NAME: Triveni Kallappa Biradar

PRN : 202401070172

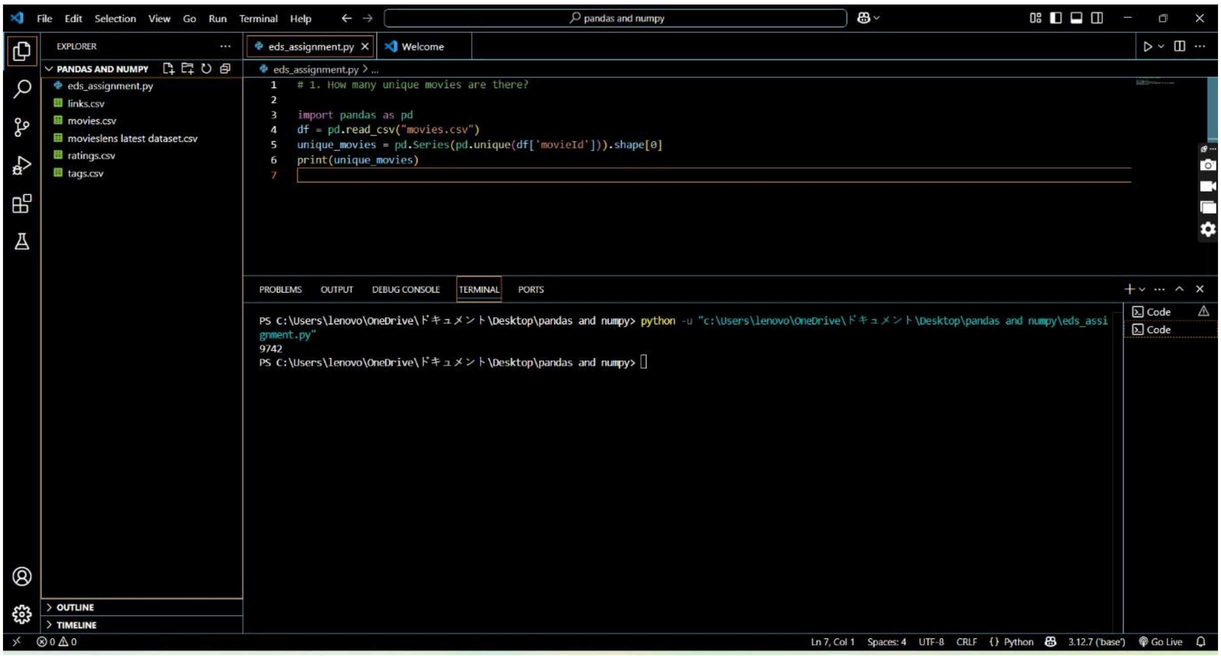
ROLLNO : ET2-44

CLASS : ET2

EDS THEORY ASSIGNMENT 1

URL: https://www.kaggle.com/datasets/shubhammehta21/movie-lens-small-latest-dataset

1.How many unique movies are there?

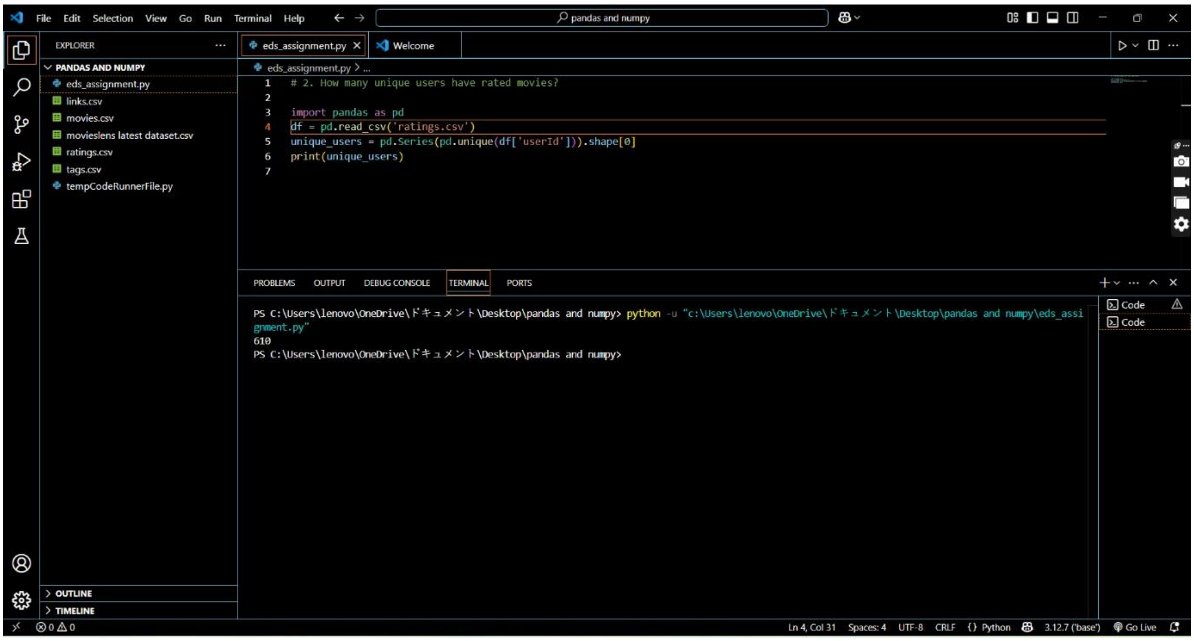

File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
OUTLINE
TIMELINE
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Code
Ln 7, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 1. How many unique movies are there?

import pandas as pd
df = pd.read_csv("movies.csv")
unique_movies = pd.Series(pd.unique(df['movieId'])).shape[0]
print(unique_movies)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
9742
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

2. How many unique users have rated movies?

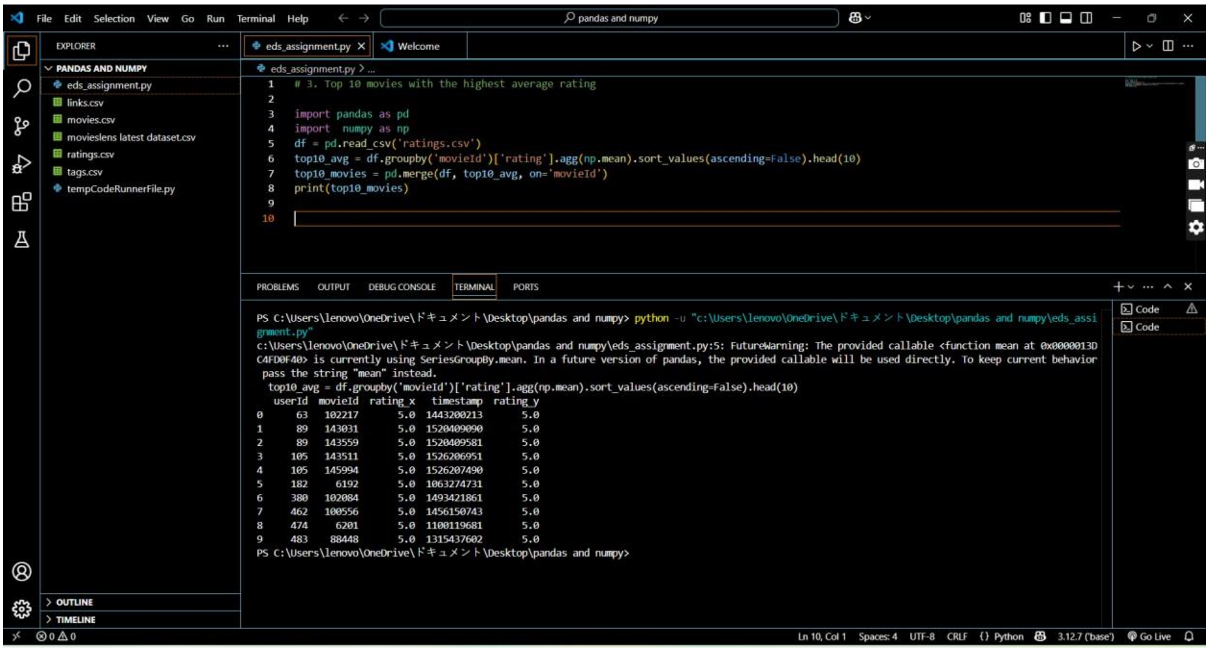

File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Code
Ln 4, Col 31 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 2. How many unique users have rated movies?

import pandas as pd
df = pd.read_csv('ratings.csv')
unique_users = pd.Series(pd.unique(df['userId'])).shape[0]
print(unique_users)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
610
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

