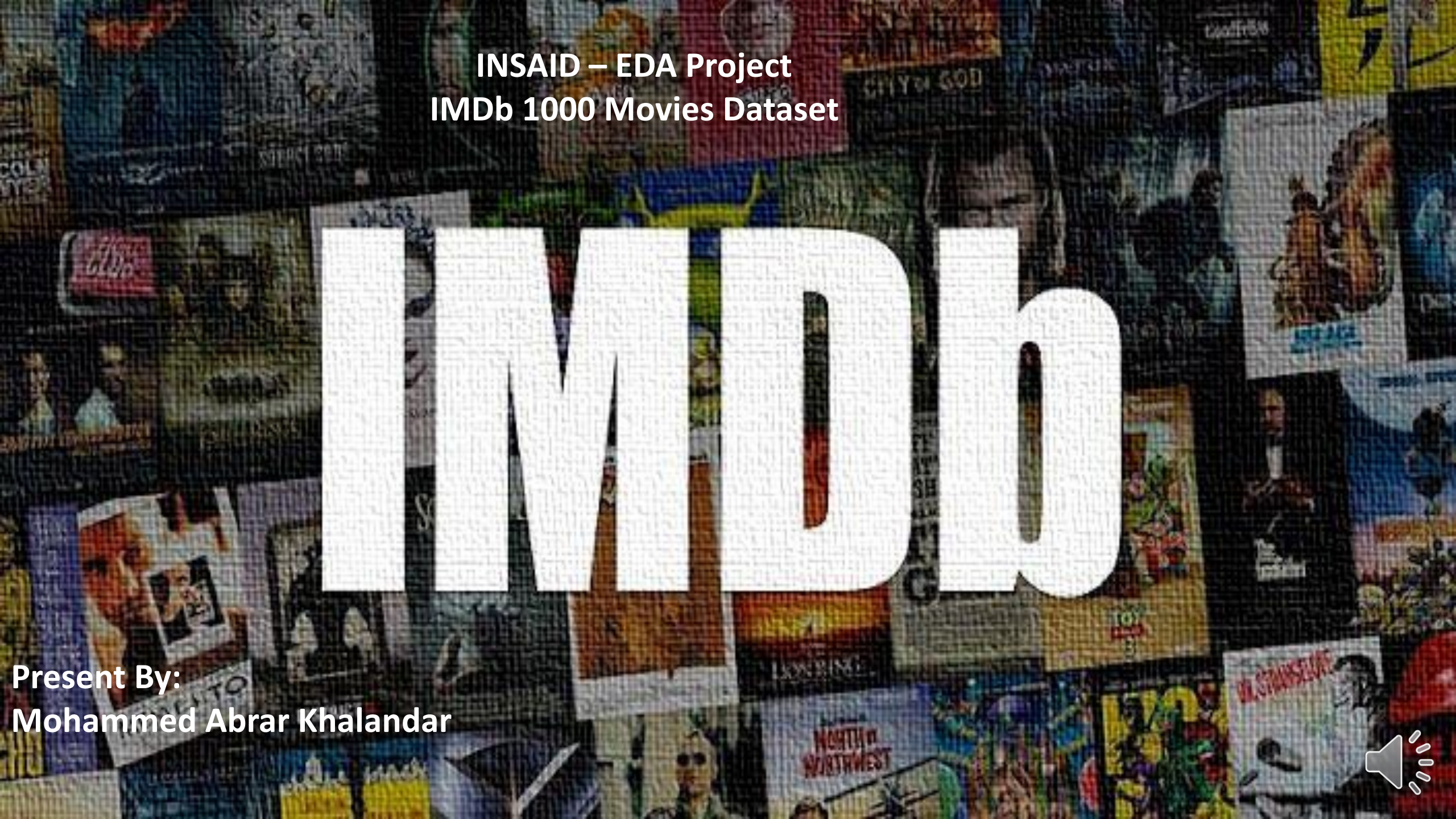
# INSAID – EDA Project
# IMDb 1000 Movies Dataset




**Present By:**
**Mohammed Abrar Khalandar**

# Outline

- Problem statement and approach
- Data description and loading
- Data profiling
- Deriving best Genre, Actors and Directors based on their Movies Revenue, Ratings and Metascore
- Conclusion

# Problem statement and approach

- A Production needs to take a decision for movie production based on IMDb dataset.
- The dataset consists of 1000 Movies of data.
- In order make a right decision the production company should choose the best Genre, Actors and Director for the Movies success.
- We shall now see how Actors, Director and the Genre is chosen based on their Revenue(Millions), Rating and Metascore

# Data description and loading

- The Dataset is derived from IMDb which consists of 1000 movies from year 2006 to 2016
- The Dataset consists of information such as Movie Title, Genre, Director, Actors, Ratings, Revenue etc.
- The Dataset Comprises of 1000 rows and 12 columns

# Datasets Columns description

| Columns | Description |
|---|---|
| Rank | Ranking of Movies |
| Title | Title of the film |
| Genre | List of genres to classify the Movies |
| Description | Brief summary of the Movies |
| Director | Movies Directors |
| Actors | Main Actors of the Movies |
| Year | Release year of the Movies |
| Runtime (Minutes) | Duration of the film in minutes |
| Rating | User ratings for the movie 0-10 |
| Votes | Number of votes |
| Revenue (Millions) | Movies revenue in millions |
| Metascore | Average of critic scores from 0-100 |

# Data profiling

- Pre-Profiling

## Dataset info

| | |
|---|---|
| **Number of variables** | 12 |
| **Number of observations** | 1000 |
| **Total Missing (%)** | 1.6% |
| **Total size in memory** | 93.8 KiB |
| **Average record size in memory** | 96.1 B |

## Variables types

| | |
|---|---|
| **Numeric** | 7 |
| **Categorical** | 4 |
| **Boolean** | 0 |
| **Date** | 0 |
| **Text (Unique)** | 1 |
| **Rejected** | 0 |
| **Unsupported** | 0 |

## Warnings

`Actors` has a high cardinality: 996 distinct values **Warning**
`Director` has a high cardinality: 644 distinct values **Warning**
`Genre` has a high cardinality: 207 distinct values **Warning**
`Metascore` has 64 / 6.4% missing values **Missing**
`Revenue (Millions)` has 128 / 12.8% missing values **Missing**
`Title` has a high cardinality: 999 distinct values **Warning**

# Data profiling

- Pre-Processing
- **Metascore** has **6.4%** missing values and hence we fill up the missing data with *Mean* values grouped by year
- **Revenue (Millions)** has **12.8%** missing values and hence we fill up the missing data with *Mean* values grouped by year

# Data profiling

- Post - Processing

## Dataset info

| | |
|---|---|
| **Number of variables** | 12 |
| **Number of observations** | 1000 |
| **Total Missing (%)** | 0.0% |
| **Total size in memory** | 93.8 KiB |
| **Average record size in memory** | 96.1 B |

## Variables types

| | |
|---|---|
| **Numeric** | 7 |
| **Categorical** | 4 |
| **Boolean** | 0 |
| **Date** | 0 |
| **Text (Unique)** | 1 |
| **Rejected** | 0 |
| **Unsupported** | 0 |

