AMAZON PRIME TITLES
DATA ANALYTICS USING POWER BI

A PROJECT REPORT

Submitted by

RAJA S   (2116250711013)

MASTER OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

Department of Computer Science and Engineering,

Rajalakshmi Engineering College, Thandalam

BONAFIDE CERTIFICATE

This is to certify that Mr ..**RAJA S**………….. of " **Computer Science and Engineering"** at Rajalakshmi Engineering college, Thandalam had undergone his training program on Data Analytics Using Power BI in our company premises from 18th November 2025.. to ,28th November 2025

Authorized signatory

TABLE OF CONTENTS

CHAPTER 1

ABSTRACT

The Amazon Prime Titles Dataset Analysis Project focuses on understanding, analyzing, and visualizing the catalog of movies and TV shows available on Amazon Prime Video. With the rapid global expansion of OTT (Over-The-Top) streaming platforms, Amazon Prime has emerged as a major competitor in digital entertainment. Analyzing its content library helps uncover trends in genres, release patterns, countries of production, and audience-oriented content strategies.

This project uses a real-world dataset containing information such as title, director, cast, release year, country, rating, duration, date added, and genre classification. Using Microsoft Excel, Python, and Microsoft Power BI, the project applies a structured workflow to clean, analyze, and visualize Prime Video content.

Through data analytics, the study provides insights into platform tendencies, genre distribution, regional content diversity, release-year clustering, and maturity rating patterns. These observations help content analysts, streaming strategists, and media researchers understand platform characteristics and global streaming preferences.

CHAPTER 2
PROBLEM STATEMENT

OTT platforms host massive content libraries, making it difficult to identify content trends, regional variations, and streaming catalog patterns. Although individual platforms maintain large datasets, integrated analysis of catalog composition is often missing.

Key challenges include:

- Understanding genre distribution and global content representation.
- Identifying content diversity across countries and languages.
- Analyzing historical release trends and maturity ratings.
- Determining how often Amazon adds content and from which regions.

The primary question addressed in this project is:

"How can data analytics and visualization tools be used to understand the structure, trends, and characteristics of Amazon Prime Video's content library?"

By analyzing the dataset through an integrated approach using Excel, Python, and Power BI, the project uncovers patterns in content addition, global representation, genre strengths, and audience- oriented classifications.

CHAPTER 3

OBJECTIVES

- To perform a comprehensive analysis of the Amazon Prime Video Titles dataset.

- To clean and prepare the dataset using Microsoft Excel for error-free visualization.

- To use Python for advanced statistical and exploratory analysis of Amazon Prime content.

- To design an interactive Power BI dashboard to visualize content categories, release patterns, and genre distribution.

- To identify top genres, frequently represented countries, and popular maturity ratings.

- To analyze content release trends across decades.

- To understand how Amazon structures its global catalog across movies and TV shows.

- To promote analytics-based decision-making in the entertainment and OTT ecosystem.

CHAPTER 4
SOFTWARE DESCRIPTION AND PURPOSE

The project uses multiple software tools to preprocess, analyze, and visualize the Amazon Prime dataset:

1.   Microsoft Excel

Used for initial cleaning—removing duplicates, fixing missing values, formatting columns, splitting duration components, and ensuring uniform structuring of fields such as genre, cast, and director.

2.   Python (via PyCharm)

Python performs deeper analysis using libraries such as pandas, matplotlib, and seaborn. It is used for:

- Descriptive statistics
- Genre frequency analysis
- Country-based distribution
- Release year clustering
- Maturity rating breakdown

3.   Microsoft Power BI

Power BI converts the cleaned dataset into dynamic dashboards that display:

- Genre distribution
- Release year timelines
- Movies vs TV Shows split
- Country-based content share
- Rating categories
- Date-added trends

4.   Amazon Prime Titles Dataset

The dataset includes fields such as show_id, type, title, director, cast, country, release_year, rating, duration, listed_in, and description.

