Business 620

Allegheny College

Meadville, Pennsylvania 16335

Analyzing Attendance Determinants of the Arizona Diamondbacks

Jacob Charles Boord

April 8, 2024

Analyzing Attendance Determinants of the Arizona Diamondbacks

by

Jacob Charles Boord

Submitted to The Department of Business and Economics

Project Advisor: Professor Onyeiwu

Second Reader: Professor Ormiston

Date: April 8, 2024

I hereby recognize and pledge to fulfill my responsibilities as defined in the Honor Code and to maintain the integrity of both myself and the College as a whole.

Jacob C. Boord

Acknowledgments

First and foremost, none of my achievements, academic and beyond, are possible without my Lord and Savior, Jesus Christ. It is to God I owe the many blessings I have in my life. From my family, to my friends, I am so blessed each and every day through the gifts of God.

To my family, thank you for supporting me every step of the way. I know that Allegheny College was a leap of faith for me, and I am so thankful for how y'all had my back each step of the way. Mom, thank you for being my rock and always calling and facetiming me when you miss me. Your love is felt everywhere I go, together or apart. Thank you for being an inspiration to further my education. Dad, thank you for showing me how to be a responsible man in the world. You inspire my work ethic, keep me honest, and push me to be my very best. Your guidance has helped me achieve more than I ever thought possible. Mike, thank you for being the best bonus-dad around. You taught me so many valuable life lessons, and I could not be more thankful that you have been there every step of the way. I am blessed to have you as another father-figure in my life. Noah, thank you for causing me stress for not going… well I won't go there for Mom, Dad, and Mike's sake. You keep me in check, push me when I need to be pushed, but most importantly, you always have my back when I need you. These four years have been difficult apart, but I believe it ultimately has brought us closer together. You keep me young at heart even though you have given me grey hairs. Never change bud.

Bella, Rosie, Dale Tucker, Dakota Rose, Pumpkin, are and were the best pets a man could have. My fur babies and crazy doggo, you keep me grounded when I need it and provide me comfort. Even though Bella was supposed to replace me when I left for college, I am thankful for each and every one of you fluffy rascals.

To Jake and Josh, my first friends I can remember, y'all will always hold a special place in my heart. All the days spent in the back yard playing wiffle-ball, getting into fights, and having fun like kids should made me into the person I am today. There is no doubt in my mind that I would not be where I am today without those experiences together.

To Jared and Tristan, thank you for being my brothers back home. I know that if I am ever in a pinch, I can count on both of you to be there right by my side. Our brotherhood is strong, and for that I am forever thankful.

To Coach Linzy, thank you for being a father figure and mentor to me away from home. You are the reason I came to Allegheny College to pursue my degree and compete in track and field. Your recruitment of me has led me to my best friends, a strong education, and so many stories I will be able to tell my children one day about this place that I called home for four years. You are a hero to me, and I am so thankful God brought us together.

To Coach Wade, thank you for pushing me and being the older brother I never had. You have been there for me through thick and thin, hell or highwater, and have never once stopped fighting for me. You brought my throwing to the next level, and I am so grateful to be your first All-American. I will never forget all the laughs, tears, smiles, frowns, fights, and conversations about the most random things we have shared over the years. I am going to miss our car rides to distant meets and our pick it up and throw practices in the middle of the day. I would not be half the man I am today without you and your unconditional love. Love you long time brother.

To Coach Mourer, Coach Kim, Coach Tara, Coach White, Coach Wilson, Coach Williams, Coach T, Coach Bri and Coach Bri, Coach Murray, Coach Marques, Coach Corey, and Coach Jerod, each of you have played a huge role as my development as an athlete, but most

importantly, as a man. Each of you provided me with the tools to succeed in my sport and beyond. Without each of you, I do not accomplish half of what I have. My journey in athletics has had its many ups and downs, but what will always remain consistent is the support I have had along the way. Thank you to each and every one of you for your belief in me.

To all of my teammates, thank you for being there for me as I have been able to live out my dream of being a collegiate javelin thrower. The support at my highs and my lows has not gone unnoticed, and I am so grateful to have shared this journey with each and every one of you. Hope and Jonah, you both believed in me unconditionally as your coach but also your teammate. I am so proud of how far both of you have come in your javelin careers. The two of you succeeding in javelin is one of my finest achievements. I am beyond proud of both of you.

Thank you to the entire Business and Economics Department as well as the Computer Science department for allowing me to flourish as a student. I have been fortunate to serve as a TA across both disciplines thanks to the belief of the staff in me as a leader within the classroom. My experiences in the classroom prepared me for success in the real world, beyond the brick roads of Allegheny College. A special thank you to Professor Onyeiwu and Professor Ormiston, the advisors of my senior project, for pushing me to perform at my absolute best. My project has undergone many changes since its beginning, and I am confident in the product that I have produced. There have been many tiring work-shifts put into this senior project in order to have it be of the quality both of you pushed me to achieve. I am thankful for this rigorous task of meeting your expectations as it allowed me to produce a product that I was unaware I was capable of completing. I would not have this quality of work without your guidance and expertise. Thank you to both of you for being a driving force in the completion of this project.

And last, but certainly not least, my fellow Disciples of Tank. Lord knows I do not have enough words to describe how grateful I am that we all came together. Jacob King (Tex), Stephen Cullinan (Steve), Jackson Bowling (Shanghai), Mac Polny (Chipper), Landon Schmader (Tank), Colm Mulligan (Sully), Peter Lantz (Pete), Matt Reynolds (Matty Rey), and Mark Stover (Edwin). You boys are my brothers until the end of time. I do not know what I did to deserve such a well-rounded group of weirdos and jock-nerds, but I will never stop trying to express what each of you means to me. I have seriously too many memories to share, so I will in the order of the names listed previously share one quick core memory of mine.

Tex, watching you set the conference record in the weight throw brought tears to my eyes. You make me incredibly proud every day and are the smartest man I know.

Steve, watching you fight like hell to try and make nationals to try and compete together one more time was inspiring. I envy your grit and determination to succeed. Thank you for pushing me to be the best I can be.

Shanghai, working our lives away over the last summer for a company that never appreciated our work allowed me to see a side of you I did not get to see beforehand. You never quit and had an unreal battery (I swear you never stopped working). Thank you for being a part of my journey and helping me make party playlists to help keep your spirit alive.

Chipper, boy did we have our ups and major downs. I will never forget the phone call you made to me after I became an All-American. You are my brother for life, and I am so thankful we went through hell together.

Tank, for your courage and will inspire me every day. I will never forget all of the days spent together talking about random things and trying to screw Chipper over in Fantasy Football.

Sully, boy where do I even begin? Most of our stories are wild, but I am so thankful you came into the room that day looking for Pete. My favorite memory of us is all the time I have spent watching you grow as a man. I am so proud of you for how far you have come. Never stop being you man.

Pete, I am so thankful we were randomly assigned to live together for all four years of college! Crazy how it all worked out, but you have been there for me through my highs and lows. I do not think I could have shared a living space with anyone better. Watching you do the coldest gator-chomp in the history of gator chomps was the most badass thing ever. You better buy that farm so I can come hunt on it.

Matty Rey, I am so proud of how far you have come as well. You have never given up along the way. My favorite memory is all the "I like it when you do it right thurr right thurrrs" and random memes and stickers or all the days we spent in good lighting. Never stop being your original self.

Edwin, I am still mad at you for leaving us here at Allegheny for Pitt. I mean who even transfers for women other than you? Jokes aside, I will never forget when you looked me up and down last year, sizing me up the very first time I met you. You, Sully, and Tex together making up the OG Giggle Squad is just a small part of our journey together.

I love each and every one of you. Y'all make me a better person and I will forever be in debt to each and every one of you.

Table of Contents

Acknowledgments iii
Table of Contents viii
List of Tables ix
List of Figures x
Abstract xi
Chapter I: Introduction 1
Chapter II: Literature Review 6
Section I: Stylized Determinants 6
Section II: Dynamic Pricing 13
Section III: Gaps in the Literature 17
Chapter III: Empirical Analysis 19
Section I: Variable Definition 19
Section II: Regression Modeling 27
Section II.A: Model without Macroeconomic Considerations 27
Section II.B: Model with Macroeconomic Considerations 34
Section II.C: Regression Pre-Dynamic Pricing with Macroeconomic Considerations 41
Section II.D: Regression Post-Dynamic Pricing with Macroeconomic Considerations 46
Section III: Case Study Analysis 52
Chapter IV: Summary, Conclusions, and Policy Recommendations 59
Section I: Senior Project Summary 59
Section II: Findings Summary 60
Section III: Further Research Implications 62
Section IV: Policy Recommendations 63
Bibliography 65

List of Tables

Table 1: Independent Variables and Source Information 26

Table 2: Regression without Economic Considerations 28

Table 3: Regression with Macroeconomic Considerations 35

Table 4: Regression Pre-Dynamic Pricing Introduction 42

Table 5: Regression Post-Dynamic Pricing Introduction 47

List of Figures

Figure 1: Total Home Attendance by Year *53*

Figure 2: Average Attendance by Opponent *54*

Figure 3: Average Attendance by Day of the Week and Start Time *55*

Figure 4: Payroll by Year *56*

Figure 5: Payroll and Attendance by Year Overlay *57*

Abstract

Current literature regarding MLB attendance fails to address different promotion types as well as discuss the implications on attendance over multiple years. This senior comprehensive project focuses on the attendance determinants of the Arizona Diamondbacks, who have struggled with declining attendance rates since 2003. The project fills the gaps in the literature that fail to address different promotion types, long term studies on the impacts of attendance, and an in-depth approach to a single team. Most attendance studies focus on the league as an aggregate, including a variety of teams; however, this project combines regression analysis with a case study approach in order to uncover the explanations behind variations in attendance levels. Advancing the research in the attendance determinant studies on MLB franchises, this study finds that the Arizona Diamondbacks experienced negative attendance rates following the introduction of dynamic pricing. Furthermore, the study finds that promotions focused on children are less effective at drawing crowds than generic promotions. This senior project provides a background on the Diamondbacks, review of prior studies and literature, and provides a basis for further research in the field along with policy recommendations for executives.

Chapter I: Introduction

Year after year, Major League Baseball (MLB) teams search for ways to draw larger crowds. What better advantage for a home team to have than a fierce home crowd? Quantifying the number of fans each year that walk through the front gates is simple: collect the information from each ticket scanned at the gates to determine the total number of fans. However, understanding the reasoning for the fans walking through the gates allows the organization to accurately understand how to curate the best experience for their fan base.

An expansion team introduced by the MLB in 1998, the Arizona Diamondbacks sold out their Opening Day crowd of 50,179 (McManaman, 2018). Despite the 1998 season ending in a last-place finish in the National League (NL) West, the following year the Diamondbacks would go on to win 100 games and make the playoffs the quickest out of any expansion team. By 2001, the Diamondbacks achieved the pinnacle of MLB, winning the World Series. In doing so, the Diamondbacks became the fastest expansion team to win a World Series, defeating the New York Yankees in seven games (Pavlovic, 2023).

When it comes to historical franchises, MLB's most storied franchises boast a rich history of success and Hall-of-Fame caliber athletes (Rottenberg, 1956). According to historical data studied by Meyer (2023), the top three franchises in history are the New York Yankees, the St. Louis Cardinals, and the Los Angeles Dodgers. Absent from this discussion are the Diamondbacks, with such a young history compared to the rest of the MLB. The Yankees, Cardinals, and Dodgers face no issues filling seats due to prestige and history; however, the Diamondbacks have struggled to fill seats since their early seasons as a franchise.

According to Baseball Reference, the New York Yankees have only had one season since 1998 with an attendance total below three million for the season (All attendance data is collected from Baseball Reference, cited in the Bibliography). This one season was the season following COVID-19. The Diamondbacks, in the same time frame, only boasted three million total fans for three seasons (Baseball Reference). The data indicates that in times of poor performance concerning winning percentage, the Diamondbacks notice a decline in attendance. With a young history of existence, the Diamondbacks cannot rely on the power of being a historic franchise, like the Yankees, Red Sox, or Giants, to fill seats in these times of poor performance. The case study portion of this senior project in *Chapter III* will highlight the areas in which the Arizona Diamondbacks have been successful in driving attendance as well as address the weaknesses the team currently faces with regard to attendance determinants.

According to the Team Marketing Report, which publishes the Fan Cost Index (FCI), the Diamondbacks have provided fans with the cheapest average ticket price for home games for 11 of the past 25 seasons (Hartweg, 2023). Recently making headlines with a new season ticket package of $299 for all 81 home games in the 2024 season, the team emphasizes fan experience and cost efficiency over profits (Corrado, 2023, Hodell 2021). Averaging out below $4 a game, the Diamondbacks implemented this pricing package to fill more seats at the expense of ticket revenues for selling these tickets at regular prices. However, the value added by filling these seats minimizes the lost revenue of failing to sell the ticket for the seat in general. This package, similar to the popular summer pass of 2022 and 2023 that provided fans with unlimited access to all summer games from June 13 to August 31 for only $99, emphasizes the theme of being the ballpark for families (Klapper, 2023).

In an interview with Fox Sports, Derrick Hall, President and CEO of the Diamondbacks, stated, “We’re excited to provide yet another affordable option for individuals and families to enjoy baseball together… you’ll see why Chase Field provides the most affordable ballpark experience in all of Major League Baseball” (Corrado, 2023). The Diamondbacks have kept this message consistent for the past 15 seasons, focusing on providing the cheapest overall fan experience despite their lackluster performance at best (Hartweg, 2023). Focusing a study on the Diamondbacks provides a unique opportunity to understand the demand for tickets for a team not known for consistent star-power relative to their competitors or impressive season results, but rather a team focusing on providing an affordable, family-friendly experience to their fans.