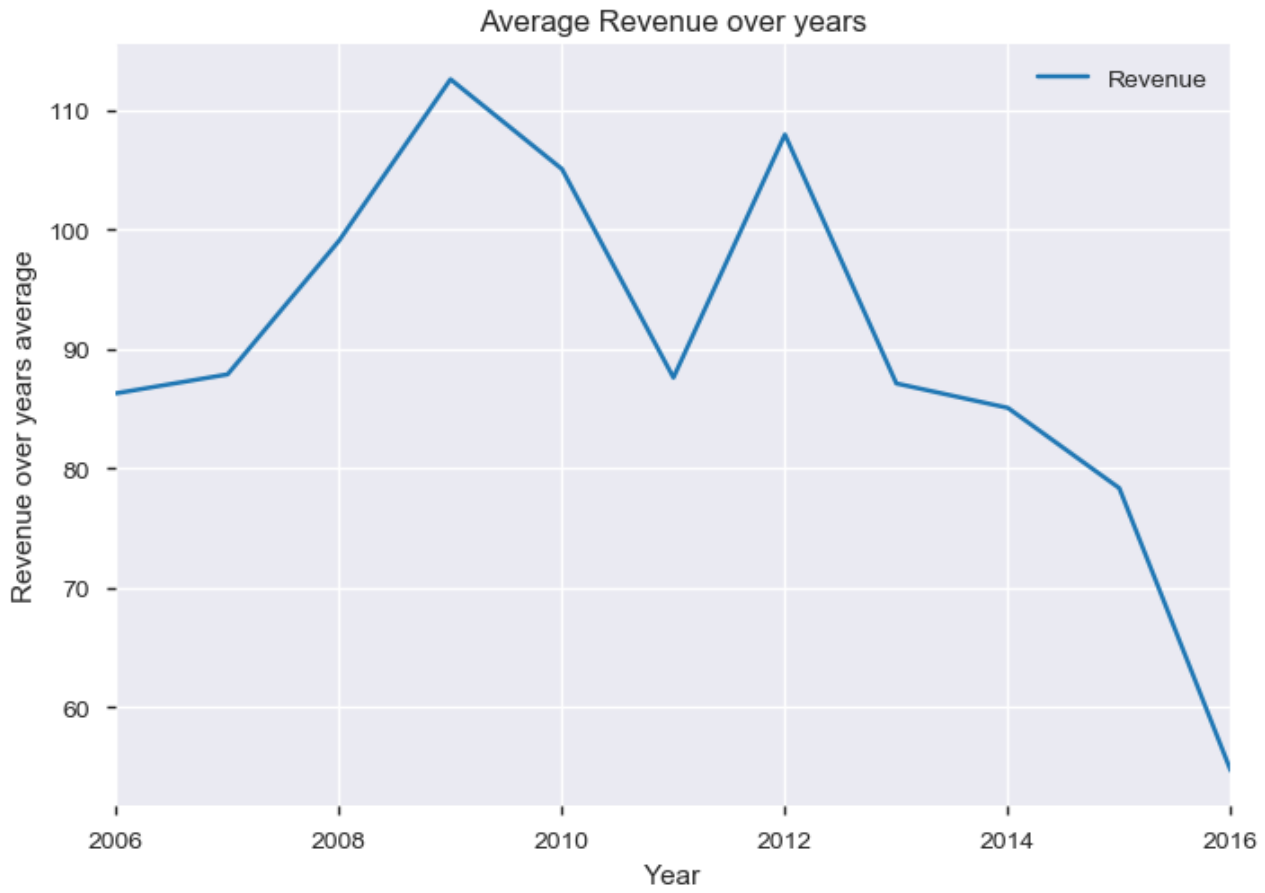
## Warnings

`Actors` has a high cardinality: 996 distinct values Warning
`Director` has a high cardinality: 644 distinct values Warning
`Genre` has a high cardinality: 207 distinct values Warning
`Title` has a high cardinality: 999 distinct values Warning

# Deriving Top Genre, Actors and Directors based on their Movies Revenue, Ratings and Metascore

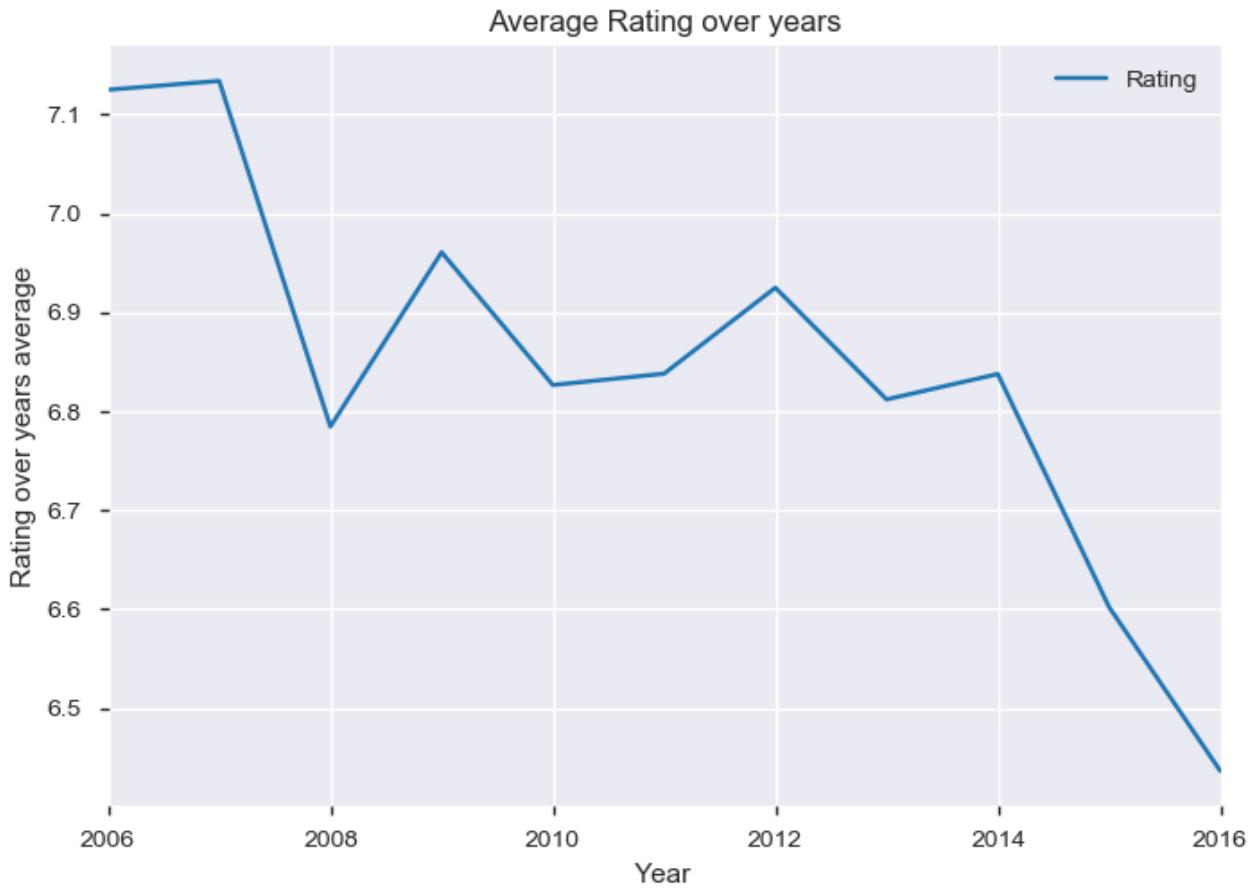
- Years in which Movies Revenues had a rise?
- Years in which the Movies Ratings had a rise?
- Years in which the Movies Metascore had a rise?

