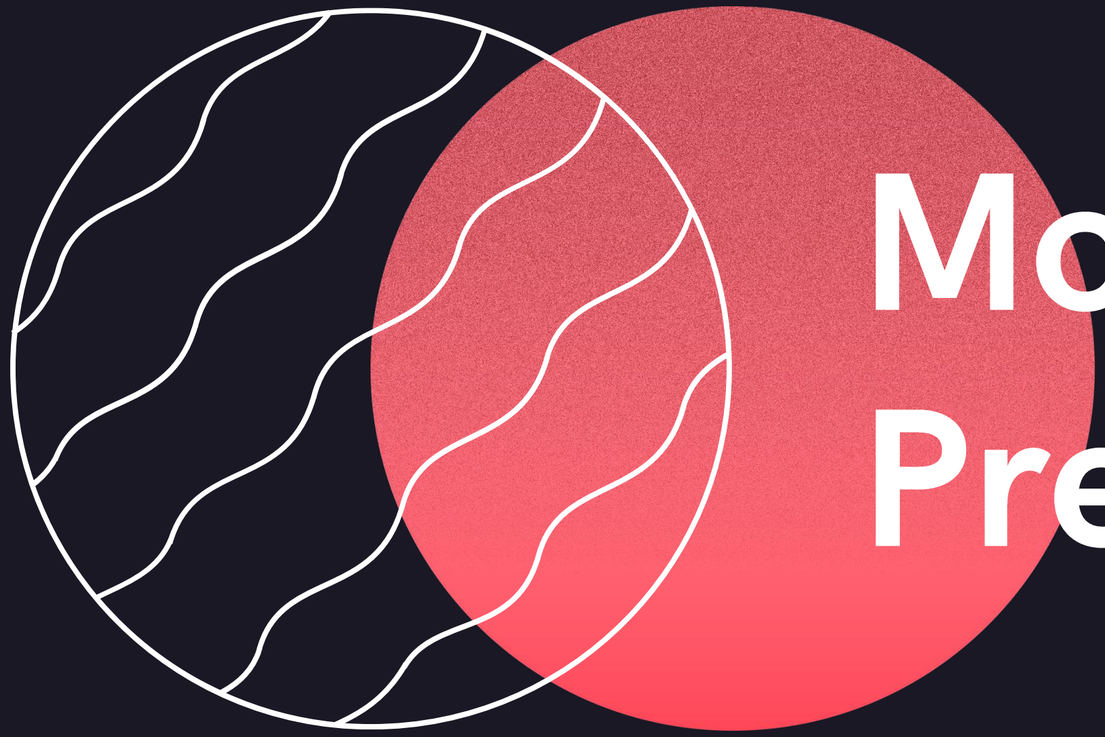


Movies Revenue Prediction

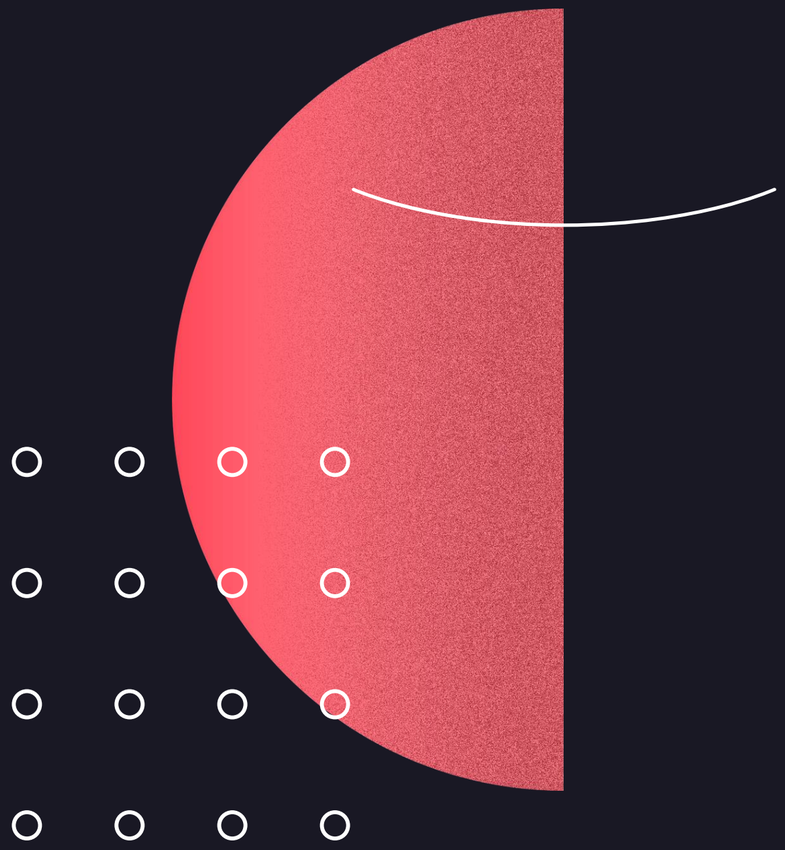
SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

Ali Aljamid

What is This Project Trying to Solve?

