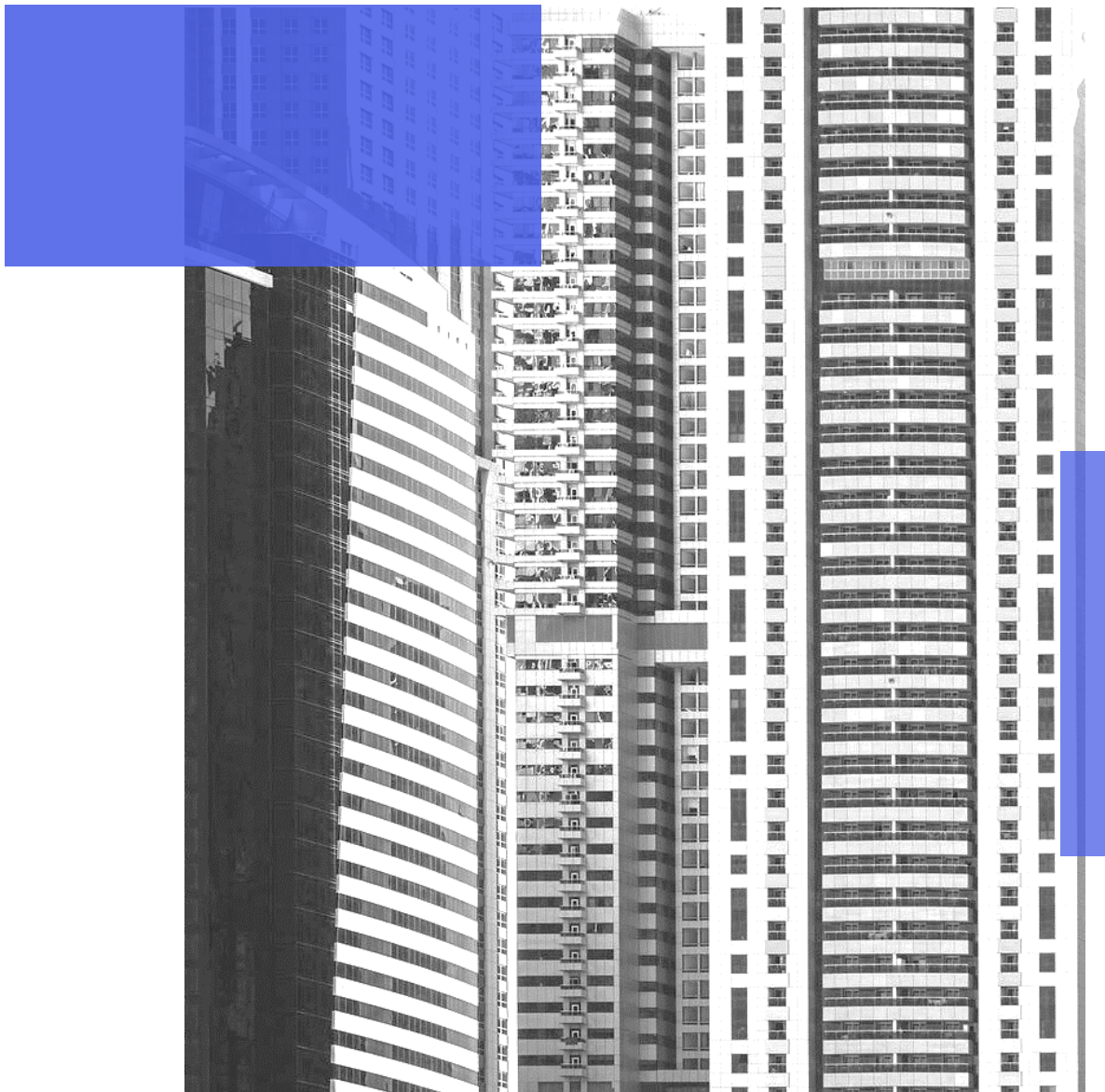


MUHAMMAD AHMED

PYTHON PROJECT





AGENDA

INTRODUCTION

DATA CLEANING

ASK QUESTIONS?

