Assignment No: 1.1.2 Name:Tony Prajapati Roll No. 79

Task:Write and execute a simple Python code cell (e.g., print "Hello, Data Science!").

```
[1]  print("Hello Data Science!")
```

```
Hello Data Science!
```

Assignment No: 1.1.3 Name:Tony Prajapati Roll No. 79

Task:Write a simple Python script that prints the current date and time.
Execute the script from the IDE.

```
import datetime
a=datetime.datetime.now()
print(a)
```

```
@gittonyp ➜/workspaces/DSLAB (main) $ /home/codespace/.python
/current/bin/python3 "/workspaces/DSLAB/Unit 1/1.1/1.1.3/time1
.py"
2024-09-03 04:35:36.297681
@gittonyp ➜/workspaces/DSLAB (main) $
```

Assignment No: 1.1.4 Name:Tony Prajapati Roll No. 79

Task:Write a brief comparison (200 words) between Jupyter Notebook and the chosen Python IDE based on your experience.

Jupyter Notebook and a traditional Python IDE like PyCharm have different strengths, which can be useful depending on what you're working on.

Jupyter Notebook is great for interactive work and experimentation. It lets you run code in small sections and see results immediately. This is really helpful for tasks like data analysis and visualization, where you can test ideas and adjust your code quickly. You can also add explanations and notes in markdown cells, which helps in creating reports and documenting your process.

On the other hand, PyCharm is better for managing larger and more complex projects. It offers features like code completion, debugging tools, and version control. These tools are useful when you need to write and maintain a lot of code, as they help you find and fix errors and keep track of changes. PyCharm also helps organize your files and handle dependencies, which is important for bigger projects.

In summary, Jupyter Notebook is best for interactive and exploratory tasks with immediate feedback, while PyCharm is better for more structured development and managing larger codebases. Both tools are valuable depending on the task you are working on.

Assignment No: 1.2.1 Name:Tony Prajapati Roll No. 79

Task: Write a 300-word essay on what Data Science is and its significance in today's world.

Data Science is a field that combines various techniques from statistics, mathematics, and computer science to extract valuable insights from data. It involves collecting, cleaning, analyzing, and interpreting large amounts of data to make informed decisions and solve complex problems. Data Science uses methods like machine learning, data mining, and predictive analytics to uncover patterns and trends that can guide strategic choices.

In today's world, Data Science is incredibly significant due to the vast amount of data generated every day. From social media interactions to financial transactions, data is everywhere, and businesses, governments, and organizations are all looking to harness this data for various purposes. For example, companies use data science to understand customer behavior, optimize marketing strategies, and improve products and services. By analyzing data, businesses can make more accurate predictions about market trends and customer needs, which helps them stay competitive.

Moreover, Data Science plays a crucial role in sectors like healthcare and finance. In healthcare, it helps in predicting disease outbreaks, personalizing treatments, and improving patient care through data analysis. In finance, it is used for risk assessment, fraud detection, and investment strategies. The ability to analyze large datasets quickly and accurately enables organizations to make better decisions and respond to challenges more effectively.

Data Science also has a growing impact on everyday life. For example, recommendation systems used by streaming services and e-commerce sites are powered by data science, offering personalized suggestions based on user behavior. In summary, Data Science is a powerful tool that drives innovation and efficiency across various fields, making it an essential component of modern society and the global economy.

Assignment No: 1.2.2 Name:Tony Prajapati Roll No. 79

Task: Identify three real-world applications of Data Science and explain each in 200 words.

1. Healthcare Diagnostics:

In the healthcare sector, Data Science is transforming how diseases are diagnosed and treated. Machine learning algorithms analyze medical records, imaging data, and genetic information to identify patterns that may be invisible to human doctors. For instance, predictive models can forecast patient outcomes, helping doctors to catch conditions like diabetes or cancer earlier. Radiologists use image recognition techniques to detect anomalies in X-rays and MRIs more accurately and quickly. Data Science also supports personalized medicine, where treatments are tailored to an individual's genetic profile and medical history. This approach not only improves the accuracy of diagnoses but also enhances treatment effectiveness and reduces side effects, leading to better patient outcomes and more efficient healthcare systems.

2. Financial Fraud Detection:

In the financial industry, Data Science plays a crucial role in detecting and preventing fraud. Financial institutions use advanced algorithms to analyze transaction patterns and identify unusual activities that may indicate fraudulent behavior. Machine learning models are trained on historical transaction data to recognize the signs of fraud, such as unusual spending patterns or anomalous transaction sizes. When a potential fraud is detected, these systems can flag the transaction for further review or automatically block it, reducing financial losses and enhancing security. By continuously learning from new data, these systems improve over time, adapting to evolving fraud tactics and ensuring better protection for consumers and financial institutions.