Beginning as early as 1999, MLB franchises began sponsoring secondary market sellers for tickets in an attempt to boost ticketing revenues (Courty and Davey p. 121, 2020). Doing so led to the adoption of variable pricing and then subsequent dynamic price modeling adoption by much of the league (Courty and Davey, 2020). By 2011, half of the MLB adopted dynamic price modeling for their tickets (Courty and Davey p. 121, 2020). Courty and Davey (2020) observed the adoption of this new methodology of ticketing led to increased team ticketing revenues along with increasing team value (Courty and Davey pp. 128-132, p. 134, 2020). While this study addressed added value, the study failed to consider the impacts on attendance, thus leaving a gap in the literature for further study. This study will confirm or refute this study along with others covered in *Chapter II* of this senior project by providing a focused look into the Diamondbacks’ attendance determinants.

The goal of my research is to uncover the modern trends of attendance as well as the attendance determinants for the Diamondbacks by splitting the data into pre- and post-dynamic pricing periods as well as with and without macroeconomic considerations. In doing so, this

paper will utilize stylized determinants of both game-by-game and season-by-season MLB attendance factors as well as introduce new variables, such as the championship leverage index (cLI) and dynamic pricing variables, to allow for a deeper analysis of the attendance in both periods of analysis as well as the franchise's history. Defined fully in *Chapter III*, the cLI is a statistic that analyzes the importance of the game's outcome in terms of the effect on the probability of the home team winning *the World Series*. Understanding these trends will build on the prior studies on MLB attendance and provide insight into the history of MLB's quickest expansion team to win a World Series Title (Pavlovic, 2023).

Chapter II of the senior project aims to discuss various relevant literature based on the research. The literature covered will provide insights into prior studies conducted on other MLB baseball teams that provide a foundational framework for the analysis discussed in *Chapter III*. The literature review will uncover and further discuss the gaps in the literature this senior project aims to fill.

Chapter III explains the collection of the data for the study and discusses the variables utilized in the analysis along with providing a holistic approach to solve the research questions laid out at the end of this introduction. The analysis in *Chapter III* combines an analytical case-study approach to regression analysis to provide a full-scale, comprehensive study. Following the analysis, *Chapter IV* will provide the conclusions, policy recommendations, and areas for further research.

By the end of this research paper, this research will address and provide in-depth insights into the following research questions:

- How do known determinants of MLB attendance change for the Arizona Diamondbacks pre- and post-dynamic pricing introduction?
- How do known determinants affect the attendance of a team focused on providing a 'cheap and affordable, family-friendly' experience?
- How can teams with 'less strong' performance maximize their attendance rates?
- How do differing promotion types affect attendance?

These findings will provide insight into the world of Major League Baseball through the lens of an expansion team focused on an affordable, family-friendly experience with lackluster performance over their 25 years of existence and provide a framework for other studies into other teams within the MLB or beyond.

Chapter II: Literature Review

This section covers the current literature on varying topics relevant to the research within this comprehensive project. The literature will be subcategorized into the overarching themes and patterns and provide gaps within each sub-category along with overall gaps that provided the positioning of this paper.

Section I: Stylized Determinants

This subsection discusses the current state of the literature on stylized determinants of MLB attendance. With a rich history as a professional sports league, literature on MLB attendance is plentiful. This subsection literature will inform the choice of independent variables for the regression analysis in Chapter IV.

Coates and Humphreys (2007) researched attendance determinants across the three major professional sports leagues in the United States: the MLB, the National Basketball Association (NBA), and the National Football League (NFL). It is the first study to analyze the FCI from the Team Marketing Report in an attempt to assess the impacts of complementary goods, such as concession prices, parking, and souvenirs, on attendance (Coates and Humphreys pp. 161-162, 2007). Analyzing data from 1991 to 2001 across all three sports leagues, Coates and Humphreys (2007) found that demand for MLB tickets is slightly inelastic concerning price amongst all three leagues due to multiple products (i.e. broadcasts to local TV stations) being sold (Coates and Humphreys p. 162, 2007). Furthermore, Coates and Humphreys (2007) found that lagged attendance, winning percentage, and population of the city in which the team resided were statistically significant (Coates and Humphreys p. 168, 2007).

In a study primarily focused on the impact of promotions on attendance in the MLB by McDonald and Drayer (2000), the researchers reviewed literature on the myriad of studies previously done on attendance determinants and found one major finding concerning attendance and promotions: promotions have a significant, positive effect on attendance (McDonald and Drayer p. 449, 2000). McDonald and Drayer (2000) discussed the idea that, while many studies have been conducted previously on factors out of the control of franchises, franchises completely control the frequency of promotions throughout the season.

An important gap McDonald and Drayer (2000) filled was the differentiation between a price promotion, a non-price promotion, and a giveaway (McDonald and Drayer p. 10, 2000). Price promotions are classified as a night where the price is directly impacted by the promotion such as a two-for-one senior night or kids attending for free day. Non-price promotions are classified as value-added promotions such as a fireworks night or post-game concert. Giveaways are classified as games where bats, merchandise, bobbleheads, or other merchandise (McDonald and Drayer p. 10, 2000). Combining these variables with other known determinants, McDonald and Drayer (2000) found that a promotional night increased attendance in the observed games by 14 percent over non-promotional games. Their research also concluded that there is a slight "watering-down" effect of too many promotional games throughout a season; however, the attendance gains from having an additional promotional game outweigh this effect (McDonald and Drayer p. 18, 2000).

The framework of the research by McDonald and Drayer (2000) derives from the ground-breaking research by Rottenberg (1956) on the uncertainty of outcome hypothesis. Rottenberg (1956) is credited with the first in-depth study into the labor market of the MLB along with laying out the complexity of the MLB. Rottenberg (1956) provided an in-depth analysis of the

intricacies that make up the entirety of the MLB from developmental and minor leagues all the way up to the major leagues (Rottenberg pp. 243-246, 1956). In doing so, Rottenberg (1956) provided a fundamental piece of all MLB research papers by providing this in-depth foundational review.

The uncertainty of outcome hypothesis that Rottenberg (1956) developed stated that in order for attendance to be driven higher, a more-or-less balanced competitive nature between teams is required (Rottenberg p. 246, 1956). Without a competitive balance, the willingness to pay the fan is lowered also (Rottenberg p. 246, 1956).

While this hypothesis made Rottenberg (1956), whose paper became prominent in the MLB research world, study also analyzed a small number of attendance factors, focusing on the population of the surrounding area, size and location of the stadium, average stadium rank, other leisure-time activities in the surrounding area, and payrolls (Rottenberg p. 246, 1956). The findings were that population, stadium size and location, and average stadium rank had a positive impact on the attendance of an MLB game across all the teams studied (Rottenberg p. 246, 1956). On the other hand, areas with high concentrations of leisure-time activities and payrolls have a net negative impact on attendance (Rottenberg pp. 246-247, 1956). Despite the MLB's best efforts to moderate the talent, even in the 1950's the New York Yankees outspent their competitors to purchase the best players' contracts in efforts to achieve the best possible team. Rottenberg's (1956) research deemed an attempt to moderate this imbalance in the league was nearly impossible without a salary cap, he recognized that each MLB franchise would have some differing attendance factors that are involved on a team level.

In a revisit to Rottenberg's (1956) perspectives, Vrooman (1996) aimed to update the study. With 40 years in between publications, Vrooman (1996) studied the shift of talent from

small-market teams to large-market teams in an attempt to attain a higher salary. Vrooman (1996) found that in the 40-year time frame, small-market teams, compared to the large-market teams, did a poor job at attracting and signing top talents in the MLB free-agency pool (Vrooman p. 347, 1996). Challenging Rottenberg's (1956) idea that teams prefer to win and teams prefer winning at a closer margin over a larger margin as it remains more profitable, Vrooman (1996) found that, "small markets clubs will be more profitable if they have an even chance to win, whereas the large market clubs will be more profitable because they have provided a quality contest with an uncertain" (Vrooman p. 342, 1996). While Vrooman (1996) updated most of Rottenberg's (1956) study, the author failed to capture the effects on attendance fully within his study, focusing more heavily on the labor strikes and free agency market (Vrooman pp. 339-341, 1996).

These findings by Rottenberg (1956), with an updated study from Vrooman (1996), support the idea that a competitive balance generally increases attendance. Neale (1964) analyzed the uncertainty of the outcome hypothesis on game attendance, television viewership, and radio listenership. Neale (1964) found evidence that supported the ideas of Rottenberg (1956), finding that fans attended, watched, and listened more to games that were competitive and exciting over one-sided games (Neale, 1964). When it comes to live attendance, Neale (1964) discovered that when standings were close between the two opposing teams or there was a possibility of a significant change within the standings, gate receipts were higher (Neale p. 3, 1964). Neale (1964) also discovered a positive trend in attendance: the more frequent standings changed throughout a season.

A study conducted by Parris et. al (2012); cited Rottenberg's (1956) ideas on performance and connected to Ahn and Lee's (2014) ideas on the impact of team image to

analyze a steep downturn in attendance by the Los Angeles Dodgers in the years from 2009 to 2011. The Los Angeles Dodgers realized a loss of nearly 10,000 fans on average per game due to poor performance and negative press (Parris et. al; 2012). Despite the historical reputation of the Los Angeles Dodgers attracting fans, Parris et. al; (2012) observed this sharp decrease in attendance directly due to declining performance but also negative press around the team owner's messy divorce (Rottenberg, 1956, Ahn and Lee, 2014). This study was the first to identify negative press directly impacting attendance for an extended period of time as a direct result of a team scandal (Parris et. al; 2012).

Coates et. al; (2014) utilized the ideas of Rottenberg (1956) and Neale (1964) to identify a major gap in previous research into MLB attendance: both researchers failed to identify the effect of consumer behavior on this model (Coates et. al; p. 960, 2014). Coates et. al; (2014) applied marginal utility to the uncertainty of outcome hypothesis and discovered that in order for fans to be driven to attend, the marginal utility of an unexpected win must be equal to or greater than the marginal utility of an unexpected loss (Coates et. al; p. 960, 2014).

Developing a model that tests the loss-aversion theory against the results of the model that runs the uncertainty of outcome hypothesis, Coates et. al; (2014) discovered that while results showed marginal increases in attendance following the uncertainty of outcome hypothesis, the loss-aversion model proved to be more significant (Coates et. al; p. 972, 2014). In doing so, the results suggested that rather than prefer a competitive balance, fans generally attended more when there was less uncertainty of outcome (Coates et. al; p. 972, 2014). These findings generally apply to a single game attendance by the marginal fan and indicate that fans typically prefer less surprise in the outcome of the game.

Extending the research done by Rottenberg (1956) another study by Ahn and Lee (2014) analyzed popular attendance determinants on a game-to-game basis from 1904 to 2012. This study provided groundbreaking insights into the long history of change throughout the MLB and why fans attended games. This study enlarged the known data set of MLB attendance through filling gaps utilizing panel data and estimates, correcting a prior study by Lee (2013) that attempted to fill gaps utilizing other estimation methods for the current data set from 1975 to 2009 (Ahn and Lee p. 452, 2014). Filling these gaps, Ahn and Lee (2014) examined the stadium effect, fan loyalty, the uncertainty of outcome hypothesis, and the effects of rule changes within the league on attendance amongst other known determinants (Rottenberg, 1956, Ahn and Lee p. 453, 2014).

Since no prior study covered attendance for teams over this length of study, Ahn and Lee (2014) decided to split the dataset into two categories. The first category from 1904 to 1957 covered the MLB, while the Dodgers and Giants were both located in New York. The second category followed the expansion of the MLB beyond the Mississippi River while the Dodgers and Giants moved to Los Angeles and San Francisco respectively. This allowed for Ahn and Lee (2014) to analyze the entire model beyond simply analyzing the data as a single panel factor model (Ahn and Lee pp. 453-454, 2014).

Contrary to the findings of Coates et. al; (2014), Ahn and Lee (2014) observed that the uncertainty of outcome hypothesis provided a significant positive impact on attendance throughout the duration of the analyzed data, showing a seven percent increase in NL games and a four percent increase in American League (AL) games. These findings imply that policy measures taken by the MLB to ensure higher levels of competitive balance amongst teams are both statistically and economically significant (Ahn and Lee pp. 473-474, 2014). Further

supporting Rottenberg's (1956) study, Ahn and Lee (2014) also observed fan loyalty remaining significant throughout the duration of the observed games.

On an overall basis, Ahn and Lee (2014) observed an increase in significant variables within the regression model increase from four to seven from the first period (1904 to 1957) to the second period (1958 to 2012) indicating that fans' preferences changed historically (Ahn and Lee pp. 458-463, p. 477, 2014, Lee pp. 583-584, 2016). One major finding indicated that the offensive presence of teams became significant in the shift from the first period to the second, indicating that teams with more offensive power attracted larger game-day crowds (Ahn and Lee p. 474, 2014). Ahn and Lee (2014) imply that teams evaluating a better batter than the league average increase attendance by providing more offensive power along with more wins, providing a double effect on attendance.

In another study conducted on common attendance factors, Lee (2016) revisited the ideas of Ahn and Lee (2014) utilizing a sequential testing method on the same dataset introduced in the study from 1904 to 2012. In doing so, the reformatting of the regression by Lee (2016) allowed for further investigation into attendance determinants. This study did not focus on policy by the league, like that of the study by Ahn and Lee (2014), and focused solely on the determinants of attendance. Lee (2016) also introduced macroeconomic features into the regression. In doing so, Lee (2016) found significant variables that directly impacted attendance through per capita GDP, the uncertainty of outcome hypothesis, multiple offensive statistics (average runs per game, average slugging percentage, and average batting average), and strikeouts (Lee pp. 594-596, 2016). These findings contradicted a finding of Ahn and Lee (2014) in that these effects mitigated the overall positive effects of fan loyalty on attendance; however, fan loyalty remained significant (Lee p. 597, 2016).

