**NBA Injury Data from 2015-2025:**

**Data Visualization of Player Injuries**

William Wright

Department of Information Technology, Purdue Global University

**Author Note**

Correspondence concerning this article should be addressed to William Wright, Department of Information Technology, Purdue Global University, 2550 Northwestern Avenue, Suite 1100 West Lafayette, IN 47906. E-mail: williamwright@student.purdueglobal.edu

**NBA Injury Data from 2015-2025**

This season's NBA playoffs for marked by numerous severe and potentially career altering injuries to a number of star players from the last two NBA finals, including 3 all-nba players suffering achilles tears– Damian Lillard, Jayson Tatum, and Tyrese Haliburon. This type of injury is considered severe and potentially career altering, with an estimated recovery between 12 to 18 months before they can return to play. Both Haliburton and Tatum are 27 or younger– a player is generally considered in their prime between 27-30 (Salmeh, 2023)–and now face a long road to recovery with questions of if they will be able to get back to the same level prior to the injuries.

Sports analysts have begun to raise questions of the rise of serious injuries in the playoffs especially those who are considered “all-stars” within the league (Reynolds, 2025). One theory for the rise in serious injuries in the playoffs include “wear and tear”, as the NBA season has continued to expand with an 82-game schedule and the addition of an in-season tournament and play-in games (Rosenstein & Srinivasan, 2025). In addition, players may be required to play multiple back-to-back games within a season along with flights across the US and occasionally internationally.

The purpose of this research project is to review NBA injury data from the past 10 years and look for any hidden patterns, such as frequent injury types, season phases with high injury rates, season ending injuries, and whether rest has an effect on player health. Three measurable key performance indicators (KPIs) were identified for this project– injury type distribution, injury frequency, and season-ending injury count. Data was scraped from the website Pro Sports Transactions (2025) using a python script I designed based on the original code by Hopkins (2020).

The data dictionary for this project was generated based on the table provided by Pro Sports Transactions with additional categories created through python. The complete data dictionary prior to cleaning features:

* Date: The date the injury update was provided (e.g. 3/16/2019)
* Team: NBA team associated with the player at time of injury (e.g. Blazers)
* Player: The player who was injured (e.g. Damian Lillard)
* Acquired: Player returns to lineup (e.g. Damian Lillard)
* Relinquished: Player placed on injured list (e.g. C.J. McCollum)
* Notes: Narrative summary of injury, status, or update (e.g. placed on IL recovering from surgery on back)

Data was extracted using a python code that utilized BeautifulSoup and Selenium/undetected-chrome driver to scrape the HTML tables across the paginated results on the Pro Sports Transaction Site. Data was then cleaned with a python code I created based on Zuvkovic (2018) and collaborated on by ChatGPT (2025) and Google Colab AI (n.d.). The dataset of NBA injuries from 2015 to 2025 underwent a multi-step cleaning and feature engineering process using Python and its data manipulation libraries such as pandas, numpy, and re. Initially, the data was loaded from a CSV file and subjected to whitespace stripping and character decoding to correct encoding artifacts (e.g., “â€¢”, “Ã”). Date fields were standardized, and key text fields such as “Acquired” and “Relinquished” were used to construct a unified “Player” field. The “Team” names were normalized to title case, and date-related features such as year, month, and weekday were extracted for temporal analysis. To interpret injury severity, the notes were categorized into a custom duration type (e.g., “out for season,” “day-to-day,” “surgery”) using string pattern recognition, and corresponding estimated durations were mapped numerically. An “InjuryKeyword” was extracted using a predefined list of common injury terms to enable categorical analysis.

The script then attempted to construct injury duration records by analyzing sequences of transaction entries for each player and team. It calculated both the actual duration of injury and the number of in-season days missed, using June 15 as the standard season-end date if a return was not recorded. Injuries with no return transaction were flagged as presumed season-ending. The output from this injury tracking logic was merged back into the main dataset. To reduce duplication and noise, rows that represented return transactions were removed, and records were deduplicated using a combination of player name, team, date, and notes. The cleaned dataset includes final indicators such as “Month,” “DayOfWeek,” and complete duration estimates, and it was exported as a CSV file for downstream analysis and visualization. The cleaned dataset now includes these additional categories in the data dictionary:

* Injury Keyword: Injury category from the description that categorize the injury (e.g. achilles)
* Injury category: The category for injury duration (e.g. day-to-day, out for season)
* Year: Year of event (e.g.2022)
* Month: Month of event (e.g. March)
* Year and Month (e.g. 2022-10)
* Estimated Days Out from notes: an estimation based on injury category (e.g. 180 days)
* Injury start: Date injury was first reported (e.g 10-29-2022)
* Injury end date: Date when player was moved to “acquired” (e.g. 2-14-2023).
* Actual injury duration: Based on the difference between relinquished/acquired (e.g. 33 days)
* Day of week: Day of the week the injury event occurred (e.g. Sunday).

