**POWER BI PROJECT REPORT**

Session 2024-25

on

**Movie Sales Analysis on Amazon**



**Team Members:**

Himanshi Gautam – 2100680100156

Aayushi Chaudhary – 2100680130002

**Submitted To:**

Mr Madhvendra Singh Sir

Department of Computer Science and Engineering

Meerut Institute of Engineering & Technology, Meerut- 250005, India

**Project Overview:**

* This project analyzes movie data from Amazon Dataset to better understand what types of movies have been the most successful in the last decade, both from a profitability and popularity perspective. Analysis of recent historical budget, revenue, genre and review data can help Microsoft determine how to best strategize their impending entrance into the movie market. [Amazon Prime Movies and TV Shows (kaggle.com)](https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows)

**Data Cleaning:**

Data cleaning was a crucial step in ensuring the accuracy and usability of the analysis. The following steps were taken:

1. **Handling Missing Values:** Missing data, particularly in fields like subsidiary of in the accounts dataset, were managed by either filling in the values where possible or excluding the irrelevant columns from analysis.
2. **Data Type Conversion:** Numeric fields such as revenue, sales price, and close value were converted to appropriate numerical data types to facilitate calculations.
3. **Duplicate Removal:** Duplicate records in datasets like movie sales were removed to avoid skewed analysis.

These steps ensured that the data was clean, consistent, and ready for analysis.

**Data Modeling:**

The data modeling involved creating relationships between the various datasets to support meaningful analysis. The primary and foreign keys (PK and FK) and relationships were defined as follows:

* **AMAZON Table:** This table contains cast, directors, duration, ratings, title and release year of the movies in IMDB platform.
* **Cast:** This field lists the actors involved in Amazon's media content, but in sales analysis, it can refer to product brands or featured creators.
* **Country:** Indicates the location or region, useful for segmenting data and analyzing sales trends by country or region.
* **Date Added:** The date when a product or content was listed on the platform, useful for tracking product lifecycle and analyzing sales growth over time.
* **Description:** A short summary of the product or content. This can help classify product categories for further analysis.
* **Director:** For media content, this refers to the director, but for sales, it can be adapted to signify a product's key influencers or stakeholders.
* **Duration**: In sales, this could represent the lifespan of a product in the marketplace, like time in stock or the duration of a sales campaign.
* **Listed In:** Categories or departments where products are listed, essential for categorizing products for analysis in Power BI.
* **Rating:** The product or content's customer rating, vital for analyzing customer satisfaction and sales performance.
* **∑ Release Year:** The year a product or content was released, important for cohort analysis, product life-cycle tracking, or understanding yearly sales trends.
* **Show ID:** A unique identifier for content or products, necessary for tracking individual sales or performance.
* **Title:** The name of the product or content, useful for identifying and grouping products during analysis.
* **Type:** Refers to the nature of the product (e.g., physical goods, digital media) and is important for segmenting sales data.

These relationships enabled a cohesive model that facilitated the creation of insightful visualizations.

**Visualizations:**

The visualizations created in this Power BI project include:

**Card:** Displays single, critical metrics like total movie sales, customer count, or average revenue per sale. It's useful for highlighting key performance indicators (KPIs) at a glance.