With many studies on attendance determinants from a macro-level to a micro-level, each study provides unique insights into the world of the MLB and its organizations. While these studies are important for understanding attendance determinants and have found many statistically significant models, none of the studies compared the period pre- and post-dynamic pricing introduction. The current studies appear to be segmented in either attendance focus or dynamic pricing focus.

Section II: Dynamic Pricing

The MLB partnered with Qcue in 2011 to bring data analytics and dynamic pricing solutions into the MLB for all franchises to utilize. According to their website, Qcue provides their customers with a dynamic pricing model that maximizes the potential of each ticket to generate the greatest ticket revenues with minimum ticket waste. Qcue's software utilizes live analytics on websites to update ticket prices with demand fluctuations based on ticket location, historical sales, and other variables consistent with driving demand changes.

A study conducted by Elmaghraby and Keskinocak (2003) analyzed the impacts of dynamic pricing beyond the airline, hotel, and utility sectors. The research focused on covering the rapid expansion of dynamic pricing and how it translated beyond the commonly studied sectors. The study itself uncovered three main phenomena behind the expansion of the application of dynamic price modeling: the increased availability of demand data, the ease of changing prices due to new technologies, and the availability of decision-support tools for analyzing demand data and for dynamic pricing (Elmaghraby and Keskinocak p. 1287, 2003).

Although the paper heavily focuses on the overall prior literature for dynamic pricing, Elmaghraby and Keskinocak (2003) find that, given a regular demand, three principles are true:

- At a given point in time, the optimal price decreases as the inventory increases
- For a given level of inventory, the optimal price rises if there is more time to sell
- More on-hand inventory and/or a longer remaining selling horizon leads to higher expected revenues (Elmaghraby and Keskinocak p. 1291, 2003).

Work by Sweeting (2012) tracked the secondary ticket market for MLB tickets in the regular season and found that sellers slashed prices by nearly 40 percent leading up to the first pitch. Coinciding with the framework laid out by Elmaghraby and Keskinocak (2003), these ideas indicate that ticket sellers look to maximize the number of tickets sold, even if it means taking a loss on the ticket itself. The study by Sweeting (2012) provided a framework for studying ticket prices over time in a dynamically changing market. Sweeting (2012) also discovered a trend that secondary sellers (sellers that were not the organization itself) utilizing dynamic pricing increased revenues by upwards of 16 percent.

Utilizing similar ideas to Sweeting (2012), a study by Drayer et. al; (2012) on the secondary seller market of National Football League (NFL) games found that ticket prices that were able to accurately reflect the willingness to pay of customers. Due to dynamic pricing, secondary sellers were able to capture upwards of $260,000 in customer surplus otherwise missed by sellers not utilizing dynamic pricing (Drayer et. al; 2012). These studies highlight the importance of the adoption of dynamic pricing in the sports ticketing markets because of the increased potential to capture missed revenues through a static pricing system.

The work by Paul and Weinbach (2015) also utilizes the first two principles of Elmaghraby and Keskinocak (2003) through their research into the dynamic factors that cause fluctuations in the highest and lowest ticket prices for MLB games. Utilizing data from four

different teams in the 2011 season, Paul and Weinbach (2015) compare and contrast the findings from each team to find the overarching similarities from team to team.

Paul and Weinbach (2015) utilized known determinants of MLB attendance (Ahn and Lee, 2014, Rottenberg, 1956, Coates and Humphreys, 2007, McDonald and Drayer, 2000) in their study to evaluate price fluctuations as a result of dynamic pricing. Utilizing weather, starting pitchers, promotions, win percentage, and days of the week and month in the regression model, Paul and Weinbach (2015) observed many significant independent variables for explaining variations in ticket price. While many of the findings correspond with the results of other MLB attendance determinants studies, a major finding of Paul and Weinbach (2015) was the magnitude of the effect of weather on price variation (Paul and Weinbach p. 164, 2015). Connecting these findings to attendance data would indicate if the changes in price through dynamic pricing were effective in attracting a larger crowd than in years without dynamic pricing.

In another study utilizing known MLB attendance determinants (Ahn and Lee, 2014, Rottenberg, 1956, Coates and Humphreys, 2007, McDonald and Drayer, 2000) on the variations in ticket pricing as a result of dynamic pricing by Shapiro and Drayer (2014) aimed to compare the results between primary (MLB Franchise) and secondary (Ticketmaster, StubHub, etc.) sellers (Shapiro and Drayer pp. 145-146, 2014). Unlike the four teams utilized by Paul and Weinbach (2015), Shapiro and Drayer (2014) only focused on the San Francisco Giants which were the first MLB franchise credited with the implementation of dynamic pricing to their tickets (Courty and Davey p. 121, 2020).

The findings of Shapiro and Drayer (2014) indicate a strong correlation between both the primary and secondary models with regard to statistically significant variables. The strongest

variations of ticket pricing observed were explained through team and individual performance along with time conditions (day and month) (Shapiro and Drayer pp. 155-158, 2014). This study draws a strong connection with the findings of Ahn and Lee (2014) and Lee (2016) who observed similar trends with regard to MLB attendance.

Partnering with Ticketmaster, Xu et. al; (2019) reverse-engineered the data utilized by independent firms to dynamically alter prices in an attempt to further understand the drivers of price variation. Initial findings indicated a loss of 0.78 percent of total revenues when the anonymous franchise utilized a flat pricing strategy (Xu et. al; p. 1, p. 4, 2019). Xu et. al; (2019) utilized the determinants studied by Shapiro and Drayer (2014) and Paul and Weinbach (2015) to engineer a more robust dynamic price modeling system.

While the findings of Xu et. al; (2019) shed light on the impacts of dynamic price modeling on team revenues, they indicate that, for an observed period in the middle of the season, the observed team's performance was well below league average (Xu et. al; pp. 4-5, 2019). Xu et. al; (2019) denoted a direct impact in the results as ticket revenues as attendance decreased. Furthermore, Xu et. al; (2019) only observed 49 games with the dynamic pricing strategy and 48 games with the flat pricing strategy. While the results indicated a strong correlation between ticket pricing and revenues, the research failed to extend the study beyond a small group of games (Xu et. al; pp. 4-6, 2019).

Parallel studies into the hotel industry on the application of dynamic price modeling in an attempt to maximize revenue by Kimes (1989) uncovered that hotel managers preferred generating maximum revenues per room over selling out the available rooms in the hotel. Kimes (1989) argued that due to dynamic pricing, the industry was able to capitalize on the previously missed willingness-to-pay by utilizing known determinants of hotel room reservation to create

fluctuations in price. This study uncovered the relationship between revenue generation and dynamic price modeling.

Drayer et. al; (2012) argued that, unlike the findings of Kimes (1989), MLB organizations attempt to maximize attendance without focusing on the revenue impacts (Drayer et. al; p. 188, 2012). Utilizing data from the San Francisco Giants, Drayer et. al; (2012) recognized dynamic pricing as having positive effects on team revenues within both a theoretical and applied setting. Drayer et. al; (2012) observed captured revenue losses otherwise not recognized before the implementation of dynamic price modeling.

Section III: Gaps in the Literature

The research of Drayer et. al; (2012) mentions the connection between attendance and revenues with the introduction of dynamic pricing; however, fully researching the connection of determinants and dynamic pricing has not been previously done. The studies on dynamic pricing (Xu et. al; 2019, Shapiro and Dayer, 2014, Paul and Weinbach, 2015) that include attendance determinants (Ahn and Lee, 2014, Rottenberg, 1956, Coates and Humphreys, 2007, McDonald and Drayer, 2000) only evaluate the impacts of the determinants on ticket prices, not attendance itself.

Furthermore, the studies conducted on MLB attendance determinants (Ahn and Lee, 2014, Rottenberg, 1956, Coates and Humphreys, 2007, McDonald and Drayer, 2000) fail to address the long-term impacts of dynamic pricing, with many studies concluding before long-term impacts can be observed. While Ahn and Lee (2014) and Lee (2016) observe changes in attendance determinants in the two time periods observed, neither cover the shift in determinants following the application of dynamic price modeling.

Finally, while research by McDonald and Rascher (2000) found a significant, positive effect of promotions on game-to-game attendance, no current research highlights the impact of different types of promotions. Research covers the aggregate of promotions, detailing just the raw number of promotional events. This research aims to fill the gap by distinguishing between the various types of promotions and their effects on attendance. Doing so will provide a framework for teams to understand the attendance impacts of kid versus non-kid-friendly promotions on game day.

Chapter III: Empirical Analysis

This chapter of the research covers an in-depth empirical analysis to address the research questions introduced in the first chapter. The research questions are as follows:

- How do known determinants of MLB attendance change for the Arizona Diamondbacks pre- and post-dynamic pricing introduction?
- How do known determinants affect the attendance of a team focused on providing a 'cheap and affordable, family-friendly' experience?
- How can teams with 'less strong' performance maximize their attendance rates?
- How do differing promotion types affect attendance?

Section I: Variable Definition

The purpose of this section is to define the variables utilized in the data analysis and regression along with providing source information. This section also features *Table 1* which lists each variable, the expected sign for regression, and sources of information.

Dependent Variable

Attendance: This represents the total number of fans entering the stadium on the day of the game. The attendance number is as reported by MLB and retrieved from Retro Sheet. The goal of utilizing attendance as the dependent variable in the regression was to determine the factors that have the highest impacts on game-to-game attendance for the Arizona Diamondbacks. Attendance is a standardized metric that is measurable by every MLB franchise, thus providing a method by which the study could be easily replicated.

Independent Variables

AZ GDP Growth Rate: This variable indicates the growth rate percentage year over year of the state of Arizona, as reported by the Bureau of Labor Statistics. Highlighting macroeconomic factors in the regression analysis allows for factors beyond the control of the Diamondbacks management to be accounted for within the model, thus explaining more of the variance in the attendance numbers. Selecting the Arizona State GDP Growth Rate allows for the health of the state's economy to be considered in this analysis. This is expected to have a positive impact on attendance due to economic growth being correlated with extra spending money.

AZ Unemployment Rate: This variable indicates the unemployment rate percentage of the state of Arizona, as reported by the Bureau of Labor Statistics. Utilizing the Arizona state unemployment rate allows for the health of the labor market within the state to be analyzed. A higher unemployment rate would indicate a less healthy economy, thus potentially hindering fans from spending income on leisure activities such as attending a baseball game. This is expected to have a negative impact on attendance due to a high unemployment rate relating to a weak economy.

Batting Average (BA): This is a lagged variable that indicates the batting average of the Diamondbacks going into the game. The first game of the season utilizes an arbitrary value of .225 to account for no prior statistics. This value indicates a hit would occur 22.5 out of every 100 at-bats. The variable is calculated by taking the total number of hits divided by the total number of at-bats for the team. Batting average, as highlighted in *Chapter II*, is a stylized determinant found to be statistically significant across many studies. The higher a team's batting average, the stronger the offense is. It is known from prior studies that a higher-power offense

will fill more seats than a lower-power offense, thus leading to the expectation of a positive impact on attendance.

Championship Leverage Index (cLI): This variable indicates a statistic created by Baseball Reference that analyzes the importance of the game's outcome in terms of the effect on the probability for the home team to *win the World Series*. A value of [1.0] indicates the *average* importance of a victory for a team's probability of winning the World Series. A value greater than [1.0] indicates an *above-average importance.* A value less than [1.0] indicates a *below-average importance*. Introducing this variable to the regression analysis allows for a new statistic to be tested in the equation for attendance and is expected to have a positive impact on attendance given the nature of the variable.

Dynamic Pricing: This is a dummy variable that constitutes whether the game features tickets that are affected by Qcue's dynamic pricing software. Adopted by the Diamondbacks in 2011, all games post-2011 will contain a [1] in this column, indicating the change from no dynamic pricing [0] before 2011. Introducing this variable allows for the full franchise history regressions and allows for the results to show if dynamic pricing is relevant for Diamondbacks fans, the impact on attendance, and overall effectiveness. This is expected to have a positive impact on attendance as it is designed to maximize revenues and attendance.

Friday – Sunday: This is a dummy variable that classifies the day of the week the game occurs. A [1] indicates the game occurred on a weekend. Accounting for this variable allows for the impact of a weekend game to be measured against games played during the week. Previous studies on attendance determinants for the MLB indicated weekend games are statistically significant in drawing larger crowds than weekday games. These games are expected to be attended more than the non-weekend games.

Games Behind: This variable indicates the total number of games behind first place the Diamondbacks are before the game starts. A *negative* number indicates how many games ahead of the *second*-place team and a *positive* number indicates how many games behind the *first*-place team the Diamondbacks are. A team with a larger number of games behind is considered to be less of a contender than a team with a fewer number of games behind. If a team is in first place, they are considered to be the best in the division, thus indicating a stronger team. Accounting for this variable in the regression allows for the impact of team performance compared to other teams in the division on attendance to be captured. Given the nature of this variable, it is expected that the regression will show a negative coefficient.

Home Opening Day: This is a dummy variable that indicates the *first* home game [1] of each regular season. Opening day in Major League Baseball is considered a celebration across the league. This variable accounts for the impact of this celebrated day on the variation in attendance levels to be captured. This game is expected to have a high, positive impact on attendance due to its holiday-like nature for baseball fans.

