

EDA on IMDB Top Movies Dataset

This notebook performs a full EDA analysis workflow on dataset of top IMDB movies. Below are the steps followed.

1) Imported the necessary libraries

→ pandas, matplotlib, seaborn, numpy

2) Initial Data Exploration

- displayed datasets basic information like shape, column names, datatypes etc using `.shape`, `.columns`, `.head()`, `.dtypes`, `.describe()`
- checked for null values using `.isna().sum()`
- counted unique values

3) Data cleaning & preprocessing

- fixed Incorrect Released-Year values
- converted Release:Year, Year to int
- cleared Runtime
- Dropped Irrelevant columns

4) Visualization and Analysis

plotted histogram for meta.score, IMDB.Rating, Runtime, Released.Year, No.of.Votes, Year

- checked for outliers using Box plot
- Bar chart of top 15 directors ~~also~~ based on avg IMDB rating

5) Exported cleaned dataset

Tharun Emeg
tharunemeg@makean
VJC