**Project Report**

**On**

**NFL Surface Analytics**



Submitted in partial fulfillment for the award of

**Post Graduate Diploma in High Performance Computing System Administration** from **C-DAC ACTS (Pune)**

**Guided by:**

**Mr. Prateek Maheshwari**

Presented by:

**Mr. Rohit Deshmukh PRN: 190840125007**

**Mr. Mrinal Raghav PRN: 190840125023**

### Mr. Prateek Tiwari PRN: 190840125035

### Mr. Vipin Soni PRN: 190840125054

**Centre of Development of Advanced Computing (C-DAC), Pune**



CERTIFICATE

### TO WHOMSOEVER IT MAY CONCERN

**This is to certify that**

**Mr. Rohit Deshmukh**

**Mr. Mrinal Raghav**

**Mr. Prateek Tiwari**

**Mr. Vipin Soni**

**have successfully completed their project on**

On

NFL Surface Analytics

# under the guidance of Mr. Prateek Maheshwari

**Project Guide Project Supervisor**

### HOD ACTS

**Mr. Aditya Kumar Sinha**



**ACKNOWLEDGEMENT**

This project “**NFL Surface Analytics**” was a great learning experience for us and we are submitting this work to Advanced Computing Training School (CDAC ACTS).

We all are very glad to mention the name of *Mr. Prateek Maheshwari* for his valuable guidance to work on this project. His guidance and support helped us to overcome various obstacles and intricacies during the course of project work.

We are highly grateful to Ms. Risha P.R. (Manager (ACTS training Centre), C-DAC), for her guidance and support whenever necessary while doing this course Post Graduate Diploma in *Big Data Analytics (PG- DBDA)* through C-DAC ACTS, Pune.

Our most heartfelt thanks goes to *Ms. Shilpi Shalini* (Course Coordinator, PG- *DBDA*) who gave all the required support and kind coordination to provide all the necessities like required hardware, internet facility and extra Lab hours to complete the project and throughout the course up to the last day here in C-DAC ACTS, Pune.

**Submitted by:**

Mr. Rohit Deshmukh (190840125007)

Mr. Mrinal Raghav (190840125023)

Mr. Prateek Tiwari (190840125035)

Mr. Vipin Soni (190840125054)

**TABLE OF CONTENTS**

1. **Abstract**
2. **Introduction and Overview of Project**
3. **Code**
4. **Data Visualization**
5. **Conclusion**
6. **Precautions**
7. **Bibliography**

# Abstract

The National Football League is America's most popular sports league, comprised of 32 franchises that compete each year to win the Super Bowl, the world's biggest annual sporting event. Founded in 1920, the NFL developed the model for the successful modern sports league, including national and international distribution, extensive revenue sharing, competitive excellence, and strong franchises across the country.

The NFL is committed to advancing progress in the diagnosis, prevention and treatment of sports-related injuries. The NFL's ongoing health and safety efforts include support for independent medical research and engineering advancements and a commitment to work to better protect players and make the game safer, including enhancements to medical protocols and improvements to how our game is taught and played. As more is learned, the league evaluates and changes rules to evolve the game and try to improve protections for players. Since 2002 alone, the NFL has made 50 rules changes intended to eliminate potentially dangerous tactics and reduce the risk of injuries.

Now, challenge was to help them examine the effects that playing on synthetic turf versus natural turf can have on player movements and the factors that may contribute to lower extremity injuries. NFL player tracking, also known as Next Gen Stats, is the capture of real time location data, speed and acceleration for every player, every play on every inch of the field.

As part of this challenge, the NFL has provided full player tracking of on-field position for 250 players over two regular season schedules. One hundred of the athletes in the study data set sustained one or more injuries during the study period that were identified as a non-contact injury of a type that may have turf interaction as a contributing factor to injury. The remaining 150 athletes serve as a representative sample of the larger NFL population that did not sustain a non-contact lower-limb injury during the study period. Challenge was to characterize any differences in player movement between the playing surfaces and identify specific scenarios (e.g., field surface, weather, position, play type, etc.).

