Football Player Injury Impact Analysis

Project Overview

This project involved cleaning, processing, and analyzing a dataset of professional football players' injuries and performance ratings. The goal was to uncover insights about how injuries affect player performance, recovery durations, and potential risks — information crucial for scouting, team management, and transfer decisions.

Data Cleaning & Pre-processing

- Handled multiple types of missing values and inconsistent entries such as N.A., nan, and empty strings.
- Standardized date formats for injury and return dates, fixing common formatting errors.
- Corrected known typos and inconsistencies in player names.
- Cleaned and converted player rating data, removing noise and symbols to prepare for analysis.
- Calculated injury durations (days between injury and return) as a key feature.

Exploratory Data Analysis (EDA)

- Investigated the distribution of injury durations to understand typical recovery times.
- Compared pre-injury and post-injury player ratings from multiple matches to quantify performance changes.
- Identified "hidden gems" — players whose performance improved significantly post-injury.
- Flagged high-risk players with long injury durations who may pose market or tactical risks.

Key Findings

- Injury durations varied widely, but longer injuries tended to correlate with larger drops in performance.
- Several players showed strong recovery and improved ratings post-injury, highlighting resilience or underappreciated talent.
- The dataset's multi-match rating structure allowed robust before-after comparisons to evaluate injury impact.
- Manual fixes and imputation were applied for missing dates to ensure analysis completeness.

Technical Approach

- Data manipulation and cleaning performed using Python's pandas library.
- Date parsing handled carefully to accommodate inconsistent formatting.
- Visualizations created with matplotlib and seaborn for insightful injury and performance patterns.
- Correlation analysis and basic linear regression model built to predict post-injury performance.
- Player-level scouting reports generated for individual insights.