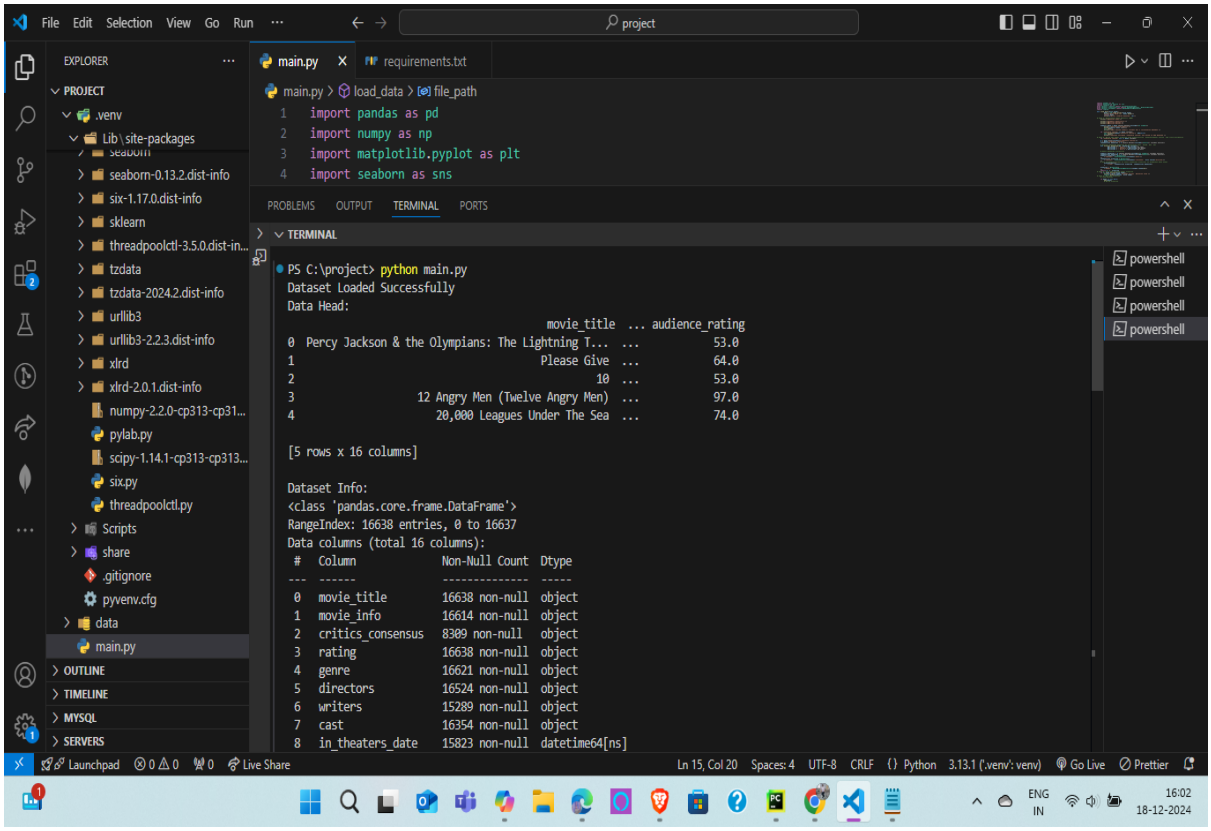
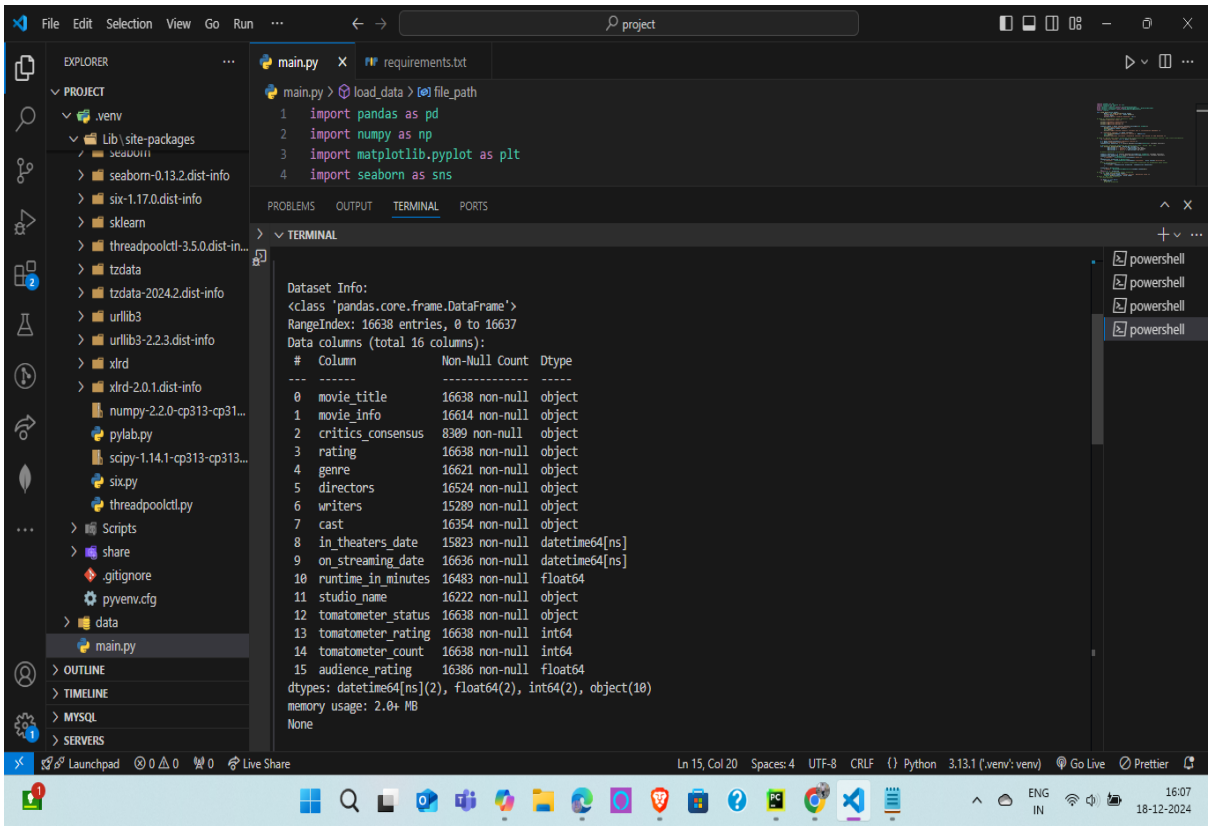
Result:


File Edit Selection View Go Run
project
EXPLORER
PROJECT
.venv
Lib\site-packages
seaborn-0.13.2.dist-info
six-1.17.0.dist-info
sklearn
threadpoolctl-3.5.0.dist-in...
tzdata
tzdata-2024.2.dist-info
urllib3
urllib3-2.2.3.dist-info
xlrd
xlrd-2.0.1.dist-info
numpy-2.2.0-cp313-cp31...
pylab.py
scipy-1.14.1-cp313-cp313...
six.py
threadpoolctl.py
Scripts
share
.gitignore
pyvenv.cfg
data
main.py
OUTLINE
TIMELINE
MYSQL
SERVERS
main.py
requirements.txt
main.py > load_data > file_path
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
PROBLEMS OUTPUT TERMINAL PORTS
TERMINAL
PS C:\project> python main.py
Dataset Loaded Successfully
Data Head:
movie_title ... audience_rating
0 Percy Jackson & the Olympians: The Lightning T... ... 53.0
1 Please Give ... 64.0
2 10 ... 53.0
3 12 Angry Men (Twelve Angry Men) ... 97.0
4 20,000 Leagues Under The Sea ... 74.0
[5 rows x 16 columns]
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16638 entries, 0 to 16637
Data columns (total 16 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 movie_title 16638 non-null object
1 movie_info 16614 non-null object
2 critics_consensus 8309 non-null object
3 rating 16638 non-null object
4 genre 16621 non-null object
5 directors 16524 non-null object
6 writers 15289 non-null object
7 cast 16354 non-null object
8 in_theaters_date 15823 non-null datetime64[ns]
powershell
powershell
powershell
powershell
Launchpad
Live Share
Ln 15, Col 20 Spaces: 4 UTF-8 CRLF Python 3.13.1 ('.venv': venv)
Go Live
Prettier
ENG
IN
16:02
18-12-2024


File Edit Selection View Go Run
project
EXPLORER
PROJECT
.venv
Lib\site-packages
seaborn-0.13.2.dist-info
six-1.17.0.dist-info
sklearn
threadpoolctl-3.5.0.dist-in...
