



NFL 1st and Future - Analytics

Can you investigate the relationship between the playing surface and the injury and performance of NFL athletes?

Author: Aleksandra Deis

📍 Sunnyvale, CA, USA

✉️ lexie@live.ru

📞 +1 650 476-3386





The Challenge

1

Examine the **effects that playing on synthetic turf versus natural turf** can have on player movements:

- characterize any differences in player movement between the playing surfaces

2

Analyze the **factors that may contribute to lower extremity injuries**

- identify specific scenarios (e.g., field surface, weather, position, play type, etc.) that interact with player movement to present an elevated risk of injury

[Competition's Homepage](#)



The Dataset



INJURY RECORD

Information on **105 lower-limb injuries** that occurred during regular season games over the two seasons: body part, surface, days missed due to the injury

Injuries can be **linked to specific records** in a player history



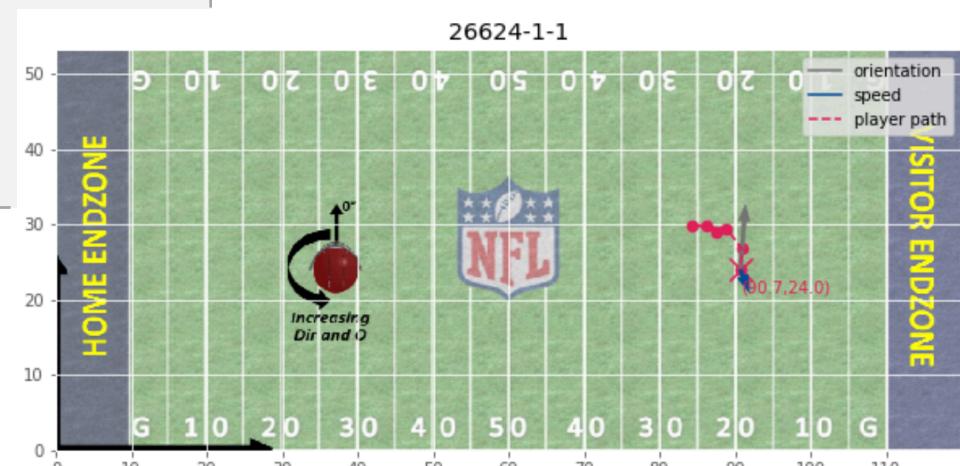
PLAY LIST

Details for the **267,005 player-plays** including: the player's assigned roster position, stadium type, field type, weather, play type, position for the play, and position group



PLAYER TRACK DATA

Player level data that describes the **location, orientation, speed, and direction of each player** during a play recorded at 10 Hz (i.e. 10 observations recorded per second)



An example of the player track visualization ([image url](#))



The Approach

I took the **statistical approach based on hypothesis testing**

- ✓ provides formal justification to the conclusions
- ✓ is not impacted by the number of injury records

MAJOR CHALLENGE:

- The dataset contains **267 005** player-plays records for **250** players and only **105** injury records
- **Less than 0.04%** of play records are associated with injury

CONSEQUENCES:

- The correlation analysis (Pearson correlation coefficient, Cramer's V, Theil's U) **doesn't show any correlation**, because any possible correlation is vanished by the number of records not related to the injury
- **Nearly impossible to build a predictive model**
 - Models converge to no injury prediction for all input
 - Resampling changes the priors significantly and doesn't help to get the accurate predictions



Player Movements and Playing Surfaces (1)

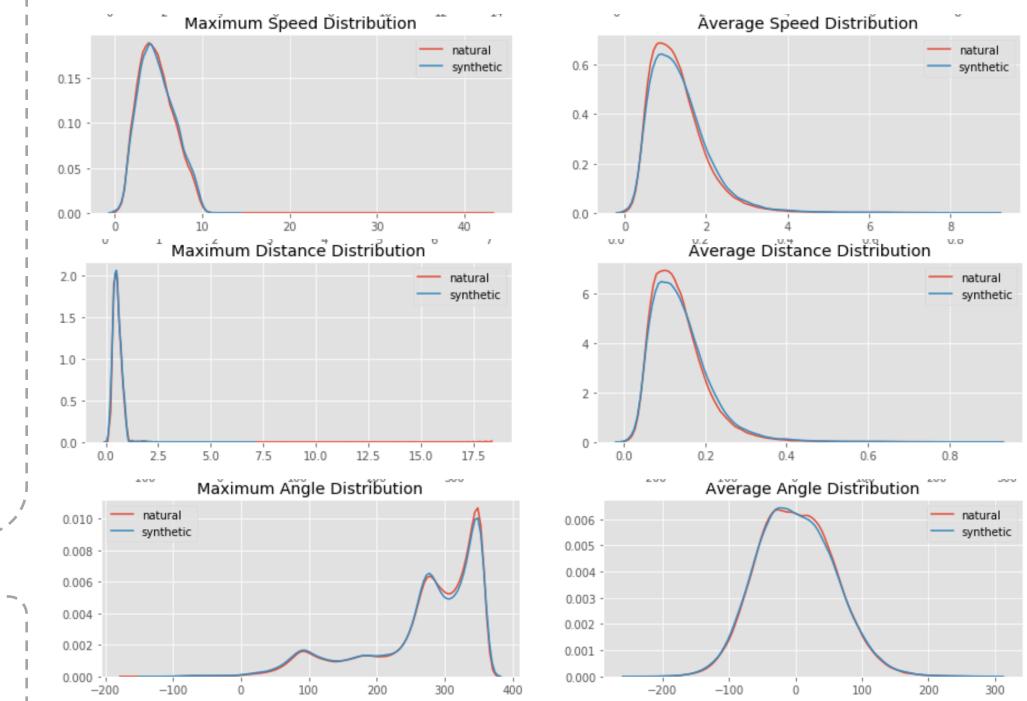
In general, **the playing surface does not affect main movement features**

