THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

A Data-Driven and Modeling Study of NFL Ticket Pricing Through Team Records and Player Injury Data.

VRAJ PATEL
SPRING 2026

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Computer Science
with honors in Computer Science

Reviewed and approved* by the following:

Conor Ryan
Assistant Professor of Economics
Thesis Supervisor

John Sampson
Associate Department Head of Computer Science & Engineering
Thesis Honors Adviser

* Electronic approvals are on file.

ABSTRACT

This thesis examines the relationship between NFL ticket prices and two key variables: team injuries and team performance as measured by win-loss records. Understanding how current team conditions affect fan demand has important ramifications for revenue optimization and pricing strategy, as dynamic pricing is becoming more and more crucial in sports economics. The main question guiding this study focuses on how much do player injuries and team records impact NFL ticket pricing during a season.

To explore this, I compiled a thorough dataset tracking weekly injury reports and win-loss records for all 32 NFL teams starting in Week 9 of the 2025 NFL season. Throughout this process, average ticket prices were gathered from a public, verified ticket selling platform, TickPick for each team's home and away games. The objective was to assess whether lower ticket prices are correlated with higher injury counts and whether this effect is mitigated or enhanced by a team's performance and the injuries to star players.

Ordinary Least Squares (OLS) regression models were employed to measure the statistical significance and strength of the link between injury factors, team records, and changes in ticket prices. Python was used to create and test multiple regression models that explored combinations of features including the number of injured starters and recent win/loss streaks. To verify the robustness of the model, several statistical variables were assessed, including R-squared values, p-values, and coefficient magnitudes.

The results provide data-driven knowledge of how off-field economics are impacted by on-field reality. By utilizing computer science methods to simulate and predict pricing strategy, these findings connect economic theory with actual consumer behavior more accurately. These observations can help shape future dynamic pricing plans for teams and add to more general conversations about fairness in sports ticketing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to Dr. Conor Ryan for all his help, encouragement, and support during this thesis process. This effort and my general interest in the topic have been shaped by his profound knowledge of economics and his considerate mentoring. I was originally motivated to investigate dynamic pricing in the NFL generally backed by computer science models, which eventually lead me to follow this subject. Our conversations, which included anything from economic ramifications to proper data collection, constantly challenged me to think more deeply and approach issues with more purpose and clarity.

Additionally, I want to thank Dr. Jack Sampson, my honors advisor, for his unwavering support during my undergraduate studies and for guiding me through the Schreyer Honors requirements. His guidance was crucial in helping me stay on course and keep perspective during this process.

Finally, I want to express my gratitude to my friends and family for their unwavering support. Even during periods of struggle throughout this process, their support kept me inspired and concentrated. I am appreciative that I had such a great support network while completing this thesis.

Chapter 1

Introduction

The National Football League (NFL) is one of the most profitable and lucrative sports organizations in the world. The NFL generates billions of dollars each year from broadcasting rights, sponsorships, merchandise, and, most importantly, for this study, ticket sales. Its 32 franchises are spread throughout major and mid-sized American cities. Eight or nine of each NFL team's 17 regular-season games are held at their home stadium, and fans frequently spend hundreds of dollars to see these games in person. NFL teams are increasingly using dynamic pricing algorithms to establish ticket prices instead of the fixed pricing models that have historically been utilized in sports. These algorithms make real-time adjustments based on parameters including the opponent, day of the week, demand, and performance indicators. This often raises concerns about fairness, transparency, and the underlying causes that most affect price fluctuations, but it also brings pricing instability and increases the possibility of revenue maximization.

The impact of team injuries and on-field performance is an understudied factors influencing these pricing models. The significance of mid-season injuries, especially to top players, on ticket pricing has not been thoroughly measured throughout the league, despite the obvious correlation between a team's win-loss record and fan excitement and demand. Furthermore, limited predictive models now in use combines ticket data, team records, and weekly injury reports to explain or predict pricing behavior. Such a strategy would provide more knowledge and transparency for consumers to determine if ticket prices represent true team value. Understanding how injury trends and performance changes impact fan behavior may help teams and ticketing platforms create more financially efficient and morally sound pricing structures. This gap serves as the driving force for this study: although dynamic pricing is common, the precise causal links between injury impact, team performance, and ticket price changes are still mostly anecdotal or conjectural. This thesis uses a computational and data-driven lens to try and make sense of that area.

1.1 Motivation and Research Gap

Few studies depict the short-term economic consequences of injuries or investigate them as variables in dynamic pricing models. It makes sense that such variables would similarly affect consumer willingness to pay in a league where a single player injury (such as to a quarterback or star wide receiver) can significantly change playoff odds, betting spreads, and team morale. However, there is little empirical data that explicitly links changes in average ticket prices to player availability, particularly when looking at a large sample of teams and games.

Furthermore, from a computer science standpoint, not much research has been done on developing predictive or explanatory models that combine team performance measures and injury-level data to predict patterns in ticket prices. Although proprietary dynamic pricing engines are used by ticket-selling platforms, fans are unaware of the factors that influence price changes due to these systems' lack of transparency. This study intends to close that research gap by creating an interpretable model, enabling fans to better comprehend the factors influencing ticket pricing. By combining computer science, data science and microeconomic theory with real-world applications in pricing strategy design and consumer-facing platforms, the approach also advances the larger area of sports analytics.

1.2 Objectives and Contributions

The primary objective of this thesis is to develop a data-driven model that assesses how NFL ticket prices relate to injuries while also focusing on win-loss record performance. Starting in Week 9 of the 2025–26 NFL season and continuing through the rest of the season for all 32 teams, the study gathers comprehensive data looking at these variables. Each week's dataset includes:

1. Injury reports (number of injured players) from ESPN's weekly injury report tracker
2. Average ticket price for each game (home and away) based on publicly available pricing data from TickPick.com

Following data collection, the project uses Python-based data science tools in a structured analytic pipeline. To classify injuries by severity and role and match ticket prices with pertinent game dates, the dataset must first be cleansed and normalized. OLS (Ordinary Least Squares) regression summaries are then created using the data to assess the statistical significance and strength of each explanatory variable.

Lastly, by experimenting with multiple linear regression models and visualizing residual distributions to assess model fit, the study integrates computational modeling approaches beyond simple regression. In addition to determining the degree of association between injuries, performance, and pricing, the goal is to evaluate predictive power: that is, whether these factors can predict price patterns in upcoming weeks or games or if fans can predict what ticket prices may look like given a specific number of injuries to a team. This thesis offers a real-world dataset and a methodological framework that can be used or expanded upon by future analysts.

Chapter 2

Literature Review

The NFL ticket market has developed into a dynamic ecosystem influenced by algorithmic pricing techniques, consumer expectations, and competitive performance. In the past, ticket prices were mostly fixed, predetermined by the opponent, seat position, and general demand projections. However, real-time price discovery is becoming more and more prevalent in the current NFL ticket market, as prices vary between the main market (team and official ticketing systems) and secondary markets (resale marketplaces) (Drayer, Shapiro, & Lee, 2012). These variations are not arbitrary; rather, they reflect the willingness of consumers to pay in an unpredictable environment where demand can change quickly due to factors like a team's win-loss record, player availability, and perceived playoff hopes. NFL ticket prices are now a quantifiable economic indicator of fan mood and anticipated game quality.

The factors that influence ticket demand are the subject of an increasing amount of sports economics studies, which usually focuses on factors like team performance, market size, star power, rivalry matchups, and stadium experience. Team performance is generally seen as a major generator of demand within that framework (Quansah, 2024). However, injury volatility adds another layer of uncertainty that is particularly severe in professional football. Injuries in the NFL are more common than in many other leagues and frequently involve high-impact positions, which affects both consumer expectations and game results. Injuries can cause abrupt demand shocks, but fans may react to performance trends gradually over the course of a season (Yilmaz et al, 2023). Despite this, in empirical ticket price research, injuries are often handled as anecdotal explanations rather than formally modeled variables (Greer, 2025). This presents a methodological gap since, particularly in a league like the NFL, injury data should be quantifiable and statistically testable whether markets respond to injuries.

Price prediction research has also grown quickly from a data science and computational standpoint due to the availability of more data and the usefulness of using regression and machine learning techniques to model markets. In many applied pricing studies, time-varying predictors are linked to pricing outcomes through structured data pipelines (Greer, 2025). The strength and significance of these predictors are then measured using statistical models. This strategy is analogous to real-world dynamic pricing systems in that computers modify prices in response to contextual data, anticipated attendance, and observed demand (Stavinova, 2023). The issue of openness still exists, though. The weighting of variables and the incorporation of changes in team conditions are rarely disclosed by ticketing systems or franchises. In that context, independently modeling the relationship between injuries, record, and ticket prices accomplishes two goals: it enhances consumer market clarity by determining which factors most consistently explain price swings, and it enriches scholarly knowledge of demand creation.