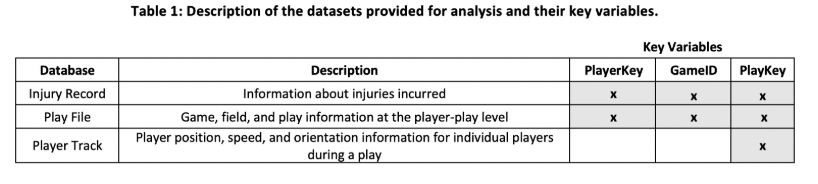


# Introduction and Overview

The datasets and variables provided to examine the effects that playing on synthetic turf versus natural turf can have on player movements and the factors that may contribute to lower extremity injuries. The data provided for analysis are 250 complete player in-game histories from two subsequent NFL regular seasons. Three different files in (.csv) format are provided, documenting injuries, player-plays, and player movement during plays. This manual describes the specifics of each variable contained within the datasets as well as guidelines on the best approach to processing the information.

There are three files provided in the dataset, as described below:

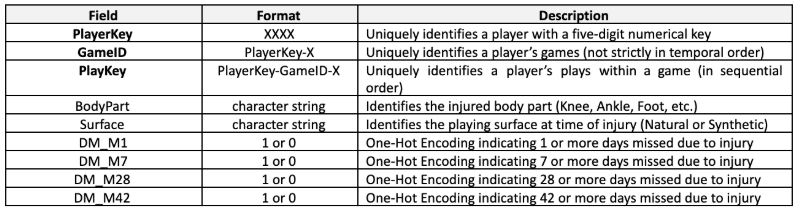
* **Injury Record**: The injury record file in .csv format contains information on 105 lower-limb injuries that occurred during regular season games over the two seasons. Injuries can be linked to specific records in a player history using the PlayerKey, GameID, and PlayKey fields.
* **Play List**: – The play list file contains the details for the 267,005 player-plays that make up the dataset. Each play is indexed by PlayerKey, GameID, and PlayKey fields. Details about the game and play include 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).



## **Field and Key Definitions**

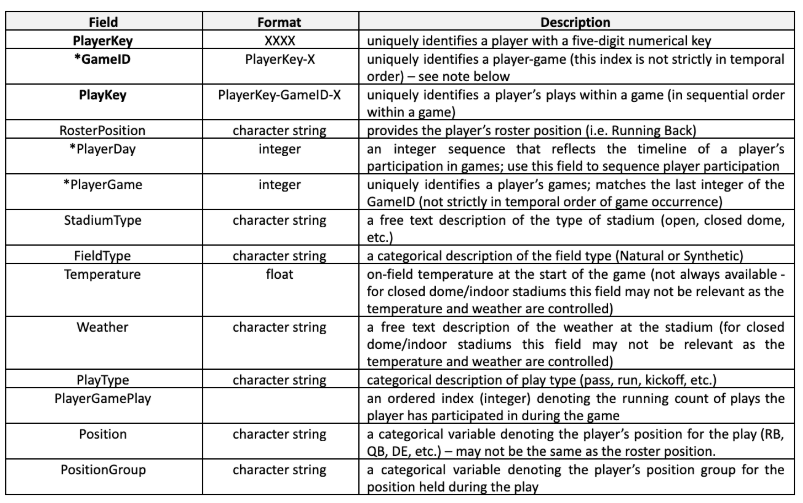
The following provides a description of each field contained within the datasets and their corresponding formats and descriptions. **Key Variables** are designated in bold.

### Injury Data File

Note that there is not a PlayKey available for every injury.

### Play List

The play file contains information about each player-play in the dataset, to include the player’s assigned roster position, stadium type, field type, weather, play type, position for the play, and position group



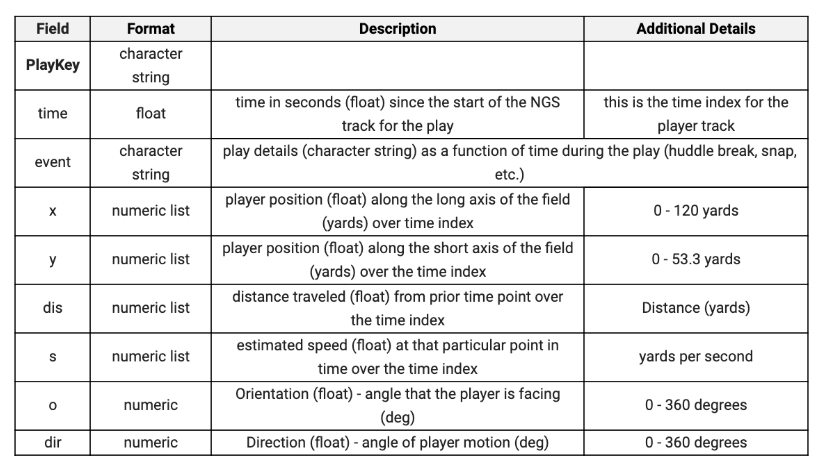
\*Important Note: The GameID field is a unique identifier of player games but does not strictly reflect the order in which the games were played. The PlayerDay is an integer sequence that provides an accurate timeline for player game participation. In order to generate an accurate timeline of an individual player’s game participation, the PlayerDay variable should be used. The interval between days in the PlayerDay field for an individual player accurately reflects the interval in days between that player’s participation in games. Every player has a PlayerDay = 1 (note that this date is not the same for all players). Some players may have negative values for PlayerDay, which simply indicates participation in a game that occurred before their individually assigned PlayerDay = 1.

