EDS Activity – 1

27-04-2005

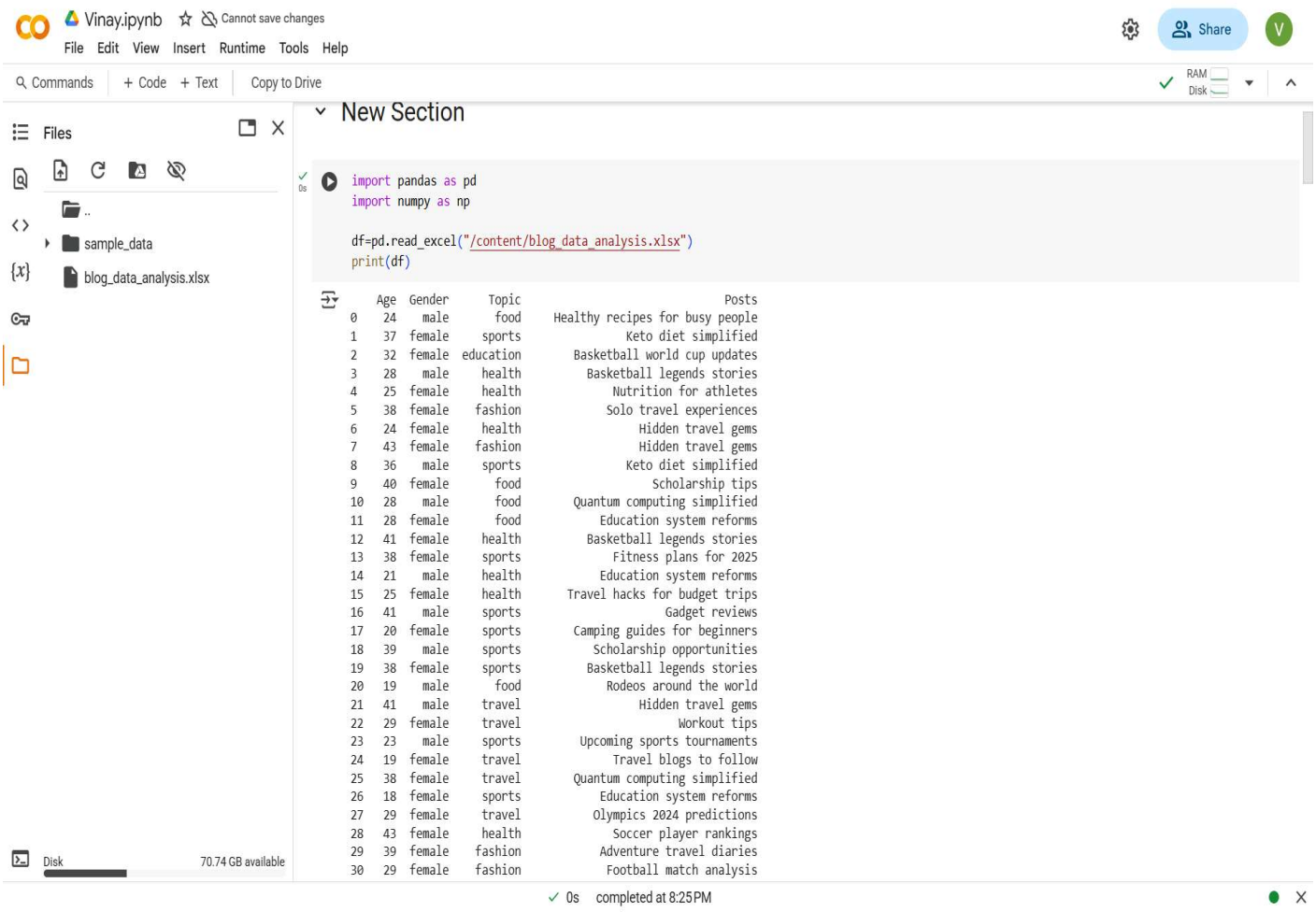
Name: - Vinay Sandip Dhake

Div.: - CS7

Roll No.: - CS7-42

PRN: - 202401110031

Dataset: - The Blog Authorship Corpus



	Age	Gender	Topic	Posts
0	24	male	food	Healthy recipes for busy people
1	37	female	sports	Keto diet simplified
2	32	female	education	Basketball world cup updates
3	28	male	health	Basketball legends stories
4	25	female	health	Nutrition for athletes
5	38	female	fashion	Solo travel experiences
6	24	female	health	Hidden travel gems
7	43	female	fashion	Hidden travel gems
8	36	male	sports	Keto diet simplified
9	40	female	food	Scholarship tips
10	28	male	food	Quantum computing simplified
11	28	female	food	Education system reforms
12	41	female	health	Basketball legends stories
13	38	female	sports	Fitness plans for 2025
14	21	male	health	Education system reforms
15	25	female	health	Travel hacks for budget trips
16	41	male	sports	Gadget reviews
17	20	female	sports	Camping guides for beginners
18	39	male	sports	Scholarship opportunities
19	38	female	sports	Basketball legends stories
20	19	male	food	Rodeos around the world
21	41	male	travel	Hidden travel gems
22	29	female	travel	Workout tips
23	23	male	sports	Upcoming sports tournaments
24	19	female	travel	Travel blogs to follow
25	38	female	travel	Quantum computing simplified
26	18	female	sports	Education system reforms
27	29	female	travel	Olympics 2024 predictions
28	43	female	health	Soccer player rankings
29	39	female	fashion	Adventure travel diaries
30	29	female	fashion	Football match analysis

Vinay.ipynb ☆ Cannot save changes

File Edit View Insert Runtime Tools Help

Share V

Commands + Code + Text Copy to Drive

RAM Disk

Files

..

sample_data

blood_data_analysis.xlsx

Disk 70.74 GB available

[2]
```
# 1) Find the average age of all bloggers.

average_age = np.mean(df['Age'])
print("Average Age:", average_age)
```

```
Average Age: 32.22
```

[3]
```
#2) Count number of male and female bloggers.

gender_count = df['Gender'].value_counts()
print(gender_count)
```

```
Gender
female    36
male      14
Name: count, dtype: int64
```

[4]
```
# 3) Find the most popular blog topic.

popular_topic = df['Topic'].mode()[0]
print("Most Popular Topic:", popular_topic)
```

```
Most Popular Topic: sports
```

```
# 4: Blogger with maximum age.

oldest_blogger = df[df['Age'] == df['Age'].max()]
print(oldest_blogger)
```

```
    Age  Gender   Topic                    Posts
33   44  female  sports    Nutrition myths busted
