



PROJECT DESCRIPTION

• This project aims to investigate the factors influencing the success of movies on IMDB, defined by high ratings. Through data cleaning and analysis, we'll explore correlations between ratings and variables like genre, director, budget, etc. Utilizing the Five 'Whys' approach, we'll delve deeper to uncover root causes. Our report will present findings and insights, aiding stakeholders such as producers, directors, and investors in making informed decisions for future projects.

Approach

- 1. Data Cleaning: Preprocessed dataset by handling missing values, duplicates, and converting data types.
- 2. Data Analysis: Explored correlations between ratings and variables using statistical techniques and visualizations in Python with libraries like Pandas, Matplotlib, and Seaborn.
- 3. Five 'Whys' Approach: Employed to dig deeper into factors influencing movie success.
- 4. Report & Insights: Presented findings, insights, and actionable recommendations using storytelling techniques and visualizations to aid stakeholders.

Tech stack used

I used Web Microsoft 365 excel as Macbook don't have excel in built





INSIGHT

- . 1. Budget & Ratings: Higher-budget movies tend to have higher ratings, possibly due to better production quality.
- 2. Genre Impact: Certain genres like drama and thriller consistently attract higher ratings than others.
- 3. Director Influence: Movies directed by renowned directors often receive higher ratings, highlighting the importance of directorial vision.
- . 4. Actor Effect: Involvement of popular actors positively impacts ratings, indicating the influence of star power.
- 5. Release Year: Recent movies generally receive higher ratings, suggesting evolving audience preferences.
- 6. Audience Engagement: Viewer experience and enjoyment significantly influence ratings, emphasizing the importance of storytelling and engagement.
- 7. Positive Reviews: Positive viewer reviews play a crucial role in a movie's success, influencing others' decisions to watch it.
- These insights can guide decision-making for stakeholders, helping them understand factors contributing to movie success on IMDB.

RESUL

130

CVS 7

.cs

250 .07

80.

ट्युर वुद्य

151.69

95%

.00

569.01

.00

199

101

10.

8°.94

82.07

088.16

18973

51.90

565

92.908

85.08

10.02

150.34

15.599

2.37

3.

38.131

17.653

2.544

6.

12.101

91,501

N2.050

7.321

21

0.088

21.661

19.250

34.609

63.81

69.593

73.601

83.155

1.10

30.932

201

23.21

96.924

0.709

36.561

72.098

369

13.65

B

28.801

12.518

B.

38.161

19.64

VO. 140

.00

601

7.7°

0.

726

03.00

970

20.00

.00

10.01

53.020

23.765

54.322

13.16

VO.38

16.606

3.

57.322

24.384

91.7A

989

43.43

32.121

81.67

00

809.16

.00

0.

55.72

111.99

8

000

80

126

5

10.

52.025

V.205

ONO

80.

23.765

96.921

69.693

10.102

17.653

16.565

32.00

ණ.

54.322

73.76

6.

20

N. 189

372

53.026

94.88

75.278

MA

517

2.

63.032

19.518

26.039

15.11

29,131

23.323

72.036

57.365

14.688

6.818

22.920

3.

201

18.98

16.198

32.029

38.713

20.282

58° OS

896

11009

986.94

166

016

108.2V

05.70

ME.08

026.46

.00

c16.V1

2007

दुल्ली कुड़ा

·cò

076.71

000,00

1689

851.1Y

ST

15.

187.48

964

.60

628

200 570,753

MAY

10,01

168,06

000

8

30

·60.

186

8

V

500

90

100

100

750

325.01

.80

0976

060 069.00

192.68

128.16

500.65

20

726.60

CO. 24

8.

CSTY CO

81.69

VIV

500.00

248 60° 17

O ST

00.

970

250

.20

MY

5 0000

912

.go

250 20.05

16.51

91.886

75.278

52.025

V.205

94.833

200

do.

5.161

2.

13.994

9.033

9A.

12.

88.75

80.

84.07

821

06.049

95.052

30.