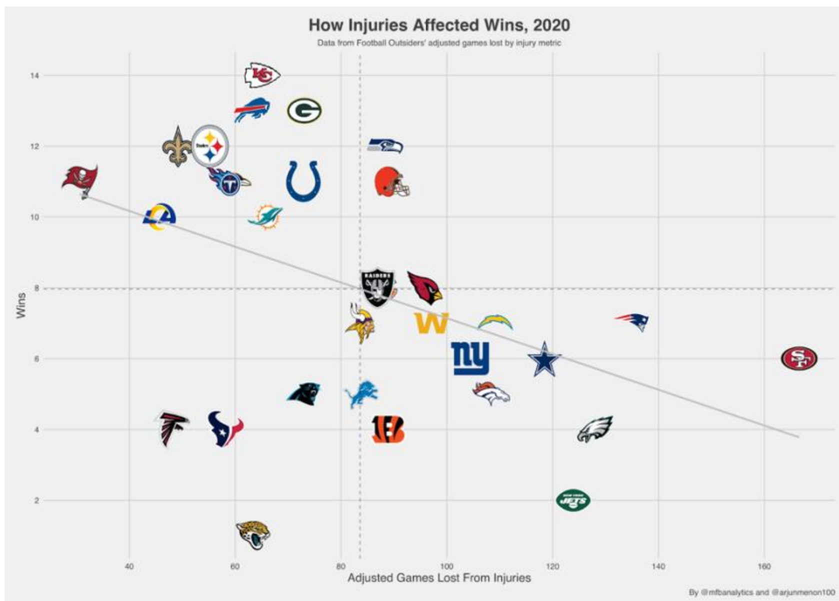

Figure 1: Injuries Impact on NFL Games

Taken from url:
https://lh5.googleusercontent.com/hPFi9GvfBT19IdAFun4yz86MMCcKq_8PB4OXYgGTWW_NS0tY1_-EZw40QzRnrmpGtMIh_rjyLMQz-Hxx1gGKXWjvlug-xeuQdV7gdorAH9e_mwrgVWlxuVmU0HpPXj2JuABRJ3uH

In this figure, we can see firsthand the impact NFL injuries have on team performance. Although this image was initially discovered on social media, it is based on information gathered by the Michigan Football Analytics Society (M-FANS), a student-run research group that specializes in football quantitative analysis. A variety of publicly accessible analytical studies based on recognized indicators, such as adjusted games lost due to injury, have been produced by M-FANS. Given that teams with higher injury loads typically record fewer victories, the graph clearly shows a negative correlation between injuries and team success. By supporting the notion that injuries have real competitive impact, this outside data supports the study's conclusions. Because weaker rosters and less successful teams diminish fan expectations and willingness to pay, these performance effects thus offer an economic channel through which injuries affect ticket demand. It is, thus, interesting to analyze the relationship between these injuries on ticket prices for all 32 teams in the NFL. When a team performs poorly every week, prices are bound to drop, and it will be interesting to see what affect and role injuries for a certain team has on ticket prices and consumer behavior.

2.1 Strengths and Limitations of Market-Based Ticket Pricing Research

One of the main advantages of the current research on sports ticket pricing is that it consistently confirms a fundamental economic principle: demand for anticipated entertainment value is reflected in ticket costs. Ticket price fluctuations offer a useful proxy for changes in demand in real time because ticket markets are consumer-facing and frequently very liquid (particularly in the NFL). By modeling ticket pricing against quantifiable indicators, like as team success, game importance, opponent strength, day-of-week scheduling, and macro factors like income and city population, researchers have taken advantage of this (Roll, 2025).

Studies on ticket prices frequently readily fit with microeconomic theory, which is a further strength. Teams in sporting environments provide a predetermined number of seats, but demand varies over time according to the anticipated caliber of the game (Choi, 2025). With limited supply and erratic demand, this structure is like traditional supply-and-demand models. Additionally, it promotes rigorous quantification: researchers may examine if rivalry games generate premiums, whether demand rises as teams win more, and whether playoff races result in "late-season surges." Numerous studies consistently show a "winning effect," in which more conflict and victory rates result in higher pricing, indicating that supporters act as logical buyers reacting to quality cues (Stavinova, 2023).

However, there are a few restrictions, particularly for studies that try to separate the effects of injuries. First, "team quality" is treated as a single construct in many ticket-pricing research, and it is measured by record, previous season results, or betting spreads. Instead of treating injuries as a separate shock, this may inadvertently bury them inside more general performance outcomes. Injuries have both direct and indirect effects (Quansah, 2024). Directly, they take away star power (fans may want to watch players), and indirectly, they reduce the likelihood that a team will succeed. The approach runs the risk of incorrectly attributing an injury-driven price shift to record or "team quality" alone if injury factors are not separated.

The second issue is temporal: the ticket market changes throughout the course of the week before a game in response to media narratives, injury reports, and news cycles. Measured "ticket prices" may represent noise rather than similar conditions if collection timing is irregular (Quansah, 2024). Unstandardized timing can lead to artifacts in research, such as gathering the price of one game right after a major injury announcement and the price of another game days before kickoff. This is especially important for injury-based models since the precise moment of damage confirmation is crucial.

Finally, although studies on ticket prices sometimes show association, injuries are still challenging. Although there may be a correlation between reduced costs and injuries, there may also be a correlation between injuries and other confounding factors like a team that is already having trouble, subpar quarterback play, or season-long pessimism (Humphreys & Johnson, 2020). Injury effects may be overestimated or underestimated in the absence of strict controls and regular measurement.

2.2 Key Research and Findings in Ticket Pricing

One of the most consistent findings is that fans are willing to pay more when the expected quality of an NFL game increases. The existence of a formidable opponent, star power, or team success are commonly used as proxies for "game quality." Because the NFL season is short and every game matters more than in leagues with more than 80 games, team success has a disproportionate impact. Because late-season games for contenders frequently turn into "must-win" situations, raising demand and driving up prices, the win-loss record becomes both a gauge of past performance and a forecast of future significance. Additionally, research highlights how behavior is shaped by expectations (Roll, 2025). A team on a winning streak may create more enthusiasm than a team with the same record but uneven performance. Perceived trajectory is just as important as current record (Stavinova, 2023). Similarly, teams with high preseason expectations could maintain demand even if early outcomes don't live up to expectations: that is, until the season definitively breaks in one direction.

A second element is that demand for tickets is frequently driven by players rather than just teams. To watch top quarterbacks, standout receivers, or athletes with historical significance, fans might pay more (Humphreys & Johnson, 2020). This idea is significant because it implies that injuries may affect demand even in cases where the team's record is steady. For instance, even if a team still has a chance to make the playoffs, losing a top quarterback could result in a steep decline in demand due to the increased uncertainty of competitiveness and entertainment value. Because they are a part of what people buy when they purchase a ticket, star players generate additional consumer surplus from an economic perspective (Page, 2023). The argument that injuries should be explicitly represented is supported by this line of investigation. Injury shocks lower the expected utility of going if fans place a high value on players (Humphreys & Johnson, 2020). This should show up in ticket pricing statistics.

Regression-based approaches continue to be fundamental in empirical pricing research because they offer interpretability; statistical tests can determine significance, and coefficients can be interpreted as marginal impacts. as estimating how much a ticket price changes as a predictor change while keeping other factors constant, OLS models are commonly employed as baseline approaches (Greer, 2025). OLS Models and machine learning pipelines for prediction are examples of more sophisticated methods. Interpretability is still important in academic work, though, particularly when the objective is explanation rather than just prediction.

2.3 Gaps and Challenges for Future Exploration

The fact that injuries are frequently not operationalized as quantifiable variables over time is a significant gap. Injuries are mentioned qualitatively in several research studies, but formal models do not incorporate injury counts, injury severity, or position-specific absence

(Humphreys & Johnson, 2020). This makes it more difficult to gauge how sensitive supporters are to the health of the roster. Injuries provide an ideal environment for researching how markets respond to uncertainty because the NFL is prone to injuries and weekly availability can fluctuate rapidly. A more robust injury-focused literature would explore whether injury sensitivity varies among fanbases or market sizes, differentiate between short-term injuries and long-term absences, and separate quarterback effects from other positions.