# 

# **Figure 1**

# *Top 20 Provinces in China by Confirmed COVID-19 Cases (Jan 25, 2020)*

# 

# *Note.* The purpose of this chart is to highlight the top 20 Provinces most affected during the initial COVID-19 outbreak in China. The data displays a large outlier, Hubei, and should be helpful to epidemiologists and decision-makers. The bar chart allows for large amounts of data to be used and compared while also drawing clear attention to Hubei’s overwhelming case numbers early in the pandemic.

# 

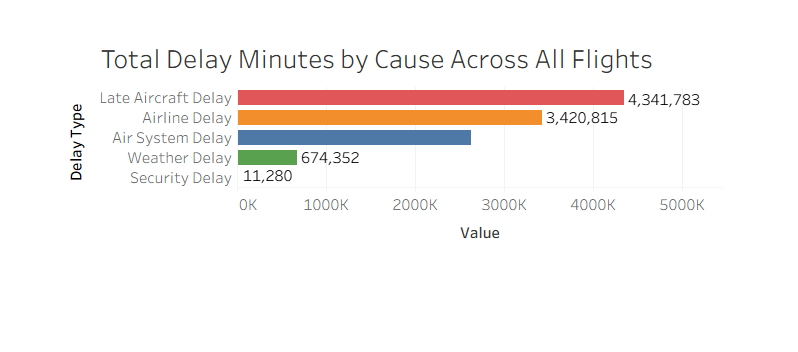
# **Figure 2**

# *Confirmed vs. Recovered Cases in U.S. States (Top 15, Jan 21, 2021)*

# *Note.* The line chart is intended to illustrate recovery trends relative to confirmed cases across the top 15 states in confirmed cases. The line chart is effective for visualizing multiple quantitative trends across categories, specifically revealing disparities in recovery relative to infection levels. The use of data will be helpful for policy makers, healthcare analysts, and researchers looking at the trends of the COVID-19 outbreak.

# **Figure 3**

# *Top 20 States By COVID-19 Testing (per 100,000, Jan 21, 2021)*



# *Note.* The heatmap shows which states had the highest COVID-19 testing rates per 100,000 across the top 20 states based on testing rate. The heatmap provides a color-based visual intensity to immediately distinguish high-testing vs low-testing states, ideal for spotting geographic testing disparities. The data should be helpful for public health agencies, journalist, and the general public in identifying testing rates of U.S. states. Those concerned about healthcare access and state-level testing will be most interested in this data.

# 

**References**

Basketball transactions search results. Pro Sports Transactions. (n.d.). https://prosportstransactions.com/basketball/Search/SearchResults.php?BeginDate=&EndDate=&ILChkBx=yes&InjuriesChkBx=yes&Player=&Team=&start=28100&submit=Search

OpenAI. (2025). ChatGPT (July 21 version) [Large language model]. https://chat.openai.com/chat

Google Colaboratory. (n.d.). Colab. https://colab.research.google.com/#scrollTo=GJBs\_flRovLc

Hopkins, G. (2020, October 30). *gboogy / nba-injury-data-scraper*. GitHub. https://github.com/gboogy/nba-injury-data-scraper/blob/master/injury\_scrapper.py

Reynolds, T. (2025, May 13)*. Injuries are becoming the story of the NBA playoffs. and not even the game’s stars are safe.* AP News. https://apnews.com/article/nba-playoffs-injuries-tatum-curry-22e9533b490480dc7f2c45e12f29d3cd

Rosenstein, G., & Srinivasan, S. (2025, June 24). *3 NBA stars hurt their Achilles tendons this postseason. is this injury becoming more common?*. NBCNews.com. https://www.nbcnews.com/health/health-news/achilles-tendon-tears-injuries-nba-playoffs-rcna214562

Salameh, T. (2023). An empirical analysis of prime performing age of NBA players; when do they reach their prime?. Bryant Digital Repository. https://digitalcommons.bryant.edu/eeb/vol16/iss1/14/

Zivkovic, J. (2018, November 23). Extensive NBA injuries Deep Dive Eda. Kaggle. https://www.kaggle.com/code/jaseziv83/extensive-nba-injuries-deep-dive-eda/report