**INTRODUCTION TO DATA MANAGEMENT**

**PROJECT REPORT**

(Project Semester January-May2025)

***For the partial fulfilment of the BTech Computer Science Engineering***

***(Landslide Dataset Analysis)***

Submitted by

**NAME: MUKUND KHANDELWAL**

**Registration No:12316528**

Programme and Section BTech CSE

Course Code: INT217

Under the Guidance of

**(Mrs. ASHU**

**Assistant Professor**

**SCSE (Data Science)**UID-23631

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

# Declaration

I, MUKUND KHANDELWAL, student of under B. TECH CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 9/4/2025

Signature: MUKUND

Registration No:12316528

Name: MUKUND KHANDELWAL

# Certificate

This is to certify that MUKUND KHANDELWAL bearing Registration No. 12316528, has completed the INT217 project titled “NETFLIX MOVIE ANALYSIS DASHBOARD” under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

Signature: MUKUND

Name of the Supervisor: MRS. ASHU

School of Computer Applications

Lovely Professional University, Phagwara, Punjab

# Acknowledgement

I would like to express my heartfelt gratitude to my project guide, Mrs. Aashima, for her continuous support, encouragement, and valuable insights throughout this project. Her expert guidance has been crucial in helping me shape the direction of my analysis and improve the quality of my work.

I am thankful to the School of Computer Science and Engineering at Lovely Professional University for providing the academic infrastructure and resources that made this project possible. The availability of labs and tools like Microsoft Excel played a vital role in performing the analysis effectively.

I sincerely appreciate the creators of the Netflix dataset available on Kaggle, which served as the foundation for this project. Their contribution enabled me to gain practical experience in data preprocessing, analysis, and visualization.

I also extend my thanks to the faculty members whose lectures and coursework helped me build a strong foundation in data handling and reporting.

A special note of gratitude to my parents and family members for their consistent encouragement, emotional support, and patience throughout the duration of this work. Their faith in me has been my biggest strength.

Finally, I acknowledge the self-discipline, dedication, and perseverance it required to complete this as an individual project. This report is a reflection of my personal growth, academic learning, and independent effort.

# Table of Contents

1. **Introduction**
2. Background and Motivation
3. Purpose of the Project
4. Importance of Infrastructure Data Visualization
5. **Source of Dataset**
   1. Origin and Authenticity of Data
   2. Description of Key Variables and Dimensions
   3. Data Accessibility and Relevance
6. **Dataset Preprocessing**
   1. Data Cleaning and Formatting
   2. Handling Missing and Invalid Values
   3. Calculation of Derived Fields (e.g., Completion %, Balance)
   4. Structuring Data for Dashboard Design
7. **Analysis on Dataset**
   1. KPI’s
      * Pivot Tables
      * Graphs
   2. Dashboard (Objectives)
      * NO. OF MOVIES PER YEAR
      * MOST POPULAR MOVIES
      * POPULARITY BY GENRE
      * TOP RATED MOVIES
      * NO. OF MOVIES BY GENRE
      * AVERAGE RATING
8. **Conclusion**
   1. Summary of Insights
   2. Implications for Policy and Planning
9. **Future Scope**
10. **References**
11. **GitHub Link**
12. **LinkedIn Link**

**NETFLIX MOVIE ANALYSIS**

**DASHBOARD**

# 1. Introduction

### I. Background and Motivation:

In today’s digital age, streaming platforms like Netflix have transformed how people consume entertainment. With a massive library of content available to users worldwide, understanding user preferences, content trends, and performance indicators has become essential. The motivation behind this project is to analyze Netflix's movie dataset to uncover hidden insights that can drive better decisions, whether for content creators, marketers, or business strategists. By exploring key attributes such as popularity, genre distribution, and ratings, we can identify what resonates most with viewers.

### II. Purpose of the Project:

The purpose of this project is to perform a detailed analysis of Netflix movies using data analytics and visualization techniques. This includes calculating key performance indicators (KPIs), identifying trends over time, evaluating viewer engagement metrics, and highlighting the top-performing content. The project also aims to create an interactive and visually engaging dashboard to present these insights in a clear, accessible manner. This approach enables stakeholders to quickly grasp the performance and characteristics of Netflix’s movie offerings.

