Understanding Broadway Gross: Key Factors and Growth Strategies*

Optimizing Revenue Per Seat While Addressing Modest Holiday Gains and Revenue Declines in Tony Award Months

Xuanel Zhou

December 2, 2024

This study examines factors influencing weekly gross revenue on Broadway, focusing on key predictors such as average ticket price, seating capacity, number of performances, holiday weeks, and Tony Award months. A $1 increase in average ticket price adds $9,168 to weekly revenue, while an additional performance contributes $70,688, and each new seat generates $699.82, underscoring the importance of pricing and demand strategies over simply expanding seating capacity. Holiday weeks boost revenue by $16,684, but this modest increase relative to NYC's tourism surge shows untapped potential for Broadway to enhance its appeal as a premier holiday destination. Conversely, Tony Award months are linked to a $7,140 decline in weekly revenue, highlighting the need to reposition the awards season as an industry-wide promotional opportunity to benefit all productions.

Table of contents

*Code and data are available at: Broadway_Gross.

1 Introduction

Broadway, located in the heart of New York City's Theater District, comprises 41 theaters in Midtown Manhattan, extending from 42nd to 53rd Streets and between 6th and 8th Avenues (Labor 2021). Renowned as a cultural and economic landmark, Broadway plays a pivotal role in shaping the city's global reputation and driving its economy. For stakeholders, understanding the drivers of Broadway revenue is essential, as it directly informs strategic decisions on pricing, scheduling, and resource allocation to maximize profitability and sustain operations. This study seeks to identify and quantify the key factors influencing Broadway's weekly gross revenue, utilizing datasets created by Alex Cookson (Cookson 2020), which are derived from data originally provided by Playbill (Playbill 2024).

The estimand of interest in this study is the average change in weekly gross revenue for Broadway productions, attributed to specific predictors such as ticket price, seating capacity, number of performances, holiday weeks, and Tony Award months, while holding other factors constant. This measure captures the independent impact of each variable on revenue. While the true estimand is inherently unknown due to potential unobserved confounders, it is approximated using multiple linear regression, relying on the available data and model assumptions. Clearly defining the estimand ensures the analysis aligns with the research objectives, facilitates interpretation of the relationships between predictors and revenue, and addresses potential biases from factors such as omitted variables or measurement errors, ensuring the findings remain meaningful and relevant for revenue optimization.

The analysis shows that ticket pricing, seating capacity, performance frequency, and holiday weeks positively impact Broadway's weekly gross revenue. Specifically, each additional performance generates $70,688, while holiday weeks add $16,684 to weekly revenue. Conversely, Tony Award months are associated with a $7,240 decline in weekly gross revenue, underscoring the need to adjust strategies to better capitalize on this period.

These findings highlight both the opportunities and challenges Broadway faces in maximizing its revenue. Prioritizing revenue per seat over simply increasing seating capacity presents a more effective approach to boosting Broadway's financial performance. Additionally, significant potential can be unlocked through strategies such as holiday packages, partnerships with tourism boards, and enhancing the on-site experience with festive activities and merchandise. To address the negative impact of Tony Award months, initiatives exclusive ticket bundles, targeted promotions, and expanded media engagement can help capitalize on the event's visibility.

The remainder of this paper is structured as follows: Section 2 provides an overview of the dataset and visualizations of the variables. Section 3 details the modeling approach, including predictor selection and the regression model structure. Section 4 presents the outcomes of the model. Finally, Section 5 explores the key findings, highlights the main takeaways, addresses the study's weaknesses, and outlines potential next steps. Section A contains the idealized survey designed to gather more detailed information, while Section B provides supplementary

information about the current dataset. Lastly, Section C includes diagnostic analyses of the model.

2 Data

2.1 Overview

This study utilizes R packages (R Core Team 2023) for data cleaning and analysis, incorporating libraries from tidyverse (Wickham et al. 2019), ggplot2 (Wickham 2016), knitr (Xie 2024), arrow (Richardson et al. 2024), here (Müller 2020), scales (Wickham, Pedersen, and Seidel 2023), and modelsummary (Arel-Bundock 2022). The **tidyverse** package is employed for data manipulation, such as filtering, grouping, and reshaping datasets. **ggplot2** is used for data visualization, creating a range of plots to explore and present trends in the data. **knitr** facilitates the integration of analysis and reporting by generating dynamic reports in R Markdown. **arrow** is used for efficient handling of large datasets, particularly for reading and writing data in Parquet format to optimize storage and processing. The **here** package ensures reproducibility by standardizing file paths in the project directory, making it easier to locate and load data files. Finally, **scales** is employed to enhance the readability of plots by formatting axes, labels, and scales, ensuring clarity in the presentation of numeric and categorical data.

The dataset, compiled by Alex Cookson (Cookson 2020) and derived from data originally provided by Playbill (Playbill 2024). The data cleaning process involved grouping observations and removing missing values. The final dataset consists of 14,519 observations across 9 variables: `week_number`, `weekly_gross`, `avg_ticket_price`, `seats_in_threatre`, `performances`, `year`, `month`, `holiday_week` and `Tony_Award`. The dataset used for this analysis spans from the beginning of 2010 to the end of 2019.

2.2 Measurement and Considerations

The dataset for this study originates from Playbill (Playbill 2024), which collects information about Broadway shows through direct reporting and partnerships with industry organizations. A significant portion of the data, particularly box office grosses and performance metrics, is provided by The Broadway League (League 2024b), a national trade association for the Broadway industry.

The transformation of real-world phenomena into the dataset begins with data collection. The Internet Broadway Database (IBDB), maintained by The Broadway League, compiles detailed records of Broadway productions using theater programs to document cast and crew information ((IBDB) 2024). Additional data is supplemented by newspaper and magazine reports, theatrical textbooks, interviews with theater professionals, and The League's archival

records. This information is then processed into key metrics, including ticket sales, attendance, revenue, and theater utilization, which are published weekly. Before release, The League's research department verifies the data's accuracy and removes personal information to ensure privacy.

However, converting real-world phenomena into publicly available dataset entries inevitably leads to the loss of important contextual details in such circumstances. For example, productions that include premium tickets or standing-room sales may disproportionately increase weekly grosses compared to those without these options. While the dataset includes a top ticket price metric, it does not specify the number of tickets sold at this price, leaving gaps in understanding revenue dynamics. Similarly, discounts and promotions, such as rush tickets or group sales, can lower the average ticket price, potentially underestimating the perceived value of tickets sold. Additionally, although the dataset includes weekly metrics, it does not specify whether performances took place on weekdays or weekends, limiting the analysis of gross patterns. These limitations underscore the challenges of translating complex industry data into simplified metrics for analysis.

2.3 Outcome variables

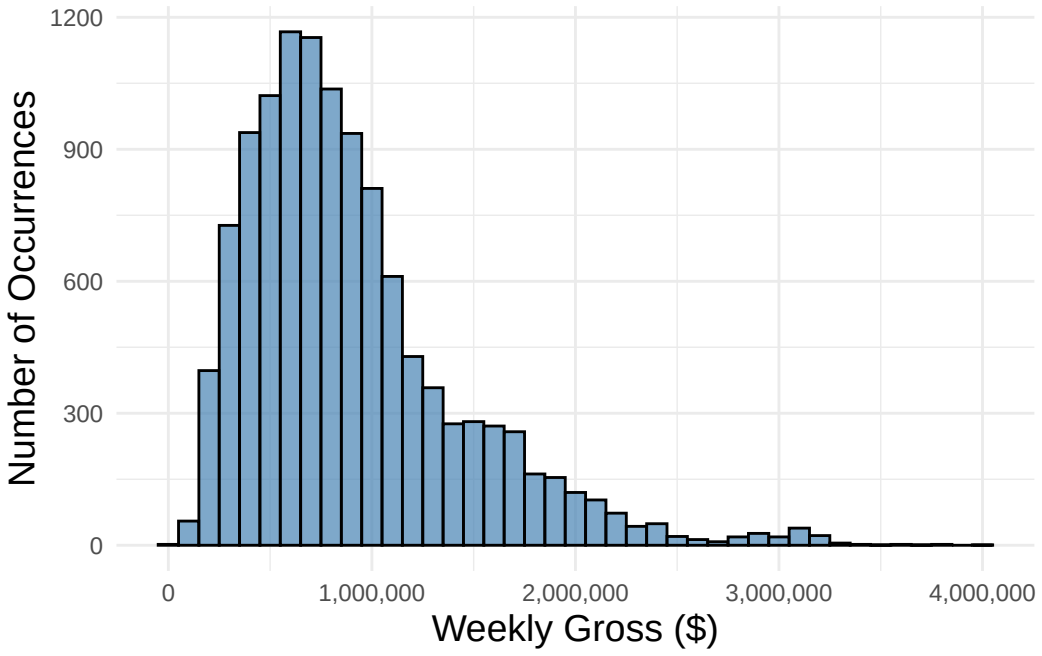

Figure 1: Distribution of Broadway weekly gross. Weekly gross revenues predominantly fall between $500,000 and $2,000,000, with the largest concentration near $700,000 and few instances surpassing $3,000,000.