PRESENT THE RESULTS

INTRO TO DATA

```
[23]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 10866 entries, 0 to 10865
```

```
Data columns (total 22 columns):
```

#	Column	Non-Null Count	Dtype
0	id	10866 non-null	int64
1	imdb_id	10856 non-null	object
2	popularity	10866 non-null	float64
3	budget	10866 non-null	int64
4	revenue	10866 non-null	int64
5	original_title	10866 non-null	object
6	cast	10790 non-null	object
7	homepage	2936 non-null	object
8	director	10822 non-null	object
9	tagline	8042 non-null	object
10	keywords	9373 non-null	object
11	overview	10862 non-null	object
12	runtime	10866 non-null	int64
13	genres	10843 non-null	object
14	production_companies	9836 non-null	object
15	release_date	10866 non-null	object
16	vote_count	10866 non-null	int64
17	vote_average	10866 non-null	float64
18	release_year	10866 non-null	int64
19	budget_adj	10866 non-null	float64
20	revenue_adj	10866 non-null	float64
21	percentage	10866 non-null	object

```
dtypes: float64(4), int64(6), object(12)
```

```
memory usage: 1.8+ MB
```

INTRO

df.describe()

	id	popularity	budget	revenue	runtime	vote_count	vote_average	release_year	budget_adj	revenue_adj
count	10866.000000	10866.000000	1.086600e+04	1.086600e+04	10866.000000	10866.000000	10866.000000	10866.000000	1.086600e+04	1.086600e+04
mean	66064.177434	0.646441	1.462570e+07	3.982332e+07	102.070863	217.389748	5.974922	2001.322658	1.755104e+07	5.136436e+07
std	92130.136561	1.000185	3.091321e+07	1.170035e+08	31.381405	575.619058	0.935142	12.812941	3.430616e+07	1.446325e+08
min	5.000000	0.000065	0.000000e+00	0.000000e+00	0.000000	10.000000	1.500000	1960.000000	0.000000e+00	0.000000e+00
25%	10596.250000	0.207583	0.000000e+00	0.000000e+00	90.000000	17.000000	5.400000	1995.000000	0.000000e+00	0.000000e+00
50%	20669.000000	0.383856	0.000000e+00	0.000000e+00	99.000000	38.000000	6.000000	2006.000000	0.000000e+00	0.000000e+00
75%	75610.000000	0.713817	1.500000e+07	2.400000e+07	111.000000	145.750000	6.600000	2011.000000	2.085325e+07	3.369710e+07
max	417859.000000	32.985763	4.250000e+08	2.781506e+09	900.000000	9767.000000	9.200000	2015.000000	4.250000e+08	2.827124e+09

DATA CLEANING

```
df = df[df["cast"].notnull()]
df = df[df["genres"].notnull()]
df = df.query("budget_adj != 0 and revenue_adj != 0")
df
```

Calculated Field

```
df['profit'] = 100 * (df['revenue_adj'] - df['budget_adj']) / df['budget_adj']
df['profit'] = df['profit'].map(lambda x: f"{x:.2f}%")
df
```


“ WHICH MOVIES ARE WITH THE HIGHEST AND LOWEST BUDGET? ”

```
[28]: if 'budget_adj' in df.columns and 'original_title' in df.columns:

    highest_budget_movie = df[df['budget_adj'] == df['budget_adj'].max()]

    lowest_budget_movie = df[df['budget_adj'] == df['budget_adj'].min()]

    print("Highest Budget Movie:")
    print(highest_budget_movie[['original_title', 'budget_adj']])

    print("\nLowest Budget Movie:")
    print(lowest_budget_movie[['original_title', 'budget_adj']])

else:
    print("The required columns 'budget_adj' and 'original_title' are not available in the dataset.")
```