3. Top 10 movies with the highest average rating


File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Code
Ln 10, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

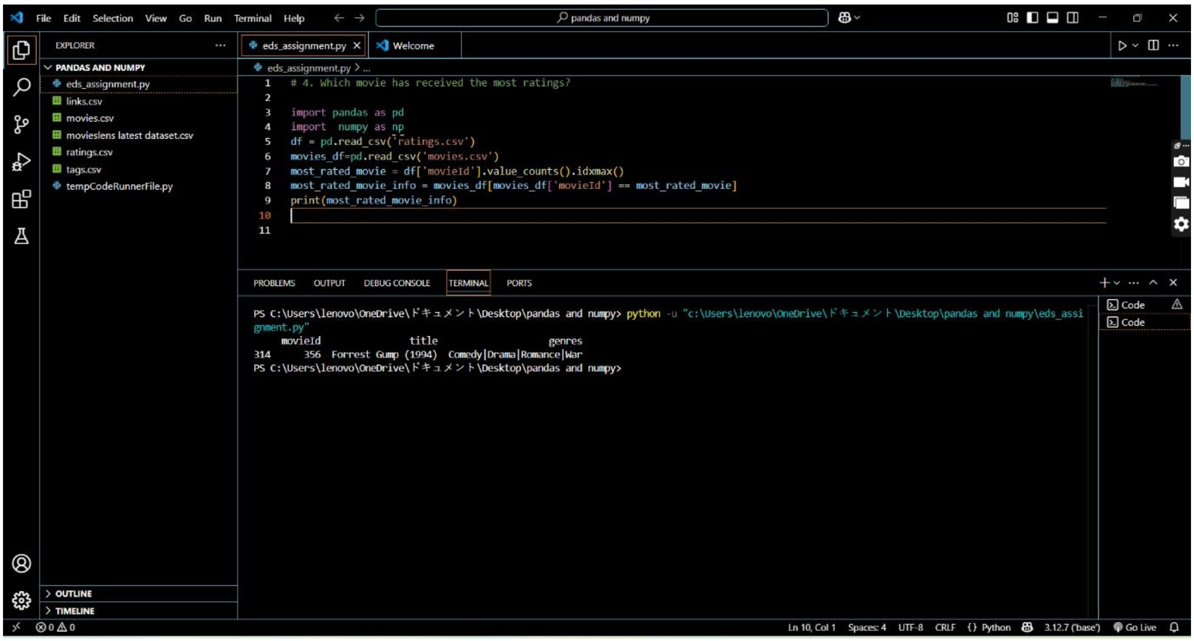
```python
# 3. Top 10 movies with the highest average rating

import pandas as pd
import  numpy as np
df = pd.read_csv('ratings.csv')
top10_avg = df.groupby('movieId')['rating'].agg(np.mean).sort_values(ascending=False).head(10)
top10_movies = pd.merge(df, top10_avg, on='movieId')
print(top10_movies)

```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assignment.py:5: FutureWarning: The provided callable <function mean at 0x0000013D
C4FD0F40> is currently using SeriesGroupBy.mean. In a future version of pandas, the provided callable will be used directly. To keep current behavior
 pass the string "mean" instead.
  top10_avg = df.groupby('movieId')['rating'].agg(np.mean).sort_values(ascending=False).head(10)
   userId  movieId  rating_x   timestamp  rating_y
0      63   102217       5.0  1443200213       5.0
1      89   143031       5.0  1520409090       5.0
2      89   143559       5.0  1520409581       5.0
3     105   143511       5.0  1526206951       5.0
4     105   145994       5.0  1526207490       5.0
5     182     6192       5.0  1063274731       5.0
6     380   102084       5.0  1493421861       5.0
7     462   100556       5.0  1456150743       5.0
8     474     6201       5.0  1100119681       5.0
9     483    88448       5.0  1315437602       5.0
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

4. Which movie has received the most ratings?


File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Code
Ln 10, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

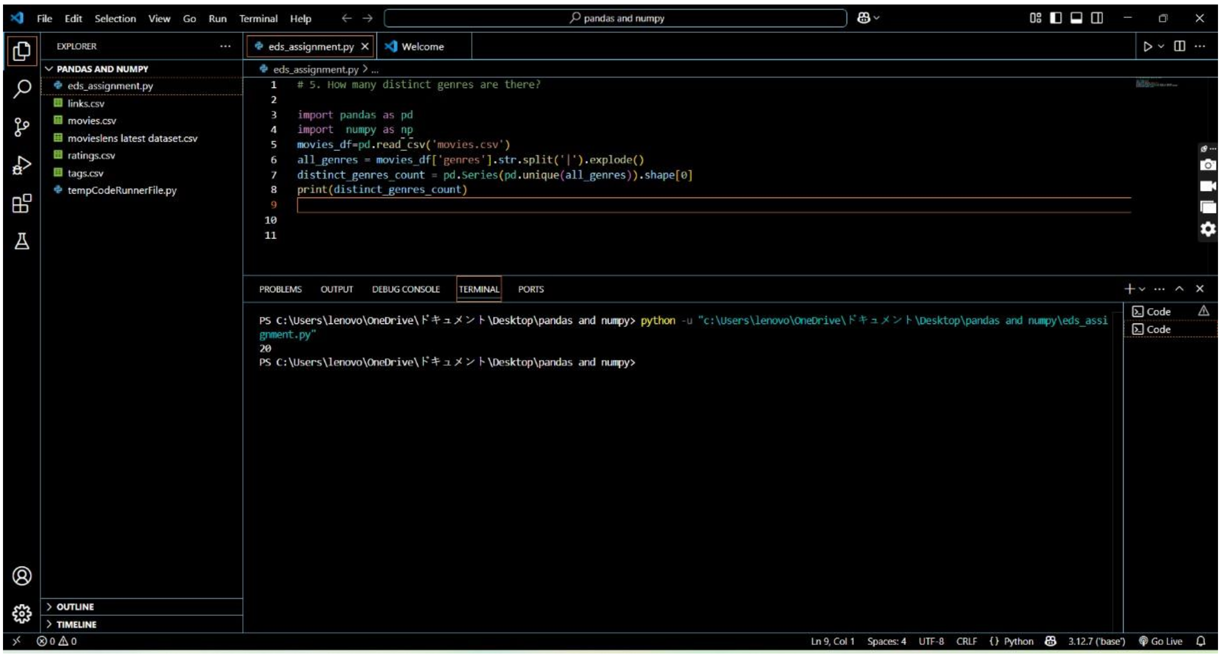
```python
# 4. Which movie has received the most ratings?

import pandas as pd
import  numpy as np
df = pd.read_csv('ratings.csv')
movies_df=pd.read_csv('movies.csv')
most_rated_movie = df['movieId'].value_counts().idxmax()
most_rated_movie_info = movies_df[movies_df['movieId'] == most_rated_movie]
print(most_rated_movie_info)


```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
     movieId                title                    genres
314      356  Forrest Gump (1994)  Comedy|Drama|Romance|War
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

5. How many distinct genres are there?

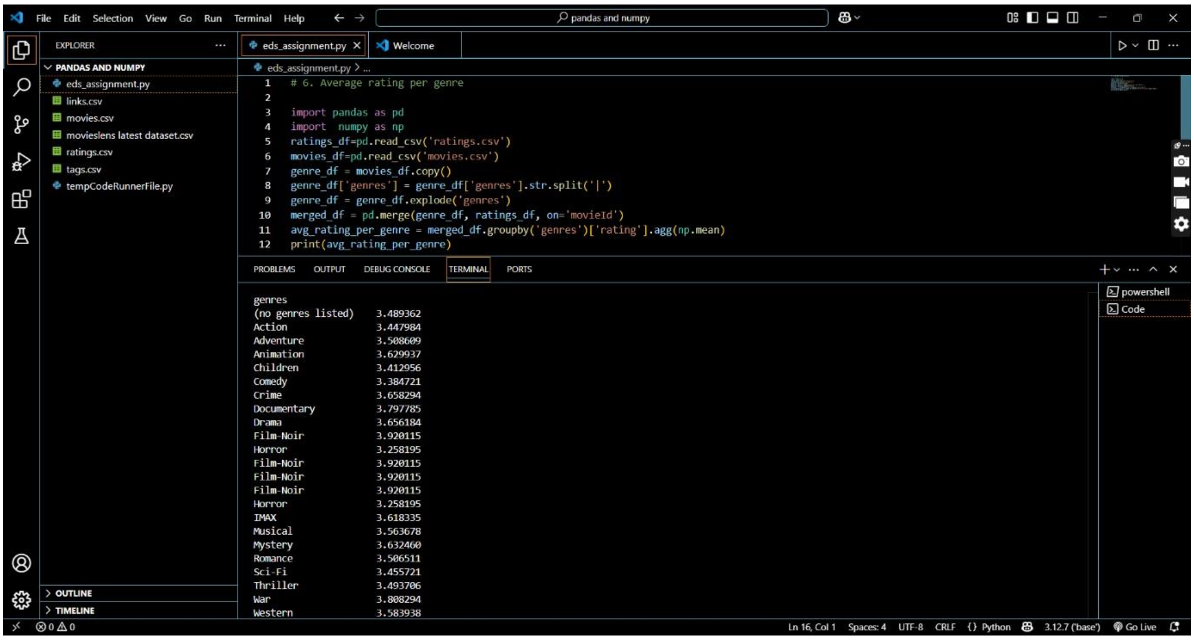

File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Code
Ln 9, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 5. How many distinct genres are there?

