Data Cleaning Using String Functions for Variables and Data Structures

Assignment I

Submitted By: Ankit Shrestha

Submitted To: Samriddha Pathak

Date: July 6, 2025

Table of Contents

Easy 4

1. Remove Spaces: 4
2. Remove Leading Characters: 4
3. Remove Trailing Characters: 5
4. Capitalize a Sentence: 5
5. Title Case a Name: 6
6. Clean List of Names: 6
7. Remove Custom Characters: 7
8. Capitalize All Names in List: 7
9. Clean Dictionary Values: 8
10. Title Case Sentences in List: 8

Intermediate 9

1. Clean and Title Case: 9
2. Clean List of Emails: 9
3. Remove Leading Numbers: 10
4. Clean Nested List: 10
5. Capitalize After Cleaning: 11
6. Clean Dictionary Keys: 11
7. Clean and Deduplicate Names: 12
8. Remove Multiple Characters: 12
9. Conditional Cleaning in List: 13
10. Clean and Group by First Letter: 13

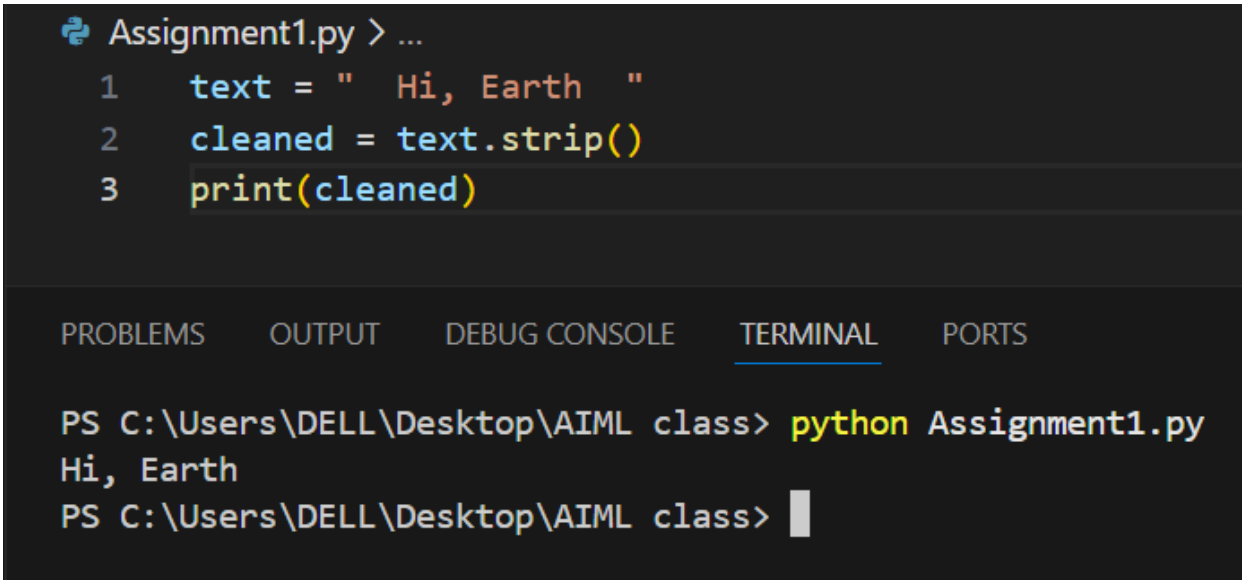
Hard 14

1. Clean Set of Strings: 14
2. Complex Nested Cleaning: 14
3. Custom Title Function: 15
4. Clean and Format Emails: 15

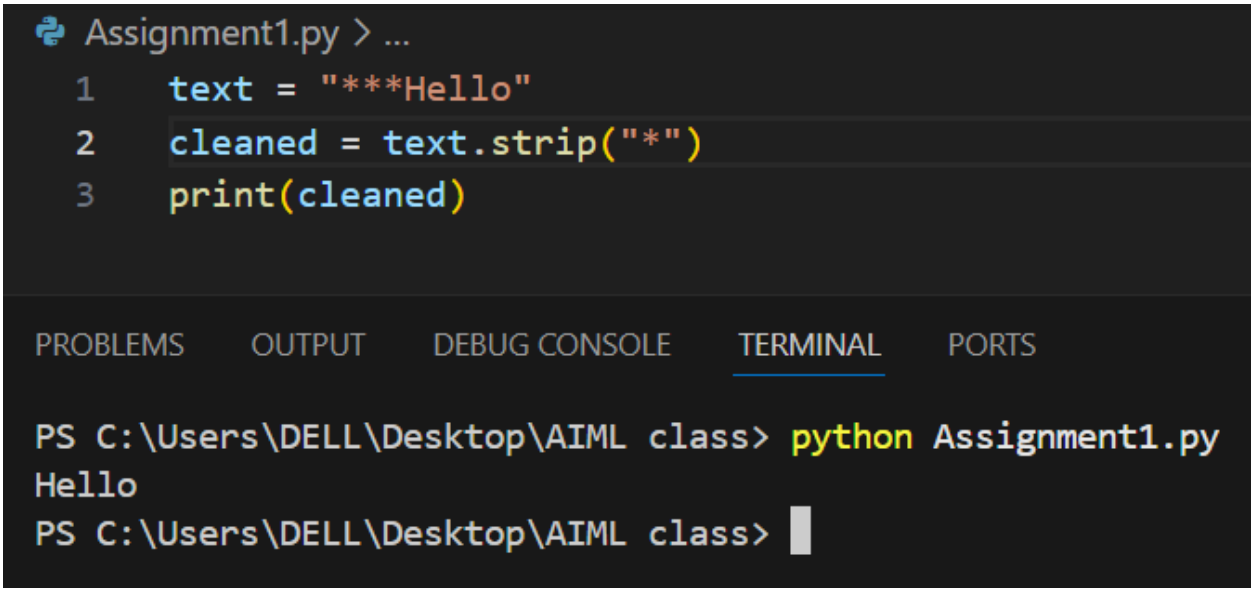
5. Multi-Step Cleaning: 16
6. In-Place Cleaning: 16
7. Clean and Count Unique Words: 17
8. Clean Dictionary Sentences: 17
9. Selective Character Removal: 18
10. Batch Clean and Sort: 18

Easy

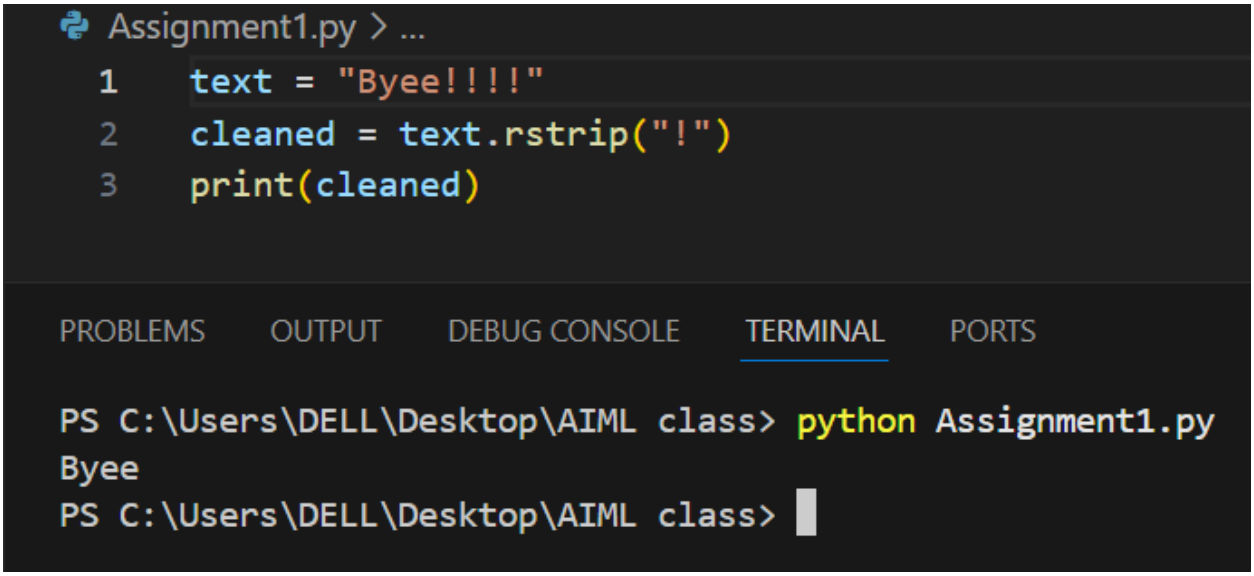
1. Remove Spaces:



