SPORTS INFO SOLUTIONS

FOOTBALL ANALYTICS CHALLENGE

FOR THE

BOYS & GIRLS CLUBS OF AMERICA

Elijah Cavan

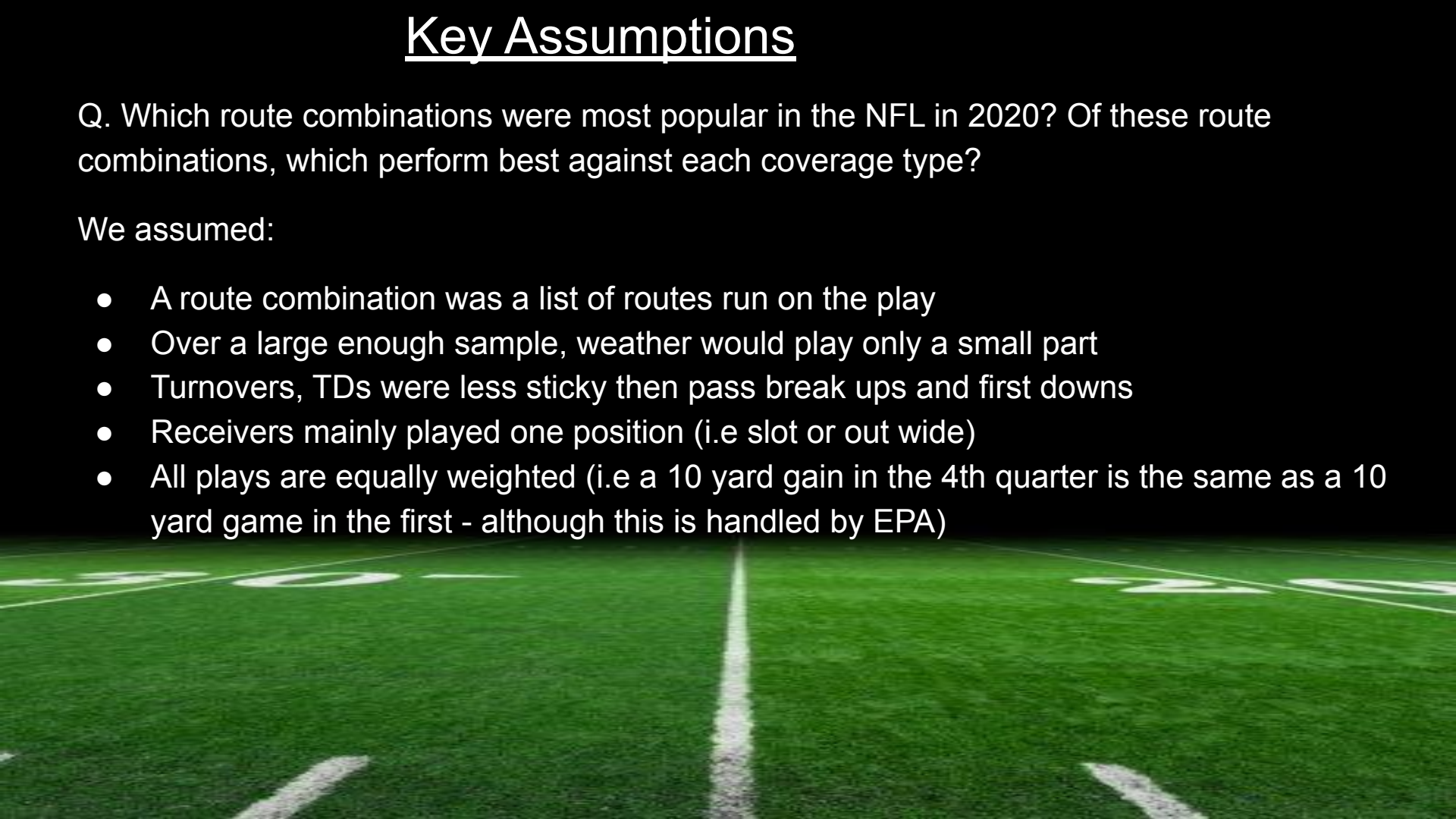
William Ngana

Key Assumptions

Q. Which route combinations were most popular in the NFL in 2020? Of these route combinations, which perform best against each coverage type?