Phoenix (Metro) GDP Per Capita: This variable indicates the GDP per capita of the surrounding area of the Diamondbacks' home field, as reported by the Bureau of Labor Statistics. Introducing this variable into the regression allows for the health of the economy to be considered in the variation in attendance. The Phoenix Metro region is the region surrounding and including the home stadium, Chase Field, of the Diamondbacks. Utilizing the health of the local area in the regression should show that the healthier the economy, the greater the total attendance at Diamondbacks' games. A higher GDP per capita is expected to have a positive impact on attendance.

Promotion: This is a dummy variable that indicates a [1] when there is a promotion present. Promotional data was collected from the Diamondbacks season archive found on Now Hitting (https://www.nowhitting.com/index.php?view=article&id=160&catid=39). Promotions, as indicated in *Chapter II*, are statistically significant when it comes to boosting attendance. Introducing this variable to the regression allows for the impact promotions have on attendance to be captured and is expected to have a positive impact on attendance.

Promotion U15: This dummy variable indicates a promotion that is available for kids only. A [1] indicates that the promotion targets children, and a [0] indicates a regular promotion. Promotional data was collected from the Arizona Diamondbacks season archive found on Now Hitting. The Diamondbacks classify their promotions as 'for kids' when the promotion targets an audience under the age of 15. This is a new variable introduced into an attendance determinant study for the MLB as no previous study covered different types of promotions. This variable will directly fill the gap in the literature on promotion type discussed in *Chapter II*. This variable is expected to have a positive impact on attendance.

Stolen Bases (SB): This is a lagged variable that indicates the total number of *stolen bases* by the Diamondbacks in the game prior. Stolen bases in the MLB provide a competitive advantage for teams and can be momentum-changing moments in games. The MLB has been in discussions of trying to increase the number of stolen bases per game as it adds excitement to the game. A higher number of stolen bases in this analysis would provide more base running action than games with lower numbers of stolen bases. Therefore, this regression will capture the impact of stolen bases on the variation in attendance levels where a positive impact on attendance is expected.

Summer: This is a dummy variable that indicates if the game occurs in a summer month. Summer for the purpose of this study is classified as June, July, and August. A [1] indicates the game is played in a summer month and a [0] indicates the game is played outside a summer month. Including this variable in regression allows for games outside of June, July, and August to be compared against the earlier and later games in the year. Understanding these trends will allow executives to make more informed decisions on trying to boost attendance. These games are expected to have higher attendance than games in the non-summer months.

Time of Day: This is a dummy variable that constitutes the time of day the game is taking place. In the regression model, a [0] indicates a day game and a [1] indicates a night game. Time of day is defined as found in the Retro Sheet data set available for public download. Accounting for the difference in the time of day allows for a more comprehensive understanding of the impact of the game start time on attendance. Typically, games played in the middle of the day feature a smaller attendance total during the week; however, this is not as significant of a factor on the weekend. The regression with this variable will allow for these patterns to be tested for their level of significance and overall impact on the variation in attendance. Night games, therefore, are expected to have a positive impact on attendance.

Win Percentage: This variable takes into account the win percentage based on the running record of the team before the home game. Winning percentage is a common variable used in regressions with regard to attendance. Previous studies covered in *Chapter II* highlight the continued finding of winning percentage being a significant driver in attendance, with studies showing that a team with a higher winning percentage tends to fill more seats than a team with a lower winning percentage. Introducing this variable allows for the impact on the variation of

attendance to be measured for the Diamondbacks, where a positive impact on attendance is expected. This statistic was calculated using the common formula:

Win Pct. = Total Wins / Total Games

Table 1: Independent Variables and Source Information

Variable	Expected Sign	Source
AZ GDP Growth Rate	+	Bureau of Labor Statistics
AZ Unemployment Rate	-	Bureau of Labor Statistics
Batting Average (BA)	+	Baseball Reference
Championship Leverage Index (cLI)	+	Baseball Reference
Dynamic Pricing	+	*
Friday – Sunday	+	Baseball Reference
Games Behind	-	Baseball Reference
Home Opening Day	+	Baseball Reference
Phoenix Metro GDP Per Capita	+	Bureau of Labor Statistics
Promotion	+	Now Hitting
Promotion U15	+	Now Hitting
Rank	-	Baseball Reference
SB (Stolen Bases)	+	Baseball Reference
Summer	+	Baseball Reference
Time of Day	+	Baseball Reference
Win Percentage	+	**

* Dynamic Pricing variable was user-created with games before 2011 having a [0]

** Win Percentage was calculated using the win/loss totals from Baseball Reference using the basic win percentage equation (win percentage = total wins / total games)

Section II: Regression Modeling

This section will cover the regression modeling, with *attendance* as the dependent variable. The models will aim to explain the variance of attendance on a game-to-game basis. The regression will be estimated with and without macroeconomic considerations along with two models that contain macroeconomic factors from 1998-2010 and 2011-2023. This will evaluate how the models change with and without external factors beyond the game of baseball and the games played before and after the introduction of dynamic pricing (before the 2011 season) to the Arizona Diamondbacks. This is done in an attempt to construct a sound model for explaining the variance in attendance along with answering the research questions posed at the beginning of the chapter.

In total, there were 2025 home games played by the Diamondbacks from 1998 to 2023. The 2020 MLB season was truncated (only 60 games) due to COVID-19 with many COVID-19-related restrictions on travel and social distancing. Therefore, this regression excludes the 2020 MLB season due to COVID-19 impacts on the game and prevention of fans from attending games in person.

Section II.A: Model without Macroeconomic Considerations

The purpose of the model without the macroeconomic considerations is to model attendance given only game-to-game factors. *Table 2* outlines the results from the regression analysis. With an adjusted R squared of .4504, the model is responsible for explaining around 45 percent of the variance in game-to-game attendance. Paired with a significant p-value, the model is statistically significant. Given the regression results, some results were expected and others were a surprise.

Table 2: Regression without Economic Considerations

	Coefficients	***Standard Error***	***t Stat***	***P-Value***
Intercept	16115	2227.8658	7.2334	6.6652E-13
cLI	-259	338.2107	-0.7666	0.4434
Dynamic Pricing	-7247***	325.1092	-22.2903	1.5220E-98
Division Rival	713**	313.7656	2.2728	0.0231
Promotion U15	-2210***	654.1929	-3.3785	0.0007
Promotion	3786***	375.1865	10.0919	2.1406E-23
Home Opening Day	17220***	1439.3163	11.9641	6.4571E-32
Time of Day	749**	368.4008	2.0336	0.0421
SB	155	174.9993	0.8873	0.3750
BA	38642***	8762.5689	4.4099	1.0889E-05
Win Pct	1073	1790.7550	0.5991	0.5492
Friday – Sunday	6258***	359.3030	17.4173	2.1643E-63
Summer	1423***	320.1849	4.4456	9.2411E-06
Games Behind	-77***	21.6905	-3.5693	0.0004
R-Squared	0.4539			
Adjusted R-Squared	0.4504			
Significance F	1.0351E-252			

Significance Codes:
*** The variable is significant with a one percent level
** The variable is significant with a five percent level
* The variable is significant with a ten percent level

The equation for total attendance on a game-to-game basis, based on the regression results, is as follows:

Regression Equation:

Attendance = 16115 – 259 (cLI) – 7247 (Dynamic Pricing) + 713 (Division Rival) – 2210 (Promotion U15) + 3786 (Promotion) + 17220 (Home Opening Day) + 749 (Time of Day) + 155 (SB) + 38642 (BA) + 1073 (Win Pct) + 6258 (Friday – Sunday) + 1423 (Summer) -77 (Games Behind)

The first statistically significant variable from the regression is the dynamic pricing dummy variable. Dynamic pricing is statistically significant with a 99 percent confidence interval. With a coefficient of –7247, games featuring dynamic pricing for the tickets led to a *decrease* of 7247 fans, all else being equal, when compared to games that did not feature dynamically priced tickets. *Table 1* shows that the expected sign for the dynamic pricing variable is positive, thus making this result surprising.

Introduced to the team in 2012, all games before this did not feature any dynamically changed ticket prices. The marketing behind Qcue's dynamic price modeling was attendance maximization paired with revenue maximization. This finding indicates that when strictly looking at attendance on a game-to-game basis, dynamic pricing failed to draw a larger crowd than in the era before dynamically priced tickets.

This finding, however, could be explained through the attendance trends of the Diamondbacks. These findings could be explained by the downturn of attendance from the early years of the franchise, as the Arizona Diamondbacks had attendance levels of over 3,000,000 total in only three of their first five years as a franchise. Since then, the Diamondbacks have

never passed 3,000,000 total fans for a season. Basic economic theory implies that front offices will focus on revenue maximization at the expense of attendance numbers. Since dynamic pricing allows for previously uncaptured revenues, lower attendance numbers have the potential to meet or surpass attendance revenues before dynamically priced tickets.

The second statistically significant variable found in the regression model is the division rival variable. The variable is statistically significant with a 95 percent confidence level. With a coefficient of 713, the regression indicates that games played against teams within the division add 713 fans, all else being equal, as compared to games played against opponents outside of the division.

Given the schedule for an MLB season, teams play against other divisional teams most frequently throughout the year. The division rival having a positive coefficient is important for the Diamondbacks because these opponents visit the ballpark most frequently. This finding, however, means that the front office must look to other ways, like promotions, to increase attendance if the opponent is not within the same division as the Diamondbacks.

The regression also found both the under-15 promotion and promotion variables statistically significant with a 99 percent confidence level. One of the research questions outlined for this senior project is the effects of different promotions on game-to-game attendance. With a coefficient of 3786 on the promotion variable, the regression finds that a game with a promotion will draw 3786 more fans, all else being equal, than a game without a promotion.

This positive effect on attendance, however, is dampened when the promotion is geared towards kids. The coefficient of –2210 for the under 15 promotion variable indicates that games with promotions geared towards children draw 2210 *fewer* fans, all else being equal, than generic

promotions. When combining these effects, there is an *overall net gain* of 1576 fans, all else being equal, when having a promotion geared towards kids.

The positive effect on attendance through promotions is an expected result of the regression. The literature on promotions outlined in *Chapter II* shows that previous studies have found a similar effect. When considering the promotions focused on children, the net effect remains positive; however, the magnitude of the effect comes as a surprise. This effect indicates that while promotions are successful at drawing larger crowds than games with no promotions, the games that feature children-focused promotions only draw 40 percent of the fans that games with promotions do.

This finding could be explained through the idea that promotions focused on children specifically target a subdivision of the fans. This means that only fans with children who will directly benefit from the promotion (children aged 15 and younger based upon the classification of the Diamondbacks on children-based promotions) are targeted with these types of promotions. It is also worth noting that the Diamondbacks feature children-focused promotions outside of the game-day promotions, such as community events and days at the park when the team is not playing. These promotions outside of attending a game could also dampen this effect.

The next statistically significant variable in the regression is the home Opening Day promotion. Statistically significant with a 99 percent confidence interval, the home Opening Day variable features a coefficient of 17220. This indicates that the home Opening Day game will draw 17220 more fans, all else being equal, than a game not on Opening Day.

While this finding is not a surprise, the magnitude indicates that this is a significant attractant for fans. This finding could be explained by the celebrations held by teams across the

MLB for the first home series of the season. Opening Day weekend in the MLB is considered a massive celebration of the season to come. These celebrations tend to lead to some of the largest crowds of the season as excitement for the team is at its highest before the season begins.

The time-of-day variable is also statistically significant in the regression. Statistically significant at a 95 percent confidence level, the time-of-day variable has a coefficient of 749. This variable indicates that all else being equal, a night game will attract 749 more fans than a day game. This finding comes as no surprise and could be explained through fewer restrictions, such as work or school, for fans to attend night games when compared to games that start after the common end times for these activities.

In terms of offensive statistics, the batting average variable is statistically significant with a 99 percent confidence level. With a coefficient of 38642, understanding the effect on attendance requires some math. A common average of .225, which means 22.5 hits out of every 100 at-bats for a team, would lead to adding 8694 fans, all else being equal. A higher team batting average of .250 would lead to adding 9661 fans, all else being equal. A lower team batting average of .200 would only add 7728 fans, all else being equal.

Understanding this result allows for team executives to make decisions regarding additions to the team. As shown by the regression, a higher batting average leads to an increase in fans. These findings are not a surprise result and could be explained by fans' desires to see offensive power in the game. The higher a team's batting average, the more likely a team is to have more hits. With an increased chance for hits, there is an increased chance for fans to see multi-base hits or even home runs. Therefore, the higher this statistic, the more fans would anticipate witnessing more offensive production, thus keeping the game objectively more exciting.

The next statistically significant variable is the Friday – Sunday variable. Statistically significant with a 99 percent confidence level, the Friday – Sunday has a coefficient of 6258. This indicates that all else being equal, a weekend game will draw 6258 more fans than a game during the week. This finding is not a surprise finding of the regression as weekends typically have fewer restrictions, as stated with the time-of-day variable when compared to games played during the week.

Another statistically significant variable found in the regression is the summer variable. Statistically significant with a 99 percent confidence level, the coefficient of 1423 indicates that all else being equal, summer games will draw 1423 more fans than games played in the non-summer months. Similar to the time-of-day and Friday – Sunday variable, there are commonly fewer restrictions for fans during the summer months, such as vacations and no school, that could allow for fans to attend compared to the non-summer months.

Games Behind is the final statistically significant variable in the model found in the model without macroeconomic considerations. Statistically significant with a 99 percent confidence level, the games behind variable features a coefficient of –77. This implies that, all else being equal, each game behind the division leader the Diamondbacks fall, there will be 77 fewer fans. Given the structure of the data, a negative number for games behind indicates how many games the Diamondbacks are ahead of the second-place team. This means that for every game ahead of the second-place team, all else being equal, 77 more fans will attend.

