

## WASHINGTON COMMANDERS QUANT ASSESSMENT

### Part 1. Model Design

You are given a data set with players' college measurables (height, weight, speed, etc.) and position, and are asked to rank them with respect to how likely they are to become a special teams contributor (excluding punters/kickers). Some of the players have missing measurables. How would you approach this problem? What other problems would you expect to arise?

**Please save as “LastNameFirstName\_ModelDesign.pdf”. Do not include your name in your submission. Your submission should be no more than 1000 words.** In your response, clearly identify how you would address the problem of missing data, how you would quantify special teams contributions, how you would evaluate players, and other problems you would expect to arise.

### Part 2. Coding Exercise

The attached dataset labeled “CommandersQuantTestData.csv” includes real play calls from a NCAA football team. The data has had identifying information masked but has been masked consistently to retain information necessary for prediction. See below for a short explanation of the included columns.

Key:

- PLAYID: A unique identifier for each individual play
- DOWN: The play down
- DIST: Distance to the first down
- LOS: Absolute line of scrimmage from 1 to 99, with the number indicating how many yards the offense is from scoring a touchdown (i.e. a 10 yard gain at LOS = 10 would be a touchdown)
- GAIN: Number of yards gained as a result of the play
- FORMATION: Offensive formation alignment
- PLAYCALL: The offensive play call, masked to obscure team information
- PLAYTYPE: Run or pass
- PASSER: Quarterback on the play
- SCOREDIFF: Score differential in terms of offense minus defense
- DEFTEAM: Opponent team name, masked to obscure team information
- KEY PLAYER NUMBER: The main player on the play, usually the ball carrier on a run play and the receiver on a pass play
- KEY PLAYER POSITION: The aligned position of the key player for the play
- PASSRESULT: The result of a pass play, either complete, incomplete, interception, sack, or scramble
- FUMBLE and INTERCEPTION: Binary indicators of when a play results in a fumble or interception, respectively

Assume all penalty plays have been removed and that there is no other data available.

## WASHINGTON COMMANDERS QUANT ASSESSMENT

Your task is to create predictions for the number of yards gained on each play in which the column is marked as MISSING. **You will be graded solely based on the accuracy of your predictions, as measured by two measures:**

- **Root Mean Square Error between your predictions and the true results of the plays**
- **A 90% prediction interval for the sum of yards gained on all the MISSING plays**

### Requirements:

- You must complete this project using R or Python.
- Your predictions may only use the columns listed above, excluding the PLAYID column, which should only be used for identification purposes.
- You are allowed to create new features from the columns listed above, so long as they are derived strictly from the information included in the provided file.
- **While your code will be judged and compared with others based on the measures described above, uncommented code, as well as incomplete submissions will be immediately disqualified.**

You may use any methodology to obtain your estimates, but you must include a short (300 words or less) explanation of your methodology. Please submit your solution in the form of a .csv with exactly two columns: PLAYID and GAIN. There should be 125 rows, corresponding to the plays marked as MISSING.

Your submission should include the following three items:

- An R or Python script that can generate/replicate your predictions, named “LastNameFirstName\_CodingScript.R” or “.py”
- Your actual predictions, named “LastNameFirstName\_CodingPredictions.csv”.
- Your explanation of methodology, named “LastNameFirstName\_CodingExplanation.pdf”