**Business Objectives**

Through in-depth analysis and machine learning techniques, my objective is to showcase the data science skills I have acquired by providing analyses of a sample of 20,000 chess games from Lichess.org. This sample was made available on Kaggle. My hope is that others will be able to use these analyses to gain helpful insights.

I will use Python, R and Tableau for my project, which will include data wrangling and creating visuals in addition to performing analyses. I will then present my work to my instructor.

**Background**

This project was chosen because the data offers a chance to highlight a variety of data skills, including regression and machine learning, highly sought skills in the data science field.

I also am very interested in the dynamics of chess.

**Scope**

This project will analyze the chess game data to discover whether a variety of variables influence the winner (black or white), including ratings of players, starting move and number of moves. It also will include analyses of whether a variety of variables affect game outcome (mate, draw, resignation, time out). A machine learning model will be created that will use a variety of variables to predict the game winner. It also will include descriptive statistics such as mean game time and mean player rating.

I may do some further analysis, but that is not required.

I will use the programs Python, R, and Tableau. I may use SQL or another program if needed, but that is not required.

**Functional requirements**

Data wrangling: This project will require a lot of data wrangling. A new column of game length will need to be created; any unnecessary and/or potentially identifying columns will need to be dropped (i.e. player usernames), and some columns may need to be converted from character to numeric and vice-versa. In addition, the times need to be converted.

Data analysis: Multiple logistical regression analyses will need to be conducted on the data set to answer the project questions. The project also includes conducting an independent chi-square and a machine learning model. Simple descriptive analyses also will be run.

Data visualization: The creation of graphs and other visuals will be conducted using Tableau or other software. The visualizations will need to be clear, easy to understand at a glance, and visually appealing. They will include labels, keys, and other text information to explain the graphics.

Presentation: When the project is complete, a presentation will be given that includes a breakdown of the data, explanation of questions and appropriate analyses, how data was wrangled, how the analyses were conducted, and the findings.

**Personnel requirements**

This project will be conducted by just myself. I will need to check in daily on Exeter LMS and/or in my planning material to be sure I am on track. Further, I will meet with my instructor each week to report on my status. I will use this opportunity to ask any questions I might have, or to seek guidance.

My instructor and mentor will be tremendous helps to completing this project, as they will offer guidance and feedback throughout.

**Delivery schedule**

Week 1: Will have completed choosing data for my project, choosing questions to answer with that data and the appropriate analyses, setting up a Github repository, and creating the project proposal.

Week 2: Will complete all known necessary data wrangling, including any subsetting required and any data transformations and recoding. Some wrangling might carry over to week three as I perform exploratory analysis and discover the need for additional wrangling.

Week 3: Will complete exploratory analysis to get a better feel for the data and the types of questions I can explore, as well as the types of analyses to run.

Week 4: Will conduct all analyses and examine and explain conclusions. Will use these results to further think about the dataset and determine whether additional analyses would be interesting/helpful.

Week 5: Will create visualizations that showcase my findings. Will use labels, etc. to ensure visualizations are easy to understand. Will begin creating presentation in Powerpoint.

Week 6: Will finalize presentation and review. Will present presentation to instructor via Zoom.

**Assumptions**

My computer, software, Internet and other technologies on which I will rely will function correctly.

**Limitations**

If I have an unanticipated event arise, it could endanger completion of the project on time. Also, if my instructor or mentor are unavailable due to illness or time off, it could inhibit on-time completion.

To ward off this possibility, I plan to work ahead as much as possible, and have a buffer.

**Risks**

Possible risks to the project completion include natural disasters, power outages, illness, and family emergencies. As stated above, I hope to work ahead and have a buffer to allow for any interruptions to the project.