A surprising finding of the regression can be found in the lack of statistical significance of the winning percentage variable. Previous literature on the stylized attendance determinants for MLB indicates that winning percentage is a commonly significant variable in game-to-game attendance models.

While these findings provide around explanation for around 45 percent of the variation of attendance for the Diamondbacks on a game-to-game basis, the model only considers a perfect scenario with baseball-only factors. The model in *Section II.B* will provide a model that features macroeconomic considerations to try and understand a larger portion of explaining the variation in attendance on a game-to-game basis.

Section II.B: Model with Macroeconomic Considerations

The model with macroeconomic considerations introduces three main variables into the regression analysis; GDP per capita of the Phoenix Metro Area, GDP growth rate percentage for the state of Arizona, and the Arizona unemployment growth rate percentage. In doing so, the model changed variable significance. With an adjusted R squared of .5547 and a p-value with statistical significance at a 99 percent confidence level, the model explains around 55 percent of the variance in attendance levels per game. This is an increase of over 10 percent from the model *without* macroeconomic considerations. *Table 3* contains the results from the regression analysis.

Table 3: Regression with Macroeconomic Considerations

	Coefficients	***Standard Error***	***t Stat***	***P-Value***
Intercept	62053	3544.9022	17.5048	5.7970E-64
cLI	78	314.0687	0.2468	0.8051
Dynamic Pricing	-7984***	299.9470	-26.6167	6.011E-134
Division Rival	1218***	283.6438	4.2930	1.8467E-05
Promotion U15	-1371**	590.2588	-2.3228	0.0203
Promotion	3122***	340.3902	9.1708	1.129E-19
Home Opening Day	17104***	1295.4748	13.2032	3.1358E-38
Time of Day	458	331.9996	1.3790	0.1681
SB	-114	158.0298	-0.7187	0.4724
BA	24278***	8067.9556	3.0091	0.0027
Win Pct	1372	1634.5989	0.8392	0.4015
Friday – Sunday	6203***	323.5494	19.1706	2.3378E-75
Summer	1405***	288.7378	4.8676	1.2179E-06
Games Behind	-103***	19.7504	-5.2166	2.0102E-07
Phoenix (Metro) GDP Per Capita	-1***	0.0572	-16.2061	1.3168E-55
AZ GDP Growth Rate	429***	58.2494	7.3580	2.7064E-13
AZ Unemployment Rate	-326***	109.0507	-2.9859	0.0029
R-Squared	0.5583			
Adjusted R-Squared	0.5547			
Significance F	0.0000			

Significance Codes:
*** The variable is significant with a one percent level
** The variable is significant with a five percent level
* The variable is significant with a ten percent level

The equation for total attendance on a game-to-game basis, based on the regression results, is as follows:

Regression Equation:

Attendance = 62053 + 78 (cLI) – 7984 (Dynamic Pricing) + 1218 (Division Rival) – 1371 (Promotion U15) + 3122 (Promotion) + 17104 (Home Opening Day) + 458 (Time of Day) – 114 (SB) + 24278 (BA) + 1372 (Win Pct.) + 6203 (Friday – Sunday) + 1405 (Summer) – 103 (Games Behind) – 1 (Phoenix Metro GDP per Capita) + 429 (AZ GDP Growth Rate) – 326 (AZ Unemployment Rate)

Similar to the model without economic considerations, the dynamic pricing variable is statistically significant with a 99 percent confidence level. In this regression, however, the coefficient is over 700 fans more negative. This indicates that all else equal, games with dynamically priced tickets led to around 8000 fewer fans than games without dynamically priced tickets. As previously mentioned in the regression model without macroeconomic considerations, this is an unexpected result of the regression.

The first model highlights the struggles of the Diamondbacks to fill seats following the first five seasons of the franchise's existence; however, with an increased magnitude of the efficiency in this model, further analysis is required to understand the true effects of this downturn on fans. *Section III* of this chapter will provide further insights into this trend through a case study approach.

Another similar result to the model in *Section II.B*, the division rival variable is statistically significant with a 99 percent confidence level. With a coefficient of 1218, this finding indicates that a game played against a division rival, all else equal, leads to an increase of

1218 fans when compared to games played outside of the division. As mentioned in the model without economic considerations, this could be explained by the willingness of fans to see the Diamondbacks play the teams that visit the ballpark the most and games with importance concerning the team's standings within the division compared to "less-meaningful" games.

Another similar result of the first regression, this model also finds both the under-15 promotion and promotion variables with a statistical significance of 95 and 99 percent confidence levels respectively. The magnitude of the promotion variable coefficient in this model, however, is 3122, indicating a decrease of over 600 fans from the first model. This could be explained by more of the variation being explained by the economic factors introduced in the model. Despite these changes, however, the finding is still an expected result of the regression since the effect is still positive on attendance.

The magnitude of the under-15 promotion variable also decreased from –2210 in the first model to –1371 in this regression model. When combining the effects, a game that features a children-focused promotion will only lead to an increase of 1751 fans, all else equal. This effect, when compared to the first model, shows a *net increase* in children-focused promotion games by almost 200 fans.

The home Opening Day variable also remained statistically significant with a 99 percent confidence level. With a coefficient of 17104, the magnitude is slightly lower than the 17220 from the first model; however, this finding shows the importance of the home Opening Day game to the franchise regarding attendance. This finding is an expected result of the regression as highlighted in the discussion of the first regression model findings.

The batting average variable is statistically significant in the model with economic considerations with a 99 percent confidence level. With a coefficient of 24278, all else equal, positive changes in batting will lead to positive changes in attendance levels. Utilizing similar measures to evaluate the effect of this variable, a team batting average of .225 would lead to 5463 fans, all else equal. A higher team batting average of .250 would lead to 6070 fans, all else equal. A lower team batting average of .200 would lead to 4856 fans, all else equal.

These findings show a dampened effect when compared to the model with baseball-only factors. When introducing the economic factors into the equation, the effect of the batting average was significantly dampened. The effect of adding .01 to the team batting average correlates with a 243-fan increase, compared to a 386 fan increase of the baseball-only model.

Another similar finding to the model without economic considerations, the Friday – Sunday variable is statistically significant with a 99 percent confidence level. With a coefficient of 6203, the regression result indicates, all else equal, weekend games will draw 6203 more fans than a game played during the week. When comparing the magnitude of the coefficient to the first model, it is observed that there is only a difference in 55 fans from the first model to this model. This further backs the importance of weekend games concerning game-by-game attendance. This finding is not a surprising result of the regression and is explained in the first regression model discussion.

The summer variable is statistically significant with a 99 percent confidence level and features a coefficient of 1405. All else equal, the regression shows that a summer game will draw 1405 more fans than a game played in a non-summer month. Similar to the Friday – Sunday variable, the magnitude of the coefficient showed little change with the introduction of the

economic factors, showing the importance of the variable in terms of attendance on a game-by-game basis.

The last variable statistically significant in both models is the games behind variable. With a statistical significance confidence level of 99 percent and a coefficient of –103, the impact of performance within the division is amplified when compared to the first model. The first model featured a coefficient of –77, thus showing a 26-fan change in the magnitude of the impact on attendance. Although seeming trivial, a Diamondbacks team 15 games behind first place, all else equal, will lose 1545 fans, compared to the first model showing a loss of only 1155 fans. A similar measure of a Diamondbacks team with a 10-game lead in the division, all else equal, will draw 1030 fans, compared to only 770 in the first model. This finding is not a surprise result, as highlighted in the discussion of the first model.

A new variable for economic consideration purposes, the Phoenix (Metro) GDP per capita variable is statistically significant with a 99 percent confidence level. With a coefficient of –1, all else equal, a GDP per capita shows a one-to-one ratio of decreasing fans within the equation for attendance. This finding comes as a surprising result, as the expected sign on the coefficient was positive, not negative.

When analyzing the trends in the GDP per capita for the Phoenix Metro region, the peak in GDP per capita occurred in 2009, the year of the financial crisis. At the same time, the Diamondbacks' season attendance averaged around 26281 fans per game. The largest fans per game average occurred in the debut season of the Diamondbacks franchise with 44571 fans per game. The same year featured a GDP per capita of 10000 less than the 2009 year. Therefore, the data collected on the Diamondbacks, while important for the model consideration of economic factors, could lead to the surprise result shown by the regression model.

The second economic factor statistically significant with a 99 percent confidence level is the Arizona GDP growth rate percentage variable. With a coefficient of 429, all else equal, an increase of one percent in the GDP growth rate would lead to an increase of 429 fans. An expected result of the regression is growth in the GDP for the home state of the Diamondbacks should have positive effects on the economy.

The final statistically significant variable in the economic consideration regression model is the Arizona unemployment rate variable. Statistically significant with a 99 percent confidence level, the unemployment rate variable features a coefficient of –326. This indicates that all else equal, a one percent increase in the unemployment rate would lead to a decrease of 326 fans. This is an expected result of the regression.

While many of the variables featured in this regression model, the time-of-day variable was no longer statistically significant. This is an unexpected result of the regression as the data shows that games played in the evening tend to feature a larger crowd than during the day. The lack of significance in this variable with the introduction of economic considerations requires further research to understand the lack of significance.

Furthermore, this regression also features a lack of statistical significance with regard to the winning percentage variable. As mentioned in the discussion of the first model, winning percentage is a common statistically significant variable in game-to-game attendance models. The lack of significance on this variable is puzzling and not expected with regard to the regression model.

Section II.C: Regression Pre-Dynamic Pricing with Macroeconomic Considerations

This section is responsible for discussing the regression model for games before the 2011 season. In 2011, the Arizona Diamondbacks adopted dynamically priced tickets for their home games in an attempt to boost attendance; however, before 2011, the tickets were priced by the Diamondbacks alone. This regression will provide insights into this era and compare the results to *Section II.D* which covers the era post-dynamic pricing introduction.

Table 4 contains the results from the regression analysis on the pre-dynamic pricing period (from the 1998 to 2010 season) for the Diamondbacks. The regression featured an adjusted R-squared of .5222, thus explaining around 52 percent of the variance in attendance levels. The following equation defines and describes the findings of the regression to calculate total attendance on a game-to-game basis:

Regression Equation:

Attendance = 82231 + 390 (cLI) + 1233 (Division Rival) – 776 (Promotion U15) + 2048 (Promotion) + 14266 (Home Opening Day) – 420 (Time of Day) – 100 (SB) – 41527 (BA) + 3835 (Win Pct) + 5247 (Friday – Sunday) + 1388 (Summer) + 3 (Games Behind) – 1 (Phoenix Metro GDP per Capita) + 466 (AZ GDP Growth Rate) – 902 (AZ Unemployment Rate)

Table 4: Regression Pre-Dynamic Pricing Introduction

	Coefficients	*Standard Error*	*t Stat*	*P-Value*
Intercept	82231	4485.2539	18.3337	3.0003E-65
cLI	390	383.6600	1.0154	0.3102
Division Rival	1233***	378.9634	3.2548	0.0012
Promotion U15	-776	634.7533	-1.2227	0.2217
Promotion	2048***	431.3335	4.7484	2.3387E-06
Home Opening Day	14266***	1676.7292	8.5082	6.1111E-17
Time of Day	-420	453.5593	-0.9269	0.3542
SB	-100	222.7145	-0.4504	0.6525
BA	-41527***	13941.7218	-2.9786	0.0030
Win Pct	3835*	2309.8378	1.6603	0.0972
Friday – Sunday	5247***	414.6534	12.6530	3.157E-34
Summer	1388***	382.0374	3.6330	0.0003
Games Behind	3	25.2054	0.1302	0.8964
Phoenix (Metro) GDP Per Capita	-1***	0.0562	-16.7386	7.4718E-56
AZ GDP Growth Rate	466***	67.7568	6.8709	1.0989E-11
AZ Unemployment Rate	-902***	156.6701	-5.7545	1.1435E-08
R-Squared	0.5291			
Adjusted R-Squared	0.5222			
Significance F	1.0854E-157			

Significance Codes:
*** The variable is significant with a one percent level
** The variable is significant with a five percent level
* The variable is significant with a ten percent level

Aligning with the macroeconomic consideration model from *Section II.B*, the division rival variable is statistically significant with a 99 percent confidence level. With a coefficient of 1233, all else equal, a game played within the division will draw 1233 more fans than a game played against an opponent outside of the division.

The second statistically significant variable in the pre-dynamic pricing model is promotion. With a statistical significance in the 99 percent confidence level, the promotion variable features a coefficient of 2048. All else equal, the regression shows that a game with a promotion will add an additional 2048 fans when compared to a game without a promotion. The magnitude of the effect, however, is over 1000 fans less than the regression model from *Section II.B*.

While the positive effect on attendance for a game featuring a promotion is an expected result of the regression, the change in magnitude could be explained by the increased frequency of promotions before the 2011 season. This data set of games before 2011 features the inaugural season of the Diamondbacks franchise. The inaugural season featured promotions in 45 of the 81 home games. The second season featured promotions in 54 of the 81 home games. In the third season of the franchise's history 45 of the 81 home games. With this massive influx of promotions, the overall effect of the promotion variable in game-to-game attendance could have been dampened.

It is also worth noting that this regression is the only reg without under 15 promotion variable not statistically significant. This could be explained similarly to the dampened effect of the promotion variable as the focus for promotions was primarily on general promotions in an attempt to grow the fan base as a whole. As mentioned, when discussing the results of the first

regression model, a promotion geared towards a specific subset is less likely to have an impact when compared to the effects of a generic promotion.