# Years in which Movies Revenues had a rise



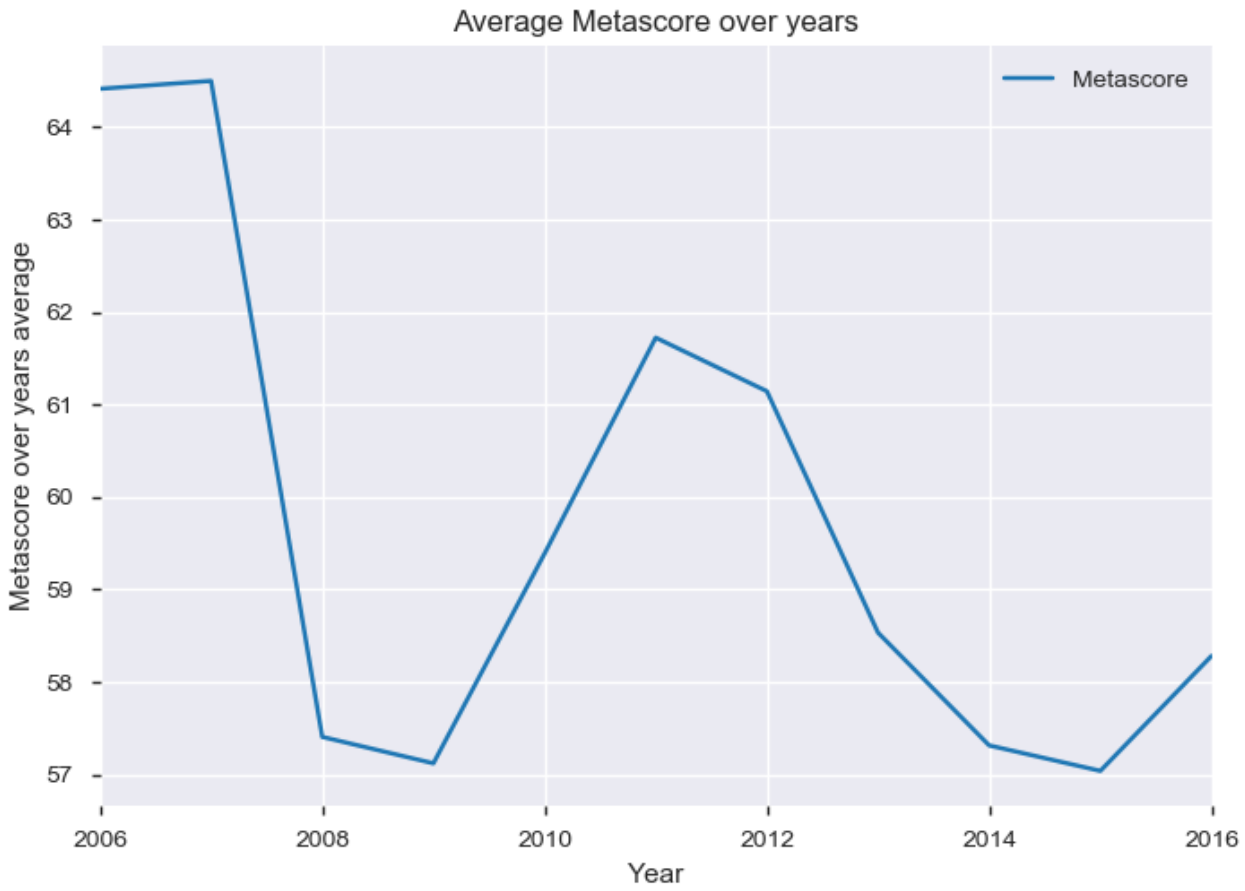

Years **2009** and **2012** has the best Revenue generated in the years 2006 - 2016

# Years in which the Movies Ratings had a rise




Years **2007**, **2009** and **2012** had the best Average ratings for Movies in the years 2006 - 2016

