Data Analytics Challenge
In Partnership with DataSense Analytics

Created By	John Dale Vacaro
Dashboard Link	https://github.com/Vacaro23/IMDb-Movies/blob/main/DICT%20Dashboard%20Output%20-%20IMDb%20Movies%20-%20John%20Vacaro%20--%20202407.pdf
Task	Participants will analyze a provided dataset to transform and shape data, model the data, uncover insights, and visualize findings. The challenge aims to assess participants' data wrangling, analysis and presentation skills.
Data Source	https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-show

Criteria	• Data Cleaning and Preprocessing (20%)
	    o Handling of missing values
	    o Data transformation and normalization
	    o Feature engineering
	• Data Modeling (20%)
	    o Data normalization
	• Exploratory Data Analysis (20%)
	    o Insightfulness of visualizations
	    o Clarity of findings
	    o Depth of analysis
	• Data Visualization (30%)
	    o Creativity and clarity of visualizations
	    o Effectiveness in communicating insights
	• Presentation and Documentation (10%)
	    o Organization and clarity of the report and presentation
	    o Reproducibility of the analysis

High-Level Summary

- **Total Movies Analyzed**: 1000

- **Total Gross Revenues**: $56.5 billion

- **Total Votes**: 273.7 million

- **Total Genres**: 21

- **Total Directors**: 548

- **Total Stars**: 2709
- **Average IMDB Rating**: 7.5
- **Average Meta Score**: 65.7
- **Average Runtime**: 123 minutes

Top Stars

1. **Most Active Star**: Robert De Niro (17 movies)
2. **Top Grossing Star**: Robert Downey Jr. ($3.1 billion in Gross Revenue)
3. **Highest Voted Star**: Leonardo DiCaprio (10.78 million votes)
4. **Most Versatile Actor**: Matt Damon (13 genres)
5. **Star with Most Directors Worked With**: Robert De Niro (12 directors)
6. **Star with Most Co-Stars**: Robert De Niro (44 co-stars)
7. **Highest Average IMDB Rating**: William Sadler (9.0)
8. **Highest Average Meta Score**: Wendell Corey (100.0)
9. **Longest Average Movie Runtime**: Tigmanshu Dhulia and Richa Cahada (310 minutes)

Top Movies

1. **Top Voted Movie**: The Shawshank Redemption (2.34 million votes)
2. **Highest Grossing Movie**: The Dark Knight ($0.53 billion in Gross Revenue)
3. **Highest IMDB Rated Movies**: The Dark Knight, The Godfather, The Shawshank Redemption (9.0)

Top Directors

1. **Director with Highest Gross vs. Average Votes**: Christopher Nolan (81%)
2. **Director with Highest Gross vs. Average IMDb Rating**: Steven Spielberg (94%)
3. **Director with Highest Gross vs. Meta Scores**: Steven Spielberg (90%)

Top Genres

1. **Most Popular Genre by Number of Movies Released**: Drama (72%)
2. **Highest Grossing Genre**: Adventure ($28 billion from 196 movies)

Detailed Analysis

Star Performance:

- Robert De Niro has the most movies (17) and has worked with the most directors (12) and co-stars (44).

- Robert Downey Jr. leads in total gross revenue ($3.1 billion).

- Leonardo DiCaprio has the highest total votes (10.78 million).

- Matt Damon is noted for his versatility across 13 different genres.

- William Sadler and Wendell Corey stand out with the highest average IMDB rating (9.0) and Meta Score (100.0) respectively.

- Tigmanshu Dhulia's movies have the longest average runtime (310 minutes).

Movie Performance:

- "The Shawshank Redemption" is the top voted movie with 2.34 million votes.

- "The Dark Knight" is the highest-grossing movie with $0.53 billion.

- Top-rated movies include "The Dark Knight," "The Godfather," and "The Shawshank Redemption," each with an IMDB rating of 9.0.

Director Performance:

- Steven Spielberg tops in terms of gross revenue vs. average votes and meta scores, indicating high audience engagement and critical acclaim.

- Christopher Nolan is also highly rated, leading in average votes received per movie.

Genre Performance:

- Drama dominates as the most popular genre by the number of movies released (72% of the 1000 movies).