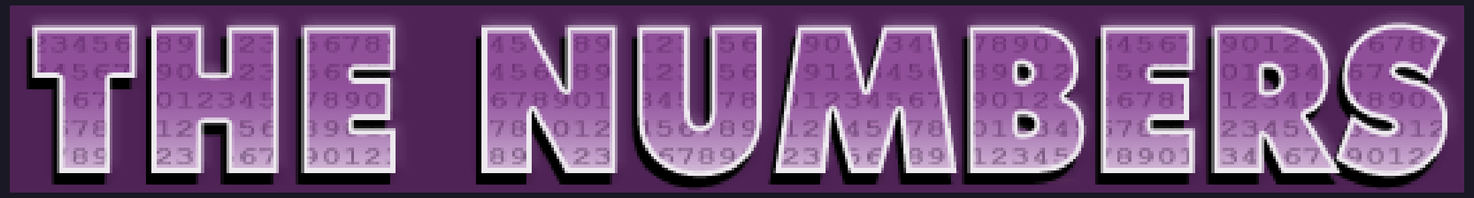
- Predict the revenue of the selected movies

What Does Movie Revenue Actually Depend on?



- Actor/Actress
- Script
- Budget

CAN WE USE MOVE BUDGET TO PREDICT MOVIE REVENUE?



DATA WAS COLLECTED FROM THE-NUMBERS.COM

DATA SET (DIRTY)

Six Features

- Rank
- Release Date
- Movie Title
- Production Budget ($)
- Worldwide Gross ($)
- Domestic Gross ($)

Movies Data

```
df
```

	Rank	Release Date	Movie Title	Production Budget ($)	Worldwide Gross ($)	Domestic Gross ($)
0	5293	8/2/1915	The Birth of a Nation	$110,000	$11,000,000	$10,000,000
1	5140	5/9/1916	Intolerance	$385,907	$0	$0
2	5230	12/24/1916	20,000 Leagues Under the Sea	$200,000	$8,000,000	$8,000,000
3	5299	9/17/1920	Over the Hill to the Poorhouse	$100,000	$3,000,000	$3,000,000
4	5222	1/1/1925	The Big Parade	$245,000	$22,000,000	$11,000,000
...	...	...	...	...	...	...
5386	2950	10/8/2018	Meg	$15,000,000	$0	$0
5387	126	12/18/2018	Aquaman	$160,000,000	$0	$0
5388	96	12/31/2020	Singularity	$175,000,000	$0	$0
5389	1119	12/31/2020	Hannibal the Conqueror	$50,000,000	$0	$0
5390	2517	12/31/2020	Story of Bonnie and Clyde, The	$20,000,000	$0	$0

5391 rows × 6 columns

```
df.shape
```

```
(5391, 6)
```

DATA SET (CLEAN)

Two Features

- Production Budget ($)
- Worldwide Gross ($)

	prediction_budget_usd	worldwide_gross_usd
0	110000.0	11000000.0
2	200000.0	8000000.0
3	100000.0	3000000.0
4	245000.0	22000000.0
5	3900000.0	9000000.0
...	...	...
5378	55000000.0	376856949.0
5379	40000000.0	166893990.0
5380	185000000.0	561137727.0
5381	175000000.0	140012608.0
5382	42000000.0	57850343.0

5034 rows × 2 columns

DATA SET (CLEAN)