# Years in which the Movies Metascore had a rise

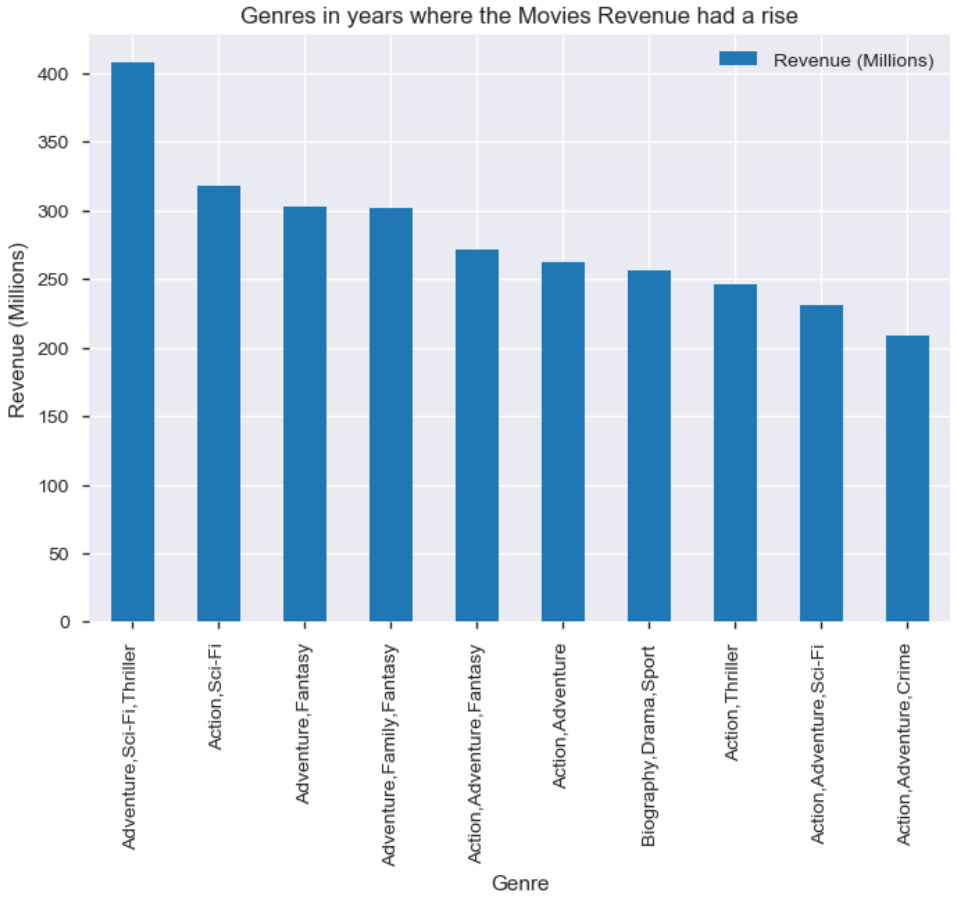



Year **2007** and **2011** had the best average Metascore in years 2006 - 2016

# Deriving best Genre from the Dataset

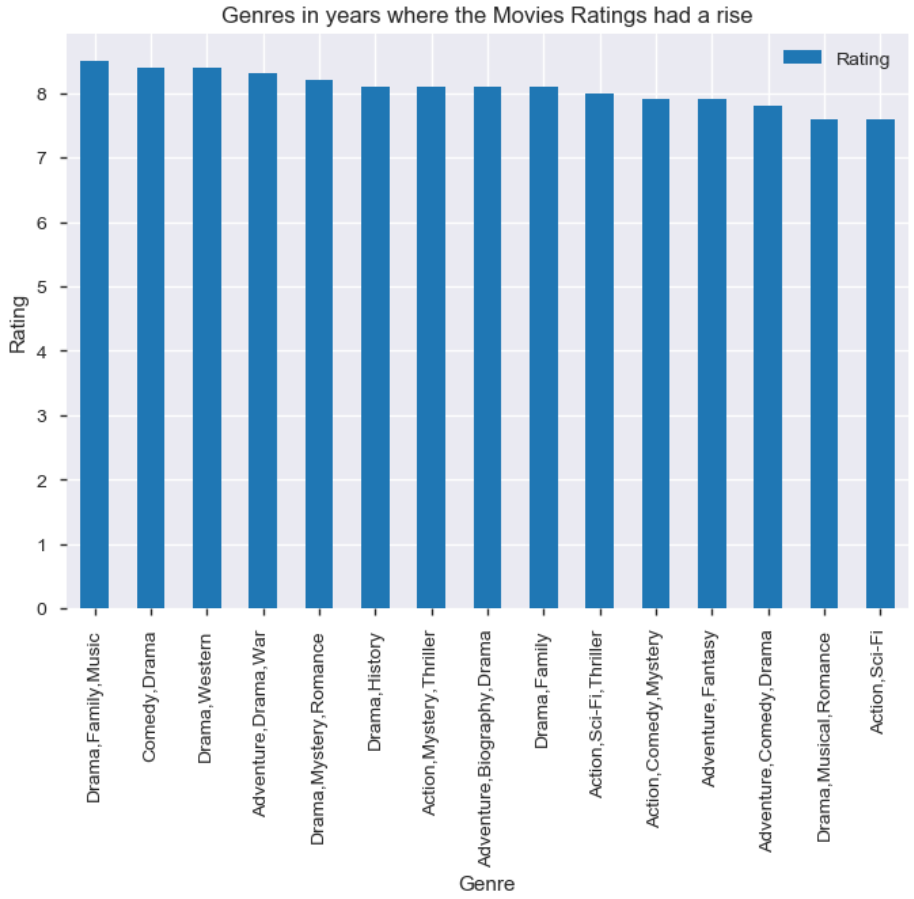
- Which are the top most Genres in years where the Movies Revenue had a rise?
- Which are the top most Genres in years where the Movies Ratings had a rise?
- Which are the top most Genres in years where the Movies Metascore had a rise?
- Which are the top 10 Genres based on Revenue, Ratings and Metascore?

# Which are the top most Genres in years where the Movies Revenue had a rise



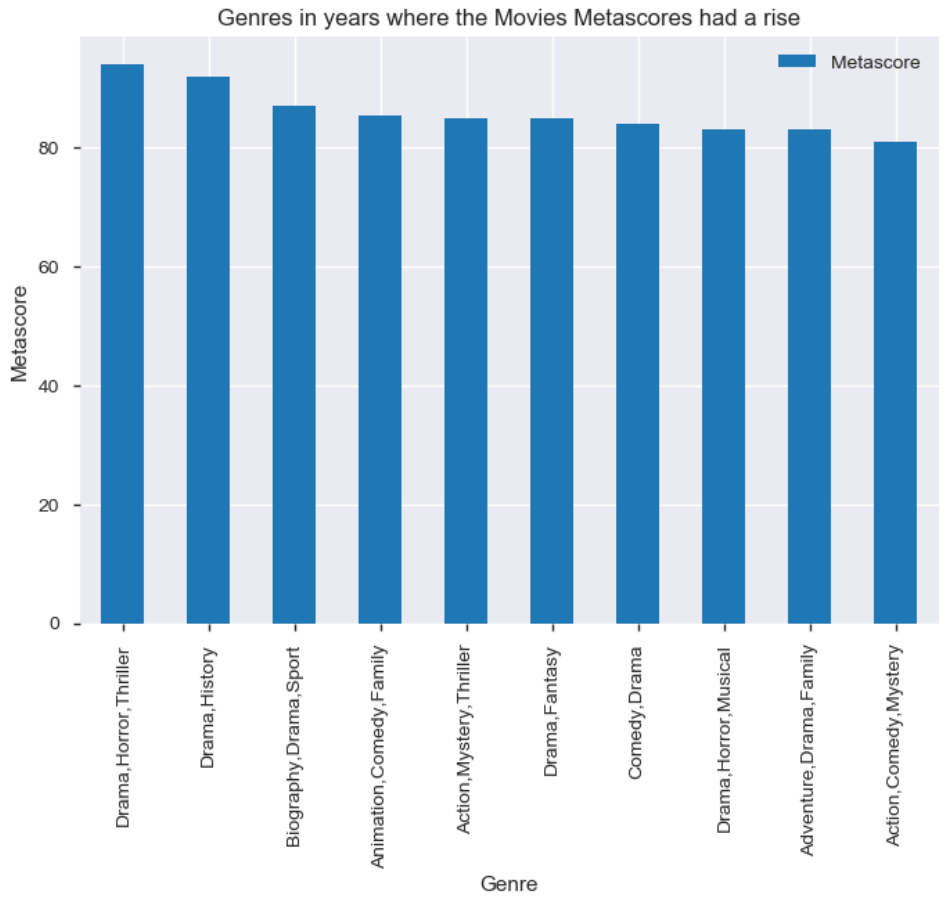

As years **2009** and **2012** had the highest revenue, Movies with Genre combination **Adventure, Sci-Fi, Thriller** tops the chart based on Revenue (Millions)