The outcome variable of interest for this study is the weekly gross revenue for Broadway productions. The distribution, shown in Figure 1, exhibits a right-skewed pattern. Most weeks have gross revenue ranging between $500,000 and $2,000,000, with a declining frequency as revenue increases. with the highest frequency observed around $700,000. A small number of weeks report revenues exceeding $3,000,000, which may represent potential outliers or rare events that warrant further investigation. These high-revenue weeks could stem from unusual circumstances, such as special promotions, holiday seasons, or significant market shifts.

2.4 Predictor variables

2.4.1 Year

Figure 2 captures the Broadway industry's growth over the decade. The yearly averages were calculated by averaging the weekly gross revenue across all Broadway shows for each year. The chart shows a steady increase in average yearly gross revenue, reflecting overall industry growth. From 2013 to 2015, revenues remain relatively stable, followed by sharp growth from 2016 onward, peaking in 2019. A slight dip in 2020 likely reflects the impact of the COVID-19 pandemic on Broadway operations.

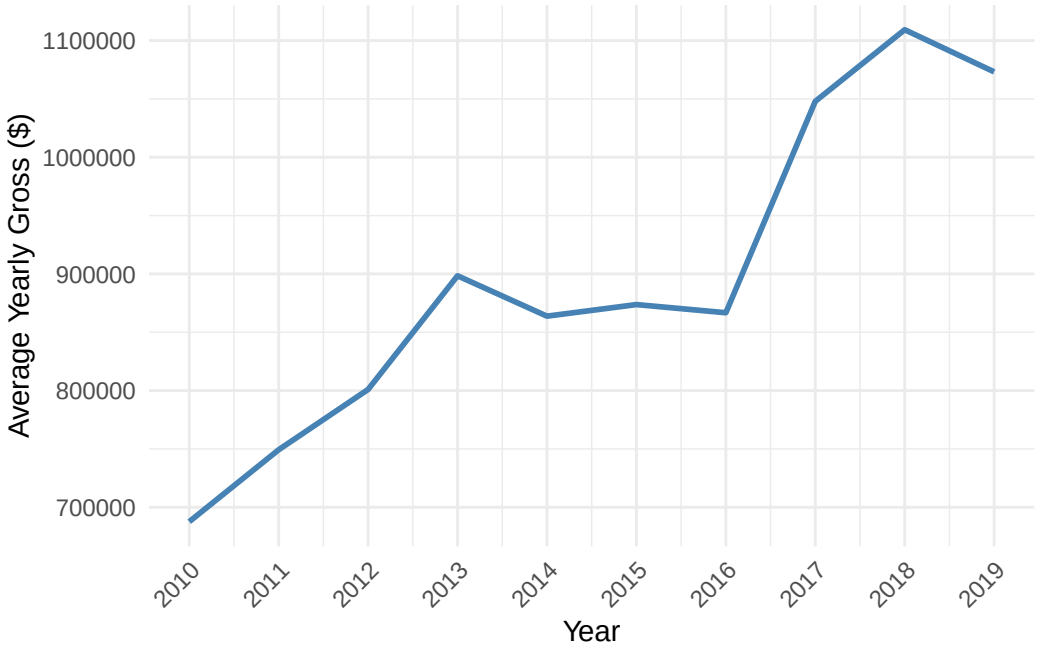

Figure 2: Yearly Gross Revenue Trends for Broadway Productions, spanning from 2010 to 2019. Average yearly gross revenue remained relatively stable between 2013 and 2015. Outside of this period, the industry experienced growth.

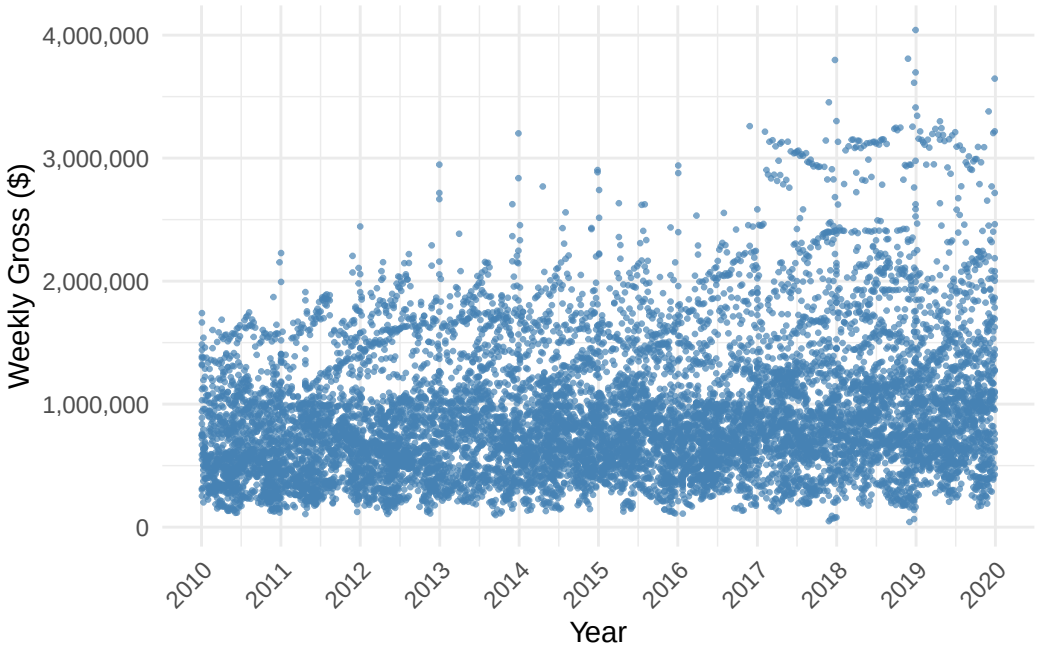

Figure 3: Weekly gross for each Broadway show over time, spanning from 2010 to 2019. Each point represents the gross for a particular week for a specific show. Most productions have weekly gross revenues ranging from $300,000 to $2,000,000, though the data becomes more scattered in later years.

Figure 3 provides detailed information on weekly gross revenues, showing that most values clusters predominantly in the range of $300,000 to $2,000,000, which aligns with the results shown in Figure 1. However, there is a gradual upward trend in the maximum weekly gross over the years, indicating that the Broadway industry experienced growth during this period.

Earlier years (2010–2012) show a compact distribution of weekly gross. From 2015 onward, the range widens, with both higher peaks and broader variation between the lowest and highest values, reflecting increased performance variability among Broadway shows. Notably, some shows at some weeks after 2015 exceed $3,000,000, indicating more outliers and growing disparities between top-performing shows and others, likely influenced by audience preferences, pricing, and industry growth. These peaks are more commonly observed at the end of the year, suggesting a seasonal effect, likely tied to holiday demand and special events.

2.4.2 Average price of tickets sold

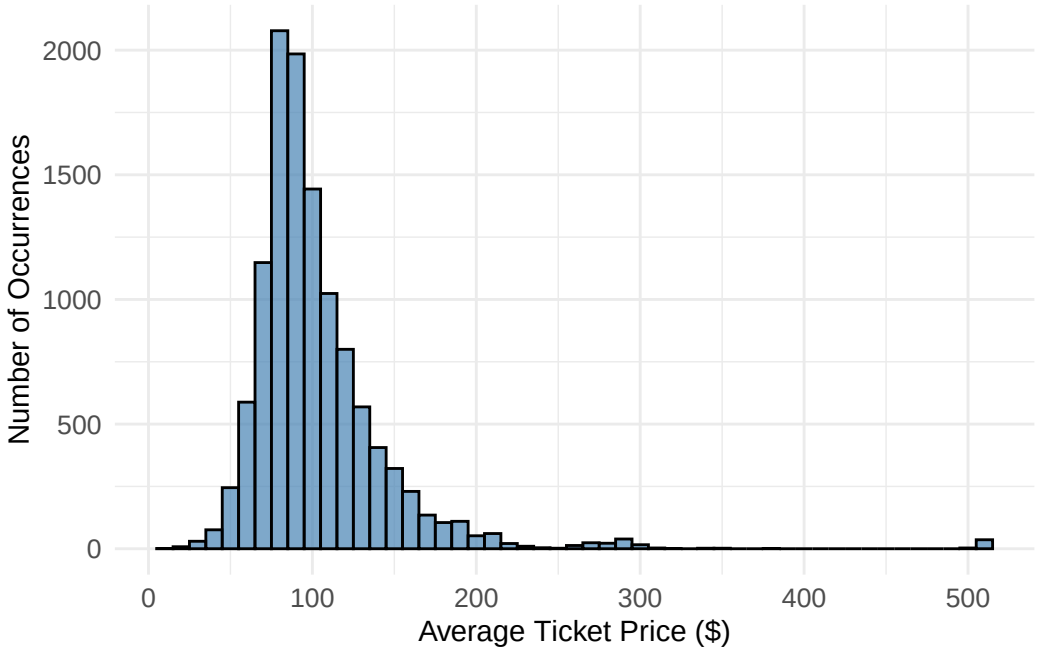

Figure 4: The distribution of average Broadway ticket prices shows that most prices cluster around $100, with a few shows exceeding $500.

