"The only sign we have in the locker room is from 'The Art Of War': Every battle is won before it is fought."

- Bill Belichick.

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)
- A receiver is anyone who runs a route - including TEs and RBs

What position excels against what coverage?

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. This makes sense since TEs are probably being covered by LBs in these schemes.
- We also see how the various personnel groupings fare against a given coverage; for example 21 personnel is most efficient at the combination and tampa 2 coverages (and useless for the cover type - i.e cover 0, cover 1 coverages)

EPA for Different Receiver Positions for a given Coverage Scheme

EPA by Personnel against a given Coverage Scheme
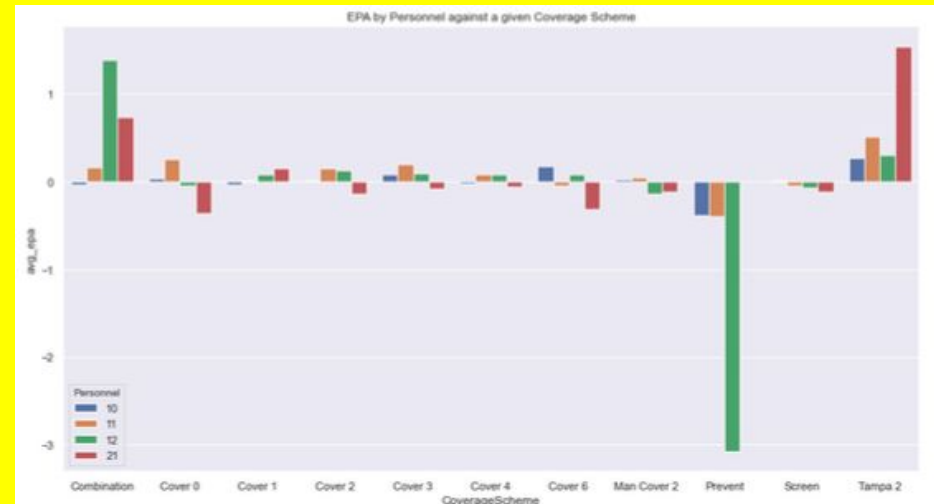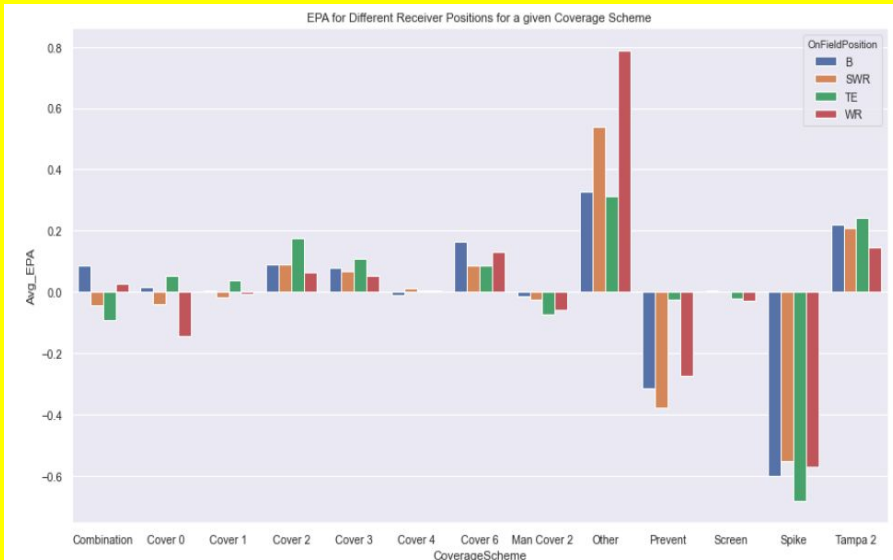
How to tell which routes are the best?

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
```

We also looked at most common routes by personnel grouping:

```
Most Common Routes 10 personnel
[(('Run Fake', 'rpo'), 48), (('Curl', 'Curl', 'Curl', 'Curl'), 36), (('rpo', 'Run Fake'), 36), (('Curl', 'Curl', 'Curl', 'Curl', 'Curl'), 35), (('Screen - RB',), 24)]

Least Common Routes 10 personnel
[(('Go/Fly', 'Screen - RB'), 2), (('Screen - Shovel',), 1), (('rpo',), 1), (('Fade',), 1)]
-------------------------------------------------------------------------

