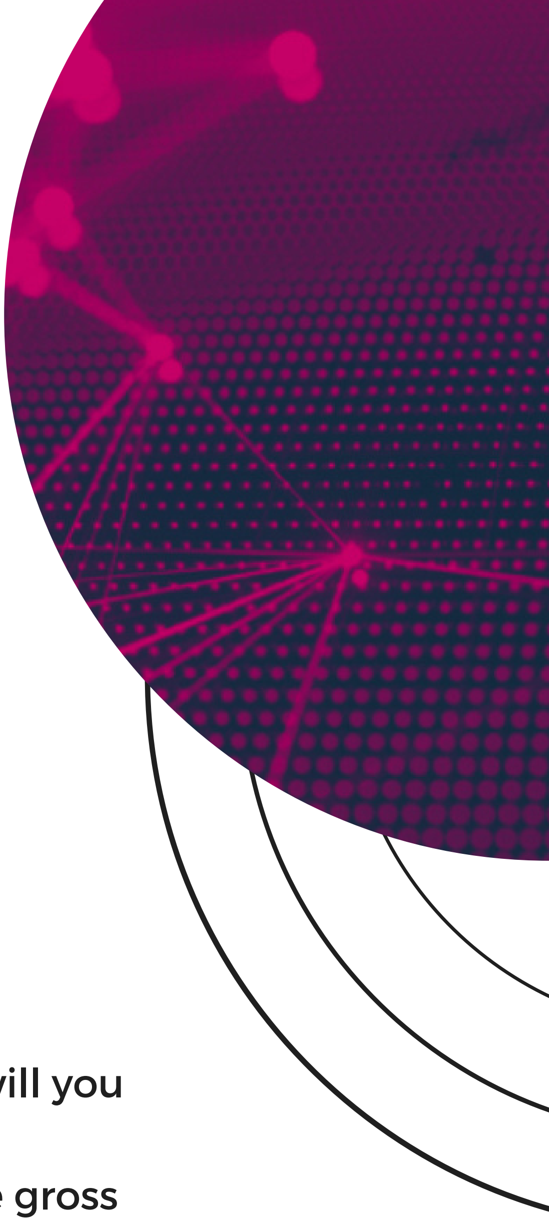
OCTOBER, 2021

MOVIES REVENUE PREDICTION PROPOSAL

ORGANISED BY:
Ali Aljamid

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

PROJECT PROPOSAL



QUESTION/NEED:

What is the framing question of your analysis, or the purpose of the model/system you plan to build?

- To predict movie box office revenue with Linear Regression

DATA DESCRIPTION:

What dataset(s) do you plan to use, and how will you obtain the data?

- Movies title, production budget, worldwide gross
- It was obtained from (www.the-numbers.com)

What is an individual sample/unit of analysis in this project?

- Movies budget versus the worldwide gross

What characteristics/features do you expect to work with?

- Movie box office revenue

TOOLS:

01 How do you intend to meet the tools requirement of the project?

- Clean the data by Pandas
- Leverage scikit-learn to apply Linear Regression

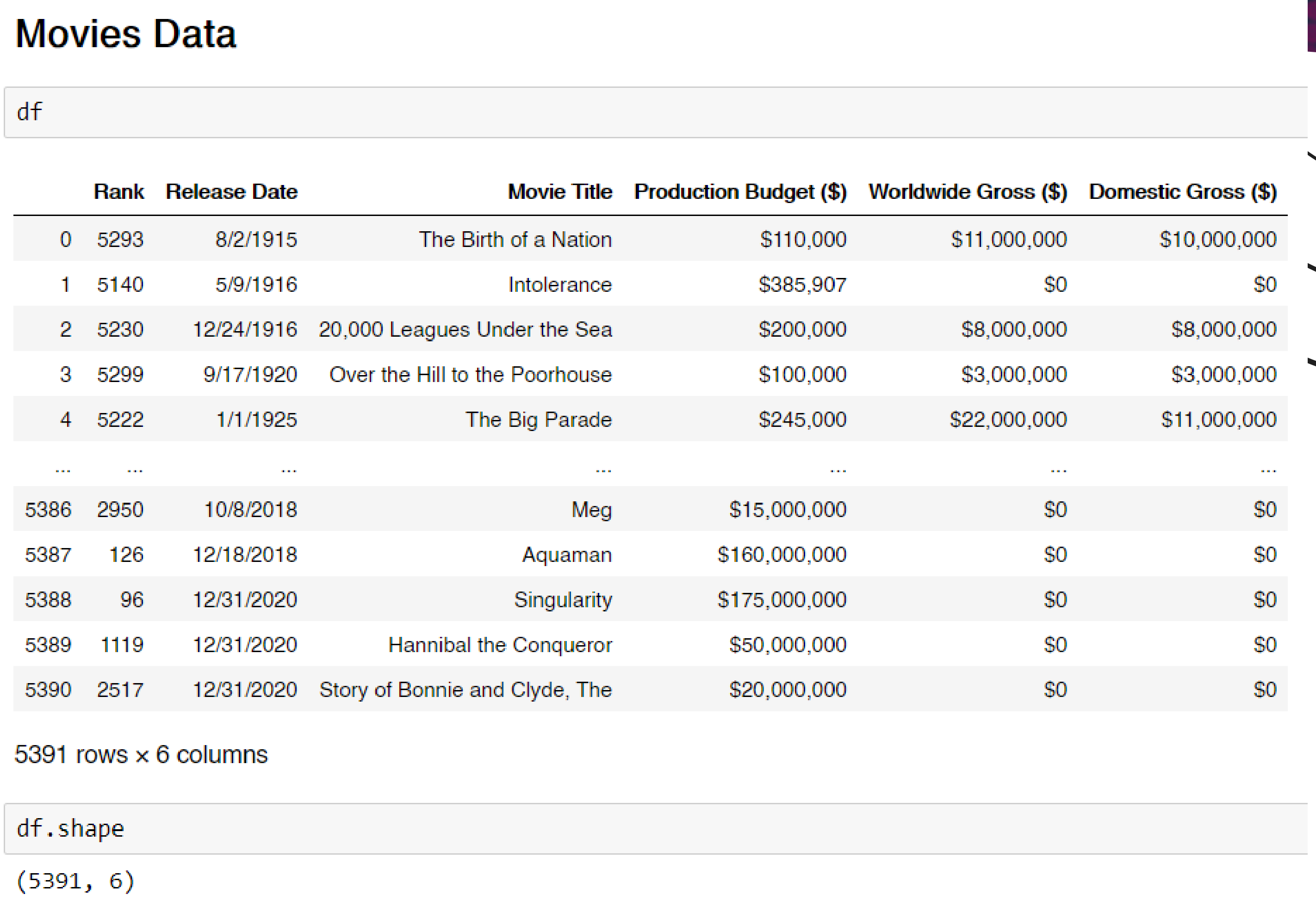
02 Are you planning in advance to need or use additional tools beyond those required?

- No

PROJECT PROPOSAL

EXPLORATORY DATA ANALYSIS

To understand our data further and to put the initial hypothesis



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)
```

- The dataset contains six features (Rank, Release Date, Movie Title, Production Budget ($), Worldwide Gross ($), Domestic Gross ($))
- A total of 5391 data points
- Zero values for Worldwide Gross ($), Domestic Gross ($)