answer here

In [251…
```
df = pd.DataFrame(midterm_scores, columns=['midterm_scores'])
```

In [253…
```
df.shape
```

Out[253…
```
(3, 1)
```

In [255…
```
df.sort_values(by='midterm_scores', ascending=False)
```

Out[255…

	midterm_scores
0	96
1	90
2	85

Coding style (5 pts)

Please make sure your code follows the coding style as defined here.

1. Comments on functions.
2. Additional comments on parts of code that maybe difficult to understand
3. Remove any commented code i.e. your solution should only include the code required by the assignment. All experiemental code needs to removed on the submitted file.
4. Code Readibility (if part of your code is long, use multiple lines)

Submission on Gradescope

On canvas left menu -> click on Gradescope

Submit the jupyter notebook, and a pdf version of this notebook.

To create a pdf of this notebook : In your browser open print, and save as pdf. Name the pdf LastNameFirstName_pset7.pdf example: DoeJohn_pset7.pdf

Name this jupyter notebook with the same format LastNameFirstName.ipynb

Make sure that your notebook has been run before creating pdf. Any outputs from running the code needs to be clearly visible. We need both .ipynb, and pdf of this notebook to assign you grades.

Drop all the files in gradescope under PSET 7: Python OO Programming Exercise

```
In [ ]:
```