

# WASHINGTON COMMANDERS QUANT ASSESSMENT

## Part 1. Model Design

I will systematically rank players based on their likelihood of becoming special team contributors using their college measurables (height, weight, speed, etc.) and positions. This involves handling missing data, defining and quantifying special teams' contributions, evaluating players through feature engineering and modeling, and anticipating potential problems.

### Handling Missing Data

To address the problem of missing data, I will follow these steps:

1. **Identify Missing Data:** The first step is to identify which columns have missing data and the extent of these missing values. This helps in deciding the appropriate imputation method.
2. **Impute Missing Values:**
  - **Numerical Data:** For numeric columns such as height, weight, and speed, I would use median imputation. The median is robust against outliers and provides a stable central tendency.
  - **Categorical Data:** For categorical columns like position, mode imputation can be applied. However, if a significant portion of the data is missing, predictive modeling or advanced imputation techniques like K-Nearest Neighbors (KNN) imputation can be employed to estimate missing values.

### Quantifying Special Teams Contributions

Quantifying special teams' contributions involves defining key performance metrics that accurately reflect a player's impact. The key metrics I will focus on include:

1. **Tackles:** The number of tackles a player makes directly indicates their effectiveness on special teams.
2. **Blocks:** The number of blocks executed which can prevent the opposing team from gaining yardage.
3. **Returns:** Yards gained from kickoff and punt returns, showcasing a player's ability to advance the ball.
4. **Participation:** The number of special team plays the player participates in highlights their experience and utility.

To determine the significance of each metric, I will analyze historical data to understand which metrics most strongly correlate with successful special teams' performance. Based on this analysis, I will assign weights to each metric. For example, if historical data indicates that tackles have a greater impact on special teams' success than blocks, then tackles will be weighted more heavily in the evaluation.

### Evaluating Players

Evaluating players involves creating a composite score that reflects their potential contribution to special teams. Here are the steps I will follow:

1. **Feature Engineering:**

## WASHINGTON COMMANDERS QUANT ASSESSMENT

- **Composite Metrics:** I will create new metrics such as Body Mass Index (BMI) from height and weight data and a speed score combining weight and 40-yard dash time. These metrics provide a more holistic view of a player's physical attributes.
  - **Normalize Data:** Standardize the data to ensure that each feature contributes equally to the model. This involves scaling numerical features to have a mean of 0 and a standard deviation of 1.
2. **Model Training:**
- **Predictive Modeling:** I will train machine learning models, starting with logistic regression to establish a baseline and then exploring more complex models like random forest or gradient boosting. These models will predict the likelihood of a player becoming a special team contributor.
  - **Cross-Validation:** To prevent overfitting, I will use cross-validation techniques. Cross-validation involves dividing the dataset into multiple folds and training the model on different subsets while validating it on the remaining data. This ensures that the model generalizes well to new data.
3. **Scoring and Ranking:** Once the model is trained, I will use its predictions to assign a contribution likelihood score to each player. Players will be ranked based on these scores, with higher scores indicating a higher likelihood of making significant contributions to special teams.

### Anticipating Potential Problems

Several potential problems could arise during this process, and it's important to anticipate and address them:

1. **Data Quality:** Inaccuracies or inconsistencies in the data can lead to incorrect predictions. It's crucial to clean the data, handle outliers, and ensure the accuracy of entries.
2. **Model Overfitting:** Overfitting occurs when a model performs well on training data but poorly on unseen data. To prevent this, I would use regularization techniques such as L1 or L2 regularization and robust cross-validation practices.
3. **Subjectivity:** Defining contribution metrics and assigning weights can be subjective. To minimize this, I would use objective, data-driven criteria and be transparent about the methodology and rationale behind metric selection and weighting.
4. **Model Interpretability:** Complex models can be difficult to interpret, making it challenging for stakeholders to understand the results. Using interpretable models like logistic regression initially can help, as they provide coefficients that can be easily understood to show the impact of each feature.

### Conclusion

This systematic approach ensures a comprehensive and robust analysis for ranking players based on their likelihood to become special teams' contributors. We can develop a reliable and interpretable model by carefully handling missing data, defining and quantifying key performance metrics, evaluating players through feature engineering and model training, and anticipating potential problems. This model will provide valuable insights to decision-makers, helping them identify players who can significantly impact special teams' performance.