466 Project Proposal

Team: Braeden Alonge, Lucas Summers, Xiuyuan Qiu, Nathan Lim

Dataset Info:

- Board Games Dataset
- 20k+ board games, 19 million ratings, 411k users
- Rich dataset with many entries and many categories/attributes of study
- Each entry typically contains:
    - Game attributes (e.g., year published, min/max players, mechanics, categories, play time, complexity rating).
    - Numerical data such as average user rating, rank, and number of ratings.
    - Textual data (descriptions, user reviews) – though for brevity, we may limit ourselves primarily to structured attributes rather than full text reviews.

Questions to Study:

- Which attributes of a game produce the highest user rating?
    - For example, does the game's complexity rating, recommended player count, or year of publication most influence its average rating?
    - We aim to build a **predictive model** (e.g., regression or classification) to estimate rating levels based on these features.
- Can we cluster board games into meaningful categories based on mechanics and complexity?
- Can we predict a "value" metric (e.g., average rating divided by average price or total playing time) for a board game from its attributes?