tzdata
tzdata-2024.2.dist-info
urllib3
urllib3-2.2.3.dist-info
xlrd
xlrd-2.0.1.dist-info
numpy-2.2.0-cp313-cp31...
pylab.py
scipy-1.14.1-cp313-cp313...
six.py
threadpoolctl.py
Scripts
share
.gitignore
pyvenv.cfg
data
main.py
OUTLINE
TIMELINE
MYSQL
SERVERS
main.py
requirements.txt
main.py > load_data > file_path
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
PROBLEMS OUTPUT TERMINAL PORTS
TERMINAL
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16638 entries, 0 to 16637
Data columns (total 16 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 movie_title 16638 non-null object
1 movie_info 16614 non-null object
2 critics_consensus 8309 non-null object
3 rating 16638 non-null object
4 genre 16621 non-null object
5 directors 16524 non-null object
6 writers 15289 non-null object
7 cast 16354 non-null object
8 in_theaters_date 15823 non-null datetime64[ns]
9 on_streaming_date 16636 non-null datetime64[ns]
10 runtime_in_minutes 16483 non-null float64
11 studio_name 16222 non-null object
12 tomatometer_status 16638 non-null object
13 tomatometer_rating 16638 non-null int64
14 tomatometer_count 16638 non-null int64
15 audience_rating 16386 non-null float64
dtypes: datetime64[ns](2), float64(2), int64(2), object(10)
memory usage: 2.0+ MB
None
powershell
powershell
powershell
powershell
Launchpad
Live Share
Ln 15, Col 20 Spaces: 4 UTF-8 CRLF Python 3.13.1 ('.venv': venv)
Go Live
Prettier
ENG
IN
16:07
18-12-2024

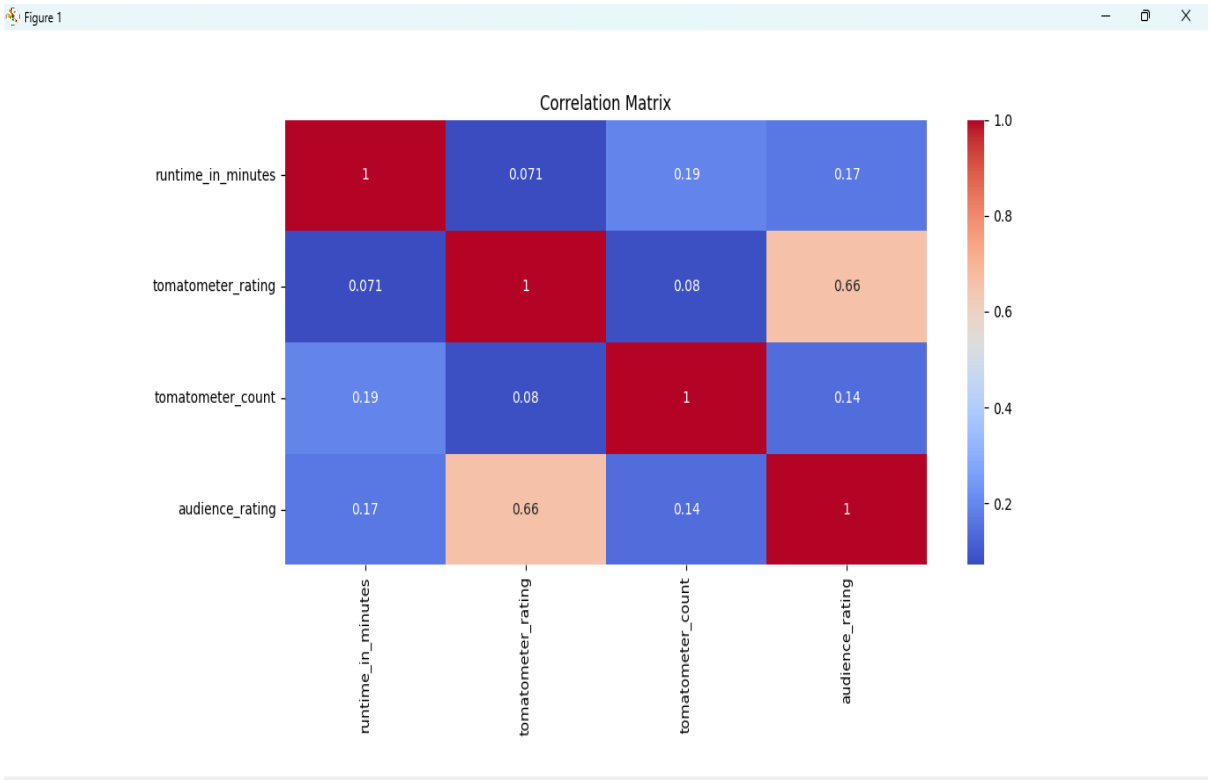
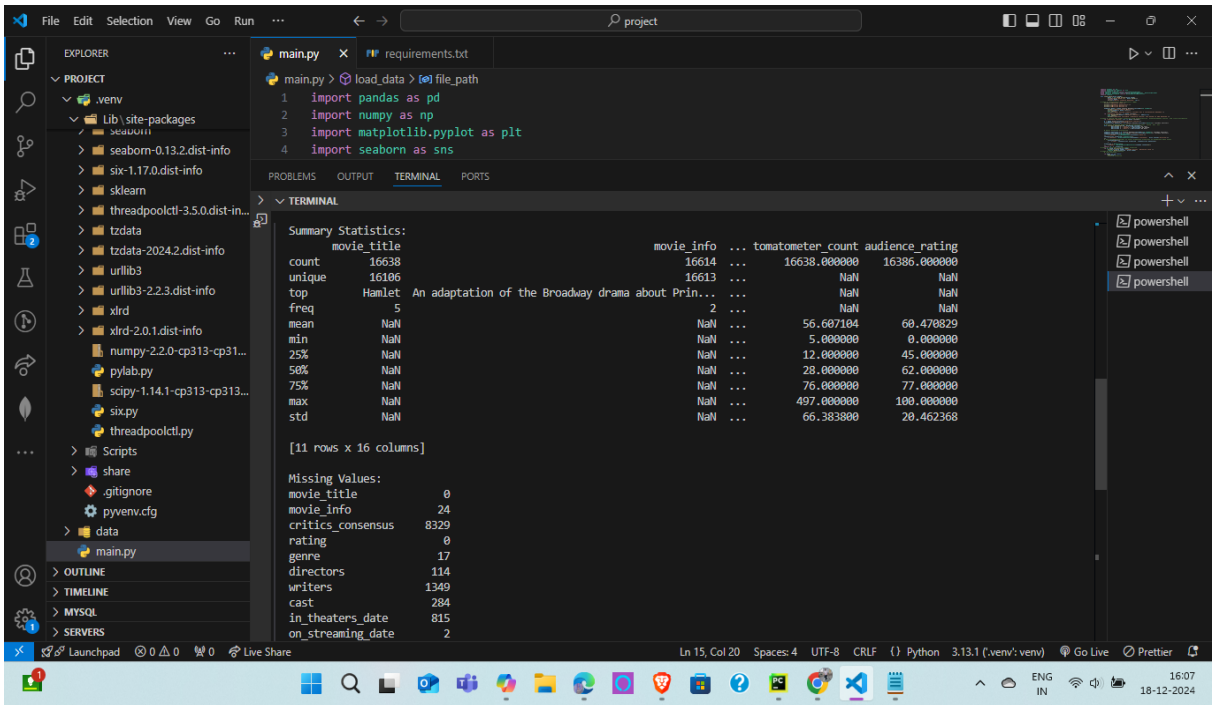
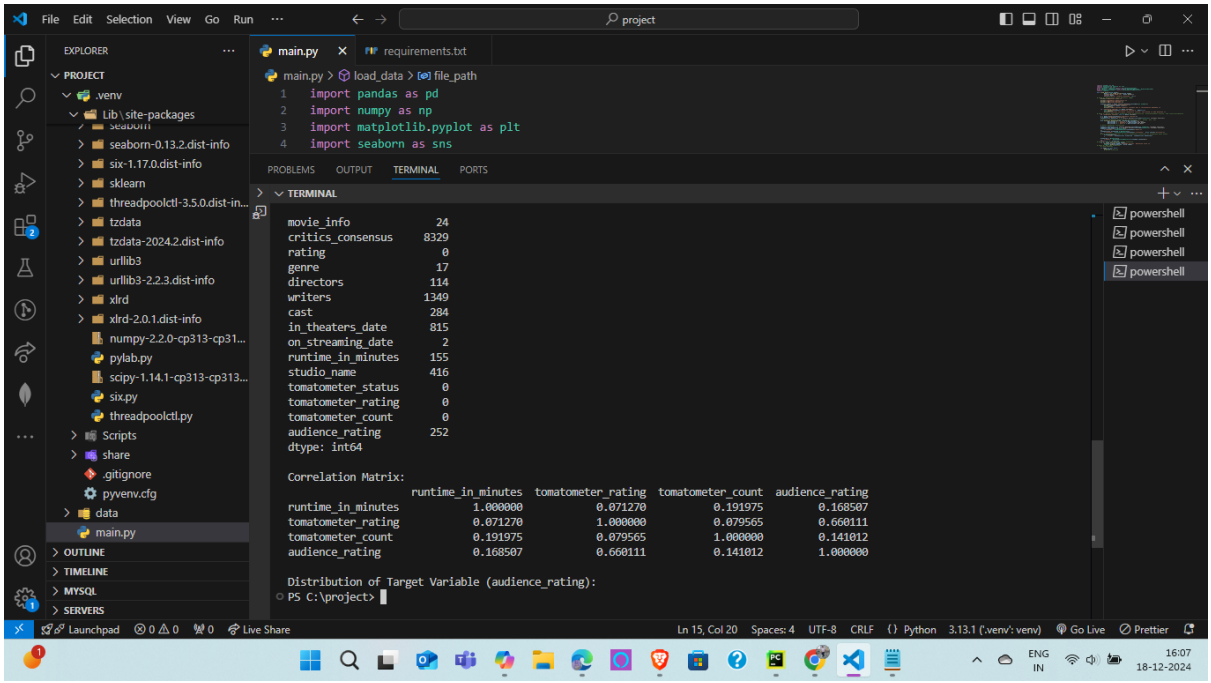


Figure 1
Correlation Matrix
runtime_in_minutes
tomatometer_rating
tomatometer_count
audience_rating
1
0.071
0.19
0.17
0.071
1
0.08
0.66
0.19
0.08
1
0.14
0.17
0.66
0.14
1
1.0
0.8
0.6
0.4
0.2
(x, y) = (,)
ENG
IN
16:02
18-12-2024


File Edit Selection View Go Run
project
EXPLORER
PROJECT
.venv
Lib\site-packages
main.py
requirements.txt
main.py > load_data > file_path
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
PROBLEMS OUTPUT TERMINAL PORTS
TERMINAL
powershell
Summary Statistics:
