Theory Activity – 1

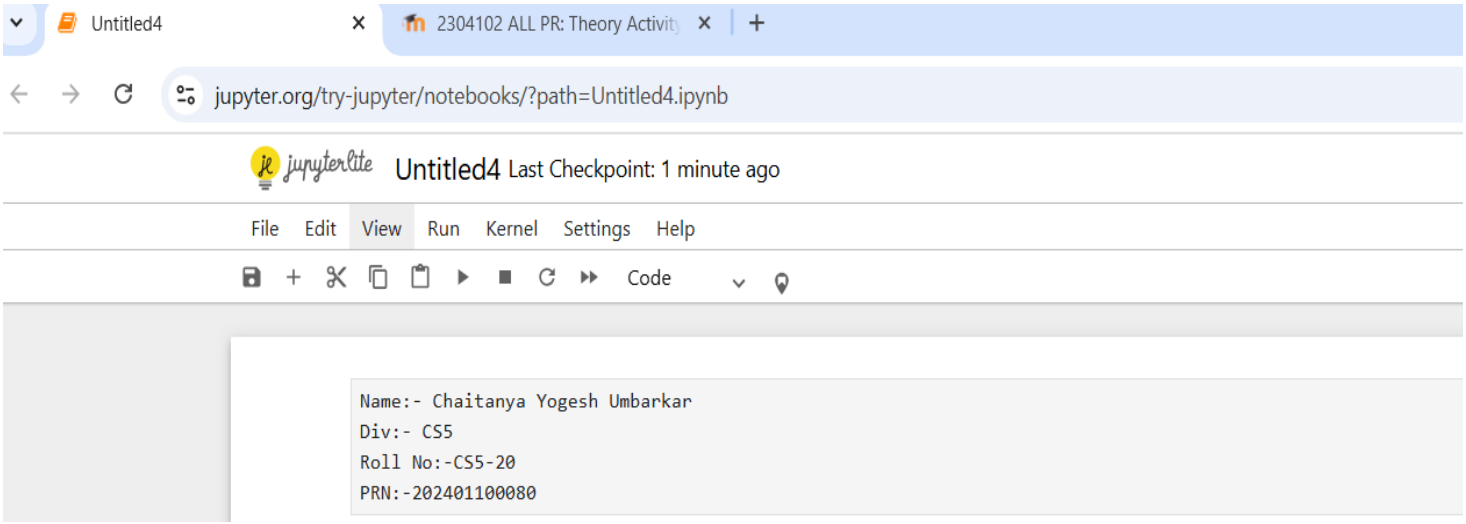
Name: - Chaitanya Yogesh Umbarkar

Div: - CS5

Roll: - CS5-20

PRN: - 202401100080

Dataset: - Paper Review



Creation of dataset on Paper Review:

[21]:
```
import numpy as np
import pandas as pd


pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', 1000)
pd.set_option('display.colheader_justify', 'center')


data = {
    'Paper_ID': [1,2,3,4,5,6,7,8,9,10],
    'Reviewer': ['Alice', 'Bob', 'Charlie', 'David', 'Eva', 'Frank', 'Grace', 'Heidi', 'Ivan', 'Judy'],
    'Paper_Score': [8.5, 7.2, 9.0, 5.5, 6.8, 8.0, 9.5, 7.0, 6.5, 8.8],
    'No_of_Pages': [12, 8, 15, 7, 10, 9, 16, 11, 5, 14],
    'Submission_Year': [2023, 2022, 2024, 2023, 2021, 2024, 2022, 2021, 2023, 2022],
    'Field': ['AI', 'Data Science', 'Cybersecurity', 'IoT', 'Cloud Computing', 'AI', 'Robotics', 'Data Science', 'Cybersecurity', 'AI'],
    'Revision_Required': ['Yes', 'No', 'No', 'Yes', 'Yes', 'No', 'No', 'Yes', 'Yes', 'No']
}

df = pd.DataFrame(data)
scores = np.array(df['Paper_Score'])
```