This study's practical implication is that, although ticket prices seem to effectively account for things like team success and injuries, customers are frequently ignorant of the precise causes of these shifts. Due to the private nature of ticket pricing systems, the methods underpinning these modifications remain opaque even when prices appropriately reflect underlying shifts in demand (Humphreys & Johnson, 2020). Fans may, therefore, see price swings without completely comprehending whether they are caused by injuries, shifts in the team's performance, or typical market timing impacts. This opens the door for more open, data-driven tools that analyze price changes rather than contesting their accuracy. A consumer-facing model, for instance, might assist in determining if a price reduction is most likely the result of a surge in injuries, a drop in anticipated performance, or more general market factors like an increase in ticket supply. Thus, the results of this study validate the effectiveness of ticket markets and the need of enhancing customer interpretability.

Extending the investigation beyond ticket markets to include economic effects at the city-level could be an essential step to study in the future. Local hotels, bars, restaurants, the demand for transportation, and temporary jobs can all be impacted by NFL home games (Humphreys & Johnson, 2020). Injuries and team performance may have an indirect impact on game-day economic activity if they have an impact on ticket pricing and attendance (Humphreys & Johnson, 2020). For instance, a team removed from postseason contention might not generate as much tourist as a late-season contender hosting a significant game. In a similar vein, an injury to a prominent player may result in fewer fans wanting to travel or lower overall spending in the stadium's entertainment sector.

Likewise, models beyond OLS may be tested in additional studies. The effects of injuries can be nonlinear; for example, losing a quarterback could result in a significant discontinuous decline, whereas losing a backup lineman might not matter. Additionally, interaction effects may be important. For example, injuries may have less of an impact when a team is winning but more of an impact when a team is already having trouble. Such nonlinearities could be modeled by machine learning techniques (random forests, gradient boosting) that compare predictive performance to classic regression while also explaining feature relevance using interpretability techniques.

Chapter 3

Methodology

This thesis employs a systematic empirical methodology based on data collection, statistical modeling, and computational modeling to determine how NFL ticket prices react to team injuries and performance. The objective is to measure the correlation between ticket prices (dependent variable) and team injuries (primary independent variable). This research stresses methodical timing, consistency of sources, and repeatability of analysis through Python-based modeling because ticket markets can shift quickly in response to breaking news and weekly outcomes. The selection of the dataset, the software environment used for data analysis, the regression models and modeling configuration selections, statistical measures used to test model strength, and visualization approaches used to understand results and confirm model assumptions are all described in this chapter.

3.1 Dataset Selection and Preparation

3.1.1 Rationale for Selecting TickPick and ESPN Injury Tracker

This research requires consistent measures of ticket prices and injury history across all 32 NFL teams. Two public sources were selected due to its systematic approach, repeatability, and relevance to fan-facing markets:

1. TickPick (Ticket Price Data): TickPick was chosen to offer a consistent “get-in” ticket price for every game. Because it reflects the lowest cost of attendance at the time of measurement, the “get-in” price, which is the lowest available ticket price for a particular matchup, is a useful proxy for baseline fan demand. The get-in metric also lessens distortion from outliers and more accurately represents the entry-level amount a typical fan would pay when compared to averages based on premium seating or limited sections. This approach does have some restrictions, however. When a single exceptionally cheap

ticket is available while the bulk of tickets are far more expensive, the lowest advertised ticket may occasionally be an outlier in and of itself. In some instances, this could result in an underestimation of overall market values. The get-in price is a useful and relevant metric for examining relative shifts in demand over time because, despite this possible flaw, it continues to be a standard and popular benchmark across teams and games.

2. ESPN Injury Tracker (Injury Data): ESPN's injury tracker served as the main source for team injury status to record injury implications. Weekly tracking of player availability changes over time is made possible by ESPN's organized and regularly updated injury listing for teams. Since the relationship between injuries and ticket prices is the focus of this thesis, injuries were considered the primary explanatory feature, with record acting as a control variable rather than the study's primary driver. ESPN Injury Tracker was also used to check which star players were injured on a week-by-week basis.

The study goal of creating a repeatable dataset that monitors pricing and injury patterns over time for each NFL team is supported by these sources taken together.

3.1.2 Scope and Time Window

The study covers the second half of an NFL regular season, starting in Week 9 and ending in Week 18. The start point for Week 9 is the point at which systematic data gathering was stable and practical. Crucially, there is significant variance in both independent variables currently:

- Injury patterns typically intensify as the season progresses, with more players missing time and cumulative injuries affecting roster strength.
- Team performance becomes clearer by midseason, shaping playoff contention narratives and fan demand.
- Market behavior becomes more sensitive because late-season outcomes (playoff contention, elimination, seeding) increase the stakes for fans.

Thus, this window offers a good opportunity to evaluate whether a rising injury load is correlated with lower ticket prices while taking into consideration how a team's record may mitigate or magnify that effect. In total, 1,671 data points and observations were collected throughout this process across 9 weeks in the NFL and weekly check-ins on of how many players were injured each week.

3.1.3 Data Collection Procedure

Implementing an organized and methodical collecting schedule to increase reliability and reduce noise was a crucial methodological choice in this project. Tickets and injury information for the following week's game were gathered on a regular basis for every NFL week, starting the day after the previous week's game ended. This data was collected by hand every week through TickPick and ESPN's injury tracker and confirmed through AI that all data points looked valid. This data was collected on the Sunday and Monday of each NFL Week leading up to a football game. For example, if an NFL game was on an upcoming Sunday, the data was collected on Monday right after the prior week's game ended. This timing restriction was created to minimize inconsistencies brought on by gathering at various times throughout the week and to standardize the market snapshot across teams. For instance, since the market changes over the week due to injuries, weather reports, and media storylines, the pricing would not be comparable if one team's statistics were gathered on Monday morning and another on Saturday night.

The dataset is better positioned to assess how changes in performance and injuries are reflected in pricing behavior because data is collected at the same time each week. For every team, each week, the following measures were recorded in a master spreadsheet: ticket "get-in" price and injury count per team. Ticket Prices were collected for each NFL Team across Week 9 and 18 of the NFL Season. After recording the get-in ticket price, ESPN's injury tracker was consulted to check the number of total injuries and key injuries for NFL teams. This process was followed thoroughly throughout the NFL Weeks and verified by hand across every week. Essentially, the dataset was constructed by repeatedly observing the same future games over time. For example, if I began collecting my data for a Week 10 game, I would also collect data points for that team's remaining schedule all the way up to Week 18. This means that each game had several price observations and can show how prices fluctuated over time through nine weeks. This repeated procedure produced a structured panel-style dataset across weeks and teams.

3.1.4 Data Integrity and Quality Checks

Several measures were taken to avoid recording errors and guarantee data fidelity because the dataset was constructed manually through weekly observation. This means that information about every game was gathered and documented on a weekly basis by directly consulting publically accessible sources, like TickPick for ticket prices and ESPN for injury reports after each game week ended. Prior to ending the session, values were first entered into the spreadsheet during collection and checked against the same sources. Second, the dataset was examined after each week's data was gathered to make sure there were no missing entries, improper team allocations, or mismatched week/opponent combinations. Lastly, team records were examined over the course of several weeks to make sure they developed rationally (for example, a team could not win twice in one week). To ensure that injury counts accurately reflected updates rather than transcribing errors, they were also examined for significant, unexpected changes.

3.2 Software Environment

3.2.1 Hardware and Operating Setup

The Mac (macOS) environment was used for all primary work. Since the modeling technique mostly uses regression-based statistical analysis rather than deep learning architectures, the analysis pipeline was made to be lightweight and repeatable without requiring specific GPU resources.

3.2.2 Data Management Tools and Libraries/Packages

Two main tools were used:

- Microsoft Excel (Data Collection and Storage): The main database for the weekly data collecting was Excel. During manual collection, it facilitated weekly organizing, structured tabular entry, and rapid validation checks. When inconsistencies were found, Excel's flexibility allowed for quick evaluation and adjustment. The final Excel Data file included week by week price and IR Injury counts as well as a number from one to five reflecting the amount of key injuries to players to further test that impact on NFL Ticket Pricing.
- Python (Statistical Modeling and Reporting): Regression modeling, data cleansing, exploratory analysis, and diagnostic visualization were all done with Python. Both hosted notebook environments and the terminal were used to run scripts locally.

These libraries and packages were used as well:

- pandas for tabular manipulations, cleaning, and data ingestion.
- NumPy for creating features and performing numerical calculations.
- statsmodels for statistical summaries (coefficients, p-values, R2 diagnostics) and OLS regression estimates.
- scikit-learn for optional predictive evaluation procedures and regression support.
- matplotlib for data visualization

3.2.3 Python Execution Environments

Python was executed in two primary contexts:

- Local Execution (Mac Terminal): Repeatable pipelines, rapid debugging, and model iteration were all done via local scripts.
- Cloud Based Notebooks (Google Colab): Python was executed in a structured notebook workflow using Google Colab. This environment facilitates consistent generation of

model summaries and graphics in a controlled workspace, support reproducibility, and enable clean reporting outputs.