- Adventure is the highest-grossing genre, with $28 billion from 196 movies, highlighting its commercial success.

Insights

1. Revenue and Popularity Trends

- **High Correlation Between Star Power and Revenue**:

- Stars like Robert Downey Jr. and Leonardo DiCaprio not only have high grossing movies but also rank high in total votes, indicating a strong correlation between star power and box office success.

- **Blockbuster Directors**:

    - Steven Spielberg and Christopher Nolan consistently produce high-grossing and critically acclaimed movies, making them key influencers in the industry.

2. Genre Analysis

- **Adventure and Drama Dominate**:

    - Adventure movies, though fewer in number compared to Drama, generate the highest revenue. This suggests that while dramas are more common, adventure films tend to be higher budget and appeal to wider audiences.

- **Critical Reception of Genres**:

    - The high average IMDb ratings and Meta Scores are Western and Animation suggests these genres are critically acclaimed, despite not being the highest grossing.

3. Star Metrics

- **Most Prolific Actors**:

    - Robert De Niro and Matt Damon are highly versatile and prolific, appearing in a wide variety of genres and working with numerous directors and co-stars. This versatility likely contributes to their enduring careers.

- **Star Run Time**

    - Tigmanshu Dhulia and Richa Cahada have significantly higher movie runtime than others, indicating a preference for longer storytelling formats in their films.

4. Director Influence

- **Director's Box Office Success**:

    - Directors like Steven Spielberg, who consistently produce high-grossing films, set benchmarks for commercial success in the industry.
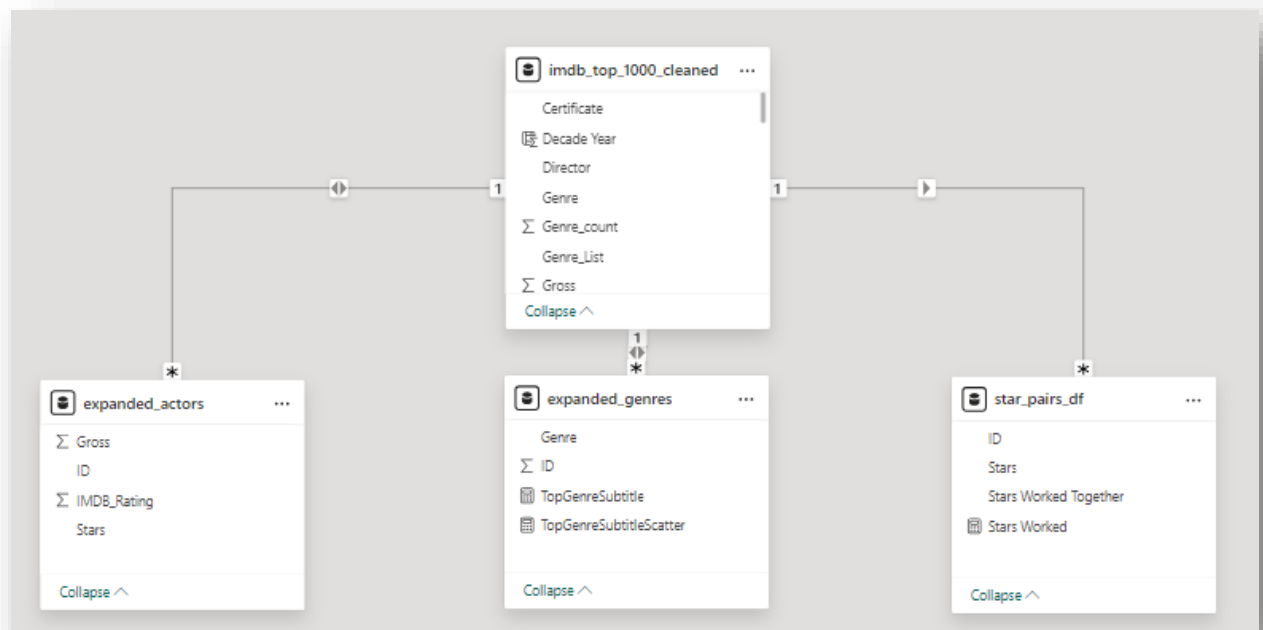
5. Ratings and Scores

- **High IMDB and Meta Scores**:

    - Films with high IMDB and Meta Scores like "The Dark Knight" and "The Godfather" highlight the importance of critical acclaim in achieving long-term popularity and high audience engagement.