# Which are the top most Genres in years where the Movies Ratings had a rise



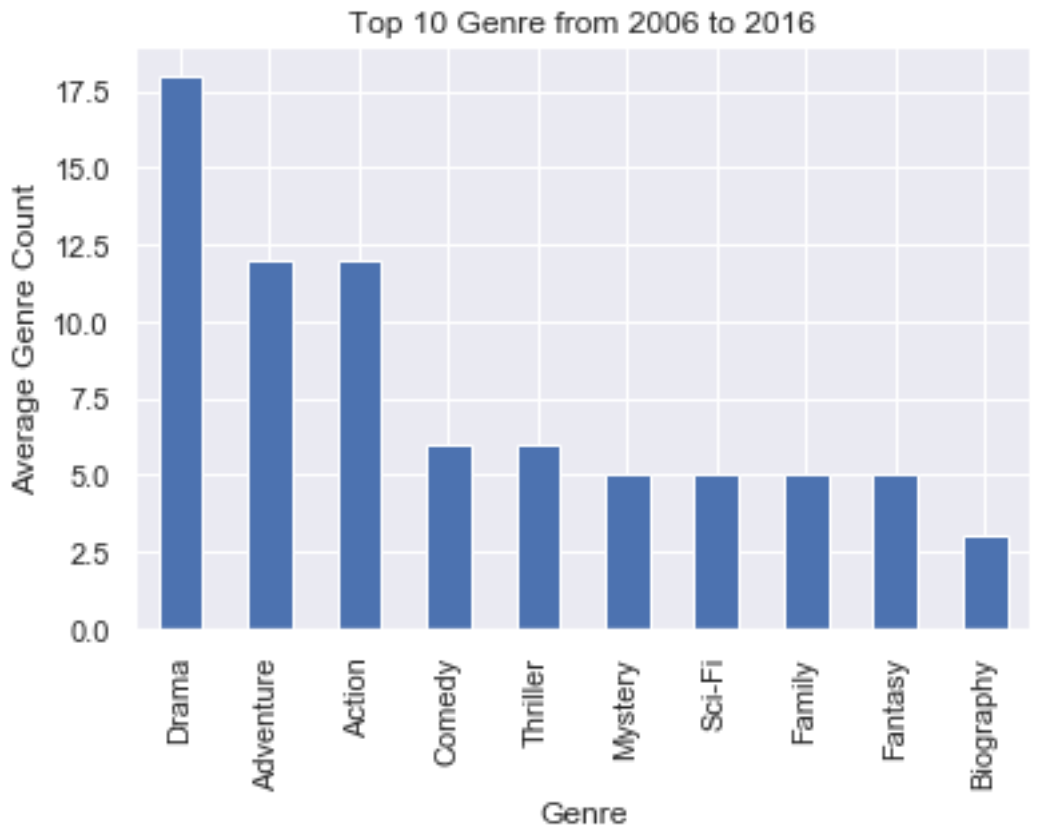

As years **2007**, **2009** and **2012** had the highest Rating, Movies with **Drama, Family, Music** tops the chart based on ratings

# Which are the top most Genres in years where the Movies Metascore had a rise




As years **2007** and **2011** had the highest Metascore, Movies with Genre Combination **Drama, Horror and Thriller** tops the chart based on Metascore

# Which are the top 10 Genres based on Revenue, Ratings and Metascore

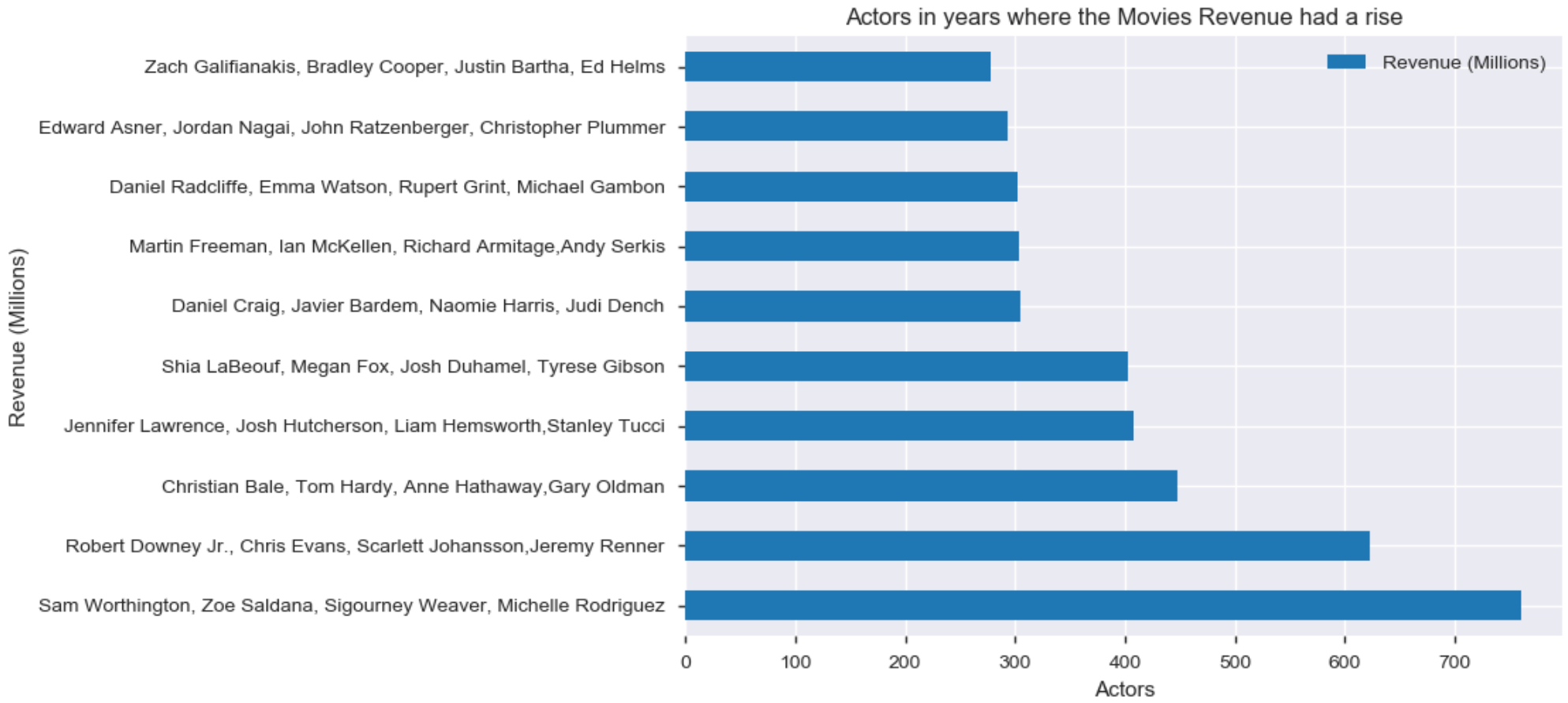



**Top 10 Genres based on Revenue, Ratings and Metascore**

# Deriving best Actors from the Dataset

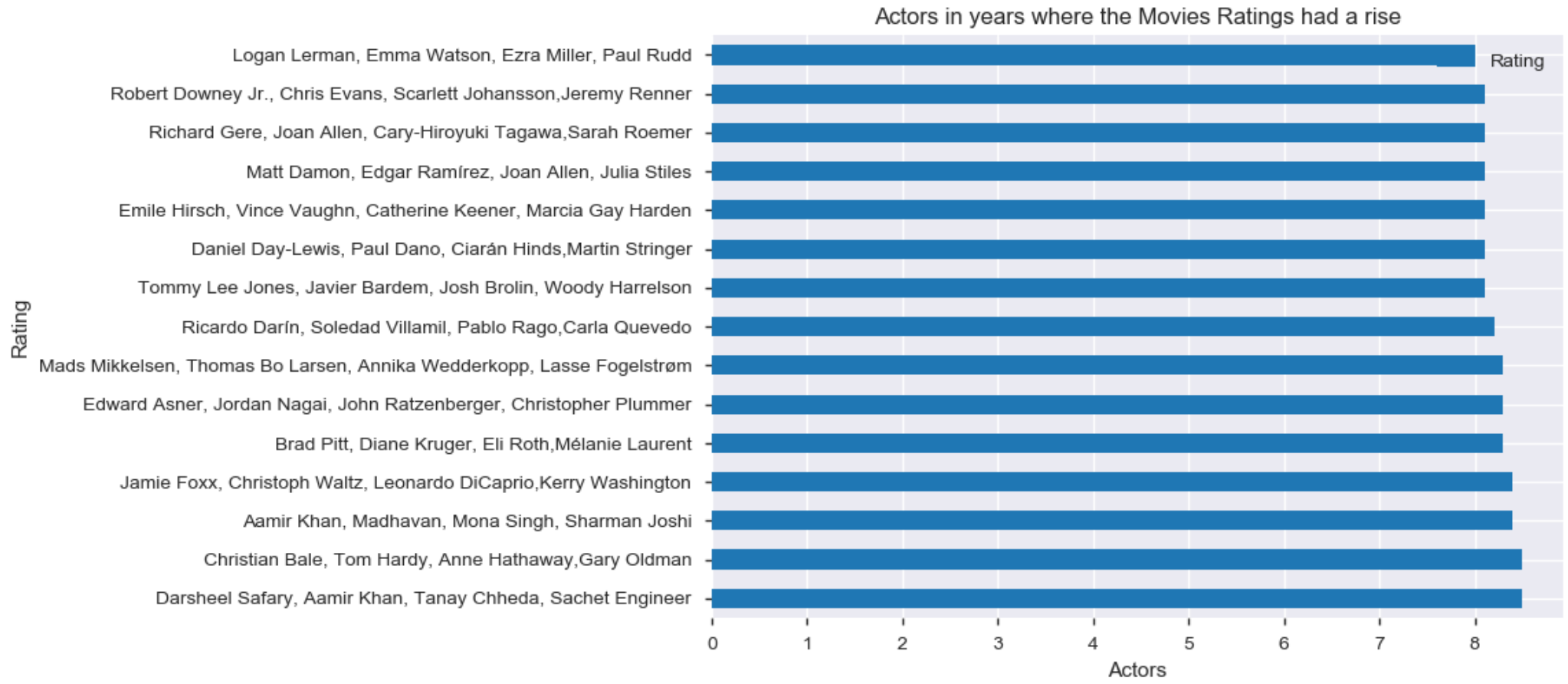
- Which are the top most Actors in years where the Movies Revenue had a rise?
- Which are the top most Actors in years where the Movies Ratings had a rise?
- Which are the top most Actors in years where the Movies Metascore had a rise?
- Which are the top 10 Actors based on Revenue, Ratings and Metascore?

