ShotQuality Comprehensive Technical Evaluation

Please spend no more than four hours on this challenge. Contact Ethan Carpenter
(410-992-2955) for any questions.

Modeling

You have been asked to build a model that predicts the probability a shot is made for the
Euroleague. You have access to play-by-play for all professional leagues globally, and location
data (x and y coordinates for all ten players at the moment of the shot) for the Euroleague's
most recent season.

Discuss your **full** approach to the model, be sure to consider the following:
- How would you approach the modeling question?
- Data cleaning
- Feature engineering
- Model selection
- Model training
- Model maintenance
- Production workflow (i.e. automated prediction, retraining)

Hardware constraints can cause problems for model development, especially when working with
many models for different leagues, or large historic datasets. What are some issues that could
arise when we push this into production, and how could you address them?

Database

You are building a new service that collects data from ESPN, NBA.com, NCAA Live Stats, and
Basketball Reference, and delivers a single stream of box scores and play-by-play to external
API users.

What are some key considerations in designing a database for this need? Be sure to include a
discussing of the following:
- Database design
- Security
- Performance and cost considerations
- Data quality
- Data integrity
- Robust ID system (i.e. games, players, teams)

Data Processing

You noticed that the play-by-play and box score for a game do not always agree. For example,
the box score credits a player with two steals, but only one appears in play-by-play.

Describe how you would implement a process to monitor and resolve these issues. Be sure to consider the following:
- Error detection
- Delivery of potentially erroneous data (i.e. how do we handle erroneous data from a product perspective?)
- Error resolution

For this problem, assume unlimited time and resources. For example, if your solution is to drive to the league offices and petition them to fix the data for you, that is within the scope of this problem.