IHOOP Insights

Injury Prediction

Shashank Guda
Rithika Gurram
Vishnu Charugundla
Varshin Bhaskaran

Agenda

1. Introduction
2. Data Overview
3. Exploratory Data Analysis (EDA)
4. Statistical Significance
5. Injury Risk Model
6. Risk Assessment
7. Key Takeaways & Recommendations

Introduction

- Analyze injury risk in basketball players based on key performance metrics.

- Identify and assess risk factors contributing to injuries (e.g., muscle imbalances, performance metrics).

- Provide actionable insights for injury prevention strategies.

Data Overview

- **Dataset:** 2604 records, 14 unique players.

- **Date Range:** January 1, 2023 - December 30, 2023.

- **Key Features:**

  o Performance metrics (e.g., distance, speed, jump load, heart rate).

  o Muscle imbalances (e.g., hamstring-to-quad ratio, calf imbalance).

  o Injury information (e.g., injury type, body part, recovery time).

- **Missing Data**: High percentage of missing values in injury-related columns.

Exploratory Data Analysis (EDA)

- **Injury Rate:** 226 injuries (8.68% injury rate).

- *Hamstring-to-Quad Ratio* and *Muscle Imbalances* are significant predictors of injuries.

- **Position Analysis**: Guards appear to have higher injury risk based on risk scores.

- **Injury Type**: Muscle strains and tendonitis are most common injuries.

Injury Distribution by Position


Injury Distribution by Position
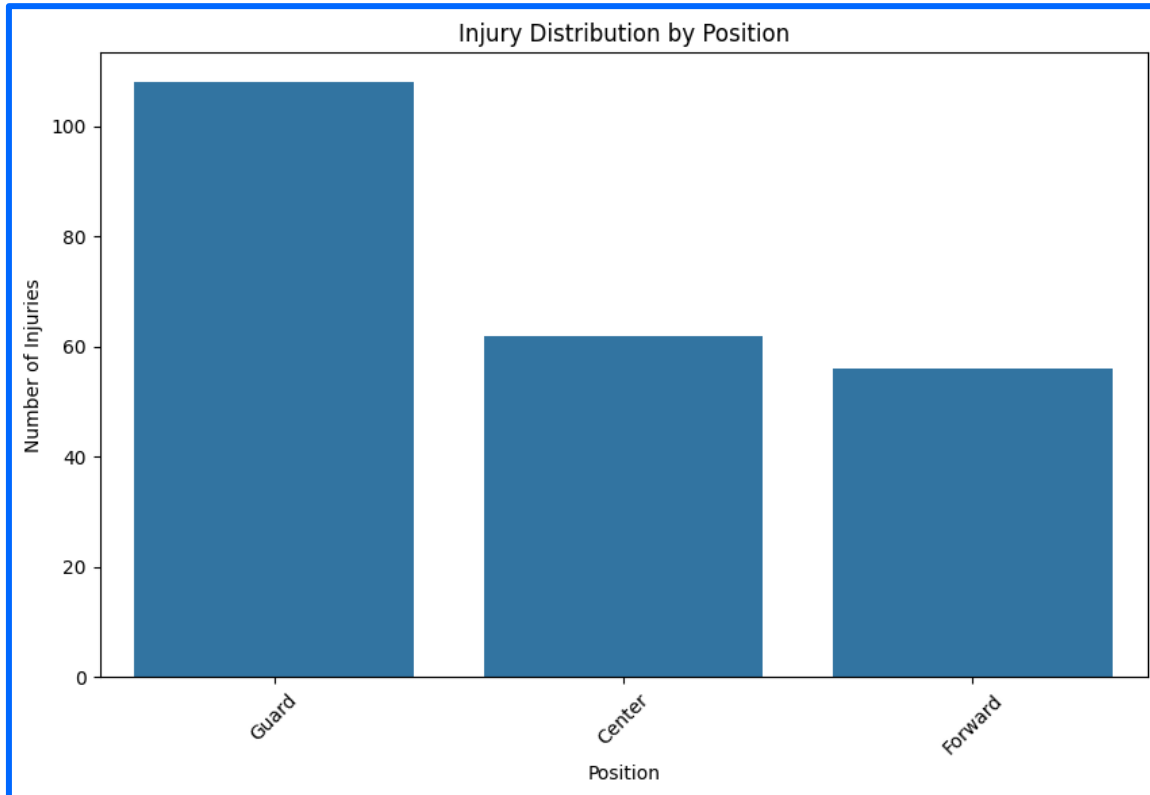
- The bar chart shows the distribution of injuries across different player positions: Guard, Center, and Forward.

- Guard positions have the highest number of injuries, followed by Center, and Forward with the least.