CHAPTER 5

WORKING PRINCIPLE

The project follows a structured data analytics pipeline:

Stage 1: Data Cleaning (Excel)

- Import dataset into Excel

- Remove duplicates and empty rows

- Standardize column formats

- Convert date_added into datetime

- Separate duration into numeric value + unit

- Format genre and country fields

Stage 2: Data Analysis (Python)

- Load cleaned dataset using pandas

- Analyze:

  o Genre occurrence

  o Country contribution

  o Release year patterns

  o Ratings and maturity distribution

  o Movie vs TV Show composition

- Create plots such as bar charts, histograms, and year-wise trends

- Export final dataset for BI visualization

Stage 3: Visualization (Power BI)

- Import processed data

- Build dashboards showing:

  - Genre spread

  - Geographic representation

  - Timeline charts for release years

  - Ratings distribution

  - Movies vs TV Shows ratio

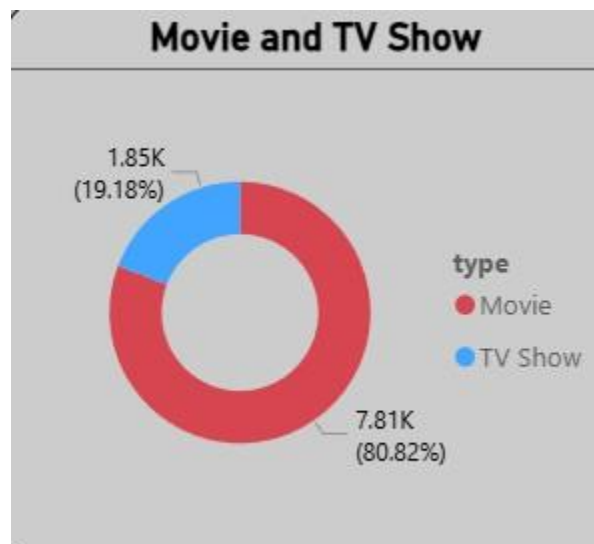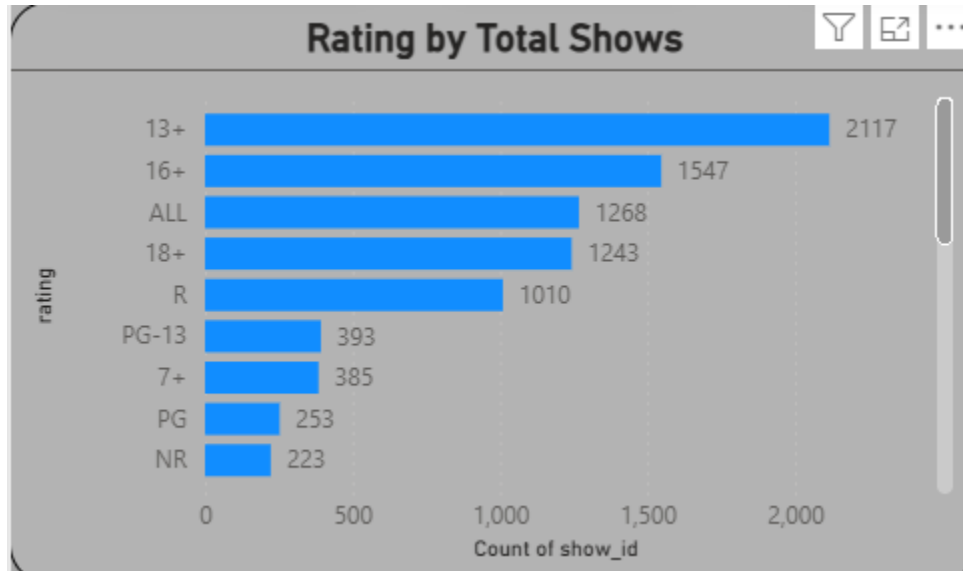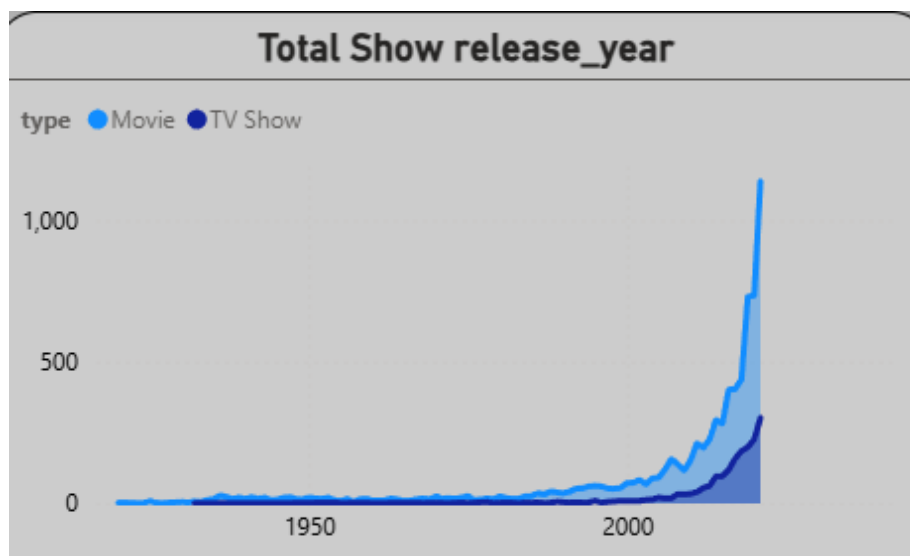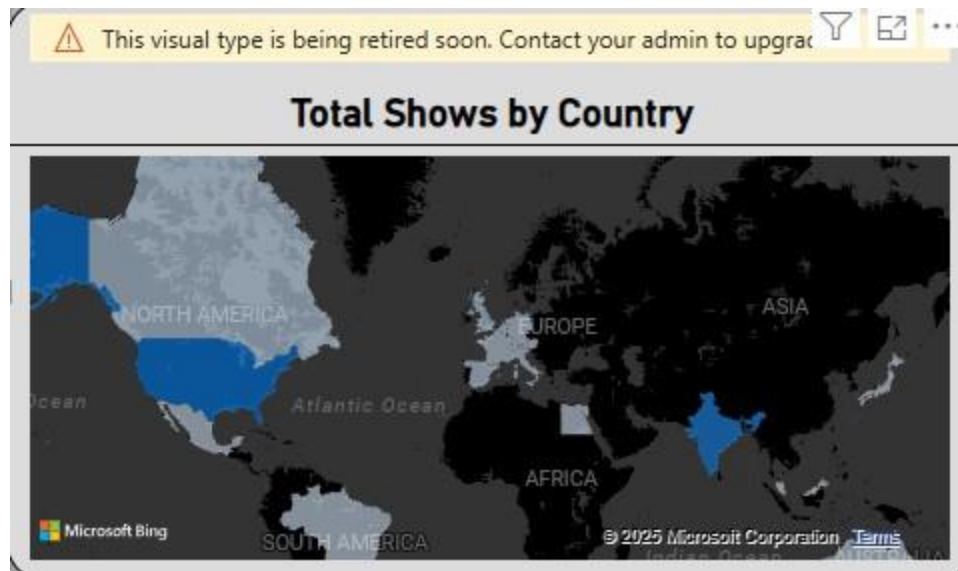- Use slicers for filtering by country, year, type, and rating