Highest Budget Movie:

	original_title	budget_adj
2244	The Warrior's Way	425000000.0

Lowest Budget Movie:

	original_title	budget_adj
3581	Love, Wedding, Marriage	0.969398

WHICH MOVIES ARE WITH THE HIGHEST AND LOWEST REVENUE? IS THE BIGGEST RETURN THE BIGGEST PROFIT?

```
if 'revenue_adj' in df.columns and 'original_title' in df.columns:

    highest_budget_movie = df[df['revenue_adj'] == df['revenue_adj'].max()]

    lowest_budget_movie = df[df['revenue_adj'] == df['revenue_adj'].min()]

    print("Highest revenue Movie:")
    print(highest_budget_movie[['original_title', 'revenue_adj']])

    print("\nLowest revenue Movie:")
    print(lowest_budget_movie[['original_title', 'revenue_adj']])

else:
    print("The required columns 'revenue_adj' and 'original_title' are not available in the dataset.")
```

Highest revenue Movie:

	original_title	revenue_adj
1386	Avatar	2.827124e+09

Lowest revenue Movie:

	original_title	revenue_adj
5067	Shattered Glass	2.370705

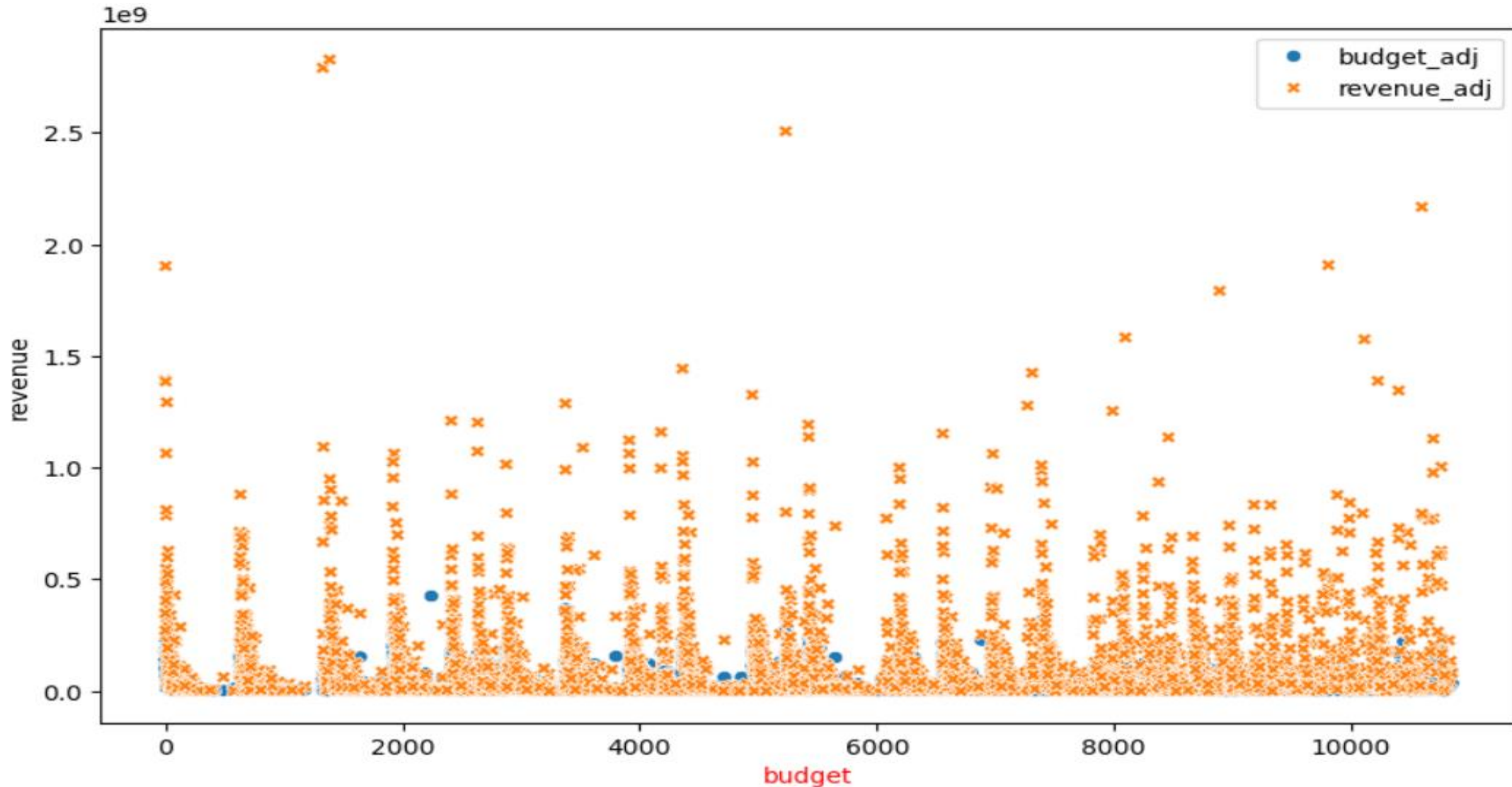
Highest profit Movie:

	original_title	profit
4376	Ted	998.74%

Lowest profit Movie:

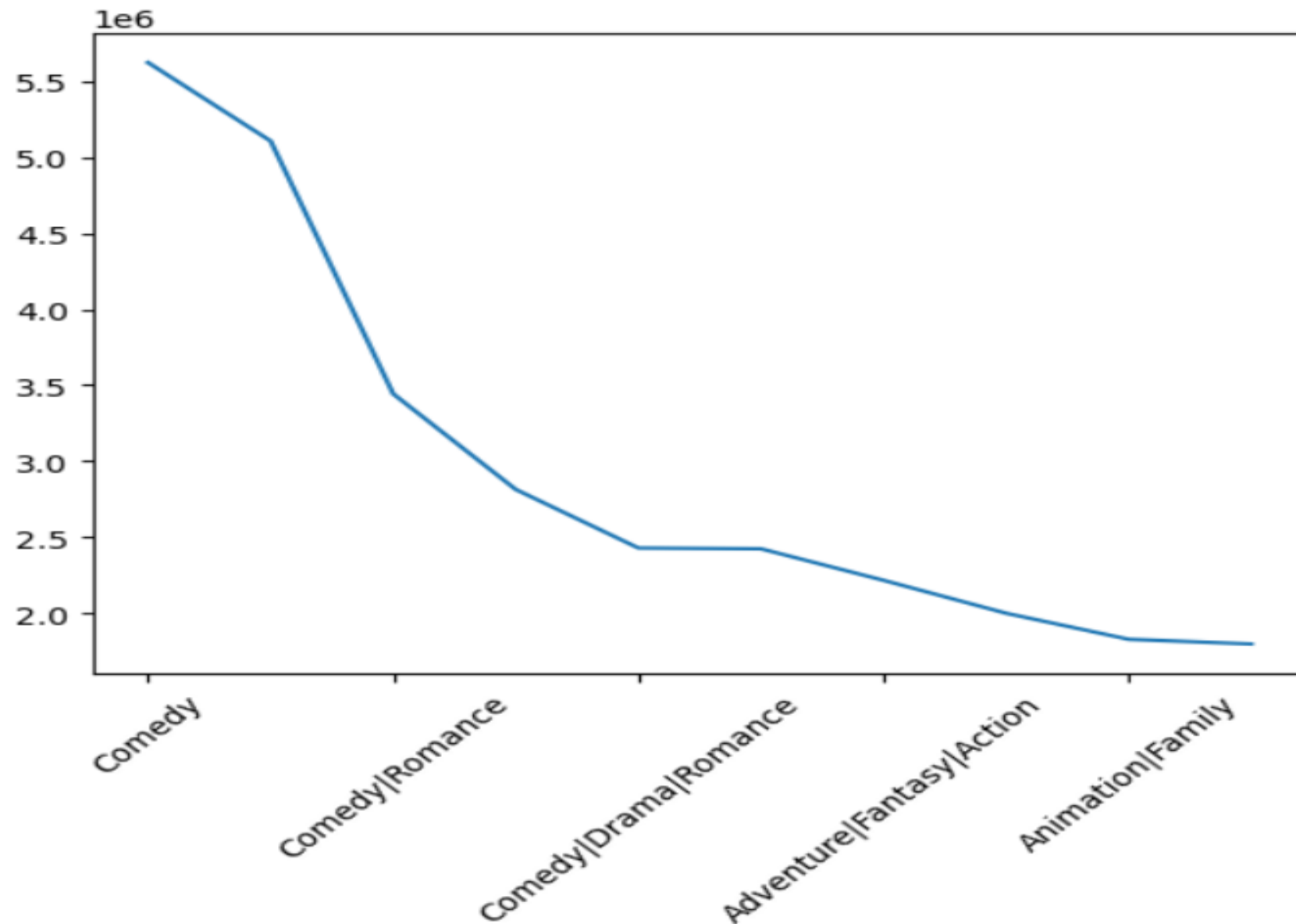
	original_title	profit
7050	Win a Date with Tad Hamilton!	-0.12%