We assumed:

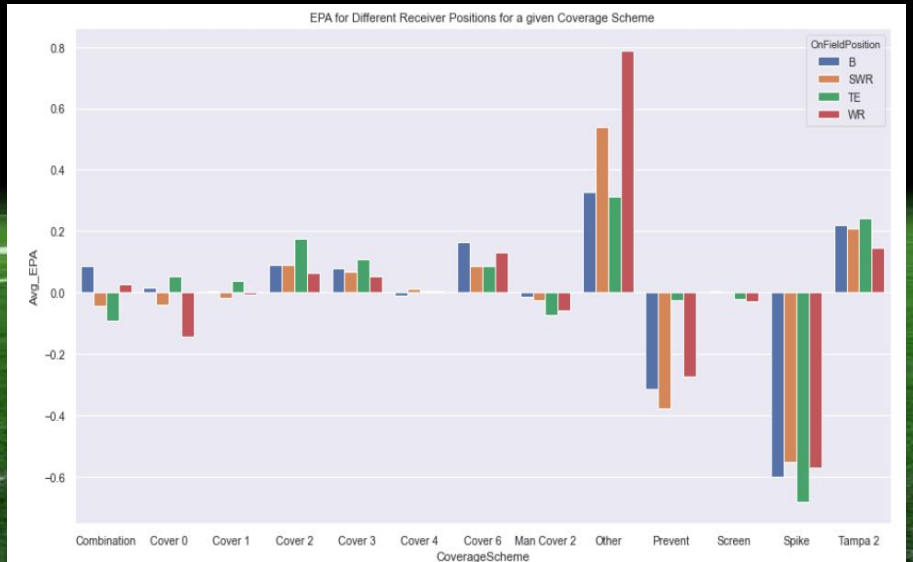
- A route combination was a list of routes run on the play
- Over a large enough sample, weather would play only a small part
- Turnovers, TDs were less sticky then pass break ups and first downs
- Receivers mainly played one position (i.e slot or out wide)
- All plays are equally weighted (i.e a 10 yard gain in the 4th quarter is the same as a 10 yard game in the first - although this is handled by EPA)

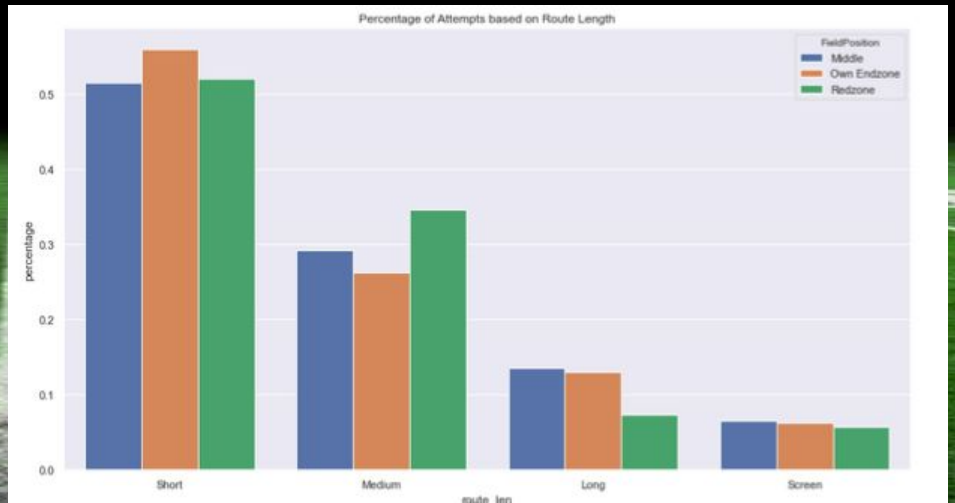


Exploratory Data Analysis (EDA)

Insights:

- Slot receivers are generally better than wide outs against Tampa 2. TEs are the best receivers in general against cover 0, cover 1 & cover 2.
- There is not much variation in terms of route length decisions between any part of the field





EDA II

These are the best routes/route combinations based by EPA. Here we are showing Cover 2 as an example.