3.3 Model Selection and Configuration

3.3.1 Rationale for Regression-Based Modeling

Quantifying the relationship between changes in injuries and changes in ticket pricing while accounting for record is the main objective of the study. This thesis stresses regression-based modeling rather than black-box prediction alone since interpretability is crucial, especially when describing the link in terms of consumer demand and economics. The baseline model treats ticket price as a continuous dependent variable and injuries as the primary explanatory variable. This baseline establishes whether injury levels alone show a negative association with price. To account for the fact that teams with higher records may maintain demand even when injured, win-loss was incorporated as a control variable. That model examines whether injuries are still statistically significant when performance is considered. Controlling for win percentage enhances interpretability and lessens omitted variable bias because performance can both directly influence demand and indirectly construct injury narratives (for example, injured contenders still attract fans).

Although there is a high and persistent correlation between injuries and ticket prices according to the regression results, these estimates shouldn't be taken as strictly causal. Even though injuries are frequently viewed as unexpected shocks, there are instances where team context and strategic decision-making are related to injury designations. Teams who are no longer in the running for the playoffs, for instance, would be more inclined to rest important players or disclose injuries cautiously, which could artificially inflate the number of injuries reported while also reflecting reduced underlying demand. In these situations, injuries are not totally unrelated to anticipated team performance or audience engagement, which raises the risk of unnoticed confounding variables. Therefore, rather than representing a completely causal effect, the coefficients should be understood as expressing a strong association. However, the consistency of results across many specifications, including models that account for fixed effects and team performance, increases confidence that injuries have a significant and independent impact on ticket pricing.

3.3.2 Model Variation and Further Exploration

Multiple variations were tested to validate stability:

- Alternative Performance Encoding: win percentage against raw wins
- Examining if injuries have a different impact on prices for winning teams than losing teams is an example of interaction terms.

The objective was to investigate if the injury-price association holds true under plausible alternative specifications rather than just fitting a model once.

3.3.3 Python Model Implementation

Python was used to implement each model, which was periodically run when variables were changed. Because the statsmodels framework generates comprehensible summary tables with standard errors, t-statistics, p-values, confidence intervals, and global fit statistics, it was utilized for OLS models. The more relevant values from this output were then formatted within Excel to output in Chapter 4. The model code itself can be referenced in Appendix A as well.

3.4 Evaluation Procedure and Metrics

3.4.1 OLS Summary Statistics

Standard OLS diagnostic outputs were used to assess the model's strength and interpretability. These statistics provide evidence for whether injuries and record meaningfully predict ticket prices in the dataset:

1. Valid Coefficient estimates (β-values): direction and magnitude of injury and performance effects.
2. p-values: whether coefficients are statistically significant relative to a chosen alpha threshold.
3. R-squared and Adjusted R-squared: depicts strength between variables and helps explain the relationship.
4. F-statistic: explains variance between variable groups.

3.4.2 Robustness and Sensitivity Checks

To ensure the validity of results, sensitivity checks on the computational models were often performed. This process helps strengthen the credibility of findings by demonstrating if the dataset was adequate to draw conclusions for this study:

- Re-running models to see the same values and check for outliers.

3.5 Visualization and Reporting

3.5.1 Exploratory Visualizations

Verifying model assumptions and comprehending the relationship between variables were the two functions of visualization. Important images included:

- Regression fit lines may be used in scatter plots of injuries versus ticket prices.
- Trend charts for teams that display injury numbers and ticket costs by week.
- Distribution plots to comprehend skew and variance

3.5.2 Reporting Framework

Lastly, tables and figures appropriate for thesis delivery were created using the standardized model outputs:

1. Quantitative MA consistent format for OLS summaries (key coefficients and significance levels).
2. Tables comparing model variants and fit metrics.
3. Visuals that highlight the directionality of injury effects under different team performance levels.

This methodical reporting strategy guarantees that individuals can compare models and understand findings with clarity.

Chapter 4

Analysis Results

This chapter depicts the model code and results for NFL Ticketing Analysis. Through this section, the code will be explained in-depth, and the regression results will outline the various regressions utilized to present the results. For regression analysis, another column was made labeled Key Injuries – limited to starting quarterback, wide receiver, offensive lineman, defensive end, and defensive back – to offer further analysis on how these key injuries also impact NFL ticket pricing. The results are organized by code and its respective plots and OLS summaries, alongside qualitative visualizations. Each section also highlights model-specific strengths and limitations. By comparing each model's outputs side by side, this chapter lays the groundwork for the broader discussion and implications explored in Chapter 5.

Building a comprehensive injury tracking system for all thirty-two NFL teams over the season was a significant part of the dataset used in this analysis. The publicly accessible reports from ESPN's NFL injury tracker, which provide weekly updates on injured players and their status, were the main source of injury statistics. Every week, the number of injuries sustained by each team was documented and included to the dataset along with ticket prices. A secondary variable called Key Injuries was added to capture differences in injury importance. This variable specifically monitored injuries to five high-impact positional groups: defensive end, defensive back, offensive lineman, wide receiver, and starting quarterback. On a scale of zero to five, the number of injuries impacting these crucial positions was noted for each game observation. This differentiation made it possible for the analysis to distinguish between roster injuries in general and injuries that are more likely to affect fan demand and perceived game quality.

4.1 Model Code

Python was employed to create several regression models to assess the connection between NFL ticket pricing and team injuries. The pandas package was used to import the manually produced dataset, which was saved as a structured CSV file, into the analysis environment. To guarantee consistency and dependability, data cleansing techniques were first used. Ticket prices (GetInPrice), overall injury counts (IR_Count), and key injury counts (Key_Injuries) were all translated into numeric format. Additionally, observations with missing values were not utilized. In this process of collecting data, (since I was tracking prices over every week), once a NFL game had passed for a given week, that cell would be empty for a ticket price when it came to recording down the ticket prices for a certain week. For example, if we consider a Week 9 game, in my data collection table, cells after that game had occurred would simply be null because at that point, there were no ticket prices for that game as it had already happened. By removing these observations, it was ensured that only completely observed and pertinent data points were used to estimate all regression models.

To correct for the skewed distribution of prices and lessen the impact of extreme values, a logarithmic transformation of ticket prices was developed after this preparation. Regression coefficients can be interpreted in percentage terms thanks to this converted variable (LogPrice), which is common in economic modeling of price dynamics. The statsmodels package, which offers comprehensive Ordinary Least Squares (OLS) summaries including coefficient estimates, standard errors, p-values, and goodness-of-fit measures, was used to implement all regression models.

To compare various injury measurements, multiple regression specifications were computed. The impact of total injuries on ticket pricing was investigated in the first model. To determine whether injuries to key positional players had a greater impact, the second model substituted key injury counts for total injuries. Other models included fixed effects for teams and weeks, resilient standard errors, and logarithmic price transformations. These enlarged models aid in accounting for seasonal and team-specific variables that could affect ticket costs apart from injuries.

To make statistical analysis easier, the dataset utilized for the regression models was arranged concisely. Variables including team name, week of the season, average ticket price, total injury count, and key injury count were all included in each observation, which represented a single team-game instance. The analysis was able to investigate the relationship between changes in ticket market valuations and variations in team health over the course of the season by merging injury and ticket price data into a single structured dataset. Additionally, this structure made it possible to use Python for effective regression modeling and diagnostic testing. Model code for this study is available in the Appendix A section for reference.

4.2 Regression Results

To comprehend the fundamental structure of the dataset and find possible trends between injuries and ticket prices, exploratory analysis was done prior to generating the regression models. According to preliminary summary figures, ticket prices vary significantly between teams and weeks, indicating variations in opponent quality, team performance, and market size. Throughout the season, injury totals also varied significantly, especially in the final weeks when cumulative injuries tend to rise. These differences provide a helpful setting for determining whether variations in ticket pricing could be statistically linked to shifts in team health.

Model 1: To first examine the relationship between injuries and ticket prices, an Ordinary Least Squares (OLS) regression was performed using total injuries (IR_Count) as the independent variable and ticket price (GetInPrice) as the dependent variable. The findings show a statistically significant inverse association between ticket prices and injuries. The coefficient on total injuries was −5.7577 ($p < 0.001$), which indicates that the average ticket price decreased by almost $5.76 for every extra injury. The idea that more injuries lower fan demand and ticket prices is supported by this finding. With an R-squared value of 0.043, the model's explanatory power was constrained, meaning that total injuries only account for 4.3% of the variation in ticket prices. This specific model can be found in Appendix A.

