Eleanor Miranda

Data 102-CO1

<div align="center">**Broadway Revenue and Attendance Analysis**</div>

Abstract

This paper explores how Broadway show characteristics relate to revenue and attendance. Interactive Excel and Tableau dashboards were created to analyze a dataset of 31,296 weekly records from 1990 to 2016. The results of this analysis indicate that musicals have a higher weekly average gross revenue and attendance than plays. Further, there is a strong positive correlation between attendance and gross revenue. Overall, it can be assumed that show type is a main predictor of success in Broadway shows.

Introduction

Broadway is a multi-billion-dollar industry that provides a significant impact on New York City's economy. The economic success of Broadway shows depends on attendance and ticket prices. There has been an increase in gross revenue over the past few decades due to inflation and high demand. There is a high demand since there are only a limited number of seats in a theatre and the fact that the shows can only be seen in person. While there is an increase in success, it is also found that about 80% of Broadway productions fail to earn back their investments in creating the shows, since ticket prices tend to be higher than most people are willing to pay (Dent). Thus, there is a need for an analysis of what characteristics of a Broadway show have a positive impact on revenue and attendance. Overall, it is hypothesized that musicals tend to have a higher revenue than plays due to higher attendance at musicals rather than plays. Given these factors, this investigation will aim to answer the question: How do Broadway show characteristics, such as show, show type, and theatre relate to their average weekly gross revenue and attendance?

Methodology

The dataset used for this analysis is from Kaggle and includes 31,296 observations from the years 1990-2016. The full dataset was used without sampling to ensure a comprehensive analysis. Seven of the twelve variables were used, including both quantitative and categorical variables. The quantitative variables include year, weekly gross, weekly attendance, and weekly capacity. The categorical variables include show name, show type, and theatre. When incorporating the weekly gross, weekly attendance, and weekly capacity variables in the analysis, the averages of these values was taken rather than the sums for a fair comparison. Otherwise, the sums would impact the data since there are different run times/lengths for different shows. The data was analyzed using both Excel and Tableau, and interactive dashboards were created using each program. In Excel, a histogram and stacked bar graph were created to show the distribution of the average weekly gross and the average capacity by show type and theatre. The visuals created in Tableau include a scatter plot, line graphs, and a bar graph to compare average weekly gross and average weekly attendance by show types. Both dashboards include filters, such as show type and year, as well as numerical statistics. In addition, a table of descriptive statistics was created in Excel to further analyze the data.
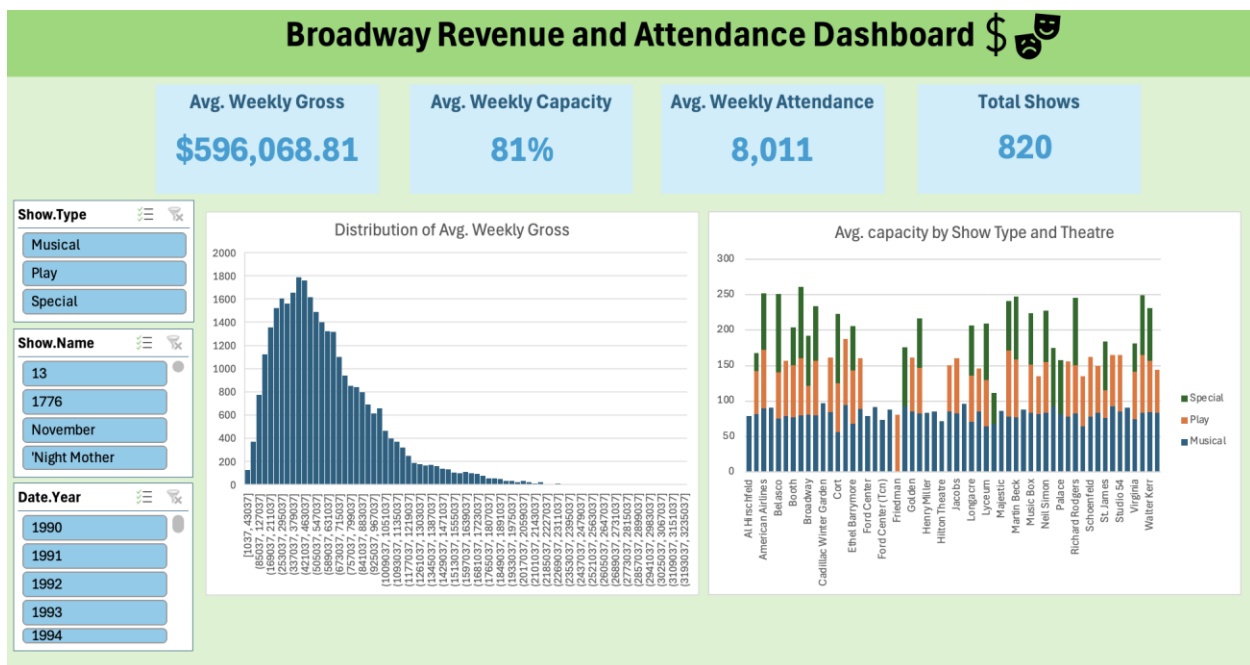
Data Analysis and Results