Positions

Forward

Center

Guard

Low Risk

Medium Risk

High Risk

7

Top 10 Most Common Injury Types
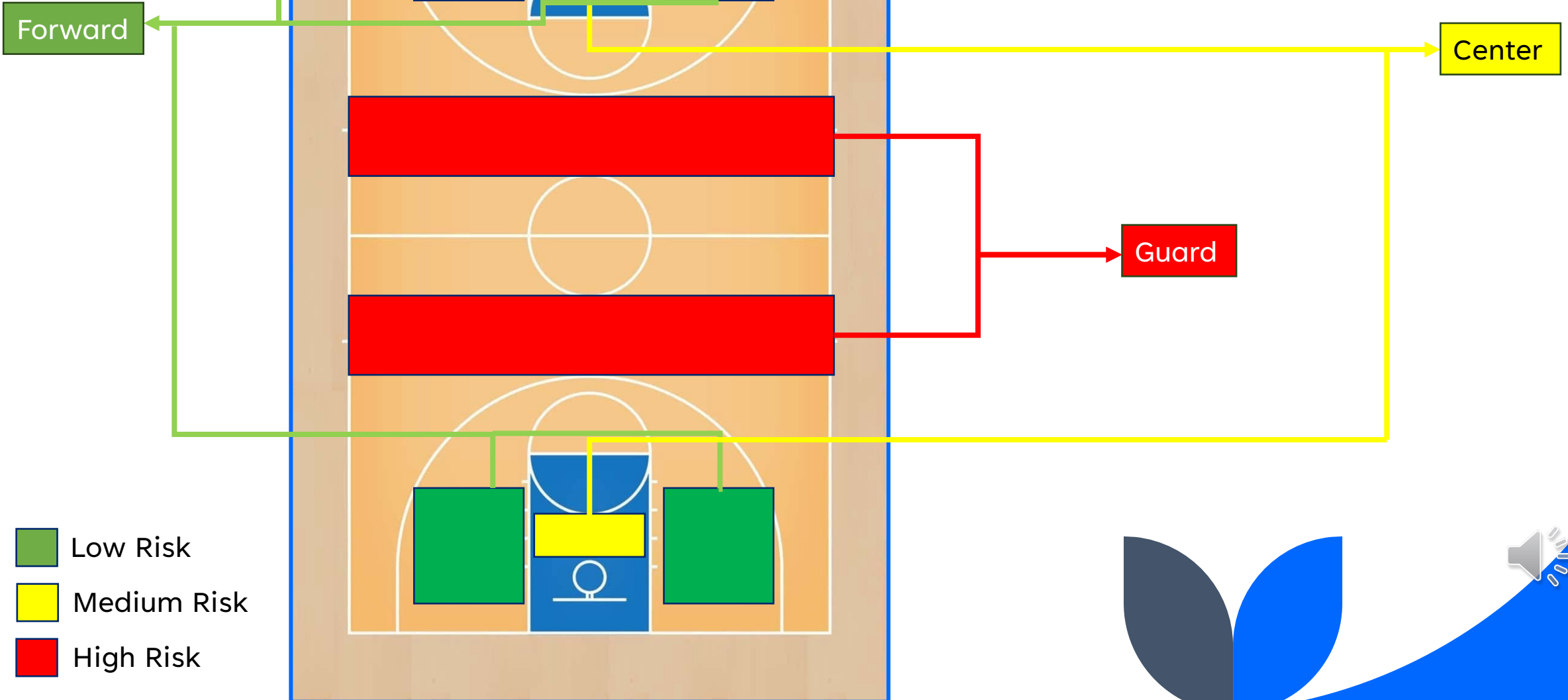
Top 10 Most Common Injury Types
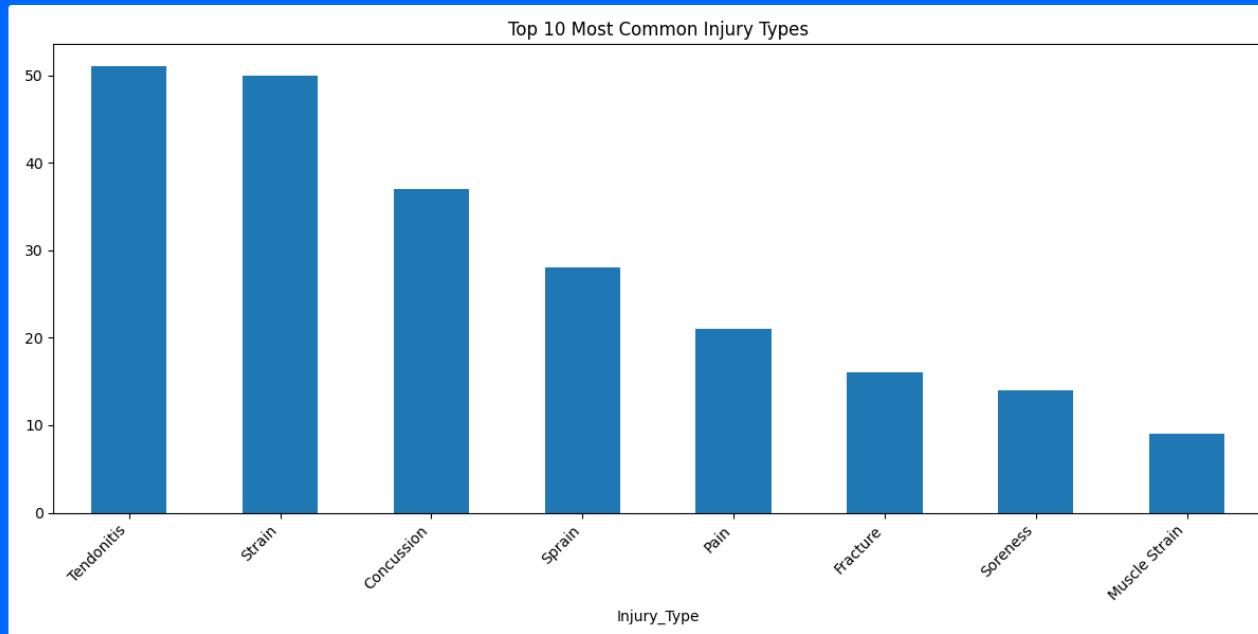
- The bar chart shows the top 10 most common injury types