Most Common Routes 11 personnel
[(('rpo', 'Run Fake'), 70), (('Run Fake', 'rpo'), 44), (('Jet Sweep Pass',), 18), (('Run Fake', 'Screen - Tunnel'), 16), (('Out', 'Out', 'Curl', 'Out', 'Curl'), 15)]

Least Common Routes 11 personnel
[(('Swing - Right', 'Screen - RB'), 2), (('Go/Fly', 'Screen - TE'), 2), (('Screen - Tunnel', 'Flat - Right'), 2), (('Screen - Drag',), 1)]
-------------------------------------------------------------------------

Most Common Routes 12 personnel
[(('rpo', 'Run Fake'), 44), (('Run Fake', 'rpo'), 24), (('Curl', 'Check & Release', 'Curl'), 18), (('Curl', 'Curl', 'Flat - Right', 'Flat - Left'), 16), (('Curl', 'Curl', 'Flat - Left'), 15)]

Least Common Routes 12 personnel
[(('Screen - Bubble', 'Run Fake'), 2), (('Screen - RB', 'Deep Cross'), 2), (('Slant',), 1), (('Flat - Left',), 1)]
-------------------------------------------------------------------------

Most Common Routes 21 personnel
[(('rpo', 'Run Fake'), 10), (('Run Fake', 'Whip', 'Comeback', 'Deep Cross'), 8), (('Curl', 'Run Fake', 'Chip - Flat', 'Dig'), 8), (('Curl', 'Curl', 'Flat - Right', 'Flat - Left'), 8), (('Run Fake', 'Post', 'Out'), 6)]

