```
import pandas as pd
```

```
Start coding or generate with AI.
```

```
! pip install pandas
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: numpy>=1.23.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.
```

```
df_xlsx = pd.read_excel("/content/Movies_Dataset.xlsx")
print(df_xlsx)
```

```
     Movie Name      Genre  Release_Year  IMDb_Rating  Revenue
0       Movie 1     Sci-Fi          2000          6.0  1680.53
1       Movie 2  Adventure          2000          2.2   987.24
2       Movie 3    Romance          2002          5.2  1122.13
3       Movie 4    Romance          2001          4.5   581.84
4       Movie 5     Action          2012          6.6  1205.51
..          ...        ...           ...          ...      ...
995   Movie 996     Sci-Fi          2013          8.7   456.39
996   Movie 997  Adventure          1993          6.2  1255.27
997   Movie 998    Fantasy          2016          8.1  1161.96
998   Movie 999     Action          1996          8.9   941.67
999  Movie 1000     Horror          1995          9.4  1390.70

[1000 rows x 5 columns]
```

```
df_xlsx.head()
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue
0	Movie 1	Sci-Fi	2000	6.0	1680.53
1	Movie 2	Adventure	2000	2.2	987.24
2	Movie 3	Romance	2002	5.2	1122.13
3	Movie 4	Romance	2001	4.5	581.84

```
df_xlsx.tail()
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue
995	Movie 996	Sci-Fi	2013	8.7	456.39
996	Movie 997	Adventure	1993	6.2	1255.27
997	Movie 998	Fantasy	2016	8.1	1161.96
998	Movie 999	Action	1996	8.9	941.67

```
df_xlsx.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Movie Name    1000 non-null   object 
 1   Genre         1000 non-null   object 
 2   Release_Year  1000 non-null   int64  
 3   IMDb_Rating   1000 non-null   float64
```

```
 4   Revenue       1000 non-null   float64
dtypes: float64(2), int64(1), object(2)
memory usage: 39.2+ KB
```

```
df_xlsx.shape
```

```
(1000, 5)
```

```
df_xlsx[df_xlsx["Genre"] == "Action"]
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue
4	Movie 5	Action	2012	6.6	1205.51
34	Movie 35	Action	2009	8.7	545.60
64	Movie 65	Action	2023	7.7	1137.79
65	Movie 66	Action	1999	3.1	1335.85
73	Movie 74	Action	2001	4.1	350.80
...	...	...	...	...	...
930	Movie 931	Action	1994	8.6	1863.66
944	Movie 945	Action	2004	7.5	1709.81
980	Movie 981	Action	2022	5.2	28.40
988	Movie 989	Action	2023	4.1	231.67
998	Movie 999	Action	1996	8.9	941.67

```
df_xlsx[df_xlsx["Release_Year"] > 2000]
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue
2	Movie 3	Romance	2002	5.2	1122.13
3	Movie 4	Romance	2001	4.5	581.84
4	Movie 5	Action	2012	6.6	1205.51
6	Movie 7	Mystery	2021	5.1	1683.99
7	Movie 8	Comedy	2005	4.0	1062.13
...	...	...	...	...	...
988	Movie 989	Action	2023	4.1	231.67
989	Movie 990	Fantasy	2004	1.6	1056.58
991	Movie 992	Thriller	2004	9.2	251.25
995	Movie 996	Sci-Fi	2013	8.7	456.39
997	Movie 998	Fantasy	2016	8.1	1161.96

```
df_xlsx[df_xlsx["IMDb_Rating"] > 8.5]
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue
34	Movie 35	Action	2009	8.7	545.60
51	Movie 52	Mystery	1998	9.0	1089.93
56	Movie 57	Comedy	1993	9.2	1586.13
74	Movie 75	Adventure	1999	8.6	1688.49
81	Movie 82	Mystery	1997	8.6	1238.32
...	...	...	...	...	...
991	Movie 992	Thriller	2004	9.2	251.25
992	Movie 993	Horror	1985	9.6	951.23
995	Movie 996	Sci-Fi	2013	8.7	456.39
998	Movie 999	Action	1996	8.9	941.67
999	Movie 1000	Horror	1995	9.4	1390.70