Questions On Numpy with Solution:

```
print("-----Numpy questions with solution-----")
print()
#1.Find the average score of all papers.
print("1. Average Paper Score:", np.mean(scores))

#2.Find the maximum paper score.
print("2. Maximum Paper Score:", np.max(scores))

#3.Find the minimum paper score.
print("3. Minimum Paper Score:", np.min(scores))

#4.Find the standard deviation of the paper scores.
print("4. Standard Deviation of Paper Scores:", np.std(scores))

#5.Find the sum of all paper scores.
print("5. Sum of Paper Scores:", np.sum(scores))

#6.Find how many papers have a score greater than 8.
print("6. Number of Papers with Score > 8:", np.sum(scores > 8))

#7.Find the median score of the papers.
print("7. Median of Paper Scores:", np.median(scores))

#8.Find the total number of pages submitted.
print("8. Total Number of Pages:", np.sum(df['No_of_Pages']))

#9.Find the range (max - min) of paper scores.
print("9. Range of Paper Scores:", np.max(scores) - np.min(scores))

#10.Count how many papers were submitted in 2023.
print("10. Papers Submitted in 2023:", np.sum(df['Submission_Year'] == 2023))
print()
```

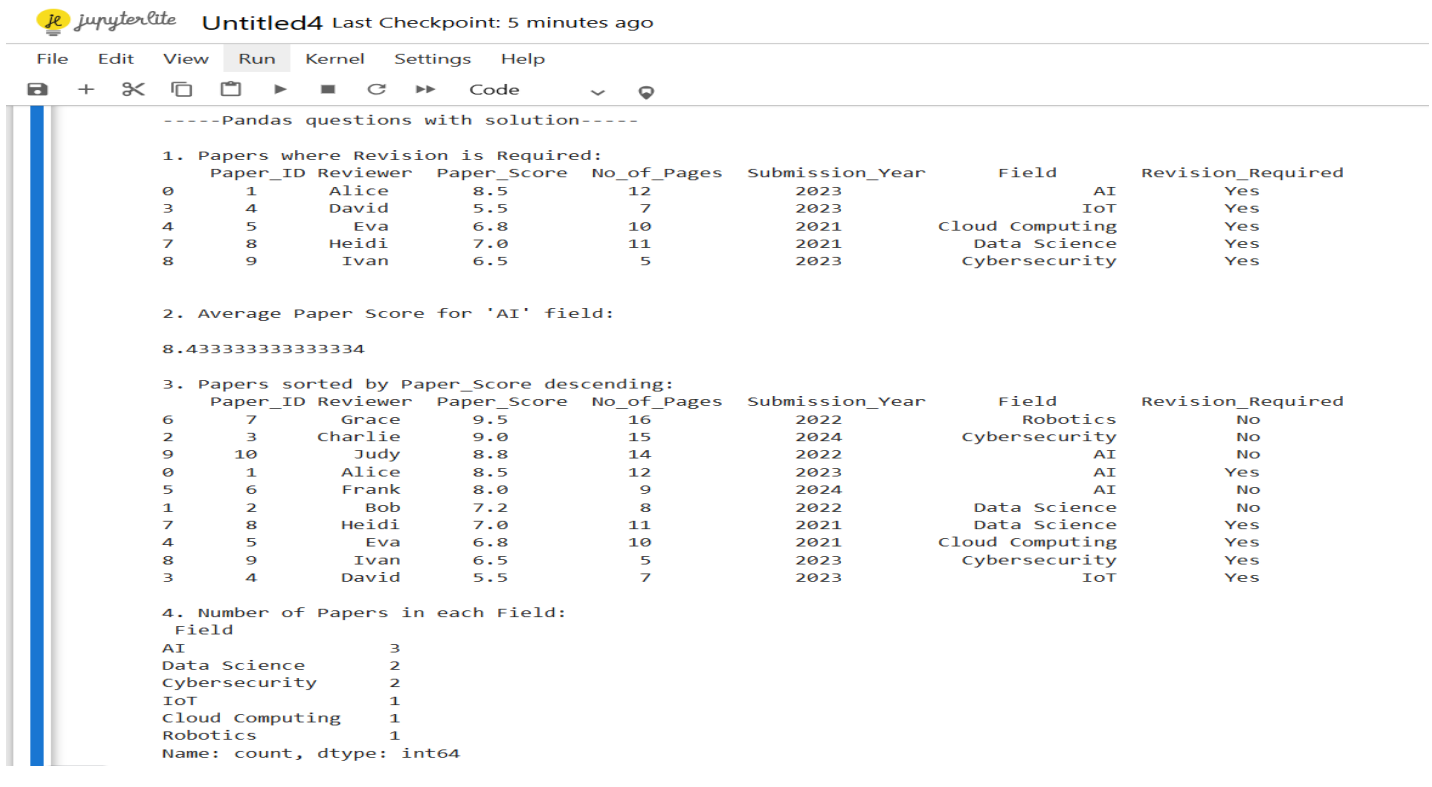
Output of the Applied Solutions:

```
-----Numpy questions with solution-----

1. Average Paper Score: 7.68
2. Maximum Paper Score: 9.5
3. Minimum Paper Score: 5.5
4. Standard Deviation of Paper Scores: 1.212270596855339
5. Sum of Paper Scores: 76.8
6. Number of Papers with Score > 8: 4
7. Median of Paper Scores: 7.6
8. Total Number of Pages: 107
9. Range of Paper Scores: 4.0
10. Papers Submitted in 2023: 3
```