IS THERE A RELATIONSHIP BETWEEN BUDGET_ADJ&REVENUE_ADJ

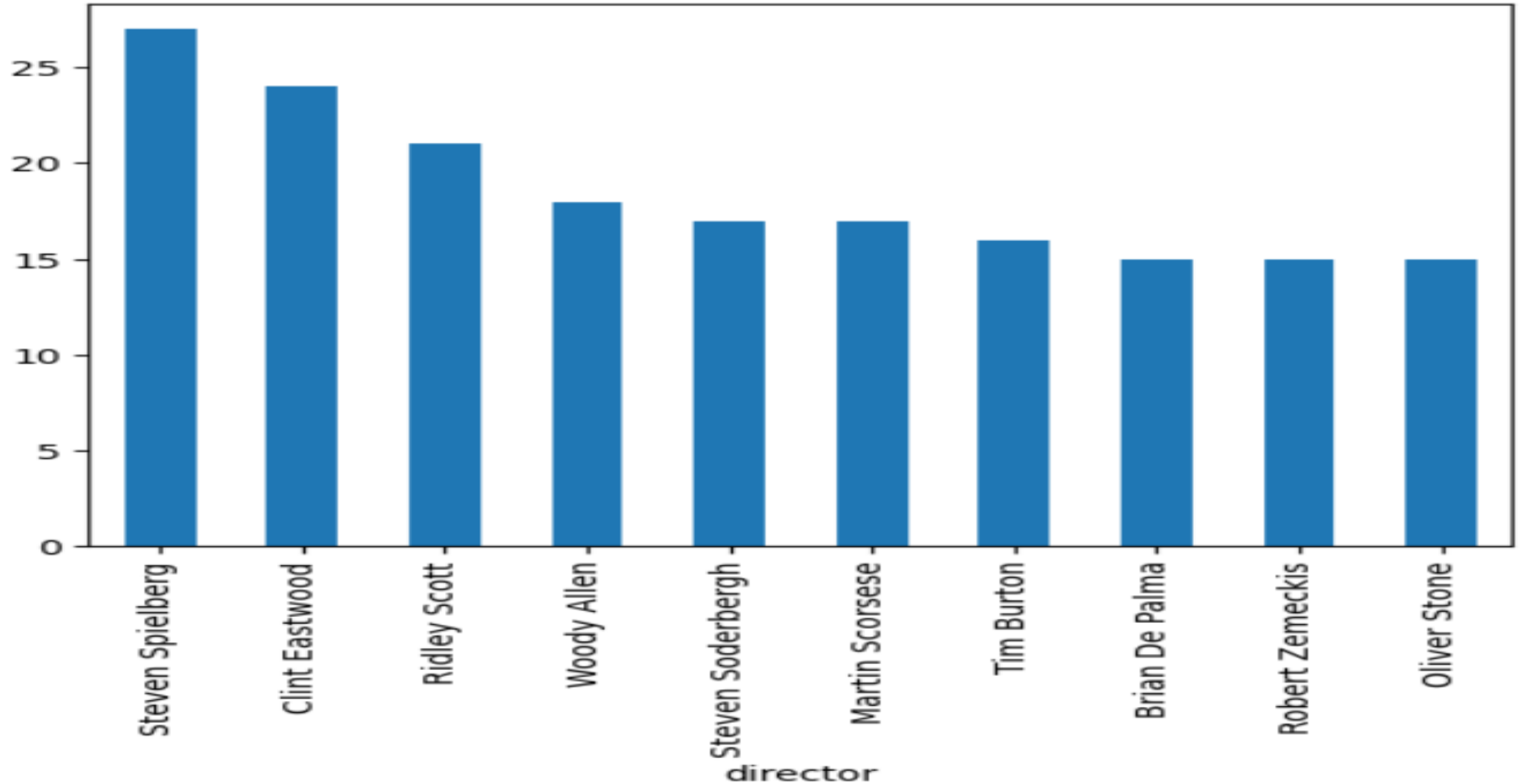