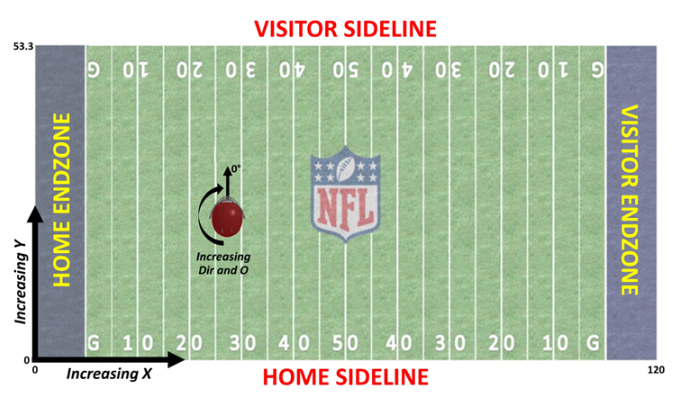
### 3. Player Track Data

The player track file in .csv format includes player position, direction, and orientation data for each player during the entire course of the play collected using the Next Gen Stats (NGS) system. This data is indexed by PlayKey (which includes information about the player and game), with the time variable providing a temporal index within an individual play.

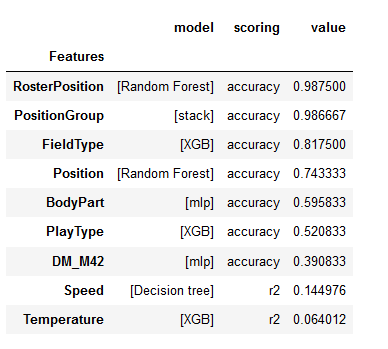
When processing the player track data, it is recommended to calculate velocity using the x, y position data and use those calculated velocities for any analysis (although we have provided the speed variable reported by the NGS system). The origin for the x and y coordinates is defined as the corner of the home endzone and home sideline (see Figure 1). The angles defined by orientation and direction are referenced from the y-axis of the coordinate system.

Note that the orientation variable should not be considered to be a reliable indicator of the actual direction a player is facing. The records for this study come from multiple seasons of the NFL during which different systems were used to calculate and record a player’s orientation. Within a play, and across plays in a game, an individual player’s orientation will be recorded consistently, but the “geography” (i.e. the actual direction the player is facing) of the reported orientation may not match the “geography” of the direction variable in the same play. For those players that participated in multiple seasons, the geography used to record the orientation variable is not consistent across seasons. The orientation variable can be used to characterize how much a player is turning or pivoting during the course of a play. The “geography” of the direction variable does remain consistent across the study horizon.

