Name : Ishan Anand Kulkarni

Class : ET1   Batch : E13   Roll no : ET1-61   PRN : 2020401070156

Dataset used :

https://drive.google.com/file/d/1LlPOy6q4IWnHxynXmBbtRNjghY_1ar4l/view?usp=sharing

1) Q1. Load the dataset and display the first 5 records.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("First 5 Records:")
print(df.head())
```

```
First 5 Records:
  PaperID                        Title      Reviewer  Rating                                    Comments Recommendation
0    P001    Deep Learning in Healthcare  Alice Johnson       4  Insightful research with solid experiments.          Accept
1    P002      Quantum Computing Advances      Bob Smith       3            Interesting topic but lacks depth.  Minor Revision
2    P003               AI for Social Good    Charlie Lee       5  Excellent paper with strong societal impact.          Accept
3    P004      Blockchain in Supply Chain       Dana Kim       2     The concept is good but poorly presented.  Major Revision
4    P005    Neural Networks Optimization      Eva Brown       4              Well-written with good results.          Accept
PS D:\Python>
```

Q2) Display column names of the dataset

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Column Names:")
print(df.columns)
```

```
PS D:\Python> python -u "d:\Python\abc.py"
Column Names:
Index(['PaperID', 'Title', 'Reviewer', 'Rating', 'Comments', 'Recommendation'], dtype='object')
PS D:\Python>
```

Q3. Get basic info about the dataset.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Dataset Info:")
print(df.info())
```

```
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   PaperID         10 non-null     object
 1   Title           10 non-null     object
 2   Reviewer        10 non-null     object
 3   Rating          10 non-null     int64
 4   Comments        10 non-null     object
 5   Recommendation  10 non-null     object
dtypes: int64(1), object(5)
memory usage: 612.0+ bytes
None
PS D:\Python>
```

Q4. Display summary statistics for numeric columns.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Summary Statistics:")
print(df.describe())
```

```
PS D:\Python> python -u "d:\Python\abc.py"
Summary Statistics:
          Rating
count  10.000000
mean    3.300000
std     1.337494
min     1.000000
25%     2.250000
50%     3.500000
75%     4.000000
max     5.000000
```

Q5. Show unique values in the 'Recommendation' column.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Unique Recommendations:")
print(df['Recommendation'].unique())
```

```
PS D:\Python> python -u "d:\Python\abc.py"
Unique Recommendations:
['Accept' 'Minor Revision' 'Major Revision' 'Reject']
PS D:\Python>
```

Q6. Count number of each type of recommendation.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Recommendation Counts:")
print(df['Recommendation'].value_counts())
```

```
> python -u d:\Py
Recommendation Counts:
Recommendation
Accept             5
Minor Revision     2
Major Revision     2
Reject             1
Name: count, dtype: int64
PS D:\Python>
```

Q7. Find the average rating.

```
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Average Rating:")
print(df['Rating'].mean())
```

```
            > pyth
Average Rating:
3.3
PS D:\Python> []
```

Q8. Display papers with rating greater than 3.

```
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Papers with Rating > 3:")
print(df[df['Rating'] > 3])
```

```
Papers with Rating > 3:
    PaperID                          Title      Reviewer  Rating                                   Comments Recommendation
0    P001       Deep Learning in Healthcare  Alice Johnson      4  Insightful research with solid experiments.         Accept
2    P003                 AI for Social Good   Charlie Lee      5  Excellent paper with strong societal impact.    Accept
4    P005        Neural Networks Optimization    Eva Brown      4              Well-written with good results.    Accept
7    P008  Natural Language Processing Trends  Henry Scott      5          Very comprehensive and up-to-date.    Accept
9    P010        Augmented Reality Interfaces    Jack Adams      4          Engaging topic with great visuals.    Accept
PS D:\Python> []
```

Q9. Display titles of papers that were rejected.

```
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Rejected Papers:")
print(df[df['Recommendation'] == 'Reject']['Title'])
```

```
Rejected Papers:
6     IoT in Smart Cities
Name: Title, dtype: object
PS D:\Python> []
```

Q10. How many papers got the highest rating (5)?

```
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Papers with Rating 5:")
print((df['Rating'] == 5).sum())
```

```
Papers with Rating 5:
2
PS D:\Python>
```