Model 2: Then, with key injuries (Key_Injuries) as the sole independent variable, the regression was conducted again. The association is significantly stronger, according to the results. The coefficient was −26.2959 ($p < 0.001$), meaning that the average ticket price dropped by about $26.30 for every additional major injury. When compared to total injuries, the explanatory power more than doubled as the R-squared rose to 0.094. This implies that the demand for tickets is considerably more affected by injuries to crucial positions. This can also be found in Appendix A.

Model 3: next, a multiple regression that included key injuries and total injuries was computed. The statistical significance of both variables was maintained ($p < 0.001$). With a value of −25.6071, key injuries had a greater effect than overall injuries, which had a coefficient of −5.4193. Combining both injury measurements increases explanatory power, as evidenced by the R-squared rising to 0.132. This demonstrates that although total injuries are important, major injuries have a far greater impact on ticket costs.

Log-Linear Models: Log-linear models were estimated due to the right-skewed ticket prices. Model 4: Log Price vs. Total Injuries: According to the coefficient of -0.0638 ($p < 0.001$), the cost of tickets decreases by about 6.38% for every additional injury. Compared to the linear model, the R-squared improved to 0.065. Model 5: Log Price vs. Key Injuries: With a coefficient of −0.2300 ($p < 0.001$), key injuries had an even greater impact, resulting in a 23.0% drop in ticket prices per key injury. R-squared went up to 0.088. Both factors continued to be significant in Model 6: Log Price vs. Both Injury Measures. The greater impact was still seen in key injuries. With an R-squared of 0.146, this model has the maximum explanatory power. This is the model that is the strongest overall.

Regarding pricing behavior in secondary ticket marketplaces, the log-linear models' performance aligns with economic predictions. Percentage variations in price better reflect how markets react to new information since ticket prices frequently show multiplicative rather than just additive increases. In this situation, injuries serve as a signal that affects consumer expectations regarding the caliber of games and the competitiveness of teams. As a result, the log-linear framework offers a more accurate depiction of how ticket markets take injury data into account when fans choose whether to buy tickets. These models and their results are referenced in Appendix A.

Now, I will include the overall tables grouping Model 1, 2, and 3 as the Linear Model and a second table grouping Models 4, 5, and 6 as the Log Models. These models also have an overall equation that explain the regression corresponding to the overall Linear and Log models. Table 1 reports the linear regression specifications relating ticket prices to injury measures. Table 2 reports the log-linear specifications, which provide a better fit and allow for percentage-based interpretation of injury effects:

Table 1: Linear Price Models			
Dependent variable = GetInPrice ($). Entries are coefficients with standard errors in parentheses.			
Panel A. Linear models			
Injury Count	-5.758***	—	-5.419***
	(0.663)		*(0.632)*
Key Injuries	—	-26.296***	-25.607***
		(2.002)	*(1.962)*
Home Game	—	—	—
Week FE	No	No	No
Team FE	No	No	No
Intercept	162.914***	198.472***	237.132***
	(5.424)	*(6.353)*	*(7.683)*
R-squared	0.043	0.094	0.132
N	1,671	1,671	1,671
*Notes: *** p < 0.001. Model 1 uses total injury count only. Model 2 uses key injuries only. Model 3 includes both injury measures. These specifications represent the baseline OLS models without team or week fixed effects; fixed effects models are presented separately later in this section.*			

Table 1: Linear Price Models Grouped

$$Price_i = 237.132 - 5.419\, IR\_Count_i - 25.607\, Key\_Injuries_i$$

A strong foundation for comprehending the relationship between injuries and ticket costs is provided by the linear parameters listed in Table 1. The hypothesis that more injuries lower consumer demand for NFL games is supported by the negative and statistically significant coefficients on injury variables in all three models. Models 2 and 3 depict that key injuries have a significantly greater impact on prices, whereas Model 1 indicates that the overall injury count has a moderate effect. This implies that fans are giving high-impact athletes' ailments more weight rather than responding to all injuries equally. Even though injuries are significant, the comparatively low R-squared values show that they only account for a small percentage of the overall price fluctuation, leaving opportunity for other market considerations.

Table 2: Log Price Models				
Dependent variable = log(GetInPrice). Entries are coefficients with standard errors in parentheses.				
Panel B. Log models				
Variable / Specification	**(4)**	**(5)**	**(6)**	**(7)**
Injury Count	-0.0638***	—	-0.0608***	—
	(0.006)		*(0.006)*	
Key Injuries	—	-0.2300***	-0.2223***	-0.2082***
		(0.018)	*(0.018)*	*(0.018)*
Home Game	—	—	—	—
Week FE	No	No	No	No
Team FE	No	No	No	No
Intercept	4.970***	5.180***	5.614***	5.061***
	(0.049)	*(0.058)*	*(0.069)*	*(0.057)*
R-squared	0.065	0.088	0.146	0.079
N	1,671	1,671	1,671	1,604
*Notes: *** p < 0.001. Model 7 is the outlier-removed robustness specification using the IQR method. All specifications represent baseline log-linear OLS models without team or week fixed effects; fixed effects models are presented separately later in this section.*				

Table 2: Log Price Models Grouped

$$\log(Price_i) = 5.614 - 0.0608\, IR\_Count_i - 0.2223\, Key\_Injuries_i$$

Table 2's log-linear parameters offer a more robust and accurate depiction of pricing behavior. The models capture proportional rather than absolute increases by converting ticket prices into logarithmic form, which more accurately depicts how secondary ticket markets function. All injury factors continue to have negative and statistically significant coefficients, which is consistent with the linear models. Key injuries lower ticket prices by roughly 20–23% each occurrence, according to the current interpretation of the impacts' magnitude in percentage terms. Furthermore, all these models' higher R-squared values show stronger explanatory power, indicating that the log transformation more closely matches the underlying data. This confirms that the best specification for examining the dynamics of ticket prices is log-linear models.

Model Following Outlier Removal: The model was re-estimated after outliers were eliminated to guarantee robustness. The statistical significance of the relationship was maintained, and the coefficient remained negative with a similar magnitude, indicating that extreme values are not driving the observed effect. To further account for unobserved heterogeneity, a fixed-effects model was also estimated including both team and week fixed effects, which control for time-invariant team characteristics (such as the market size or fan base attendance) and common weekly shocks across the league (season timing, playoff implications, etc.). Even after incorporating these controls, key injuries remained statistically significant, reinforcing that the relationship is not driven by differences across teams or specific weeks.

While the dataset is collected over time, the "dynamic" aspect is partially captured through the inclusion of week fixed effects, which account for systematic changes in pricing patterns as the season progresses. However, the model does not explicitly control for the exact number of days or weeks remaining until each game at the time of price observation. As a result, some short-term pricing dynamics like late-week price adjustments driven by resale activity or last-minute demand shocks may not be fully captured. Including a more precise time-to-game variable represents a valuable extension for future research.

Variable	Coefficient	Std Error	R^2	F-stat	N
Intercept	5.0608	0.057	0.079	137.5	1604
Key_Injuries	-0.2082	0.018			

Variable	Coefficient	Std Error	z	P>|z|	CI Lower	CI Upper
Week 9	3.7208	0.073	51.065	0	3.578	3.864
Week 10	3.7331	0.073	51.237	0	3.59	3.876
Week 11	3.6789	0.073	50.497	0	3.536	3.822
Week 12	3.6683	0.073	49.991	0	3.524	3.812
Week 13	3.6786	0.075	49.192	0	3.532	3.825
Week 14	3.7037	0.077	48.302	0	3.553	3.854
Week 15	3.7029	0.079	46.842	0	3.548	3.858
Week 16	3.6887	0.08	46.031	0	3.532	3.846
Week 17	3.7087	0.079	47.01	0	3.554	3.863
Week 18	3.7325	0.098	38.243	0	3.541	3.924
Key_Injuries_z	-0.2007	0.014	-14.847	0	-0.227	-0.174
HomeGame	0.0199	0.03	0.663	0.507	-0.039	0.079

Variable	Coefficient	Std Error	R^2	F-stat	N
Intercept	3.416***	(0.066)	0.485	4539	1671
Key Injuries	0.089***	(0.011)			
Weeks_To_Game	0.0029	(0.005)			
Home Game	0.0198	(0.030)			
Team FE	Yes				
Week FE	No				

Tables 3, 4, 5: Model After Outlier Removal, Week by Week Summary, Dynamic Log Price Model

The stability of the correlation between injuries and ticket prices is further supported by the outlier-removed model. The coefficient on important injuries is still negative, statistically significant, and comparable in size to the first log-linear values after removing extreme observations. This consistency suggests that the results represent a more widespread and enduring pattern in the data rather than being influenced by a few exceptionally high- or low-ticket prices. Therefore, it can be said that the detrimental effect of injuries on ticket pricing is resistant to different sample specifications.

