INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

A Data-Driven Analysis of IMDb Movie Ratings

INNOMATICS
RESEARCH LABS

ABOUT US:

We are a team of passionate data enthusiasts with diverse backgrounds, united by our interest in extracting valuable insights from data.

- **Avula Nikhil** is a data science graduate with a deep passion for analytics and deriving actionable insights.

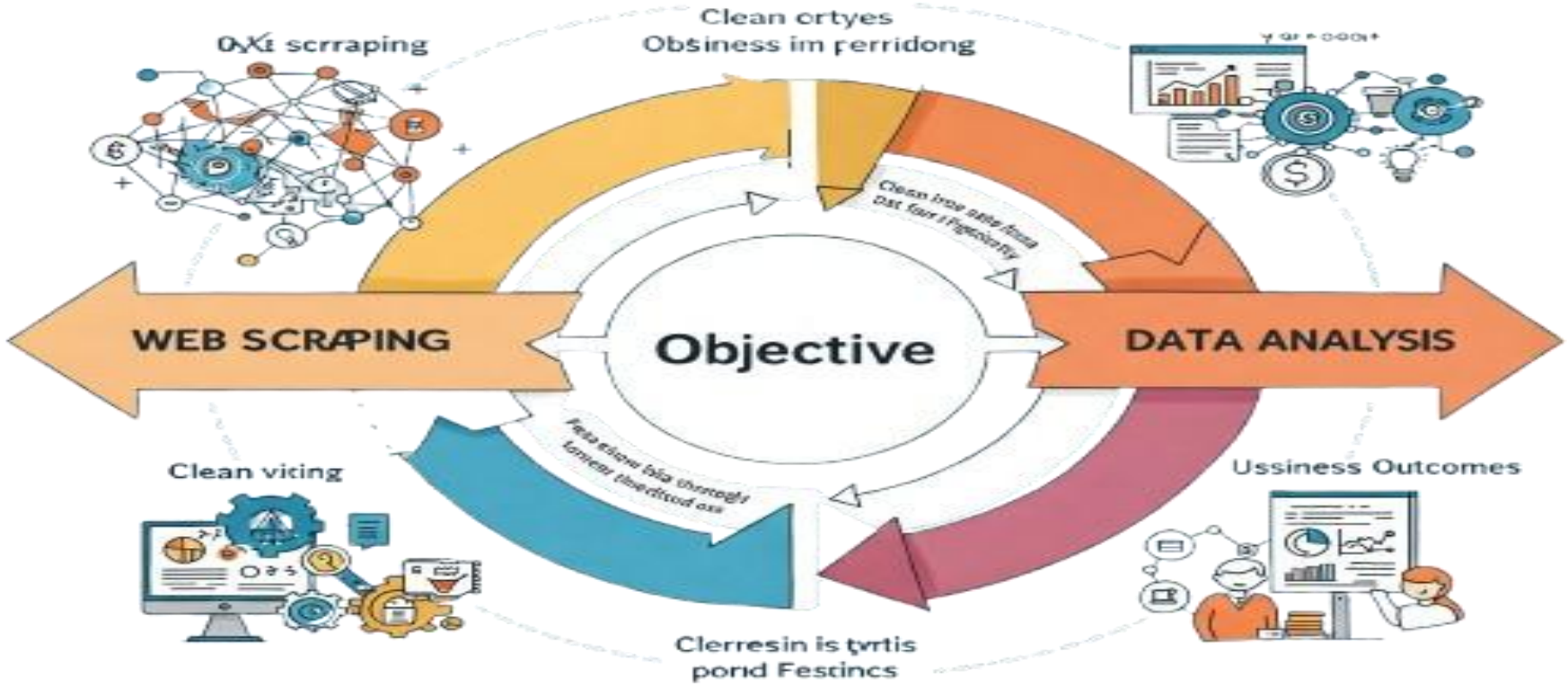
- **Vallala Koushik**, an engineering graduate, brings a strong analytical mindset and a problem-solving approach to our projects.