The next statistically significant variable in the regression is the home Opening Day variable. Significant with a 99 percent confidence level, the coefficient of 14266 indicates that all else equal, the game played on Opening Day will draw 14266 more fans than the rest of the games played throughout the season. Of the regression models, this regression featured the lowest magnitude of the coefficient for the home Opening Day variable. This reduction in the magnitude of around 3000 fans from the other models could be explained when analyzing the data on these games.

Of the seasons with the largest attendance, three of the first five seasons of the Diamondbacks' franchise featured the largest average attendance per game numbers. While this does not directly correlate with the home Opening Day variable, the increase in average attendance on a game-to-game basis could dampen the magnitude of the effect of the first game. The lower the variance in attendance from game to game, the less drastic the Opening Day game is when comparing the data.

The next statistically significant variable in the pre-dynamic pricing regression model is the batting average variable. With a statistical significance of 99 percent significance level, the coefficient of –41527 indicates that all else equal, a 0.01 percent increase in the team batting average leads to a 415-fan reduction in attendance. This finding comes as a surprise result, given the other regression models and prior literature find this variable to have a positive impact on attendance. Analyzing the data, there are no outliers or extraneous values, thus not explaining this surprising result.

Statistically significant with a 90 percent confidence level, the winning percentage features a coefficient of 3835. This indicates that all else equals, a 0.01 raw increase in winning percentage would lead to an increase of 38 fans. This regression model is the only regression model in the senior project that features the winning percentage as a statistically significant variable. This finding indicates that a .500-winning percentage would lead to 1918 fans attending.

Similar to the other regression models, the Friday – Sunday variable features statistical significance with a 99 percent confidence level. Featuring a coefficient of 5247, all else equal, a weekend game will draw 5247 more fans than a game played during the week. The magnitude of this effect is around 1000 fans less than the total regression. This change in magnitude could be explained by the findings detailed with regard to the change in magnitude of the Opening Day variable.

The next variable found statistically significant in the regression model is the summer variable. Featuring a coefficient of 1388 and a statistical significance with a 99 percent confidence level, all else equal, a game played in the summer draws 1388 more fans than a game played in non-summer months.

Similar to the regression model found in *Section II.B* each of the three variables introduced into the regression to account for the economy were statistically significant with a 99 percent confidence level. The Phoenix (Metro) GDP per capita featured a –1 coefficient, identical to the previous regression model. This means that all else equal, for every one-dollar increase in GDP per capita, there will be one less fan attending the game. Similar to the last model, the sign on the coefficient makes this a surprise result that is explained in the discussion of the regression model for *Section II.B*.

The Arizona GDP growth rate featured a coefficient of 466, thus meaning that all else equal, for every one percent increase in GDP growth rate, 466 more fans will attend the game. This finding is an expected result of the regression. The final statistically significant variable, the Arizona unemployment rate featured a coefficient of –902. This indicates that all else equal, a one percent increase in the unemployment rate leads to 902 fewer fans attending the game. This also was an expected result of the regression model.

This model serves as the baseline for comparison in *Section II.D* of the senior project. Comparing and contrasting the results of these regression models will directly answer the first research question posed by the senior project as stated in *Chapter I* and *Chapter III*.

Section II.D: Regression Post-Dynamic Pricing with Macroeconomic Considerations

The purpose of this section is to compare and contrast the regression results against those in *Table 4* to analyze the shifts in attendance determinants with regard to dynamic pricing. This will directly answer the first research question of this senior project.

Table 5 contains the results from the regression of the games from the 2011 season through the end of the 2023 season, excluding the 2020 season. The regression featured an adjusted R-squared of .5352, thus explaining around 54 percent of the variance in attendance levels.

Table 5: Regression Post-Dynamic Pricing Introduction

	Coefficients	*Standard Error*	*t Stat*	*P-Value*
Intercept	-140744	37151.2504	-3.7884	0.0002
cLI	-748	507.3657	-1.4748	0.1406
Division Rival	931**	374.7009	2.4834	0.0132
Promotion U15	-4404***	1207.1874	-3.6482	0.0003
Promotion	3499***	485.2453	7.2114	1.1248E-12
Home Opening Day	18391***	1767.4092	10.4057	4.2679E-24
Time of Day	545	427.8779	1.2735	0.2031
SB	-27	196.9591	-0.1366	0.8914
BA	61465***	9842.3005	6.2450	6.3670E-10
Win Pct	2813	2324.6181	1.2100	0.2266
Friday – Sunday	7539***	451.6661	16.6908	4.4226E-55
Summer	1357***	384.1389	3.5316	0.0004
Games Behind	-147***	29.9369	-4.9143	1.0479E-06
Phoenix (Metro) GDP Per Capita	3***	0.8182	4.0603	5.3016E-05
AZ GDP Growth Rate	-1762***	216.1188	-8.1550	1.0904E-15
AZ Unemployment Rate	538**	247.5739	2.1732	0.0300
R-Squared	0.5424			
Adjusted R-Squared	0.5352			
Significance F	1.405E-150			

Significance Codes:
*** The variable is significant with a one percent level
** The variable is significant with a five percent level
* The variable is significant with a ten percent level

The following equation describes the findings of the regression to calculate total attendance on a game-to-game basis, post-dynamic pricing introduction:

Regression Equation:

Attendance = – 140744 – 748 (cLI) + 931 (Division Rival) – 4404 (Promotion U15) + 3499 (Promotion) + 18391 (Home Opening Day) + 545 (Time of Day) – 27 (SB) + 61465 (BA) + 2813 (Win Pct) + 7539 (Friday – Sunday) + 1357 (Summer) – 147 (Games Behind) + 3 (Phoenix Metro GDP per Capita) – 1762 (AZ GDP Growth Rate) + 538 (AZ Unemployment Rate)

The first statistically significant variable of the post-dynamic pricing introduction regression model is the division rival variable. With a statistical significance level of 95 percent, the variable carries a coefficient of 931. This indicates that all else equal, a game played against a division rival will attract 931 more fans than games played against non-divisional rivals. Comparing this result to the regression model in *Section II.C*, the magnitude of the positive impact on attendance is nearly 300 fans less than in the era before dynamically priced tickets.

Unlike the pre-dynamic pricing era regression model, the under-15 promotion variable is statistically significant in the post-dynamic pricing regression model with a 99 percent confidence level. The coefficient of –4404 indicates that all else equal, a game with a promotion focused on children will have 4404 fewer fans than a game with a generic promotion. Combining this finding with the promotion variable, which also is statistically significant with a 99 percent confidence level and a coefficient of 3499, produces a *net negative* impact of 905 fans. This means that the true effect of an under-15 promotion is a loss of 905 fans in the post-dynamic pricing era. This is an unexpected result of the regression.

Analyzing the data set, there is no clear trend in the games from 2011 to 2023 that indicates the under-15 promotion variable would lead to a net negative effect on game-to-game attendance. Therefore, there is no clear reason for the attendance to drop when having a promotion focused on kids, and requires further research on teams with a similar time constraint to understand this result on a larger scale. The impact of a standard promotion, however, remained an expected result of the regression. As previously mentioned, the promotion variable's coefficient of 3499 indicates that all else equal, 3499 more fans attend a game with a promotion than a game without a promotion.

As found in all of the regression models, the home Opening Day variable was statistically significant with a 99 percent confidence level. Featuring a coefficient of 18391, all else equal, the home Opening Day game will have 18391 more fans than the rest of the games played throughout the season. Comparing the results to the pre-dynamic pricing model, the magnitude of the impact of the Opening Day variable on attendance is nearly 4000 fans larger than pre-dynamic pricing. This finding is an expected result; however, the change in magnitude could be explained by analyzing the data.

The Diamondbacks were consistently in the lower third of MLB for average and total attendance per game in the time period analyzed for this regression. Since the home Opening Day games typically feature some of the largest crowds for the entire season, these years faced more variance in attendance per game than the games observed in the pre-dynamic pricing era.

Another statistically significant variable with a 99 percent confidence level is the batting average variable. The batting average variable features a coefficient of 61465, thus indicating that all else equal, a .01 increase in team batting average would lead to a 615 fan increase in attendance. Comparing this finding to the surprise finding of a negative coefficient in the pre-

dynamic pricing model, this regression finds an expected result. Furthermore, the positive impact of team batting average on attendance correlates to an increase in the importance of offensive power for the MLB as a whole with regard to attendance.

Both the Friday – Sunday and Summer variables remained statistically significant with the post-dynamic pricing model with a 99 percent confidence level. With a coefficient of 7539, weekend games, all else equal, show a 7539 fan increase when compared to games played during the week. From model to model, however, the magnitude of the positive impact on attendance for weekend games increase by over 2000 fans. This change in magnitude could be explained by a greater variance in attendance levels on a game-to-game basis following the 2010 season.

The summer variable coefficient of 1357 indicates that all else equal, a game played in the summer will bring in 1357 more fans than a game played in non-summer months. Comparing this finding to the pre-dynamic pricing model, the magnitude of the coefficient is only slightly less than the pre-dynamic pricing regression model. This comparison indicates little change in the impact of summer games on attendance between the time frames and is an expected result of the regression.

Unlike the pre-dynamic pricing model, the games behind the variable is statistically significant in the post-dynamic pricing model with a 99 percent confidence level. With a coefficient of –147, this finding indicates that all else equal, every additional game behind the division leader the Diamondbacks fall, 147 fewer fans show up. The contrary is true, for every additional game ahead of the rest of the division the Diamondbacks are in, 147 more fans attend the game, based on the regression results.

Finally, all three of the economic consideration variables remained statistically significant. The Phoenix (Metro) GDP per capita variable is statistically significant with a 99 percent confidence interval. Unlike the unexpected findings of the regression models in *Section II.B* and *Section II.C*, the coefficient of 3 for the variable indicates, all else equal, that every 1 dollar increase in the GDP per capita of the Phoenix Metro region leads to an increase of 3 fans.

While this is an expected result of the regression, this change in sign of the variable could be explained by the economic recovery following the financial crisis in 2009. In the years following 2009, economies around the United States were recovering from the crisis within the housing market. This economic downturn likely impacted the number of fans in the seasons following the recession. This era captures the stabilization in the data not captured by the pre-dynamic pricing model. With less variance in the GDP per capita values, this is likely the cause for the change in sign and magnitude.

The Arizona GDP growth rate variable is also statistically significant with a 99 percent confidence level. Featuring a coefficient of –1762, all else equal, a 1 percent increase in the GDP growth rate would lead to a decrease of 1762 fans. This is an unexpected result of the regression and the opposite sign of the pre-dynamic pricing model.

Statistically significant with a 95 percent confidence level, the Arizona unemployment rate features a coefficient of 538. This indicates that all else equal, a 1 percent increase in the unemployment rate leads to an increase of 538 fans. This result is an unexpected result of the regression and of the opposite sign when compared to the findings of the pre-dynamic pricing regression model. Analyzing the data collected, there are no clear trends that would

Multicollinearity tests performed on the dataset show no collinearity issues between the three chosen economic factors. Many initial variables were cut from the regressions in order to avoid issues of multicollinearity, thus providing a more robust regression. This further indicates a confusing finding of a positive coefficient on the Arizona unemployment rate variable and a negative coefficient on the Arizona GDP growth rate percentage variable.

Dividing the models into pre-dynamic pricing and post-dynamic pricing highlighted changes regarding the magnitudes of many of the variables. These changes allow for a new time frame to be studied within MLB literature. *Section III* will take a further deep dive beyond the regression to provide insight into the attendance trends of the Diamondbacks.

Section III: Case Study Analysis

The purpose of this section is to provide more context to the franchise beyond the regression analysis. Understanding the full picture beyond just a regression allows for the research questions to be answered thoroughly. This section will also explain some of the expected and unexpected results from the regression analysis, providing further insights.

Throughout the Arizona Diamondbacks' 25 years as a franchise, 57,715,514 fans have attended home games (Baseball Reference). In that time, the Diamondbacks have achieved winning seasons 13 times, making the playoffs seven times. Of those seven times, the Diamondbacks have played in the World Series twice, once in 2001 and once in 2023, winning in 2001 against the New York Yankees and losing in 2023 to the Texas Rangers.

The first year of the franchise's history (1998 season) saw 3,610,290 fans attend home games for an average of 44,571 fans per game, the second most in the National League for the

year. Since 1998, the Diamondbacks have yet to eclipse this attendance total for a season as well as average fans per game. *Figure 1* shows the trend of home game attendance by year.