Least Common Routes 21 personnel
[(('Post', 'Screen - RB'), 2), (('Slant', 'Out'), 2), (('Slant', 'Slant'), 2), (('Post', 'Quick'), 2)]
```
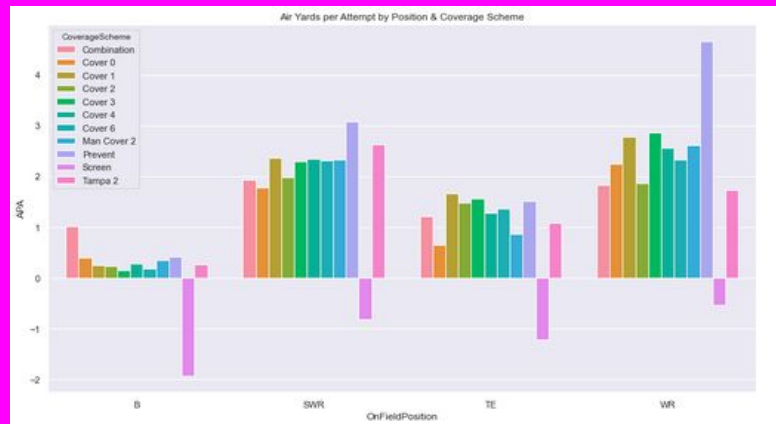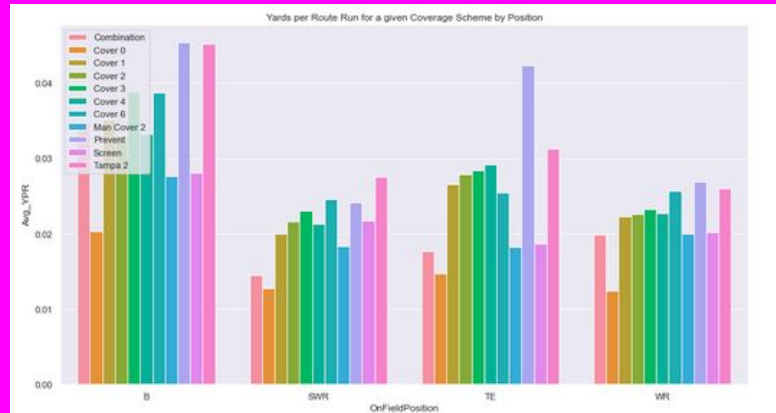
Whose catching the Ball?

How do you know what position to throw the ball to? Seems like an easy answer- but we just saw last slide that sometimes it's smarter to throw it to a RB or TE. Do we even need WRs?

On the top plot, it looks like WRs don't excel at yards per route run. However, they are on the field more (running routes at least, compared to RBs and TEs who may block) and their catches are more difficult. How do we know that? They catch the ball further down the field (bottom plot).

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
- Receiver class (is a receiver especially good at running long routes, short routes, ect)
- Route length (is a given route long or short)

How long is that route?

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. This was done to simplify looking at the routes- there are less combinations if we replace the route names with the route length.
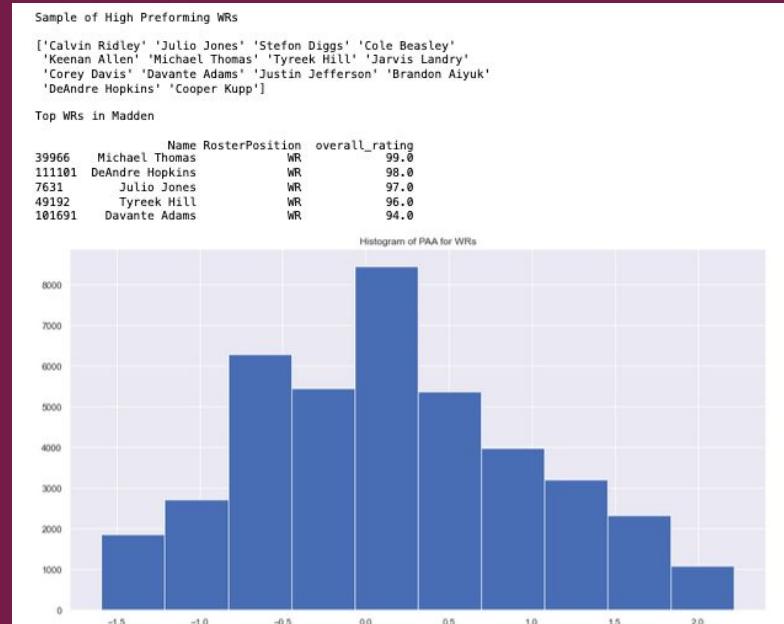
Stacking Coverage Schemes Against One-Another

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)