# Which are the top most Actors in years where the Movies Revenue had a rise?




As years **2009** and **2012** had the highest revenue, Movie with combination **Sam Worthington, Zoe Saldana, Sigourney Weaver, Michelle Rodriguez** tops during period (2006 - 2016)

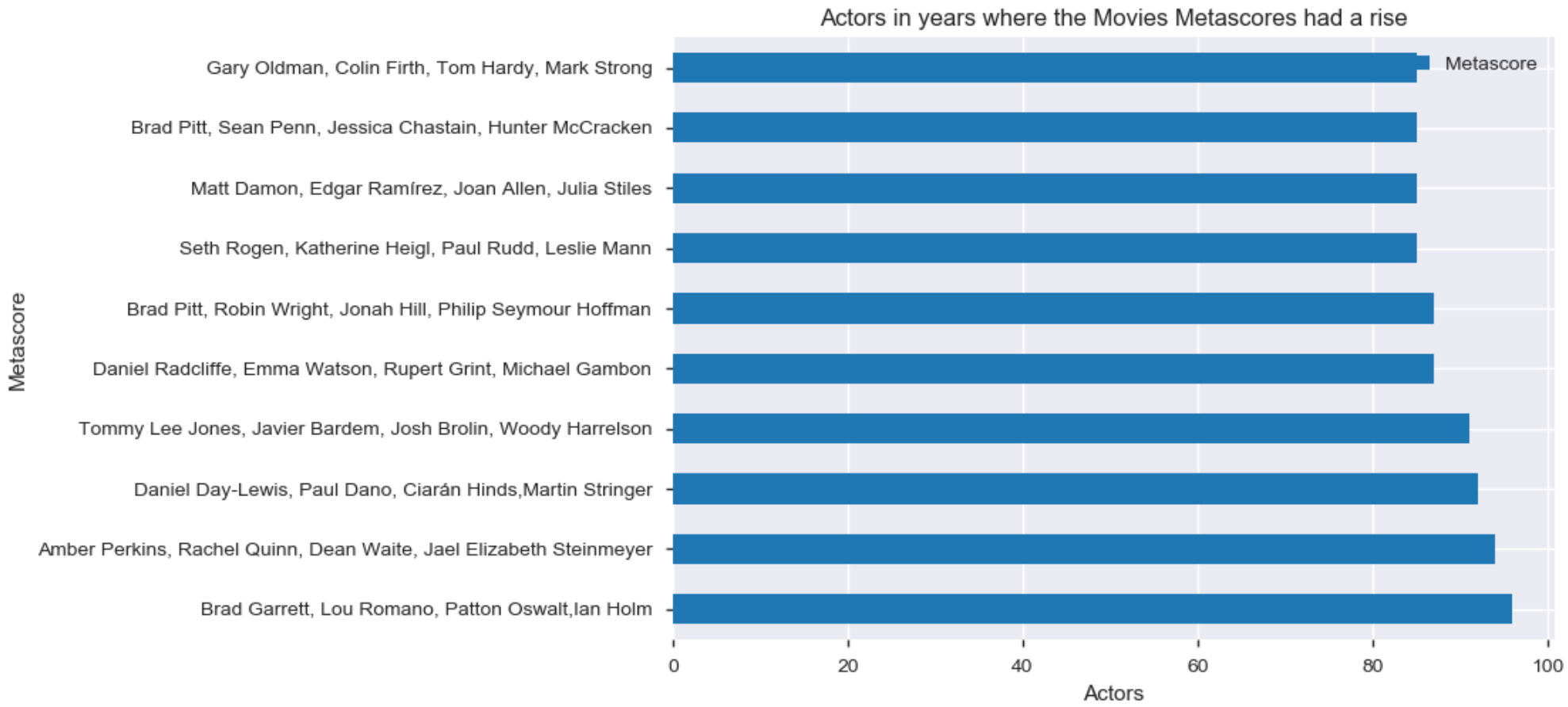
# Which are the top most Actors in years where the Movies Ratings had a rise




As years **2007**, **2009** and **2012** had the highest Rating, Actors with combination:

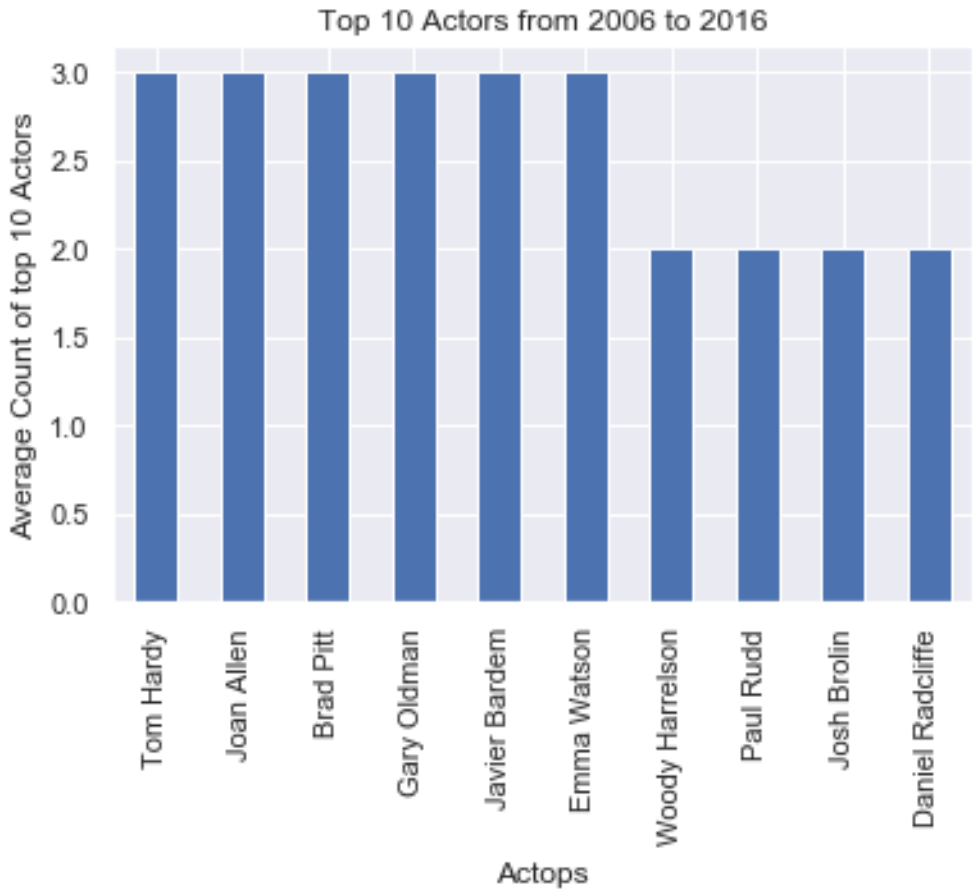
1. **Darsheel Safary, Aamir Khan, Tanay Chheda, Sachet Engineer**
2. **Christian Bale, Tom Hardy, Anne Hathaway, Gary Oldman** tops during the period (2006 - 2016)

