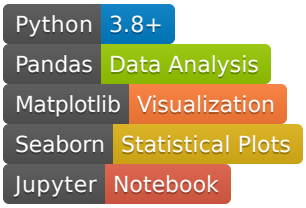
🎬 Netflix Movie Data Analysis Dashboard



📊 Project Overview

A comprehensive data analysis project that explores Netflix movie trends, ratings, and genre distributions using advanced Python data science techniques. This project demonstrates proficiency in data manipulation, statistical analysis, and interactive visualization.

🚀 Key Features

🔍 Advanced Analytics

- **Statistical Analysis**: Comprehensive descriptive statistics and data distribution analysis
- **Feature Engineering**: Custom rating categorization (Popular, Average, Below Average)
- **Time Series Analysis**: Release date patterns and temporal trends
- **Genre Analysis**: Multi-genre movie classification and popularity metrics

📈 Data Processing Excellence

- **Data Cleaning**: Robust preprocessing pipeline with error handling
- **Data Transformation**: Genre explosion for granular analysis
- **Feature Creation**: Dynamic rating classification based on vote averages
- **Date Processing**: Intelligent date parsing and year extraction

🎨 Interactive Visualizations

- **Distribution Plots**: Vote count, popularity, and rating distributions
- **Correlation Analysis**: Heatmaps showing feature relationships
- **Genre Insights**: Bar charts and pie charts for genre popularity
- **Trend Analysis**: Time-based visualizations for release patterns

Technical Stack

Technology	Purpose	Proficiency Level
Python	Core Programming	Advanced
Pandas	Data Manipulation	Expert
NumPy	Numerical Computing	Advanced
Matplotlib	Data Visualization	Advanced
Seaborn	Statistical Visualization	Advanced
Jupyter Notebook	Interactive Development	Expert

Project Structure

```
netflix-movie-analysis/
│
├── movie-data-analysis-netflix-1.ipynb    # Main analysis notebook
├── mymoviedb.csv                          # Dataset
├── README.md                              # Project documentation
└── requirements.txt                       # Dependencies
```

Analysis Highlights

Dataset Overview

- **9,827 movies** analyzed across multiple dimensions
- **Comprehensive features**: Title, Genre, Popularity, Vote Count, Release Date
- **Rating scale**: 0-10 with statistical distribution analysis
- **Genre diversity**: Multiple genres per movie with detailed breakdown

Key Insights Discovered

- **Popularity Trends**: Average popularity score of 40.3 with high variance
- **Genre Distribution**: Action, Adventure, and Science Fiction dominate
- ⭐ **Rating Patterns**: Mean rating of 6.44 with normal distribution
- **Release Trends**: Concentrated analysis of recent releases (2021-2022)

Getting Started

Prerequisites

```
pip install pandas numpy matplotlib seaborn jupyter
```

Running the Analysis

```
# Clone the repository
git clone https://github.com/yourusername/netflix-movie-analysis

# Navigate to project directory
cd netflix-movie-analysis

# Launch Jupyter Notebook
jupyter notebook movie-data-analysis-netflix-1.ipynb
```

Key Methodologies

Data Science Techniques Applied

1. **Exploratory Data Analysis (EDA)**
   - Comprehensive statistical summaries
   - Data quality assessment
   - Missing value analysis
2. **Feature Engineering**
   - Custom rating categorization algorithm
   - Genre separation and analysis
   - Date feature extraction
3. **Statistical Analysis**
   - Descriptive statistics calculation
   - Distribution analysis
   - Correlation studies
4. **Data Visualization**
   - Multi-plot dashboard creation
   - Interactive chart generation
   - Statistical plot interpretation

Professional Impact

Business Value Delivered

- **Content Strategy Insights**: Genre popularity analysis for content acquisition
- **Audience Preferences**: Rating distribution patterns for recommendation systems
- **Market Trends**: Release date analysis for optimal launch timing
- **Performance Metrics**: Popularity scoring for content evaluation

Technical Skills Demonstrated

- ✅ Advanced Python programming
- ✅ Professional data cleaning workflows
- ✅ Statistical analysis and interpretation
- ✅ Data visualization best practices
- ✅ Jupyter Notebook development
- ✅ Documentation and code organization

Sample Visualizations

The notebook includes professional-grade visualizations:

- **Distribution Plots**: Vote count and rating distributions
- **Heatmaps**: Feature correlation matrices
- **Bar Charts**: Genre popularity rankings
- **Pie Charts**: Rating category distributions
- **Time Series**: Release date trends

Future Enhancements

- [ ] Machine Learning model for rating prediction
- [ ] Interactive dashboard with Plotly/Dash
- [ ] Real-time data pipeline integration
- [ ] Advanced NLP analysis on movie overviews
- [ ] Recommendation system development

Contributing

Contributions are welcome! Please feel free to submit a Pull Request.