import pandas as pd
import  numpy as np
movies_df=pd.read_csv('movies.csv')
all_genres = movies_df['genres'].str.split('|').explode()
distinct_genres_count = pd.Series(pd.unique(all_genres)).shape[0]
print(distinct_genres_count)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
20
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

6. Average rating per genre

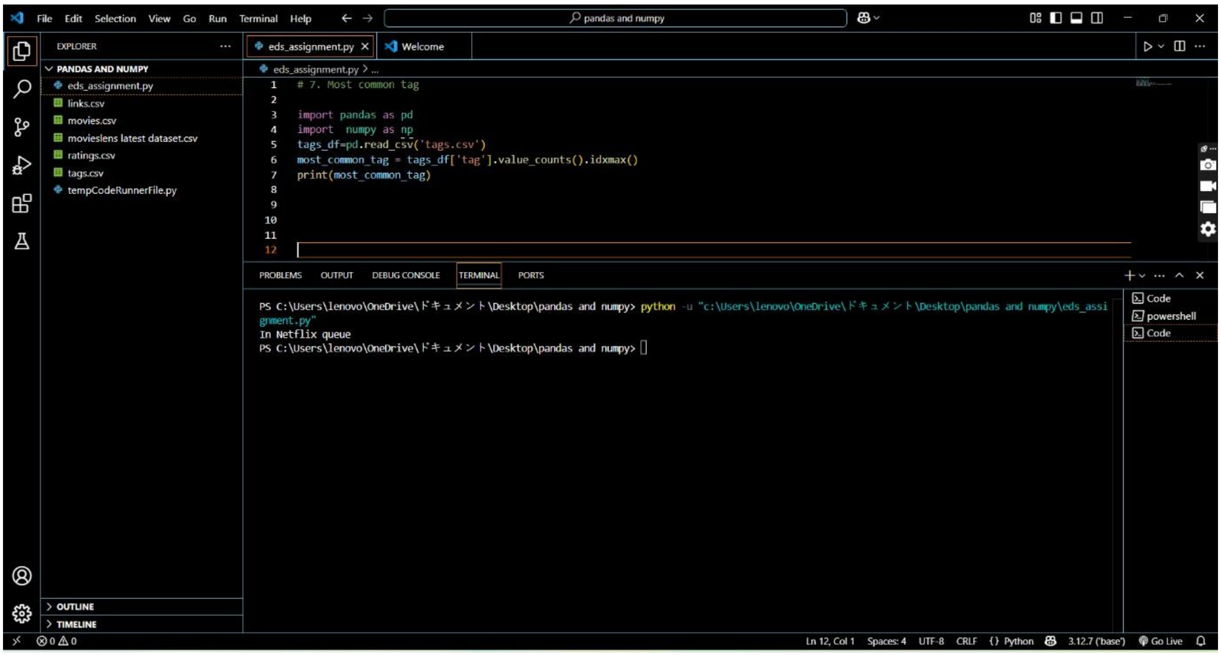

File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
powershell
Code
Ln 16, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 6. Average rating per genre

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
movies_df=pd.read_csv('movies.csv')
genre_df = movies_df.copy()
genre_df['genres'] = genre_df['genres'].str.split('|')
genre_df = genre_df.explode('genres')
merged_df = pd.merge(genre_df, ratings_df, on='movieId')
avg_rating_per_genre = merged_df.groupby('genres')['rating'].agg(np.mean)
print(avg_rating_per_genre)
```

```
genres
(no genres listed)    3.489362
Action                3.447984
Adventure             3.508609
Animation             3.629937
Children              3.412956
Comedy                3.384721
Crime                 3.658294
Documentary           3.797785
Drama                 3.656184
Film-Noir             3.920115
Horror                3.258195
Film-Noir             3.920115
Film-Noir             3.920115
Film-Noir             3.920115
Horror                3.258195
IMAX                  3.618335
Musical               3.563678
Mystery               3.632460
Romance               3.506511
Sci-Fi                3.455721
Thriller              3.493706
War                   3.808294
Western               3.583938
```

7. Most common tag

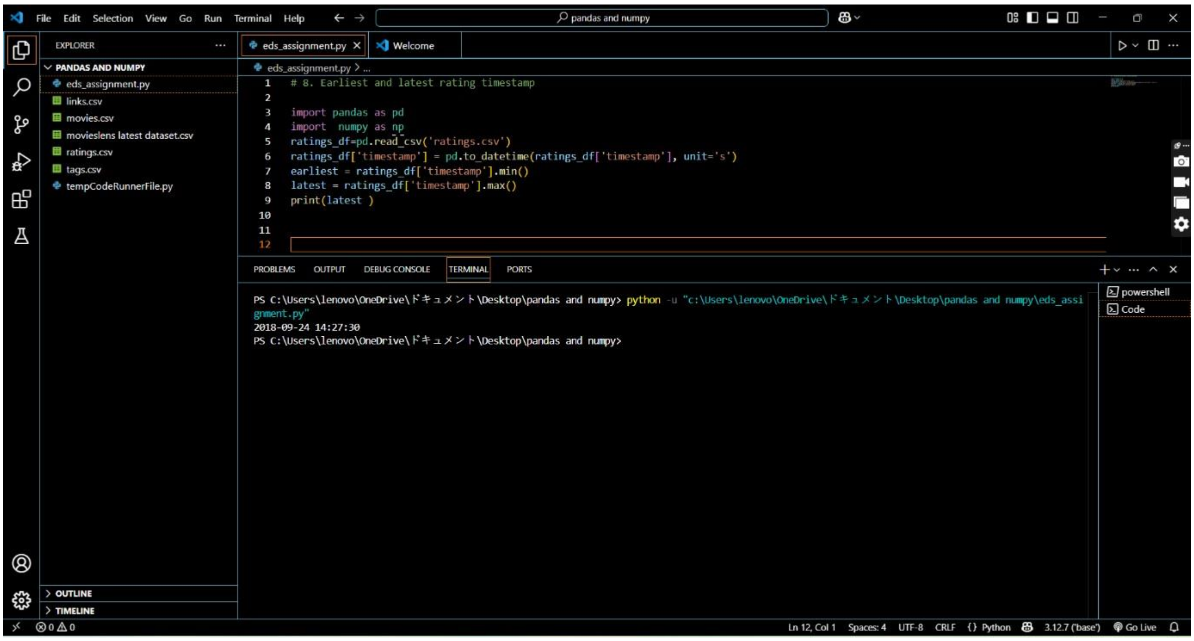