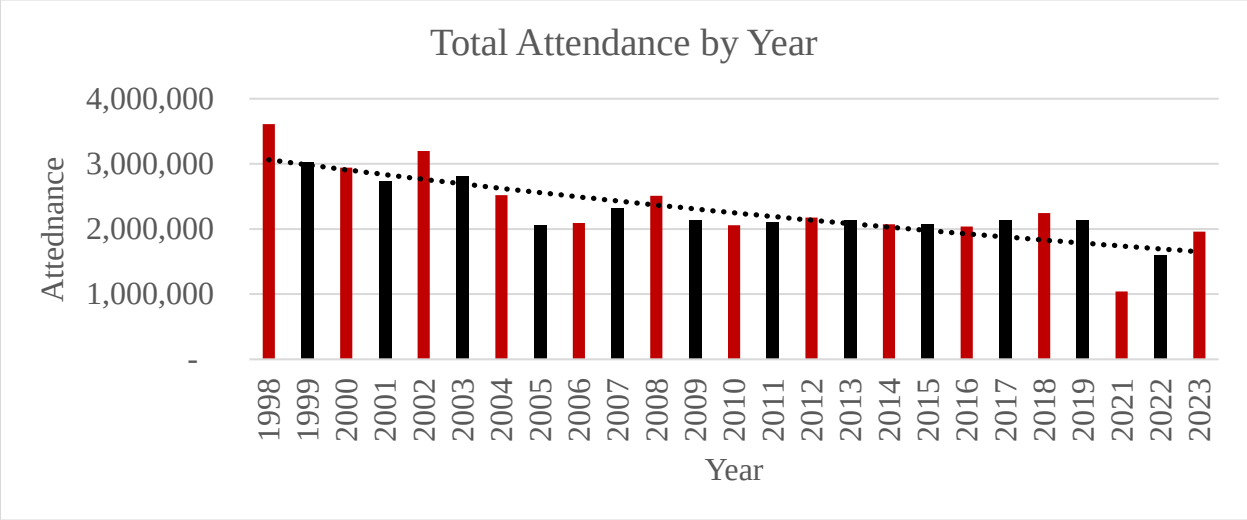

Figure 1: Total Home Attendance by Year

Source: Baseball Reference

The Arizona Diamondbacks recorded their lowest attendance year in 2021, with a total of 1,043,010 fans for an average of 12,877 fans per home game. This placed the Diamondbacks as the 13th least-attended franchise in the National League. This could be explained by the Diamondbacks having the worst season in franchise history, finishing with a record of 52 wins and 110 losses and 55 games behind the leader in the division along with the fears of COVID-19. Given this was the year following the ban of fans attending games, fears still existed of spreading COVID-19, thus hindering the willingness of fans to attend.

Breaking it down by team, the highest average attendance per game for an opposing team was the New York Yankees, with an average of 40,678 over the course of their 11 matchups in Arizona. On the other hand, in the eight matchups between the Arizona Diamondbacks and the Tampa Bay Rays since 2007 (Tampa Bay rebranded from the Devil Rays to the Rays for the 2007 season), there have only been an average of 19,578 fans per game. Similarly, when matched up against the Miami Marlins (Miami rebranded from the Florida Marlins upon their stadium move in 2012), there have only been an average of 19,106 fans at the 35 games. This data is shown in *Figure 2*.

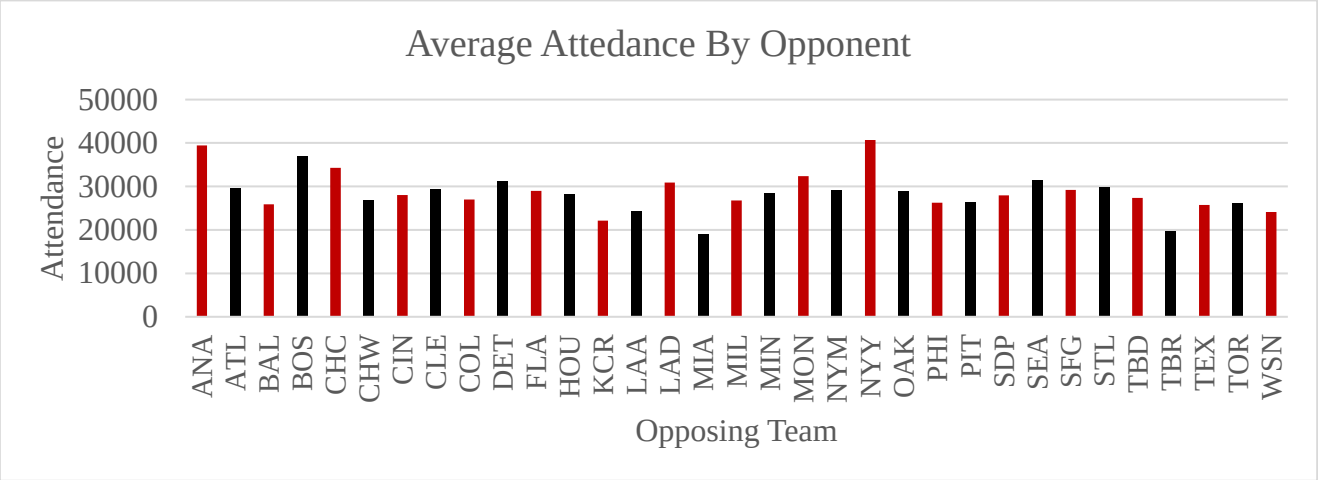

Figure 2: Average Attendance by Opponent

Source: Baseball Reference

When analyzing the day of the week, average attendance peaks when the game starts at night on a Friday, as shown in *Figure 3*. The average attendance for a Friday night game is 48,734 fans, compared to that of a Monday, day game only draws an average of 24,860 fans. *Figure 3* shows the relationship between each day of the week and the game start time. Given the findings of each of the regression models, weekend games are statistically significant in adding fans when compared to non-weekend games.

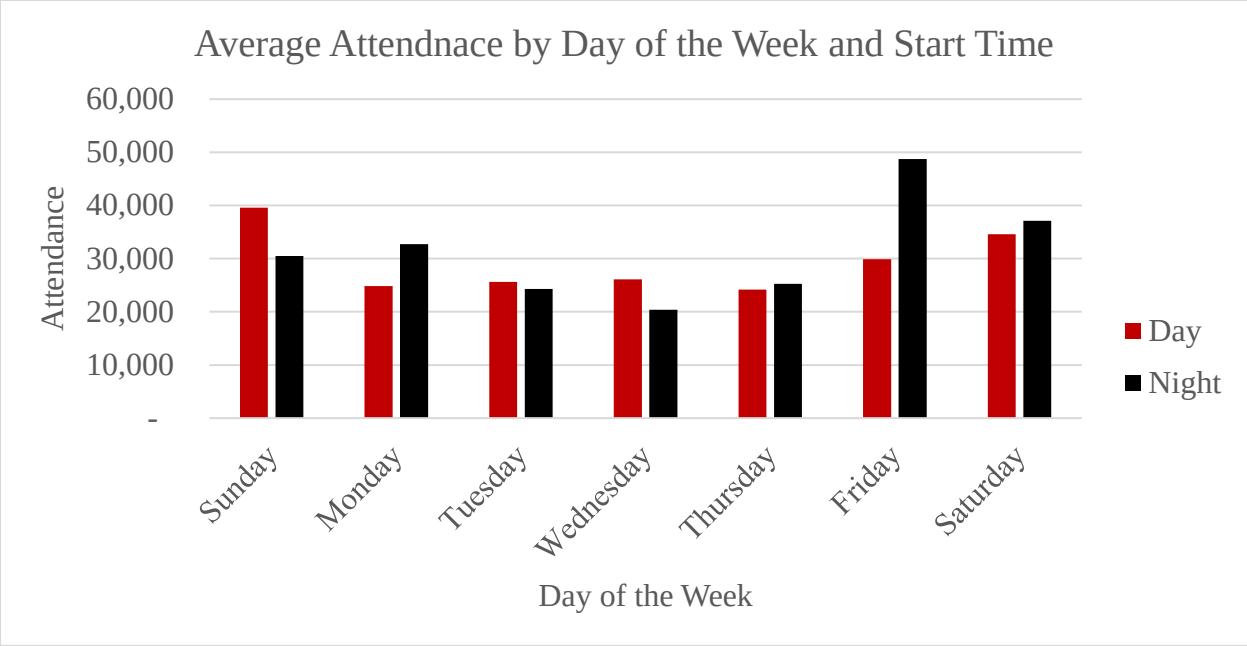

Figure 3: Average Attendance by Day of the Week and Start Time

Source: Baseball Reference

When it comes to payroll by year, the Arizona Diamondbacks have only been in the top half of salary for six (1999 to 2004 and 2014) of their 25 years as a franchise. In 2002, the Diamondbacks peaked with the fourth highest salary in the MLB, with only the New York Yankees, Boston Red Sox, and Texas Rangers spending more than the Diamondbacks. This could be explained by 2002 being the year following the World Series victory of the Diamondbacks over the New York Yankees. The increase in spending was likely due to an attempt by the front office to maintain the championship roster. *Figure 4* shows the trend of payroll by year.

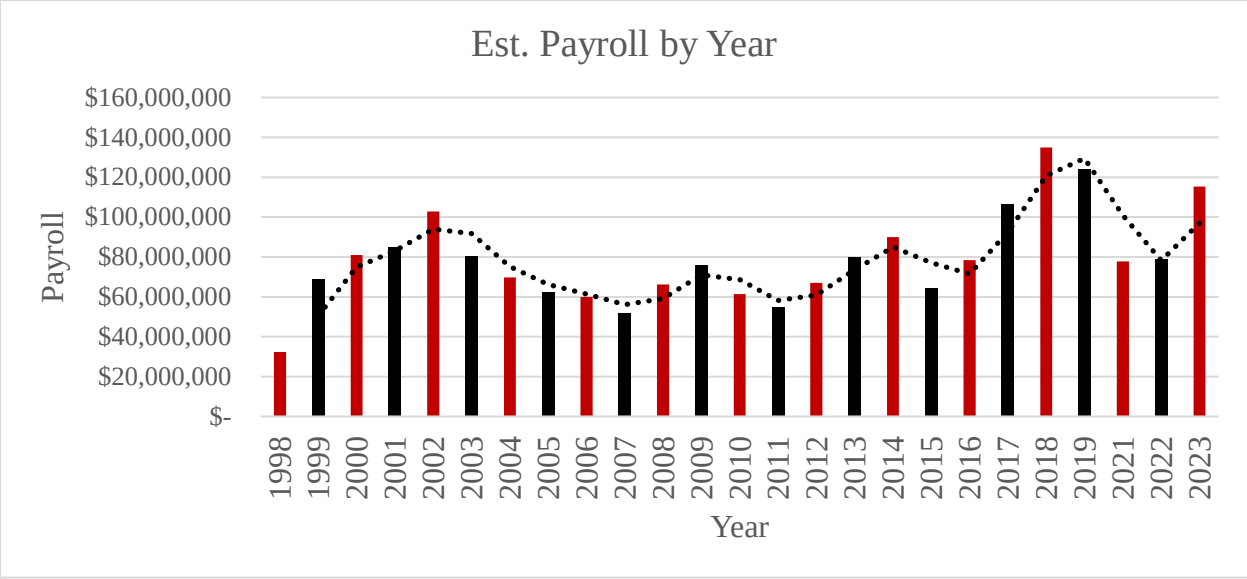

Figure 4: Payroll by Year

Source: Baseball Reference, MLB.com

Fans typically associate a higher payroll with better team performance. Many of the top-performing players receive more valuable contracts than common day-to-day players. Fan bases criticize general managers for spending time on performance. *Figure 5* aligns payroll with attendance to provide insight into these assumptions.

When aligning payroll with attendance, *Figure 5* shows the trendline of attendance stacked against the payroll figure. While it is expected that the higher the payroll, the more talented players on the team, the Arizona Diamondbacks have not followed this trend. An increase in overall MLB revenue year over year along with inflation could explain the lack of payroll dollars impacting the strength of the team. The MLB features a revenue-sharing program, similar to other professional leagues, that divides portions of overall revenues out towards the teams in the league. Given the regression analysis, it is significant that the better the team batting average, the more attended the game will be, these findings of payroll and attendance are unexpected.

However, as previously mentioned, the Arizona Diamondbacks, when compared to other teams in the National League, and subsequently the rest of the MLB, are typically not seen as a heavy spender. Since 2014, the Diamondbacks have failed to reach beyond the 17th-highest payroll. This approach has led to payroll having little correlation to attendance factors.

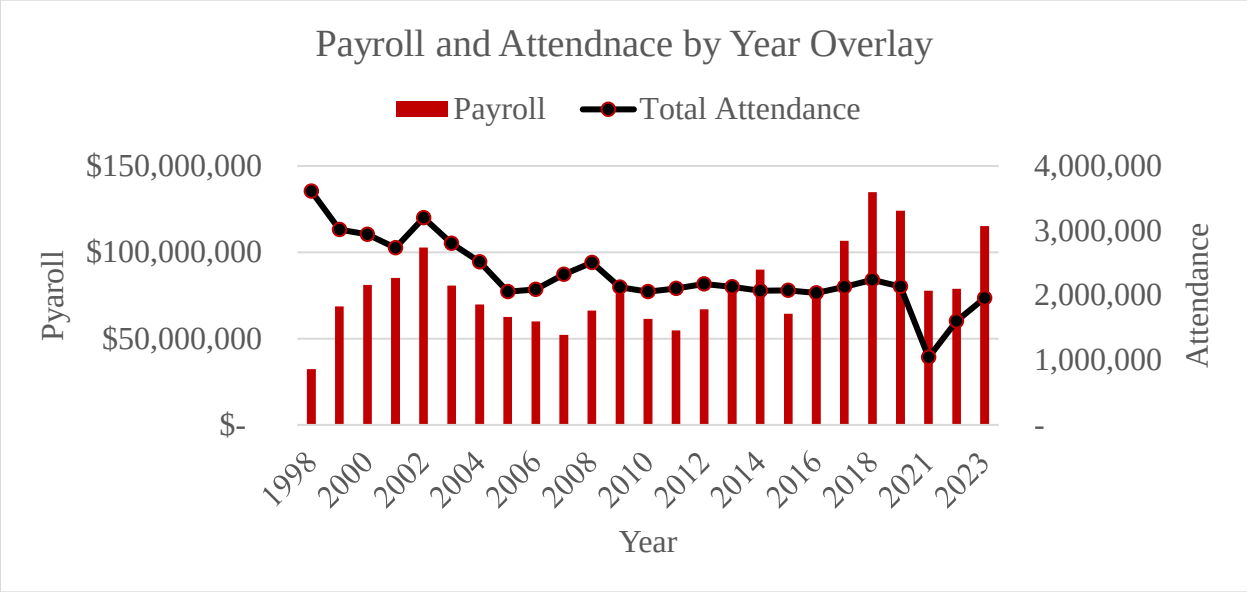

Figure 5: Payroll and Attendance by Year Overlay

Source: Baseball Reference, MLB.com

Figure 5 also shows that in 2002, when the Arizona Diamondbacks had the fourth-highest payroll in the MLB, they also saw their second-highest attendance. This, however, could be explained by 2002 being the year after the Diamondbacks won their only World Series title in 2001.