6. Voting Trends

- **Audience Engagement**:
    - Movies with high vote counts on IMDB indicate strong audience engagement. Comparing vote counts with box office revenue and ratings can reveal how audience engagement correlates with financial success and critical acclaim.

Data Modeling

Data Cleaning and Preprocessing

- Created new column ID to have primary key for every movie
- Remove "mins" for Run time
- Updated correct data types for Released Year, Gross, Meta Score, No of Votes, Runtime and IMDB Rating
- Fill blank values from Certificate to "Unrated"

Data Normalization - First Normal Form (1NF)

- Created new table called "expanded_actors"

ID	Stars
19	Lin-Manuel Miranda
19	Phillipa Soo
19	Leslie Odom Jr.
19	Renée Elise Goldsberry
21	Suriya
21	Madhavan
21	Paresh Rawal

- o Created new table called "expanded_genres"

ID	Genre
1	Drama
2	Drama
3	Drama
4	Drama
5	Drama
6	Drama
7	Drama
8	Drama

- o Created new table called "star_pairs_df"

ID	Stars	Stars Worked Together
225	Aamir Bashir	Naseeruddin Shah
225	Aamir Bashir	Jimmy Sheirgill
225	Aamir Bashir	Anupam Kher
58	Aamir Khan	Sanya Malhotra
248	Aamir Khan	Rachel Shelley
241	Aamir Khan	Saif Ali Khan
248	Aamir Khan	Gracy Singh
248	Aamir Khan	Raghuvir Yadav
241	Aamir Khan	Preity Zinta
65	Aamir Khan	Madhavan
65	Aamir Khan	Mona Singh
65	Aamir Khan	Sharman Joshi
66	Aamir Khan	Amole Gupte

Parameters

- Created values parameter – Gross, IMDb Rating, Meta Score, Runtime, Votes

```
1  ShowValuesBy = {
2      ("Gross", NAMEOF('imdb_top_1000_cleaned'[@Gross]), 0),
3      ("IMDB Rating", NAMEOF('imdb_top_1000_cleaned'[@IMDB Rating]), 1),
4      ("Meta Score", NAMEOF('imdb_top_1000_cleaned'[@Meta Score]), 2),
5      ("Runtime", NAMEOF('imdb_top_1000_cleaned'[@Runtime]), 3),
6      ("Votes", NAMEOF('imdb_top_1000_cleaned'[@Votes]), 4)
7  }
```

ShowValuesBy	ShowValuesBy Fields	ShowValuesBy Order
Gross	'imdb_top_1000_cleaned'[@Gross]	0
IMDB Rating	'imdb_top_1000_cleaned'[@IMDB Rating]	1
Meta Score	'imdb_top_1000_cleaned'[@Meta Score]	2
Runtime	'imdb_top_1000_cleaned'[@Runtime]	3
Votes	'imdb_top_1000_cleaned'[@Votes]	4

- Created view parameter – Stars, Genre, Director, Certificate, Released Year

```
1  View by = {
2      ("Stars", NAMEOF('expanded_actors'[Stars]), 0),
3      ("Genre", NAMEOF('expanded_genres'[Genre]), 1),
4      ("Director", NAMEOF('imdb_top_1000_cleaned'[Director]), 2),
5      ("Certificate", NAMEOF('imdb_top_1000_cleaned'[Certificate]), 3),
6      ("Released Year", NAMEOF('imdb_top_1000_cleaned'[Released Year]), 4)
7  }
```

