Business Understanding. C6

Background

My previous studies were in the field of media, which is probably why I have picked this topic among many others suggested on Kaggle.

Business Goals

The primary goal of this project is not tied to a specific business or client but rather to find highlights that could be useful for media researchers. The analysis aims to:

1. Determine the most successful periods for different genres over time, called "golden ages”.
2. Figure out if modern-day most popular genres are also the ones most produced
3. Study combinations between genres

Potential findings could be explained via technological or cultural events, for example, the shift from cable television to internet streaming.

Business Success Criteria

The project will be considered successful if:

- It is possible to identify "golden ages" of TV genres.
- Put together insightful plots.
- Determined the most popular genre combinations

Inventory of Resources

A publicly available Kaggle dataset containing over 10,000 TV shows.

The TMDb (The Movie Database) API is used to retrieve official genre names.

Python with Pandas, Numpy, and matplotlib.

Jupyter Notebook for analysis and visualization.

Requirements, Assumptions, and Constraints

The dataset is assumed to be accurate and representative of the population. On Kaggle, it is ranked as: Completeness – 100%, Credibility – 100%, Compatibility – 100%

The project is constrained by:

1. Course deadline (Autumn 2025/26),
2. Public dataset limitations
3. API request restrictions.

Data license: CC BY-SA 4.0.

API token usage is for educational purposes only

Risks and Contingencies

Missing, incomplete, or incorrect data may distort results.

Tiny sample sizes for specific categories can lead to incomplete results.

Terminology

Golden Age – A time period during which a genre of shows was both produced at their own average during other periods, and shows during this period received higher voting ratings

Genre Combination – A set of different genres assigned to one TV show.

Weighted Rating – A Bayesian-inspired adjustment of ratings based on small vote count.

Costs and Benefits

- No financial costs
- Probably no actual benefits

Data-mining goals

- Analysis of TV show production counts and audience ratings by genre across years
- Calculate the weighted rating for each show, rather than the one provided by the dataset.
- Detect statistically significant peaks of "golden ages.”
- Analyse frequency of genre combinations
- Compare the popularity and production volume for the shows of the past two years.

Data-Mining Success Criteria

- Golden Age Detection: Golden age periods are successfully identified for genres that have at least twenty entries in the dataset
- Weighted Rating Model: The Bayesian-weighted rating formula is implemented, and the calculated score is used for further analysis
- The top 5 most frequent genre combinations are determined.
- Generated a ranked list of TV show genres' popularity of the past two years and volume of genres produced in the same period. These lists are compared.