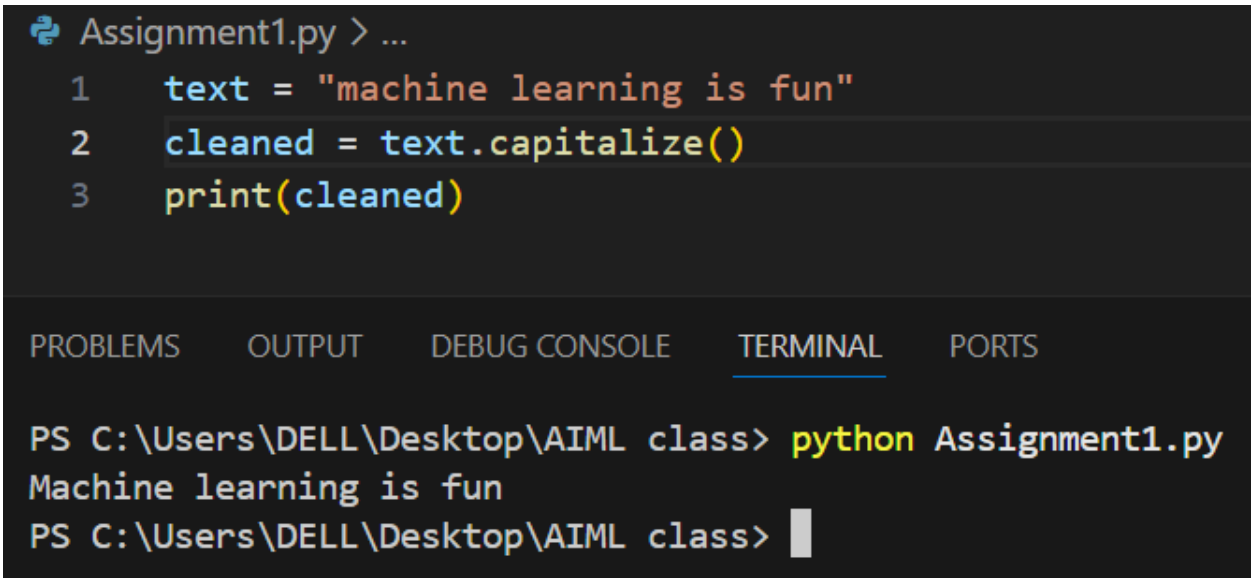
2. Remove Leading Characters:



3. Remove Trailing Characters:



4. Capitalize a Sentence:



5. Title Case a Name:

```
Assignment1.py > ...
1    text = "ankit shrestha"
2    cleaned = text.title()
3    print(cleaned)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
Ankit Shrestha
PS C:\Users\DELL\Desktop\AIML class>
```

6. Clean List of Names:

```
Assignment1.py > ...
1    names = [" Alisha", "Amit ", " Prasun "]
2    cleaned = [n.strip(" ") for n in names]
3    print(cleaned)
4

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
['Alisha', 'Amit', 'Prasun']
PS C:\Users\DELL\Desktop\AIML class>
```