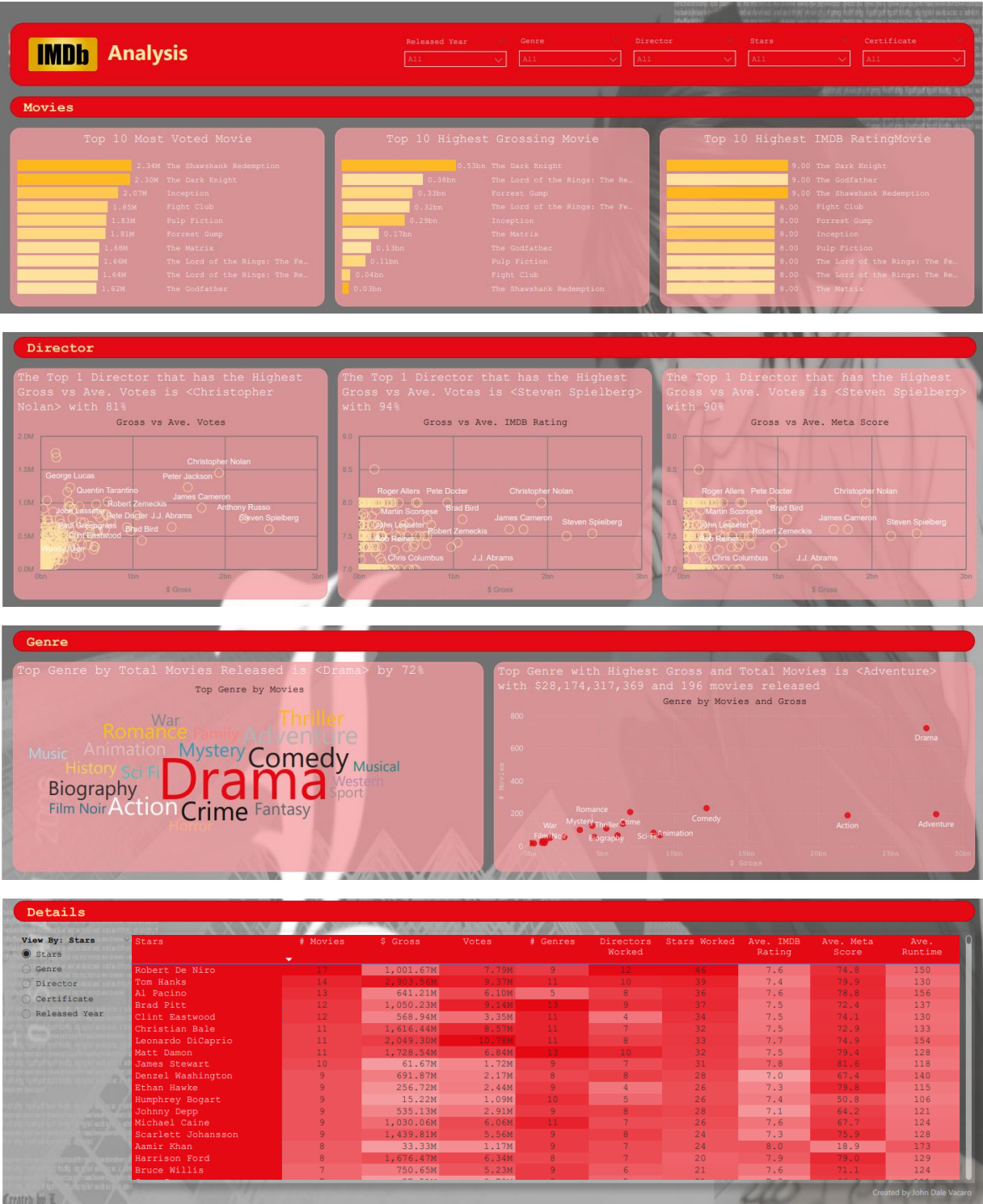| Parameter	Parameter Fields	Parameter Order	ParameterDisplayName
Stars	'expanded_actors'[Stars]	0	Stars
Genre	'expanded_genres'[Genre]	1	Genre
Director	'imdb_top_1000_cleaned'[Director]	2	Director
Certificate	'imdb_top_1000_cleaned'[Certificate]	3	Certificate
Released Year	'imdb_top_1000_cleaned'[Released Year]	4	Released Year