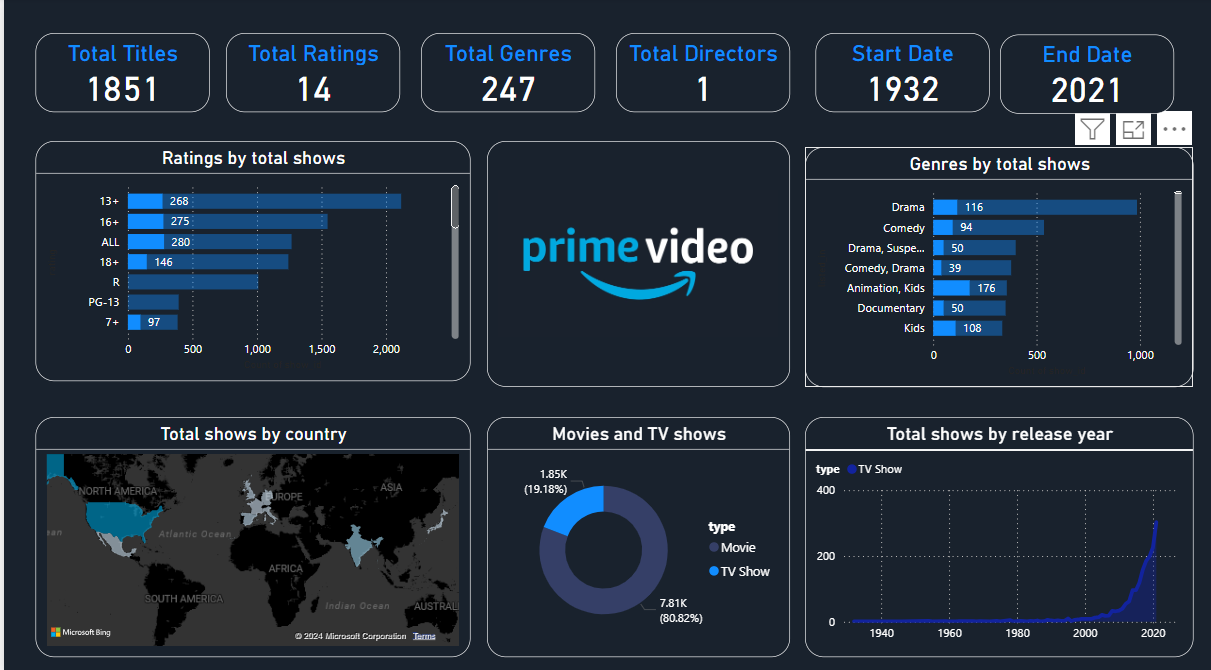
**Bar Graph:** Useful for comparing movie sales by category, genre, country, or month. This helps identify top-performing genres or regions based on sales volume.

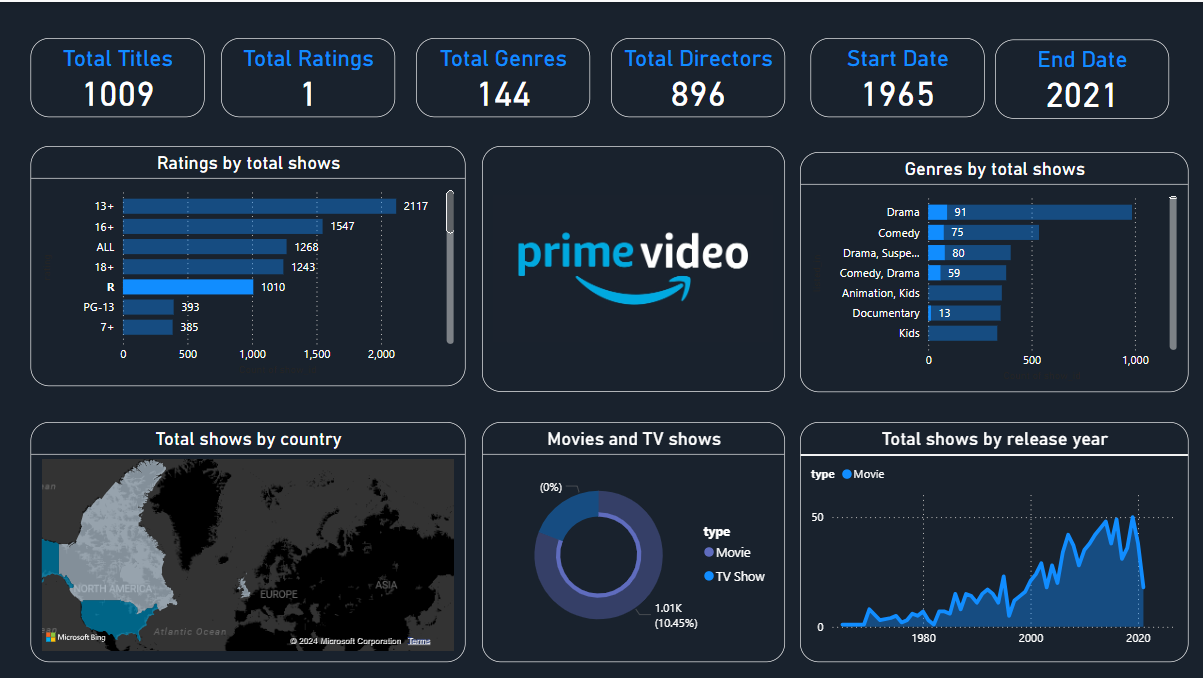
**Filled Map:** Displays sales data geographically, showing which countries or regions generate the highest movie sales. This visual is ideal for location-based trend analysis.

