🎮 Steam Game Analysis Proposal

Analyzing game features, release trends, and key factors behind high-popularity games on Steam

---

🎯 1. Project Objective

This project analyzes the Steam game market to understand characteristics that contribute to game success. Key objectives include:

- Studying game release trends and seasonal patterns
- Exploring relationships between game attributes (price, popularity score, review ratio, release year)
- Identifying top-performing genres and publishers/developers
- Evaluating value-for-money games using metrics like popularity per dollar
- Understanding the price-quality relationship and genre competitiveness

🗂️ 2. Dataset Source

Source: Kaggle 🔗 Steam Games Dataset – fronkongames

Size: 110,000+ game titles released on Steam through May 2025

Key Features:

- Game Information: Name, price, release date, required age, genre
- Developer & Publisher: Studio names
- Popularity Indicators: Estimated owners, total reviews, recommendations
- Derived Metrics: Popularity score, review ratio, value score, popularity per dollar
- Genre/Seasonal: Encoded genres, release season, price category

📌 Why This Dataset?

- **Up-to-date:** Reflects recent trends in the Steam marketplace through May 2025
- **Comprehensive:** Wide range of features enabling deep multi-dimensional analysis
- **High Quality:** Clean, structured data with minimal missing values
- **Real-world Value:** Helps developers, gamers, and analysts understand market behavior

🧼 3. Data Preprocessing

- **Data Loading:** Check structure, dropped irrelevant fields like media URLs,...
- **Data Cleaning:** Handle missing values, duplicates, convert data types, check outliers and handle (if any)
- **Row Filtering**:
    - Removed non-game entries (e.g., SDKs, demos, servers)

    - Excluded games with missing genres or release dates
    - Dropped entries before 2004 or after early 2025
- **Feature Engineering**:
    - Extracted release year/month/season
    - Calculated important metrics
- **Text Normalization:** Standardize genre
- Apply One-Hot Encoding for genre variable

📊 4. Exploratory Data Analysis (EDA)

- **General Overview:** Total games, genres, free vs. paid ratio, average price, popularity score, review ratio.
- **Release Trends Over Time:** Games released per year and peak periods, genre popularity trends by year.
- **Price & Value Analysis:** Price vs. Popularity Score scatter plot, price segments (Free, Under $10, $10-30, $30-60, Over $60), average popularity score by price segment, "Popularity per Dollar" metric for best deals, top value-for-money games by genres.
- **Genre Analysis:** Top 10 genres by popularity, average popularity score per genre, niche genres with strong popularity, genre combinations analysis
- **ublisher/Developer Analysis:** Top publishers/developers by quantity vs. quality, AAA vs. Indie comparison (based on game volume), price, genre, and popularity differences between groups
- **Legendary Games:** High-popularity older games with lasting appeal.
- **Correlation Analysis:** Heatmap of price, popularity score, review, release_year relationships

🎨 All visualizations are integrated within the Jupyter notebook, presenting interactive charts and dashboards for each EDA section.

💡 7. Recommendations & Conclusions

- **Research Questions:** Are free games highly rated? Which publishers consistently perform well? Do release timing and pricing affect success?...
- **Expected Findings:** Top-rated genres, pricing strategies, publishers/developers performance patterns, succes factors in games industry,...
- **Applications:** Help developers optimize pricing and release strategies, assist gamers in game selection.
- **Limitations:**
    - **Estimated Owners Are Ranges**: Reduces precision in analyzing player base and revenue.
    - **Steam-Only Scope**: Limits insights to PC market, excluding console and mobile platforms.
    - **No Discount Tracking**: Price field does not reflect promotional pricing or sales, which may skew value analysis.
    - **Lack of Player Behavior Data**: Missing metrics like actual playtime, installs, and in-game transactions restricts engagement and monetization analysis.