Q11. Display reviewer names who gave rating 1.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Reviewers who gave Rating 1:")
print(df[df['Rating'] == 1]['Reviewer'])
```

```
Reviewers who gave Rating 1:
6    Grace Liu
Name: Reviewer, dtype: object
PS D:\Python>
```

Q12. Sort the dataset by rating in descending order.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Sorted by Rating Descending:")
print(df.sort_values(by='Rating', ascending=False))
```

```
Sorted by Rating Descending:
   PaperID                        Title        Reviewer  Rating                                          Comments  Recommendation
7    P008  Natural Language Processing Trends  Henry Scott       5              Very comprehensive and up-to-date.          Accept
2    P003                   AI for Social Good  Charlie Lee       5  Excellent paper with strong societal impact.          Accept
9    P010          Augmented Reality Interfaces   Jack Adams       4              Engaging topic with great visuals.          Accept
0    P001          Deep Learning in Healthcare  Alice Johnson      4    Insightful research with solid experiments.          Accept
4    P005          Neural Networks Optimization    Eva Brown       4               Well-written with good results.          Accept
1    P002            Quantum Computing Advances    Bob Smith       3     Interesting topic but lacks depth.  Minor Revision
5    P006          Cybersecurity Threat Modeling  Frank White      3            Needs more empirical validation.  Minor Revision
3    P004            Blockchain in Supply Chain     Dana Kim       2    The concept is good but poorly presented.  Major Revision
8    P009                  Robotics and Ethics     Ivy Green       2     Ethical arguments need more support.  Major Revision
6    P007                  IoT in Smart Cities    Grace Liu       1               Lacks novelty and clarity.          Reject
```

Q13. Create a new column 'Rating_Squared'.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
df['Rating_Squared'] = df['Rating'] ** 2
print("Added 'Rating_Squared' column:")
print(df[['Rating', 'Rating_Squared']].head())
```

```
Added 'Rating_Squared' column:
   Rating  Rating_Squared
0       4              16
1       3               9
2       5              25
3       2               4
4       4              16
PS D:\Python>
```

Q14. Replace 'Minor Revision' with 'Revise'.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
df['Recommendation'] = df['Recommendation'].replace('Minor Revision', 'Revise')
print("Updated Recommendations:")
print(df['Recommendation'].unique())
```

```
> python -u "d:\Python\abc.py"
Updated Recommendations:
['Accept' 'Revise' 'Major Revision' 'Reject']
PS D:\Python>
```

Q15. Group by recommendation and get average rating.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Average Rating per Recommendation:")
print(df.groupby('Recommendation')['Rating'].mean())
```

```
Average Rating per Recommendation:
Recommendation
Accept            4.4
Major Revision    2.0
Minor Revision    3.0
Reject            1.0
Name: Rating, dtype: float64
PS D:\Python>
```

Q16. Display number of characters in each comment.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Length of Comments:")
print(df['Comments'].apply(len).head())
```

```
> python -u d:\Python\
Length of Comments:
0    43
1    34
2    44
3    41
4    31
Name: Comments, dtype: int64
PS D:\Python>
```

Q17. Use NumPy to calculate standard deviation of ratings.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
import numpy as np
print("Rating Standard Deviation:")
print(np.std(df['Rating']))
```

```
Rating Standard Deviation:
1.2688577540449522
PS D:\Python>
```

Q18. Check if any rating is below 2.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Any rating below 2?")
print((df['Rating'] < 2).any())
```

```
PS D:\Python> python -u "d:\Python\abc.py"
Any rating below 2?
True
PS D:\Python>
```

Q19. Convert the 'Rating' column to NumPy array.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
ratings_array = df['Rating'].to_numpy()
print("Ratings Array:")
print(ratings_array)
```

```
PS D:\Python>



                > python -u "d:\Python\abc.py"
Ratings Array:
[4 3 5 2 4 3 1 5 2 4]
PS D:\Python>
```

Q20. Filter papers with 'AI' in the title.

```python
import pandas as pd
df = pd.read_csv('Paper_Review_Dataset.csv')
print("Papers with 'AI' in Title:")
print(df[df['Title'].str.contains('AI')]['Title'])
```

```
Papers with 'AI' in Title:
2      AI for Social Good
Name: Title, dtype: object
PS D:\Python>
```