

# NFL Catch Probability Analysis

---

By: Tyler Parker & Colby Beliveau



# Motivation

---

- Utilize tracking data from 2023 NFL Season provided by the NFL Big Data Bowl to estimate catch probabilities while the ball is in the air, based on certain features calculated before the ball is thrown and while the ball is in the air.
- Ultimately trying to answer the question: How do certain things that receivers and defenders do while the ball is in the air affect the catch probability?

# Data Filtering and Cleaning

- Want to look at throws  $\geq 20$  yards downfield, not negated by a penalty
- Assigned each play a unique identifier by combining game\_id and play\_id
- Converted time left in game to seconds to make it easier to index to specific plays

# Features

- Separation at throw
- Separation at arrival
- Defenders within 3 yards
- Maximum receiver speed during route
- Height difference (Receiver-Defender)
- Distance advantage at throw

catch	sep_at_throw	sep_at_arrival	defenders_within_3	receiver_speed	height_diff	advantage_at_throw	game_play
0	4.7356520	4.4732986	0	16.15905	2	1.8340795	2023090700_101
1	11.0195553	10.4419347	0	16.77269	-1	10.4920850	2023090700_1154
0	0.6356099	0.6768309	2	13.45906	4	0.3455791	2023090700_1422
0	0.8736704	0.9160786	1	13.68406	3	0.3214955	2023090700_1494
0	1.0282509	0.9476286	1	18.04087	-2	-0.5588539	2023090700_1588
1	4.2137988	4.8513606	0	19.00223	4	3.1784139	2023090700_1679

# Fitting Our Model

- Used RandomForest to evaluate each play
- As we expected, some features contribute more than others.

	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
sep_at_throw	10.883584	0.9695287	15.6464477	186.571758
sep_at_arrival	16.046107	7.5696667	30.0845883	199.043828
defenders_within_3	4.353687	-4.9273366	0.8009656	7.864803
advantage_at_throw	4.495792	9.1980064	11.7248659	213.768366
receiver_speed	-6.405054	-5.8336033	-8.5704978	168.917500
height_diff	-4.991480	-2.9481029	-5.7168897	95.713396

# Best Catch

```
best_catch = model_df %>% # Just to provide a little intuition. This was a Russel Wilson hail mary, so makes sense it has a l
  filter(catch == 1) %>%
  arrange(predicted_prob) %>%
  head(1)
kable(best_catch)
```

catch	sep_at_throw	sep_at_arrival	defenders_within_3	receiver_speed	height_diff	advantage_at_throw	game_play	predicted_prob	predicted_class
1	0.7340981	0.6315061	1	15.87269	4	-0.3015153	2023120302_1246	0.0395778	0