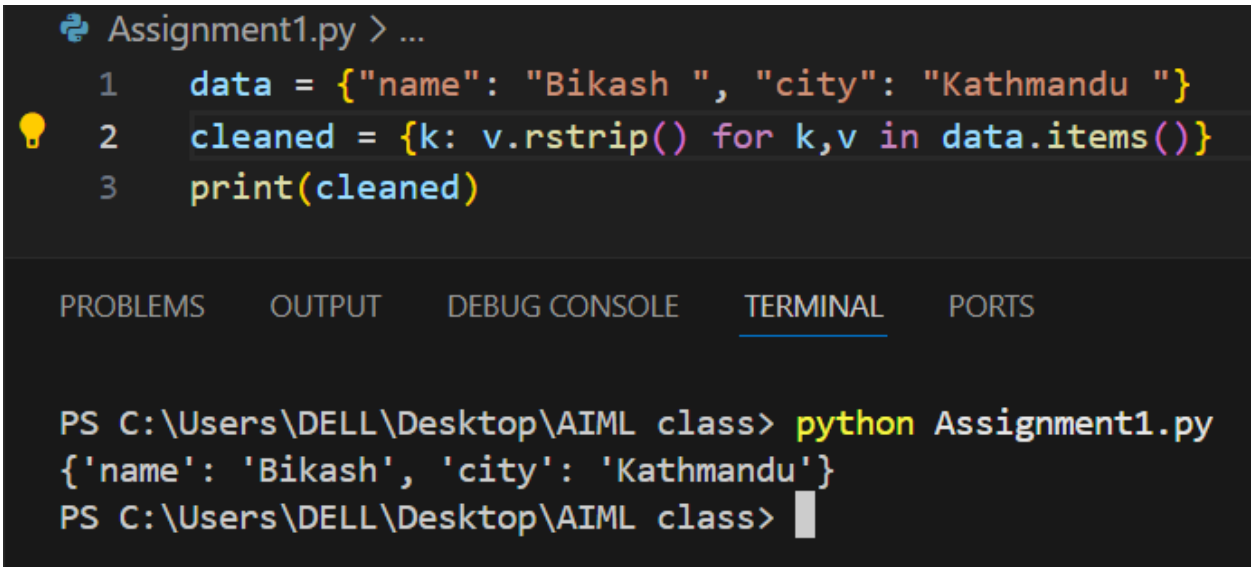
7. Remove Custom Characters:

```
Assignment1.py > ...
1    text = "#$^Hello@!#"
2    print(text.strip("#$^@!"))

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
Hello
PS C:\Users\DELL\Desktop\AIML class>
```