3. E-Commerce Recommendations:

E-commerce platforms leverage Data Science to enhance user experience through personalized recommendations. By analyzing browsing history, purchase behavior, and user preferences, recommendation algorithms suggest products that a user is likely to be interested in. For example, if a customer frequently searches for fitness gear, the platform might recommend related items like workout clothes or supplements. These recommendations are powered by collaborative filtering and content-based algorithms that analyze vast amounts of data to predict what products a user might prefer. This personalization increases user engagement and sales, as customers are more likely to find products that match their interests and needs, leading to a more satisfying shopping experience and increased revenue for the retailer.

Assignment No: 1.2.3 Name:Tony Prajapati Roll No. 79

Task: List five common tools or libraries used in Data Science. Provide a brief description of each tool/library.

1. Pandas:
Pandas is a powerful library for data manipulation and analysis in Python. It provides data structures like DataFrames and Series that make it easy to work with structured data. Pandas offers functions for reading and writing data, handling missing values, and performing operations like merging, grouping, and aggregating. It's widely used for data cleaning and preprocessing tasks, making it a fundamental tool for any data science workflow.
2. NumPy:
NumPy (Numerical Python) is a library that provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays. It forms the foundation for many other data science libraries by offering efficient array operations and mathematical functions, which are essential for numerical computations and data manipulation.
3. Scikit-learn:
Scikit-learn is a machine learning library for Python that provides simple and efficient tools for data mining and data analysis. It includes a wide range of algorithms for classification, regression, clustering, and dimensionality reduction. Scikit-learn also offers tools for model selection and evaluation, making it a go-to library for implementing and testing machine learning models.
4. Matplotlib:
Matplotlib is a plotting library for Python that enables users to create static, animated, and interactive visualizations. It provides a flexible interface for generating various types of plots, including line charts, bar graphs, histograms, and scatter plots. Matplotlib is often used to visualize data, explore relationships, and present results in a clear and understandable way.
5. TensorFlow:
TensorFlow is an open-source machine learning framework developed by Google. It is widely used for building and deploying deep learning models. TensorFlow supports a range of tasks from training neural networks to deploying models in production environments. It provides tools for designing complex neural networks and offers extensive support for numerical computations and parallel processing, making it suitable for large-scale machine learning projects.

Assignment No: 1.2.4 Name:Tony Prajapati Roll No. 79

Task: Analyze a case study where Data Science was used to solve a significant problem. Summarize the case study in 300 words.

Predicting Patient Readmissions with Data Science

Overview:

A prominent healthcare provider, Cleveland Clinic, faced a significant challenge with high patient readmission rates, which not only affected patient outcomes but also incurred substantial costs. To address this issue, the clinic implemented a data science approach to predict which patients were at risk of being readmitted within 30 days of discharge. The goal was to reduce readmission rates by intervening early and improving patient care.

Approach:

The Cleveland Clinic employed a machine learning model to analyze historical patient data, including demographics, medical history, hospital stay details, and follow-up care records. They used features such as previous admissions, chronic conditions, and discharge summaries to build a predictive model. The team utilized Scikit-learn to develop and test various algorithms, eventually selecting a logistic regression model due to its effectiveness in predicting binary outcomes.

Data preprocessing involved cleaning and transforming data to handle missing values and normalize features. The model was trained on a large dataset of patient records and validated using cross-validation techniques to ensure accuracy and robustness.

Results:

The predictive model achieved high accuracy in identifying patients at risk of readmission. The Cleveland Clinic implemented targeted interventions based on model predictions, such as enhanced discharge planning, follow-up appointments, and personalized patient education. As a result, the clinic observed a significant reduction in readmission rates, improving patient outcomes and reducing costs associated with readmissions.

Impact:

This data science-driven approach not only enhanced patient care but also demonstrated the potential of predictive analytics in healthcare. By leveraging historical data and advanced modeling techniques, the Cleveland Clinic successfully addressed a critical issue, showcasing how data science can lead to actionable insights and operational improvements in the healthcare industry.

Assignment No: 1.3.1.1 Name:Tony Prajapati Roll No. 79

Write a Python program to find the factorial of a number..

```
def fact(n):
    if n==1:
        return 1;
    return n*fact(n-1)
i=int(input())
print(fact(i))
```

```
@gittonyp ➜/workspaces/DSLAB (main) $ /home/codespac
/current/bin/python3 "/workspaces/DSLAB/Unit 1/1.3/1.
rial.py"
5
120
@gittonyp ➜/workspaces/DSLAB (main) $
```

Assignment No: 1.3.1.2 Name:Tony Prajapati Roll No. 79

Write a Python program to check if a number is prime.