By accounting for unobserved variations between teams and weeks, the fixed-effects model offers a more stringent test. Key injuries continue to be quite significant and show a negative correlation with ticket pricing even after taking these things into consideration. The significant rise in R-squared indicates that a significant amount of price volatility may be explained by team-specific traits and weekly circumstances. Beyond these controls, however, the injury effect's durability suggests that injuries have a steady and independent impact on ticket pricing. This supports the idea that ticket market pricing decisions actively take injury information into account.

A dynamic specification that included a time-to-game variable was estimated to expand the analysis. A more direct evaluation of whether short-term pricing changes affect the relationship between injuries and demand is made possible by this model, which explicitly considers how far in advance ticket prices were seen. The findings indicate that the time-to-game variable is not statistically significant, indicating that the proximity of the game at the time of observation does not significantly influence pricing. However, the coefficient on important injuries turns positive when team fixed effects are added to this dynamic component. This change suggests that injury variance may capture other phenomena, such as late-season contention effects or pricing rigidity, rather than just lowering willingness to pay when correcting for persistent team-level demand disparities. This emphasizes that the interpretation of injury effects depends on model specification and shows the significance of taking underlying team demand into account, even though it does not reverse the baseline finding that injuries are linked to lower prices.

When combined, the different model specifications show a pattern that is consistent across different analytical techniques. The association between injuries and ticket prices is nevertheless statistically significant whether basic linear regressions, log-linear transformations, or models that take team and week effects into consideration are used. The results' consistency across many specifications increases trust in the findings' dependability. Additionally, the greater magnitude linked to major injuries implies that supporters react more strongly to injuries involving players who are highly visible or strategically significant.

All things considered, the regression study offers compelling statistical proof that injuries considerably lower NFL ticket costs. The effect of key injuries is significantly greater than that of total injuries, indicating that injuries affecting key players and positions are the ones that fans react to the most. The best fit was obtained by the log-linear models, suggesting that damage effects are proportional rather than absolute. These results provide credence to the theory that injury data is factored into ticket market price and has a quantifiable impact on ticket value.

Chapter 5

Discussion

To comprehend how injuries affect NFL ticket pricing, this section analyzes the regression results. Clear trends about the connection between team health and ticket demand were found by looking at several regression specifications, such as linear, log-linear, and fixed-effects models. The findings offer statistical support as well as useful information on how the ticket market reacts to shifts in the makeup of the squad, especially when important players are injured.

Understanding how the statistical findings translate into observable behavior in the ticket marketplace is a crucial component of this conversation. Secondary resale platforms, where prices swiftly adapt to new information, are the primary means of operation for NFL ticket markets. Throughout the week before a game, there are often updates on player availability, depth chart modifications, and injury news. The ticket market frequently responds right away because these adjustments affect fans' expectations regarding team competitiveness and entertainment value. Therefore, the regression results represent a more comprehensive economic mechanism wherein customer purchase decisions are influenced by publicly available performance information.

5.1 Overall Performance Trends

Injuries showed a consistent and statistically significant negative correlation with ticket prices across all regression models. Reductions in ticket prices were linked to both total injuries and key injuries, with the impacts of major injuries being significantly greater. Log-linear models outperformed linear models in terms of fit, suggesting that injuries had a proportionate rather than a fixed dollar impact on ticket pricing. Furthermore, models that included both injury variables explained more variation than models that just included one, indicating that the availability of key players and the general health of the squad both influence ticket prices. Even though the total R-squared values were low, this is to be expected in real-world economic systems where a wide range of other factors, like market size, playoff implications, and opponent quality, affect ticket pricing.

The significance of information timing throughout the NFL season is another important consideration. As the season goes on, injuries mount, making player availability more unclear in later weeks. Because of this, later in the season, when playoff placement or divisional standings are at stake, fans might become more sensitive to injury information. The persistence of the negative association implies that ticket markets routinely incorporate this information as the season progresses, even if the regression models mainly concentrate on injury counts.

5.2 Specific Observations

The regression analysis produced several important findings. First, the impact of crucial injuries was four to five times greater than that of general ailments, indicating that fan demand is disproportionately impacted by injuries to starting quarterbacks and other important players. Second, the adoption of log-linear modeling in ticket pricing analysis was supported by the log-price models' greater statistical performance and better residual distributions. Third, even after correcting for team and weekly effects and eliminating outliers, the association between injuries and ticket prices remained statistically significant, suggesting that the results are reliable and unaffected by outliers. These findings show that injury information is regularly factored into ticket prices.

This trend may be explained by the fact that spectators appreciate watching well-known athletes and players more. In professional football games, starting quarterbacks and other standout skill-position players are frequently the main points of attention, and their absence from a game can drastically lower the perceived value of going in person. Because of this, injuries to these players may change what fans expect from the matchup's overall level of competition or entertainment value. This interpretation is supported statistically by the higher coefficients found for significant injuries in the regression models.

5.3 Implications for NFL Ticket Pricing

The analysis's conclusions show that NFL ticket costs react strongly to injury news, especially when those injuries affect important players like starting quarterbacks and other starting positional players. Economically speaking, this represents shifts in customer willingness to pay because injuries lower perceived game quality and entertainment value as well as predicted team performance. When star players aren't available, fans are less likely to buy tickets at higher costs, which causes quantifiable drops in market pricing (Humphreys & Johnson, 2020). The assumption that not all injuries have the same economic weight, and that demand is disproportionately shaped by the absence of high-impact players is supported by the key injury coefficient's bigger magnitude when compared to general injuries.

To comprehend the extent of these consequences, it is especially crucial to distinguish between total injuries and critical injuries. While minor injuries to depth players are common for clubs during the season, not all injuries have the same economic impact. Fans are more likely to notice injuries to starting quarterbacks, offensive playmakers, or key defense players, which could affect their expectations for the outcome of the game. The data shows that ticket markets react most strongly to injuries that significantly impact team performance by clearly differentiating crucial positional injuries from overall injury numbers.

These findings have significant ramifications for teams' and secondary ticket marketplaces' dynamic pricing tactics. A real-time, measurable variable that might enhance pricing precision and market responsiveness is injury data (Drayer, Shapiro, and Lee, 2017). Teams and platforms may be able to modify prices more successfully in response to anticipated demand if they include injury severity and key player availability in pricing models. Furthermore, the ability of injury data to function as a trustworthy predictor in predictive pricing algorithms is suggested by the durability of statistically significant associations across various model specifications.

It is also possible to predict the possible financial impact of injuries on team revenue using the regression results. A straightforward extrapolation can be made by combining the projected price effects with normal stadium attendance levels, even though the models directly assess changes in ticket prices rather than total revenue. For instance, a single important injury is linked to a roughly 20 - 23% drop in ticket prices using the preferred log-linear formulation. Assuming an average NFL ticket price of approximately $150 and a stadium attendance of 70,000, this corresponds to a revenue decline of roughly $2.1 - $2.4 million per game. Even under conservative assumptions, this represents a substantial financial impact, highlighting how changes in team health can translate directly into measurable economic losses for franchises. This analysis emphasizes the substantial financial effects of player injuries even if it ignores variables like dynamic pricing adjustments, secondary market behavior, or attendance fluctuations. These results imply that injuries have significant financial ramifications for clubs and the larger sports industry in addition to having an impact on on-field performance.

More generally, this study demonstrates how sports markets swiftly incorporate fresh data and modify pricing in accordance with economic ideas of efficient markets. Ticket prices react dynamically to shifting expectations during the season rather than just reflecting a team's static reputation. This shows how real-time performance analytics is becoming increasingly important in sports economics and shows that computer science-based modeling techniques may effectively capture these linkages. In the end, injuries are a significant and quantifiable factor that affects ticket prices and adding injury data to prediction models can improve both theoretical knowledge and real-world pricing applications in the NFL ticket market.

5.4 Challenges and Potential Improvements

Determining the scope of the thesis in a way that was both analytically sound and practically achievable was one of the main difficulties I faced with this study. The original version of the thesis sought to investigate dynamic pricing in the NFL in general, including its impact on fan behavior, team revenue, and even city-level economic implications. Although theoretically appealing, this scope turned out to be overly broad given the limitations on data availability and the difficulty of separating causal links between various economic layers. As the project progressed, more literature review and exploratory data analysis showed that a more accurate and testable research framework was made possible by focusing on team-level factors, particularly injuries. The study's validity and clarity were greatly improved by reframing ticket prices as the dependent variable and using injuries and performance as the main explanatory variables.