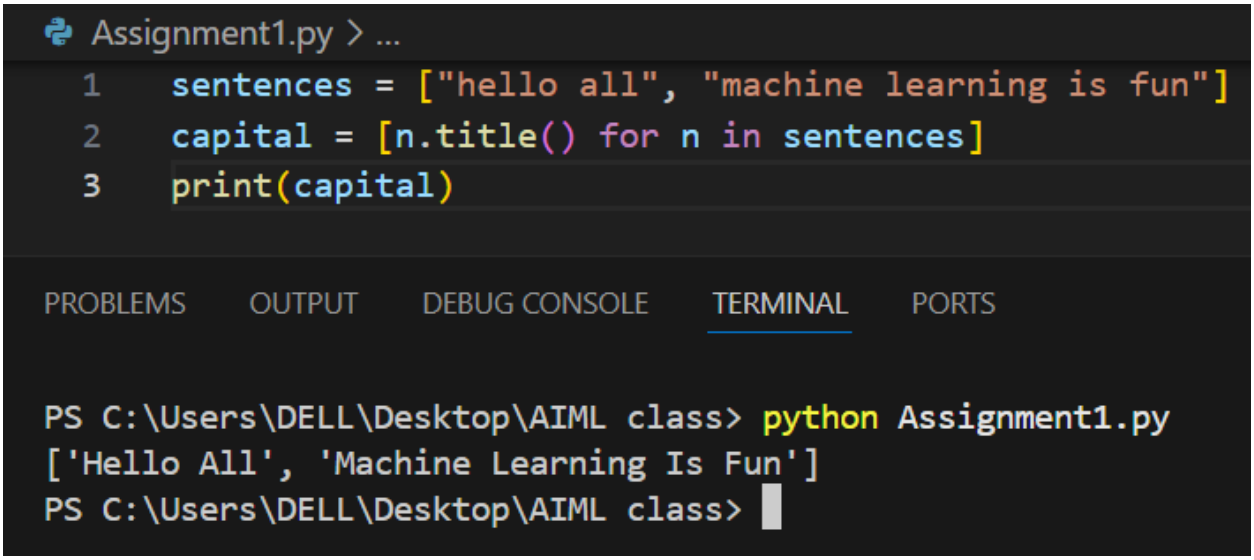
8. Capitalize All Names in List:

```
Assignment1.py > ...
1    names = ["alisha", "amit", "prasun"]
2    capital = [n.capitalize() for n in names]
3    print(capital)
4

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
['Alisha', 'Amit', 'Prasun']
PS C:\Users\DELL\Desktop\AIML class>
```