34   44  female    food  Adventure travel diaries
```

0s completed at 8:25PM

Vinay.ipynb ☆ Cannot save changes

File Edit View Insert Runtime Tools Help

Share V

Commands + Code + Text Copy to Drive

RAM Disk

[2]
```
# 1) Find the average age of all bloggers.

average_age = np.mean(df['Age'])
print("Average Age:", average_age)
```

```
Average Age: 32.22
```

```
#2) Count number of male and female bloggers.

gender_count = df['Gender'].value_counts()
print(gender_count)
```

```
Gender
female    36
male      14
Name: count, dtype: int64
```

[4]
```
# 3) Find the most popular blog topic.

popular_topic = df['Topic'].mode()[0]
print("Most Popular Topic:", popular_topic)
```

```
Most Popular Topic: sports
```

[5]
```
# 4: Blogger with maximum age.

oldest_blogger = df[df['Age'] == df['Age'].max()]
print(oldest_blogger)
```

```
    Age  Gender   Topic                    Posts
33   44  female  sports    Nutrition myths busted
34   44  female    food  Adventure travel diaries
```

0s completed at 8:25PM

Vinay.ipynb ☆ Cannot save changes

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Copy to Drive

RAM Disk

```
[6] # 5: Blogger with minimum age.

youngest_blogger = df[df['Age'] == df['Age'].min()]
print(youngest_blogger)
```

```
    Age  Gender   Topic                    Posts
26   18  female  sports  Education system reforms
```

```
[7] # 6: Number of unique blogging topics.

unique_topics = df['Topic'].nunique()
print("Unique Topics:", unique_topics)
```

```
Unique Topics: 6
```

```
# 7: Bloggers who are older than 30.

older_than_30 = df[df['Age'] > 30]
print(older_than_30)
```

```
    Age  Gender      Topic                            Posts
1    37  female     sports              Keto diet simplified
2    32  female  education       Basketball world cup updates
5    38  female    fashion            Solo travel experiences
7    43  female    fashion                 Hidden travel gems
8    36    male     sports              Keto diet simplified
9    40  female       food                  Scholarship tips
12   41  female     health         Basketball legends stories
13   38  female     sports             Fitness plans for 2025
16   41    male     sports                    Gadget reviews
18   39    male     sports          Scholarship opportunities
19   38  female     sports         Basketball legends stories
21   41    male     travel                 Hidden travel gems
25   38  female     travel       Quantum computing simplified
```