In 2014, when the Diamondbacks had a payroll rank of 11th in the MLB, they only saw just over two million fans. Thus, providing more evidence for the inconsistencies with the Diamondbacks payroll and their attendance.

Overall, these findings help paint the total picture for the Arizona Diamondbacks and their attendance factors. *Chapter IV* will further discuss the implications of these findings and provide follow-up research potential and policy recommendations.

Note: Payroll information as reported to MLB

Chapter IV: Summary, Conclusions, and Policy Recommendations

The purpose of this chapter is to provide the conclusions of the senior, outline a basis for further research to extend this work and provide policy recommendations to executives.

Section I: Senior Project Summary

The focus of this senior project was to take a deep dive into the Arizona Diamondbacks franchise to compare the statistically significant determinants for attendance in comparison to other stylized determinant studies in Major League Baseball. *Chapter I* of the senior project discusses important background information on the Diamondbacks as well as provides the research questions discussed in *Section II* of this chapter.

The Arizona Diamondbacks, which officially became a franchise in 1998, experienced a quick rise to a World Series title in 2001. Since then, the Diamondbacks have struggled performance-wise. This study aims to explain the variation in attendance levels on a game-to-game basis. Utilizing a combination of regression analysis and a case study approach, the senior project provides a full picture of the 25 seasons as a franchise for the Diamondbacks.

Chapter II of the senior project provides an in-depth literature review of various studies concerning sports attendance. The literature review is divided into two sections: stylized determinant studies and dynamic pricing studies. This allows for both portions of the senior project to be covered, as the project aims to analyze attendance before and after the introduction of dynamic pricing along with a whole franchise history view.

Chapter II of the senior project also provides the foundation for the research by highlighting these previous studies. The prior studies explain variable selection for the regression analysis conducted in *Chapter III* along with explaining the current gaps in the literature that the

research questions are derived from. The main gaps in the literature include the lack of studies on the long-term impacts of the introduction of dynamic pricing and the differentiation of promotional types.

This senior project bridges these gaps by providing 12 years of data beyond the season that the Arizona Diamondbacks introduced dynamic pricing (2011) to follow the long-term impacts. Furthermore, through the division of the regression on the games before and after dynamic pricing, the study analyzes how the determinants of attendance change. This models the prior research in Chapter II that analyzes changes in attendance determinants in significant periods across the MLB like the introduction of free agency.

This senior project also explores promotion types in attendance determinant studies. This establishes a new variable that can be utilized in future research on promotion impact on attendance. *Section II* of this chapter will summarize the findings from *Chapter III* to further discuss these additions to the attendance determinant study. *Section III* of this chapter will also provide future research implications, and *Section IV* will provide policy recommendations for executives.

Section II: Findings Summary

Summarizing the findings from *Chapter III*, the stylized determinants of MLB attendance moderately translated to the case study of the Arizona Diamondbacks. In the regression with macroeconomic considerations, the winning percentage was not a statistically significant factor. In many of the prior attendance determinant studies, winning percentage is a main driver in attendance.

The regression also derived a statistical significance in the dynamic pricing dummy variable. These findings indicated a negative impact of dynamic pricing on attendance for the Diamondbacks. Unlike the findings on dynamic pricing research outlined in *Chapter II*, these findings show that the Diamondbacks have not realized gains in attendance concerning dynamic pricing.

Similar to other stylized determinant studies outlined in *Chapter II*, the regression found results consistent with summer and weekend games having a positive impact on attendance. The regression also realized a large increase in fan attendance for the home Opening Day game. These findings are consistent across the MLB, with opening day attendance rates consistently higher along with summer and weekend games being the highest attended with regards to a whole season.

The promotional findings indicate that consistent with other attendance studies with a promotional focus, there is a significant, positive impact of promotions on attendance. When considering promotions geared towards kids, as determined by the classification of an under 15 promotion per the Diamondbacks, there is a slight decrease in the overall impact on attendance, but for the total regression still a *net positive* impact on attendance.

Overall, with the findings of the regression analysis paired with the case study, the study indicates the Arizona Diamondbacks are somewhat conforming with the rest of the MLB when it comes to attendance studies. The findings of the lack of statical significance for the winning percentage and

Section III: Further Research Implications

Further research into the effects of promotions for kids across the league would provide a deeper insight into the true impacts. While this study analyzed the impacts of promotions for kids over the 25 years of the Arizona Diamondbacks being a franchise, the research only covers the Diamondbacks. Expanding the research beyond a single team will broaden the pool of games analyzed. This would also allow for the current studies on promotions to be updated with a new level of understanding.

Introducing other promotional data points, such as the impact of price-based promotions on tickets or food-based promotions within the ballpark could uncover interesting findings in terms of what type of promotion is most effective at drawing the largest crowd. While this study is the first to separate promotions into general and children-focused promotions, there are varying types of promotions found across the MLB. Due to time constraints and data limitations, these areas were not able to be explored.

Not only would conducting a similar study on more franchises increase the data points within the data set, but it would also allow for a more robust regression analysis. As highlighted in *Section I*, the Arizona Diamondbacks regression found non-traditional results when compared with the rest of the MLB. Conducting the regression on a broader scale will eliminate these unpredictable findings and provide a clearer picture on a wide-scale basis of the MLB.

With regard to dynamic pricing, there are many data limitations. Many of the current studies that analyze dynamic pricing accumulate data throughout the season. Ticketing partners with the MLB, like Ticketmaster, StubHub, and Seat Geek only allow for forward-looking data to be analyzed. An in-depth study into dynamic pricing would entail a large-scale data mining endeavor throughout the season or a partnership with one of the major ticketing partners of the

MLB. This data would allow for fluctuations in price before the game to be captured as well as average prices. Collecting this data would allow for the long-term impacts of dynamic pricing to be analyzed and fill that gap in the literature.

This could lead to two different avenues of research concerning the impacts of different promotional types. A researcher could analyze the impact of varying promotions on ticket prices to capture that avenue of dynamic pricing. The second avenue would be applying a similar study approach with a larger regression that includes more in-depth dynamic pricing models.

The final suggestion for further research stems from a budding rule change in the MLB. At the beginning of the 2023 season, the MLB instituted a pitch clock across the league. The MLB determined that there was a decline in attendance from game to game across the league as a result of games taking too long. Given the findings of this study, it appears that the duration of games did not affect attendance (Riccardelli and Appel, 2023). While the time period is too early to evaluate, studies that focus on this rule change could provide an analysis of these rule changes and their effects on attendance.

Section IV: Policy Recommendations

Overall, the findings of the research provide a framework for understanding the attendance determinants of the Arizona Diamondbacks over the course of their 25-year history. Understanding these findings will allow for decision-makers to understand what draws a crowd and make necessary adjustments to variables that can be controlled by decision-makers.

The findings indicate that promotions during the summer months will generally lead to an increase in attendance. When curating a promotional schedule, if there is a significant downturn in the non-summer months, an attendance boost could be realized with the addition of a

promotion to the season. The findings also show that when considering a promotion, decision-makers should aim to feature a neutral promotion over a promotion geared toward kids.

In times of poor performance, the decision-makers of the Diamondbacks can be assured that the winning percentage, in the case of this study, is not statistically significant in drawing fans. These findings indicate a loyal fan base that enjoys attending Diamondbacks games. Furthermore, the effectiveness of dynamic pricing should be called into question by decision-makers. The benefits of dynamic pricing concerning capturing the missed revenue from non-dynamically priced tickets must be weighed against the effects it has on game-to-game attendance. If front office executives truly care about a “family-friendly” experience, then the Diamondbacks’ management must evaluate where their priorities lie: revenue maximization or attendance maximization.

Bibliography

Ahn, S. C., & Lee, Y. H. (2014). Major League Baseball Attendance: Long-Term Analysis Using Factor Models. Journal of Sports Economics, 15(5), 451-477. https://doi.org/10.1177/1527002514535171

Baseball Reference

Coates, Dennis & Humphreys, Brad. (2007). Ticket Prices, Concessions, and Attendance at Professional Sporting Events. International Journal of Sport Finance. 2. 161-170.

Coates, D., Humphreys, B.R. And Zhou, L. (2014), Reference-Dependent Preferences, Loss Aversion, And Live Game Attendance. Economic Inquiry, 52: 959-973. https://doi.org/10.1111/ecin.12061

Corrado, B. (2023, November 27). D-Backs Selling Home Game Season Pass For $299. FOX 10 Phoenix. https://www.fox10phoenix.com/sports/d-backs-selling-home-game-season-pass-for-299

Courty, P., & Davey, L. (2020). The Impact of Variable Pricing, Dynamic Pricing, and Sponsored Secondary Markets in Major League Baseball. Journal of Sports Economics, 21(2), 115-138. https://doi.org/10.1177/1527002519867367

Drayer, J., Rascher, D., & McEvoy, C. (2012). An Examination of Underlying Consumer Demand and Sport Pricing Using Secondary Market Data. *Sport Management Review*, 15, 448-460.

Drayer, Joris; Shapiro, Stephen L.; and Lee, Seoki, (2012). "Dynamic Ticket Pricing in Sport: An Agenda for Research and Practice". *Human Movement Sciences Faculty Publications*. 17. https://digitalcommons.odu.edu/hms_fac_pubs/17

Elmaghraby, W. & Keskinocak, Pinar. (2003). Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions. Management Science. 49. 1287-1309. 10.1287/mnsc.49.10.1287.17315.

Hodell, N. (2021, September 16). Report: Diamondbacks Offer the Most Affordable MLB Fan Experience. Arizona Sports. https://arizonasports.com/story/2818207/report-says-diamondbacks-offer-the-most-affordable-mlb-fan-experience/

Kimes, S. E. (1989). The Basics of Yield Management. Cornell Hotel and Restaurant Administration Quarterly, 30(3), 14-19.

Klapper, C. (2023, May 15). Arizona Diamondbacks Summer Pass on Sale For $99. ABC15 Arizona in Phoenix (KNXV). https://www.abc15.com/sports/arizona-diamondbacks-summer-pass-on-sale-for-99

Lee, Y. H. (2013). Estimation of Temporal Variations in Fan Loyalty: Application of Multifactor Models. In P. Rodriguez, S. Kesenne, & J. Garcia (Eds.), The econometrics of sport (pp. 135–153). Cheltenham, England: Edward Elgar.

McDonald, Mark & Rascher, Daniel. (2000). "Does Bat Day Make Cents? The Effect of Promotions on the Demand for Major League Baseball," MPRA Paper 25739, University Library of Munich, Germany.

Meyer, J. (2023, May 8). Most successful MLB franchises of All Time. The Poolside Post. https://www.runyourpool.com/articles/2023/05/08/most-successful-mlb-franchises-of-all-time/

Neale, W. C. (1964). “The Peculiar Economics of Professional Sports.” Quarterly Journal of Economics, 78(1), 1–14.

Parris, D. L., Drayer, J., & Shapiro, S. L. (2012). Developing a Pricing Strategy for the Los Angeles Dodgers. Sport Marketing Quarterly, 21(4), 256-264. Retrieved from https://www.proquest.com/scholarly-journals/developing-pricing-strategy-los-angeles-dodgers/docview/1324536397/se-2

Pavlović, S. (2023, October 25). When was the last time the Diamondbacks went to the World Series? Diario AS. https://en.as.com/mlb/when-was-the-last-time-the-diamondbacks-went-to-the-world-series-n/

Qcue. https://www.qcue.com/dynamic-pricing

Ricciardelli, M., & Appel, M. (2023, March 30). *Poll shows enthusiasm for ’23 MLB season, aided in part by pitch clock and shorter games; MLB fans favor pitch clock by more than 4 to 1. - Seton Hall University*. Poll Shows Enthusiasm for ’23 MLB Season, Aided in Part by Pitch Clock and Shorter Games; MLB fans favor Pitch Clock by more than 4 to 1. - Seton Hall University. https://www.shu.edu/business/news/poll-shows-enthusiasm-for-new-baseball-season.html

Rottenberg, S. (1956). The Baseball Players’ Labor Market. *Journal of Political Economy*, *64*(3), 242–258. http://www.jstor.org/stable/1825886

Shapiro, Stephen L. and Drayer, Joris. (October 24, 2014). An Examination of Dynamic Ticket Pricing and Secondary Market Price Determinants in Major League Baseball. Sport Management Review 17 (2014) 145–159, Available at SSRN: https://ssrn.com/abstract=2457074

Steve the Ump. https://www.stevetheump.com/Payrolls.htm

Summers, Jeff. Now Hitting. https://www.nowhitting.com/index.php

Sweeting, A. (2012). Dynamic Pricing Behavior in Perishable Goods Markets: The Case of Secondary Markets for Major League Baseball Tickets. Journal of Political Economy, 120(6), 1133-1172.

Vrooman, J. (1996). The Baseball Players' Labor Market Reconsidered. Southern Economic Journal, 63(2), 339–360. https://doi.org/10.2307/1061172

Xu, Joseph Jiaqi, Fader, Peter S., Veeraraghavan, Senthil (2019) Designing and Evaluating Dynamic Pricing Policies for Major League Baseball Tickets. Manufacturing & Service Operations Management 21(1):121-138. https://doi.org/10.1287/msom.20