9. Clean Dictionary Values:

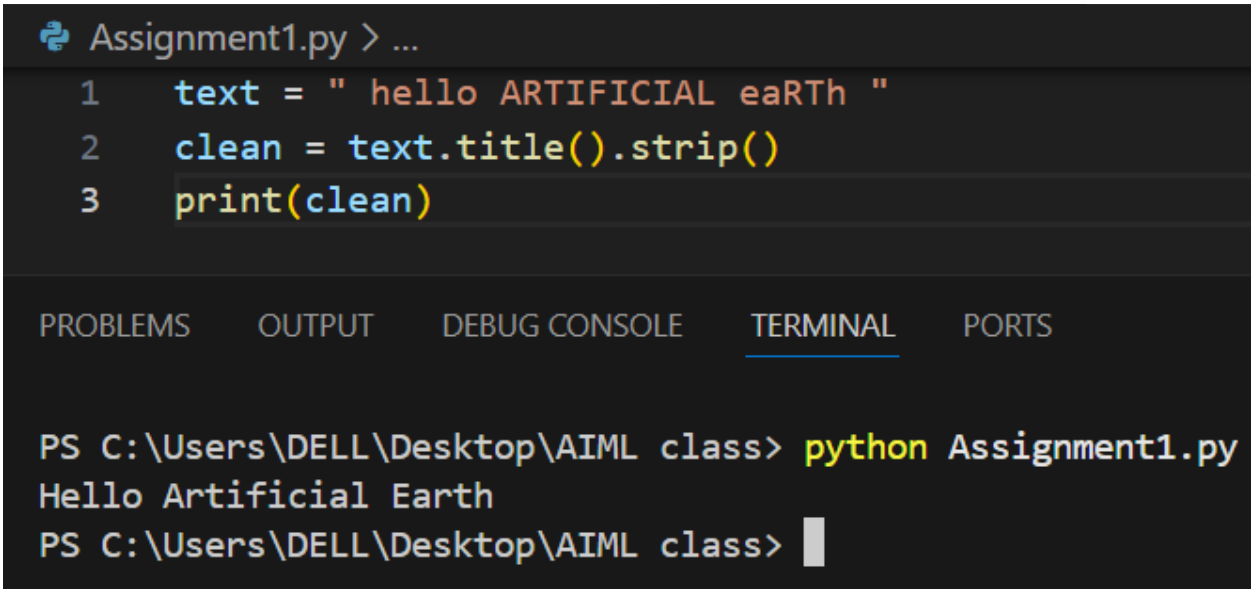


10. Title Case Sentences in List:

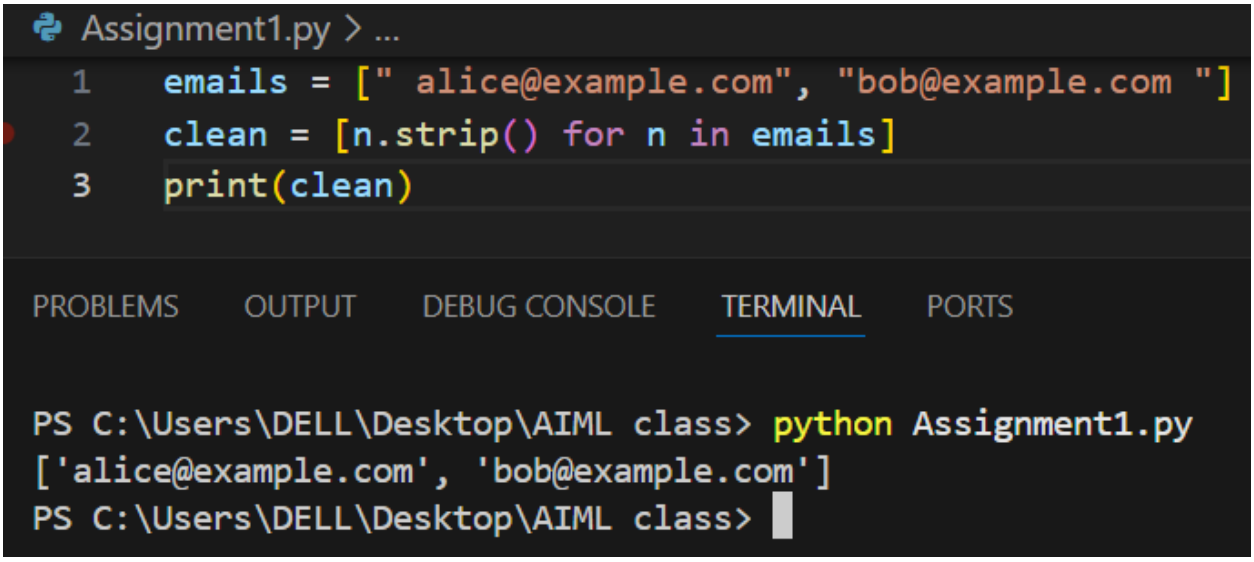


Intermediate

1. Clean and Title Case:



2. Clean List of Emails:



3. Remove Leading Numbers:

```
Assignment1.py > ...
1   text = "1245Hello"
2   clean = text.lstrip("1245")
3   print(clean)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
Hello
PS C:\Users\DELL\Desktop\AIML class>
```

4. Clean Nested List:

```
Assignment1.py > ...
1   nested = [[" apple", "banana "], [" cherry "]]
2   clean = [[n.strip() for n in lists] for lists in nested]
3   print(clean)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
[['apple', 'banana'], ['cherry']]
PS C:\Users\DELL\Desktop\AIML class>
```

5. Capitalize After Cleaning:

```
Assignment1.py > ...
1   text = "  prasun shrestha   "
2   clean = text.strip().capitalize()
3   print(clean)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
Prasun shrestha
PS C:\Users\DELL\Desktop\AIML class>
```

6. Clean Dictionary Keys:

```
Assignment1.py > ...
1   data = {"name_": "Alice", "age_": 30}
2   clean = {k.strip("_"): v for k,v in data.items()}
3   print(clean)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
{'name': 'Alice', 'age': 30}
PS C:\Users\DELL\Desktop\AIML class>
```

7. Clean and Deduplicate Names:

```
Assignment1.py > ...
1   names = [" alisha ", "AMIT", "amit", "ALISHA"]
2   clean = [n.strip().capitalize() for n in names]
3   deduplicated = list(set(clean))
4   print(deduplicated)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
['Amit', 'Alisha']
PS C:\Users\DELL\Desktop\AIML class>
```