```python
# 7. Most common tag

import pandas as pd
import  numpy as np
tags_df=pd.read_csv('tags.csv')
most_common_tag = tags_df['tag'].value_counts().idxmax()
print(most_common_tag)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
In Netflix queue
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> 
```

8. Earliest and latest rating timestamp

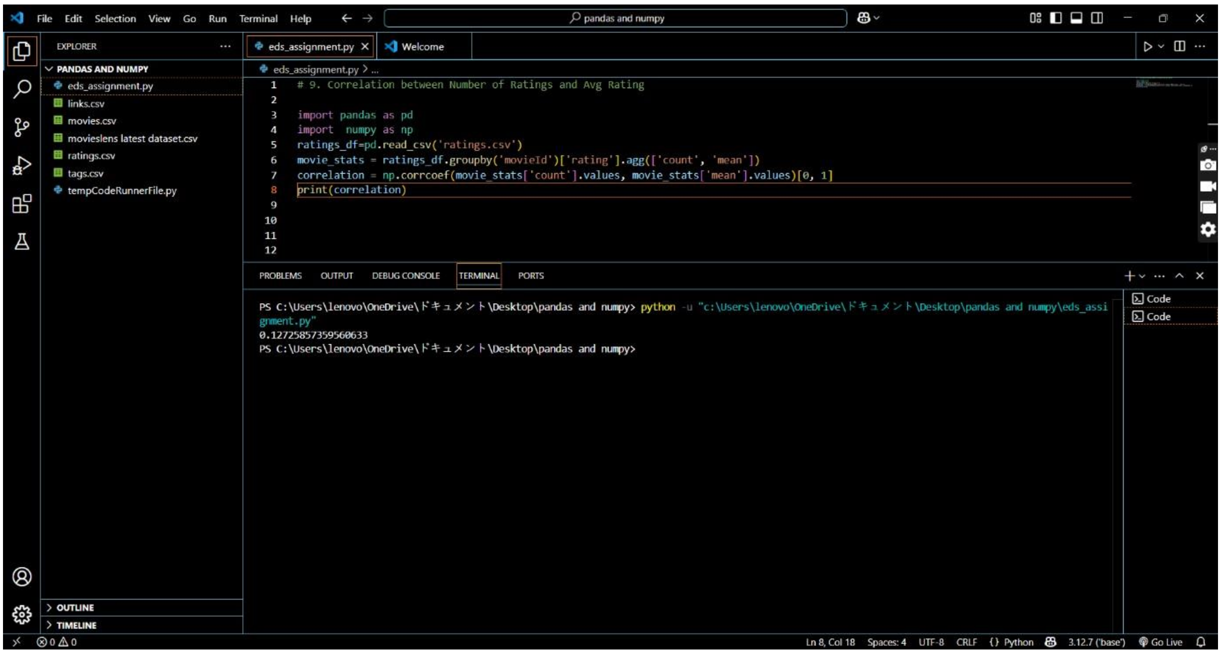


```python
# 8. Earliest and latest rating timestamp

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
ratings_df['timestamp'] = pd.to_datetime(ratings_df['timestamp'], unit='s')
earliest = ratings_df['timestamp'].min()
latest = ratings_df['timestamp'].max()
print(latest )
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
2018-09-24 14:27:30
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

9. Correlation between number of ratings and average rating

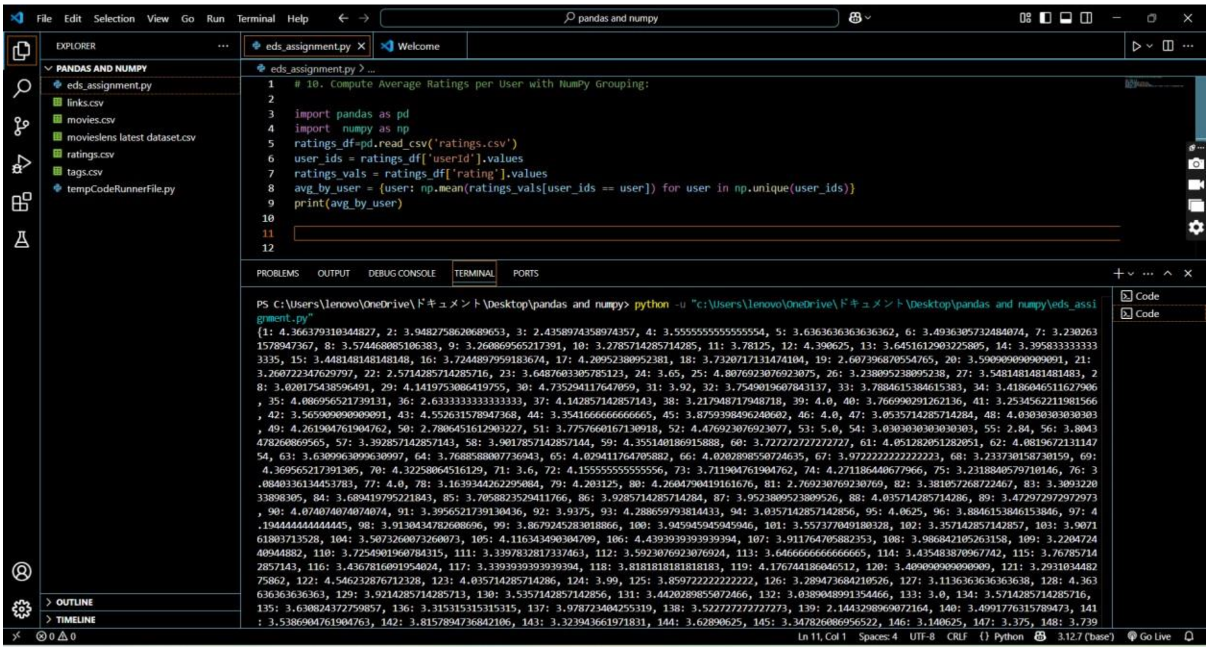

File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Code
OUTLINE
TIMELINE
Ln 8, Col 18 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 9. Correlation between Number of Ratings and Avg Rating

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
movie_stats = ratings_df.groupby('movieId')['rating'].agg(['count', 'mean'])
correlation = np.corrcoef(movie_stats['count'].values, movie_stats['mean'].values)[0, 1]
print(correlation)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
0.12725857359560633
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

10. Compute Average Ratings per User with NumPy Grouping:

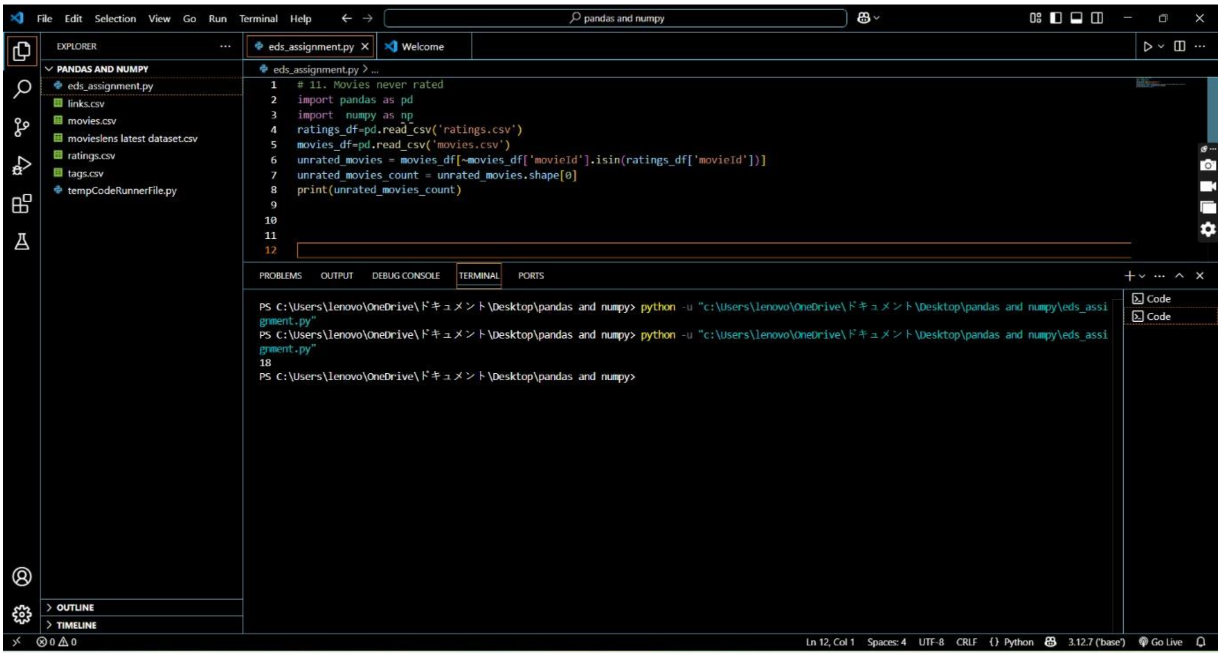

File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Code
OUTLINE
TIMELINE
Ln 11, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 10. Compute Average Ratings per User with NumPy Grouping:

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
user_ids = ratings_df['userId'].values
ratings_vals = ratings_df['rating'].values
avg_by_user = {user: np.mean(ratings_vals[user_ids == user]) for user in np.unique(user_ids)}
print(avg_by_user)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
{1: 4.366379310344827, 2: 3.9482758620689653, 3: 2.4358974358974357, 4: 3.5555555555555554, 5: 3.6363636363636362, 6: 3.4936305732484074, 7: 3.230263
1578947367, 8: 3.574468085106383, 9: 3.260869565217391, 10: 3.2785714285714285, 11: 3.78125, 12: 4.390625, 13: 3.6451612903225805, 14: 3.395833333333
3335, 15: 3.4481481481481481, 16: 3.7244897959183674, 17: 4.20952380952381, 18: 3.7320717131474104, 19: 2.607396870554765, 20: 3.590909090909091, 21:
3.260722347629797, 22: 2.5714285714285716, 23: 3.6487603305785123, 24: 3.65, 25: 4.8076923076923075, 26: 3.238095238095238, 27: 3.5481481481481483, 2
8: 3.020175438596491, 29: 4.1419753086419755, 30: 4.735294117647059, 31: 3.92, 32: 3.7549019607843137, 33: 3.7884615384615383, 34: 3.4186046511627906
, 35: 4.086956521739131, 36: 2.6333333333333333, 37: 4.142857142857143, 38: 3.217948717948718, 39: 4.0, 40: 3.766990291262136, 41: 3.2534562211981566
, 42: 3.565909090909091, 43: 4.552631578947368, 44: 3.3541666666666665, 45: 3.8759398496240602, 46: 4.0, 47: 3.0535714285714284, 48: 4.03030303030303
, 49: 4.261904761904762, 50: 2.7806451612903227, 51: 3.7757660167130918, 52: 4.476923076923077, 53: 5.0, 54: 3.0303030303030303, 55: 2.84, 56: 3.8043
478260869565, 57: 3.392857142857143, 58: 3.9017857142857144, 59: 4.355140186915888, 60: 3.727272727272727, 61: 4.051282051282051, 62: 4.0819672131147
54, 63: 3.6309963099630997, 64: 3.7688588007736943, 65: 4.029411764705882, 66: 4.0202898550724635, 67: 3.9722222222222223, 68: 3.233730158730159, 69:
 4.369565217391305, 70: 4.32258064516129, 71: 3.6, 72: 4.155555555555556, 73: 3.711904761904762, 74: 4.271186440677966, 75: 3.2318840579710146, 76: 3
.0840336134453783, 77: 4.0, 78: 3.1639344262295084, 79: 4.203125, 80: 4.2604790419161676, 81: 2.769230769230769, 82: 3.381057268722467, 83: 3.3093220
33898305, 84: 3.689419795221843, 85: 3.7058823529411766, 86: 3.9285714285714284, 87: 3.9523809523809526, 88: 4.035714285714286, 89: 3.472972972972973
, 90: 4.074074074074074, 91: 3.3956521739130436, 92: 3.9375, 93: 4.288659793814433, 94: 3.0357142857142856, 95: 4.0625, 96: 3.8846153846153846, 97: 4
.194444444444445, 98: 3.9130434782608696, 99: 3.8679245283018866, 100: 3.945945945945946, 101: 3.557377049180328, 102: 3.357142857142857, 103: 3.9071
61803713528, 104: 3.5073260073260073, 105: 4.116343490304709, 106: 4.4393939393939394, 107: 3.911764705882353, 108: 3.986842105263158, 109: 3.2204724
40944882, 110: 3.7254901960784315, 111: 3.3397832817337463, 112: 3.5923076923076924, 113: 3.6466666666666665, 114: 3.435483870967742, 115: 3.76785714
2857143, 116: 3.4367816091954024, 117: 3.3393939393939394, 118: 3.8181818181818183, 119: 4.176744186046512, 120: 3.409090909090909, 121: 3.2931034482
75862, 122: 4.546232876712328, 123: 4.035714285714286, 124: 3.99, 125: 3.8597222222222222, 126: 3.289473684210526, 127: 3.1136363636363638, 128: 4.363
636363636363, 129: 3.9214285714285713, 130: 3.5357142857142856, 131: 3.4420289855072466, 132: 3.0389048991354466, 133: 3.0, 134: 3.5714285714285716,
135: 3.630824372759857, 136: 3.315315315315315, 137: 3.978723404255319, 138: 3.522727272727273, 139: 2.1443298969072164, 140: 3.4991776315789473, 141
: 3.5386904761904763, 142: 3.8157894736842106, 143: 3.323943661971831, 144: 3.62890625, 145: 3.347826086956522, 146: 3.140625, 147: 3.375, 148: 3.739
```

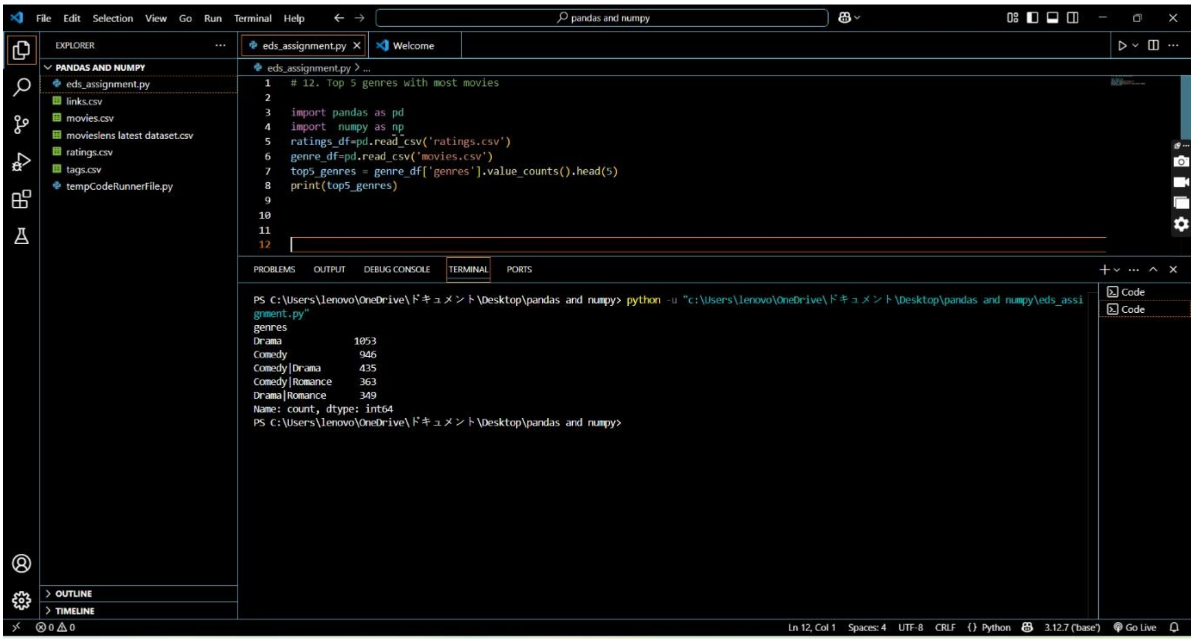
11. Movies never rated



```python
# 11. Movies never rated
import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
movies_df=pd.read_csv('movies.csv')
unrated_movies = movies_df[~movies_df['movieId'].isin(ratings_df['movieId'])]
unrated_movies_count = unrated_movies.shape[0]
print(unrated_movies_count)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
18
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