- The most common injuries are Tendonitis and Strain, followed by Concussion, Sprain, and others.
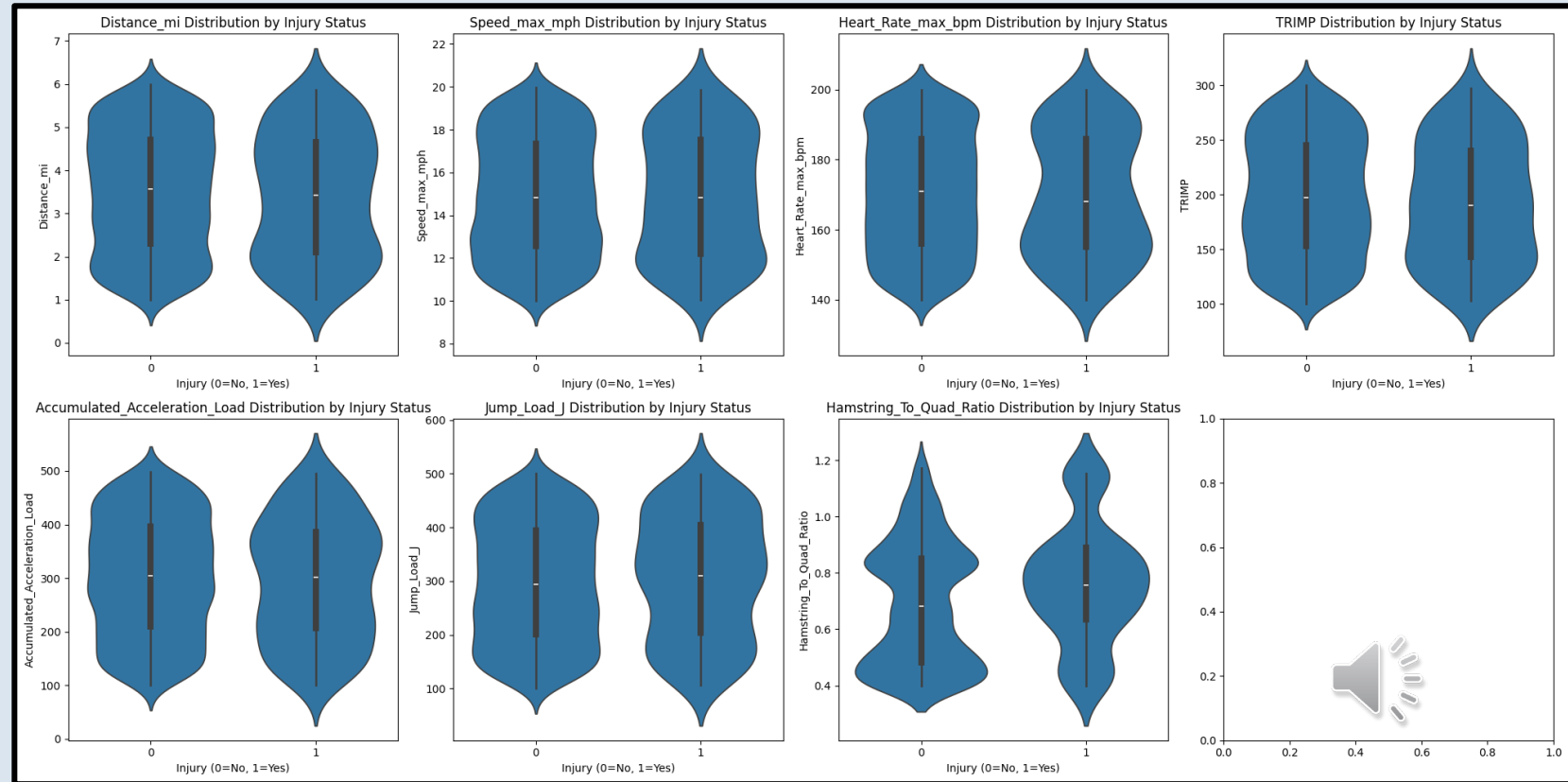
Key metrics distribution for injured vs non-injured players

Distance_mi: The distribution of distance covered by injured and non-injured players appears to have some overlap, but non-injured players tend to have a wider range of values.

Speed_max_mph: There is a clear separation between injured and non-injured players, with non-injured players generally achieving higher speeds.