# Which are the top most Actors in years where the Movies Metascore had a rise




As years **2007**, and **2011** had the highest Metascore, Actors with combination:
**Brad Garrett, Lou Romano, Patton Oswalt, Ian Holm** tops during the period (2006 - 2016)

# Which are the top 10 Actors based on Revenue, Ratings and Metascore

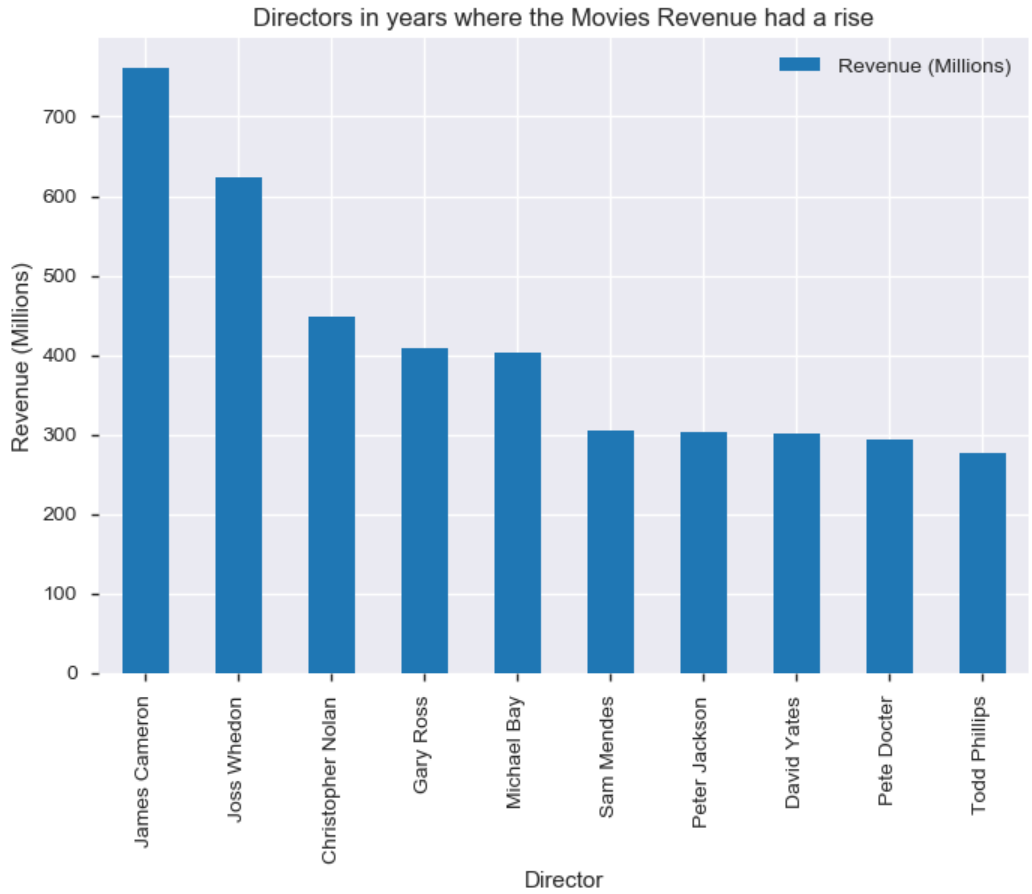



**Top 10 Actors based on Revenue, Ratings and Metascore**

# Deriving best Director from the Dataset

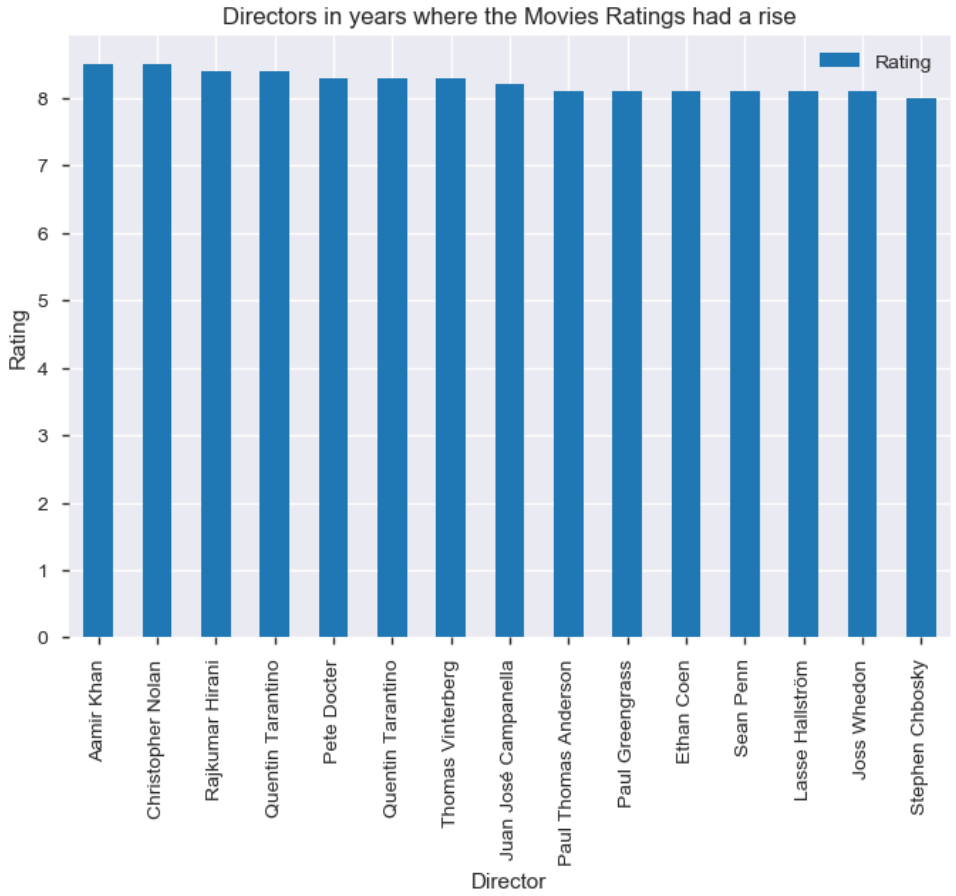
- Which are the top most Director in years where the Movies Revenue had a rise?
- Which are the top most Director in years where the Movies Ratings had a rise?
- Which are the top most Director in years where the Movies Metascore had a rise?
- Which are the top 10 Director based on Revenue, Ratings and Metascore?