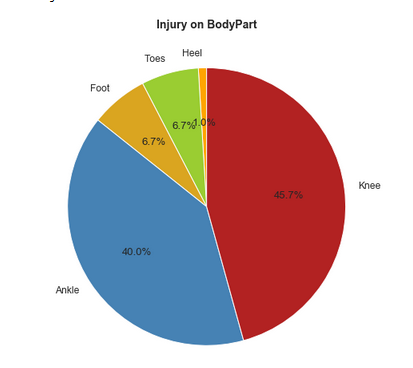


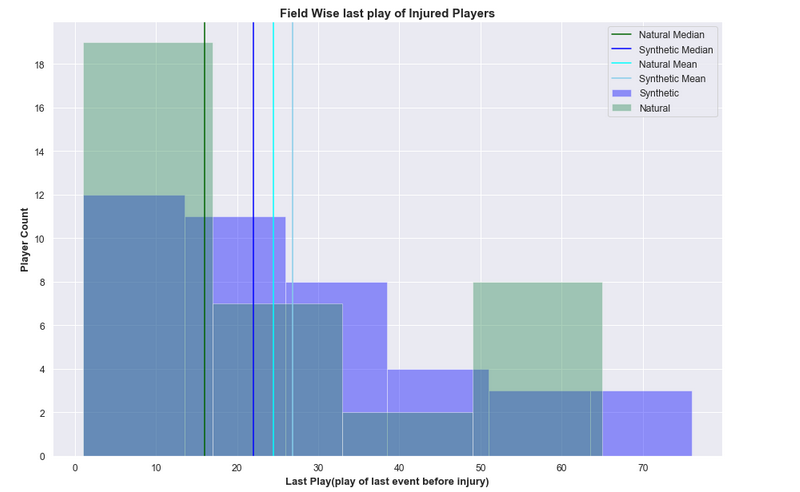


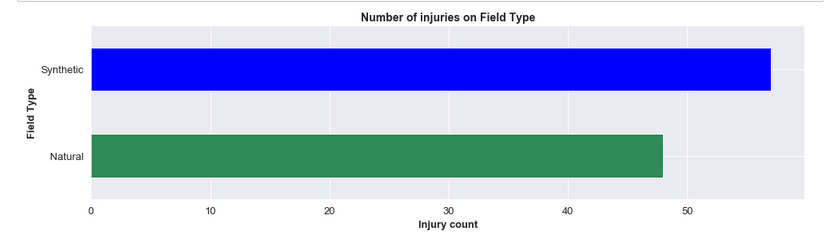
1. **Model Evaluation**

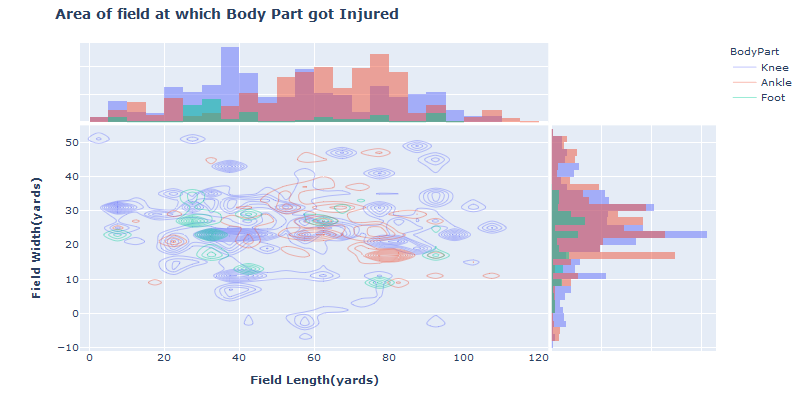


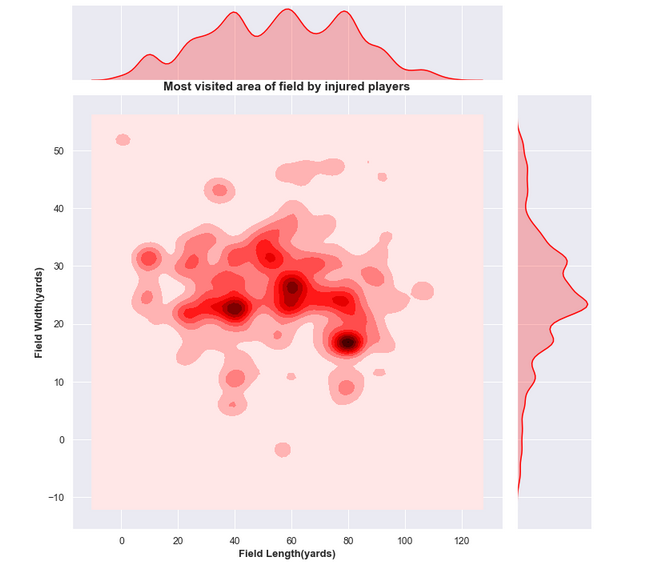
**4. Data Visualization**

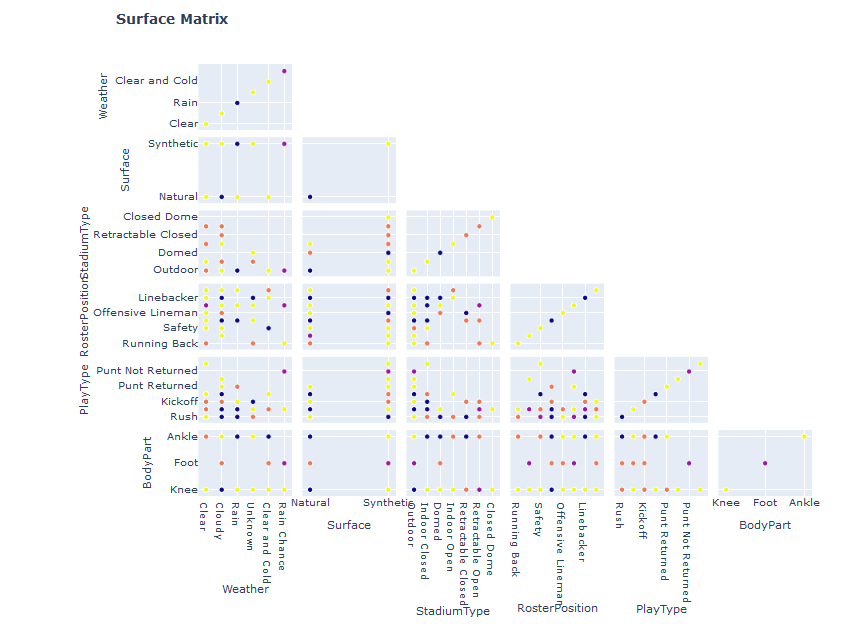












# Conclusion

# Final Comparisons between Synthetic and Natural playing Surface:

1. Maximum Injuries(body part) on Synthetic Surface -> Ankle

Maximum Injuries (body part) on Natural Surface -> Knee

1. Most Injured Roster Position on Synthetic -> Linebacker & Wide Receiver

Most Injured Roster Position on Synthetic -> Linebacker

1. Play Type as Pass and Rush were running at low speeds on Synthetic

Play Type as Punt were running at high speeds on Synthetic

 Play Type as Pass and Rush were running at low speeds on Natural

Play Type as Punt were running at high speeds on Natural

1. No matter whether the speed at which the players were running was high or low, the recovery of injuries caused on Synthetic surface took more time compared to the ones on Natural
2. Maximum Injuries happened on Outdoor and Indoor Stadium type on Synthetic Surface. Maximum Injuries happened on Outdoor Stadium type on Synthetic Surface
3. Temperature at which most injuries were observed on Synthetic Surface were in the range -> 68 to 88 Degree Fahrenheit

Temperature at which most injuries were observed on Natural Surface were in the range -> 60 to 75 Degree Fahrenheit

1. Fastest Player injured on Natural Surface was a Defensive Lineman which was observed to be the Slowest Player on Synthetic
2. 53% of the Injuries occurred on the Synthetic Surface

47% of the Injuries occurred on the Natural Surface

1. Indoor High Temperature were observed to be injury prone where as on

outdoor the temperature was observed to be low when injuries were incurred