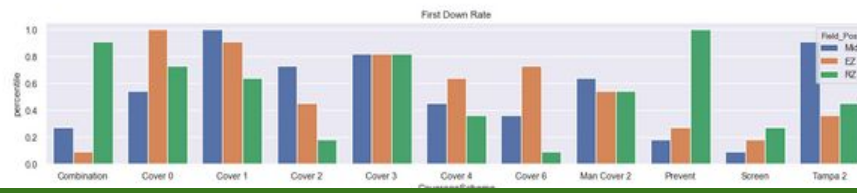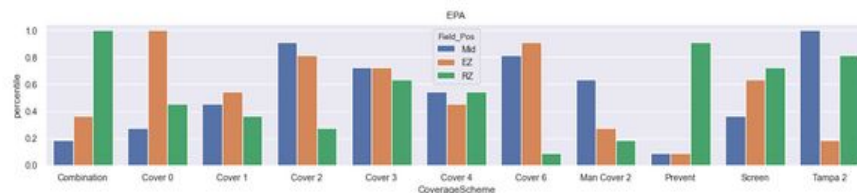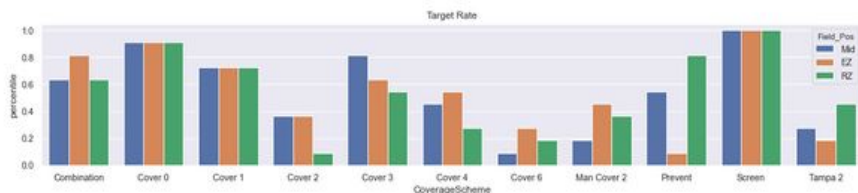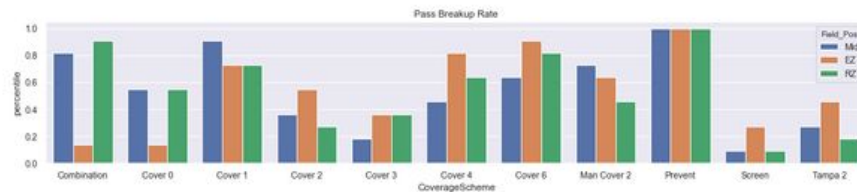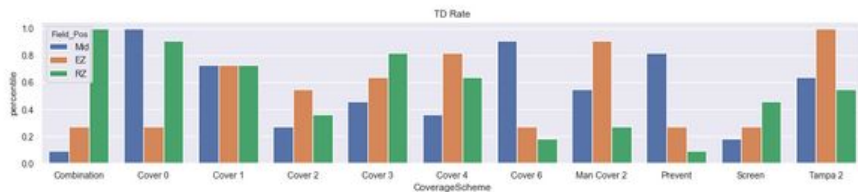
Percentile Plots: Insights

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.
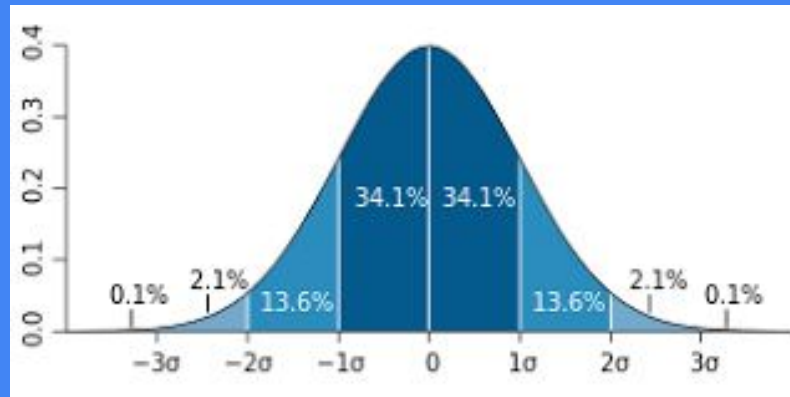
ML Models: General Info

- Logistic regression and random forests regression models used
- Features used:
    - Team stats class: (ex. Point differential) allows the model to determine the difference between the two teams playing (including how good one offense is compared to the opposing defense)
    - Player stats (ex. Points above average) allows the model to distinguish between a good QB throwing to a good receiver v.s a bad QB throwing to a bad receiver
    - Cover Scheme class: coverage scheme statistics such as the average pass break ups or pressure rates for a given coverage scheme
    - Routes class: the route combinations and route lengths used on the play. Including if a receiver is running a 'specialist' route (a route the receiver is much better than the average receiver in running)

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 compared to Mean Offensive Yardage
