Predicting UK Racehorses' Risk of Injury Based on Historical Race and Biometric Data

Sabrina del Rosal

Topic: Predicting UK Racehorses' Risk of Injury and Performance Decline

- **Problem**: Racehorses undergo intense physical strain during races, leading to frequent injuries and performance decline.
- **Opportunity**: By predicting injuries, trainers, jockeys, and owners can optimize training schedules, reduce injury risks, and improve performance.
- **Affected Stakeholders**:
    - **Horse Trainers**: Better training management.
    - **Jockeys**: Understand horse performance on race day.
    - **Owners**: Reduce costs and increase success rates.

Vision for Tackling the Problem Using Data Science

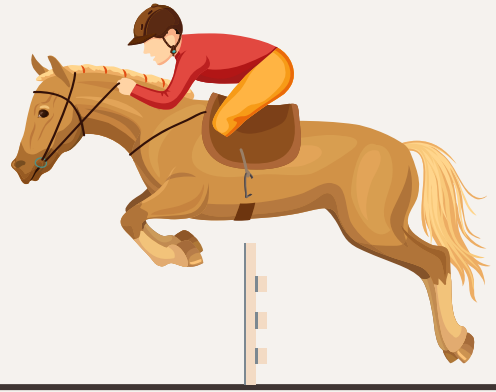
Proposed Solution:

- **Objective 1**: Build a model to predict a horse's race position based on historical race data before biometric factors are added.
- **Objective 2**: Develop a model to assess injury risk by analyzing performance trends and biometric data.

1. Linear Regression (predicting race position)

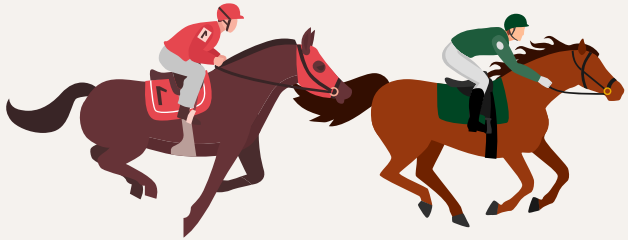
2. Logistic Regression (injury risk classification)

3. Decision Trees (injury risk classification)



Potential Impact of the Solution

- **Optimized Training**: Improved schedules based on injury risk predictions, leading to better race results.
- **Reduced Injury Frequency**: By predicting injury risk, horses can be better managed and protected.
- **Financial Impact**: Owners and trainers can save money by reducing injury-related costs and improving horse performance.



Introduction to the Dataset, Data Quality Concerns, and Preliminary EDA

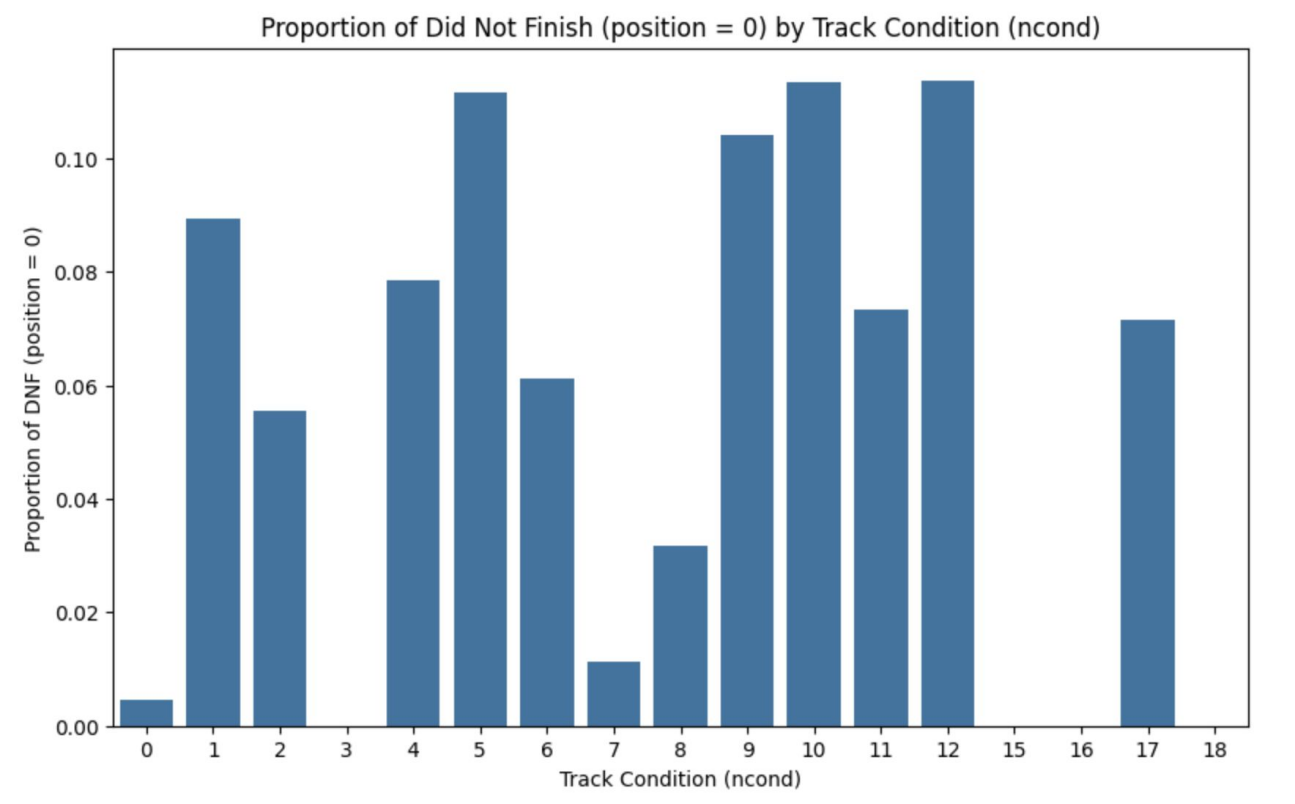
Dataset Overview:

- Pre-Race Parameters: Information collected before the race (odds, trainer data, horse statistics)
- **Race Parameters**: Data on race conditions, times, and prize money.
- **Horse Parameters**: Horse-specific data (weight, jockey, race positions, historical performance)

Data Quality Concerns:

- **Missing Data**
- **Inconsistent Data**: Variability in how data is recorded and labelled
- Both Race & Horse **csv files separated by years** (which years to choose and how many)

Preliminary Findings from EDA:



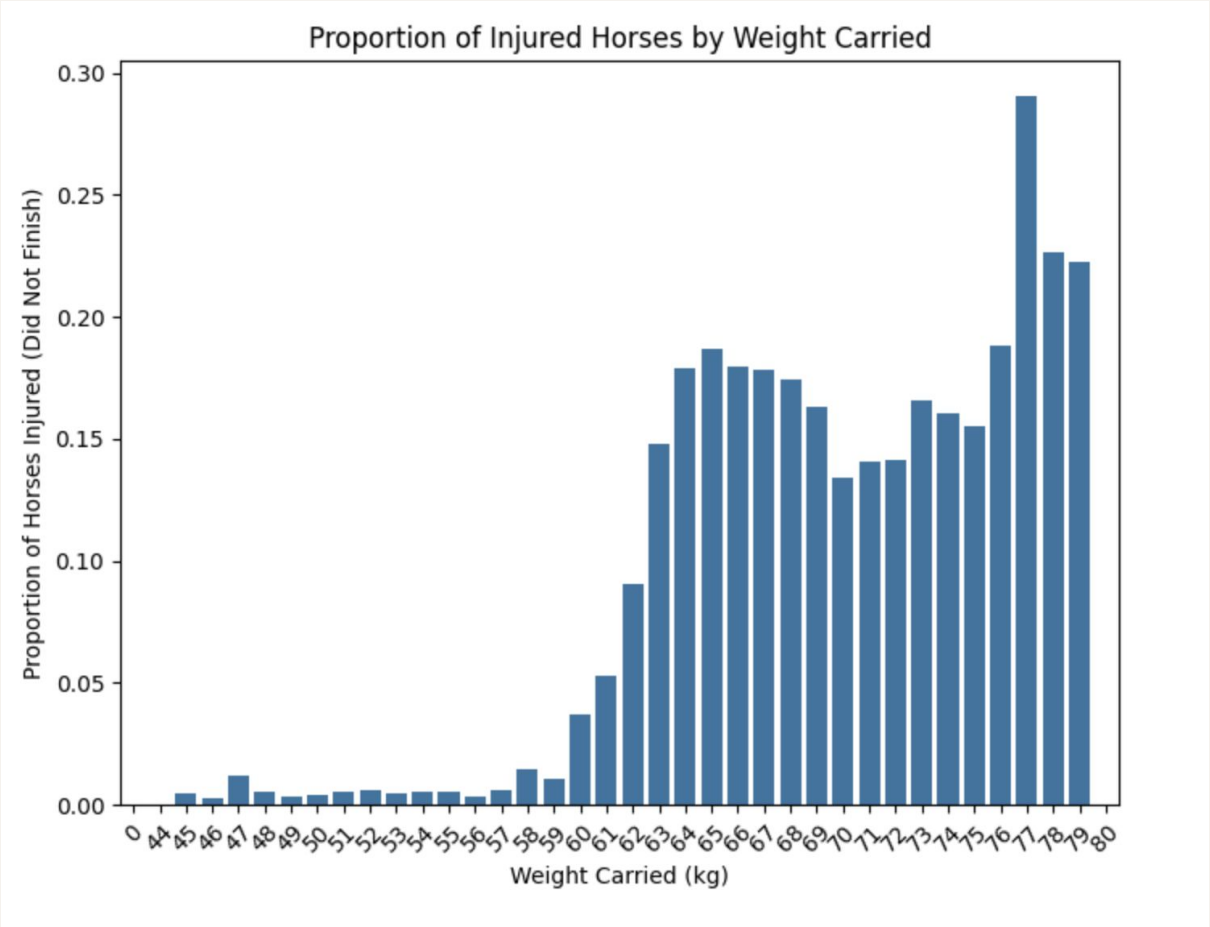
11% DNF for conditions :

