FDS Visualisation and Interpretation Exercise

s2654998

16th December 2025

1 Dataset Overview

I am using the a football player injury dataset dataset on kaggle [1] . It comprises information on premier league footballers capturing their injury history and performance/physical attribute. Each entry in the database corresponds to a specific injury giving information about the player (including name, demographic and physical details including age, height, wight, BMI and nationality along with performance related metrics from FIFA such as pace physicality and work rate). For each entry the database also includes variables about prior injury metrics for that player (e.g. the average days injured in previous seasons and cumulative days injured) and playing history attributes (e.g. mintues per game in previous seasons). It also includes information about the players involvement in the current season (e.g. season minutes played or season days injured).

There is a previous study using to evaluate football injuries [2]. However it is not really relevant to the study I will carry out.

2 Data Exploration

To explore the data I used pairplots to see which attributes seemed to show a correlation with the severity of an injury (measured by the total number of days injured). To do this I identified 3 major areas where I wanted to see if they had an effect on injury severity: exposure/workload, physical profile of players, and previous injury history. For each of these categories, I created pairplots to visualize these relationships and heatmaps to assess correlations for exposure, workload and physical condition. However from analysis I noticed that there are many problems with the dataset. Firstly a lot of the attributes may actually be consequences of the injury rather than causes meaning it may be hard to extract meaningful interpretations. For example if a player has a serious injury it leads to the amount of games played in that season being little meaning it is hard to see if overplaying caused the injury. Furthermore previous history data may be unreliable because of the fact that severe injuries may carry onto the current season. Also average days injured per season has values exceeding 365 which may mean that the data is mislabeled or erroneous. Because of these potential flaws in the dataset I decided to focus on physical attributes about the player. I noticed some weak correlations and i wanted to investigate this further.

3 Question

Does a players physical rating influence the severity of an injury?

4 Visualisation and Interpretation

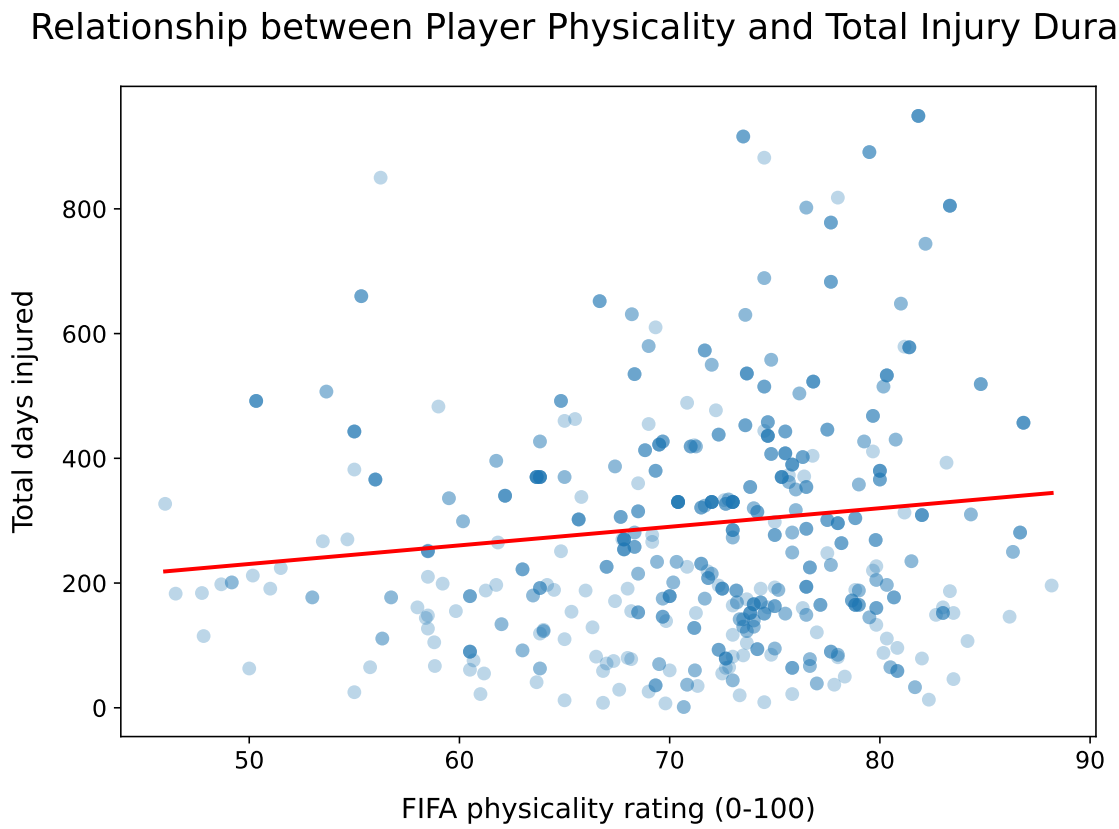

Figure 1: Relationship between FIFA physicality rating and total injury duration (in days). The line shows the best-fit linear relationship between FIFA physicality rating and total days injured

This graph examines the relationship between a football players FIFA physicality rating and their total number of days injured. The physicality score is a composite measure with range 0-100 combining the following physical attributes: strength, aggression, stamina, physical presence in duels.

The regression line shows a very weak increasing gradient, indicating a weak positive relation between physicality and injury duration. This may be because players with high physicality rating may succept themselves to more frequent aggressive and dangerous tackles leading to a higher injury risk in these tackles.

However the spread of the data is highly variable across all physicality levels with no strong clustering. This is shown with there being a correlation coefficient of 0.123(3dp). This suggests that a player's physicality score may not actually be a reliable predictor of injury duration. Implying that other factors including previous injury history, position or workload may also contribute to the injury severity

References

[1] *dataset.injuries-football*. 2025. URL: https://www.kaggle.com/datasets/francocanosamiranda/dataset-injuries-football/data.

[2] FRANCO CANOSA MIRANDA. *IoT, Performance Analytics and Injury Prediction in Football Players*. Last accessed 16th Nov 2025. 2025. URL: https://www.kaggle.com/code/francocanosamiranda/iot-performance-analytics-injury-prediction-in.