```
df_xlsx[["Movie Name", "Genre"]]
```

	Movie Name	Genre
0	Movie 1	Sci-Fi
1	Movie 2	Adventure
2	Movie 3	Romance
3	Movie 4	Romance
4	Movie 5	Action
...	...	...
995	Movie 996	Sci-Fi
996	Movie 997	Adventure
997	Movie 998	Fantasy
998	Movie 999	Action
999	Movie 1000	Horror

```
df_xlsx[(df_xlsx["Release_Year"] >= 1990) & (df_xlsx["Release_Year"] <= 2010)]
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue
0	Movie 1	Sci-Fi	2000	6.0	1680.53
1	Movie 2	Adventure	2000	2.2	987.24
2	Movie 3	Romance	2002	5.2	1122.13
3	Movie 4	Romance	2001	4.5	581.84
5	Movie 6	Fantasy	1997	2.4	485.22
...	...	...	...	...	...
991	Movie 992	Thriller	2004	9.2	251.25
994	Movie 995	Sci-Fi	1997	2.0	342.98
996	Movie 997	Adventure	1993	6.2	1255.27
998	Movie 999	Action	1996	8.9	941.67
999	Movie 1000	Horror	1995	9.4	1390.70

```
df_xlsx["Genre"].value_counts()
```

Genre	count
Drama	113
Comedy	110
Fantasy	107
Thriller	107
Action	104
Adventure	96
Sci-Fi	93
Romance	91
Horror	90
Mystery	89

```
df_xlsx.loc[df_xlsx["IMDb_Rating"].idxmax()]
```

	106
Movie Name	Movie 107
Genre	Romance
Release_Year	2007
IMDb_Rating	10.0
Revenue	1837.45

```
df_xlsx["IMDb_Rating"].mean()
```

```
5.458600000000001
```

```
df_xlsx["Revenue"].sum()
```

```
1049716.26
```

```
df_xlsx[df_xlsx["Revenue"] > 500]
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue
0	Movie 1	Sci-Fi	2000	6.0	1680.53
1	Movie 2	Adventure	2000	2.2	987.24
2	Movie 3	Romance	2002	5.2	1122.13
3	Movie 4	Romance	2001	4.5	581.84
4	Movie 5	Action	2012	6.6	1205.51
...	...	...	...	...	...
993	Movie 994	Adventure	1989	7.3	1273.65
996	Movie 997	Adventure	1993	6.2	1255.27
997	Movie 998	Fantasy	2016	8.1	1161.96
998	Movie 999	Action	1996	8.9	941.67
999	Movie 1000	Horror	1995	9.4	1390.70

```
df_xlsx["IMDb_Rank"] = df_xlsx["IMDb_Rating"].rank(method="dense", ascending=False)
```

```
df_xlsx["IMDb_Rank"] = df_xlsx["IMDb_Rating"].rank(method="dense", ascending=False)
```