THE MOST GENRES REVENUE_AJ

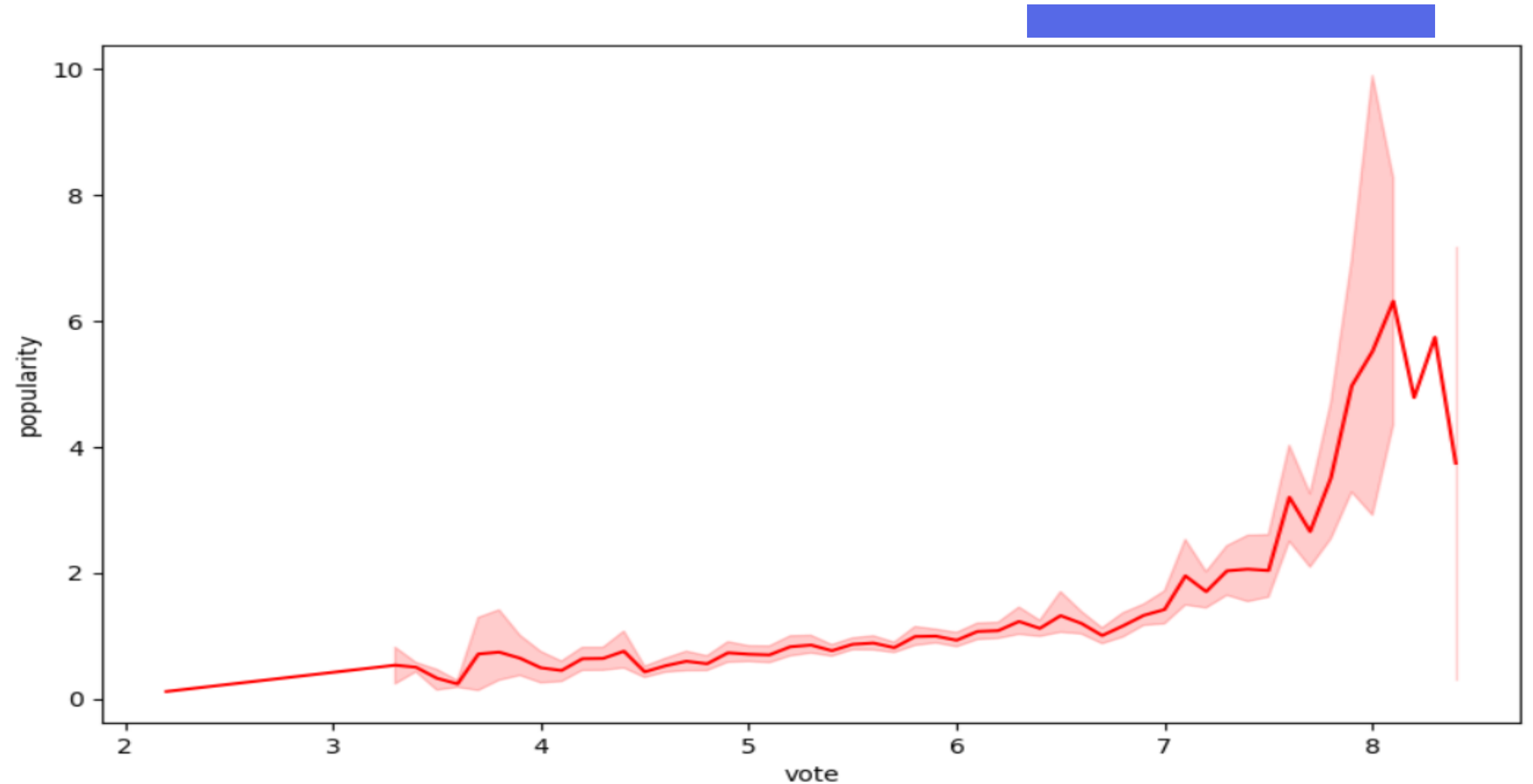
6]: <Axes: >

