Capstone Project Ideas

1. Video Game Recommendations

https://www.kaggle.com/tamber/steam-video-games/data
https://www.kaggle.com/gregorut/videogamesales

Steam is one of the world's most popular hubs for PC gaming. It includes a vast selection of games, from the most popular to small indie titles that most haven't heard of. Many users buy and play games using Steam, which means that user choices can lead onto interesting discoveries. For example, what games can you recommend users based on the games that they play and their preferences? When do people prefer to buy games? Are there lots of people who buy games and never play them? These are just some of the questions that can potentially be answered by analyzing this data.

The purpose of this project would be to answer the question: What game can be recommended to a particular user based on game play history? The two links above have two different data sets. The first includes user id, name of the game, whether the user has purchased the game, and how long the user has played it. The second includes a lot of other information, including game title, genre, platform, etc. The combination of both of these should be sufficient to answer the question.

2. Anime Recommendations

https://www.kaggle.com/CooperUnion/anime-recommendations-database

Anime is a Japanese animation that could refer to a movie or tv show. Many of these are adapted from manga, which is the Japanese form of comic books. A lot of people from around the world love to watch anime, and it would be interesting to see if their choices can lead to certain discoveries. This includes whether they have a preference for a particular genre, if it is possible to recommend certain movies or tv shows that they will definitely watch, etc.

The purpose of this project would be to answer the question: What anime can be recommended to a particular user based on viewing history? The link above includes a data set that has information on user preferences (around 70k users, and around 12k anime titles). The data includes user id, name and genre of anime, and whether the user watched and rated the anime.

3. Tennis Odds

https://www.kaggle.com/dmpierre/world-tennis-odds-database

Tennis is a popular sport around the world, that consists of singles or doubles matches. Many people like to bet on matches, which means that player history could help determine odds of winning or losing. This could range from analyzing win/loss ratios against particular opponents, or even go into detail such as first serve %, or number of unforced errors made.

The purpose of this project is to answer the question: Is player A going to beat player B? The link above includes a data set that has information on players that includes player names, tournament name, score, etc.

https://github.com/caesar0301/awesome-public-datasets#sports

The link above has a different database of information, that includes rankings, results, and stats for ATP (men), WTA (women), and Grand Slams (4 big tournaments in a year). This could be used instead of the previous data set, since the first one seems to have less detail. There is also a match charting project database that has shot-by-shot data, which means that whether a player wins/loses can possibly be determined by specific shot data (rather than just past win/loss history against the player).