12. Top 5 genres with most movies

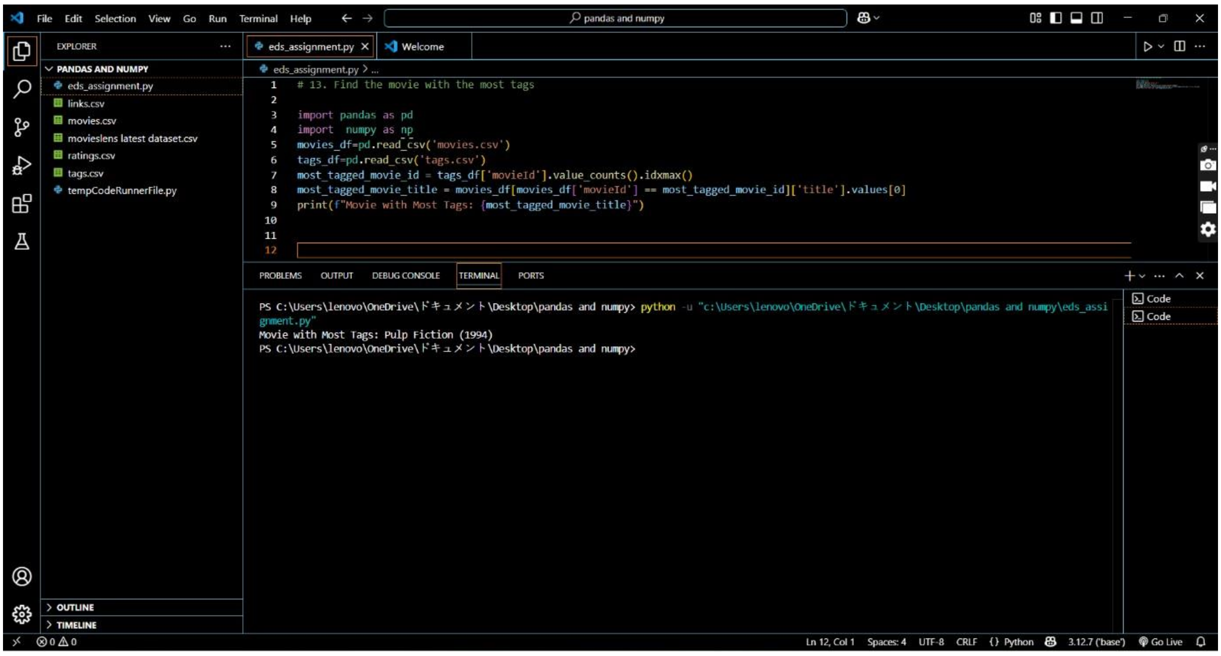


```python
# 12. Top 5 genres with most movies

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
genre_df=pd.read_csv('movies.csv')
top5_genres = genre_df['genres'].value_counts().head(5)
print(top5_genres)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
genres
Drama             1053
Comedy             946
Comedy|Drama       435
Comedy|Romance     363
Drama|Romance      349
Name: count, dtype: int64
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

13. find the movie with the most tags

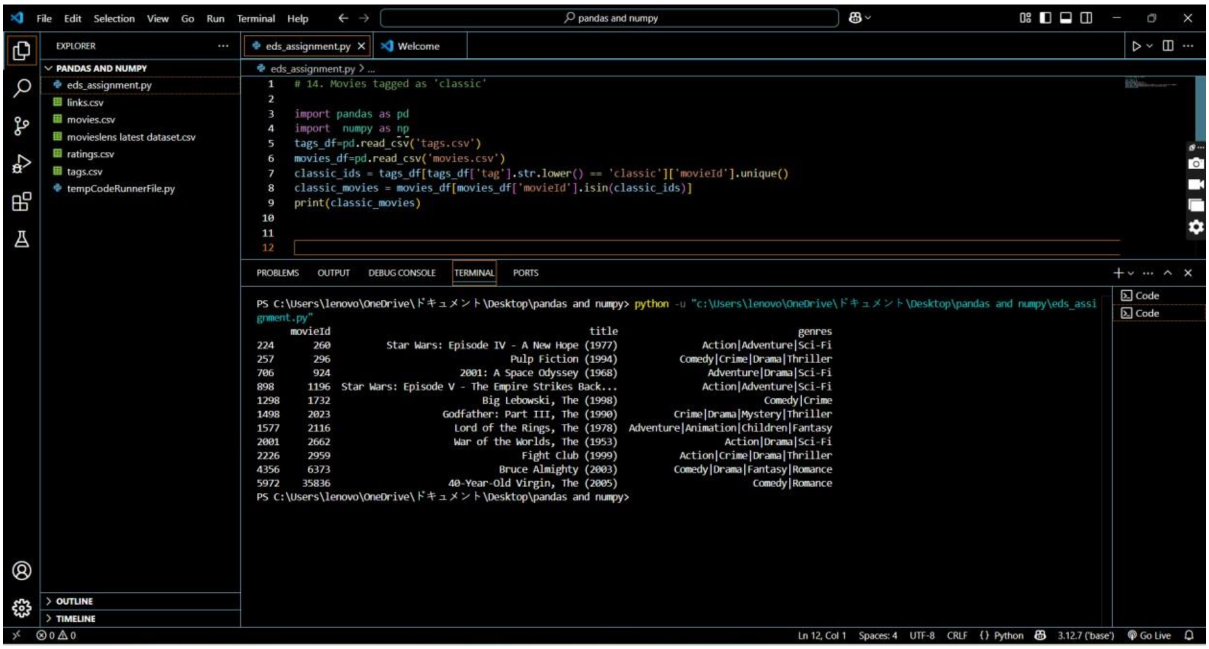


```python
# 13. Find the movie with the most tags

import pandas as pd
import  numpy as np
movies_df=pd.read_csv('movies.csv')
tags_df=pd.read_csv('tags.csv')
most_tagged_movie_id = tags_df['movieId'].value_counts().idxmax()
most_tagged_movie_title = movies_df[movies_df['movieId'] == most_tagged_movie_id]['title'].values[0]
print(f"Movie with Most Tags: {most_tagged_movie_title}")
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assignment.py"
Movie with Most Tags: Pulp Fiction (1994)
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