MOVEMENT FEATURES:

- I analyzed the following features, which characterize movement:
 - The maximum and the average **speed** per play
 - The maximum and the average **distance** covered per play
 - The maximum and the average **angle between direction and orientation** of the player per play

DISTRIBUTION:

- The main movement features have **identical distributions** on both natural and synthetic turf



The distribution of movement features on different playing surfaces



Player Movements and Playing Surfaces (2)

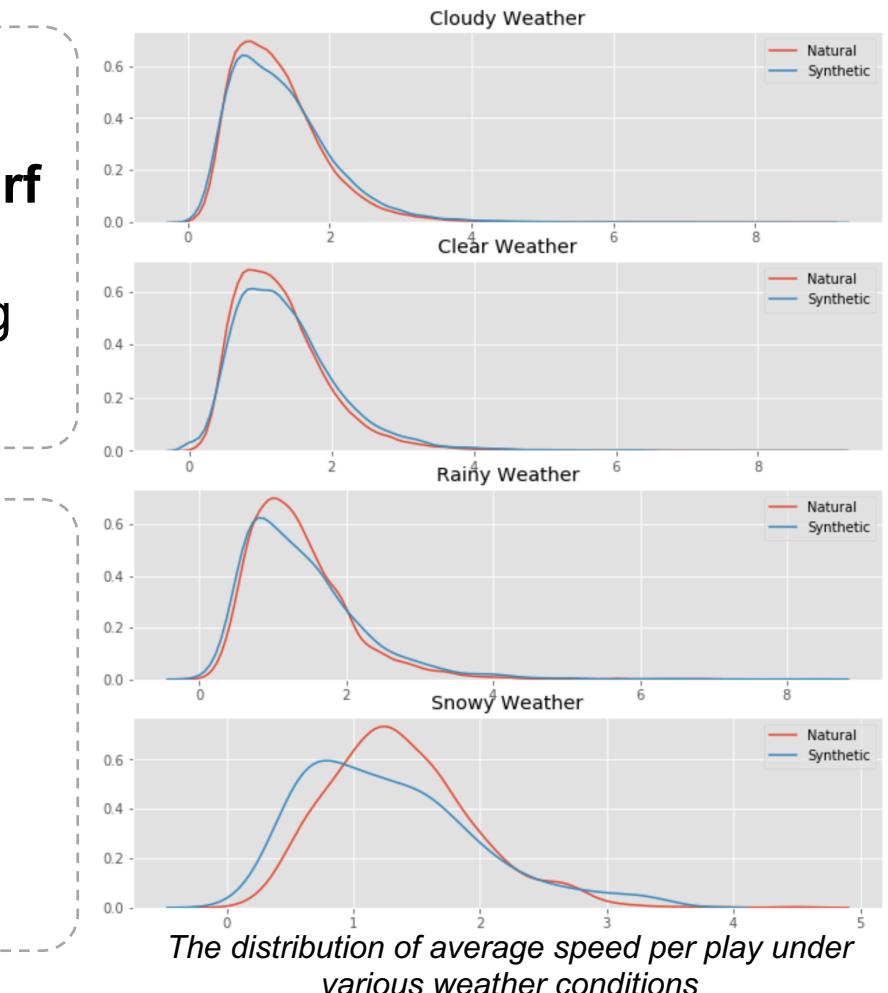
The players tend to have **lower average speed per play on the synthetic turf under severe weather conditions like rain or snow**

WEATHER CONDITIONS:

- At least **68% of games are played outdoors**, and almost **17% of games are played outdoors on the synthetic turf**
- Are there any differences in player movements under extreme weather conditions like rain or snow when playing on synthetic turf?

