2515 project ideas

* Baseball
* Finance/stock market
* Spotify

Baseball

1. Predicting batter swing decisions
2. Bayesian pitcher performance modelling
3. For a pitcher, deciding which pitch to pitch and which location in (out of the) strikeszone should aim at

Available dataset

* <https://www.kaggle.com/datasets/pschale/mlb-pitch-data-20152018> (the one I used last year)
* Statcast, query the data according to the project topic (suggested by chat gpt)
  + For batter:
    - Pitch type (fastball, curveball, slider, etc.)
    - Pitch location (x, y coordinates relative to the strike zone)
    - Pitch speed and spin rate
    - Count (balls and strikes)
    - Game situation (e.g., outs, runners on base, score difference)
    - Historical batter-pitcher matchups
  + For pitcher:
* Pitcher performance metrics (e.g., ERA, WHIP, strikeout rate)
* Game location (home vs. away)
* Opponent team offensive strength (e.g., runs scored per game)
* Weather conditions (temperature, humidity, wind)
* Time of year (early vs. late season to account for fatigue)
* Recent performance (last 5 games)

Prediction batter swing decisions

…

Bayesian pitcher performance modelling

…

Finance

Stock market prediction using news data + yfinance python

Dataset: YFinance + <https://www.kaggle.com/datasets/miguelaenlle/massive-stock-news-analysis-db-for-nlpbacktests/data>

* Yfinance library has some real time finance data, while we will probably focus on past history data for model building
  + Depending on the news dataset we can get

Spotify

Dataset: https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset

Song recommendation / playlist auto builder for different events (e.g. studying, Friday night, after work, stress release, after breaking up, etc.). especially in a way that expose a user to new songs that they may not heard of before (but they may like). Of course for a playlist blending in some user’s favourite songs may be crucial, but in some case e.g. date night with jazz music, I don’t typically listen to jazz music but would like to listen to some, in this case the auto builder should exclude any of my listen history

* Motivation: currently the recommendation system of spotify is not too impressive. It mostly recommend what you have already listen to, or add some related songs at most. Of course there is a function called smart shuffle out there, but that is based on the playlist you have created (sometimes it is not precise too, but that’s not our focus). The purpose of this project is to
  + (imagine) having an interface that let the user input their mood, what they are currently doing, is there something specific they want to listen to, etc.
    - Like a chat gpt chatbot that reads your thought and use for our playlist build later
    - Designing an interface is not the focus for this project. We may want to link to a chat gpt chatbox, or just ask for input on the IDE and preprocess the words with BOW representation etc.
  + Takes in spotify user data (which can be achieved via spotify API), songs data (some features including danceability, etc.) of a wide range of songs. Use the user input from the step above to analyze user’s tendency of what they may want to listen to