14. Movies tagged as 'classic'

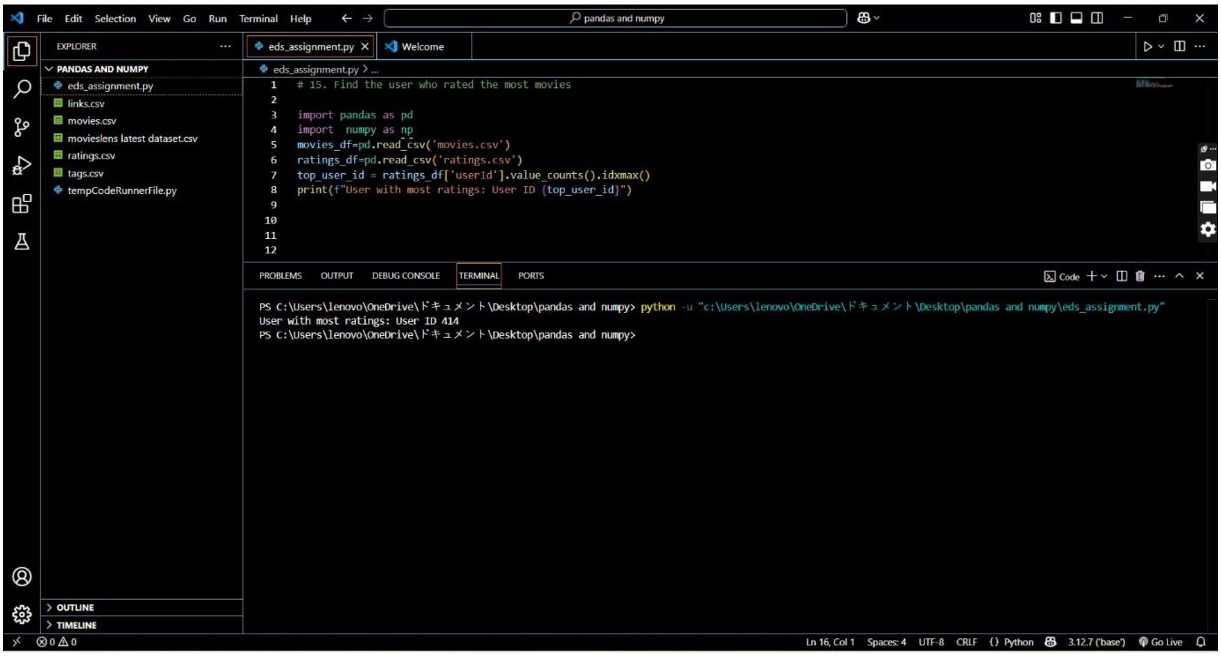


```python
# 14. Movies tagged as 'classic'

import pandas as pd
import  numpy as np
tags_df=pd.read_csv('tags.csv')
movies_df=pd.read_csv('movies.csv')
classic_ids = tags_df[tags_df['tag'].str.lower() == 'classic']['movieId'].unique()
classic_movies = movies_df[movies_df['movieId'].isin(classic_ids)]
print(classic_movies)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assignment.py"
      movieId                                          title                                   genres
224       260          Star Wars: Episode IV - A New Hope (1977)            Action|Adventure|Sci-Fi
257       296                                Pulp Fiction (1994)          Comedy|Crime|Drama|Thriller
706       924                       2001: A Space Odyssey (1968)             Adventure|Drama|Sci-Fi
898      1196  Star Wars: Episode V - The Empire Strikes Back...            Action|Adventure|Sci-Fi
1298     1732                          Big Lebowski, The (1998)                        Comedy|Crime
1498     2023                  Godfather: Part III, The (1990)         Crime|Drama|Mystery|Thriller
1577     2116                     Lord of the Rings, The (1978)  Adventure|Animation|Children|Fantasy
2001     2662                      War of the Worlds, The (1953)                Action|Drama|Sci-Fi
2226     2959                                Fight Club (1999)          Action|Crime|Drama|Thriller
4356     6373                            Bruce Almighty (2003)         Comedy|Drama|Fantasy|Romance
5972    35836                   40-Year-Old Virgin, The (2005)                       Comedy|Romance
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

15. Which user has rated the most movies?

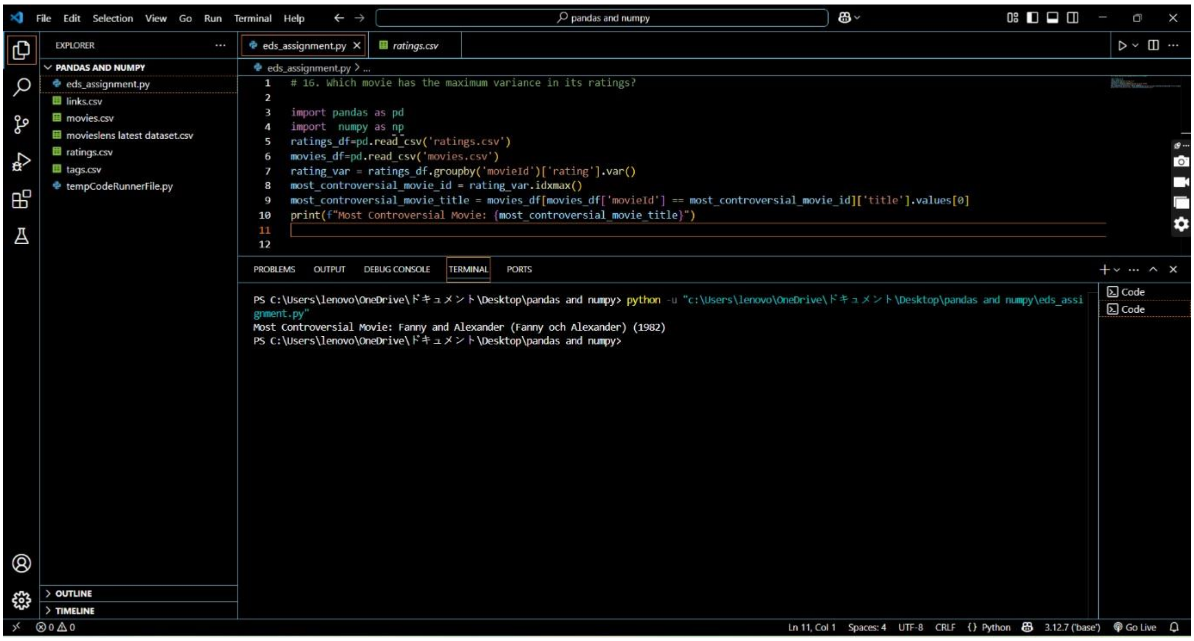

File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
Welcome
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Ln 16, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 15. Find the user who rated the most movies

import pandas as pd
import  numpy as np
movies_df=pd.read_csv('movies.csv')
ratings_df=pd.read_csv('ratings.csv')
top_user_id = ratings_df['userId'].value_counts().idxmax()
print(f"User with most ratings: User ID {top_user_id}")
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assignment.py"
User with most ratings: User ID 414
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

16.which movie has the maximum variance in its rating?


File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
ratings.csv
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Code
Ln 11, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

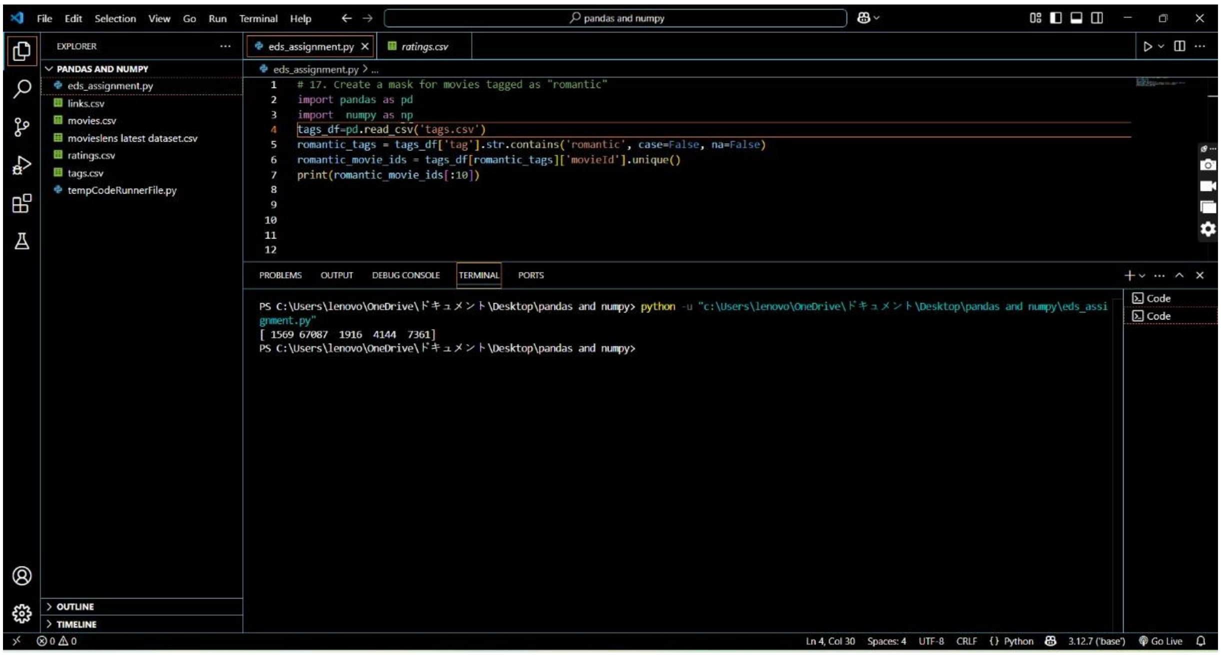
```python
# 16. Which movie has the maximum variance in its ratings?

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
movies_df=pd.read_csv('movies.csv')
rating_var = ratings_df.groupby('movieId')['rating'].var()
most_controversial_movie_id = rating_var.idxmax()
most_controversial_movie_title = movies_df[movies_df['movieId'] == most_controversial_movie_id]['title'].values[0]
print(f"Most Controversial Movie: {most_controversial_movie_title}")

```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
Most Controversial Movie: Fanny and Alexander (Fanny och Alexander) (1982)
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

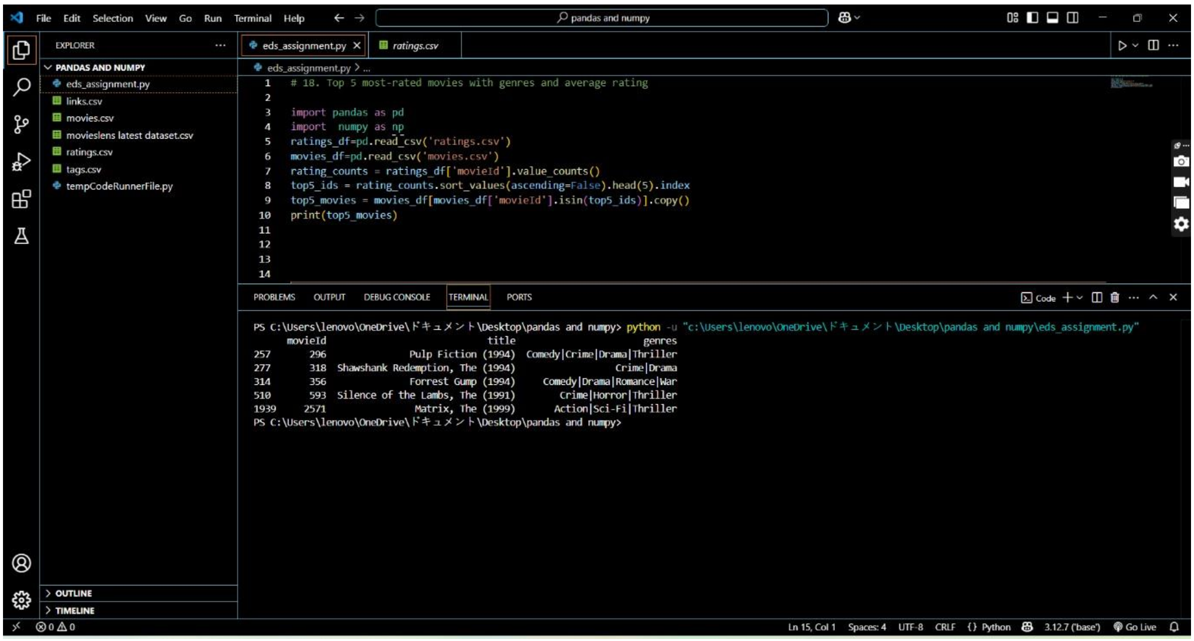
17. create a mask for movies tagged as "romantic"



