# DS 3000 – Data Analysis Plan

In this deliverable, you will submit a short summary (approximately three/four paragraphs) of your project, including a brief introduction to the problem and its significance, a statement of research questions/hypotheses and a description of the dataset and data analysis plan (i.e., the techniques you are planning to use to answer your research questions and/or test your hypotheses). Think of this as a first step in putting together all the pieces of your project.

**What to submit**

* Complete the sections on the following page.
* Before submitting your completed document, **delete this first page** containing the instructions. Otherwise, 1 point will be deducted from your overall grade for this FP deliverable.
* **Submit your completed document** on Canvas.
  + Submit a PDF document.

## FP3 – Data Analysis Plan

**Overview <Andrew>**

Your summary should begin with an overview of your topic (this is the final topic that you are committing to investigating in your final report), including a description of the problem. Here are a few questions that you should answer.

* What is the topic of your project?
  + Basketball Contract Value
* Why is it important to tackle this problem in your project?
  + I will copy from FP1
* What is your dataset about? How many samples (rows) does it have?
  + I can fill this in later (stats of players along with cap % of team)
* What are your features? What is your target variable?
  + See dataset

**Questions/Hypotheses <Jeffery>**

* What questions have you formulated regarding your dataset? These questions may tie back to the problem statement (just present them as questions this time). You should have at least one question tapping into the comparison of multiple ML algorithms in predicting your target variable from your feature variables.
  + Correlation between player performance & contract value?
  + Which stats are more important in the process
  + Which players are outperforming / underperforming their contract value (prediction)
* What hypotheses are you going to test? You should have at least one hypothesis test in your final report.
  + What hypotheses have you formulated regarding potential causal relationships among some of the variables in your dataset?
    - PPG plays an important role in predicting contract size
    - +/- doesn’t have an effect on predicting contract size
    - The model highly weighs age in the prediction algorithm
    - Advanced stats do not have a high correlation to contract size

**Data Analysis Plan <Ryan Kennedy>**

In this section, describe some of the analyses you are planning to conduct. This is tentative and can change later, but you should have a clear idea of what to do for your ML problem. Answer the following questions:

* Is your project tackling a classification problem or a regression problem?
  + Regression Problem
* Which ML algorithms are you planning to use?
  + TODO
* Which feature extraction/engineering techniques are you planning to use?
  + GridSearch w/ MIN/MAX Scaling
* What are some variables that would be useful to visualize? What types of visualizations (e.g., scatter plots, bar graphs, etc.) are you planning to make? Name at least two different visualizations and explain why you want to produce them.
  + Scatter plots for advanced stats (we can expect simpler stats to be very correlated to contract size), PPG as well, look at age as well (check for veterans contracts and rookie deals)
  + Violin: Every statistic as X, Cap space % as Y