- What is the problem you want to solve?

The project will attempt to address the relationship between NBA player performance and salaries. It would include accounting for a variety of player metrics and player salaries as all change over the course of players' careers. The problem to be solved is can players and teams better understand how players should be compensated throughout their careers.

- Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

There are two potential groups of clients for this problem. The first would be NBA owners and general managers who may find it useful to determine when to best invest in a player based on their career performances in particular metrics. For an NBA team, maximizing their investments in their players is crucial to their financial success. This project would allow them to have a better idea of how much it's worth to pay a player at different points in his career

The second group would be NBA players themselves who could use the information to not only determine when their career performances may peak, but also to determine when to pursue maximum contracts during the course of their careers. If a player can better understand when he will be likely to experience a drop in performance, he can use that information to his advantage to sign more lucrative contracts at an appropriate point in his career. Players may also be able to determine if there is a significant salary drop in a player's later career and whether getting paid well early in one's career leads to a higher overall career salary average.

- What data are you using? How will you acquire the data?

The data comes from two different sources. The first is salary data between 1991 and 2017. It lists the yearly salary of every player in the league except for those players who were signed late in the season, cut early, or only on 10-day contracts. The data was acquired from a basketball statistics website.

The second set of data is from Kaggle and lists statistics of every player between 1950 and 2017. The statistics include yearly totals as well as career totals in over 40 categories. It also includes when players' careers began and ended.

- Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

Since the aim of the project is to trace a player's salary and performance over their entire career trajectory, and since the salary dataset, which is over a small timespan, begins in 1991, we

would need to cut out all players whose careers began before 1991. That dataset would then be merged with the salary dataset, which would also have to drop all data before 1991. All missing values would then have to be accounted for by either dropping certain columns if too much data is missing or filling in the missing values.

Once the data has been cleaned and merged into a single dataset, then we'll begin attempting to establish relationships between player salaries and traditional metrics, including points, assists, and rebounds, as well as more advanced metrics such as player efficiency rating and offensive and defensive plus/minus. It is hoped that this analysis will shed some light as to how certain metrics increase and decrease over a player's career and which metrics are worth paying for. We may also do a similar comparison for a smaller subset of the highest paid players of the given timespan to determine whether the performances of these "elite" players differ from the rest and whether that justifies paying them more.

- What are your deliverables? Typically, this includes code, a paper, or a slide deck.

The best deliverables for this project would likely be a paper paired with a slide deck. A paper would allow for a more detailed breakdown of the project, highlighting how the project was constructed from start to finish, including the results. However, slide deck would allow for a more streamlined presentation of the most important conclusions.