# Which are the top most Director in years where the Movies Revenue had a rise



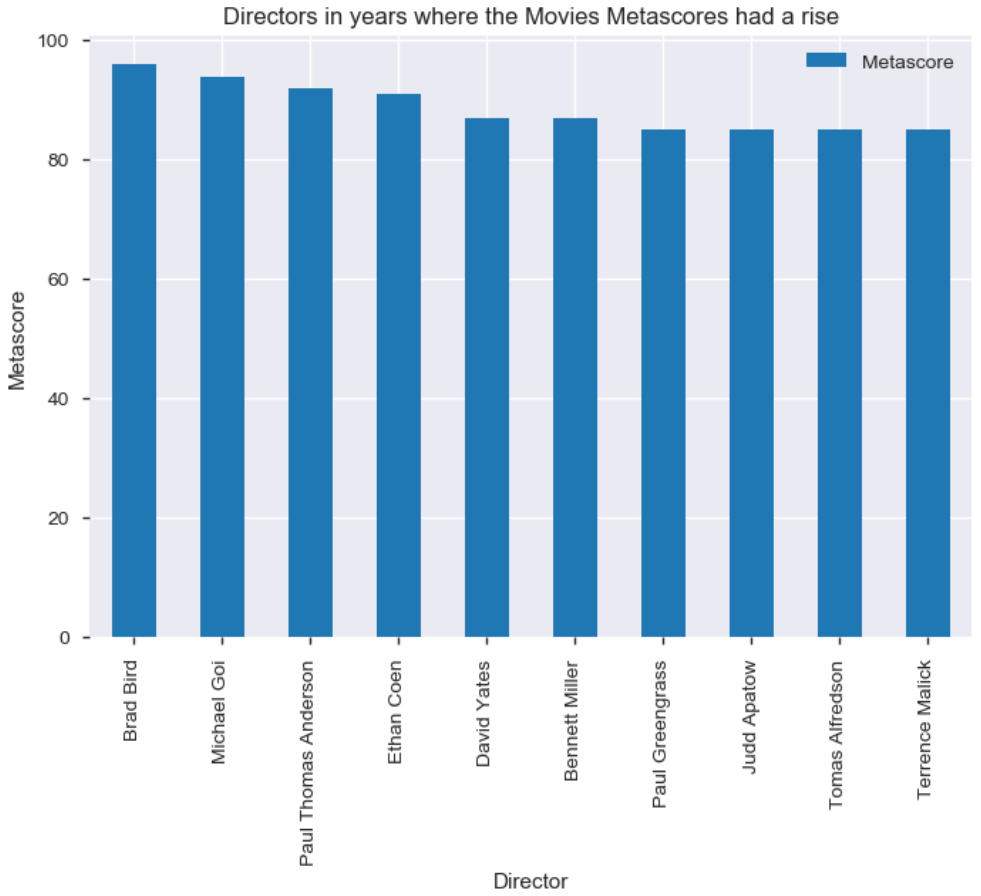

As years **2009** and **2012** had the highest revenue, Movies directed by **James Cameron** tops the chart

# Which are the top most Director in years where the Movies Ratings had a rise



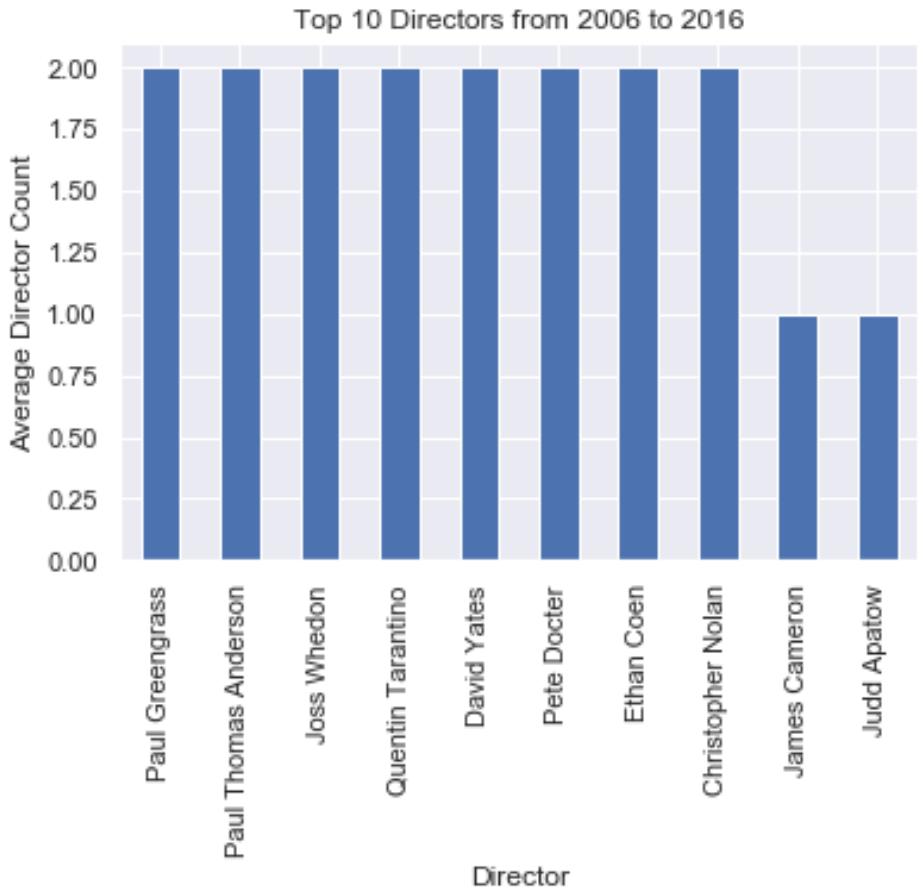

As years **2007**, **2009** and **2012** had the highest Rating, Director **Aamir Khan** tops the chart based on Rating

# Which are the top most Director in years where the Movies Metascore had a rise




As years **2007** and **2011** had the highest Metascore, **Brad Bird** tops the chart as best director based on Metascore

# Which are the top 10 Director based on Revenue, Ratings and Metascore




**Top 10 Directors based on Revenue, Ratings and Metascore**

# Conclusion

| Best Genre |
| --- |
| Drama |
| Adventure |
| Action |
| Comedy |
| Thriller |
| Mystery |
| Sci-Fi |
| Family |
| Fantasy |
| Biography |

| Best Actors |
| --- |
| Tom Hardy |
| Joan Allen |
| Brad Pitt |
| Gary Oldman |
| Javier Bardem |
| Emma Watson |

| Best Directors |
| --- |
| Paul Greengrass |
| Paul Thomas Anderson |
| Joss Whedon |
| Quentin Tarantino |
| David Yates |
| Pete Docter |
| Ethan Coen |
| Christopher Nolan |

Movie production with above Genre combination, Actors and Director should yield production company a good result.

# Thank You