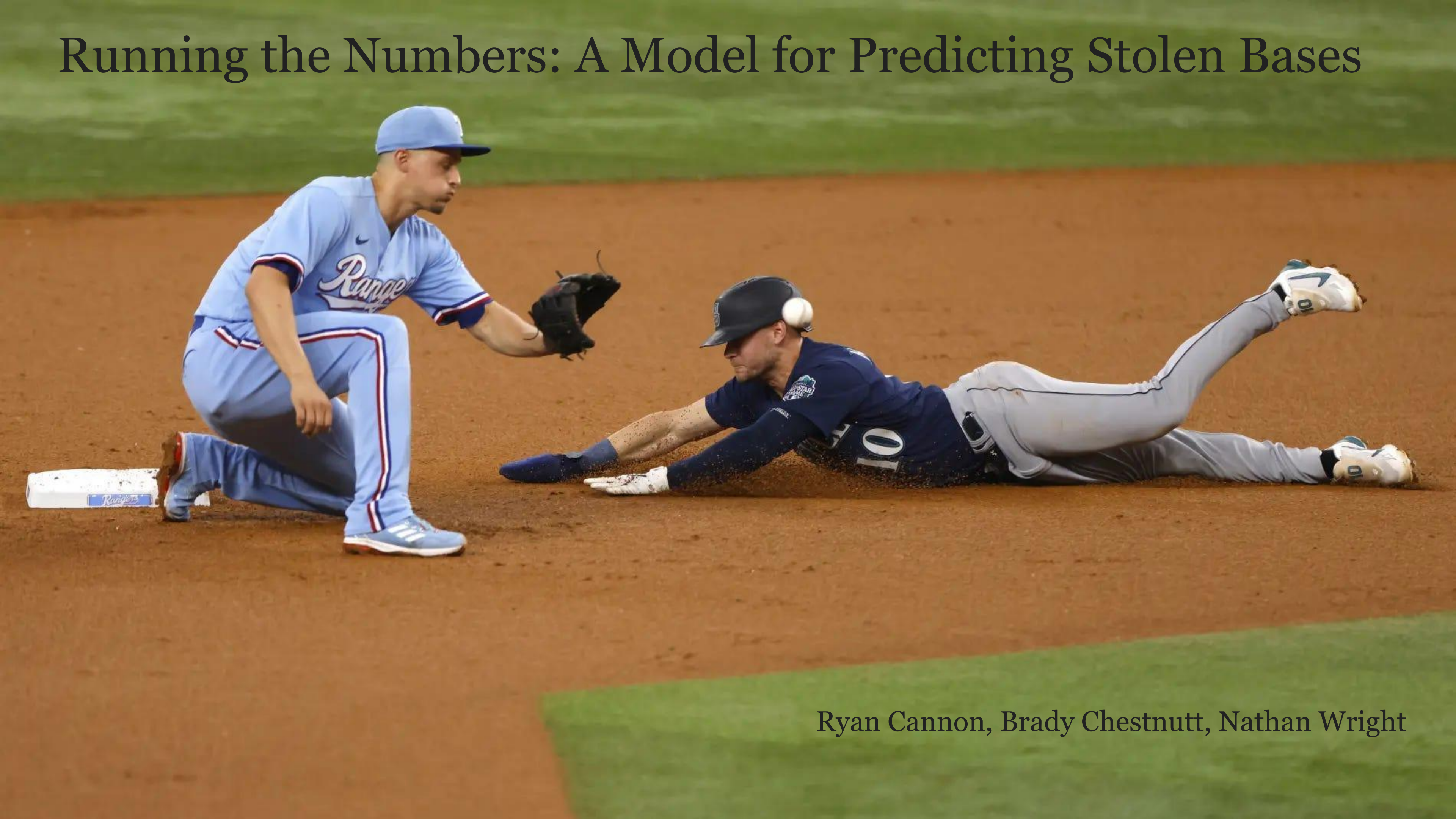
Running the Numbers: A Model for Predicting Stolen Bases



Ryan Cannon, Brady Chestnutt, Nathan Wright

Overview

In this project, we developed a model to estimate the probability of a successful stolen base attempt using Statcast data from the 2018 to 2024 MLB seasons. The model incorporates key metrics from the runner, pitcher, and catcher to evaluate each situation in context.

- Primary Goal: Create a more informed, data-driven approach to determine whether a runner should attempt to steal a base.