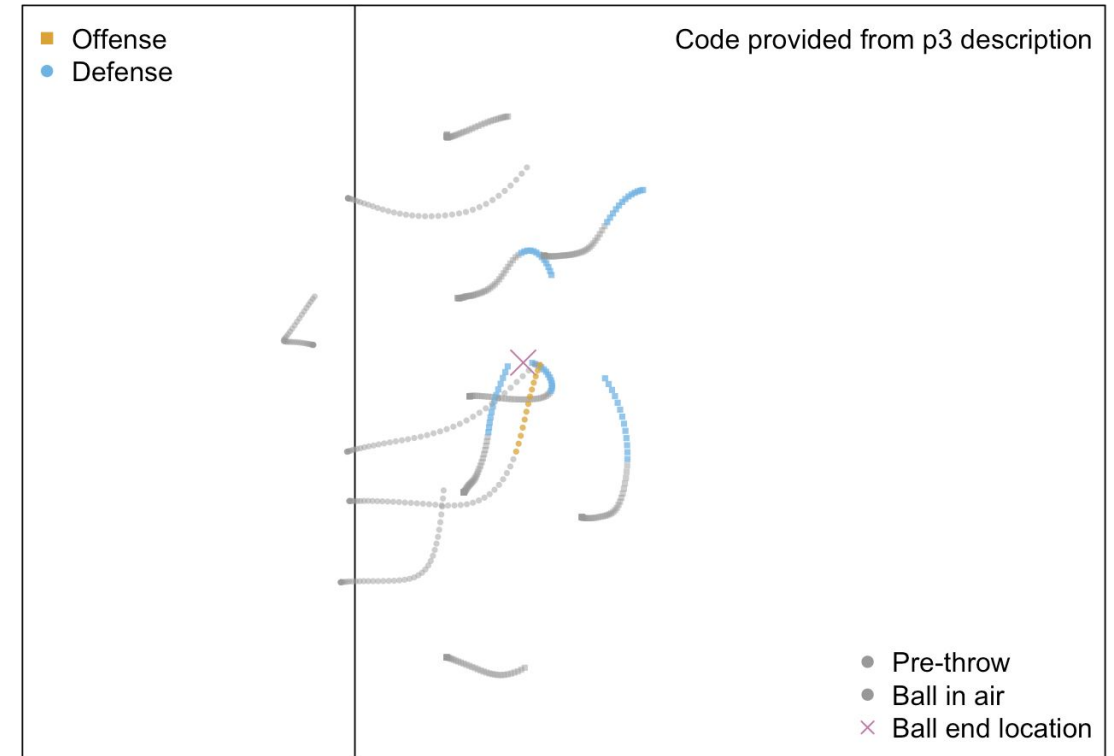
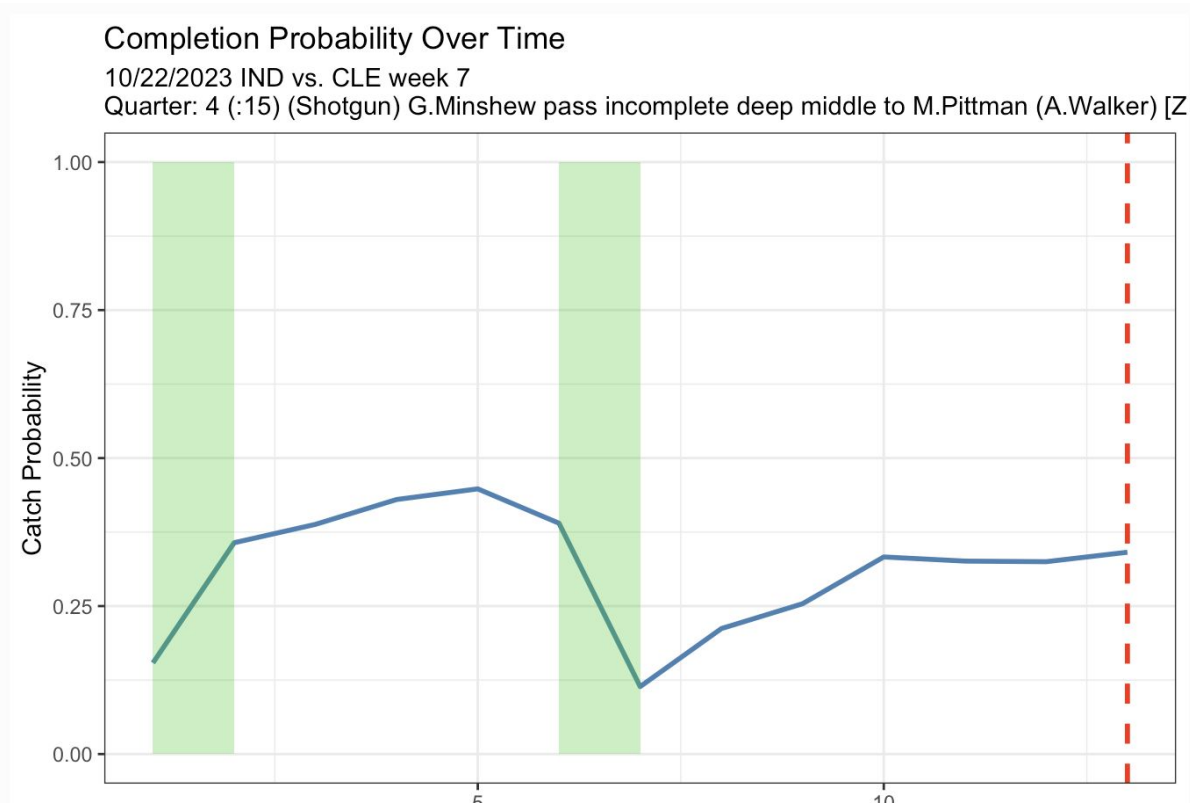
# Evaluate a Single Play

- We can run every single frame of the play through our model to get a running catch probability estimate.

frame_id	separation_at_frame	defenders_within_3	receiver_speed	height_diff	advantage_at_throw	separation_at_throw	catch_prob	prob_change	is_critical	separation_change
1	3.639025	0	15.54542	3	-0.4046564	3.527719	0.154	NA	NA	NA
2	3.754211	0	15.54542	3	-0.4046564	3.527719	0.357	0.203	TRUE	0.1151864
3	3.847662	0	15.54542	3	-0.4046564	3.527719	0.388	0.031	FALSE	0.0934507
4	3.940013	0	15.54542	3	-0.4046564	3.527719	0.430	0.042	FALSE	0.0923511
5	4.002249	0	15.54542	3	-0.4046564	3.527719	0.448	0.018	FALSE	0.0622367
6	3.855191	0	15.54542	3	-0.4046564	3.527719	0.390	-0.058	FALSE	-0.1470581



# Visualization of a play



# Critical Moments

[Show](#)

frame_id	separation_at_frame	defenders_within_3	catch_prob	separation_change
2	3.754211	0	0.357	0.1151864
7	3.288966	0	0.114	-0.5662249

- By looking at the critical moment features, we can see how an increase in separation raises the catch probability at frame 2
- And the decreasing separation lowers the probability at frame 7
- Shown in the play visualization we also see the defenders path to curl around and intercept the receiver while the ball is in the air, closing the separation gap in frame 7 and ultimately leading to an incomplete pass outcome

# Conclusion

- A limitation worth noting is the exclusion of quarterback accuracy.
- Not incorporating this in our model, however, allows us to interpret our results in a much more broad sense
- We can essentially say, that regardless of who is throwing the ball, this is the catch probability given these play and frame features.
- Overall, this project demonstrates how spatial analytics and model training methods can quantify the likelihood of a completion on long passing plays, identify the frames in which the probability meaningfully shifts, and show how defensive closing speed or receiver route adjustments affect play outcomes.
- These insights provide a foundation for evaluating route effectiveness, defender coverage quality, and the dynamic nature of passing plays.
- Future additions to this model could incorporate throw trajectory, accuracy, and quarterback decision-making to produce a more holistic model of passing success.

