

# **EDS ASSIGNMENT – 1**

TOPIC : BLOG AUTHORSHIP CORPUS

**NAME : RAJSHREE AWARDE**

**DIVISION : CS5      BATCH : CS52**

**ROLL NO : CS5-41**

**PRN : 202401100069**


```
prob1 copy 2.py > ...  
1  # Number of blog posts by bloggers aged 20-30  
2  import pandas as pd  
3  import numpy as np  
4  
5  # Load the dataset  
6  df = pd.read_csv('blogtext.csv')  
7  posts_20_30 = df[(df['age'] >= 20) & (df['age'] <= 30)].shape[0]  
8  print(posts_20_30)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree eds/prob1 copy 2.py"

321447

 prob1 copy 20.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 20. Standard deviation of ages
9  std_dev_age = df['age'].std()
10 print(std_dev_age)
11
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/Desktop/prob1 copy 20.py"

● 7.786008658769236



prob1 copy 12.py > ...

```
2 import pandas as pd
3 import numpy as np
4
5 # Load the dataset
6 df = pd.read_csv('blogtext.csv')
7
8 # 10. Calculate the average number of words in blog posts.
9 avg_words_per_post = df['text'].apply(lambda x: len(str(x).split())).mean()
10 print('10. Average words per post:', avg_words_per_post)
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

Python

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree/prob1 copy 12.py"

10. Average words per post: 200.78674238643504

prob1 copy 18.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 18. Find how many bloggers are older than 30 years.
9  bloggers_above_30 = (df['age'] > 30).sum()
10 print('18. Bloggers older than 30:', bloggers_above_30)
```

PROBLEMS

OUTPUT

DEBUG CONSOLE


TERMINAL

PORTS

yругw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yругw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yругw/OneDrive/Desktop/ra  
s/prob1 copy 18.py"

● 18. Bloggers older than 30: 123970

 prob1 copy 15.py > ...

```
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 17. Find the standard deviation of blogger ages.
9  std_dev_age = df['age'].std()
10 print('17. Age standard deviation:', std_dev_age)
```

PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   PORTS

yругw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yругw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yругw/OneDrive/Desktop/rajs

● s/prob1 copy 15.py"

17. Age standard deviation: 7.786008658769236





```
prob1 copy 16.py > ...
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 19. Find the most active blogger (the blogger with the maximum number of posts).
9  most_active_blogger = df['id'].value_counts().idxmax()
10 print(['19. Most active blogger ID:', most_active_blogger])
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/prob1 copy 16.py"

● 19. Most active blogger ID: 449628

prob1 copy 8.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 8. Distribution of astrological signs
9  astro_distribution = df['sign'].value_counts()
10 print(astro_distribution)
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree eds

• s/prob1 copy 8.py"

sign

Cancer 65048

Aries 64979

Taurus 62561

Libra 62363

Virgo 60399

Scorpio 57161

Pisces 54053

Leo 53811

Gemini 51985

Sagittarius 50036

Aquarius 49687

Capricorn 49201

Name: count, dtype: int64



prob1 copy 8.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 8. Distribution of astrological signs
9  astro_distribution = df['sign'].value_counts()
10 print(astro_distribution)
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree ed

• s/prob1 copy 8.py"

sign

Cancer 65048

Aries 64979

Taurus 62561

Libra 62363

Virgo 60399

Scorpio 57161

Pisces 54053

Leo 53811

Gemini 51985

Sagittarius 50036

Aquarius 49687

Capricorn 49201

Name: count, dtype: int64

prob1 copy 17.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7  # 17. Mean and median age by gender
8  mean_median_age_by_gender = df.groupby('gender')['age'].agg(['mean', 'median'])
9  print(mean_median_age_by_gender)
10
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS


yruwg@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yruwg/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yruwg/OneDrive/De

• s/prob1 copy 17.py"

	mean	median
gender		
female	23.855087	24.0
male	24.007529	24.0



 prob1 copy 14.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 11. Find the topic that appears most frequently.
9  most_common_topic = df['topic'].mode()[0]
10 print('11. Most common topic:', most_common_topic)
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

yru<sup>g</sup>w@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yru<sup>g</sup>w/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yru<sup>g</sup>w/Or  
.py"

● 11. Most common topic: indUnk



```
prob1 copy 4.py > ...
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 4. Number of bloggers for each gender
9  gender_counts = df['gender'].value_counts()
10 print(gender_counts)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree eds/prob1 copy 4.py"

gender

male 345193

female 336091

Name: count, dtype: int64

gender

male 345193

female 336091

Name: count, dtype: int64

prob1 copy 9.py > ...