Questions On Pandas with Solution:

```
print("-----Pandas questions with solution-----")
print()

#1.Display all papers where Revision_Required is 'Yes'.
print("1. Papers where Revision is Required:\n", df[df['Revision_Required'] == 'Yes'])
print()

#2.Find the average Paper_Score for the AI field.
print("\n2. Average Paper Score for 'AI' field:\n")
print(df[df['Field'] == 'AI']['Paper_Score'].mean())

#3.List the papers sorted by Paper_Score descending.
print("\n3. Papers sorted by Paper_Score descending:\n", df.sort_values(by='Paper_Score', ascending=False))

#4.Show the count of papers in each Field.
print("\n4. Number of Papers in each Field:\n", df['Field'].value_counts())

#5.Find the Reviewer name who reviewed the paper with the highest score.
highest_score_reviewer = df[df['Paper_Score'] == df['Paper_Score'].max()]['Reviewer'].values[0]
print("\n5. Reviewer with Highest Paper Score:\n", highest_score_reviewer)


#6.Find all papers submitted before 2023.
print("\n6. Papers submitted before 2023:\n", df[df['Submission_Year'] < 2023])

#7.Add a new column 'Score_Category' as 'High' if Paper_Score > 8 else 'Low'.
df['Score_Category'] = df['Paper_Score'].apply(lambda x: 'High' if x > 8 else 'Low')
print("\n7. Dataset with 'Score_Category' Column Added:\n", df)

#8.Replace all 'Yes'/'No' in Revision_Required with True/False.
pd.set_option('future.no_silent_downcasting', True)
print("\n8. Dataset after Replacing 'Yes/No' with 'True/False':\n")
df['Revision_Required'] = df['Revision_Required'].replace({'Yes': True, 'No': False}).infer_objects(copy=False)
print(df)
```