1 > 1.3 > 1.3.1 > Prime.py > ...

```
import math
n=int(input("Enter "))
flag=0
if n<2:
    flag=1

for i in range(2,int(math.sqrt(n)+1)):
    if n%i==0:
        flag=1

if flag==0:
    print("Prime")
else:
    print("Not Prime")
```

```
@gittonyp →/workspaces/DSLAB (main) $ /home/
/current/bin/python3 "/workspaces/DSLAB/Unit
.py"
Enter 10
Not Prime
@gittonyp →/workspaces/DSLAB (main) $ /home/
/current/bin/python3 "/workspaces/DSLAB/Unit
.py"
Enter 7
Prime
@gittonyp →/workspaces/DSLAB (main) $
```

Assignment No: 1.3.2.1 Name:Tony Prajapati Roll No. 79

Write a Python program using if-else to determine whether a number is positive, negative, or zero.

```
n=int(input("Enter "))

if n==0:
    print("Zero")
elif n>0:
    print("Positive")
else:
    print("Negative")
```

```
@gittonyp →/workspaces/DSLAB (main) $ /home
/current/bin/python3 "/workspaces/DSLAB/Unit
ive negative zero.py"
Enter 1
Positive
@gittonyp →/workspaces/DSLAB (main) $ /home
/current/bin/python3 "/workspaces/DSLAB/Unit
ive negative zero.py"
Enter -1
Negative
@gittonyp →/workspaces/DSLAB (main) $ /home
/current/bin/python3 "/workspaces/DSLAB/Unit
ive negative zero.py"
Enter 0
Zero
@gittonyp →/workspaces/DSLAB (main) $
```

Assignment No: 1.3.2.2 Name:Tony Prajapati Roll No. 79

Write a Python program using a loop to print the first 10 Fibonacci numbers.

```
n=0
n2=1

for i in range(0,10):
    print(n)
    temp=n
    n = n2
    n2=temp+n2
```

```
@gittonyp →/workspaces/DS
/current/bin/python3 "/wor
naci.py"
0
1
1
2
3
5
8
13
21
34
```

Assignment No: 1.3.3.1 Name:Tony Prajapati Roll No. 79

Create a Python function to calculate the area of a circle.

```
import math
n=int(input("Enter"))
print(n*math.pi*n)
```

```
@gittonyp →/workspaces/DSLAB (main) $ /h
/current/bin/python3 "/workspaces/DSLAB/U
of c.py"
Enter4
50.26548245743669
@gittonyp →/workspaces/DSLAB (main) $
```

Assignment No: 1.3.3.2 Name:Tony Prajapati Roll No. 79

Create a Python function that takes a list of numbers and returns the largest number in the list.

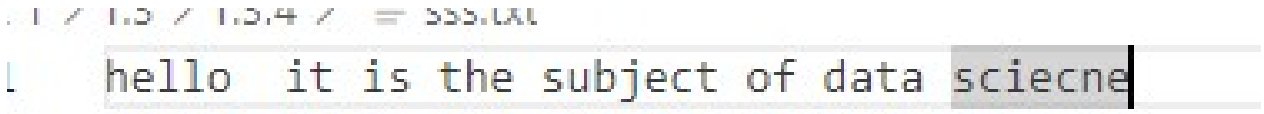
```
ll2=[1,156,13,56,77,966,138]
ll2=sorted(ll2)
print(ll2[-1])
```

```
@gittonyp →/workspac
/current/bin/python3
st.py"
966
@gittonyp →/workspac
```

Assignment No: 1.3.4.1 Name:Tony Prajapati Roll No. 79

Write a Python program to read a text file and count the number of words in it.

```
tx=open("sss.txt","rt")
print(len(tx.read().split()))
```



OUTPUT: 8

Assignment No: 1.4.1.1 Name:Tony Prajapati Roll No. 79

Write a Python program to find the second largest number in a list.

```
ll = (1, 2, -8, 0,564,897)
big = -999999
small=-99999
for i in range(len(ll)):
    if int(ll[i]) > big:
        big = int(ll[i])

for i in range(len(ll)):
    if int(ll[i]) > small and big!=ll[i]:
        small = int(ll[i])

print(small)
```

```
/current/bin/python3 "/wo
y"
564
@gittonyp →/workspaces/D
```

Assignment No: 1.4.1.2 Name:Tony Prajapati Roll No. 79

Write a Python program to merge two tuples into a dictionary.

```
ll = (1, 2, -8, 0)
ll2 = ("a", "b", "C", "d")

di={}

for i in range(0,len(ll)):
    di.update({ll[i]:ll2[i]})

print(di)
```

```
@gittonyp ➜/workspaces/DSLAB (main) $ /home/c
/current/bin/python3 "/workspaces/DSLAB/Unit 1
y"
{1: 'a', 2: 'b', -8: 'C', 0: 'd'}
@gittonyp ➜/workspaces/DSLAB (main) $
```