Connected to Python 3 Google Compute Engine backend

Vinay.ipynb ☆ Cannot save changes

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Copy to Drive

RAM Disk

```
[9] # 8: Bloggers who wrote about "food".

food_bloggers = df[df['Topic'] == 'food']
print(food_bloggers)
```

```
    Age  Gender Topic                          Posts
0    24    male  food  Healthy recipes for busy people
9    40  female  food                Scholarship tips
10   28    male  food     Quantum computing simplified
11   28  female  food         Education system reforms
20   19    male  food          Rodeos around the world
32   34  female  food          Rodeos around the world
34   44  female  food         Adventure travel diaries
40   29  female  food      Workout plans for beginners
```

```
[12] # 9: Total number of words across all posts.

total_words = df['Posts'].apply(lambda x: len(x.split())).sum()
print("Total Words:", total_words)
```

```
Total Words: 155
```

```
# 10: Which post has the maximum number of words?

longest_post_idx = df['Posts'].apply(lambda x: len(x.split())).idxmax()
print(df.loc[longest_post_idx])
```

```
Age                                24
Gender                           male
Topic                            food
Posts     Healthy recipes for busy people
Name: 0, dtype: object
```

```
# 11: Calculate the mean age for each blogging topic.
```

Connected to Python 3 Google Compute Engine backend

Vinay.ipynb ☆ Cannot save changes

File Edit View Insert Runtime Tools Help

Share V

Commands + Code + Text Copy to Drive

RAM Disk

```
[14] # 11: Calculate the mean age for each blogging topic.

mean_age_topic = df.groupby('Topic')['Age'].mean()
print(mean_age_topic)
```

```
Topic
education    32.000000
fashion      35.200000
food         30.750000
health       30.555556
sports       32.500000
travel       33.100000
Name: Age, dtype: float64
```

```
[15] # 12: How many bloggers have age between 25 and 35?

age_between = df[(df['Age'] >= 25) & (df['Age'] <= 35)].shape[0]
print("Bloggers aged 25-35:", age_between)
```

```
Bloggers aged 25-35: 16
```

```
# 14: Create new column "Post_Length" (word counts).

df['Post_Length'] = df['Posts'].apply(lambda x: len(x.split()))
print(df[['Posts', 'Post_Length']])
```

```
                              Posts  Post_Length
0     Healthy recipes for busy people            5
1               Keto diet simplified            3
2        Basketball world cup updates            4
3          Basketball legends stories            3
4              Nutrition for athletes            3
5             Solo travel experiences            3
6                   Hidden travel gems            3
```

Connected to Python 3 Google Compute Engine backend

Vinay.ipynb ☆ Cannot save changes

File Edit View Insert Runtime Tools Help

Share V

Commands + Code + Text Copy to Drive

RAM Disk

```
8               Keto diet simplified            3
9                  Scholarship tips            2
10       Quantum computing simplified            3
11           Education system reforms            3
12         Basketball legends stories            3
13              Fitness plans for 2025           4
14           Education system reforms            3
15      Travel hacks for budget trips            5
16                   Gadget reviews              2
17       Camping guides for beginners            4
18          Scholarship opportunities            2
19         Basketball legends stories            3
20            Rodeos around the world            4
21                 Hidden travel gems            3
22                      Workout tips             2
23        Upcoming sports tournaments            3
24             Travel blogs to follow            4
25       Quantum computing simplified            3
26           Education system reforms            3
27          Olympics 2024 predictions            3
28             Soccer player rankings            3
29           Adventure travel diaries            3
30            Football match analysis            3
31                   Scholarship tips            2
32            Rodeos around the world            4
33             Nutrition myths busted            3
34           Adventure travel diaries            3
35           Adventure travel diaries            3
36            Fashion week highlights            3
37            Fashion week highlights            3
38  E-learning vs traditional learning           4
39                   Gadget reviews              2
40        Workout plans for beginners            4
41           Adventure travel diaries            3
42                Trekking essentials            2
43         Basketball legends stories            3
44          Tennis grand slam winners            4
45          Exploring AI advancements            3
46             Nutrition for athletes            3
47       Quantum computing simplified            3
48                   Low-carb recipes            2
49              Latest fashion trends            3
```

Connected to Python 3 Google Compute Engine backend

Vinay.ipynb ☆ Cannot save changes

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Copy to Drive

RAM Disk

```
[18] #  15: Find topic with bloggers having highest average post length.

topic_max_post_len = df.groupby('Topic')['Post_Length'].mean().idxmax()
print("Topic with longest posts:", topic_max_post_len)
```

```
Topic with longest posts: food
```

```
# 16: Sort bloggers by Age ascending.

sorted_asc = df.sort_values(by='Age')
print(sorted_asc[['Age', 'Gender', 'Topic']])
```

```
    Age  Gender      Topic
26   18  female     sports
20   19    male       food
24   19  female     travel
17   20  female     sports
44   20  female     sports
14   21    male     health
45   22    male     travel
23   23    male     sports
6    24  female     health
47   24    male  education
0    24    male       food
15   25  female     health
4    25  female     health
49   26  female     sports
35   27    male    fashion
10   28    male       food
3    28    male     health
11   28  female       food
40   29  female       food
22   29  female     travel
27   29  female     travel
30   29  female    fashion
2    32  female  education
38   32  female     health
37   32  female  education
```