Heart_Rate_max_bpm: The distribution of maximum heart rate shows a broader range for non-injured players, whereas injured players seem to have a more concentrated range.
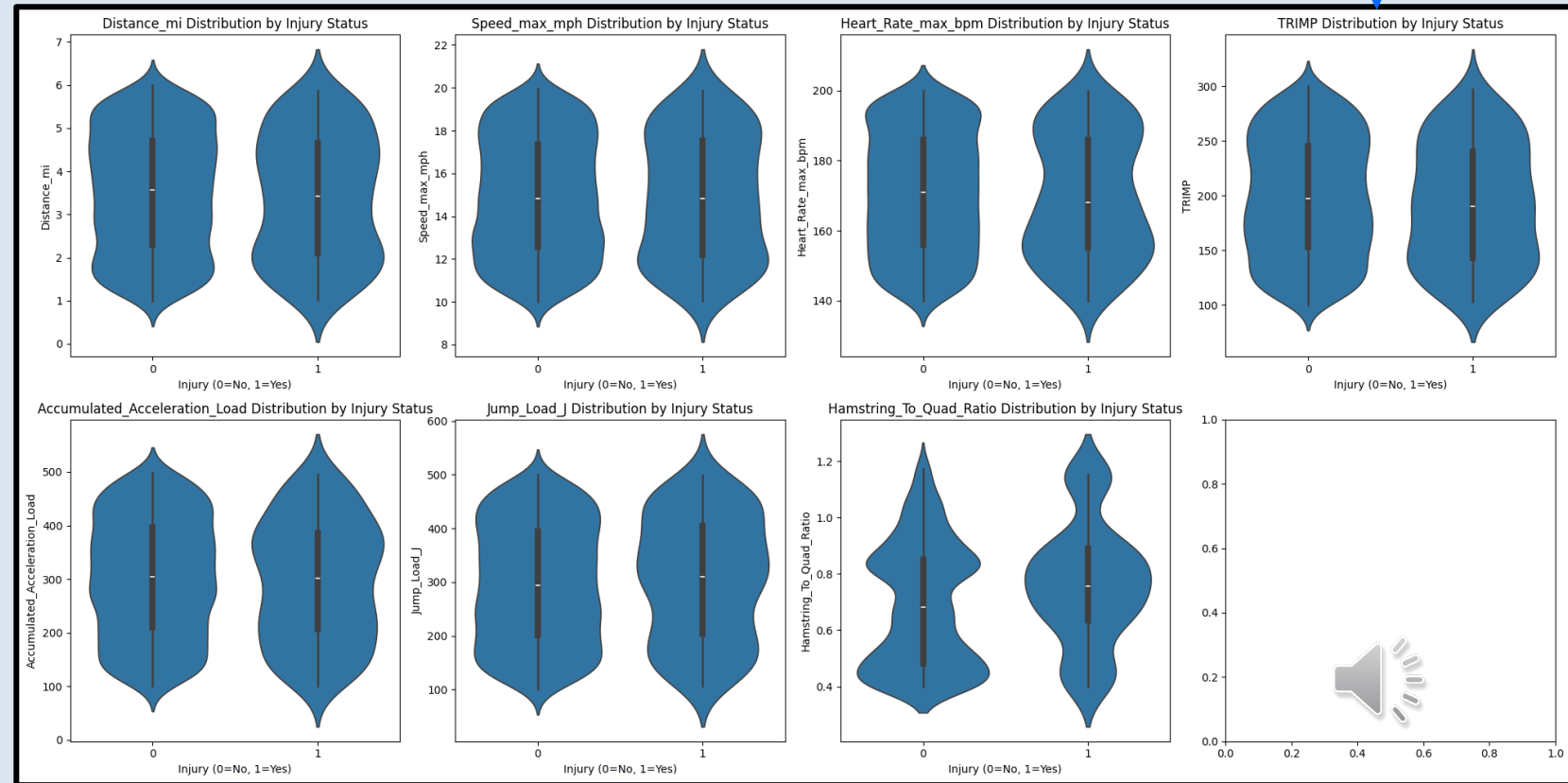
Key metrics distribution for injured vs non-injured players

TRIMP (Training Impulse): Non-injured players show a higher variance in TRIMP values compared to injured players, indicating a higher intensity or variation in their training sessions.
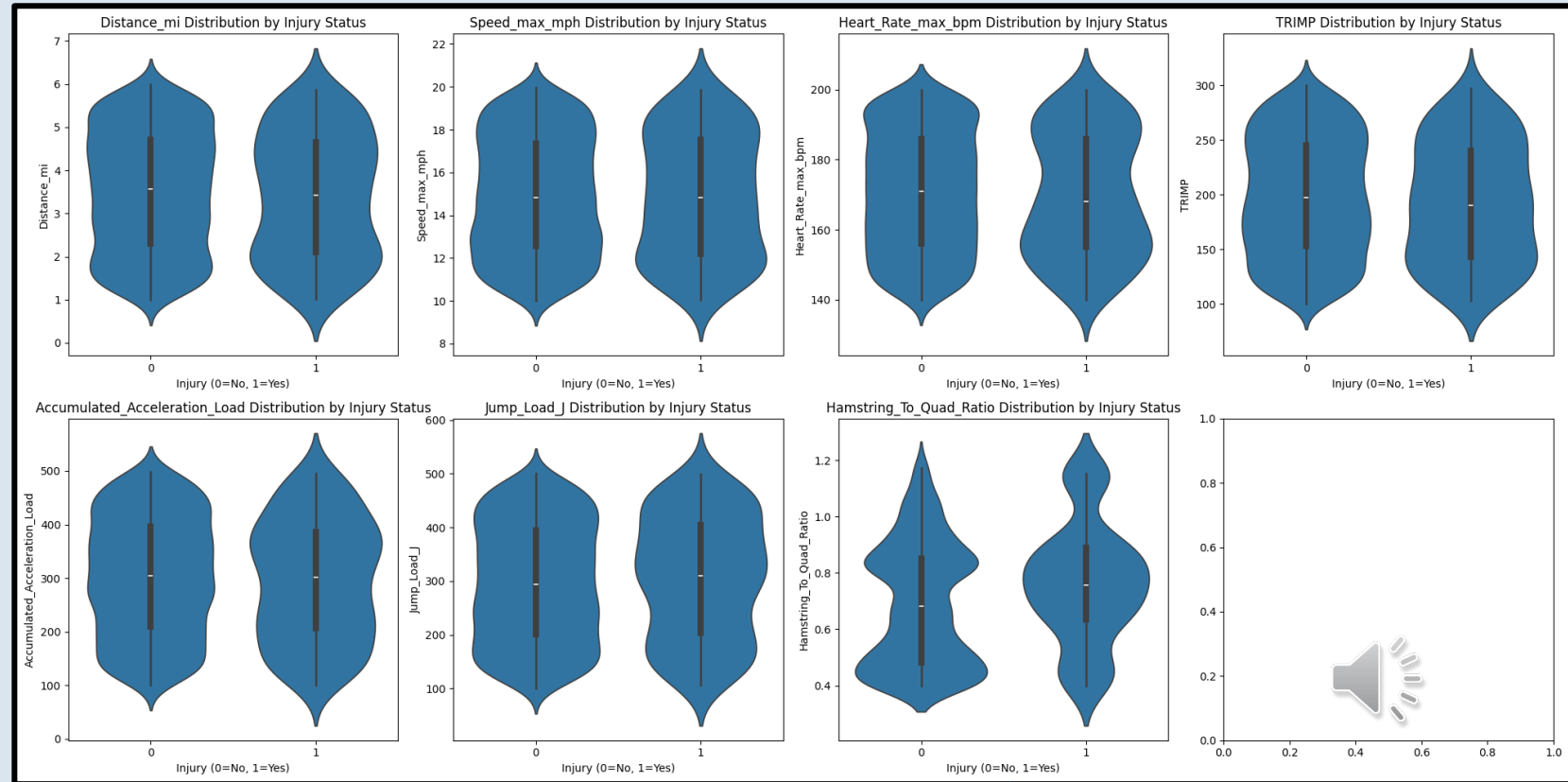
Accumulated_Acceleration_Load: Injured players tend to have a higher accumulated acceleration load, which could be related to the physical strain they experience during play.

