# **Theory Activity No. 1**

Name: Karan Ugale

Subject : EDS

**PRN**: 202401070135

**Roll No.** : ET1-62

Batch: E14

## Import Required Libraries

```
import pandas as pd
import numpy as np
```

#### Dataset

```
data = {
                                                                                                                                                                                                                                                                                                                                                                             ↑ ↓ ♦ © 🗏 💠 🗓 🔟 🗜
                            'Paper_ID': range(1, 11),
                            'Title': ['Paper A', 'Paper B', 'Paper C', 'Paper D', 'Paper E', 'Paper F', 'Paper G', 'Paper H', 'Paper I', 'Paper J'],
'Author(s)': ['Author1', 'Author2', 'Author1', 'Author3', 'Author4', 'Author5', 'Author2', 'Author6', 'Author7', 'Author3'],
'Submission_Date': pd.to_datetime(['2025-03-30', '2025-04-01', '2025-04-02', '2025-03-28', '2025-03-31', '2025-04-03', '2025-03-27', '2025-04-02', '2025-03-28', '2025-03-31', '2025-04-03', '2025-03-27', '2025-04-02', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '2025-04-03', '
                            'Review_Score': [8, 7, 9, 6, 5, 8, 7, 9, 6, 5],
'Reviewer_Comments': ['Good work', 'Needs improvement', 'Excellent', 'Average', 'Poor', 'Well done', 'Good novelty', 'Very good', 'Could be 'Decision': ['Accepted', 'Rejected', 'Rejected', 'Rejected', 'Accepted', 'Accepted', 'Accepted', 'Rejected'],
'Keywords': ['AI, ML', 'IoT', 'AI, Robotics', 'ML, Security', 'Blockchain', 'IoT, AI', 'ML', 'Cybersecurity', 'IoT, Blockchain', 'Security'
'No_of_Pages': [8, 10, 12, 9, 11, 7, 8, 10, 13, 9],
                             'Plagiarism_Score': [10, 35, 5, 20, 40, 8, 15, 10, 50, 25]
              df = pd.DataFrame(data)
<del>_</del>
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         圃
                                                               Title Author(s) Submission_Date Review_Score Reviewer_Comments Decision
                                                                                                                                                                                                                                                                                                                                             Keywords No_of_Pages Plagiarism_Score
                          Paper_ID
                                                                                                                                              2025-03-30
                                                                                                                                                                                                                                                                                                                                                       AI, ML
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ıl.
                                                                                                                                             2025-04-01
                                                2 Paper B
                                                                                                Author2
                                                                                                                                                                                                                                  Needs improvement
                                                                                                                                                                                                                                                                                                Rejected
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         1
                                                                                                Author1
                                                                                                                                              2025-04-02
                                                                                                                                                                                                                                                                                                                                        Al. Robotics
                                                        Paper C
                                                                                                                                                                                                                                                                                                Accepted
                                                                                                                                                                                                                                                                                                                                                                                                                9
                                                         Paper D
                                                                                                Author3
                                                                                                                                              2025-03-28
                                                                                                                                                                                                                      6
                                                                                                                                                                                                                                                                                                 Rejected
                                                                                                                                                                                                                                                                                                                                       ML, Security
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        20
                                                                                                Author4
                                                                                                                                              2025-03-31
                                                                                                                                                                                                                                                                                                  Rejected
                                                                                                                                                                                                                                                                                                                                           Blockchain
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        40
                5
                                                 6 Paper F
                                                                                                Author5
                                                                                                                                              2025-04-03
                                                                                                                                                                                                                                                               Well done Accepted
                                                                                                                                                                                                                                                                                                                                                       IoT, AI
                                                                                                                                              2025-03-27
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        15
                                                 7 Paper G
                                                                                                Author2
                                                                                                                                                                                                                                                                                                                                                                ML
                                                                                                                                                                                                                                                                                                                                                                                                                8
                                                                                                                                                                                                                                                     Good novelty
                                                                                                                                                                                                                                                                                                Accepted
                                                 8 Paper H
                                                                                                Author6
                                                                                                                                              2025-04-01
                                                                                                                                                                                                                                                                                                Accepted
                                                                                                                                                                                                                                                                                                                                   Cybersecurity
                                                                                                Author7
                                                                                                                                              2025-04-04
                                              10 Paper J
                                                                                                Author3
                                                                                                                                              2025-03-29
                                                                                                                                                                                                                                     Interesting concept
                                                                                                                                                                                                                                                                                                                                                  Security
```