To begin the analysis, a table of descriptive statistics (see below) was created to evaluate the three main quantitative variables: weekly gross, weekly attendance, and weekly capacity. The table includes the mean, median, standard deviation, and count for each variable. For all three variables, the count is the entire sample size of the dataset. The Broadway shows in this analysis earned a mean weekly gross of $596,069, and the median weekly gross is $514,520. Since the median gross is less than the mean gross, the data is right-skewed, meaning that there are only a small number of shows that earn significantly higher amounts than others. The weekly gross has a standard deviation of $380,980, meaning that there is a high variability when it comes to weekly revenue. The average attendance of these shows is about 8,011 people per week, and the median of the weekly attendance is 7,802. Since the median is close to the mean of weekly attendance, there is a balanced distribution when it comes to the size of the audiences. The standard deviation for weekly attendance is 3,190, meaning that there is a variation in audience size. This could be due to the different sizes of the theatres. Lastly, the mean of the percentage of
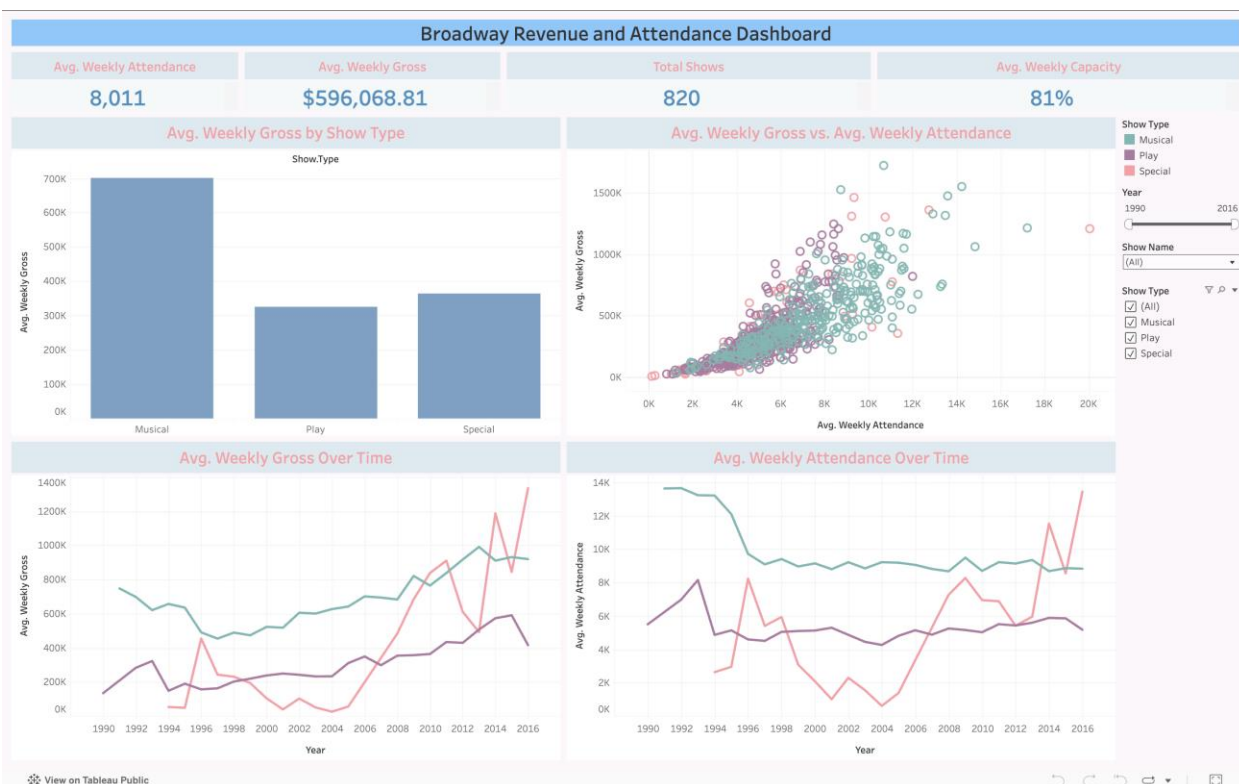
the weekly capacity that was utilized is 81%, while the median is 83%. The standard deviation is 19%, indicating that there is not much variation in the percentage of each theatre being filled each week.