Jump_Load_J: Injured players have a more concentrated distribution of jump load, possibly suggesting they are more prone to injury due to higher exertion in jumps.

Key metrics distribution for injured vs non-injured players

Hamstring_To_Quad_Ratio: The distribution shows that the hamstring-to-quad ratio is generally higher for non-injured players, which could suggest a more balanced muscle strength ratio that may protect them from injury.

Injuries Over Time

The number of injuries seems to peak during mid-year, with sharp drops in between, particularly in the months of January and December.
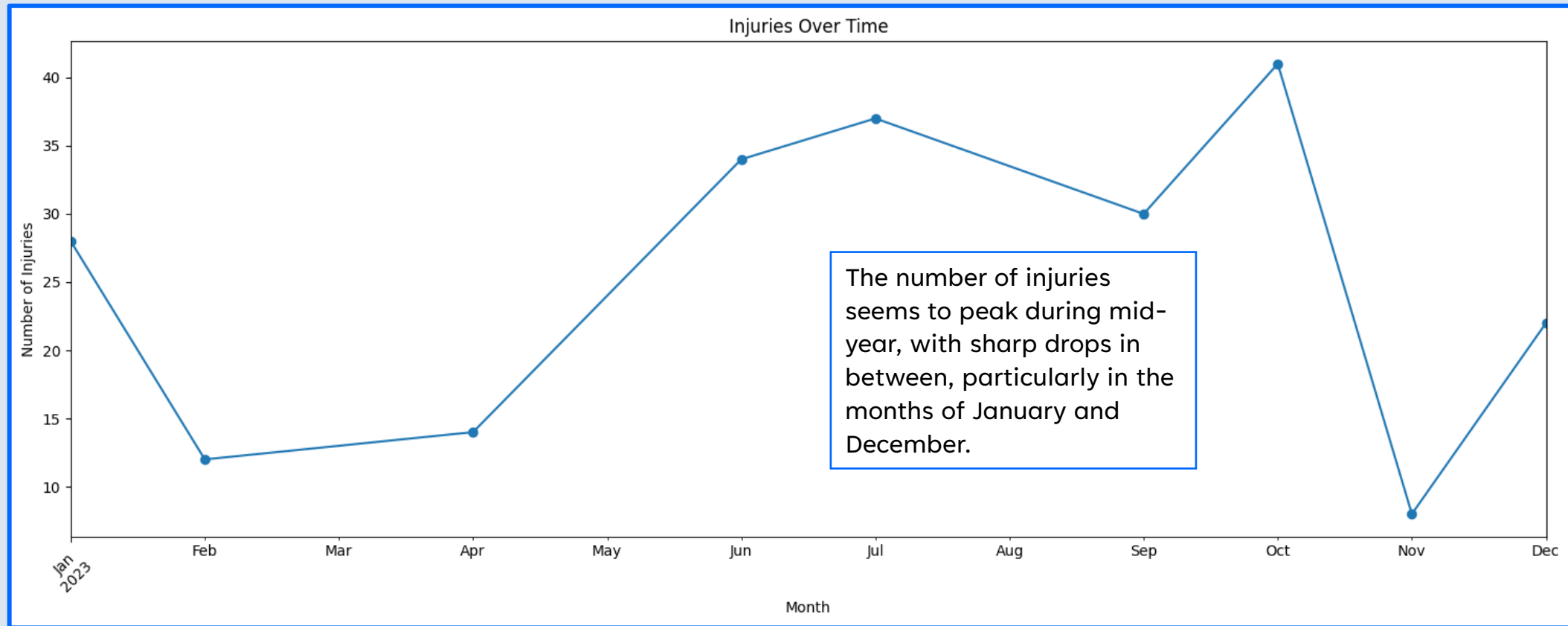
Correlation Analysis

Strong Positive Correlations:

The muscle imbalance percentages (Quad_Imbalance_Percent, HamstringImbalance_Percent, Calf_Imbalance_Percent, Groin_Imbalance_Percent) are highly positively correlated with each other, showing a strong relationship (close to 1.0). This suggests that players with higher imbalances in one muscle group tend to have higher imbalances in others.