The availability of data for the entire season was a second significant constraint. Systematic data collection started at Week 9 instead of Week 1 due to difficulties in obtaining reliable injury and pricing data early in the season. This limitation prevents the model from capturing ticket prices and injury history from Week 1 through Week 8 and early-season phenomena like preseason optimism. Increasing the dataset's size to encompass the full regular season would probably strengthen the model's reliability, predictive power, boost statistical power, and enable more detailed temporal research, potentially including contrasting the effects of injuries early and late in the season. Despite these challenges, I believe that the current modeling approach provides a strong foundation for understanding the directional and statistical significance of the relationships under study.

5.5 Broader Significance

The results of this thesis advance our knowledge of how off-field economic behavior is influenced by on-field uncertainty. This study illustrates how fan demand reacts to expectations and perceived team quality in addition to results by providing factual evidence that injuries and team performance have a significant impact on ticket pricing. In this way, ticket prices become a real-time market signal that connects behavioral economics and sports analytics by expressing collective fan sentiment, risk assessment, and readiness to pay.

This work emphasizes the importance of interpretable statistical models in applied data science from a computational standpoint. The Python-based regression approach provides a transparent substitute for assessing pricing dynamics, whereas proprietary ticketing methods are still opaque (Yilmaz, Akdeniz, and Zhang, 2023). Franchises, secondary market platforms, or consumer advocacy organizations looking for more information about demand sensitivity and pricing fairness may expand or modify such models. Crucially, this strategy shows that, when combined with reliable statistical techniques, significant insights can be extracted from publicly accessible data.

This study underscores the wider importance of data analytics in contemporary sports management in addition to its implications for ticket pricing tactics. To comprehend fan behavior and demand trends, teams, ticket platforms, and sports economists are depending more and more on big statistics. The capacity to integrate statistical modeling, ticket pricing data, and injury reports shows how publicly accessible data may be converted into insightful economic analysis. Professional sports organizations' decision-making is still influenced by this combination of economic modeling and sports analytics.

Beyond football, the findings of this study have ramifications for professional sports leagues that depend on dynamic pricing systems outside the NFL. A transportable framework for examining how performance volatility and uncertainty impact consumer demand in live entertainment markets is provided by the modeling approaches and analytical pipeline described in this thesis. Altogether, this work strengthens the notion that data-driven openness may help both producers and consumers by contributing to the expanding convergence of computer science, economics, and sports analytics. Building more equitable, effective, and knowledgeable markets will require an understanding of the quantifiable elements that influence ticket prices as they continue to change.

APPENDIX A: Model Code for Reference

This appendix includes the model code that was employed to run the different regressions for this study. This code includes several comments on what is going on per section and is a great reference point to understand the tables presented in this thesis.

```
# ==========================================
# 1. LOAD DATASET (UPDATED CSV WITH KEY INJURIES)
# ==========================================

import pandas as pd
import numpy as np

df = pd.read_csv("NFL_Ticket_Injury_Tidy_WithKeyInjuries.csv")

print("Columns:")
print(df.columns)

print("\nPreview:")
print(df.head())


# ==========================================
# 2. CLEAN DATA
# ==========================================

df["GetInPrice"] = pd.to_numeric(df["GetInPrice"], errors="coerce")
df["IR_Count"] = pd.to_numeric(df["IR_Count"], errors="coerce")
df["Key_Injuries"] = pd.to_numeric(df["Key_Injuries"], errors="coerce")

df = df.dropna(subset=["GetInPrice", "IR_Count", "Key_Injuries"]).copy()

# Log price
df["LogPrice"] = np.log(df["GetInPrice"].clip(lower=1))

# ============================================
# 3. DESCRIPTIVE STATISTICS
# ============================================

print("\nFULL SUMMARY:")
print(df.describe())

print("\nKEY INJURY SUMMARY:")
print(df["Key_Injuries"].describe())

print("\nIR COUNT SUMMARY:")
print(df["IR_Count"].describe())


# ============================================
# 4. BASELINE REGRESSIONS
# ============================================

import statsmodels.api as sm

# ----------------------------------
# Model 1: Price ~ IR_Count
# ----------------------------------

X1 = sm.add_constant(df[["IR_Count"]])
y1 = df["GetInPrice"]

model1 = sm.OLS(y1, X1).fit()

print("\n=============================")
print("MODEL 1: Price ~ IR_Count")
print("=============================")

print(model1.summary())
```

```python
# ---------------------------------
# Model 2: Price ~ Key_Injuries
# ---------------------------------

X2 = sm.add_constant(df[["Key_Injuries"]])
y2 = df["GetInPrice"]

model2 = sm.OLS(y2, X2).fit()

print("\n==============================")
print("MODEL 2: Price ~ Key_Injuries")
print("==============================")

print(model2.summary())


# ---------------------------------
# Model 3: Price ~ BOTH injury measures
# ---------------------------------

X3 = sm.add_constant(df[["IR_Count","Key_Injuries"]])
y3 = df["GetInPrice"]

model3 = sm.OLS(y3, X3).fit()

print("\n==============================")
print("MODEL 3: Price ~ IR_Count + Key_Injuries")
print("==============================")

print(model3.summary())

# ==========================================
# 5. LOG PRICE REGRESSIONS (BETTER MODEL)
# ==========================================


# LogPrice ~ IR_Count

model_log_IR = sm.OLS(
    df["LogPrice"],
    sm.add_constant(df[["IR_Count"]])
).fit()

print("\n==============================")
print("MODEL 4: LogPrice ~ IR_Count")
print("==============================")

print(model_log_IR.summary())


# LogPrice ~ Key_Injuries

model_log_key = sm.OLS(
    df["LogPrice"],
    sm.add_constant(df[["Key_Injuries"]])
).fit()

print("\n==============================")
print("MODEL 5: LogPrice ~ Key_Injuries")
print("==============================")

print(model_log_key.summary())


# LogPrice ~ BOTH

model_log_both = sm.OLS(
    df["LogPrice"],
    sm.add_constant(df[["IR_Count","Key_Injuries"]])
).fit()

print("\n==============================")
print("MODEL 6: LogPrice ~ BOTH")
print("==============================")

print(model_log_both.summary())
```

```python
# ============================================
# 6. ROBUST STANDARD ERRORS
# ============================================

print("\nROBUST MODEL: LogPrice ~ Key_Injuries")

model_log_key_robust = sm.OLS(
    df["LogPrice"],
    sm.add_constant(df[["Key_Injuries"]])
).fit(cov_type="HC3")

print(model_log_key_robust.summary())


# ============================================
# 7. OUTLIER REMOVAL (IQR METHOD)
# ============================================

def remove_outliers_iqr(data, col):

    Q1 = data[col].quantile(0.25)
    Q3 = data[col].quantile(0.75)

    IQR = Q3 - Q1

    lower = Q1 - 1.5*IQR
    upper = Q3 + 1.5*IQR

    return data[(data[col]>=lower) & (data[col]<=upper)].copy()


df_iqr = remove_outliers_iqr(df,"GetInPrice")

print("\nOutlier removal:")
print("Original:",df.shape)
print("After:",df_iqr.shape)


model_iqr = sm.OLS(
    df_iqr["LogPrice"],
    sm.add_constant(df_iqr[["Key_Injuries"]])
).fit()

print("\nMODEL AFTER OUTLIER REMOVAL")

print(model_iqr.summary())
```

```python
# ============================================
# 8. WINSORIZED MODEL
# ============================================

from scipy.stats.mstats import winsorize

df_w = df.copy()

df_w["GetInPrice_wins"] = winsorize(df_w["GetInPrice"],limits=[0.01,0.01])

df_w["LogPrice_wins"] = np.log(df_w["GetInPrice_wins"])


model_wins = sm.OLS(
    df_w["LogPrice_wins"],
    sm.add_constant(df_w[["Key_Injuries"]])
).fit()

print("\nWINSORIZED MODEL")

print(model_wins.summary())


# ============================================
# 9. FIXED EFFECTS MODEL (CLEAN VERSION)
# ============================================

import statsmodels.formula.api as smf
from sklearn.preprocessing import StandardScaler

# Normalize key injuries (fix scaling issues)
scaler = StandardScaler()
df["Key_Injuries_z"] = scaler.fit_transform(df[["Key_Injuries"]])

# Remove intercept explicitly to avoid dummy trap
model_fe = smf.ols(
    "LogPrice ~ Key_Injuries_z + HomeGame + C(Data_Week) + C(Team) - 1",
    data=df
).fit(cov_type="HC3")

print("\nFIXED EFFECTS MODEL (CLEANED)")
print(model_fe.summary())
```