Variable	Mean	Median	Standard Deviation	Count
Weekly Gross	596069	514520	380980	31296
Attendance	8011	7802	3190	31296
Capacity	81	83	19	31296

Next, an Excel dashboard (see below) was created to visualize some of these statistics. The dashboard includes a histogram of the distribution of the average weekly gross revenues. The variables used in these charts are weekly gross, weekly capacity, show type, and theatre. The histogram is right skewed, which indicates that most shows have lower average weekly grosses, while there are a few outliers that have higher average weekly revenues. A stacked bar graph is also included, which shows the average weekly capacity by show type and theatre. When comparing only musicals and plays, the graph indicates that the show type and theatre that share the highest average weekly capacity are musicals at the Cadillac Winter Garden Theatre. Similarly, the show type and theatre that share the lowest average weekly capacity, are plays at the St. James Theatre.

Lastly, an interactive Tableau dashboard (see below) was created to further visualize the data. The Tableau dashboard includes a bar graph, scatter plot, and two line graphs. The variables used in this dashboard are show type, show name, year, weekly gross, weekly attendance, and weekly capacity. The bar graph compares the average weekly gross and the show type. According to the graph, the type of show that earned the highest average weekly gross is musicals. The scatter plot compares the average weekly gross and attendance by show type. The plot indicates that average weekly gross increases with average weekly attendance, and the type of show that has both the highest average weekly gross and attendance is musicals. The two line graphs show the average weekly gross and the average weekly attendance over time. Confirming the earlier analyses, musicals consistently have significantly higher average weekly attendance and gross revenues than plays throughout the years 1990-2016. Both the Tableau and Excel dashboards also include the overall average weekly attendance (8,011), average weekly gross ($596,068.81), average weekly capacity (81%), and total number of distinct shows (820).

Conclusion

Overall, from the years 1990-2016, musicals consistently dominate over plays and special performances when it comes to Broadway revenue, attendance, and capacity. The results of this analysis indicate a strong, direct, positive relationship between average weekly attendance and revenue. Thus, musicals bring in both the highest revenue and the number of attendees on Broadway. Further studies may include analyzing the effects of the COVID-19 pandemic on Broadway attendance and revenue.

Appendix

Dataset

Mexwell. (2023). *Broadway Shows* [Dataset]. Kaggle.

https://www.kaggle.com/datasets/mexwell/broadway-shows?resource=download.

References

Dent, M. (2021, November 21). *The economics of Broadway shows*. The Hustle.

https://thehustle.co/the-economics-of-broadway-shows.

Descriptive Statistics Table

Variable	Mean	Median	Standard Deviation	Count
Weekly Gross	596069	514520	380980	31296
Attendance	8011	7802	3190	31296
Capacity	81	83	19	31296

Excel Dashboard

https://studentbridgew-my.sharepoint.com/:x:/g/personal/e2miranda_student_bridgew_edu/IQCR08YMKuMLRLgBu0f7bDmfARmPArbaenskVAVgKlI9jSk?e=dVEIRk

Tableau Dashboard

https://public.tableau.com/views/final_17660891793480/Dashboard1?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link