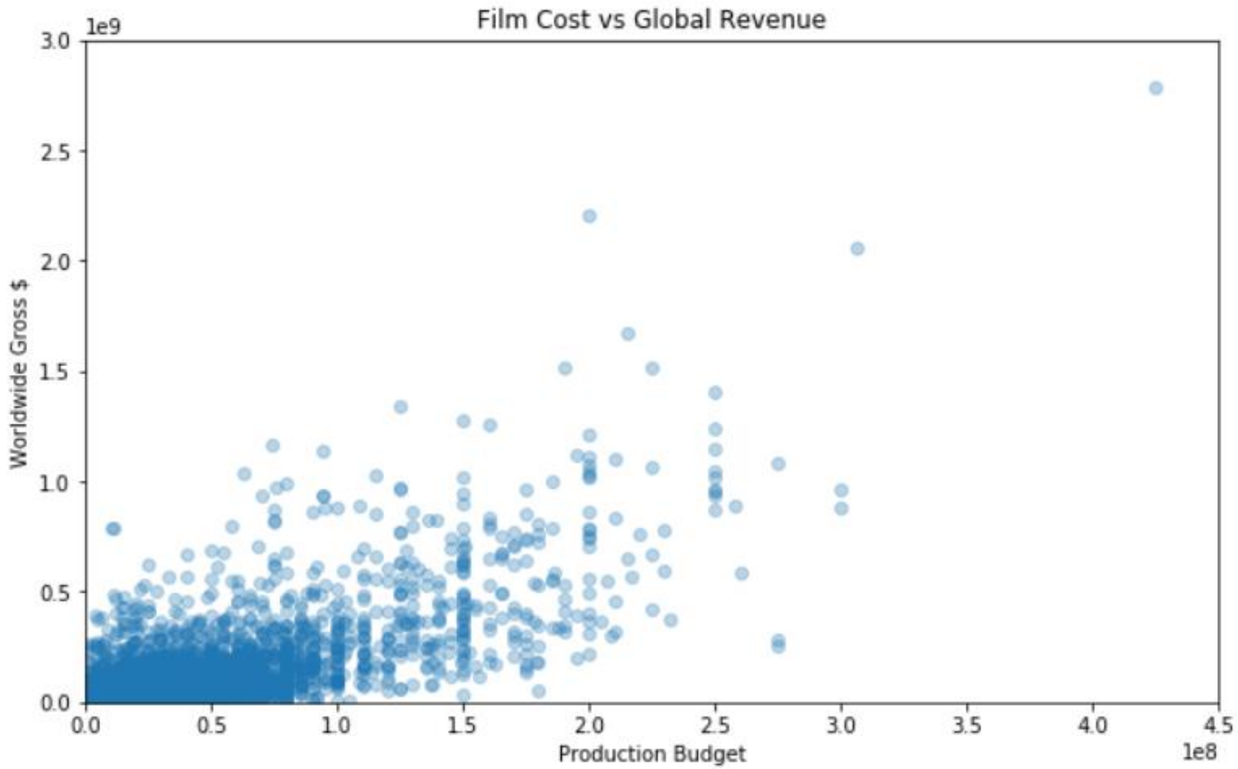
Data Overview
Using "describe()" function


```
df.describe()
```

	prediction_budget_usd	worldwide_gross_usd
count	5.034000e+03	5.034000e+03
mean	3.290784e+07	9.515685e+07
std	4.112589e+07	1.726012e+08
min	1.100000e+03	2.600000e+01
25%	6.000000e+06	7.000000e+06
50%	1.900000e+07	3.296202e+07
75%	4.200000e+07	1.034471e+08
max	4.250000e+08	2.783919e+09