Optimal route combination: [Go/Fly, Seam, Curl, Chip-Flat]

Best Route length combinations by EPA for: Cover 2

	route_len_list	mean	count
141	['long', 'screen', 'short', 'short']	0.632170	24
108	['long', 'long', 'medium', 'short', 'short']	0.586691	170
132	['long', 'medium', 'screen', 'short']	0.473960	32

Best Route combinations by EPA for: Cover 2

	Route_list	mean	count
482	['Chip - Seam', 'Curl', 'Curl', 'Post']	7.039744	4
1372	['Dig', 'Fade', 'Out']	5.386134	3
1397	['Dig', 'Go/Fly', 'Go/Fly']	4.685438	3

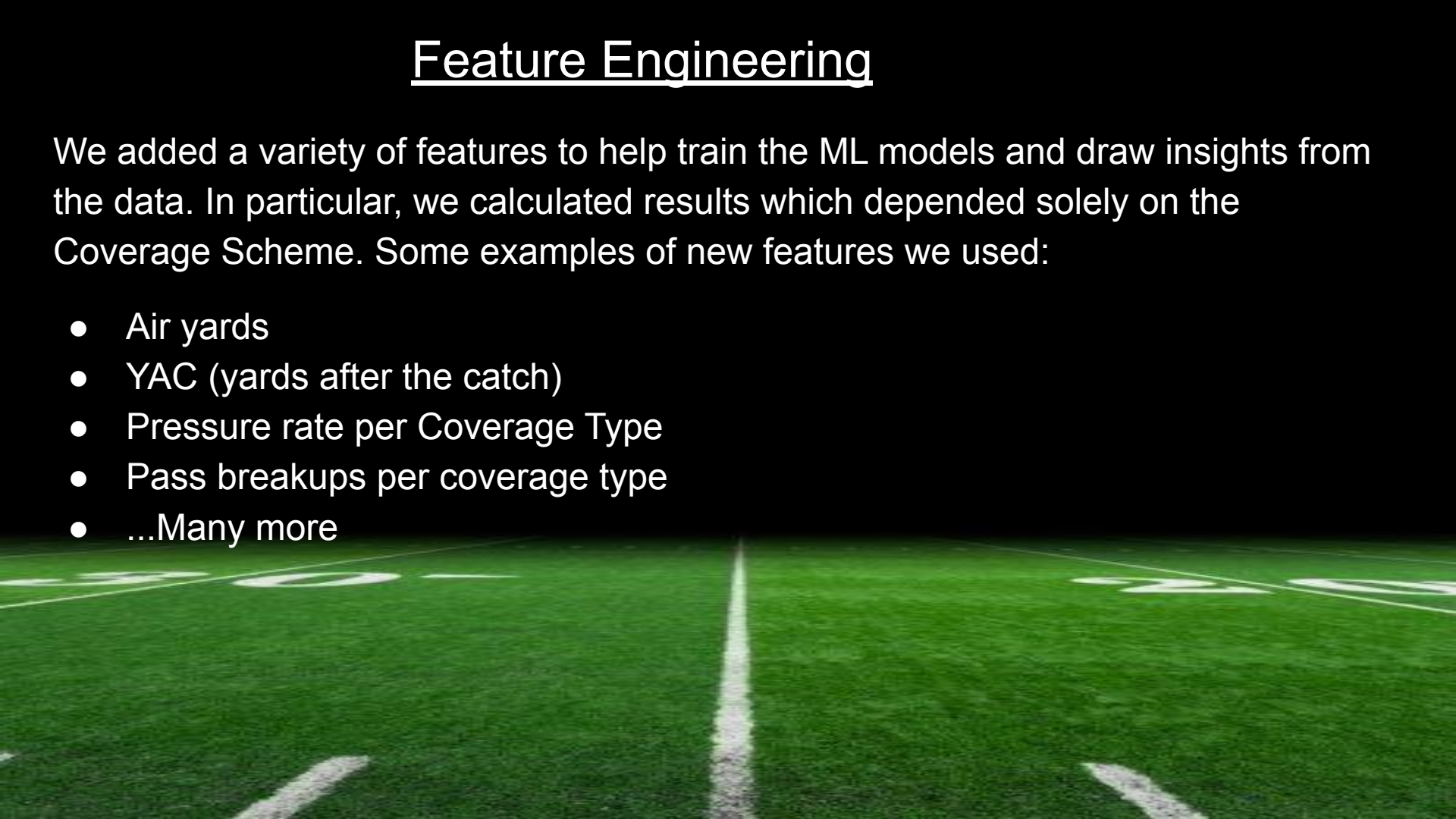
Best Routes by EPA for Cover 2

	Route	route_len	mean	count
17	Go/Fly	Long	0.208795	420
5	Chip - Flat	Short	0.194880	199
29	Seam	Medium	0.191790	325