Output of the Applied Solutions:




```
-----Pandas questions with solution-----

1. Papers where Revision is Required:
   Paper_ID Reviewer  Paper_Score  No_of_Pages  Submission_Year       Field Revision_Required
0         1    Alice          8.5           12             2023          AI              Yes
3         4    David          5.5            7             2023         IoT              Yes
4         5      Eva          6.8           10             2021  Cloud Computing         Yes
7         8    Heidi          7.0           11             2021     Data Science         Yes
8         9     Ivan          6.5            5             2023    Cybersecurity         Yes


2. Average Paper Score for 'AI' field:

8.433333333333334

3. Papers sorted by Paper_Score descending:
   Paper_ID Reviewer  Paper_Score  No_of_Pages  Submission_Year       Field Revision_Required
6         7    Grace          9.5           16             2022    Robotics              No
2         3  Charlie          9.0           15             2024    Cybersecurity         No
9        10    Judy           8.8           14             2022          AI              No
0         1    Alice          8.5           12             2023          AI             Yes
5         6    Frank          8.0            9             2024          AI              No
1         2      Bob          7.2            8             2022     Data Science         No
7         8    Heidi          7.0           11             2021     Data Science         Yes
4         5      Eva          6.8           10             2021  Cloud Computing         Yes
8         9     Ivan          6.5            5             2023    Cybersecurity         Yes
3         4    David          5.5            7             2023         IoT              Yes

4. Number of Papers in each Field:
 Field
AI                 3
Data Science       2
Cybersecurity      2
IoT                1
Cloud Computing    1
Robotics           1
Name: count, dtype: int64
```

jupyterlite Untitled4 Last Checkpoint: 5 minutes ago

File Edit View Run Kernel Settings Help

Code JupyterLab

```
5. Reviewer with Highest Paper Score:
 Grace

6. Papers submitted before 2023:
    Paper_ID Reviewer  Paper_Score  No_of_Pages  Submission_Year      Field  Revision_Required
1          2      Bob          7.2            8             2022  Data Science                No
4          5      Eva          6.8           10             2021  Cloud Computing             Yes
6          7    Grace          9.5           16             2022  Robotics                    No
7          8    Heidi          7.0           11             2021  Data Science                Yes
9         10    Judy           8.8           14             2022  AI                          No

7. Dataset with 'Score_Category' Column Added:
    Paper_ID Reviewer  Paper_Score  No_of_Pages  Submission_Year      Field  Revision_Required Score_Category
0          1    Alice          8.5           12             2023  AI                          Yes             High
1          2      Bob          7.2            8             2022  Data Science                No              Low
2          3  Charlie          9.0           15             2024  Cybersecurity               No              High
3          4    David          5.5            7             2023  IoT                         Yes             Low
4          5      Eva          6.8           10             2021  Cloud Computing             Yes             Low
5          6    Frank          8.0            9             2024  AI                          No              Low
6          7    Grace          9.5           16             2022  Robotics                    No              High
7          8    Heidi          7.0           11             2021  Data Science                Yes             Low
8          9     Ivan          6.5            5             2023  Cybersecurity               Yes             Low
9         10    Judy           8.8           14             2022  AI                          No              High

8. Dataset after Replacing 'Yes/No' with 'True/False':

   Paper_ID Reviewer  Paper_Score  No_of_Pages  Submission_Year      Field  Revision_Required Score_Category
0          1    Alice          8.5           12             2023  AI                          True            High
1          2      Bob          7.2            8             2022  Data Science                False           Low
2          3  Charlie          9.0           15             2024  Cybersecurity               False           High
3          4    David          5.5            7             2023  IoT                         True            Low
4          5      Eva          6.8           10             2021  Cloud Computing             True            Low
5          6    Frank          8.0            9             2024  AI                          False           Low
6          7    Grace          9.5           16             2022  Robotics                    False           High
7          8    Heidi          7.0           11             2021  Data Science                True            Low
8          9     Ivan          6.5            5             2023  Cybersecurity               True            Low
9         10    Judy           8.8           14             2022  AI                          False           High
```

```
9. Average Pages for Papers Requiring Revision:

9.0

10. Pivot Table of Average Paper_Score by Field:
                Paper_Score
Field
AI                 8.433333
Cloud Computing    6.800000
Cybersecurity      7.750000
Data Science       7.100000
IoT                5.500000
Robotics           9.500000
```