```
1 #
2 import pandas as pd
3 import numpy as np
4
5 # Load the dataset
6 df = pd.read_csv('blogtext.csv')
7
8 # 9. Proportion of male to female bloggers
9
10 gender_counts = df['gender'].value_counts()
11 male_female_ratio = gender_counts['male'] / gender_counts['female']
12 print(male_female_ratio)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yruwg@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yruwg/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yruwg/OneDrive/Desktop/rajshree ed

• s/prob1 copy 9.py"

1.027081951019218





prob1 copy 5.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 5. Find the number of blog posts written by male bloggers.
9  male_posts = (df['gender'] == 'male').sum()
10 print('Number of male posts:', male_posts)
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

yruwg@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yruwg/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yruwg/OneDrive/Desktop/rajshree e

• s/prob1 copy 6.py"

200.78674238643504



```
1
2 import pandas as pd
3 import numpy as np
4
5 # Load the dataset
6 df = pd.read_csv('blogtext.csv')
7 # 6. Average number of words per blog post
8 df['word_count'] = df['text'].apply(lambda x: len(str(x).split()))
9 average_words = df['word_count'].mean()
10 print(average_words)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree eds

• s/prob1 copy 6.py"

200.78674238643504



prob1 copy 13.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7
8  # 13. Age group with the most bloggers
9  df['age_group'] = pd.cut(df['age'], bins=[0,19,29,39,49,59,100], labels=['Teens','Twenties','Thirties','Forties','Fifties','Sixties+'])
10 most_common_age_group = df['age_group'].value_counts().idxmax()
11 print(most_common_age_group)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree eds/prob1 copy 13.py"

Twenties

prob1 copy 11.py > ...

```
2 import pandas as pd
3 import numpy as np
4
5 # Load the dataset
6 df = pd.read_csv('blogtext.csv')
7
8 # 11. Top 5 most common astrological signs
9 top5_signs = df['sign'].value_counts().head(5)
10 print(top5_signs)
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree eds/prob1 copy 11.py"

● sign

Cancer 65048

Aries 64979

Taurus 62561

Libra 62363

Virgo 60399

Name: count, dtype: int64



prob1 copy 19.py > ...

```
1
2 import pandas as pd
3 import numpy as np
4
5 # Load the dataset
6 df = pd.read_csv('blogtext.csv')
7
8 # 19. Average blog post length by astrological sign
9 df['word_count'] = df['text'].apply(lambda x: len(str(x).split()))
10 avg_length_by_sign = df.groupby('sign')['word_count'].mean()
11 print(avg_length_by_sign)
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDr  
.py"

```
• sign
Aquarius      212.180993
Aries         186.006464
Cancer        207.809418
Capricorn     203.671307
Gemini        207.412792
Leo           211.659103
Libra         198.088658
Pisces        194.053984
Sagittarius   201.102186
Scorpio       190.356519
Taurus        195.613561
Virgo         205.791189
Name: word_count, dtype: float64
```

prob1 copy 7.py > ...

```
1  #
2  import pandas as pd
3  import numpy as np
4
5  # Load the dataset
6  df = pd.read_csv('blogtext.csv')
7  # 7. Blogger with maximum blog posts
8  max_posts_blogger = df['id'].value_counts().idxmax()
9  print(max_posts_blogger)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree e

• s/prob1 copy 7.py"

449628



prob1 copy 3.py > ...

```
2 import pandas as pd
3 import numpy as np
4
5 # Load the dataset
6 df = pd.read_csv('blogtext.csv')
7
8 # 3. Average age of bloggers
9 average_age = df['age'].mean()
10 print(average_age)
11
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree eds/prob1 copy 3.py"

● 23.932326313255558





```
prob1 copy 10.py > ...  
5 # Load the dataset  
6 df = pd.read_csv('blogtext.csv')  
7  
8 # 7. Find the most common astrological sign among the bloggers.  
9 most_common_sign = df['sign'].mode()[0]  
10 print('7. Most common sign:', most_common_sign)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

yrugw@Deadlux MINGW64 ~/OneDrive/Desktop/rajshree eds

\$ C:/Users/yrugw/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/yrugw/OneDrive/Desktop/rajshree/10.py"

● 7. Most common sign: Cancer