- **Rajesh Madamanchi**, with an MBA and fintech background, is keen on applying data-driven strategies to optimize business outcomes. Together, we strive to turn raw data into meaningful stories.

INNOMATICS RESEARCH LABS

Objective:

- **Goal**: Analyze IMDb movie ratings to uncover patterns and insights.
- **Approach**: Leverage web scraping and advanced data analysis to derive actionable insights.



INNOMATICS RESEARCH LABS

Web scraping:

- **Website used** :

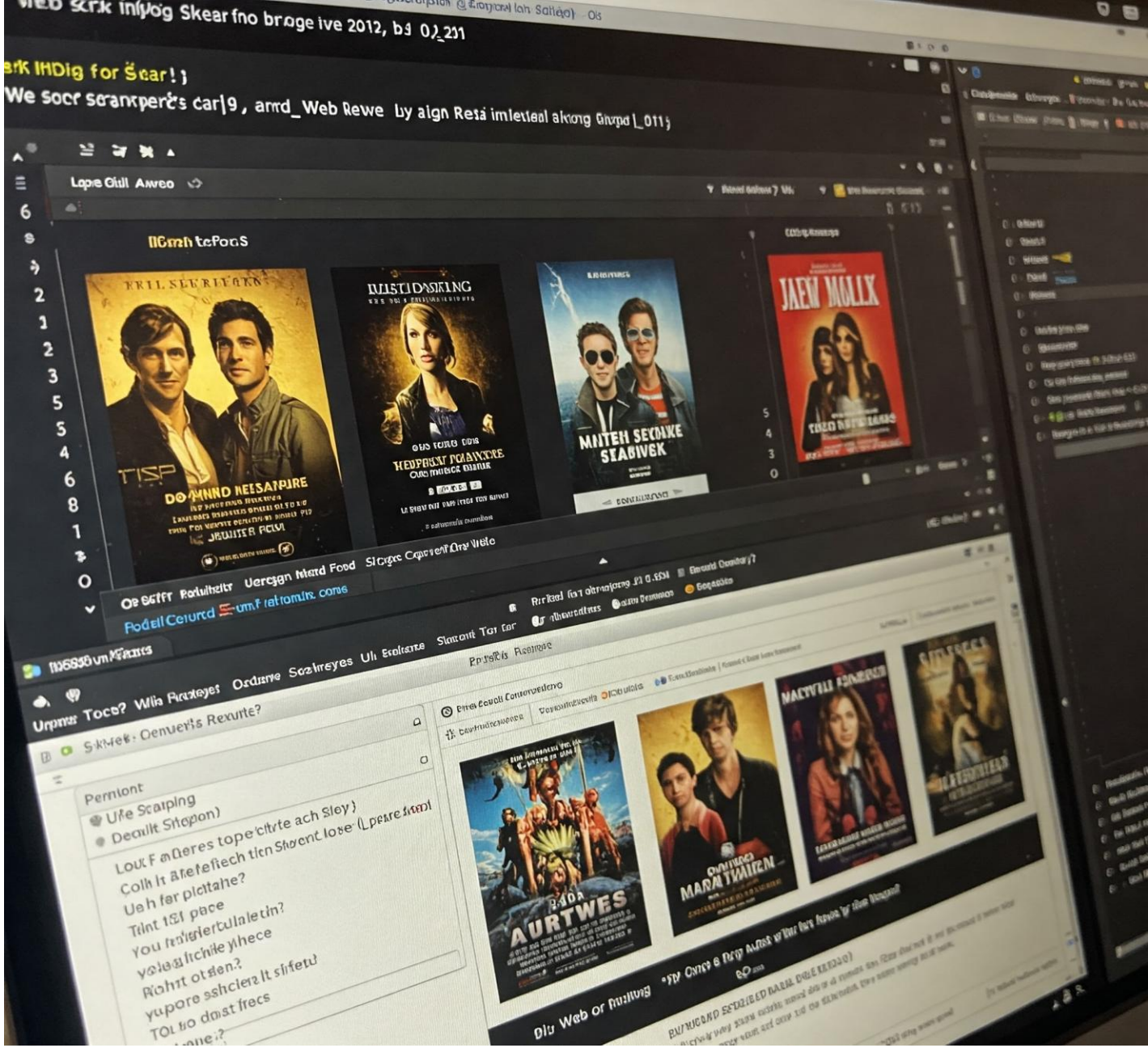
https://www.imdb.com/

- **TOOLS** :