1. Last Play on Synthetic Surface was observed between the range --> 2 to 17

Last Play on Natural Surface was observed between the range --> 2 to 26

Median of Last Play on Synthetic Surface was observed at 23

          Median of Last Play on Natural Surface was observed at 15

          Mean of Last Play on Synthetic Surface was observed at 27

          Mean of Last Play on Natural Surface was observed at 24

1. The most heated areas of the Synthetic field where maximum visits were paid by the injured players are in the range -> x = 75 to 90, y = 14 to 18

The most heated areas of the Natural field where maximum visits were paid by the injured   players are in the mid-field

1. Average Temperature on Synthetic surface and Clear weather was observed to be -> 66.08

Number of Injuries -> 12

Average Temperature on Natural surface and Clear weather was observed to be -> 72.13

Number of Injuries -> 15

Average Temperature on Synthetic surface and Cloudy weather was observed to be -> 62.27

Number of Injuries -> 15

Average Temperature on Natural surface and Cloudy weather was observed to be -> 70.83

Number of Injuries -> 12

1. **Precautions**
2. Players playing at Roster Position Linebacker on Natural and at Linebacker and Widereceiver on Synthetic must be careful throughout.
3. On Natural Surface, players must be careful as maximum chances of injuries are on Knee and on Synthetic, Ankle.
4. Defensive Lineman on Natural must be careful while running at high speeds and on Synthetic, at low speeds i.e. the player playing as Defensive Lineman must be careful about their speed according to the surface they are playing at.
5. Players must be careful in the initial gameplay.

# Bibliography

* https://www.kaggle.com/c/nfl-playing-surface-analytics
* https://docs.python.org/3/
* https://numpy.org/
* https://pandas.pydata.org/
* https://scikit-learn.org/
* https://plot.ly/