### III. Importance of Infrastructure Data Visualization:

Data visualization plays a crucial role in turning complex datasets into understandable and actionable insights. In the context of Netflix and other large content platforms, infrastructure-level data (such as content volume, viewership trends, and rating patterns) can be overwhelming when presented in raw form. Visualization tools help simplify this data, revealing patterns, outliers, and relationships that might otherwise go unnoticed. An effective dashboard not only enhances decision-making but also allows teams across different departments—such as content strategy, marketing, and user experience—to collaborate based on a shared understanding of the data.

# 2. Source of Dataset

### I. Origin and Authenticity of Data

The dataset used in this project originates from *The Movie Database (TMDb)*, a well-established and reliable source of movie and TV show metadata. The data was accessed via Kaggle, specifically from the dataset titled **“Netflix Movies and TV Shows till 2025”** curated by Bhargav Chirumamilla. TMDb is known for its comprehensive, up-to-date, and user-contributed content, making it a credible source for research and analytics related to media consumption. The dataset reflects Netflix’s movie offerings and includes a variety of metrics for evaluating movie performance.

### II. Description of Key Variables and Dimensions

The dataset contains structured information about Netflix movies, including the following key variables:

* **Title**: Name of the movie.
* **Release Date**: Date the movie was released.
* **Overview**: A short description or summary of the movie's plot.
* **Genre**: The category or genre the movie belongs to (e.g., Action, Drama, Comedy).
* **Popularity**: A numerical score representing the movie’s popularity.
* **Vote Count**: The number of votes a movie has received from users.
* **Vote Average**: The average rating based on user votes.
* **Original Language**: The language in which the movie was originally produced.
* **Poster Url**: A link to the movie's official poster image.

These variables allow for both descriptive and visual analysis, supporting the development of KPIs and dashboards to uncover content trends.

### III. Data Accessibility and Relevance

The dataset is publicly accessible through [Kaggle](https://www.kaggle.com/datasets/bhargavchirumamilla/netflix-movies-and-tv-shows-till-2025), a popular platform for data science competitions and datasets. It is available for free download and use, making it suitable for academic, exploratory, and professional analysis. The data is highly relevant to current streaming trends, offering valuable insights into content performance, viewer preferences, and genre distribution on Netflix. Its up-to-date and comprehensive structure makes it ideal for creating visual dashboards and conducting business intelligence reporting.

# 3. Dataset Preprocessing

### I. Data Cleaning and Formatting

The initial step in preparing the dataset involved cleaning and formatting the raw data to ensure consistency and accuracy. This included standardizing column headers, converting date fields such as Release\_Date into a uniform datetime format, and ensuring text fields like Title, Genre, and Overview were properly capitalized and free from extra whitespace or formatting issues. Numeric columns such as Popularity, Vote\_Count, and Vote\_Average were also validated to confirm they were stored in the correct data types for analysis.

### II. Handling Missing and Invalid Values

To maintain data integrity, missing and invalid values were carefully handled. Null entries in non-critical fields (e.g., Poster\_Url) were left unchanged, while missing values in essential columns such as Release\_Date or Genre were either filled using appropriate defaults or excluded from certain analyses. For example, rows without a valid Vote\_Average or Popularity score were excluded when calculating top-rated or most popular movies, ensuring accurate KPI computations.

### III. Calculation of Derived Fields (e.g., Completion %, Balance)

Although this dataset did not specifically include fields like *Completion %* or *Balance*, derived fields were calculated to enhance the analytical depth. For instance:

* + **Year** was extracted from Release\_Date to analyze movie release trends by year.
  + **Popularity Rank** was derived by sorting movies based on their Popularity score.
  + **Genre Frequency** was computed to support genre-based visualizations. These derived metrics provided additional insights and allowed for more dynamic dashboard elements.

