***Final Project Proposal***

***What:*** I am choosing to create a database model to represent and explore the world of professional chess. If possible, I will complete the database with actual data and use a series of queries to uncover any noteworthy anomalies or outliers.

***Why:*** In January 2021, one of my friends convinced me to begin playing chess online on Chess.com. As my university was not offering in-person classes during that period due to the pandemic, I spent a lot of time at home playing chess. This quickly became my go-to pastime when I needed to occupy myself to avoid boredom, and I soon surpassed the abilities of the friend who introduced the idea to me.

Then, in the latter stages of 2022, I discovered chess content creators such as International Master Levy Rozman, who then in turn exposed me to the exploits of top chess players. I quickly became interested in the developments in the chess world, especially as this was transpiring as the Carlsen-Niemann cheating controversy began to unfold. Since then, I have been following major events involving elite players regularly, including online and over-the-board tournaments.

Recently, more controversy has arisen. Some big names in chess, including Grandmaster Fabiano Caruana and former World Champion Vladimir Kramnik, have expressed concern with the prevalence of cheating on Chess.com. Their attention was particularly directed towards the weekly tournament named Titled Tuesday that occurs on the platform. Their concerns were not alleviated by recent comments made by Chess.com Chief Chess Officer Danny Rensch, who claimed just three percent of titled players have cheated on the site. Kramnik and Caruana estimate the true figure is much higher.

Thus, I am interested in constructing a database model, preferably filled with real-world data as well, to facilitate any relevant analysis of this issue, as well as to any other future claims to be investigated.

***How:*** My pessimistic plan is to construct a complex ER diagram with some real-world data. Entities would include a player table with biographical information, two ratings tables (one for Chess.com ratings and another for FIDE ratings), a table for OTB blitz results, another table for Titled Tuesday results, more tables for other Chess.com and over-the-board statistics, a titles table, and anything else I can brainstorm. Some data I already have access to and will be able to include, such as the Titled Tuesday dataset. I would also ensure everything is normalized.

My optimistic plan is to fill out every table with accurate information to then perform many queries. I would attempt to identify outlier players who perform much better online than over-the-board. I could do this by performing significance testing. I could also investigate any other curiosities that I discover as a part of some exploratory data analysis, such as if younger players tend to overperform more online compared to their older counterparts.