Figure 4 shows the majority of average ticket prices are concentrated around $100, with a peak frequency between $80 and $120. The right skewed distribution, indicating that while most ticket prices are relatively affordable, there are instances of higher ticket prices extending beyond $300, with a few exceeding $500.

2.4.3 Theatre seat capacity

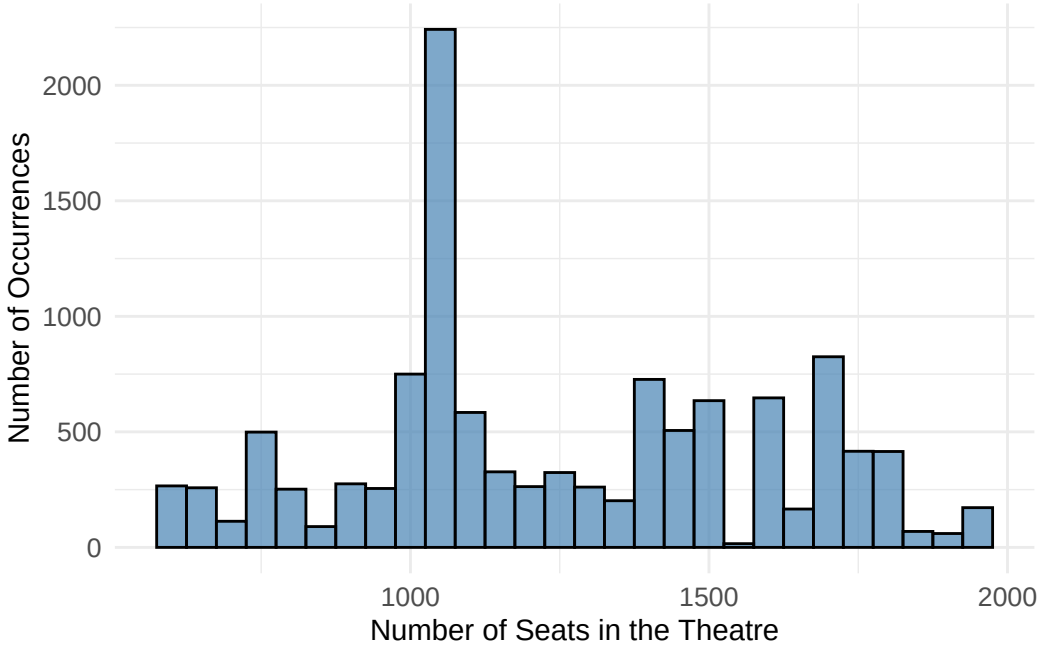

Figure 5: Distribution of Theatre Sizes Based on Number of Seats. Most productions are held in theaters with seating capacities around 1,200. However, there is considerable variation in theater size choices across different productions.

Figure 5 illustrates the distribution of theatre sizes on Broadway, measured by the number of seats. The most common theatre size is around 1,000 seats, with a significantly higher frequency compared to other sizes.

2.4.4 Number of performances in the week

Figure 6 highlights the consistency in Broadway's performance schedules, with the vast majority of shows performing 8 times per week, reaching a frequency of over 10,000 occurrences. A small number of shows have fewer or occasionally more than 8 performances per week, likely due to special events, holidays, or production adjustments.

2.4.5 Holiday Week

The Holiday Weeks was identified based on the major holidays of the year, as referenced in Stanley (2013), and includes the following:

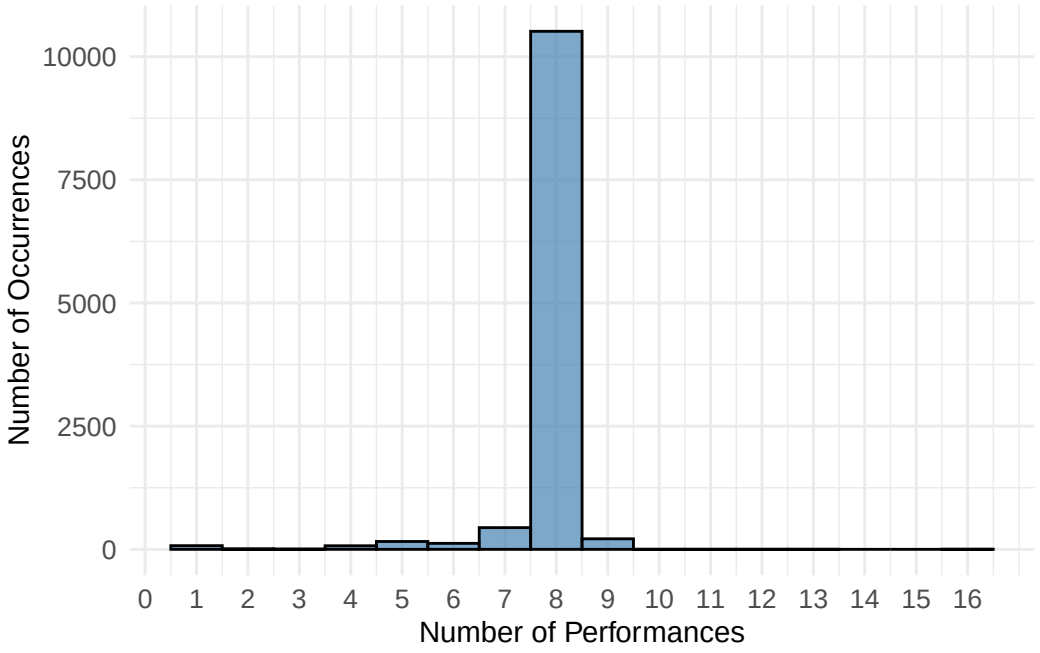

Figure 6: Distribution of Performances. Over 10,000 weeks of shows, most productions adhere to the standard schedule of 8 performances per week. Only a small portion of productions deviate from this, with either more or fewer performances in a given week.

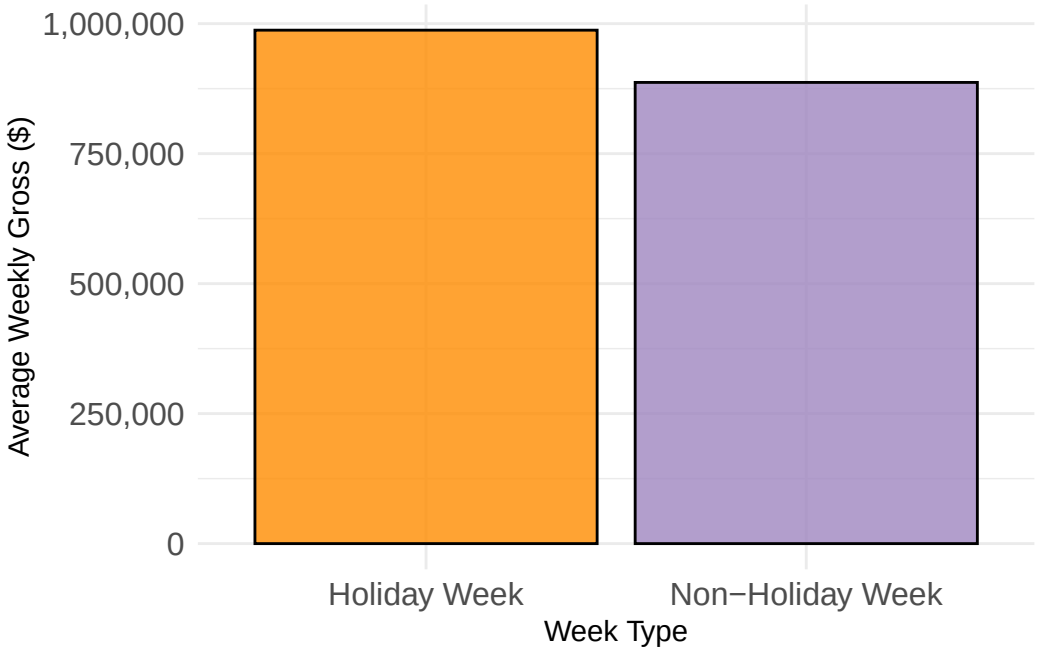

Figure 7: Comparison of Average Weekly Gross Between Holiday and Non-Holiday Weeks. This indicates that holiday weeks tend to result in higher weekly gross revenues for Broadway compared to non-holiday weeks.