Feature Engineering

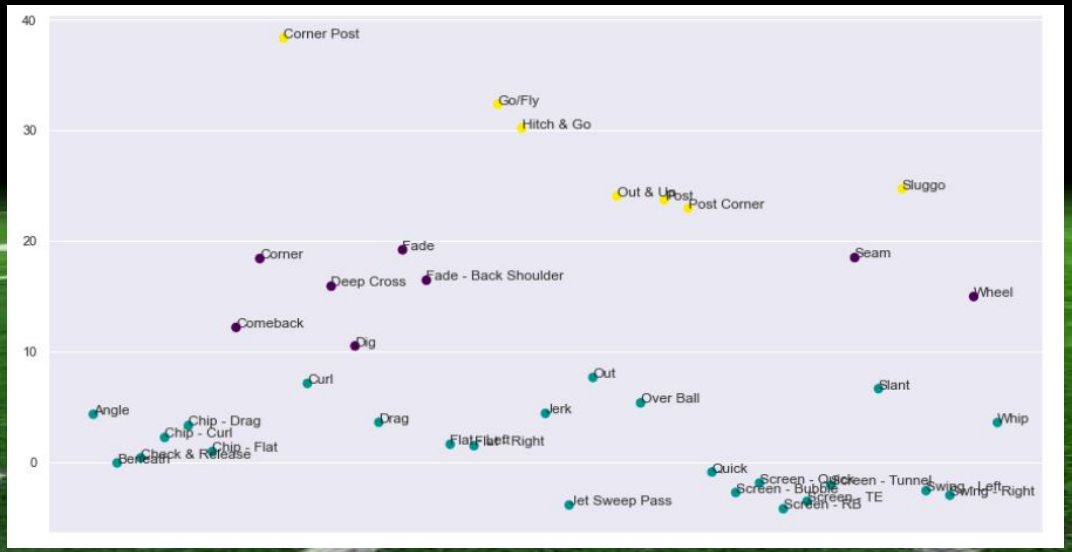
We added a variety of features to help train the ML models and draw insights from the data. In particular, we calculated results which depended solely on the Coverage Scheme. Some examples of new features we used:

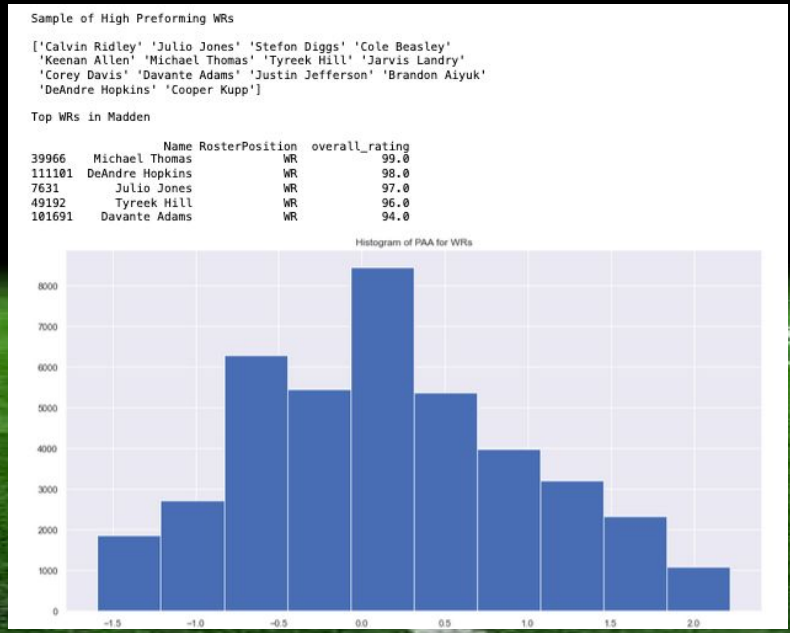
- Air yards
- YAC (yards after the catch)
- Pressure rate per Coverage Type
- Pass breakups per coverage type
- ...Many more