8. Remove Multiple Characters:

```
Assignment1.py > ...
1   text = " ***-welcome-*** "
2   clean = text.strip(" *-")
3   print(clean)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
welcome
PS C:\Users\DELL\Desktop\AIML class>
```

9. Conditional Cleaning in List:

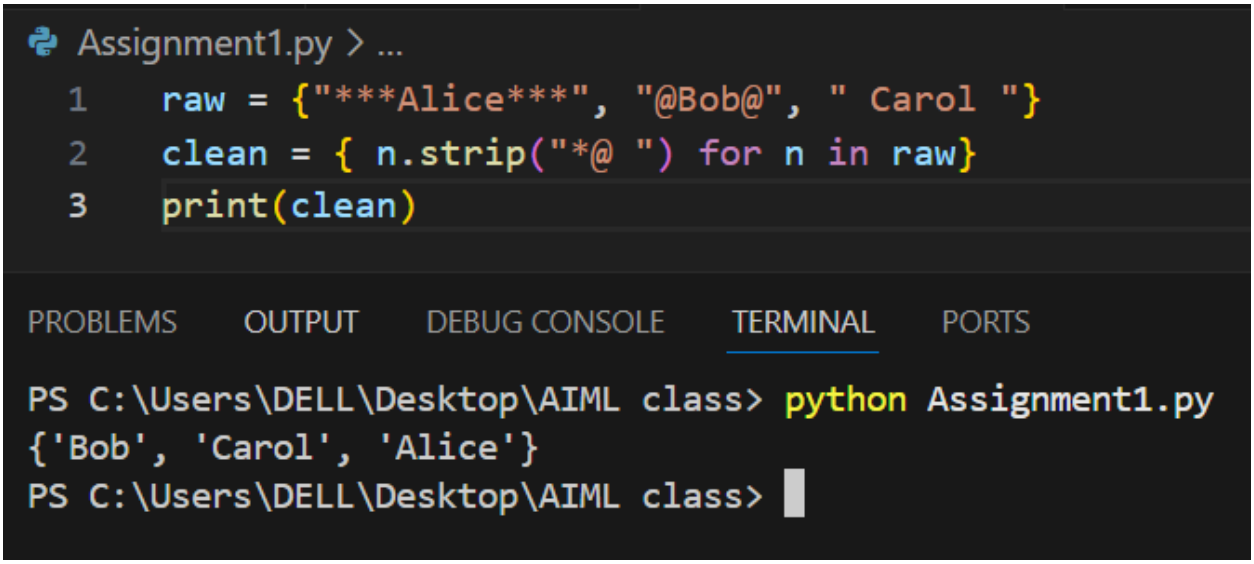
```
Assignment1.py > ...
1   tags = ["#python", "java", "#c++"]
2   clean = [ n.lstrip("#") if n.startswith("#") else n for n in tags ]
3   print(clean)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
['python', 'java', 'c++']
PS C:\Users\DELL\Desktop\AIML class>
```