- New Year's Week: The week containing January 1st.
- Independence Day Week: The week of July 4th.
- Labor Day Week: The week containing Labor Day, which falls on the first Monday of September.
- Thanksgiving Week: The week containing the fourth Thursday of November.
- Christmas Week: The week containing December 25th.
- New Year's Week: The week of the January 1st.

Non-holiday weeks are defined as all other weeks that do not overlap with these predefined holiday periods. In addition, the average weekly gross revenue for each category was was calculated by summing all weekly gross revenues within the category and dividing by the total number of weeks in that category.

As shown in Figure 7, the average weekly gross revenue during holiday weeks is slightly higher than during non-holiday weeks, though the difference is relatively small. This suggests that holiday weeks may attract slightly larger audiences or higher ticket sales, due to increased leisure time and tourism. However, non-holiday weeks still maintain a comparable level of average weekly gross, demonstrating consistent revenue throughout the year.

2.4.6 Tony Award Period

According to Hodge (2020), the Tony Awards, established in 1947 by the American Theatre Wing, are Broadway's highest honors, held annually in June. The "Best Musical" award is the most prestigious, significantly enhancing a production's success, with nominees 60% less likely to close and winners three times more likely to remain open, solidifying its status as the pinnacle of Broadway achievement.

The variable Tony Award Month refers to the month of the awards, which often sees heightened media attention and anticipation. Off-Award Months include all other months, serving as a baseline for comparison.

Interestingly, Figure 8 shows that the average weekly gross revenue during off-award months is slightly higher than during the Tony Award month. This unexpected outcome may stem from factors like limited new production launches or audience focus shifting toward the awards themselves rather than attending shows.

3 Model

3.1 Model Selection

The objective of this study is to assist Broadway in strategically managing and making decisions to increase gross revenue, making it essential to choose a model that accurately explains

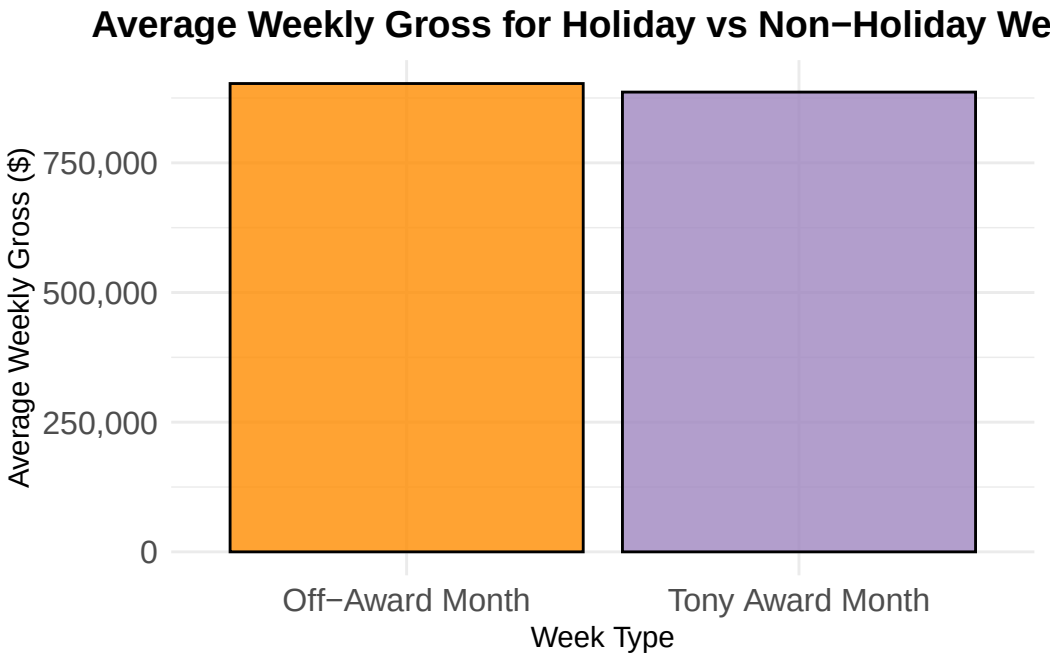

Figure 8: Comparison of Weekly Gross Revenue During Tony Award and Non-Award Months. This indicates that the occurrence of the Tony Awards in a given month does not significantly impact Broadway gross revenues.

the factors influencing revenue. To achieve this goal, several models were evaluated before selecting the final approach, including simple linear regression, decision trees, and multiple linear regression, each with its own strengths and weaknesses.

Simple Linear Regression is straightforward to implement, computationally efficient, and easy to interpret. It provides a clear understanding of the linear relationship between one independent variable and the dependent variable. However, the model is limited to analyzing the effect of a single predictor at a time and cannot account for the influence of multiple interacting variables. This oversimplification can lead to omitted variable bias in the context of a multivariable dataset like this one.

Despite Decision Trees excel at capturing non-linear relationships and interactions between variables, they are prone to overfitting, especially in moderately sized datasets, leading to poor generalization on unseen data.

Based on these considerations, **Multiple Linear Regression** was chosen. While it has limitations, such as reliance on key assumptions and sensitivity to outliers, MLR strikes a balance between simplicity, interpretability, and the ability to analyze the combined effects of multiple predictors. Its coefficients provide a clear quantification of the impact of each independent variable on weekly gross revenue, making it well-suited for understanding the underlying relationships in the data.

To ensure robust evaluation and avoid overfitting, the cleaned dataset was divided into training (70%) and testing (30%) subsets. The training set was used to fit the model, while the testing set was reserved for evaluating the model's performance on unseen data.

3.2 Model Set-up

Key Assumptions:

- Linearity: The relationship between the dependent variable and each predictor is linear.
- Independence: Observations are independent of one another.
- Homoscedasticity: Variance of residuals is constant across all levels of the independent variables.
- Normality of Residuals: The residuals are normally distributed.
- No Multicollinearity: Independent variables are not highly correlated with each other.

The model in this study is designed to predict weekly gross revenue using the following predictors:

- Average Ticket Price (`avg_ticket_price`): The average price of tickets sold during the week.
- Seats in the Theater (`seats_sold`): The seating capacity of the theater.
- Number of Performances (`performances`): The total number of performances held during the week.

- Holiday Week Indicator (`holiday_week`): Whether the show was performed during a holiday week (1 = holiday week, 0 = otherwise).
- Tony Award Month Indicator (`Tony_Award`) : Whether the show was performed during the Tony Award month (1 = Tony Award month, 0 = otherwise).
- Year (`year`): The year in which the show was performed, ranging from 2010 to 2019.

The model takes the form:

$$\begin{aligned} \text{weekly\_gross}_i = \beta_0 + \beta_1 \cdot \text{avg\_ticket\_price}_i + \beta_2 \cdot \text{seats\_in\_theatre}_i & \quad (1) \\ + \beta_3 \cdot \text{performances}_i + \beta_4 \cdot \text{holiday\_week}_i & \quad (2) \\ + \beta_5 \cdot \text{Tony\_Award}_i + \beta_6 \cdot \text{year}_i \epsilon_i & \quad (3) \\ \epsilon_i \sim \text{Normal}(0, \sigma^2) & \quad (4) \end{aligned}$$

Where:

$$\begin{aligned} \beta_0 \text{ is the intercept term} & \quad (5) \\ \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6 \text{ are the coefficients for each predictor} & \quad (6) \\ \sigma^2 \text{ is the variance of the error term} & \quad (7) \end{aligned}$$

Background details and diagnostics are included in Appendix C.

We run the model in R (R Core Team 2023) using the `stats` package from base R. The lm() function from this package was used to fit the multiple linear regression model.

