Capstone 1 Project Ideas

1. **Predict survival time of brain cancer patients**

Brain cancer patients have different expected survival times based on the type of brain cancer they are diagnosed with, characteristics of their cancer, as well as personal and genetic characteristics. This capstone project will try to predict the survival times of patients diagnosed with brain cancer. I have found 2 different public brain cancer databases. The first is the REMBRANDT study - a large collection of genomic data from brain cancer patients (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108474). The second is the Chinese Glioma Genome Atlas - a collection of over 2000 brain tumor samples (www.cgga.org.cn/index.jsp). Regardless of which database I choose to use, I will most likely only use the clinical data.

A possible problem is that both datasets have a significant number of survival times that are right censored (where the study ends prior to the patient's death or the patient withdraws from the study early). I may be able to use survival analysis methods to solve this problem. Another possible problem is that both datasets have many missing data values. If these limitations considerably complicate the project, I may need to find a better database to work with.

2. **Predict whether a player drafted by the NBA will become an all-star caliber player within their first 5 seasons**

In this capstone project I will try to predict whether a player drafted by the NBA (with at least one year of college basketball experience) will become an all-star caliber player (based on objective advanced NBA stats, i.e. Win Shares) within 5 years. Statistics from a player's final college season as well as other player data/characteristics will be used to make the prediction. The college basketball player statistics are available here (www.sports-reference.com/cbb), while NBA player statistics are available here (www.basketball-reference.com). I will use this third-party open-source API (sportsreference.readthedocs.io/en/latest/index.html) to get the necessary data from both sites (same company) so that I can create my dataset relatively easily.

3. **Attempt to create a profitable predictive sports betting model for NFL games using deep learning**

My aim in this capstone project is to try to create a profitable predictive sports betting model for NFL games using deep learning. I will most likely need to create my dataset from multiple sources with the use of APIs.

FiveThirtyEight's NFL game prediction model (github.com/fivethirtyeight/nfl-elo-game) is relatively accurate, so I will most likely use the data from their model as one of the predictors in mine. I will get historical game results and advanced NFL statistics from www.pro-football-reference.com using the same third-party API that I referenced previously. In order to get NFL opening and pre-game spreads, I will use the API of a betting website (i.e. Betfair). I may also decide that I need additional data to create an effective model.