Connected to Python 3 Google Compute Engine backend

Vinay.ipynb ☆ Cannot save changes

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Copy to Drive

RAM Disk

```
8    36    male     sports
46   36  female     health
39   36  female     travel
1    37  female     sports
42   37  female     travel
25   38  female     travel
13   38  female     sports
19   38  female     sports
5    38  female    fashion
48   38  female     travel
18   39    male     sports
29   39  female    fashion
9    40  female       food
41   40  female  education
21   41    male     travel
12   41  female     health
16   41    male     sports
31   42  female     sports
43   42  female     travel
28   43  female     health
7    43  female    fashion
34   44  female       food
33   44  female     sports
```

```
[20] # 17: Find the median age of bloggers.

median_age = np.median(df['Age'])
print("Median Age:", median_age)
```

```
Median Age: 32.5
```

```
[21] # 18: Group bloggers by gender and find average Post_Length.

avg_post_gender = df.groupby('Gender')['Post_Length'].mean()
print(avg_post_gender)
```

```
Gender
female    3.111111
male      3.071429
Name: Post_Length, dtype: float64
```

Connected to Python 3 Google Compute Engine backend

Vinay.ipynb Cannot save changes

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Copy to Drive

RAM Disk

```
[21] # 18: Group bloggers by gender and find average Post_Length.

avg_post_gender = df.groupby('Gender')['Post_Length'].mean()
print(avg_post_gender)
```

```
Gender
female    3.111111
male      3.071429
Name: Post_Length, dtype: float64
```

```
[22] # 19: Bloggers writing about travel.

travel_bloggers = df[df['Topic'] == 'travel']
print(travel_bloggers)
```

```
    Age  Gender   Topic                       Posts  Post_Length
21   41    male  travel           Hidden travel gems            3
22   29  female  travel                Workout tips            2
24   19  female  travel        Travel blogs to follow            4
25   38  female  travel  Quantum computing simplified            3
27   29  female  travel     Olympics 2024 predictions            3
39   36  female  travel                Gadget reviews            2
42   37  female  travel          Trekking essentials             2
43   42  female  travel    Basketball legends stories            3
45   22    male  travel     Exploring AI advancements            3
48   38  female  travel              Low-carb recipes            2
```

```
[23] # 20: Bloggers with posts having less than 6 words.


travel_bloggers = df[df['Topic'] == 'travel']
print(travel_bloggers)
```

```
    Age  Gender   Topic                       Posts  Post_Length
21   41    male  travel           Hidden travel gems            3
22   29  female  travel                Workout tips            2
24   19  female  travel        Travel blogs to follow            4
25   38  female  travel  Quantum computing simplified            3
```

Connected to Python 3 Google Compute Engine backend

Vinay.ipynb Cannot save changes

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Copy to Drive

RAM Disk

```
    Age  Gender   Topic                       Posts  Post_Length
21   41    male  travel           Hidden travel gems            3
22   29  female  travel                Workout tips            2
24   19  female  travel        Travel blogs to follow            4
25   38  female  travel  Quantum computing simplified            3
27   29  female  travel     Olympics 2024 predictions            3
39   36  female  travel                Gadget reviews            2
42   37  female  travel          Trekking essentials             2
43   42  female  travel    Basketball legends stories            3
45   22    male  travel     Exploring AI advancements            3
48   38  female  travel              Low-carb recipes            2
```

[23] # 20: Bloggers with posts having less than 6 words.

```
travel_bloggers = df[df['Topic'] == 'travel']
print(travel_bloggers)
```

```
    Age  Gender   Topic                       Posts  Post_Length
21   41    male  travel           Hidden travel gems            3
22   29  female  travel                Workout tips            2
24   19  female  travel        Travel blogs to follow            4
25   38  female  travel  Quantum computing simplified            3
27   29  female  travel     Olympics 2024 predictions            3
39   36  female  travel                Gadget reviews            2
42   37  female  travel          Trekking essentials             2
43   42  female  travel    Basketball legends stories            3
45   22    male  travel     Exploring AI advancements            3
48   38  female  travel              Low-carb recipes            2
```

```
# 21: Find the proportion of bloggers aged above 30.

above_30 = (df['Age'] > 30).sum()
proportion_above_30 = above_30 / df.shape[0]
print("Proportion of bloggers aged above 30:", proportion_above_30)
```

```
Proportion of bloggers aged above 30: 0.56
```

Connected to Python 3 Google Compute Engine backend