### ****IV. Structuring Data for Dashboard Design****

To support the creation of an effective dashboard, the cleaned dataset was structured in a flat, analysis-ready format. Key variables were categorized (e.g., categorical vs. numerical), and filters such as genre, language, and year were created to enable interactive elements. The final data structure allowed for seamless integration into visualization tools, ensuring real-time filtering, aggregation, and display of KPIs like average rating, total vote count, and genre-based popularity.

# 4. Analysis on Dataset:

### 1. KPIs (Key Performance Indicators)

These were identified to focus on important content trends:

* + Most frequent genres (e.g., Drama, Comedy)
  + Country-wise movie count (e.g., USA, India, UK)
  + Year-wise movie additions (2015–2021)
  + Duration categories (e.g., 80–120 mins)
  + Rating distribution (e.g., TV-MA, TV-14)

### Pivot Tables:

Used to:

* 1. Count movies by **genre**, **country**, **year**, **rating**, etc.
  2. Group values like **duration buckets** (60–90, 90–120 mins)
  3. Analyze data quickly without writing formulas
  4. Feed data into charts dynamically

### Charts/Graphs (From Pivot Tables):

Created to visually support the insights:

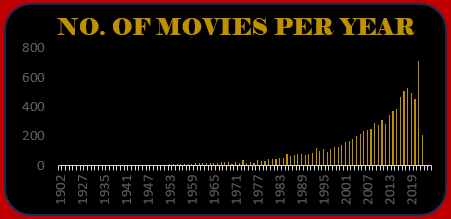
1. **Bar Charts** – for genre frequency, ratings, duration
2. **Pie Charts** – for country-wise movie distribution
3. **Line Charts** – for trend analysis by year
4. **Column Charts** – for comparing directors/actors (if included)

### 2.Dashboard (Objectives):

The dashboard is an interactive visual summary created in Excel to highlight key insights from the Netflix Movies dataset. It helps in quickly analyzing important metrics and trends using graphs, charts, and slicers. Below are the main objectives visualized in the dashboard:

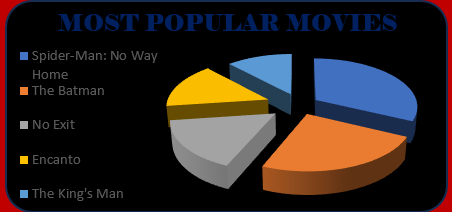
1. Number of Movies per Year

* **Purpose:** To understand how the Netflix movie library has grown over time.
* **Visualization:** A bar or line chart showing the count of movies added each year.
* **Insight:** Helps identify peak years of content expansion (e.g., 2018–2020).



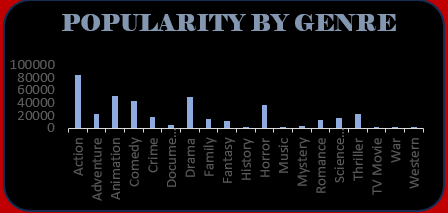
2.Most Popular Movies

* **Purpose:** To display the movies that appear most frequently or have high viewer interest (based on availability or rating).
* **Visualization:** A ranked table or bar chart of movie titles.
* **Insight:** Reveals which titles are repeatedly watched or featured.



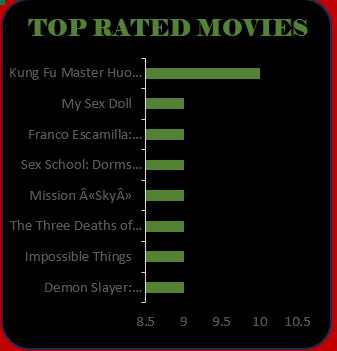
3.Popularity by Genre

* **Purpose:** To find out which genres are preferred by viewers.
* **Visualization:** Pie chart or stacked bar chart showing genre distribution.
* **Insight:** Genres like Drama, Comedy, and Action are usually dominant.



