

```

import pandas as pd

data = pd.read_csv("C:/Users/anuru/OneDrive/Desktop/Luminar DSML/Data
Science/Pandas1/ASSIGNMENT 3/movies_cleaned_pandas.csv",header = None)

df = pd.DataFrame(data)

df.columns = ['no','name','year','rating','duration']

print(df.head())
print("="*100)

# 1. Find row count
print('Q.1')
print("Number of Rows = ",len(df))
print("="*100)

# 2. Remove duplicates and find row count
df.drop_duplicates(inplace=True)
print('Q.2')
print("Number of Rows after removing duplicates = ",len(df))
print("="*100)

# 3. Sort data set by release year in des order
print('Q.3')
print(df.sort_values(by='year',ascending=False).head())
print("="*100)

# 4. Find rating mxm 5 movies name,year,rating
print('Q.4')
print(df.sort_values(by='rating',ascending=False)[['name','year','rating']].head(5))
print("="*100)

# 5. Find rating minimum 3 movies name,year,rtaing
print('Q.5')
print(df.sort_values(by='rating',ascending=True)[['name','year','rating']].head(3))
print("="*100)

# 6. Find Each year release movie count [count desc order]
print('Q.6')
print(df.groupby('year')['year'].count().sort_values(ascending=False))
print("="*100)

# 7. Each rating count [count desc order]
print('Q.7')
print(df.groupby('rating')['rating'].count().sort_values(ascending=False))
print("="*100)

```

```
# 8. 2008 and rating above 3 [collect]
print('Q.8')
print('Number of rows of movies of year 2008 and rating above 3 = ',len(df.loc[(df['year'] == 2008) &
(df['rating'] > 3)]))
print('='*100)
```

```
# 9. Find duration mxm 1 movies name,year,rating,duration
print('Q.9')
print(df.sort_values(by='duration',ascending=False)[df.columns[1:]].head(1))
print('='*100)
```

```
# 10. Find rating mnm 1 movies name,year,rating,duration
print('Q.10')
print(df.sort_values(by='rating',ascending=True)[df.columns[1:]].head(1))
print('='*100)
```

```
# 11. Rating above 4 and relase year above 2005
#   A. Rating mxm movies full data
#   B. Rating mnm movies full data
print('Q.11')
print(df.loc[(df['rating'] > 4) & (df['year'] > 2005)].sort_values(by='rating',ascending=False))
print('='*100)
print('='*100)
print(df.loc[(df['rating'] > 4) & (df['year'] > 2005)].sort_values(by='rating',ascending=True))
print('='*100)
```

```
# 12. 2008 movies count
print('Q.12')
print("2008 movies count = ",len(df.loc[df['year']==2008]))
print('='*100)
```

```
# 13. 1975-2000 movies collect
#   A. Row count
print('Q.13')
print('1975-2000 movies row count = ',len(df.loc[(df['year'] >= 1975 ) & (df['year'] < 2000)]))
print('='*100)
```

```
# 14. 1975-2000 and rating above 3.5 total row count
print('Q.14')
print('1975-2000 movies and rating > 3.5 row count = ',len(df.loc[(df['year'] >= 1975 ) & (df['year'] <
2000) & (df['rating'] > 3.5)]))
print('='*100)
```