Selenium automates web browsers, allowing for dynamic content scraping, while BeautifulSoup excels at parsing HTML. Together, these tools efficiently collect and process data from complex webpage structures.

- **Data Scope** :

Collected information on the top 500 movies based on IMDb ratings

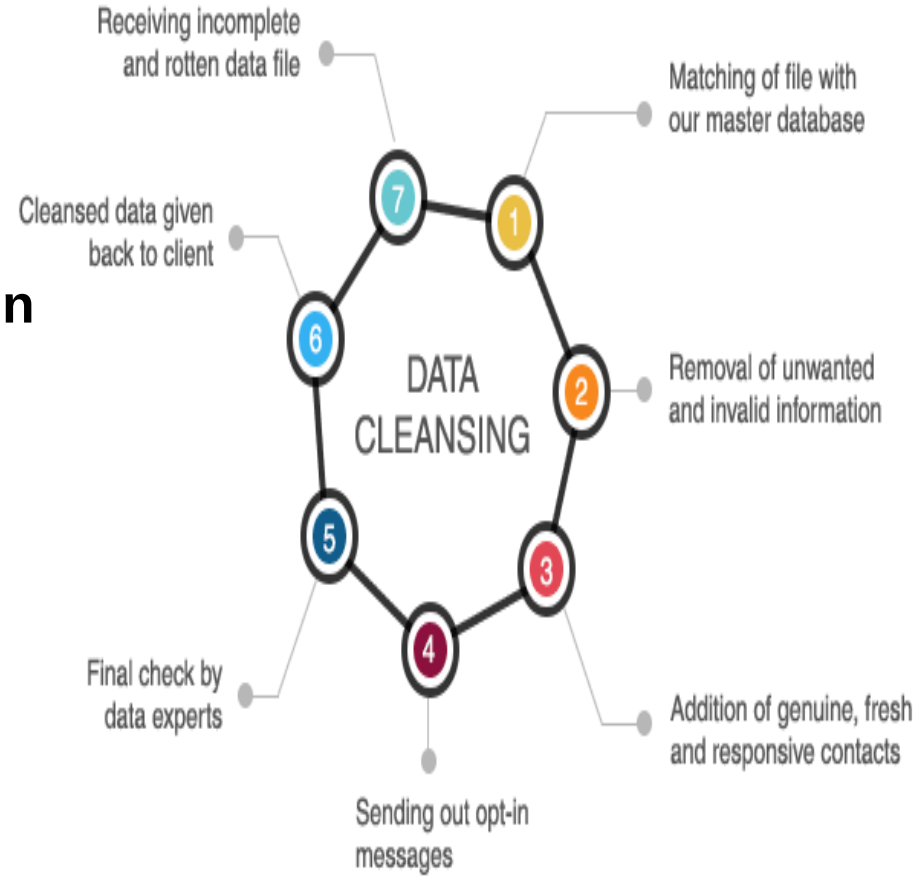


INNOMATICS RESEARCH LABS

Exploratory Data Analysis (EDA)

Data Cleaning:

- **Hours have been converted into minutes for simplified calculations and analysis**
- **To facilitate efficient calculations and analysis all figures in millions and thousands are converted into numerical values.**
- **Replaced null values with means .**