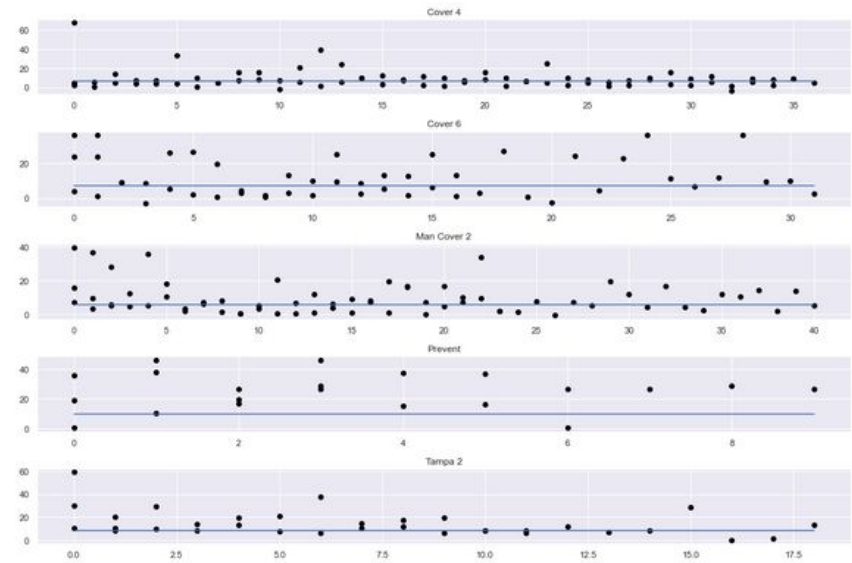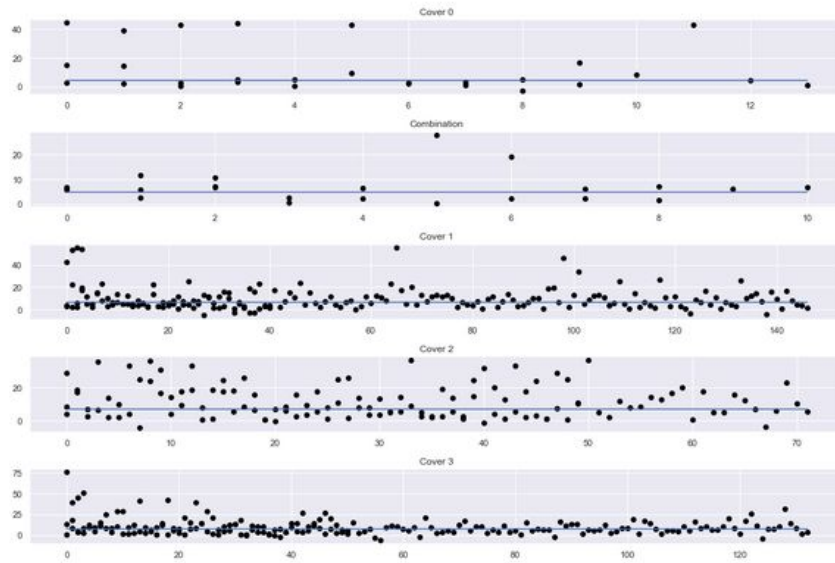
Expected Yards: Insights
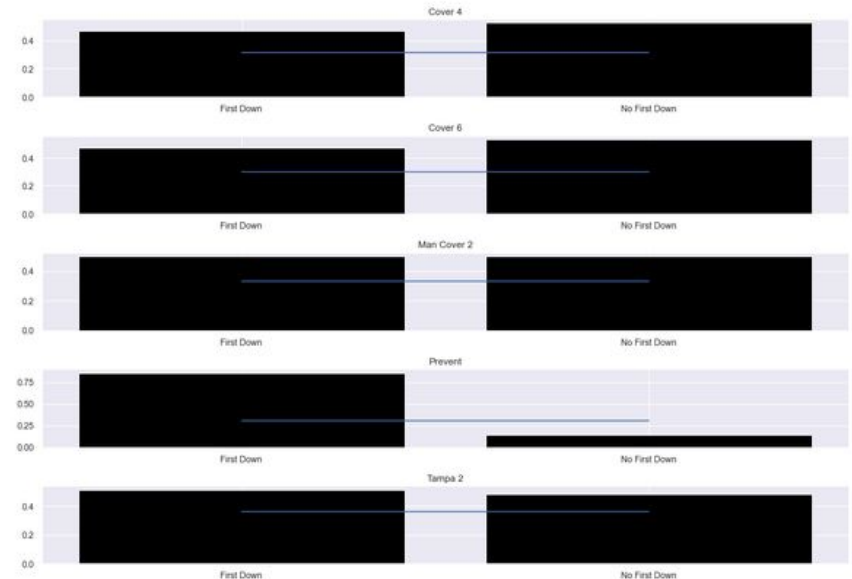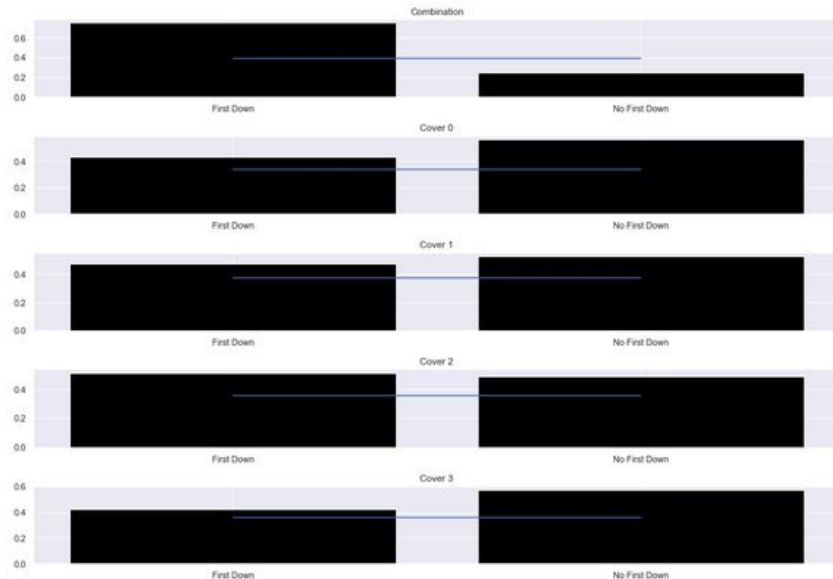
We have determined that:

- 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

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)
- Only 1 year of data
- 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.

$$\lim_{x \to a} \overbrace{f(x)}^{\text{function}} = \underbrace{L}_{\substack{\text{"What is the} \\ \text{y-value getting} \\ \text{closer to?"}}}$$

"As you approach a along the x-axis"

Calcworkshop.com

The Answer to the Question

Optimal route combination against:

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

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

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

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

Combination: [Curl, Slant, Corner]

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

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

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

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

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

Screen: [corner, fade, jet sweep pass]

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!