Project1:

Title: What Makes a Movie Popular? — A Deep Dive into Film Trends

---

The Problem Statement

In the digital streaming era, platforms like Netflix, Amazon Prime, and Disney+ constantly analyze content performance to understand what types of movies attract viewers and drive revenue.
In this project, students will analyze a real-world movie dataset to explore:

- What characteristics (e.g., genre, budget, runtime, cast, release month) influence a movie's popularity or box office success?
- How do viewer ratings relate to movie length, production budget, or release year?
- Are certain genres consistently more profitable or better rated?

Through exploratory data analysis and visualization, students will uncover trends and build a compelling narrative around what makes movies successful.

Skills Juniors Will Learn

- Data Preprocessing:
    - Handling missing data (e.g., missing revenue or budget values)
    - Parsing nested columns (genres, cast) and date/time fields
    - Cleaning categorical fields and standardizing units
- Exploratory Data Analysis & Visualization:
    - Creating line charts, bar graphs, scatter plots, and heatmaps
    - Trend analysis (yearly revenue, genre frequency, seasonal patterns)
    - Correlation studies (e.g., budget vs revenue)
- Bonus Extensions (Optional):
    - Use SQL (via SQLite or BigQuery) for querying large datasets
    - Build an interactive dashboard using Tableau or Power BI to visualize key insights like:
        - Top genres by revenue
        - Budget vs revenue trends
        - Movie releases by month/season

Why This Project Will Be Engaging for Juniors

 Relatable Topic: Everyone loves movies — working with familiar content keeps motivation high.

 Visually Rich: Data lends itself to great visualizations, charts, and dashboards.

 Balanced Learning: Combines technical skills with storytelling, a key real-world data science skill.

 Industry-Relevant: Mimics real work done by streaming platforms and film producers — makes the project feel meaningful and career-applicable.

This project introduces juniors to the analytical thinking and storytelling mindset that is crucial in every data science job.

Example Dataset Link

- **TMDb (The Movie Database) 5000 Movie Metadata:** https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata?select=tmdb_5000_movies.csv

Project 2:

Title: "GameChanger: Predicting Player Performance & Segmenting Athletes Across Sports"

Problem Statement

In the world of competitive sports, teams and franchises are increasingly relying on data to drive decisions—be it auction picks, player retention, or match strategies. In this project, students will use historical player performance data from **a sport of their choice** (e.g., cricket, football, basketball, etc.) to:

- **Predict** how well a player might perform in the next season (regression)
- **Classify** players into "high-value" vs. "low-value" or "retain" vs. "release" groups (classification)
- **Cluster** players based on performance patterns and roles (clustering + dimensionality reduction)

This can even simulate a **real-world auction or team selection system**, giving the project a real-world edge.

Skills Juniors Will Learn

- **Data Analysis & Visualization**: Understanding trends, correlations, and role-based performance (e.g., bowlers vs. batters, strikers vs. defenders)
- **Preprocessing**: Handling missing stats, encoding player roles, standardizing features
- **Regression**: Using models to forecast performance metrics like goals, points, wickets, etc.
- **Classification**: Identifying top-performing players for hypothetical team decisions

- **Clustering**: Segmenting similar players using K-Means or DBSCAN
- **Dimensionality Reduction**: Using PCA or t-SNE to visualize player groups in 2D
- **Ensemble Methods**: Improving accuracy with Random Forests or Gradient Boosting

---

Why This Project Will Be Engaging

- **Sport is relatable**—whether it's IPL, NBA, FIFA, or the Premier League, everyone follows some sport. This builds **immediate curiosity** and motivation.
- **Choice-driven learning**—students can pick **any dataset and sport** they love, making the experience personal.
- **Real-world application**—Mimics what real sports franchises do for auctions, player scouting, or performance benchmarking.
- **Collaborative learning**—Different teams can present insights from different sports, encouraging cross-domain sharing and healthy competition.

Example Datasets to Explore

- **Basketball (NBA)**: https://www.kaggle.com/datasets/drgilermo/nba-players-stats
- **Football (FIFA)**: https://www.kaggle.com/datasets/stefanoleone992/fifa-21-complete-player-dataset?select=players_15.csv
- **Other Sports**: Kaggle also has datasets on Tennis, Baseball, and even Olympic sports.