Figure 2: Model Code (Multiple Figures)

```
==============================
MODEL 1: Price ~ IR_Count
==============================
                            OLS Regression Results
==============================================================================
Dep. Variable:             GetInPrice   R-squared:                       0.043
Model:                            OLS   Adj. R-squared:                  0.043
Method:                 Least Squares   F-statistic:                     75.36
Date:                Sun, 01 Mar 2026   Prob (F-statistic):           9.14e-18
Time:                        19:22:03   Log-Likelihood:                -9842.4
No. Observations:                1671   AIC:                         1.969e+04
Df Residuals:                    1669   BIC:                         1.970e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        162.9144      5.424     30.034      0.000     152.275     173.554
IR_Count      -5.7577      0.663     -8.681      0.000      -7.059      -4.457
==============================================================================
Omnibus:                      327.823   Durbin-Watson:                   1.296
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              561.832
Skew:                           1.259   Prob(JB):                    9.99e-123
Kurtosis:                       4.315   Cond. No.                         21.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 3: Price and IR Count OLS Summary

```
==============================
MODEL 2: Price ~ Key_Injuries
==============================
                            OLS Regression Results
==============================================================================
Dep. Variable:             GetInPrice   R-squared:                       0.094
Model:                            OLS   Adj. R-squared:                  0.093
Method:                 Least Squares   F-statistic:                     172.5
Date:                Sun, 01 Mar 2026   Prob (F-statistic):           1.42e-37
Time:                        19:22:03   Log-Likelihood:                -9797.1
No. Observations:                1671   AIC:                         1.960e+04
Df Residuals:                    1669   BIC:                         1.961e+04
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const          198.4724      6.353     31.242      0.000     186.012     210.933
Key_Injuries   -26.2959      2.002    -13.135      0.000     -30.223     -22.369
==============================================================================
Omnibus:                      288.117   Durbin-Watson:                   1.396
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              468.602
Skew:                           1.143   Prob(JB):                    1.75e-102
Kurtosis:                       4.228   Cond. No.                         10.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 4: Price and Key Injuries OLS Summary

```
==============================
MODEL 4: LogPrice ~ IR_Count
==============================
                            OLS Regression Results
==============================================================================
Dep. Variable:               LogPrice   R-squared:                       0.065
Model:                            OLS   Adj. R-squared:                  0.064
Method:                 Least Squares   F-statistic:                     115.5
Date:                Sun, 01 Mar 2026   Prob (F-statistic):           4.23e-26
Time:                        19:22:04   Log-Likelihood:                -1961.3
No. Observations:                1671   AIC:                             3927.
Df Residuals:                    1669   BIC:                             3938.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.9701      0.049    102.406      0.000       4.875       5.065
IR_Count      -0.0638      0.006    -10.748      0.000      -0.075      -0.052
==============================================================================
Omnibus:                       26.744   Durbin-Watson:                   1.236
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               27.458
Skew:                          -0.302   Prob(JB):                     1.09e-06
Kurtosis:                       2.832   Cond. No.                         21.0
==============================================================================
```

```
==============================
MODEL 5: LogPrice ~ Key_Injuries
==============================
                            OLS Regression Results
==============================================================================
Dep. Variable:               LogPrice   R-squared:                       0.088
Model:                            OLS   Adj. R-squared:                  0.087
Method:                 Least Squares   F-statistic:                     160.1
Date:                Sun, 01 Mar 2026   Prob (F-statistic):           4.13e-35
Time:                        19:22:04   Log-Likelihood:                -1940.7
No. Observations:                1671   AIC:                             3885.
Df Residuals:                    1669   BIC:                             3896.
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const            5.1803      0.058     89.803      0.000       5.067       5.293
Key_Injuries    -0.2300      0.018    -12.654      0.000      -0.266      -0.194
==============================================================================
Omnibus:                       54.312   Durbin-Watson:                   1.306
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               44.506
Skew:                          -0.323   Prob(JB):                     2.17e-10
Kurtosis:                       2.529   Cond. No.                         10.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
==============================
MODEL 3: Price ~ IR_Count + Key_Injuries
==============================
                            OLS Regression Results
==============================================================================
Dep. Variable:             GetInPrice   R-squared:                       0.132
Model:                            OLS   Adj. R-squared:                  0.131
Method:                 Least Squares   F-statistic:                     126.7
Date:                Sun, 01 Mar 2026   Prob (F-statistic):           5.89e-52
Time:                        19:22:03   Log-Likelihood:                -9761.2
No. Observations:                1671   AIC:                         1.953e+04
Df Residuals:                    1668   BIC:                         1.954e+04
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const          237.1319      7.683     30.863      0.000     222.062     252.202
IR_Count        -5.4193      0.632     -8.569      0.000      -6.660      -4.179
Key_Injuries   -25.6071      1.962    -13.054      0.000     -29.455     -21.760
==============================================================================
Omnibus:                      305.132   Durbin-Watson:                   1.430
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              514.576
Skew:                           1.179   Prob(JB):                    1.83e-112
Kurtosis:                       4.354   Cond. No.                         33.4
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figures 5, 6, 7: Additional OLS Summaries

BIBLIOGRAPHY

1. 1075 The Fan. (2023). The average home ticket price for each NFL team from highest to lowest. https://1075thefan.com/618350/the-average-home-ticket-price-for-each-nfl-team-from-highest-to-lowest/
2. Bryant University Digital Commons. (n.d.). Economic analysis of sports ticket pricing. https://digitalcommons.bryant.edu/eeb/vol16/iss1/12/
3. CNBC. (2025). CNBC's official NFL team valuations 2025. https://www.cnbc.com/2025/09/04/cnbcs-official-nfl-team-valuations-2025.html
4. Choi, J. (2025). Ticket price elasticity and attendance in professional sports markets. Journal of Sports Economics. https://www.mdpi.com/2813-2203/4/4/34
5. DiscoverPHL. (2023). Philadelphia executive summary: 2022 year-end and monthly 2023 July report. https://www.discoverphl.com/wp-content/uploads/2023/07/Philadelphia_Executive-Summary_2022-Year-end_and_Monthly_2023_July.pdf
6. Drayer, J., Shapiro, S. L., & Lee, S. (2017). Dynamic ticket pricing in sport: An agenda for research and practice. Sport Marketing Quarterly, 26(4), 184–194. https://www.researchgate.net/publication/264697964_Dynamic_Ticket_Pricing_in_Sport_An_Agenda_for_Research_and_Practice
7. ESPN. (n.d.). NFL injuries. https://www.espn.com/nfl/injuries
8. Greer, J. (2025). Consumer perceptions of dynamic ticket pricing in professional sports markets. Journal of Sport Management. https://scholarworks.bgsu.edu/cosma/vol2/iss1/4/
9. Humphreys, B. R., & Johnson, B. K. (2020). The demand for superstar athletes in professional sports. https://ideas.repec.org/a/sae/jospec/v21y2020i2p152-175.html
10. Inside Injuries. (n.d.). The impact of injuries in sports performance. https://www.insideinjuries.com/articles/impactofinjuries
11. National Library of Medicine. (2021). Injury incidence and performance outcomes in professional sports. https://pmc.ncbi.nlm.nih.gov/articles/PMC8365220/

12. Page, J. (2023). Strategic pricing decisions in professional sports ticket markets. Journal of Revenue and Pricing Management.
13. Quansah, K. (2024). Determinants of football ticket pricing and demand. Journal of Sports Economics.
14. Roll, J. (2025). The economic determinants of spectator turnout: An investigation into professional sports attendance. National High School Journal of Science.
15. https://nhsjs.com/wp-content/uploads/2025/06/The-Economic-Determinants-of-Spectator-Turnout-An-Investigation-into-Professional-Sports-Attendance.pdf
16. Spofford, M. (2024). Packers finances remain on solid ground. Green Bay Packers. https://www.packers.com/news/packers-finances-remain-on-solid-ground-2024
17. Spofford, M. (2025). Packers' finances staying in good shape. Green Bay Packers. https://www.packers.com/news/packers-finances-staying-in-good-shape-2025
18. Stavinova, A. (2023). Predictive algorithms for ticket pricing optimization in sports analytics. https://ijrmeet.org/predictive-algorithms-for-ticket-pricing-optimization-in-sports-analytics/
19. TickPick. (n.d.). TickPick ticket marketplace. https://www.tickpick.com/
20. Visit Philadelphia. (2023). Center City Philadelphia hotel performance continued to improve in 2022. https://www.visitphilly.com/media-center/press-releases/center-city-philadelphia-hotel-performance-continued-to-improve-in-2022/
21. Yahoo Sports. (2023). How much injuries really impact an NFL season. https://sports.yahoo.com/article/much-injuries-really-impact-season-183417051.html
22. Yilmaz, O., Akdeniz, M. B., & Zhang, Y. (2023). Data-driven pricing strategies in sports ticket markets. Journal of Revenue and Pricing Management https://link.springer.com/article/10.1057/s41272-023-00427-9