Feature Engineering II

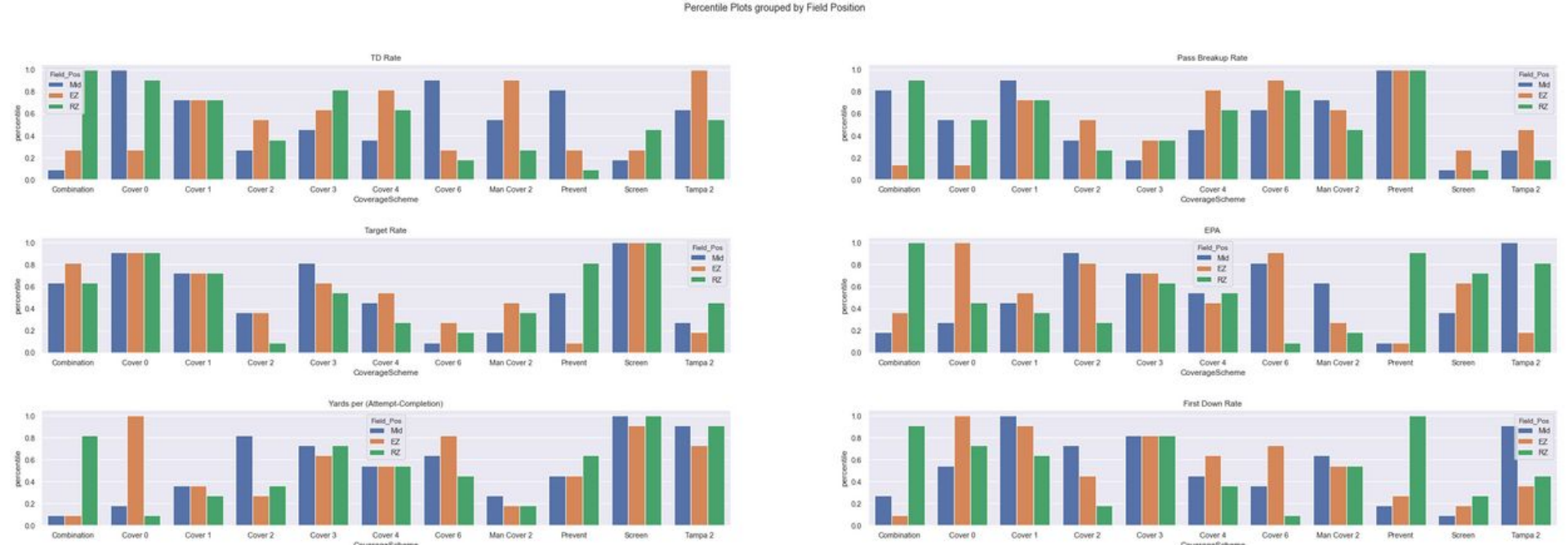
In addition to defining new features, we also changed some of the pre existing columns. Here we are showing a scatter plot of the KMeans clusters for routes based on air yards (these are classified as Short, Medium, Long routes). As well as a histogram of Points above Average for WRs.





Percentile Plots

Plotting the most important features per coverage type based on percentiles relative to other coverages. Further split by Field Position (Redzone, Own Endzone, Middle of the Field)