INNOMATICS RESEARCH LABS

DATA ANALYSIS:

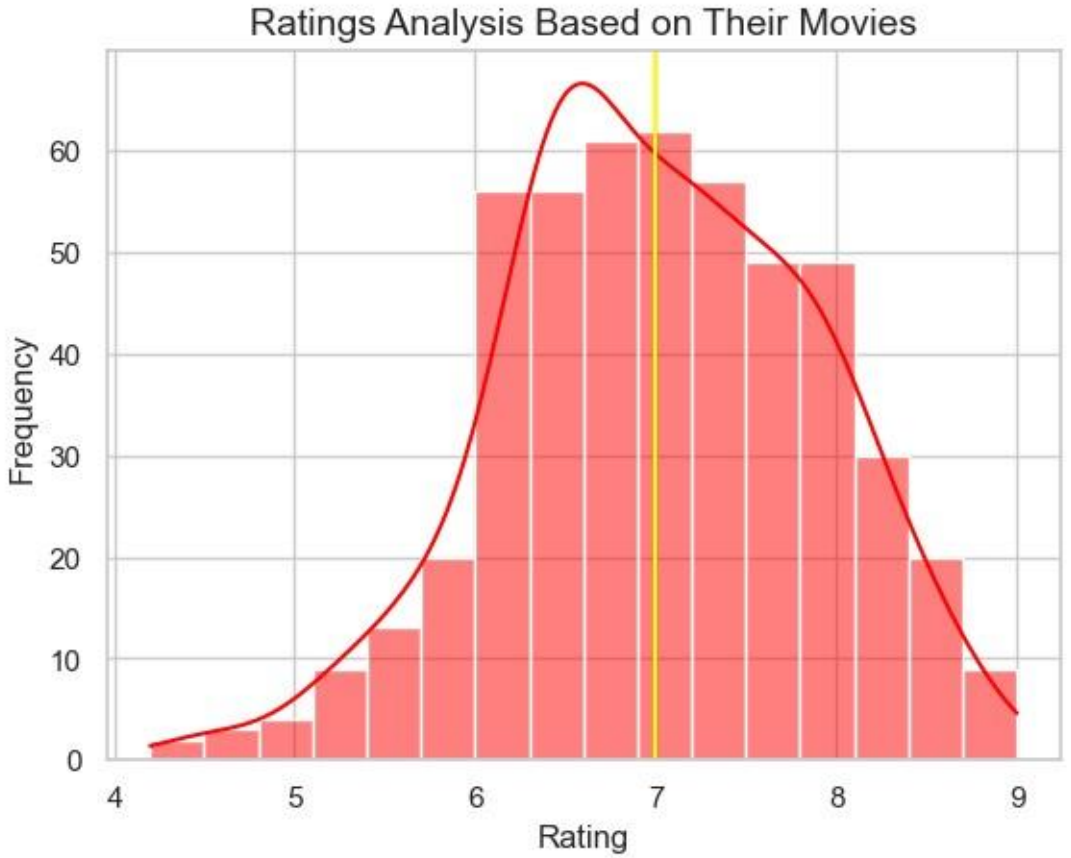
UNIVARIATE

BIVARIATE

MULTIVARIATE

UNIVARIATE ANALYSIS:

- Most movies have ratings between 6 and 8, with the highest frequency around 7, indicating generally favorable reviews.
- The distribution is left-skewed, with fewer low-rated movies, suggesting a positive bias in the ratings.
- The median (yellow line) is slightly lower than the mean (red line), reflecting the presence of some higher-rated movies pulling the mean upwards.