5/3

OV.

30.

SIA

21.96

D.

6.

69.932 13

493

30.

2.

9.012

2

92

VV.

2.

III

51.734

30

31.

97.397

31.780

30.534

39.889

1.171

68.971

25.500

12.926

13.291

N. 003

24.589

90,736

82.623

8.

18.988

5508

65.50

66.281

293

16.911

1.18

16.

63.129

92.79

8

111.89 910

8

506.67

1200

280.86

167.61

026.27

40°.24

10

896.00

210.11

0000

W.14

688

Who

80

076.46

165.2h

465.46

064

15.00 of

CYC, 21

134

8

096

.60

1100 305

550

.0,

10.61

10.

10

ON

.80

000 A6.9A

205,4

10.01

50.763

8/2

10.

960

100

. Su

602

CV.

.cs

120

84.18

1.283

513

86.11

89.789

31.159

5.92.

31.12

10.

00.

140

8%.

103

2.

80.

3.

93.668

63.575

51.987

21.224

51.65

3.

55.082

67.307

14.661

76.076

81.375

10.021

152.098

278

92.199

N.018

39.695

98.913

8.

10.532

23.662

VO.130

050.235

85.05

5A.0AA

15.981

088.46

01/2

250

00/

126

.co

20

6.

80.00

178.00 300

64015

2.00

2005 200

THES.

200

204

.28

100 .80

.

CST

92

69.306

36.01

3.

56.878

47.330

29.913

72.691

01.255

82.89A

N2.903

35.31

23.164

13.003

29.608

3A.108

2.

26.354

23.72

84.859

202

8.00

39.74

83.104

77.016

80.

13.9h

14.000

99.883

72.732

82.0A

10.00

16.693

8.202

59.182

23

100

20.00

.co

25.78

864

18.973

63.84

69.593

13.

10.102

59.34

294

17.653

16.198

32.029

38.13

0.

23.323

12.030

57.345

14.888

6.818

7.805

No.020

67.217

2.291

00.

0602

13.013

032

18. 29. NA

20.039

672

81.91

55.0A

97.891

81.382

34.918

31.251

28.119

966

94.883

62.317

3.

32.902

10.0%

39.16

1.579

30

1.550

V.

5

6.

15.11

20.13

32.11

28.00

16.898

30.18

16.64

18.152

86.560

160

W. 818

37.16

77.355

31.27

3

1.

13

29.03

81.617

3

13.97

20

18.9Th

3.55

8.

372

6.129

6.

W. 124

15.158

15.293

22.662

95.757

66.50

8.579

2.

15.914

232

17.27

15)

2.

23.264

13.088

510

W1.05V

821

0.

30.

64.555

83.679

133

10.814

92

84.

V1.8V8

84.55G

18.03

10.316

17.062

250

Vy. 180

6.

M

64. V2

29,131

2.

20.282

12.101

91:01

7.321

0.088

21.664

18ª

16.293

10.

22.21

65.634

V2.23

19.29

34.609

989

30.

3

10.

SAL

10.555

92.908

85.08

2010

83.155

170

100

81.201

B.

45.00

30.932

96.924

0.