10. Clean and Group by First Letter:

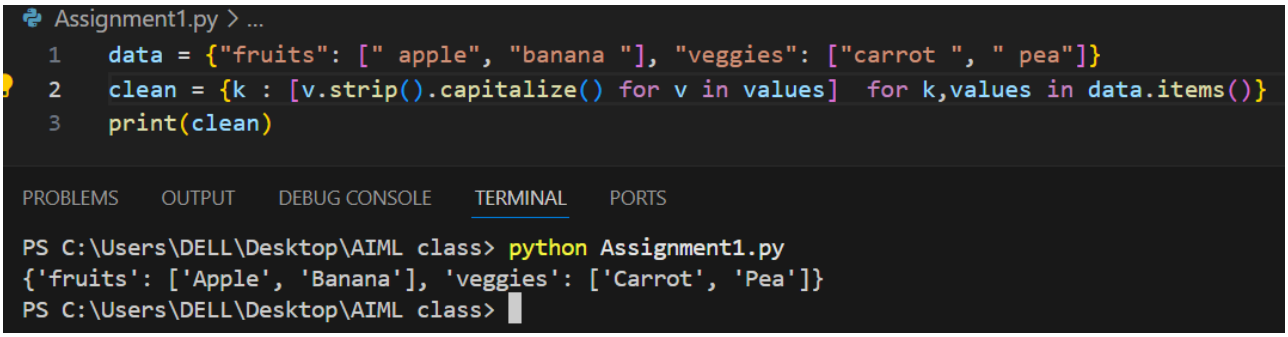
```
Assignment1.py > ...
1   products = [" red", "-ROAD", "aPPle", "area ", "read"]
2   clean = [ n.strip(" -").capitalize() for n in products]
3   group = {}
4   for n in clean:
5       key = n[0].upper()
6       group.setdefault(key, []).append(n)
7
8   print(group)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
{'R': ['Red', 'Road', 'Read'], 'A': ['Apple', 'Area']}
PS C:\Users\DELL\Desktop\AIML class>
```

Hard

1. Clean Set of Strings:



2. Complex Nested Cleaning:



3. Custom Title Function:

```
Assignment1.py > ...
1   def capitalize(sentance):
2       words = sentance.split(" ")
3       capital = [n.capitalize() for n in words]
4       return " ".join(c for c in capital)
5
6   print(capitalize("my name is ankit shrestha"))
7
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
My Name Is Ankit Shrestha
PS C:\Users\DELL\Desktop\AIML class>
```

4. Clean and Format Emails:

```
Assignment1.py > ...
1   emails = [" ALICE@Example.com", "bob@EXAMPLE.COM "]
2   clean = [n.strip().lower().capitalize() for n in emails]
3   print(clean)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
['Alice@example.com', 'Bob@example.com']
PS C:\Users\DELL\Desktop\AIML class>
```

5. Multi-Step Cleaning:

```
Assignment1.py > ...
1   text = "1444machine leaRnINg!!!!"
2   clean = text.strip("14!").title()
3   print(clean)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
Machine Learning
PS C:\Users\DELL\Desktop\AIML class>
```

6. In-Place Cleaning:

```
Assignment1.py > ...
1   names = [" alice ", "BOB", " charlie"]
2   for i in range(len(names)):
3       names[i] = names[i].strip().capitalize()
4   print(names)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
['Alice', 'Bob', 'Charlie']
PS C:\Users\DELL\Desktop\AIML class>
```

7. Clean and Count Unique Words:

```
Assignment1.py > ...
1   sentences = [" hello world ", "Hello python "]
2   clean = [n.strip().title() for n in sentences]
3   words = set()
4   for s in clean:
5       words.update(s.split())
6   print(words)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
{'Hello', 'World', 'Python'}
PS C:\Users\DELL\Desktop\AIML class>
```

8. Clean Dictionary Sentences:

```
Assignment1.py > ...
1   data = {"msg1": " hello world ", "msg2": "python is fun"}
2   clean = {k : v.strip().capitalize() for k,v in data.items()}
3   print(clean)

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
{'msg1': 'Hello world', 'msg2': 'Python is fun'}
PS C:\Users\DELL\Desktop\AIML class>
```

9. Selective Character Removal:

```
Assignment1.py > ...
1    text = "_-Hello-_"
2    print(text.strip("_- "))

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
Hello
PS C:\Users\DELL\Desktop\AIML class>
```

10. Batch Clean and Sort:

```
Assignment1.py > ...
1    codes = [" code3 ", "CODE1", " code2", "    code4"]
2    print(sorted([c.strip().title() for c in codes]))

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS C:\Users\DELL\Desktop\AIML class> python Assignment1.py
['Code1', 'Code2', 'Code3', 'Code4']
PS C:\Users\DELL\Desktop\AIML class>
```