Assignment No: 1.4.2.1 Name:Tony Prajapati Roll No. 79

Write a Python program to count the frequency of each word in a given string using a dictionary.

```
str=input()
str=str.lower
str=str.split()
dicp={}

for i in str:
    if dicp.get(i)==None:
        dicp.update({i:1})
    else:
        dicp.update({i:dicp.get(i)+1})

print(dicp)
```

```
@gittonyp →/workspaces/DSLAB (main) $ /home/codespace/.
/current/bin/python3 "/workspaces/DSLAB/Unit 1/1.4 new/1
ict2.py"
hello my name is tony
{'hello': 1, 'my': 1, 'name': 1, 'is': 1, 'tony': 1}
```

Assignment No: 1.4.2.2 Name:Tony Prajapati Roll No. 79

Write a Python program to find the union and intersection of two sets.

```
ll={1,10,156,74,19}
ll2={13,56,77,966,138}
print(ll & ll2)
print(ll | ll2)
```

```
@gittonyp ➜/workspaces/DSLAB (main) $ /home
/current/bin/python3 "/workspaces/DSLAB/Unit
et op.py"
set()
{1, 966, 74, 10, 138, 77, 13, 19, 56, 156}
@gittonyp ➜/workspaces/DSLAB (main) $
```

Create a DataFrame from a dictionary of lists.

```
import pandas as pd
import numpy as np



data = {
    "calories": [420, 380, 390],
    "duration": [50, 40, 45]
    }

a=pd.DataFrame(data)
print(a)
```

```
@gittonyp →/workspaces/DSLAB (main) :
/current/bin/python3 "/workspaces/DSL
py"
   calories  duration
0       420        50
1       380        40
2       390        45
@gittonyp →/workspaces/DSLAB (main) :
```

Perform basic operations like selecting rows/columns, filtering, and sorting.

```
import pandas as pd
import numpy as np
from sklearn import datasets
iris = datasets.load_iris()
pds=pd.DataFrame(iris.data)
a=pds.sort_values(by=2)
print(a)
print(a[0])
print(a.head(3))
```

```
      0    1    2    3
22   4.6  3.6  1.0  0.2
13   4.3  3.0  1.1  0.1
14   5.8  4.0  1.2  0.2
35   5.0  3.2  1.2  0.2
16   5.4  3.9  1.3  0.4
..   ...  ...  ...  ...
131  7.9  3.8  6.4  2.0
105  7.6  3.0  6.6  2.1
117  7.7  3.8  6.7  2.2
122  7.7  2.8  6.7  2.0
118  7.7  2.6  6.9  2.3
```

```
22     4.6
13     4.3
14     5.8
35     5.0
16     5.4
      ...
131    7.9
105    7.6
117    7.7
122    7.7
118    7.7
...
```

```
      0    1    2    3
22   4.6  3.6  1.0  0.2
13   4.3  3.0  1.1  0.1
14   5.8  4.0  1.2  0.2
```

Assignment No: 1.5.2.1 Name:Tony Prajapati Roll No. 79

Create a NumPy array and perform element-wise operations.

```
import numpy as np

a=np.array([1,2,3,4,5,6,7,8,9,10])
print(a)
a=a+10
print(a)
a=a-10
print(a)
a=a*10
print(a)
a=a/10
print(a)
a=a%10
print(a)
```

```
@gittonyp →/workspaces/DSLAB (main) $ /home
on/current/bin/python3 "/workspaces/DSLAB/Un
um1.py"
[ 1  2  3  4  5  6  7  8  9 10]
[11 12 13 14 15 16 17 18 19 20]
[ 1  2  3  4  5  6  7  8  9 10]
[ 10  20  30  40  50  60  70  80  90 100]
[ 1.  2.  3.  4.  5.  6.  7.  8.  9. 10.]
[1. 2. 3. 4. 5. 6. 7. 8. 9. 0.]
@gittonyp →/workspaces/DSLAB (main) $
```

Use NumPy to perform matrix multiplication.

```
import numpy as np

a=[[2,1,3,5],
[0,5,2,7],
[1,1,2,9]]
b=[[1]*3]*4

a=np.array(a)
b=np.array(b)
c=np.matmul(a,b)
print(c)
```

```
@gittonyp →/workspaces/DSLAB (main)
on/current/bin/python3 "/workspaces/
um2.py"
[[11 11 11]
 [14 14 14]
 [13 13 13]]
@gittonyp →/workspaces/DSLAB (main)
```