Hypothetical:

You are a GM for a team that is evaluating several older free agents. These free agents have performed well, but you want to make sure that they will continue to perform at a high enough level as they age. Therefore, you want a method to valuate these players that takes into account performance decline due to age.

Loose Project Steps:

1. Feature selection to determine which stats are most predictive of player value (validate/build-on Moneyball approach – maybe use more advanced technique than just basic linear regression)
2. Build model to predict player contract valuation
3. Do some analysis to determine the decline in performance of aging players – compare survival analysis and time series models
   1. Incorporate survival analysis models or a time series approach
4. Extensions:
   1. Clustering of undervalued players
   2. Metrics to consider:
      1. Prediction of player performance decline
      2. Some way to compare our valuation to actual contract value played and player performance?
   3. Others?

Resources and Data:

<https://swetankpathak.medium.com/its-all-about-time-survival-analysis-part-a-injury-perspective-c7ca818575d4>

<https://sportsmedicine-open.springeropen.com/articles/10.1186/s40798-022-00465-4>

Injury data (starting in 2020):

<https://www.fangraphs.com/roster-resource/injury-report?timeframe=all&season=2023>

More data to look at: <https://github.com/robotallie/baseball-injuries?tab=readme-ov-file>

Python survival analysis library:

<https://scikit-survival.readthedocs.io/en/stable/user_guide/00-introduction.html>