



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

T5 Bootcamp

Data scientists

Rawan Al-Qahtani - ru2468@hotmail.com

Instructor : Mejdal Al-Qahtani

Project Proposal

Movie Dataset

- **Question/Problem statement :**

What's the prediction 5 top rating movies from year 1920 to 2021?

- **Data Description :**

Main Data source is IMDb <https://www.imdb.com/> 'Top 1000' (Internet movie database)

- **Dataset Size:** 1,000 rows and 8 features contains (Name, Production Year, Movie Type, Rate the Site, Website visitors Vote, Duration, Us Gross, Meta Score.
- **Data Type:** Integer, Date, String.

- **Tools :**

- **Programs :**

We will use Python ,(HTML,CSS) optional.

- **Libraries :**

Pandas, matplotlib, seaborn, numpy, datetime, BeautifulSoup, requests.

- **Functions :**

We will use several functions that we define in the project.

- **Plots :**

We will use one of them (lineplot, Scatter, bar, hits, pie, circle).

- **MVP Goal:**

The goal of this project is to prediction the 5 most rating movies from year 1920 to 2021.