INNOMATICS RESEARCH LABS

➢ Plot 1: Density Plot

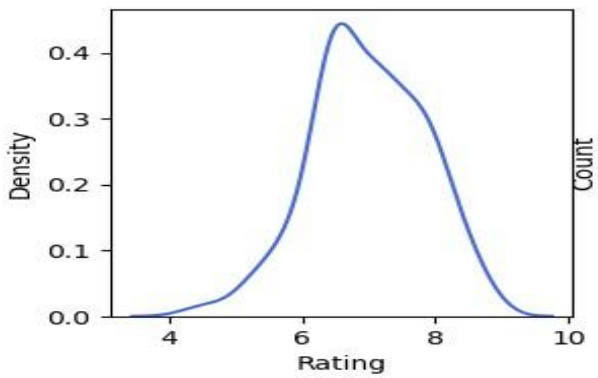
•**Shape:** The curve is unimodal (has one peak) and slightly skewed to the right. This means the majority of ratings cluster around a central value, but there's a tail extending towards higher ratings.
•**Central Tendency:** The peak suggests the most frequent rating (mode) is around 6.5-7.
•**Spread:** The curve's narrowness indicates ratings are clustered, not widely spread.
•**Interpretation:** Movies on this platform likely receive moderately positive ratings, with a slight lean towards higher scores.

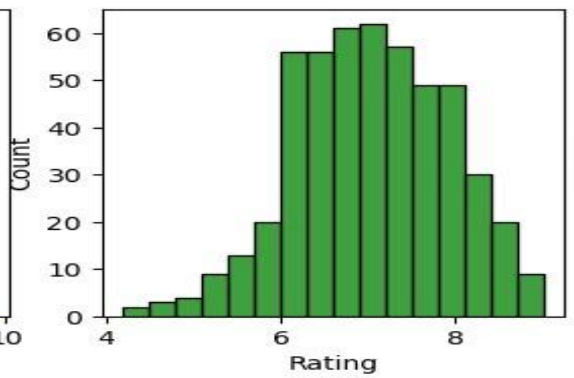
➢ Plot 2: Histogram

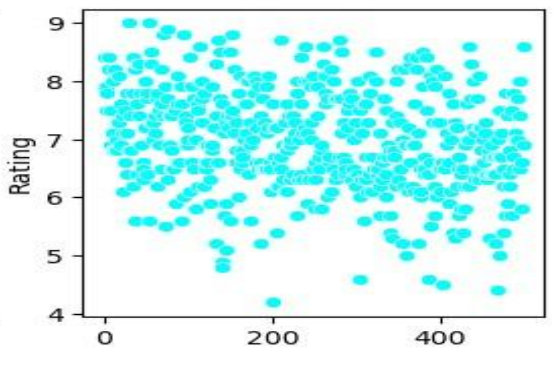
•**Data Representation:** Shows frequency of ratings within specific ranges (bins).
•**Symmetry/Skewness:** Similar to the density plot, it shows a slight positive skew.
•**Peak/Mode:** The highest bar confirms the mode around 6.5-7.
•**Frequency:** Provides the actual count of movies within each rating range.
•**Interpretation:** Reinforces the density plot's observations, giving a discrete view of the rating distribution.

➢ Plot 3: Scatter Plot

•**Relationship:** No clear pattern or trend, indicating no relationship between movie order and rating.
•**Spread:** Ratings are scattered but clustered in a horizontal band.
•**Interpretation:** Movie order doesn't influence ratings; they appear independently assigned.