3.3 Interpretation of Coefficients

- Intercept (β_0): Represents the predicted value of the weekly gross revenue when all predictors are 0. Since it is unrealistic for all predictors the intercept serves as a theoretical starting point for the model.
- Average Ticket Price (β_1): Measures the change in weekly gross revenue for each one-dollar increase in average ticket price, assuming all other predictors remain constant.
- Seats in the Theater (β_2): Represents the additional revenue generated by each additional seat available in the theater, assuming all other factors remain constant.
- Number of Performances (β_3): Indicates the increase in weekly gross revenue for each additional performance in the week for the show, controlling for other variables.
- Holiday Week (β_4): Quantifies the difference in weekly gross revenue between holiday weeks and non-holiday weeks, all else being equal.
- Tony Award Month (β_5): Captures the difference in weekly gross revenue between weeks during the Tony Award month and other months, assuming other predictors remain unchanged.
- Year (β_6): Reflects the annual trend, representing the average change in weekly gross revenue for each one-year increase, controlling for other variables.

3.4 Model justification

We expect a negative relationship between average ticket price and weekly gross revenue, as Broadway tickets are a form of entertainment and not a necessity. As ticket prices increase, demand is likely to decrease significantly because consumers may prioritize essential expenses over discretionary spending, especially for shows that are not well-known or during off-peak periods.

The seating capacity of the theater and the number of performances are also expected to have a positive relationship with weekly gross revenue, as larger venues or more performances each week can accommodate more audience members, resulting in higher total revenue.

For holiday weeks, we predict a positive effect on weekly gross revenue, as holidays often bring increased audience demand due to leisure time and tourism. Conversely, the effect of Tony Award months is less certain, as the difference in average weekly gross revenue during Tony Award months compared to other months is minimal.

Finally, the year variable is included to account for long-term trends, with an expected positive effect reflecting the steady growth of Broadway's gross revenue over the years.

4 Results

The results of the multiple linear regression model are presented in Figure 9, with more detailed information summarized in Appendix C. While some findings align with our expectations, others highlight notable deviations.

As anticipated, average ticket price, seating in theater, and number of performances exhibit significant positive relationships with weekly gross revenue. Specifically, increases in ticket prices and theater size lead to higher gross revenue, consistent with the notion that larger venues and premium ticketing contribute to revenue growth. Similarly, the significant positive impact of the number of performances aligns with our expectation that additional shows increase overall revenue by accommodating more audiences. Notably, the confidence intervals for these estimates are relatively narrow, suggesting that the relationships are estimated with high precision and are robust.

The findings for holiday weeks also meet initial expectations, showing a positive relationship with weekly gross revenue. This result suggests that increased demand during holiday periods effectively boosts ticket sales, contributing to higher overall revenue.

The results for Tony Award months diverge from our initial expectations. Contrary to the assumption that Tony Award months would boost revenue due to increased demand or publicity, the coefficients for these variables are negative, suggesting a slight decrease in weekly gross revenue during these periods. Additionally, the confidence interval for this variable is wide and crosses zero, indicating uncertainty about the direction and magnitude of the effect.

This suggests that the observed negative coefficient may not represent a consistent trend and warrants caution when interpreting the impact of Tony Award months on revenue.

Lastly, the year variable exhibits a small but positive effect, consistent with the expectation of a gradual upward trend in Broadway revenue over the years. The relatively narrow confidence interval for this variable provides strong support for the observed trend, likely reflecting the broader growth of the industry due to factors such as inflation or expanding audience engagement. Overall, the model confirms some hypotheses while highlighting discrepancies in the data, prompting further discussion and analysis in the next section.

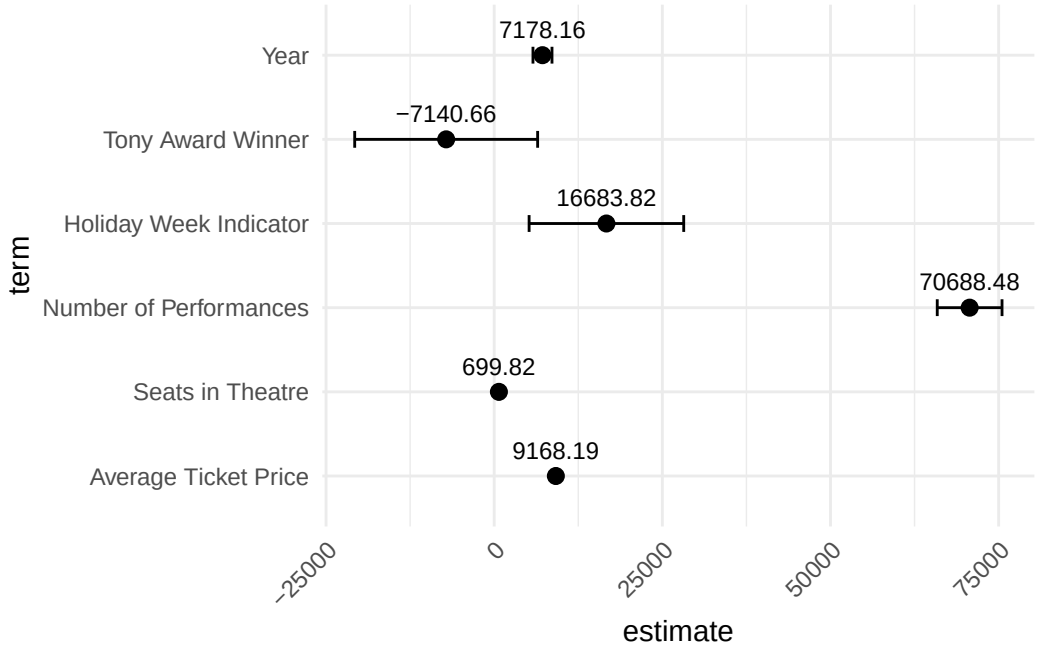

Figure 9: Coefficient plot illustrating the estimated effects of predictor variables on the response variable. Each point represents the estimated coefficient, and the horizontal lines denote the 95% confidence intervals for each coefficient. Positive values indicate an increasing effect on the response variable, while negative values indicate a decreasing effect.

5 Discussion

5.1 Key Findings

This paper presents a multiple linear regression analysis to examine the factors influencing weekly gross on Broadway. Key predictors, including average ticket price, seating capacity,

and number of performances, were found to significantly impact revenue, while holiday weeks and Tony Award months showed unexpected or non-significant effects. These findings highlight the complexity of audience behavior and the importance of operational decisions in optimizing revenue.

5.2 Prioritizing Per-Seat Value Over Expanding Seating Capacity

While increasing seating capacity and performance frequency predictably boosts revenue, the marginal gains from these factors differ significantly depending on audience demand and pricing strategy. The revenue per additional seat, estimated at $699.82, is substantially lower than the per-show contribution of $70,688.48 or the impact of a $1 increase in ticket price, which generates $9,168. This result reflects that simply increasing seating capacity, without addressing demand or ticket pricing, results in limited revenue gains. The impact of additional seats becomes most meaningful only when shows consistently sell out, underscoring the importance of strategies that drive full occupancy.

The balance between ticket price and capacity is particularly significant, as a smaller venue with higher ticket prices can outperform a larger theater with lower-priced tickets. For instance, Leslie (2004) demonstrates that optimizing ticket prices across seating sections can generate substantial revenue, often surpassing strategies focused solely on increasing seating capacity. Initiatives such as providing premium seating with added amenities, backstage tours, or bundled packages that include dining and entertainment can further enhance revenue while improving the audience experience. As emphasized by Ozhegova and Ozhegov (2018), understanding demand elasticity is essential, as premium seating typically exhibits lower price sensitivity, making it a key factor in strategic pricing decisions.

5.3 Maximizing Holiday Demand

The model results show that holiday weeks increase weekly gross revenue by an average of $16,684, reflecting a notable seasonal boost and aligning with expectations of higher audience availability during festive periods. Holiday weeks are a time when families often travel or plan special outings, and New York City, home to Broadway, becomes a global attraction with iconic events like the Rockefeller Center Christmas Tree Lighting, bustling holiday markets, and the New Year's Eve Ball Drop in Times Square. This festive atmosphere draws millions of tourists. However, the relatively modest increase of $16,684 per week suggests that Broadway may not be fully capitalizing on its unique position as a holiday destination. Studies on tourist behavior during holidays indicate that travelers are more inclined to seek unique and memorable experiences during festive periods (Pearce and Packer 2013), creating a prime opportunity for Broadway to enhance its gross revenue and establish itself as a must-visit holiday experience.