Injury-Related Insights:

had_injury shows a moderate negative correlation with performance metrics such as Speed_max_mph and Heart_Rate_max_bpm, suggesting that injured players may exhibit lower performance in these areas. There is a weak correlation between had_injury and the other performance-related metrics like Distance_mi, TRIMP, and Jump_Load_J, suggesting that these metrics do not strongly predict injury status.

Imbalance-Performance Relationship:

The imbalance metrics (Quad_Imbalance_Percent, HamstringImbalance_Percent, etc.) show some moderate correlations with performance metrics like Distance_mi and Speed_max_mph, which could indicate that muscle imbalances may affect overall performance.

Correlation Matrix of Key Metrics
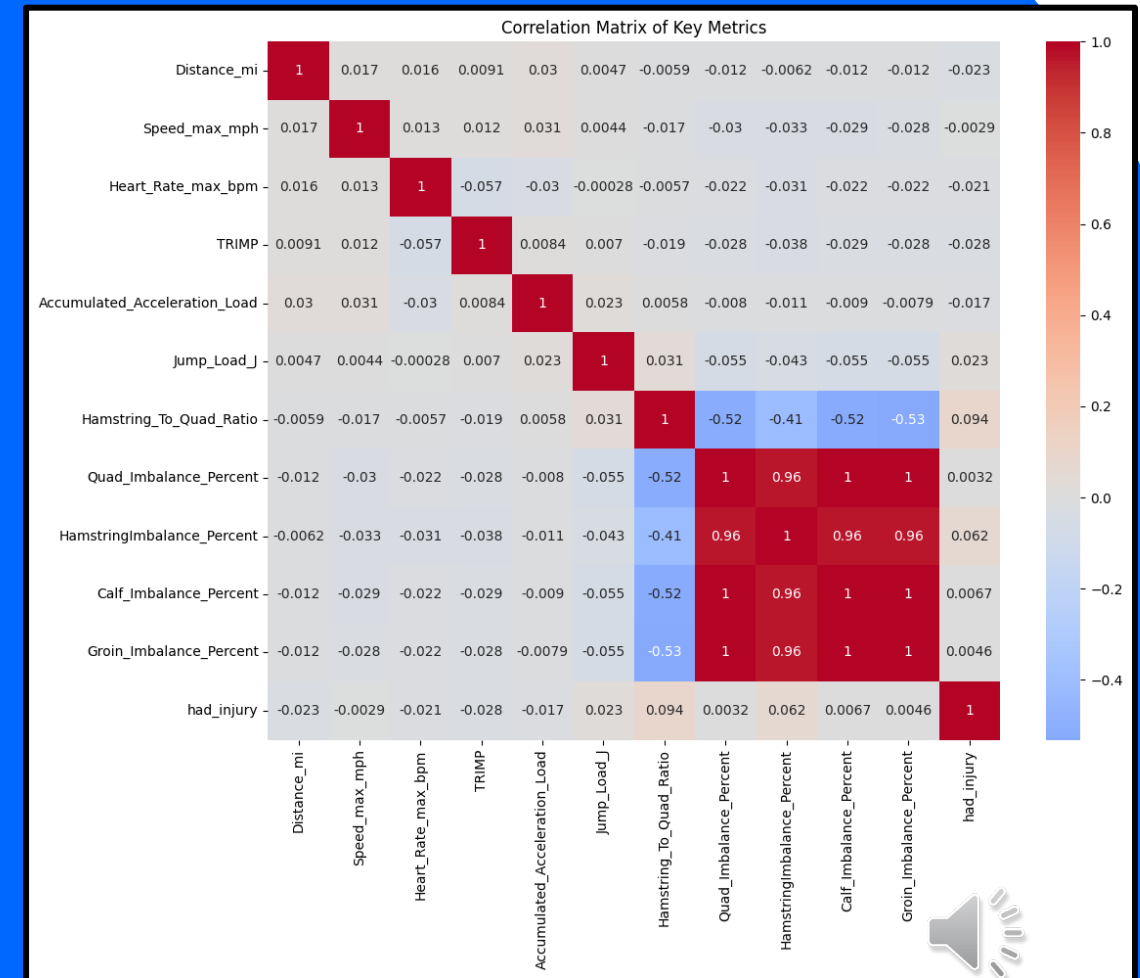
Statistical Significance

- p-values show significant differences in Hamstring-to-Quad Ratio for injured vs. non-injured players.

- Other features like Speed, Distance, Heart Rate show no significant differences between injured and non-injured players.

- Key Insight: Muscle imbalance ratios are highly significant in injury prediction.

```
Statistical Tests (Injured vs Non-injured):
Distance_mi: p-value = 0.2352
Speed_max_mph: p-value = 0.8834
Heart_Rate_max_bpm: p-value = 0.2864
TRIMP: p-value = 0.1590
Accumulated_Acceleration_Load: p-value = 0.3802
Jump_Load_J: p-value = 0.2334
```

Injury Risk Model

Model: Random Forest Classifier trained to predict injury risks.