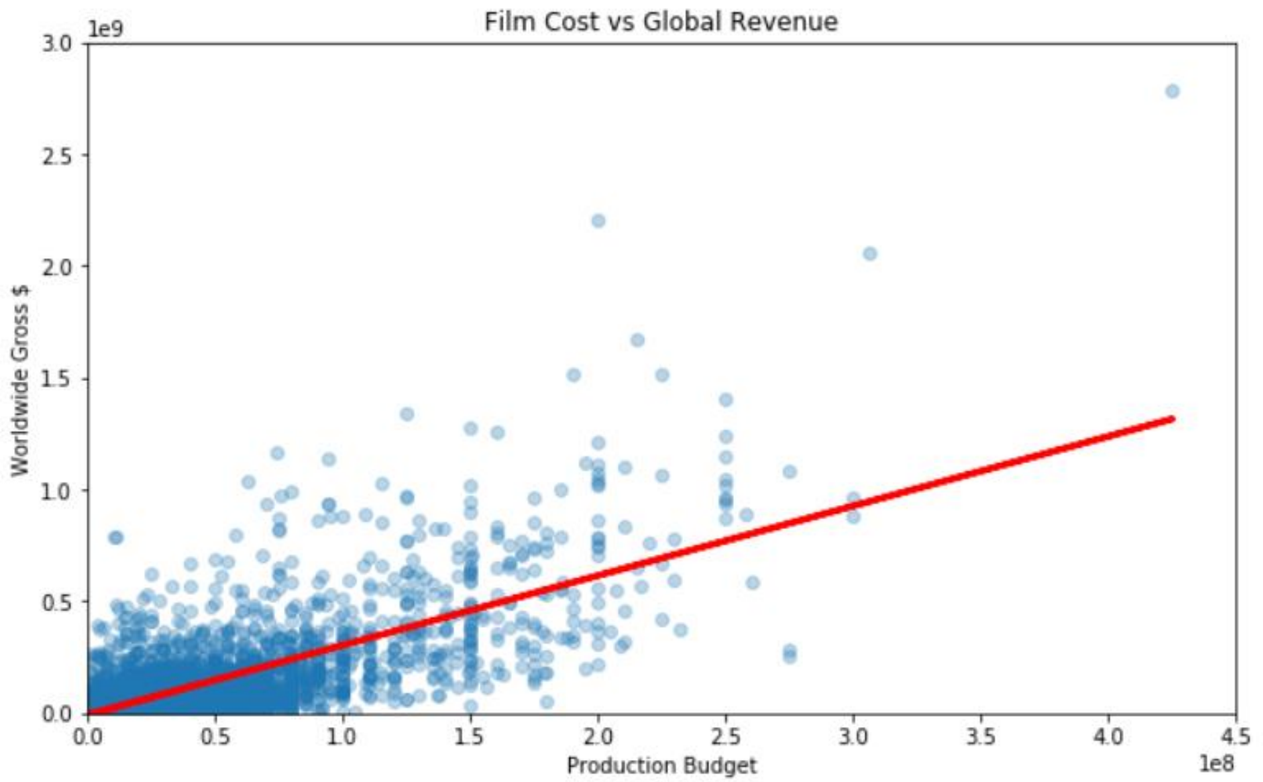
DATA SET (CLEAN)

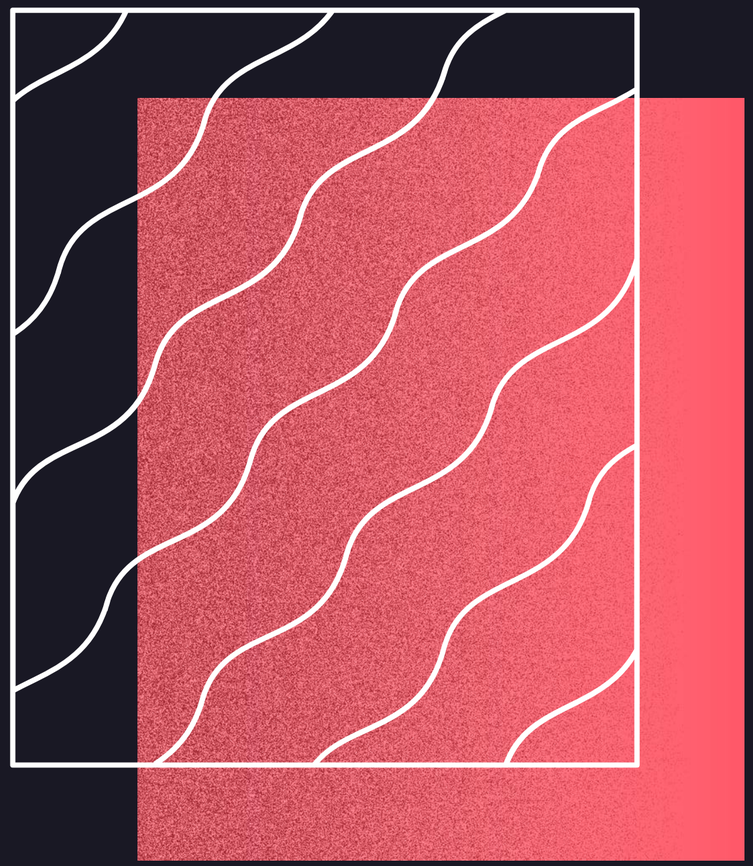
Data Visualization
Using Matplotlib library



DATA SET (CLEAN)

Data Visualization
Using Matplotlib library and sklearn to drow a plot





EVALUATION AND ANALYSIS

- Slope coefficient is: 3.11150918 which is a positive relationship
- Each dollar spent on producing a movie, there is ~ $3.1 revenue
- Intercept : -7236192 which means a movie with a 0 budget is losing ~ 7 million (not realistic)
- Prediction formula = (3.111 x budget) - 7,236,192 (148,338,807 = 3.111 x 50,000,000 -7,236,192)

- The overall score is: ~ 55%