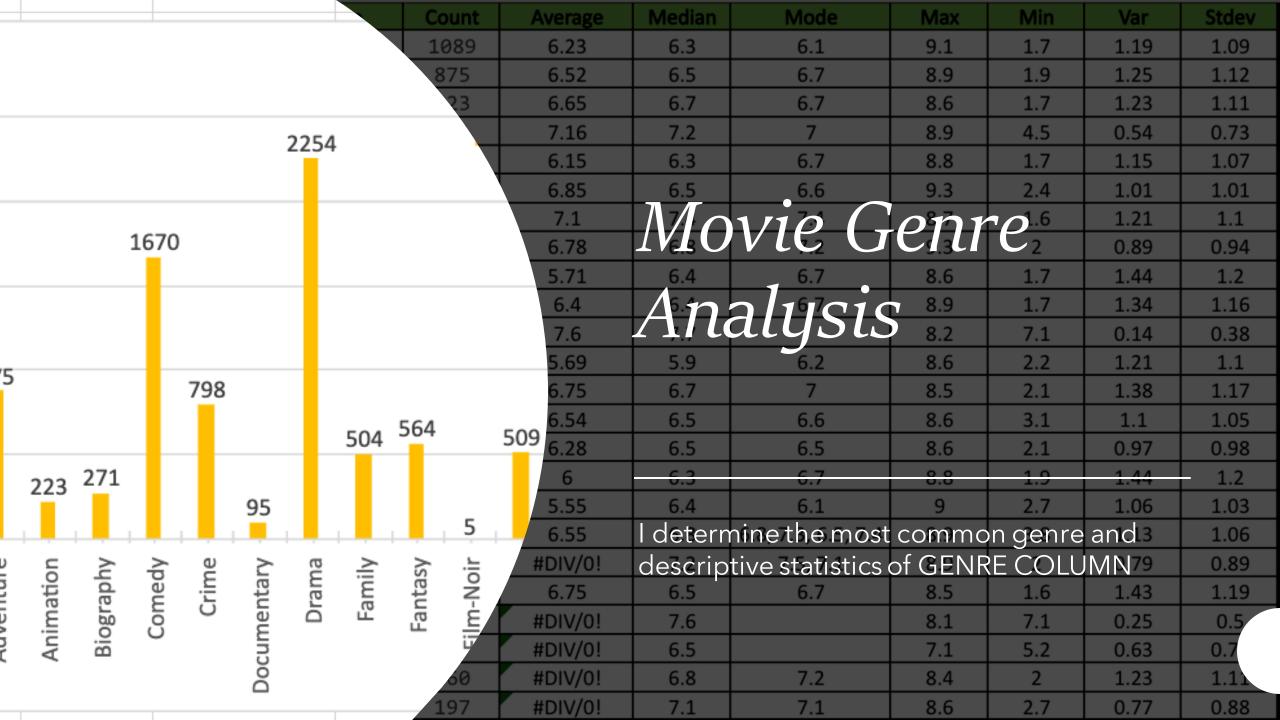
121

AA.AT

331

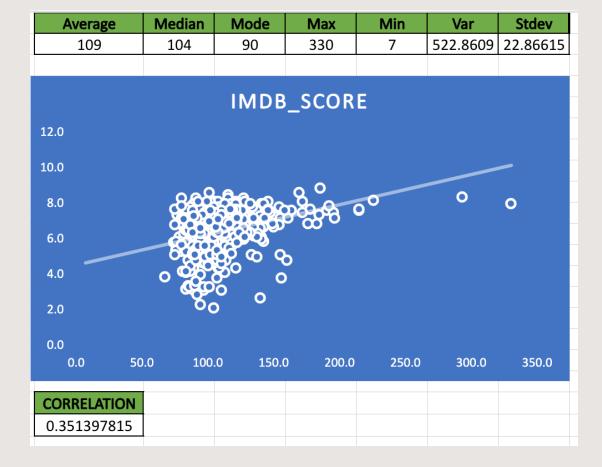
cyl

.0



Movie Duration Analysis

Did duration descriptive analysis and try to find the correlation between duration and IMDB score - which seems like moderate positive correlation.

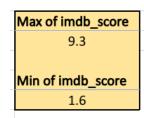


LANGUAGE ANALYSIS

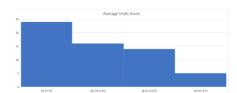
• ENGLISH is the most common language used in IMDB movies , we also perform the descriptive statistics