Feature Importance
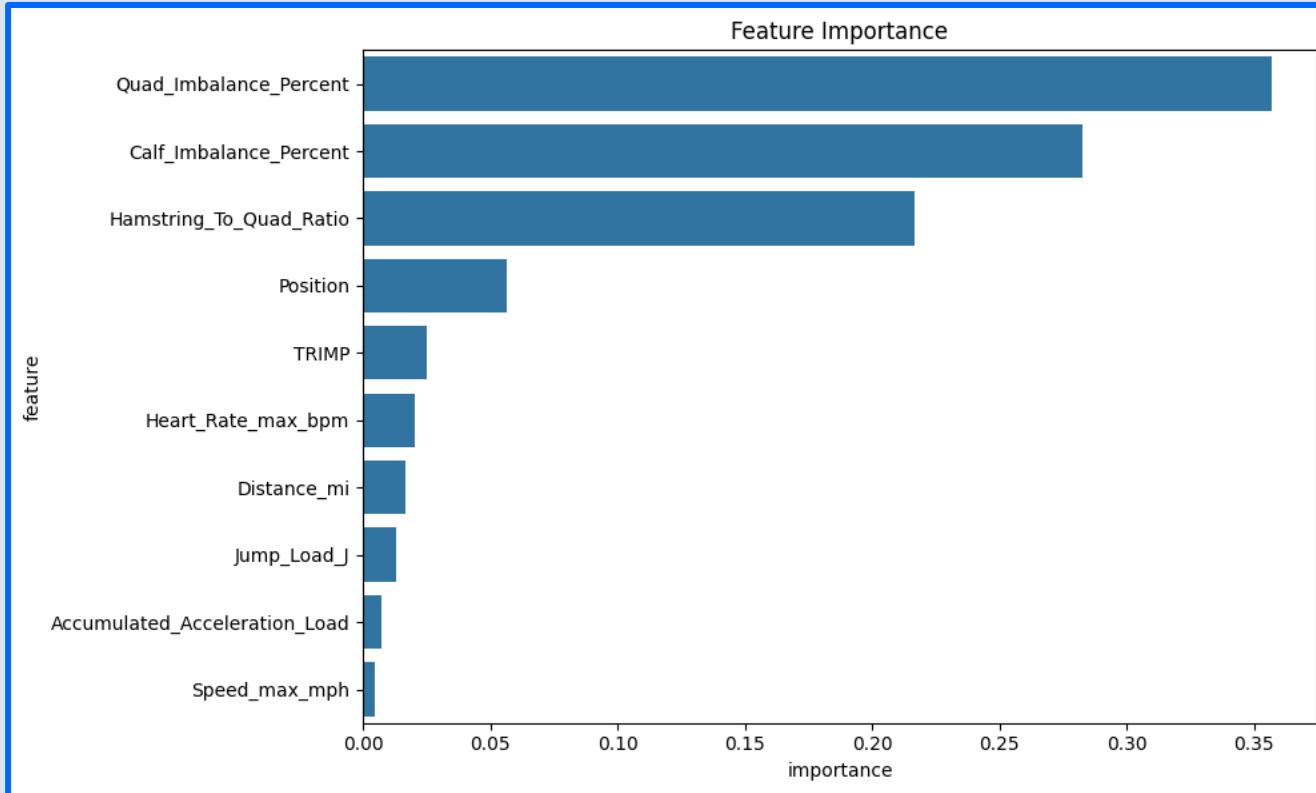
Key Features:

- Hamstring-to-Quad Ratio
- Quad and Calf Imbalances
- Position and Performance Metrics (e.g., speed, heart rate).

Injury Risk Model

Model: Random Forest Classifier trained to predict injury risks.

```
Model Performance Metrics:

Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.82      0.89       476
           1       0.31      0.84      0.45        45

    accuracy                           0.82       521
   macro avg       0.64      0.83      0.67       521
weighted avg       0.92      0.82      0.86       521


Adjusted Classification Report (with threshold 0.3):
              precision    recall  f1-score   support

           0       1.00      0.37      0.54       476
           1       0.13      1.00      0.23        45

    accuracy                           0.42       521
   macro avg       0.57      0.68      0.38       521
weighted avg       0.92      0.42      0.51       521


ROC AUC Score: 0.8979458450046685
```

Metrics:

- Precision for injured players: 0.32
- Recall for injured players: 0.98
- ROC AUC Score: 0.9

Risk Assessment

Risk Assessment for Players

Risk Scores

- Risk Scores assigned based on model predictions.

- Risk Levels: Very Low, Low, Moderate, High

- Example of High-Risk Players:
    1. Malik Robinson (Guard): Risk Score 0.76
    2. Brandon Mitchell (Guard): Risk Score 0.75
    3. Anthony Lopez (Center): Risk Score 0.72

- Risk Level Distribution: Mostly High and Moderate risks.