Power BI Dashboard Snapshot

Cover Page

Overview Page

Analysis Page

IMDb Analysis

Released Year: All | Genre: All | Director: All | Stars: All | Certificate: All

Movies

Top 10 Most Voted Movie
- 2.34M The Shawshank Redemption
- 2.30M The Dark Knight
- 2.07M Inception
- 1.85M Fight Club
- 1.83M Pulp Fiction
- 1.81M Forrest Gump
- 1.68M The Matrix
- 1.66M The Lord of the Rings: The Fe...
- 1.64M The Lord of the Rings: The Re...
- 1.62M The Godfather

Top 10 Highest Grossing Movie
- 0.53bn The Dark Knight
- 0.38bn The Lord of the Rings: The Re...
- 0.33bn Forrest Gump
- 0.32bn The Lord of the Rings: The Fe...
- 0.29bn Inception
- 0.17bn The Matrix
- 0.13bn The Godfather
- 0.11bn Pulp Fiction
- 0.04bn Fight Club
- 0.03bn The Shawshank Redemption

Top 10 Highest IMDB RatingMovie
- 9.00 The Dark Knight
- 9.00 The Godfather
- 9.00 The Shawshank Redemption
- 8.00 Fight Club
- 8.00 Forrest Gump
- 8.00 Inception
- 8.00 Pulp Fiction
- 8.00 The Lord of the Rings: The Fe...
- 8.00 The Lord of the Rings: The Re...
- 8.00 The Matrix

Director

The Top 1 Director that has the Highest Gross vs Ave. Votes is <Christopher Nolan> with 81%

Gross vs Ave. Votes

The Top 1 Director that has the Highest Gross vs Ave. Votes is <Steven Spielberg> with 94%

Gross vs Ave. IMDB Rating

The Top 1 Director that has the Highest Gross vs Ave. Votes is <Steven Spielberg> with 90%

Gross vs Ave. Meta Score

Genre

Top Genre by Total Movies Released is <Drama> by 72%

Top Genre by Movies

War, Thriller, Romance, Family, Adventure, Music, Animation, Mystery, Comedy, History, Sci Fi, Drama, Musical, Biography, Western, Sport, Film Noir, Action, Crime, Fantasy, Horror

Top Genre with Highest Gross and Total Movies is <Adventure> with $28,174,317,369 and 196 movies released

Genre by Movies and Gross

Details

View By: Stars
- ○ Stars
- ○ Genre
- ○ Director
- ○ Certificate
- ○ Released Year

Stars	# Movies	$ Gross	Votes	# Genres	Directors Worked	Stars Worked	Ave. IMDB Rating	Ave. Meta Score	Ave. Runtime
Robert De Niro	17	1,001.67M	7.79M	9	12	46	7.6	74.8	150
Tom Hanks	14	2,903.56M	9.37M	11	10	39	7.4	79.9	130
Al Pacino	13	641.21M	6.10M	5	8	36	7.6	78.8	156
Brad Pitt	12	1,050.23M	9.14M	13	9	37	7.5	72.4	137
Clint Eastwood	12	568.94M	3.35M	11	4	34	7.5	74.1	130
Christian Bale	11	1,616.44M	8.57M	11	7	32	7.5	72.9	133
Leonardo DiCaprio	11	2,049.30M	10.78M	11	8	33	7.7	74.9	154
Matt Damon	11	1,728.54M	6.84M	13	10	32	7.5	79.4	128
James Stewart	10	61.67M	1.72M	9	7	31	7.8	81.6	118
Denzel Washington	9	691.87M	2.17M	8	8	28	7.0	67.4	140
Ethan Hawke	9	256.72M	2.44M	9	4	26	7.3	79.8	115
Humphrey Bogart	9	15.22M	1.09M	10	5	26	7.4	50.8	106
Johnny Depp	9	535.13M	2.91M	9	8	28	7.1	64.2	121
Michael Caine	9	1,030.06M	6.06M	11	7	26	7.6	67.7	124
Scarlett Johansson	9	1,439.81M	5.56M	9	8	24	7.3	75.9	128
Aamir Khan	8	33.33M	1.17M	9	7	24	8.0	18.9	173
Harrison Ford	8	1,676.47M	6.34M	9	7	20	7.9	79.0	129
Bruce Willis	7	750.65M	5.23M	9	6	21	7.6	71.1	124