## 1) Find the number of papers submitted

2) Calculate the average review score

```
df['Review_Score'].mean()

rp.float64(7.0)
```

3) Find the paper(s) with the highest review score



4) Find the number of papers accepted

```
df[df['Decision'] == 'Accepted'].shape[0]
5
```

5) List the titles of papers that were rejected



6) Find the percentage of papers with plagiarism score above 30%

```
(df[df['Plagiarism_Score'] > 30].shape[0] / df.shape[0]) * 100
→ 30.0
```

7) Average number of pages for accepted papers

8) Identify papers submitted after the deadline (2025-04-01)



9) Find top 5 papers with the lowest plagiarism scores



10) List unique keywords used across all papers

11) Correlation between review score and plagiarism score

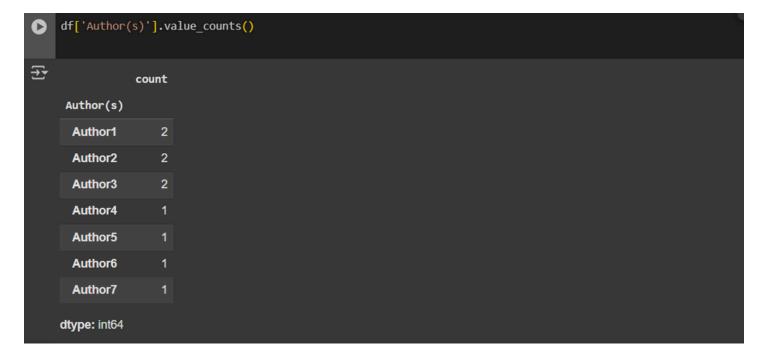


12) Find the standard deviation of the review scores

```
df['Review_Score'].std()

1.4907119849998598
```

### 13) Number of papers submitted by each author



14) Find papers with 'novel' in reviewer comments



15) Identify papers with outlier review scores



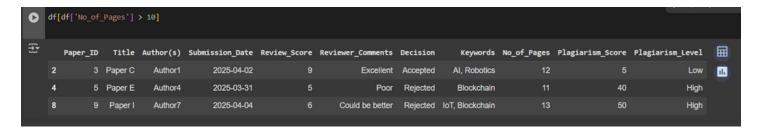
16) Group papers by decision and calculate the average review score



17) Create a new column categorizing papers based on plagiarism score

```
conditions = [
         (df['Plagiarism_Score'] <= 15),</pre>
         (df['Plagiarism_Score'] > 15) & (df['Plagiarism_Score'] <= 30),</pre>
         (df['Plagiarism_Score'] > 30)
    choices = ['Low', 'Moderate', 'High']
    df['Plagiarism_Level'] = np.select(conditions, choices, default='Unknown')
    df[['Paper_ID', 'Plagiarism_Score', 'Plagiarism_Level']]
<del>___</del>
        Paper_ID Plagiarism_Score Plagiarism_Level
                                                           圙
     0
                                  10
                                                    Low
                                                           ılı
     1
                                  35
                                                    High
     2
                3
                                   5
                                                    Low
     3
                                  20
                                               Moderate
     4
                                                    High
     5
                6
                                                    Low
     6
                                  15
                                                    Low
     7
                8
                                  10
                                                    Low
     8
                9
                                  50
                                                    High
     9
               10
                                  25
                                               Moderate
```

18) Find how many papers have more than 10 pages



19) Check for missing values in the dataset





Earliest Submission Date: 2025-03-27 00:00:00 Latest Submission Date: 2025-04-04 00:00:00