Risk Level Distribution by Position
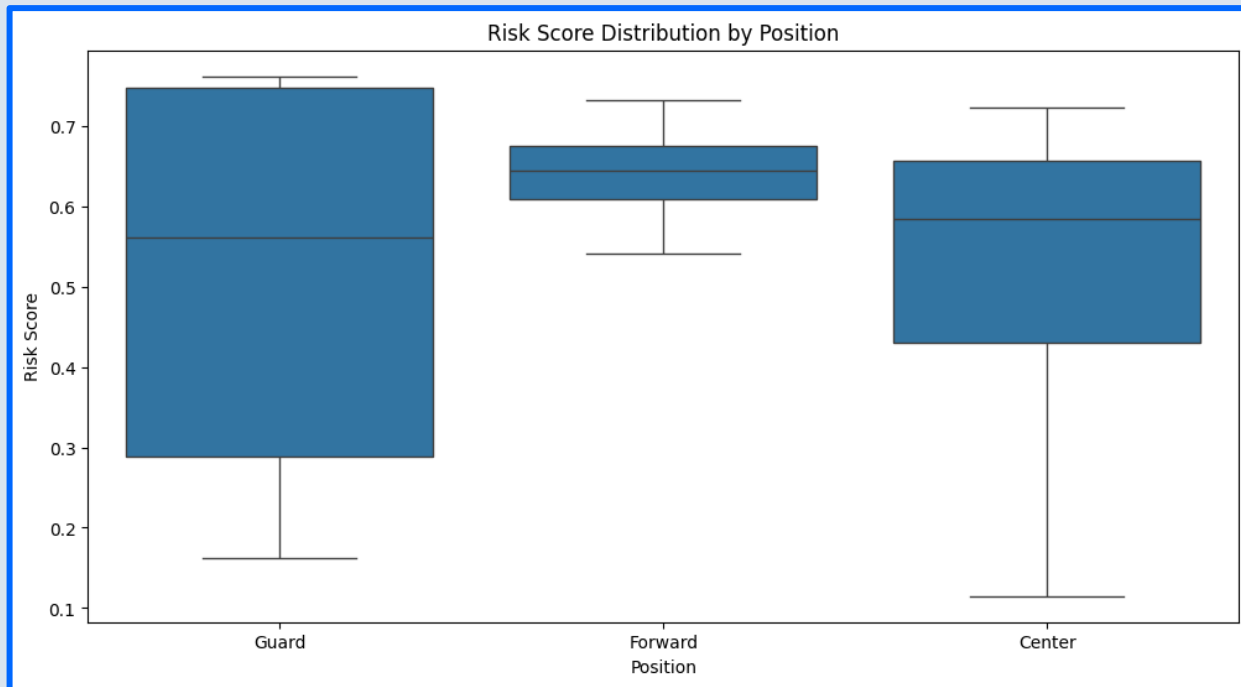
- Guards have higher average risk scores compared to Forwards and Centers.

- Guards tend to be at higher risk due to their performance metrics and muscle imbalances.

Player_ID	Name	Position	Risk_Score	High_Risk	Risk_Level	Hamstring_To_Quad_Ratio	Quad_Imbalance	Calf_Imbalance
103	Malik Robinson	Guard	0.7614873272794657	TRUE	High	0.755814198	14.82441443	14.50325707
115	Brandon Mitchell	Guard	0.7570655050122038	TRUE	High	1.153349615	3.715052264	3.918576225
114	Julian Simmons	Forward	0.7311497148524093	TRUE	High	0.633234294	-1.67453351	-2.119009024
112	Anthony Lopez	Center	0.7229757837977023	TRUE	High	0.895997304	-10.25503063	-10.40782022
105	Noah Bradley	Guard	0.7162885142190292	TRUE	High	0.697620809	-3.696758443	-2.828126327
104	Isaiah Thompson	Forward	0.6559937285633534	TRUE	High	0.627009989	0.181450254	0
107	Cameron Howard	Center	0.6342858069437868	TRUE	High	0.911572089	6.646572473	7.030502167
101	Jordan Matthews	Forward	0.6308905218885021	TRUE	High	0.610768283	5.481294557	5.409497408
110	Kyle Saunders	Forward	0.5411667652832344	TRUE	Moderate	0.571715554	5.459875813	5.55300358
109	Miles Richardson	Center	0.5350797975330448	TRUE	Moderate	0.725251624	6.624618422	6.412319566
108	Xavier Foster	Guard	0.4063770882135166	TRUE	Moderate	1.035582025	-8.842791857	-8.757703979

High Risk Players

IHOOP Insights

Key Takeaways

- Muscle Imbalances (especially Quad and Calf) are the most significant contributors to injury risk.

- Forwards & Guards are at higher risk compared to Forwards and Centers.

- Injury prevention strategies should focus on improving muscle balance, particularly in the hamstring-to-quad ratio.

- The Injury Risk Model is effective for identifying high-risk players, but further calibration and threshold adjustments can improve precision.

Final Insights

Based on the results from the analysis, we can conclude the following:

• **High Risk Players**: Players in the **Guard** & **Forward** position tend to have the highest **Risk Scores.** This indicates that **Guards** might be at a higher risk for injuries in the dataset, particularly given their high **Hamstring-To-Quad Ratio**, **Quad Imbalance**, and **Calf Imbalance**.

• **Risk Factors**: The **Hamstring-To-Quad Ratio, Quad Imbalance**, and **Calf Imbalance** are the key contributors to a player's **Risk Score**. High-risk players (with a risk score of 0.7 and above) typically have significantly higher values in these metrics.

• **Overall Distribution**: The **distribution of risk levels** is relatively skewed toward the **High** and **Moderate** categories, with only **2 players** classified as **Very Low Risk**. This shows that most players in the dataset have some degree of injury risk, and muscle imbalances are a significant contributor.

Recommendations

Focus on muscle imbalance rehabilitation:

Targeted exercises to address quad and calf imbalances.

Position-specific training:

Customize injury prevention strategies for Guards and Forwards, who are at higher risk.

Ongoing Monitoring:

Regular assessments to track players' injury risk levels and adjust training plans accordingly.

Thank you

Shashank Guda

Rithika Gurram

Vishnu Charugundla

Varshin Bhaskaran

IHOOP Insights

Detailed Analysis is present in the Jupyter Notebook File