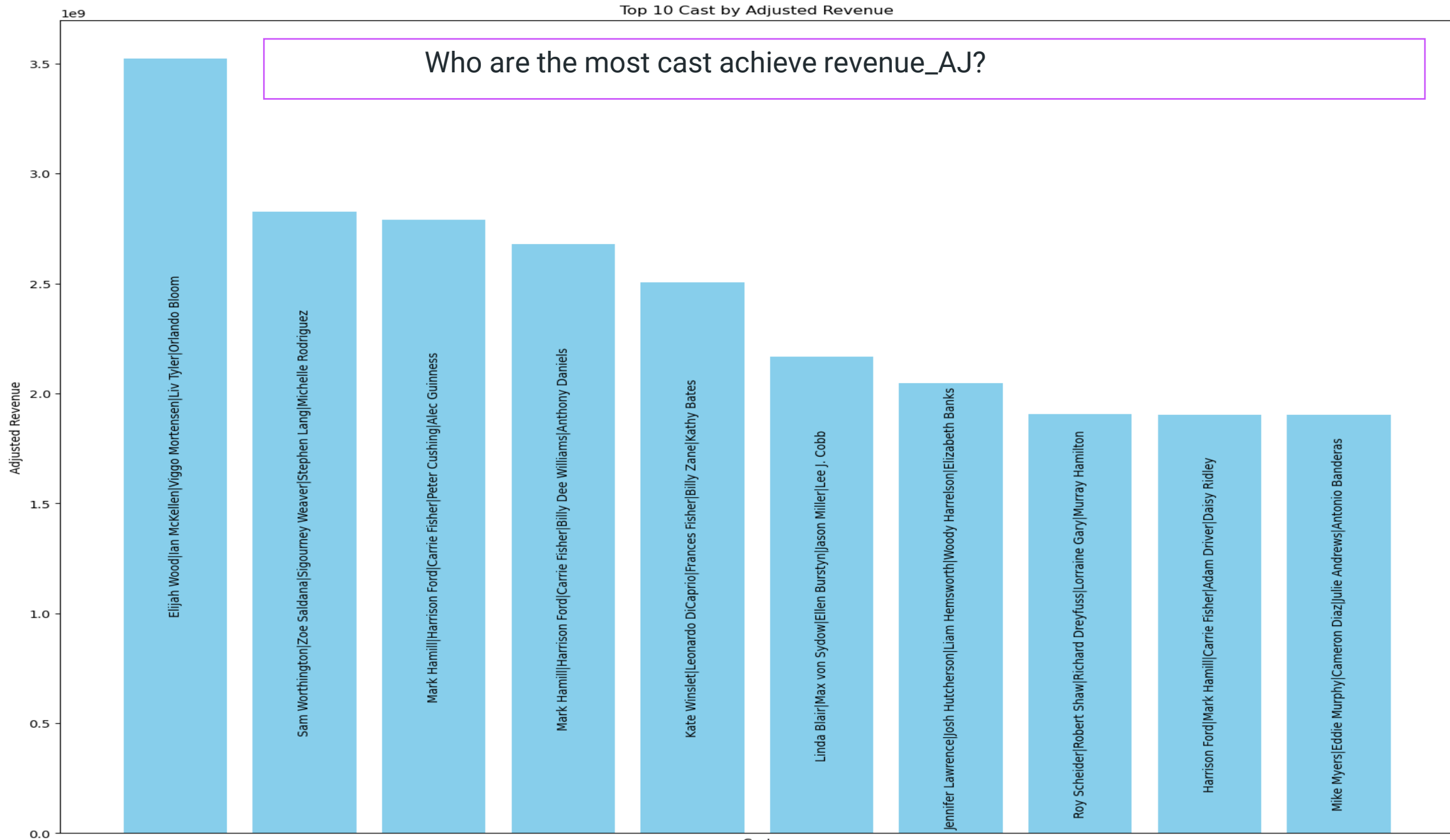


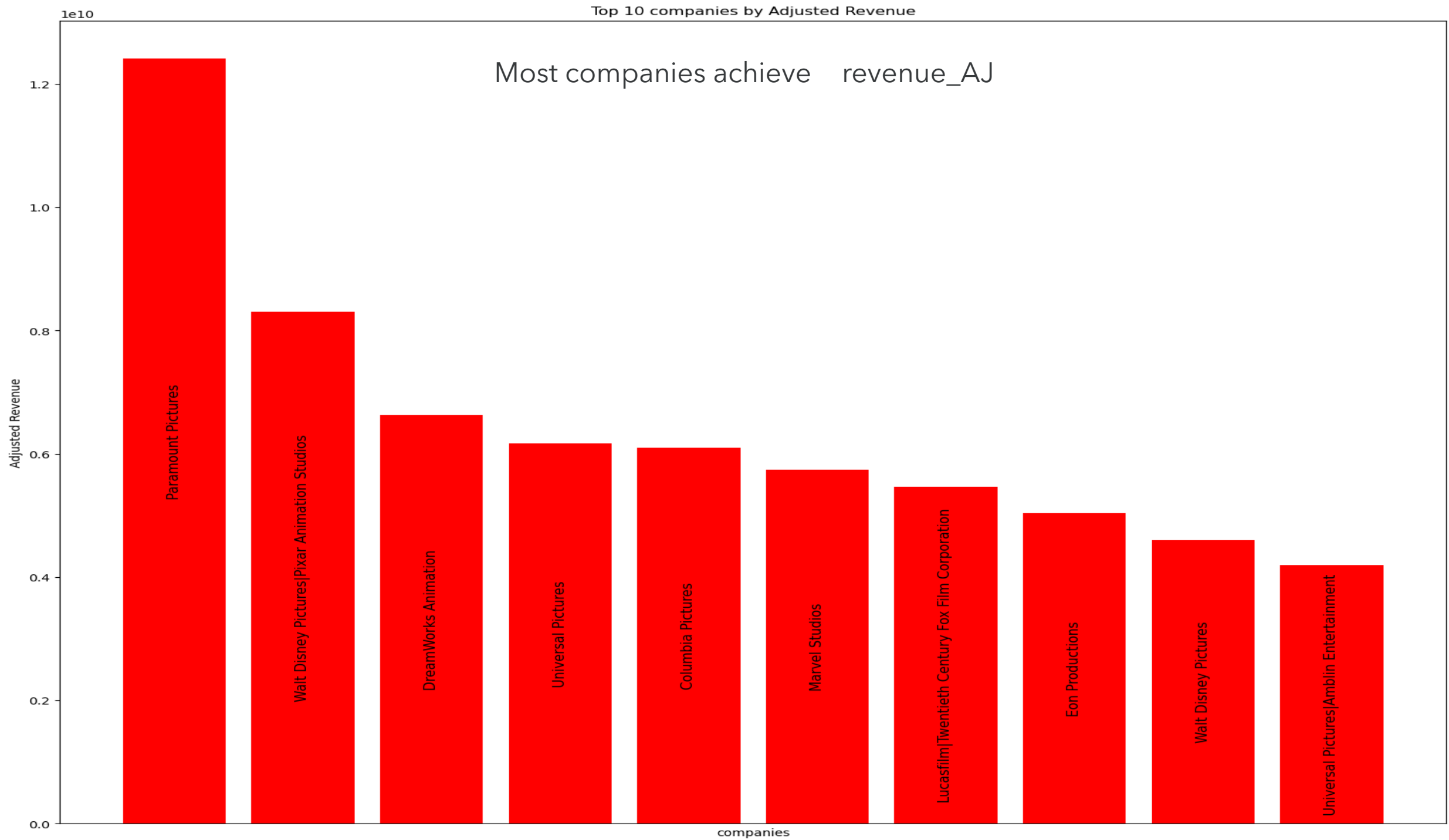
WHICH DIRECTORS HAVE PRODUCED THE MOST SUCCESSFUL FILMS?



IS THE MOST VOTED MORE POPULAR?









THANK YOU

Muhammad Ahmed