movie_title movie_info ... tomatometer_count audience_rating
count 16638 16614 ... 16638.000000 16386.000000
unique 16106 16613 ... NaN NaN
top Hamlet An adaptation of the Broadway drama about Prin... ... NaN NaN
freq 5 2 ... NaN NaN
mean NaN NaN ... 56.607104 60.470829
min NaN NaN ... 5.000000 0.000000
25% NaN NaN ... 12.000000 45.000000
50% NaN NaN ... 28.000000 62.000000
75% NaN NaN ... 76.000000 77.000000
max NaN NaN ... 497.000000 100.000000
std NaN NaN ... 66.383800 20.462368
[11 rows x 16 columns]
Missing Values:
movie_title 0
movie_info 24
critics_consensus 8329
rating 0
genre 17
directors 114
writers 1349
cast 284
in_theaters_date 815
on_streaming_date 2
OUTLINE
TIMELINE
MYSQL
SERVERS
Launchpad
Live Share
Ln 15, Col 20 Spaces: 4 UTF-8 CRLF Python 3.13.1 ('.venv': venv) Go Live Prettier
ENG
IN
16:07
18-12-2024


File Edit Selection View Go Run
project
EXPLORER
PROJECT
.venv
Lib\site-packages
main.py
requirements.txt
main.py > load_data > file_path
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