DISTRIBUTION:

- The maximum and the average player speed for different surfaces come from **different distributions when the weather is rainy or snowy**
- The p-value under the null hypothesis of the Kolmogorov-Smirnov two-sided test is less than 1% for the average speed samples when rainy or snowy





Risk of Injury on Synthetic vs Natural Turf (1)

There is a higher risk of injury when playing on the synthetic turf

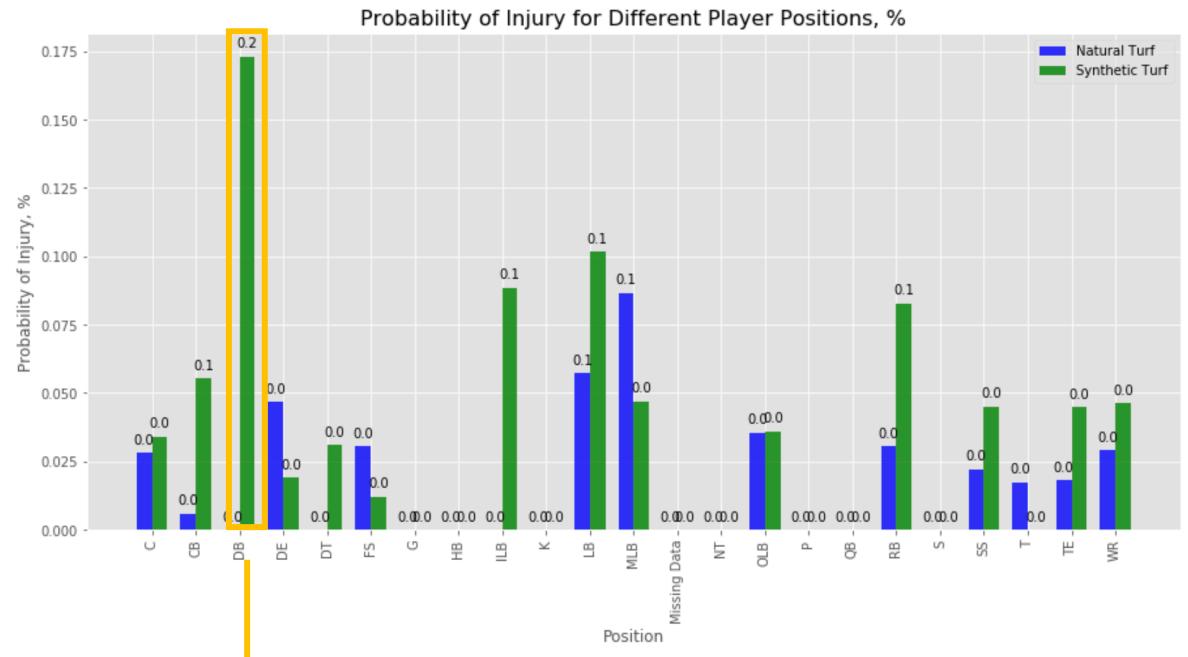
- The risk of injury on the synthetic turf (1.4%) is **63% higher** than the risk of injury on the natural turf (2.4%)
- The hypothesis testing (the two-sided z-test) proves that the **higher risk of the injury on the synthetic turf is statistically significant** (the p-value is less than 2%)
- There is **no statistical proof of a higher risk of injury under certain weather conditions** for certain field surface types



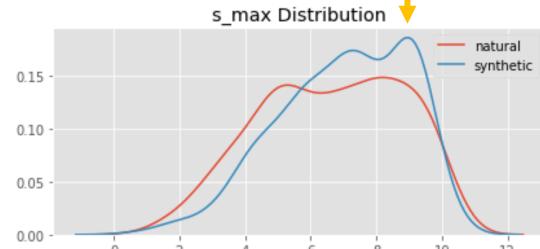
Risk of Injury on Synthetic vs Natural Turf (2)

The players have higher speed on synthetic surface in DB position, which might cause elevated risk of injury

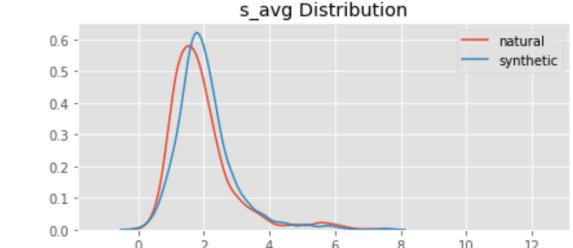
- We can observe that there is a **higher risk of injury for synthetic turf for certain player positions**: DB (defensive back), ILB (inside linebackers)
- We **could not prove that this elevated risk is statistically significant** with hypothesis testing, but we can look closer at the movement features for players in DB position
- I discovered that **the maximum and average speed for players in DB position come from different distributions depending on the field surface** (the p-value under the null hypothesis of the Kolmogorov-Smirnov two-sided test is less than 1%)



The risk of injury for different player positions



The distribution of maximum and average speed per play for injured players in DB position





Conclusions

- There is a **higher risk of injury when playing on the synthetic turf**
- **Under severe weather conditions (rain or snow) players on the synthetic turf tend to have a lower speed** than the players on natural turf
- **Probably, the synthetic turf can't sustain the high speed.** It makes the players slow down when the weather is bad. Probably, moving too fast on synthetic turf could even raise the risk of injury when playing in a certain position



References

- [Link to the Kaggle notebook](#) with the analysis
- Images:
 - [NFL logo image source](#)
 - [Cover image source](#)