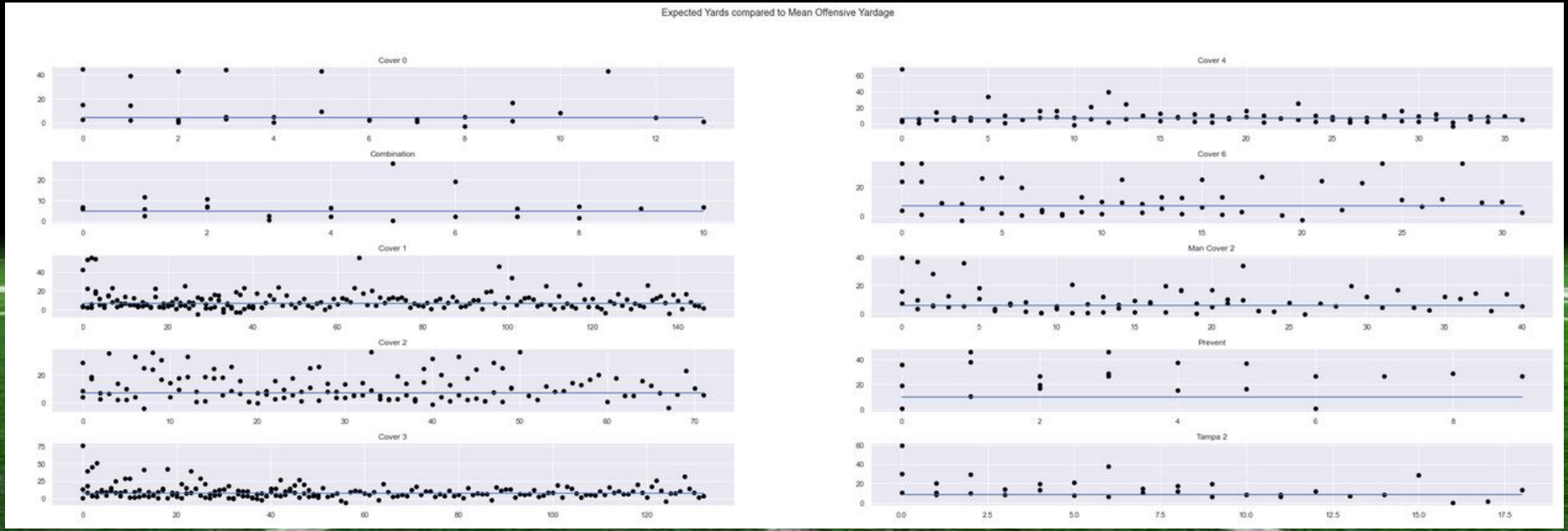
Percentile Plots: Insights

We have determined that:

- The cover schemes (cover 0, cover 1 for example) tend to be more steady, while schemes like Tampa 2, Combination can wildly fluctuate in effectiveness based on field position
- The more DBs you have in coverage tends to limit targets; but you end up giving up high EPA plays/ many first downs
- Prevent has a high pass break up rate; but also a high target rate. This tells us that game situation plays a huge role in what play schemes and coverages are run.

ML Model: Expected Yards

The blue line is the average Offensive Yardage for all route combinations against each coverage. The black dots are the models output for the combinations of routes that have the highest average EPA (i.e see EDA II slide)



Expected Yards: Insights

We have determined that:

- The optimal routes are generally more effective routes than the average route that is run.
- The cover schemes (eg. cover 0, cover 1) seem to give up more big plays compared to the Combination and Tampa 2 schemes
- Oftentimes, even running the most optimal routes results in plays that perform below average (although we have not selected for what route was targeted)

ML Model: Probability of First Down

ML model which calculates the probability of a first down based on the route being run on the play (as well as a number of other features). The plots show the probability of a first down given that an optimal route is being run, compared to the average probability for all routes.

Probability of First Down: Insights

We have determined that:

- For every coverage scheme, running the optimal routes is more effective (obvious since they are chosen based on EPA). But for some coverages these optimal routes are more effective than for others (see EDA plot, receivers are not always the best position player to target for a given coverage).
- The cover schemes (as opposed to combination schemes, Tampa 2, ect) are the most effective schemes at limiting first downs

The Answer to the Question

Optimal route combination against:

Cover 0: [Go/Fly, Curl, Swing-right,Deep Cross]

Cover 1: [Go/Fly, Deep Cross, Pick, Flat Left]

Cover 2: [Go/Fly, Seam, Curl, Chip-Flat]

Tampa 2: [over ball, flat-right, Go/fly, curl]

Cover 3: [run fake, Go/fly, Seam, swing-left]

Combination: [Curl, Slant, Corner]

Cover 4: [curl, curl, flat-right, corner post]

Screen: [corner, fade, jet sweep pass]

Cover 6: [comeback, Go/fly, Chuck & release]

Man Cover 2: [slant, Go/fly, Post, Curl]

Prevent: [chuck & release, Go/fly, Deep Cross, Dig]

Limitations and Improvements

There are several ways we could improve this analysis. We were also limited by the data.

- No tracking data (given us alternative ways to classify routes)
- We did not take into account how often a receiver plays, or how a good receiver might ‘warp’ the defence
- We did not take into account how running a play multiple times decreases its effectiveness
- We didn’t look at different formation types (21 personnel v.s 12 personnel for example)
- We did not account for the strength of a given DB and how that might affect plays run by the offense
- Is a route effective or is the defense conceding the play. For example, if I get an 8 yard gain against a ‘prevent’ coverage; is this an effective play or am I doing what the defense wants me to do.

References: Additional Datasets Used

Madden ratings: https://maddenratings.weebly.com/

Strength indicators: https://www.pro-football-reference.com/years/2020/index.htm

That’s all Folks!