12 soft to heavy

10 good to soft

5 soft condition

"Horses with sensitive **tendons**, will feel more at ease on **hard surfaces** because they facilitate their locomotion. Other horses, such as those with **joint** problems, will move more smoothly on **soft surfaces**."



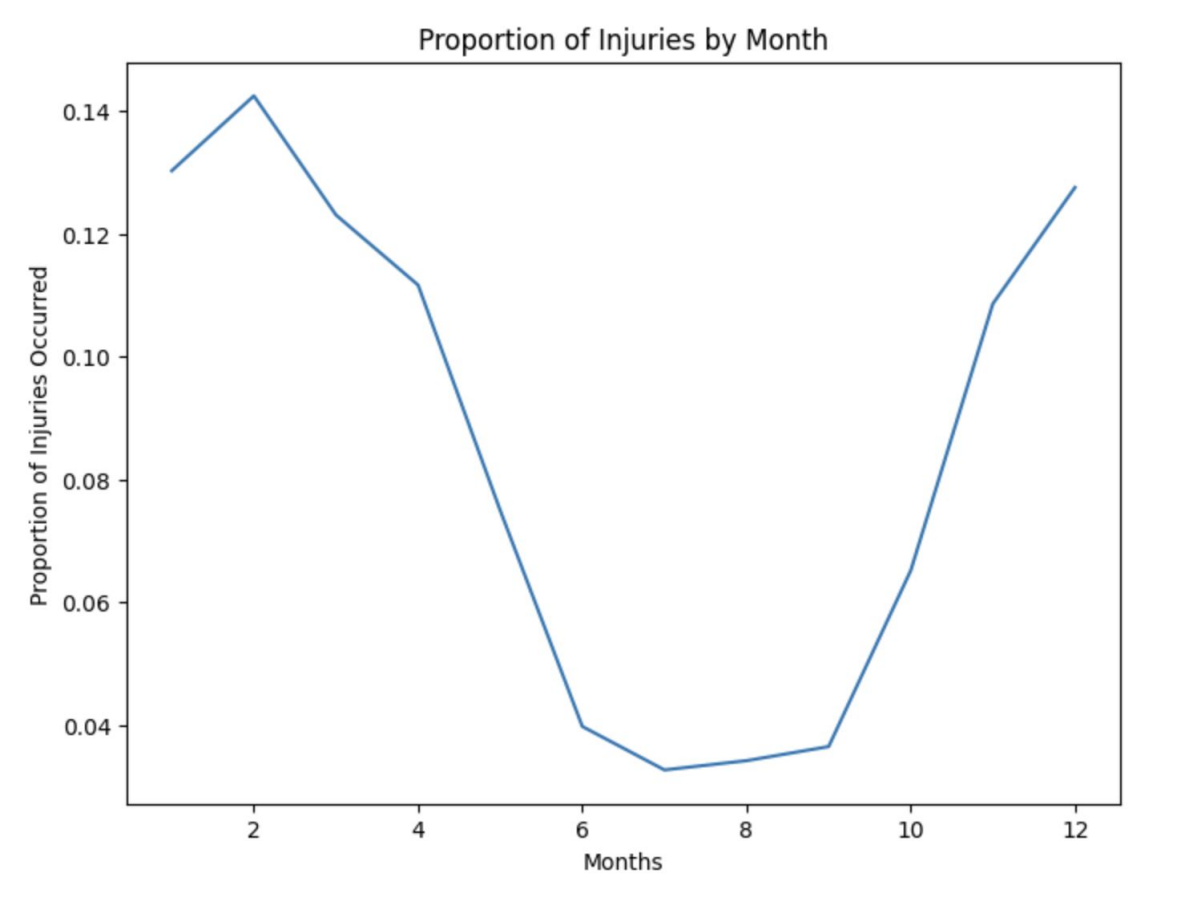
Handicap System:

- Horse with the highest rating carries the heaviest weight.
- Jockey, Saddle, Equipment, Lead Weights

Flat Racing: 50 - 65 kg

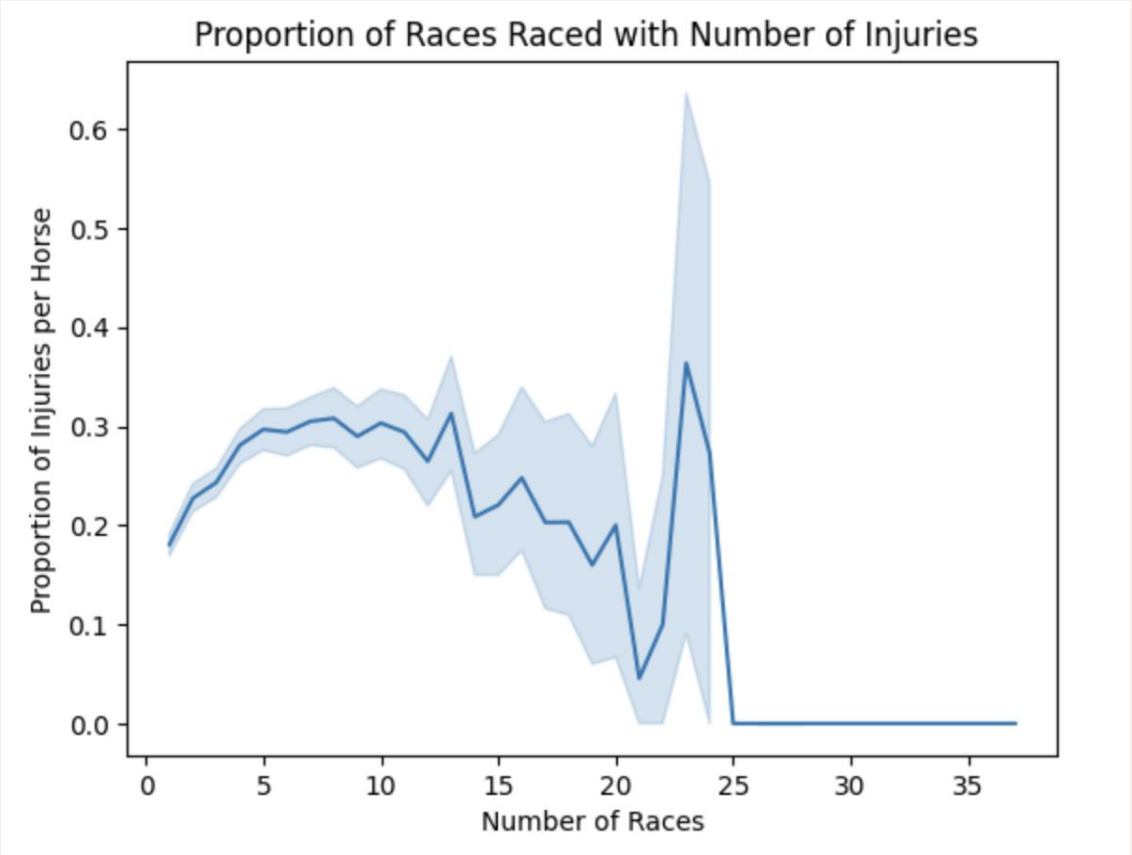
National Hunt Racing: 60 - 75 kg

77 kgs : possibly in amateur races



Winter Months:

1. Muscle Stiffness
2. Reduced Flexibility
3. Slippery
4. Longer Warm-Ups
5. Reduced Training



- “Caps” off at 24 Races
- AVG 20 races, but some horses can race up to 50 races

Thank You!