Methodology

- Develop a program to identify which metrics most strongly correlate with a runner's actual stolen base success rate
- Use these correlations to assign weights and compute a final success probability
- Generate a recommendation based on the probability and game context.

```
Attempting to steal 2nd base:
Runner: Nootbaar, Lars (Sprint Speed: 26.2 ft/s)
Pitcher: Williams, Devin (Delivery Time: 1.98s)
Catcher: Smith, Will (Pop Time: 1.94s)
Success Probability: 77.0%
Recommendation: GO!

Attempting to steal 3rd base:
Runner: Nootbaar, Lars (Sprint Speed: 26.2 ft/s)
Pitcher: Williams, Devin (Delivery Time: 1.98s)
Catcher: Smith, Will (Pop Time: 1.94s)
Success Probability: 72.2%
Recommendation: HOLD
```

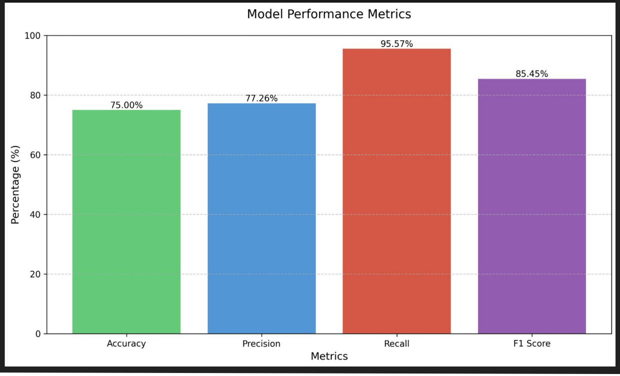
Factor Weights

For 2nd Base (2B):

- historical_success_rate: 0.155 (15.5%)
- sprint_speed: 0.280 (28%)
- pop_time: 0.140 (14%)
- arm_strength: 0.160 (16%)
- delivery_time: 0.113 (11.3%)
- pickoff_tendency: 0.038 (3.8%)
- Lead: 0.114 (11.4%)

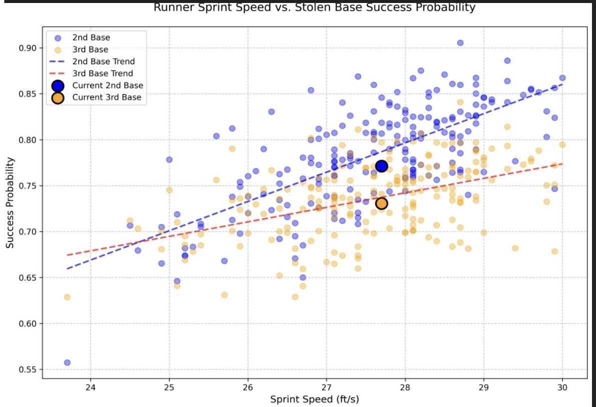
For 3rd Base (3B):

- historical_success_rate: 0.147 (14.7%)
- sprint_speed: 0.210 (21%)
- pop_time: 0.280 (28%)
- arm_strength: 0.130 (13%)
- delivery_time: 0.063 (6.3%)
- pickoff_tendency: 0.038 (3%)
- lead: 0.140 (14%)