To better take advantage of its location and the influx of holiday visitors, Broadway could implement targeted strategies to make attending shows more convenient and appealing. Collaborating with hotels, travel agencies, and tourism boards to provide bundled holiday packages that include show tickets could position Broadway as an essential part of visitors' holiday itineraries. Moreover, enhancing the on-site experience with seasonal concessions, festive decor, and exclusive holiday merchandise could create a more engaging and memorable experience, attracting more visitors and driving additional revenue during this lucrative season.

5.4 Transforming the Tony Awards as an Industry-Wide Opportunity

Tony Award months show a negligible or slightly negative impact on overall revenue, with an estimated effect of $-7,140, challenging the assumption that the awards season boosts Broadway's financial performance. However, Boyle and Chiou (2009) finds that Tony Awards increase a production's revenue by 12% in the week immediately following the event. This discrepancy likely stems from media attention and ticket sales being concentrated on a few nominated or winning shows, while most productions see little benefit. Additionally, Reddy, Swaminathan, and Motley (1998) highlights the significant role of critics' reviews in driving Broadway success, further emphasizing the missed opportunity to use the publicity generated by the Tony Awards to benefit all productions more broadly.

Thus, Broadway could transform the Tony Awards into an industry-wide promotional platform by adopting strategies that engage a broader audience and distribute the benefits more equitably, extending beyond just the nominated and award-winning productions. Collaborating with media outlets to produce exclusive behind-the-scenes content, interviews with nominees, and live coverage of performances from non-nominated shows could help expand the appeal of the awards season. Additionally, launching a "Broadway Awards Week" featuring discounted tickets, special packages, or unique experiences such as backstage tours could encourage attendance across multiple productions, leveraging the awards buzz to increase overall industry revenue. Centralized marketing campaigns highlighting the diversity and vibrancy of Broadway's productions during this period could further enhance engagement and boost ticket sales across the ecosystem. By shifting the focus from individual winners to promoting the broader Broadway experience, the industry could transform the slight revenue decline during Tony Award months into a period of growth and shared success for all productions.

5.5 Weaknesses

The paper provides a detailed analysis of revenue drivers but focuses mainly on direct factors such as ticket prices, seating capacity, and performance frequency. The dataset does not include information on marketing efforts, show popularity, repeat attendance rates, or audience demographics. As a result, the analysis overlooks other potential variables that could provide a fuller understanding of revenue dynamics.

The model assumes linear relationships between predictors and weekly gross revenue. However, factors such as ticket price or performance frequency may demonstrate diminishing returns or threshold effects that are not reflected in the analysis. Furthermore, complex interactions between variables are not explored. For instance, changes in ticket price could influence consumers' willingness to buy, potentially altering demand and revenue in ways a linear model cannot fully capture.

Additionally, the model assumes uniform effects of predictors across all Broadway shows, which may oversimplify the analysis. Different types of productions (e.g., musicals, plays, blockbusters, or niche shows) likely have distinct revenue dynamics that are not adequately addressed by a single model. Moreover, the model does not account for external factors such as economic conditions, competitor pricing, or shifts in tourism trends, all of which could significantly influence Broadway revenue. These omissions limit the model's ability to fully explain variations in weekly gross revenue and highlight areas for future improvement.

Another weakness of this study is its exclusive focus on analyzing gross revenue without considering the associated costs of producing and running Broadway shows. While understanding revenue drivers is important, gross revenue alone does not provide a complete picture of financial performance. Factors such as production expenses, marketing costs, theater rental fees, and operational overhead significantly impact profitability but are not accounted for in this analysis. Ignoring costs limits the ability to assess the net financial health of Broadway productions and could lead to incomplete or misleading conclusions about the success of certain strategies or shows.

5.6 Next Steps

Incorporating additional variables, such as marketing expenditures, customer demographics, and customer experiences, could provide a deeper understanding of the factors influencing revenue. These information could be gathered in the future through detailed surveys aimed at collecting data on audience preferences, spending habits, and satisfaction levels. With this information, Broadway could develop more targeted marketing strategies tailored to different audience groups, allowing for customized promotions, pricing models, and programming that better address the unique needs of each segment, ultimately maximizing engagement and revenue.

Future analyses could examine non-linear relationships and interactions between variables, such as the combined influence of ticket prices and theater size on attendance or how holiday weeks impact revenue differently based on theater capacity, providing a deeper understanding of revenue patterns.

Additionally, conducting separate analyses for different types of productions, such as traditional shows versus new or experimental productions, would help identify the unique revenue drivers for each category. This segmentation would enable more precise strategy recommendations tailored to the specific needs and performance characteristics of different types of

Broadway shows. Incorporating these analyses with cost considerations could provide a more detailed approach to optimizing Broadway's economic sustainability.

A Appendix: Idealized Methodology and Survey

A.1 Objective and Overview

The purpose of this appendix is to outline an idealized methodology and survey design to supplement the current analysis of Broadway's weekly gross revenue. By incorporating additional survey data, this methodology aims to enhance the understanding of revenue drivers, audience behaviors, and market dynamics. The survey is designed to capture demographic, behavioral, and experiential data that complement the operational metrics already analyzed in the study.

A.2 Core Objectives

- To gather audience demographics (e.g., age, income, location) to understand their influence on ticket purchases and attendance patterns. This information can be integrated into the model to enhance its predictive accuracy.
- To identify purchasing behaviors, such as repeat attendance, pricing sensitivity, and group sales, providing a foundation for refining marketing and pricing strategies.
- To assess audience motivations and satisfaction, focusing on factors driving show selection and overall theater experience, providing guidance on how to improve Broadway's organizational and operational strategies.

A.3 Sampling Strategy

Our sampling strategy utilizes ** simple random sampling** to ensure unbiased representation of Broadway audiences. This approach randomly selects participants from the overall population of Broadway ticket buyers and attendees, minimizing potential biases and ensuring the sample reflects the diversity of the audience.

Random sampling provides a robust foundation for generalizable analysis, capturing a wide range of perspectives and behaviors without targeting specific subgroups (Noor, Tajik, and Golzar 2022). This method aims to generate accurate and thorough findings on the factors affecting Broadway's attendance and revenue dynamics.

A minimum sample size of 1,000 responses per quarter is proposed to ensure statistical robustness across strata and account for seasonality.

A.4 Recruitment Strategy

Participants will be recruited through a combination of:

- **In-Theater Surveys**: Surveys distributed at participating Broadway theaters, targeting audiences immediately after the performance for real-time feedback.
- **Online Platforms**: Email invitations to ticket buyers, links embedded in ticket purchase confirmations, and social media outreach via official Broadway channels.
- **Tourism Partnerships**: Collaborations with NYC tourism boards, hotels, and travel agencies to engage out-of-town visitors.
- **Incentives**: Providing discounts on future ticket purchases or exclusive merchandise as incentives for survey participation.

A.5 Data Validation and Quality Assurance

- **Data Triangulation**: Responses will be cross-referenced with transactional data from ticket sales to validate self-reported behaviors.
- **Pretesting**: Pilot surveys will be conducted to test question clarity, response rates, and overall survey design effectiveness.
- **Anonymity and Privacy**: Personal data will be anonymized to encourage honest responses and comply with privacy regulations.

A.6 Survey Design Considerations

The survey is designed to balance thoroughness and brevity, aiming to achieve high response rates while collecting useful data. Important design elements include:

- **Question Types**: A combination of multiple-choice, Likert scale, and open-ended questions balances structured data collection with qualitative input.
- **Clarity**: Simple, jargon-free language to ensure accessibility for diverse audiences.

Focus Areas:

- Demographics: Age, income, location, gender, and education level.
- Attendance Behavior: Frequency of attendance, motivations, and purchasing habits.
- Experience Metrics: Satisfaction levels, likelihood of future attendance, and feedback on show quality.
- Pricing Sensitivity: Perceptions of ticket affordability and willingness to pay for premium experiences.

A.6.1 Broadway Audience Survey Form

Introduction:

Thank you for taking the time to participate in this survey. Your feedback will help us understand Broadway audiences better and improve the overall experience.

Please Note:

- All responses will be kept strictly confidential.
- Your participation is entirely voluntary.
- We kindly request that you answer all questions honestly and to the best of your knowledge.
- The survey is estimated to take approximately 10 minutes to complete.
- If you have any inquiries or concerns regarding this survey, please don't hesitate to contact the research team at **isabella.zhou@mail.utoronto.ca**.

Your contribution to this study is greatly appreciated! Each participant will be entered into a lottery with a chance to win a **$100 gift card redeemable for any Broadway show of your choice**.