CHAPTER 6

EXCEL GRAPH IMPLEMENTATION

CHAPTER 7

POWER BI GRAPH IMPLEMENTATION

Total Shows by Country

Total Show release_year

CHAPTER 8

APPLICATIONS, ADVANTAGES AND LIMITATIONS

8.1 Applications

- **OTT Platforms:** Understand catalog distribution and content strengths.

- **Media Analysts:** Identify global content representation and genre patterns.

- **Marketing Teams:** Study what type of content is released more often.

- **Recommendation Systems:** Use insights for content personalization.

- **Students & Researchers:** Learn real-world analytics on media datasets
  .

8.2 Advantages

- Covers multiple fields such as title, director, cast, country, and rating.

- Supports interactive data exploration through dashboards.

- Uses a clean, scalable analytics workflow combining Excel, Python, and Power BI.

- Helps understand global streaming trends and catalog structure.

8.3 Limitations

- Dataset does not include user ratings or watch statistics.

- Missing values for cast/director in many entries.

- Does not show content availability changes over time.

- Regional streaming rights may differ from dataset representation.

CHAPTER 9
FUTURE SCOPE, CONCLUSION, AND REFERENCES

9.1 Future Scope

- ❖ Integration with IMDb ratings for popularity analysis.
- ❖ Machine learning models for content recommendation based on dataset. Automated OTT content monitoring pipelines for real-time analysis.
- ❖ Cloud-hosted dashboards for global access.
- ❖ Combining Prime Video data with Netflix/Disney+ for comparative insights.

9.2 Conclusion

This project demonstrates how modern analytics tools—Excel, Python, and Power BI—can be applied to analyze streaming platform datasets such as Amazon Prime Titles. The study highlights genre preferences, maturity rating patterns, country-wise distribution, and release-year trends. Through interactive visualization, the project offers a clearer understanding of Amazon Prime Video's content strategy and global catalog distribution.

9.3 References

- Amazon Prime Titles Dataset

- Microsoft Power BI Documentation

- Python Pandas & Matplotlib Libraries Documentation