Netflix Content Inferential Insights

Understanding Content Popularity through Genre and IMDb ratings

DSAI3301 Section 3/4
Data Analysis & Visualization Project

Tehreem Masroor, 60302531

Presentation Outline

1. INTRODUCTION
2. DATA ACQUISITION AND PREPARATION
3. PYTHON VISUALIZATION DEMO
4. POWER BI DASHBOARD DEMO
5. CONCLUSION

Introduction

- **Project Goal:** Use inferential statistics to discover how genre and IMDb ratings affect a title's popularity on Netflix.
- **Scope:** Combined data from Netflix Top 10 and OMDb API to analyze genre-based trends.
- **Why It Matters:** Helps Netflix make data-driven content investment decisions.
    - Netflix spends over $16 billion annually on content creation. Some shows (e.g., Stranger Things or Wednesday) dominate the Top 10 for weeks.
    - Others disappear after one week despite being promoted heavily.

Business Problem and Objectives

Problem Statement:

- Identify which genres drive long-term popularity (measured by cumulative weeks in Netflix Top 10).
- Determine whether certain genres consistently receive higher IMDb ratings from viewers.

Objectives:

- Discover if genre affects popularity (Chi-Square)
- Discover if IMDb ratings vary by genre (ANOVA)
- Quantify the impact of IMDb rating and genre tags on a show's overall popularity (Regression)
- Provide actionable recommendations to Netflix

Dataset Acquisition

NETFLIX TOP 10 GLOBAL ALL TIME TITLES

- https://www.netflix.com/tudum/top10/most-popular
- **Week-wise Netflix data:** 2021 to 2025
- **Size:** 8601 rows with weekly ranking and duplicated titles, with 2637 unique titles
- **Key columns:**
  1. **week:** Date the show was in the Top 10
  2. **weekly_rank:** rank of the movie in that week
  3. **cumulative_weeks_in_top_10:** count of how many times a show has been ranked in top 10.
  4. **show_title:** name of the title

OPEN MOVIES DATABASE (OMDB) API

- https://www.omdbapi.com/
- Fetched data from the OMDb API over the course of three days.
- **Title-wise Netflix data:** fetched data based on the unique title in Netflix dataset
- **Size:** 2635 rows of unique titles metadata, the 927 titles with no metadata were hand-filled , meticulously
- **Key columns:**
  1. **genre:** content type of the show title
  2. **origin_country:** country that the show originated from
  3. **release_year :** When the show or movie was released
  4. **imdb_rating:** Audience rating of the show

Data Preparation & Preprocessing

- Data collected using the OMDb API and Netflix Global All-time Dataset were merged on the title.
- Duplicate titles across weeks were grouped and aggregated to form a clean title-level dataset.
- Aggregated fields include:
  - **max_popularity**: Max value of cumulative_weeks_in_top_10
  - **primary_genre**: first genre list
- Exploded the genre column into binary features (e.g, tag_Comedy, tag_Action)
- Popularity Band (Low, Medium, High) based on cumulative weeks
- Converted week to datetime

Python Visualization Demo

Power BI Dashboard Demo

Conclusion

- **Genres Drive Popularity**
  - Chi-Square test shows certain genres consistently perform better in Top 10 rankings.
- **Ratings can vary by Genre**
  - ANOVA reveals audience IMDb ratings vary significantly across genres.
- **Longevity Can Be Predicted**
  - Regression shows genre tags + IMDb ratings can help predict how long a title stays popular.
- **Business Outcome:** Netflix can optimize content strategy by focusing on high-performing genres, audience-preferred themes, and predictive insights to maximize long-term popularity.

Thankyou

Any Questions ?