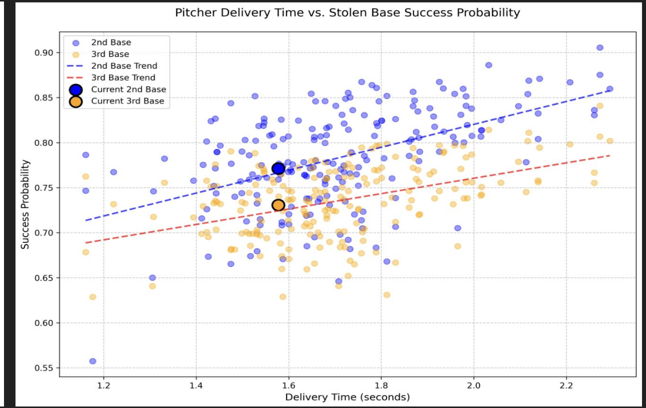


Analysis Results

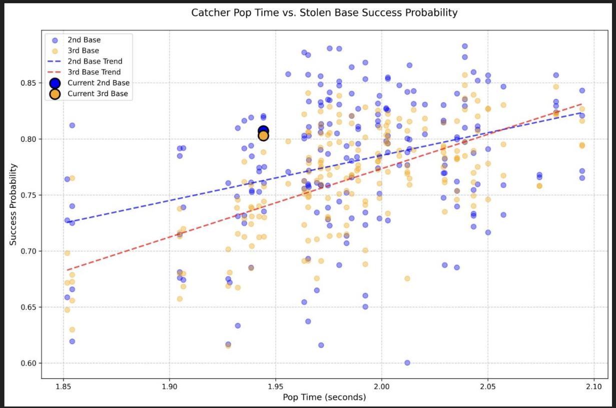
Sprint Speed Effect on Probability [Slope]



Pitcher Delivery Time Effect on Probability [slope]

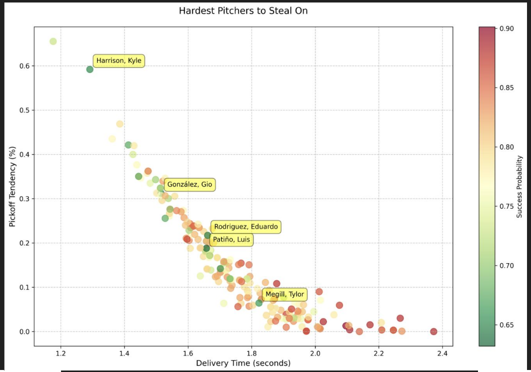


Catcher Pop Time Effect on Probability [slope]

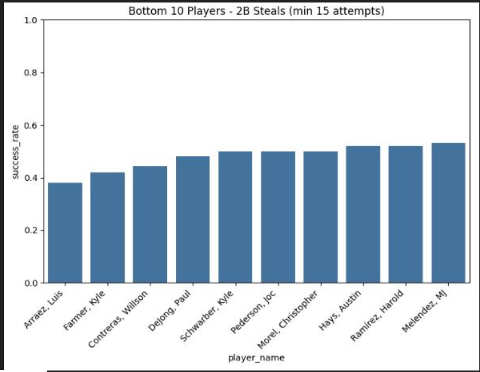


Results (cont.)

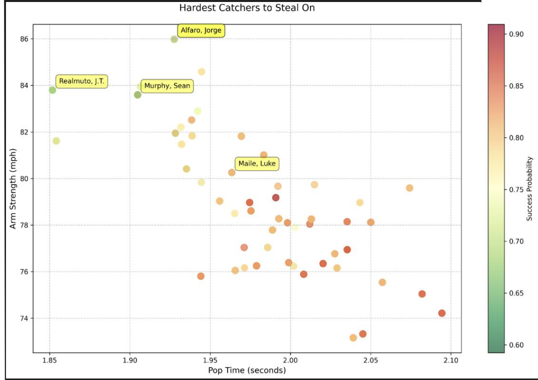
Hardest Pitchers to Steal On



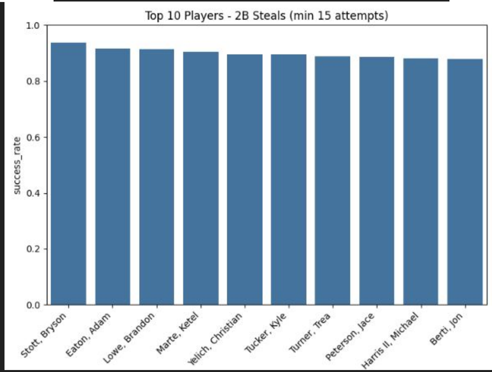
Bottom 10 Players in Steal Success Rate



Hardest Catchers to Steal On



Top 10 Players in Steal Success Rate



Limitations

Using play-specific metrics instead of aggregated averages may have improved model training and led to more accurate predictions.

- Missing data in savant dataset
- Weather affecting different factors
- Doesn’t account for hitter at the plate
- Doesn’t account for type of pitch thrown by pitcher

Application & Future Expansion

Future Expansion

- Integrate deeper situational context:
  - Batter at the plate - Probability of Ball in Play
  - Score differential
  - Pitcher’s pickoff tendencies
  - Likelihood of tipped pitches or sign stealing

Application

- Live, in-game decision-making on whether to send runners
- Identify players who are under- or over-utilizing their stealing potential
- Apply game theory: anticipate and exploit opposing teams' steal expectations

Thank You!