Section 1: About You

What is your age ?

- Under 18
- 18–24
- 25–34
- 35–49
- 50–64
- 65 or older

What sex were you assigned at birth, on your original birth certificate?

- Female
- Male

How do you currently describe yourself (mark all that apply)?

- Female
- Male
- Transgender
- I use a different term [free-text]

Where do you currently live?

- Manhattan

- Brooklyn
- Queens
- Bronx
- Staten Island
- Other New York City Borough: _____
- Outside NYC (please specify): _____

What is your annual household income (before taxes)?

- Less than $25,000
- $25,000–$49,999
- $50,000–$74,999
- $75,000–$99,999
- $100,000–$149,999
- $150,000–$249,999
- $250,000 or more

How far in advance do you typically purchase Broadway tickets?

- Less than a week
- 1–2 weeks
- 3–4 weeks
- 1–2 months
- 3 months or more

Section 2: Broadway Attendance and Ticket Purchases

How many Broadway shows have you attended in the past 12 months?

- None
- 1–2
- 3–5
- 6–10
- 11 or more

What motivates you to attend Broadway shows? (Check all that apply)

- Specific performers
- Storyline or theme of the show
- Award recognition (e.g., Tony Awards)
- Recommendations from friends or family
- Discounts or promotions
- Other (please specify): _____

How much did you pay for your most recent Broadway ticket (including fees)?

- Less than $99
- $100–$149
- $150–$199
- $200–$249
- $250–$299
- $300–$349
- $350 or more

How do you usually purchase tickets?

- Online via website or app
- At the theater box office
- Through group sales or subscriptions
- Other (please specify): _____

Section 3: Experience and Preferences

How satisfied were you with the last Broadway show you attended?

- Very satisfied
- Satisfied
- Neutral
- Dissatisfied
- Very dissatisfied

What factors most influenced your satisfaction? (Check all that apply)

- Quality of performance
- Cast or performers
- Theater amenities (e.g., seating, concessions)
- Ticket price relative to experience
- Other (please specify): _____

What would improve your Broadway experience? (Check all that apply)

- Lower ticket prices
- Easier ticket access
- More diverse productions
- Enhanced theater amenities
- Other (please specify): _____

When do you typically attend Broadway shows? (Check all that apply)

- Weekdays
- Weekends

- Holiday weeks (e.g., Thanksgiving, Christmas, New Year's Week, (please specify): _____
- Tony Award months (typically June)

Where do you typically hear about Broadway shows? (Check all that apply)

- Social Media (e.g., Instagram, Facebook, Twitter)
- Online ads or newsletters
- Word of mouth
- Print media (e.g., newspapers, magazines)
- Other (please specify): _____

Section 4: Future Engagement

How likely are you to attend another Broadway show in the next 6 months?

- Very likely
- Somewhat likely
- Neutral
- Somewhat unlikely
- Very unlikely

What factors might prevent you from attending more Broadway shows?

- High ticket prices
- Lack of time
- Lack of interest in available productions
- Other (please specify): _____

What additional suggestions do you have to improve the Broadway experience?

End Message:

Thank you for participating! Your feedback will play a vital role in enhancing the Broadway experience for future audiences. If you have any questions about this survey or would like to be informed about the results, please contact **isabella.zhou@mail.utoronto.ca**

The survey will be implemented via **Google Forms**, which provids a cost-effective platform for data collection. You can access the survey at the following link: Google Form Survey

B Appendix: Additional Broadway Surveys and Data

In addition to the dataset used for this study, Broadway conducts several key surveys to collect detailed information on various aspects of Broadway theater. However, the information and datasets collected through these surveys are not publicly accessible for analysis. If this data were made available, merging it with the current dataset could provide a more robust model for gross analysis. These surveys include **The Demographics of the Broadway Audience** (League 2023), designed to analyze the composition and behavior of Broadway theatergoers.

- **Sampling Method**: The survey is typically conducted annually, with performances selected quarterly to ensure a representative sample of Broadway's diverse productions. This approach captures seasonal variations in the audience and provides a balanced dataset
- **Recruitment Method**:
  - In-Person Distribution: Questionnaires are handed out to audience members during selected performances.
  - Online Access: Audiences can complete the survey online via a QR code provided at performances or through a link sent post-event.
  - Wi-Fi Login Prompt: Patrons connecting to a theatre's Wi-Fi are also invited to participate in the survey. These varied recruitment strategies aim to ensure a high response rate and broad representation of attendees.

Broadway also records and analyzes additional information, such as Broadway's Economic Contribution to New York City, the Audience for Touring Broadway, and the Economic Impact of Touring Broadway (League 2024a). These surveys and studies provide a broader understanding of Broadway's influence and reach, presenting a more complete view of its cultural and economic significance.

C Appendix: Model details

C.1 Diagnostics

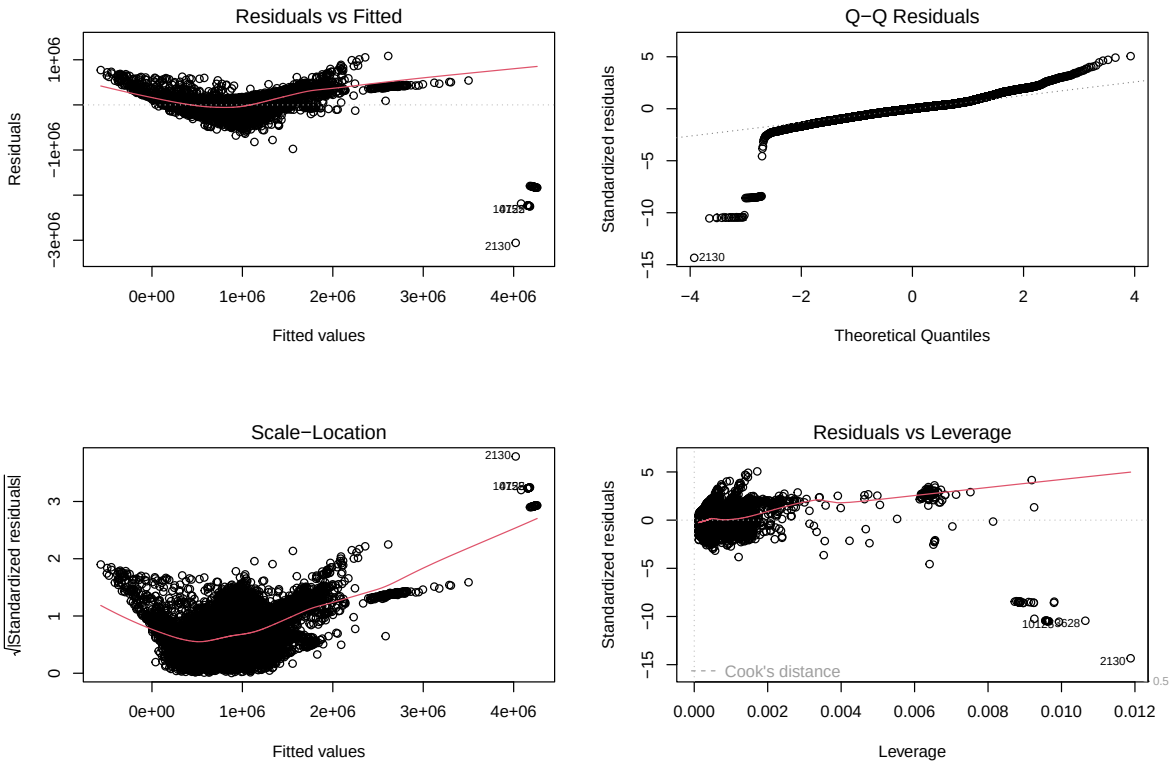

Figure 10: Diagnostic Plots for Regression Model Residuals. Residuals vs Fitted values, Q-Q Plot of residuals, Scale-Location plot, and Residuals vs Leverage, used to assess the assumptions of the multiple linear regression model.

As shown in Figure 10, these diagnostic plots indicate potential violations of the linear regression assumptions, suggesting the need for model refinements or transformations.

The curved pattern in the Residuals vs Fitted plot points to possible non-linearity, while the tails in the Q-Q plot highlight deviations from normality. Additionally, the trend observed in the Scale-Location plot suggests heteroscedasticity. Furthermore, the Residuals vs Leverage plot shows points with high leverage and large residuals, indicating potential outliers that may disproportionately influence the model.