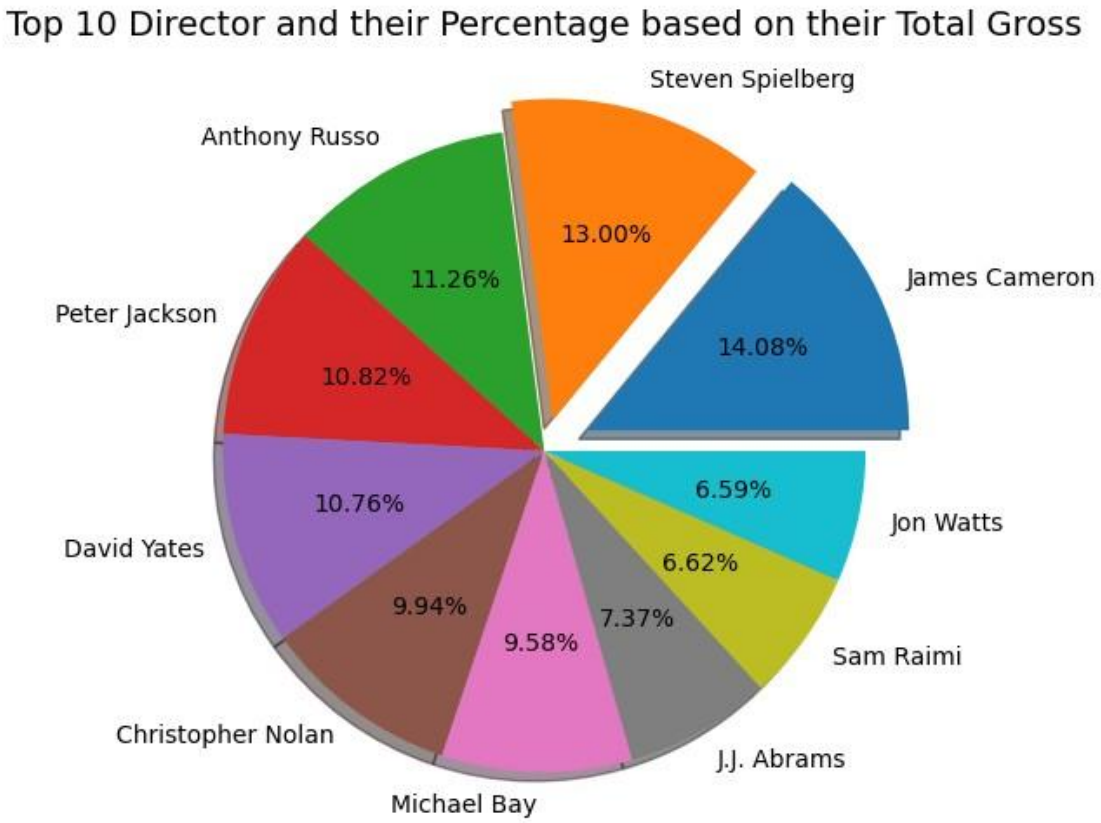




INNOMATICS RESEARCH LABS

BIVARIATE ANALYSIS:

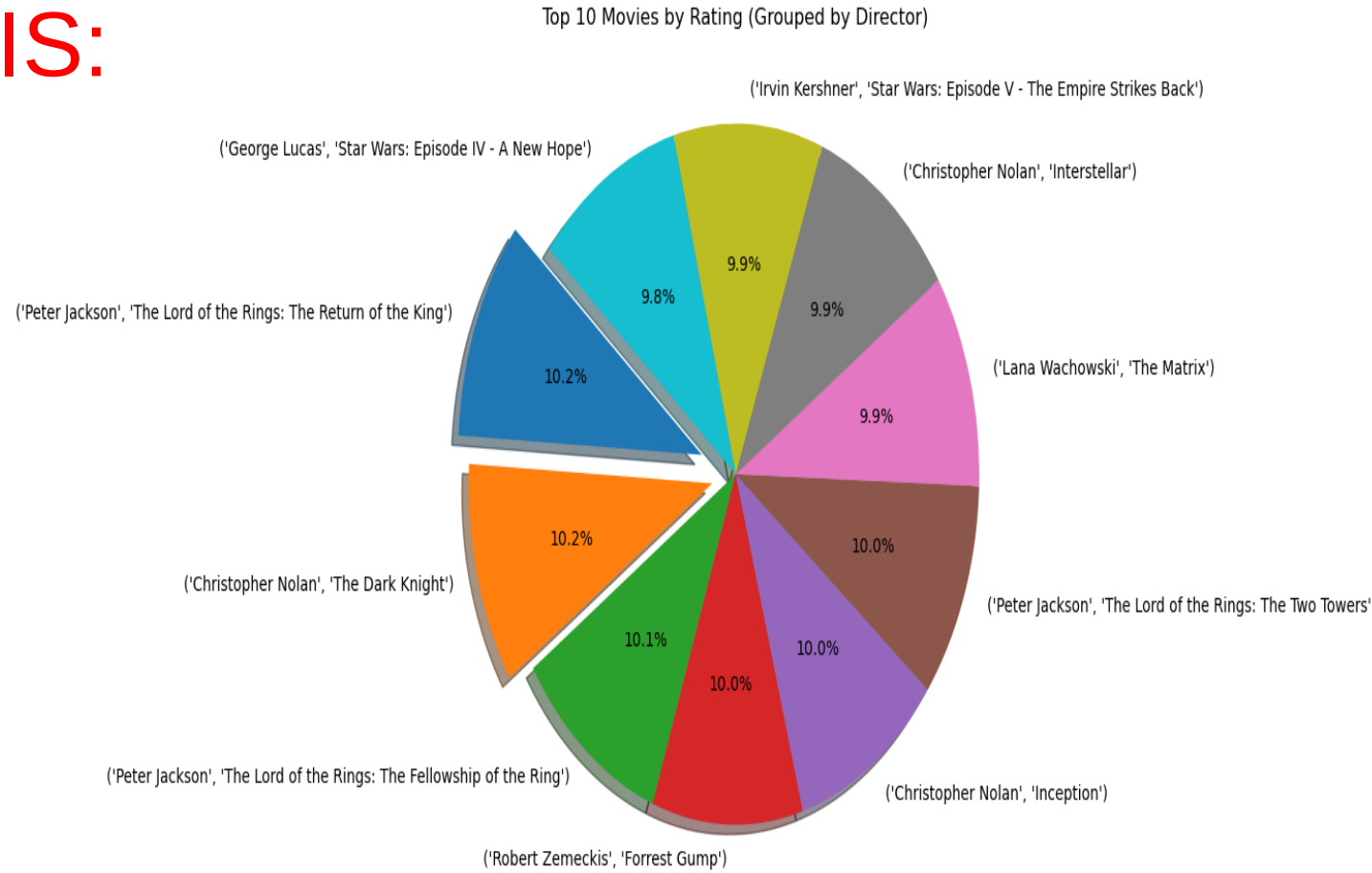
- **James Cameron leads with the highest total gross percentage, followed closely by Steven Spielberg.**
- **Anthony Russo, Peter Jackson, and David Yates also contribute significantly to the total gross, showcasing their box office dominance.**
- **Directors like Jon Watts and Sam Raimi have lower**
- **Percentages, but still rank within the top 10, reflecting notable commercial success.**



INNOMATICS RESEARCH LABS

MULTIVARIATE ANALYSIS:

- Peter Jackson and George Lucas lead with top-rated movies, mainly from the Lord of the Rings and Star Wars series.
- Christopher Nolan has multiple high-rated movies, reflecting his strong presence in the industry.
- Popular franchises and iconic directors dominate the highest-rated films, showing the audience's preference for established names



INNOMATICS RESEARCH LABS

Conclusion:

- This in-depth analysis of movie ratings offers strategic insights into audience preferences and
  Industry dynamics.

- Using advanced web scraping techniques with Selenium and BeautifulSoup,
  we systematically extracted and processed IMDb Top 500 movie data.

- The analysis examined key factors such as rating, release year, and runtime to identify trends influencing movie ratings.

- Understanding these rating patterns provides data-driven insights that can shape future content creation and marketing strategies.

- By leveraging these findings, filmmakers and production houses can optimize their approach to maximize audience engagement and commercial success.

THANK YOU

INNOMATICS
RESEARCH LABS