Most of these questions do not have a "correct" answers -- we are trying to get a sense for how you think about problems. We like to see what sorts of features and models you would try, what data you would need, how you would evaluate those models, how you would adjust for different considerations, and what edge cases you may be considering. Please reach out with any questions and we may hop on a call later to go over your answers to get a better idea of how your thought process works.

1) Using the attached dataset (mlb\_pitch\_velo\_assessment.csv), build a model that predicts whether the first pitch of a baseball game by each starting pitcher will be faster than 89.95 mph.

2) The spread for an NBA game is 76ers -7.5 at -110 and Lakers +7.5 at -110, and the over under for the game is 235. The money line markets have the 76ers at -320 and the Lakers +260.  How many points is each team expected to score?  What is the implied win probability for each team?  How much "vig" is the book taking on the spread and money line markets?

3) If the Jets are trying to decide whether to go for it on fourth down, from the 30-yard line, given the following probabilities, what conversion rate should they need to go for it?

Win Probability after made field goal: 56%

Win Probability after missed field goal: 45%

Win Probability after successful conversion: 60%

Win Probability after turnover on downs: 46%

Field Goal Make Probability: 70%

4) What kind of model/statistical tools would you use to project the probability of a kicker making a field goal based only on distance?

5) For this question, write code in any language. Write a function that takes a string as an input and counts the number of vowels and consonants in that string.

6) Each Saturday during the college basketball season there can be over 100 games for which we offer live markets. A trading manager comes to us and asks if we can help them decide which of those games we should be staffing with our best traders on any given Saturday. How would you go about building a model to help our trading manager make that decision?

7) How would you go about making an in-tournament golf win probability model that could update live throughout the weekend?

8) How would you go about identifying which row was duplicated the greatest number of times in a data set? What if you could not load the full data set into memory at one time?

9) A customer has been on a hot streak betting on our sportsbook. Looking at their betting history, how would you try to identify if they were merely lucky, or good?

10) Given a home team is expected to score 5 runs and an away team is expected to score 4.5 runs, which outcome is the most likely?

a) 10 total runs

b) 9 total runs

c) They are equally likely

Please explain your reasoning.