C.2 Mean Squared Error and Mean Absolute Error on Test Data

Table 1: Calculate Mean Squared Error (MSE) on Test Data

Table 1: Model Performance Metrics

Metric	Value
Mean Squared Error (MSE)	50,906,242,649
Mean Absolute Error (MAE)	142,488
R-squared	0.83

From Table 1, MSE measures the average squared prediction error. MAE represents the average prediction error magnitude. R-squared indicates that the model explains 83% of the variance in weekly gross revenue.

C.3 Multicollinearity Check on the Training Data

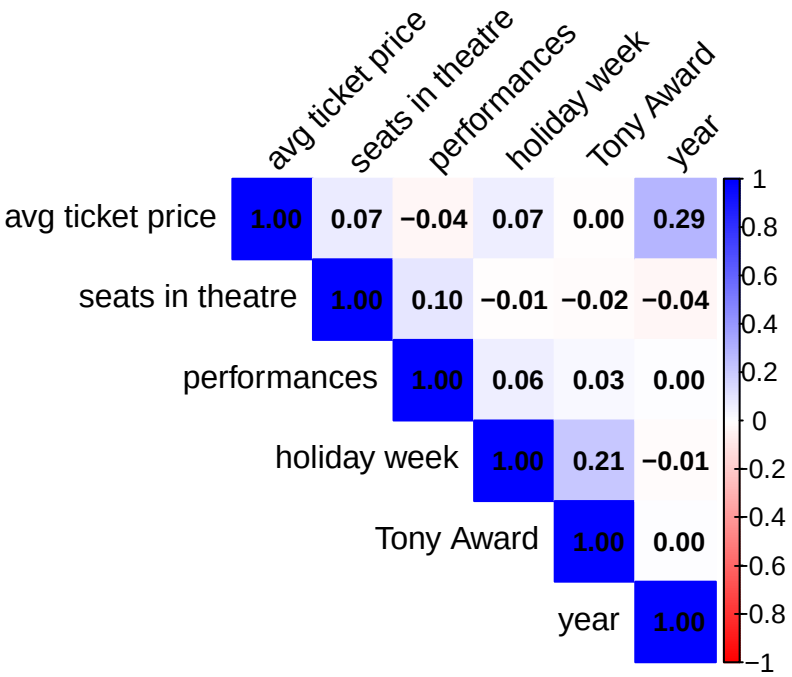

Figure 11: Correlation Matrix between variables

In Figure 11, a correlation check is performed. The matrix confirms the assumption of low multicollinearity among predictors, supporting the reliability of coefficient estimates in the regression model.

Table 2: Regression Results for Weekly Gross. This table displays the coefficients and their confidence intervalsfor the multiple linear regression model used to analyze the factors influencing weekly gross for Broadway.

	Multiple Linear Regression
Average Ticket Price	9168.189
	[9073.826, 9262.551]
Seats in Theatre	699.822
	[688.316, 711.327]
Number of Performances	70 688.482
	[65 872.308, 75 504.656]
Holiday Week Indicator	16 683.823
	[5189.541, 28 178.106]
Tony Award Winner	−7140.660
	[−20 739.270, 6457.949]
Year	7178.159
	[5764.566, 8591.752]
Num.Obs.	11 615
R2	0.839
R2 Adj.	0.839
AIC	318 149.9
BIC	318 208.7
Log.Lik.	−159 066.927
RMSE	214 490.04

C.4 Model Summary

D Additional Visulization

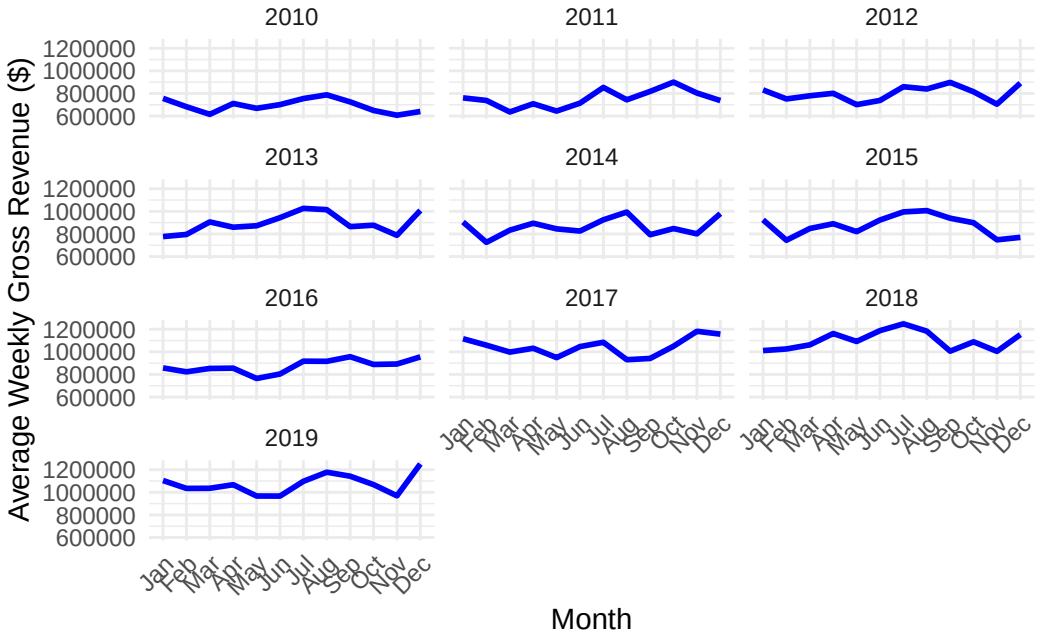

Figure 12: Average Weekly Gross Revenue Trends Across Months, spanning from 2010 to 2019. The visualization highlights both year-to-year growth and recurring seasonal patterns in Broadway's revenue performance. Notable peaks in revenue are observed around December, coinciding with the holiday season, and slight increases can also be seen during the summer months, potentially due to higher tourist activity.

E Acknowledgements

Thanks to Open AI and ChatGPT 4o is used to write the paper.

This study utilizes R packages (R Core Team 2023) for data cleaning and analysis, incorporating libraries from tidyverse (Wickham et al. 2019), ggplot2 (Wickham 2016), knitr (Xie 2024), arrow (Richardson et al. 2024), here (Müller 2020) and scales (Wickham, Pedersen, and Seidel 2023).

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Boyle, Melissa, and Lesley Chiou. 2009. "Broadway Productions and the Value of a Tony Award." *Journal of Cultural Economics* 33: 49–68.

Cookson, Thomas A. 2020. "Broadway Grosses Dataset." https://github.com/tacookson/data/tree/master/broadway-grosses.

Hodge, Matthew. 2020. "21st-Century Broadway Musicals and the 'Best Musical'tony Award: Trends and Impact." In *Arts*, 9:58. 2. MDPI.

(IBDB), Internet Broadway Database. 2024. "About - Internet Broadway Database." https://www.ibdb.com/About.

Labor, New York State Department of. 2021. "Broadway Theaters: An Economic Engine for New York." New York State Department of Labor. https://dol.ny.gov/system/files/documents/2021/03/broadway-theaters-an-economic-engine-for-new-york.pdf.

League, The Broadway. 2023. "The Demographics of the Broadway Audience: 2022–2023 Season." https://lede-admin.nyc.streetsblog.org/wp-content/uploads/sites/48/2024/07/Broadway-NY-Demographic-2022-2023.pdf.

———. 2024a. "Research Reports - the Broadway League." https://www.broadwayleague.com/research/research-reports/.

———. 2024b. "Statistics - Broadway in NYC." https://www.broadwayleague.com/research/statistics-broadway-nyc/.

Leslie, Phillip. 2004. "Price Discrimination in Broadway Theater." *RAND Journal of Economics*, 520–41.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Noor, Shagofah, Omid Tajik, and Jawad Golzar. 2022. "Simple Random Sampling." *International Journal of Education & Language Studies* 1 (2): 78–82.

Ozhegova, Alina, and Evgeniy M Ozhegov. 2018. "Heterogeneity in Demand for Performances and Seats in the Theatre." *Journal of Revenue and Pricing Management* 17: 131–45.

Pearce, Philip L, and Jan Packer. 2013. "Minds on the Move: New Links from Psychology to Tourism." *Annals of Tourism Research* 40: 386–411.

Playbill. 2024. "Grosses List." https://playbill.com/grosses.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reddy, Srinivas K, Vanitha Swaminathan, and Carol M Motley. 1998. "Exploring the Determinants of Broadway Show Success." *Journal of Marketing Research* 35 (3): 370–83.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Stanley, Richard T. 2013. *America's Favorite Holidays*. iUniverse.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Thomas Lin Pedersen, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.