,			J			,
inguages	Counts	Average	Median	Max	Min	Var
lish	4250	6.38	6.5	9.3	1.6	1.23
nch	54	7.14	7.2	8.4	5.1	0.43
nish	35	6.91	7	8.2	4.4	0.76
nese	15	7.31	7.5	8.7	5.6	1.01
sian	9	6.01	5.4	8.1	4.1	1.44
tch	4	7.43	7.45	7.8	7	0.14
bic	4	7.38	7.65	8.2	6	0.78
egian	4	7.15	7.3	7.6	6.4	
guese	6	7.72	7.95	8.7	6.1	0.65
darin	20	7.03	7.1	7.9	5.6	0.45
nian	1	4.3	4.3	4.3	4.3	0
rew	3	7.5	7.3	8	7.2	0.13
sian	3	8.13	8.4	8.5	7.5	0.2
ndic	1	6.9	6.9	6.9	6.9	0
ian	9	7.06	7	8.9	5.1	1.45
nian	1	6.4	6.4	6.4	6.4	0
ean	6	7.62	7.5	8.4	7	0.25
ai	3	6.63	6.6	7.1	6.2	0.14
onese	9	7.02	7.3	7.8	5.3	0.52
anian	2	7.2	7.2	7.9	6.5	0.49
man	15	7.59	7.7	8.5	6.1	0.47
esian	2	7.9	7.9	8.2	7.6	0.09
lu	2	7.1	7.1	7.3	6.9	0.04
imese	1	7.4	7.4	7.4	7.4	0
dish	3	7.47	7.6	8.2	6.6	0.44
ne	2	7.95	7.95	8.5	7.4	0.3
ahili	1	7.4	7.4	7.4	7.4	0
ech	1	7.4	7.4	7.4	7.4	0





UNIQUE DIRECTOR	Average_SCORE	BENCHMARK
Sadyk Sher-Niyaz	8.7	8
Cary Bell	8.7	8
Tony Kaye	8.6	8
Mike Mayhall	8.6	8
Charles Chaplin	8.6	8
Ron Fricke	8.5	8
Raja Menon	8.5	8
Majid Majidi	8.5	8
Damien Chazelle	8.5	8
Sergio Leone	8.475	8



Director Analysis

 This task involves identifying top directors based on their average IMDB scores and assessing their influence on movie success.
 We calculate the average IMDB score for each director and use Excel's PERCENTILE function to identify directors with the highest scores.
 By comparing these scores to the overall distribution, we evaluate the contribution of top directors to movie ratings, providing insights into their influence on audience perceptions and movie success.

budget		Correlation of bugdet and gross_collection		0.11					
	\$237,000,000	\$76							
	\$150,000,000	\$652,1	Maximum profit	\$523,505,847					
	\$200,000,000	\$658,672							
	\$11,000,000	\$460,935,6	Budget A	nalugie					
	\$10,500,000	\$434,949,45	Dauget	maigsis					
	\$220,000,000	\$623,279,547							
	\$45,000,000	\$422,783,777	gross_collection	1					
	\$115,000,000	\$474,544,677							
	\$185,000,000	\$533,316,061							
	\$78,000,000	\$407,999,255	 This task involves exploring the relationship between movie 						
	\$58,000,000	\$363,024,263	budgets and their financial success. We analyze the correlation between movie budgets and gross earnings using Excel's CORREL function to understand how budgets						
	\$130,000,000	\$424,645,577							
	\$63,000,000	\$356,784,000	impact financial outcomes. Additionally, we calculate the						
	\$76,000,000	\$368,049,635		profit margin for each movie (gross earnings - budget) and identify movies with the highest profit margins using Excel's MAX function. By doing so, we gain insights into the financial performance of movies relative to their budgets, aiding					
	\$58,800,000	\$350,123,553							
	\$94,000,000	\$380,838,870	performance of movies relativ						
	\$150,000,000	\$436,471,036	decision-making for investors	and producers.					
	\$94,000,000	\$377,019,25		4222.222.222					
	\$32,500,000	\$309,125	\$100,000,000 \$200,000	,000 \$300,000,000					
	\$55,000,000	\$329,69	 gross_collection Linear 	(gross_collection)					
	\$18,000,000	\$290	0						



Drive link

• <u>I am the excel file click me</u>

Thank you Trainity for the opportunity