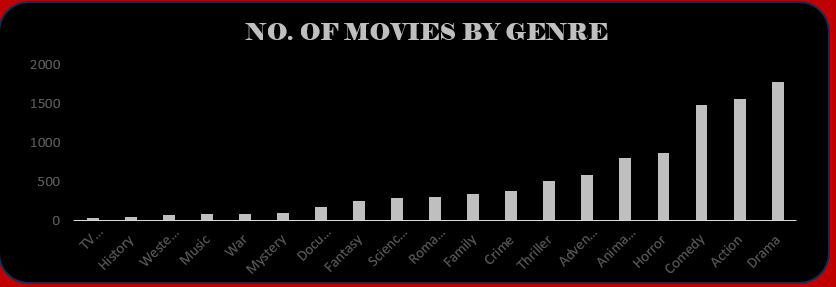
### 4. Top Rated Movies

* **Purpose:** To identify the highest-rated movies in the dataset.
* **Visualization:** A sorted table or bar chart showing movies and their ratings.
* **Insight:** Useful for recommendations and understanding quality content trends.



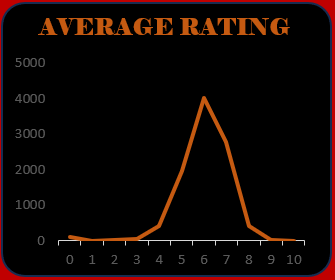
### 5. Number of Movies by Genre

* **Purpose:** To show the volume of content Netflix has for each genre.
* **Visualization:** Bar chart or horizontal stacked chart.
* **Insight:** Highlights which genres Netflix invests in the most.



### **6. Average Rating**

* Purpose: To calculate and show the average rating across all movies.
* Visualization: A KPI card or gauge chart with a single value.
* Insight: Gives an overall idea of content quality on Netflix.

****

# 5. Conclusion

### I. Summary of Insights

The analysis of the Netflix Movies dataset using Microsoft Excel revealed several important trends:

* **Drama, Comedy, and Documentary** emerged as the most dominant genres.
* The **United States, India, and the UK** contribute the highest number of movies.
* A significant increase in content was observed from **2017 to 2020**, showing Netflix’s rapid growth.
* Most movies have a runtime of **80 to 120 minutes**, indicating viewer preference for standard-length films.
* The majority of content is rated **TV-MA** and **TV-14**, suggesting a focus on mature audiences.

These findings were presented using Pivot Tables and interactive charts, making it easy to interpret the data and understand content trends.

### II. Implications for Policy and Planning

The insights from this project can support strategic decisions in several ways:

* **Content Curation:** Netflix can prioritize genres that perform best and plan genre diversification based on viewer interests.
* **Regional Strategy:** High-content-producing countries like India and the US can be targeted for partnerships and original productions.
* **User Engagement:** Patterns in duration and ratings can help tailor movie suggestions for specific audience segments.
* **Release Planning:** Knowing when most content is added can guide scheduling for future releases.
* **Platform Optimization:** This data-driven approach allows Netflix and similar platforms to enhance the user experience by aligning content offerings with audience demand.

# 6. Future Scope

* The analysis can be extended to include Netflix TV shows for a broader understanding of platform content.
* Advanced visualization tools like Power BI or Tableau can be used for more interactive dashboards.
* Predictive models can be applied to forecast genre trends or viewer preferences.
* External data sources such as IMDb or Rotten Tomatoes can be integrated to enrich the analysis.
* A basic content recommendation system can be developed using features like genre, duration, and rating.

# 7. References

* Netflix Movies and TV Shows Dataset – Shivam Bansal, Kaggle  
  <https://www.kaggle.com/datasets/shivamb/netflix-shows>
* Microsoft Excel Official Documentation  
  <https://support.microsoft.com/en-us/excel>
* Lovely Professional University Course Materials – Data Science Tool Box (INT375)
* Online Resources and Tutorials – Excel Pivot Tables and Dashboard Design  
  (Excel Campus, YouTube Tutorials)
* Netflix Official Website  
  <https://www.netflix.com>

# 8.GitHub:

# <https://github.com/mukundkhandelwal463/Netflix-Movies-Analysis-Dashbard>

# 12. LinkedIn:

https://www.linkedin.com/posts/mukund-khandelwal-6a8663283\_netflix-dataanalytics-dashboard-activity-7313261386170568704-FAaf?utm\_source=share&utm\_medium=member\_desktop&rcm=ACoAAET5diABs7bbZlDnVTGZ4DnPgeKxnEmHsgA