```python
# 17. Create a mask for movies tagged as "romantic"
import pandas as pd
import  numpy as np
tags_df=pd.read_csv('tags.csv')
romantic_tags = tags_df['tag'].str.contains('romantic', case=False, na=False)
romantic_movie_ids = tags_df[romantic_tags]['movieId'].unique()
print(romantic_movie_ids[:10])
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
[ 1569 67087  1916  4144  7361]
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

18. Top 5 most-rated movies with genres and average rating

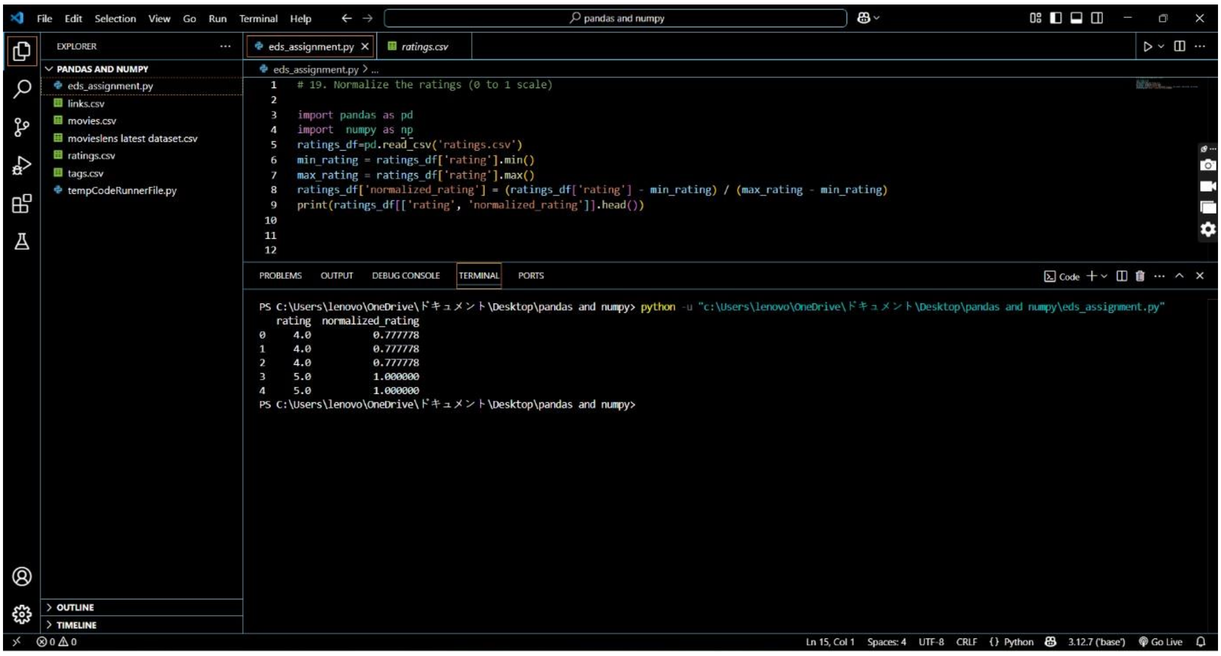


```python
# 18. Top 5 most-rated movies with genres and average rating

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
movies_df=pd.read_csv('movies.csv')
rating_counts = ratings_df['movieId'].value_counts()
top5_ids = rating_counts.sort_values(ascending=False).head(5).index
top5_movies = movies_df[movies_df['movieId'].isin(top5_ids)].copy()
print(top5_movies)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assignment.py"
      movieId                             title                       genres
257       296               Pulp Fiction (1994)  Comedy|Crime|Drama|Thriller
277       318  Shawshank Redemption, The (1994)                  Crime|Drama
314       356                Forrest Gump (1994)     Comedy|Drama|Romance|War
510       593  Silence of the Lambs, The (1991)        Crime|Horror|Thriller
1939     2571                 Matrix, The (1999)       Action|Sci-Fi|Thriller
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

19. Normalize the ratings (0 to 1 scale)



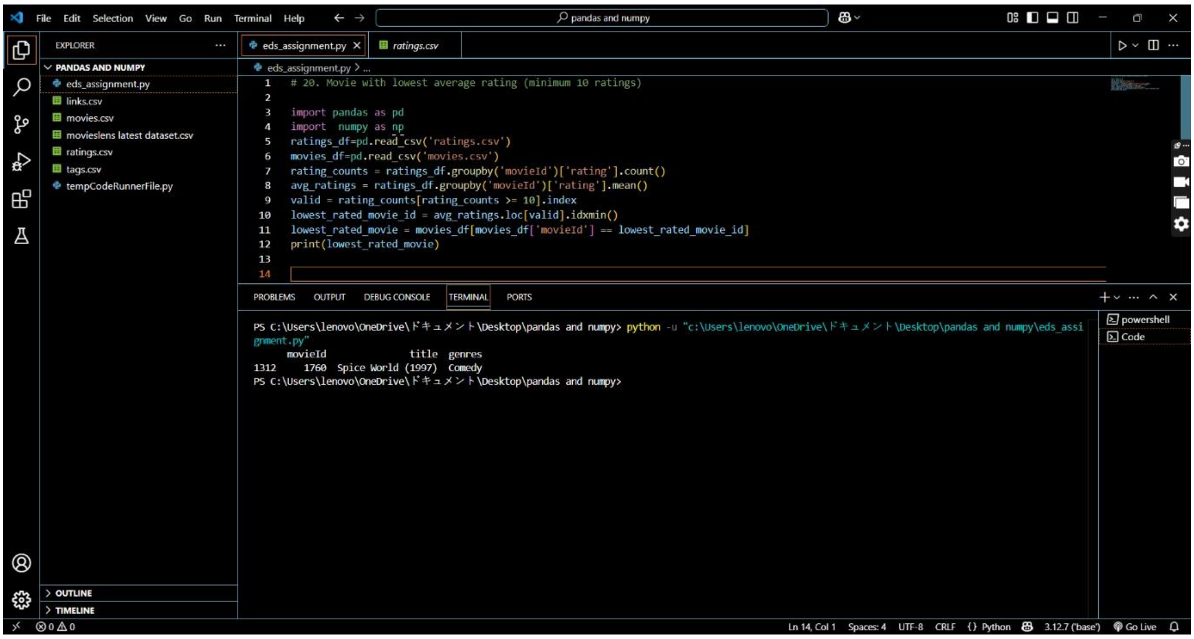
File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
ratings.csv
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Code
Ln 15, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 19. Normalize the ratings (0 to 1 scale)

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
min_rating = ratings_df['rating'].min()
max_rating = ratings_df['rating'].max()
ratings_df['normalized_rating'] = (ratings_df['rating'] - min_rating) / (max_rating - min_rating)
print(ratings_df[['rating', 'normalized_rating']].head())
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assignment.py"
   rating  normalized_rating
0     4.0           0.777778
1     4.0           0.777778
2     4.0           0.777778
3     5.0           1.000000
4     5.0           1.000000
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```

20. Movie with lowest average rating (minimum 10 ratings)



File Edit Selection View Go Run Terminal Help
pandas and numpy
EXPLORER
PANDAS AND NUMPY
eds_assignment.py
links.csv
movies.csv
movieslens latest dataset.csv
ratings.csv
tags.csv
tempCodeRunnerFile.py
OUTLINE
TIMELINE
eds_assignment.py
ratings.csv
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
powershell
Code
Ln 14, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.7 ('base') Go Live

```python
# 20. Movie with lowest average rating (minimum 10 ratings)

import pandas as pd
import  numpy as np
ratings_df=pd.read_csv('ratings.csv')
movies_df=pd.read_csv('movies.csv')
rating_counts = ratings_df.groupby('movieId')['rating'].count()
avg_ratings = ratings_df.groupby('movieId')['rating'].mean()
valid = rating_counts[rating_counts >= 10].index
lowest_rated_movie_id = avg_ratings.loc[valid].idxmin()
lowest_rated_movie = movies_df[movies_df['movieId'] == lowest_rated_movie_id]
print(lowest_rated_movie)
```

```
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy> python -u "c:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy\eds_assi
gnment.py"
      movieId               title  genres
1312     1760  Spice World (1997)  Comedy
PS C:\Users\lenovo\OneDrive\ドキュメント\Desktop\pandas and numpy>
```