IMDB - Analyzing Movies

A project that goes over the **movies** produced from **1960 to 2015**

Data provided by **Analyst Builder**. Insights and Analytics by **Divesh**

The Data We're Working With

- Movies produced between 1960 to 2015 - along with corresponding data such as Directors, Cast, Budget, IMDB Votes and many more.
- The data has both the budget used to produce the film as well as Box Office (revenue) collections allowing us to analyze success of the movie.
- IMBD Votes and the scores allow us to understand how the fans perceived the film and if a lower score impacts performance.

What We're Going To Do

- Clean up the data and check for errors - we will remove NULL values that have no significance to the dataset.
- Answer a series of questions based on the information extracted from our dataset and provide an explanation to how we reached our conclusion.
- Go over errors that could not be corrected within the dataset, any issues that occurred and future improvements that we can work on for incoming datasets.

Question 1

List the Top 3 Movies that were produced within a Low Budget

High Profit Earners with a Low Budget

1. **My Big Fat Greek Wedding**
   BUDGET: **USD 5,000,000** | BOX OFFICE: **USD 368,744,044**

2. **Crocodile Dundee**
   BUDGET: **USD 5,000,000** | BOX OFFICE: **USD 328,203,506**

3. **The Full Monty**
   BUDGET: **USD 3,500,000** | BOX OFFICE: **USD 257,850,122**

Question 2

Which Directors were involved in the Top 3 Highest Grossing Films

Box Office Record Breakers

1. **James Cameron**
   AVATAR (2009) | BOX OFFICE: **USD 2.78 BILLION**

2. **J.J. Abrams**
   STAR WARS (2015) | BOX OFFICE: **USD 2.07 BILLION**

3. **James Cameron**
   TITANIC (1997) | BOX OFFICE: **USD 1.85 BILLION**

Question 3

The Top 3 Actors with the most IMDB Votes on all their movies combined

“And the Oscar Goes To…”

1. **Leonardo DiCaprio**
   VOTES: **48,810** | IMDB AVERAGE: **6.73**

2. **Samuel L. Jackson**
   VOTES: **45,097** | IMDB AVERAGE: **6.09**

3. **Robert Downey Jr.**
   VOTES: **44,959** | IMDB AVERAGE: **6.21**

Issues & Future Improvements

- **Missing Data:** Most movies did not possess values such as budget needed to produce the movie.
- **Data Format:** All of the information such as cast members and genres were all stored in a single row. Separating all created numerous errors and multiple duplicates to be formed.
- **Improvements:** Reworking on the code to better clean the data in order to analyze it better.
- **Future Ideas:** With more recent data, we can create a visual dashboard allowing for people to see items such as high rated movies, casts, genres and many more. Adding items such as images, visual effects and a user interface would make the dashboard interactive and fun to use.

Thank you for going over my project!

GitHub

Tools Used: Python, SQL, Google Slides, Jupyter Notebooks