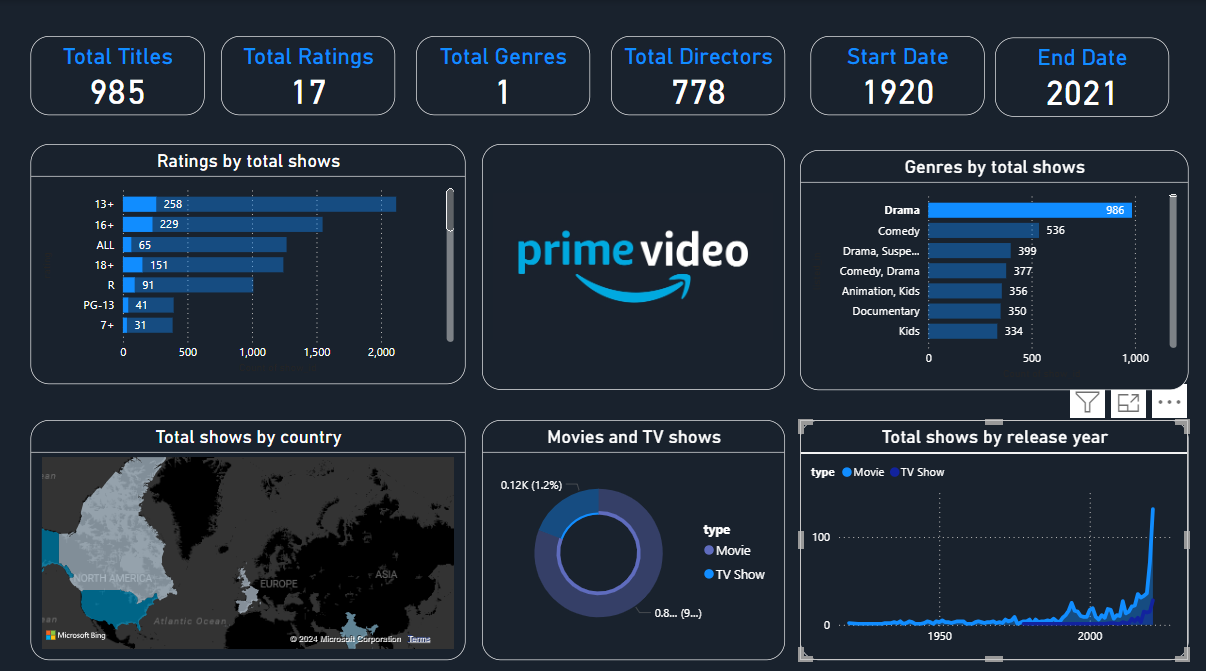
**Donut Chart:** Shows the proportion of sales across various genres, release years, or platforms. It helps visualize the distribution of sales within different categories.

**Area Chart:** Useful for tracking movie sales trends over time, such as daily, monthly, or yearly sales. It emphasizes cumulative sales growth and helps identify seasonal patterns.

These visuals provide a comprehensive overview of Amazon movie sales and help identify key insights.





****

**Conclusions and Findings:**

The analysis yielded several key insights:

1. **Sales Trends:** Identified consistent patterns in movie sales over time, with peaks during summer and holiday seasons.
2. **Genre Performance:** Certain genres (e.g., action, adventure) consistently outperformed others in terms of box office revenue.
3. **Star Power:** Movies featuring popular actors and directors showed higher sales figures, indicating the influence of star power on box office success.
4. **Top Grossing Movies:** A list of the highest-grossing films revealed key insights into audience preferences.
5. **Demographic Insights:** Analysis of viewer demographics highlighted specific age groups that favored particular genres.

These findings provide actionable insights that can guide future sales strategies and resource allocation.