PROBLEMS OUTPUT TERMINAL PORTS
TERMINAL
movie_info 24
critics_consensus 8329
rating 0
genre 17
directors 114
writers 1349
cast 284
in_theaters_date 815
on_streaming_date 2
runtime_in_minutes 155
studio_name 416
tomatometer_status 0
tomatometer_rating 0
tomatometer_count 0
audience_rating 252
dtype: int64
Correlation Matrix:
runtime_in_minutes tomatometer_rating tomatometer_count audience_rating
runtime_in_minutes 1.000000 0.071270 0.191975 0.168507
tomatometer_rating 0.071270 1.000000 0.079565 0.660111
tomatometer_count 0.191975 0.079565 1.000000 0.141012
audience_rating 0.168507 0.660111 0.141012 1.000000
Distribution of Target Variable (audience_rating):
PS C:\project>
powershell
powershell
powershell
powershell
Launchpad
Live Share
Ln 15, Col 20 Spaces: 4 UTF-8 CRLF Python 3.13.1 ('.venv': venv) Go Live Prettier
ENG IN
16:07
18-12-2024

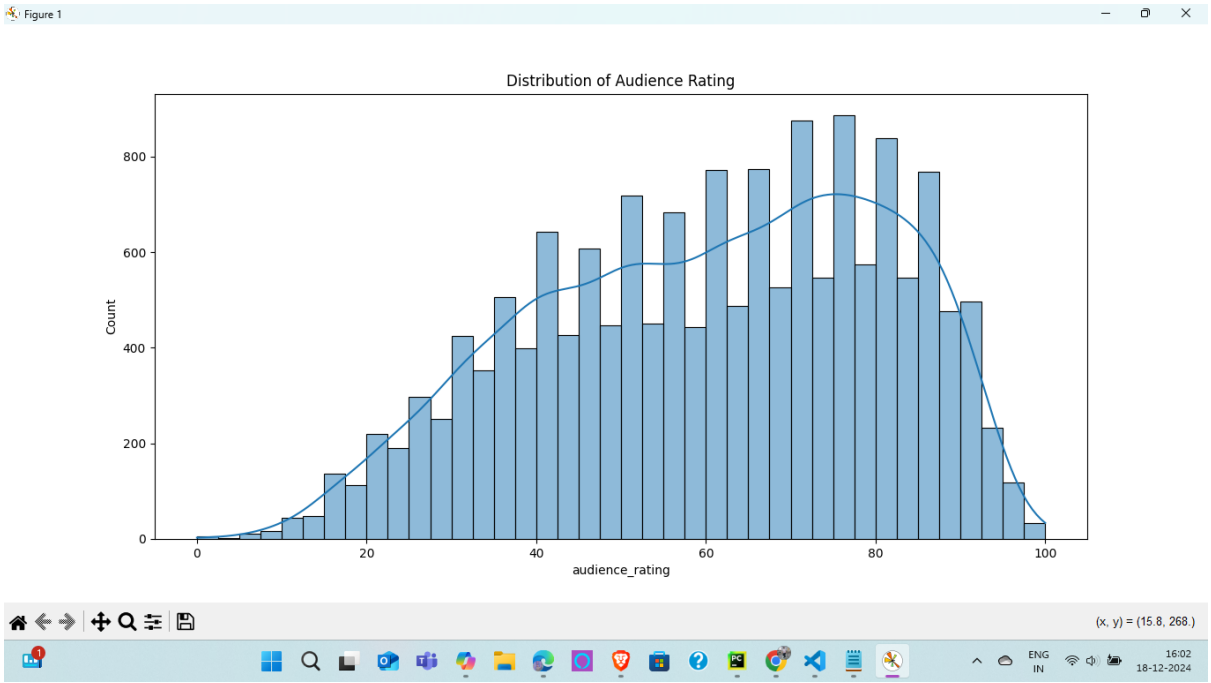


Figure 1
Distribution of Audience Rating
Count
audience_rating
0
200
400
600
800
0
20
40
60
80
100
(x, y) = (15.8, 268.)
ENG IN
16:02
18-12-2024