```
df_xlsx.sort_values(by="Release_Year", ascending=False)
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue	IMDb_Rank
320	Movie 321	Romance	2024	4.2	477.47	59.0
211	Movie 212	Comedy	2024	1.5	837.32	86.0
489	Movie 490	Thriller	2024	3.2	1536.79	69.0
808	Movie 809	Action	2024	6.0	714.96	41.0
917	Movie 918	Horror	2024	1.1	1042.80	90.0
...	...	...	...	...	...	...
509	Movie 510	Fantasy	1980	6.3	1260.11	38.0
798	Movie 799	Drama	1980	3.1	1587.13	70.0
496	Movie 497	Fantasy	1980	4.7	1683.97	54.0
794	Movie 795	Adventure	1980	4.5	1346.01	56.0
170	Movie 171	Mystery	1980	4.6	1684.78	55.0

```
df_xlsx[df_xlsx.duplicated(subset=["Movie Name"], keep=False)]
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue	IMDb_Rank
9	Movie 10	Drama	1996	3.2	1352.66	69.0
10	Movie 11	Mystery	2003	6.3	596.43	38.0
11	Movie 12	Sci-Fi	1988	3.0	1134.01	71.0
12	Movie 10	Drama	1996	3.2	1352.66	69.0
13	Movie 11	Mystery	2003	6.3	596.43	38.0
14	Movie 12	Sci-Fi	1988	3.0	1134.01	71.0
15	Movie 16	Comedy	2006	2.3	1322.79	78.0
16	Movie 17	Horror	2007	6.2	1114.31	39.0
17	Movie 18	Adventure	1998	7.9	1426.73	22.0
18	Movie 19	Comedy	1986	6.0	1783.44	41.0
19	Movie 20	Fantasy	1982	6.9	1598.08	32.0
20	Movie 16	Comedy	2006	2.3	1322.79	78.0
21	Movie 17	Horror	2007	6.2	1114.31	39.0
22	Movie 18	Adventure	1998	7.9	1426.73	22.0
23	Movie 19	Comedy	1986	6.0	1783.44	41.0

```
df_xlsx.drop_duplicates(subset=["Movie Name"])
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue	IMDb_Rank
0	Movie 1	Sci-Fi	2000	6.0	1680.53	41.0
1	Movie 2	Adventure	2000	2.2	987.24	79.0
2	Movie 3	Romance	2002	5.2	1122.13	49.0
3	Movie 4	Romance	2001	4.5	581.84	56.0
4	Movie 5	Action	2012	6.6	1205.51	35.0
...	...	...	...	...	...	...
995	Movie 996	Sci-Fi	2013	8.7	456.39	14.0
996	Movie 997	Adventure	1993	6.2	1255.27	39.0
997	Movie 998	Fantasy	2016	8.1	1161.96	20.0
998	Movie 999	Action	1996	8.9	941.67	12.0
999	Movie 1000	Horror	1995	9.4	1390.70	7.0

```
df_xlsx["Genre"] = df_xlsx["Genre"].replace("Sci-Fi", "Science Fiction")

df_xlsx.head(10)
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue	IMDb_Rank
0	Movie 1	Science Fiction	2000	6.0	1680.53	41.0
1	Movie 2	Adventure	2000	2.2	987.24	79.0
2	Movie 3	Romance	2002	5.2	1122.13	49.0
3	Movie 4	Romance	2001	4.5	581.84	56.0
4	Movie 5	Action	2012	6.6	1205.51	35.0
5	Movie 6	Fantasy	1997	2.4	485.22	77.0
6	Movie 7	Mystery	2021	5.1	1683.99	50.0
7	Movie 8	Comedy	2005	4.0	1062.13	61.0
8	Movie 9	Horror	1992	2.8	1592.74	73.0

```
df_xlsx.tail(5)
```

	Movie Name	Genre	Release_Year	IMDb_Rating	Revenue	IMDb_Rank
995	Movie 996	Science Fiction	2013	8.7	456.39	14.0
996	Movie 997	Adventure	1993	6.2	1255.27	39.0
997	Movie 998	Fantasy	2016	8.1	1161.96	20.0
998	Movie 999	Action	1996	8.9	941.67	12.0

```
df_xlsx.groupby("Genre")["IMDb_Rating"].mean()
```

Genre	IMDb_Rating
Action	5.670192
Adventure	5.012500
Comedy	5.442727
Drama	5.173451
Fantasy	5.200000
Horror	5.814444
Mystery	5.778652
Romance	5.938462
Science Fiction	5.445161
Thriller	5.267290

```
df_xlsx["Release_Year"].value_counts()
```

Release_Year	count
2024	32
1989	30
2000	29
2007	29
1998	29
1985	28
1995	27
2013	26
1992	26
1996	26
1980	26
1999	25
2022	25