Movie Theater Insights:

Applying Statistical Analysis to Understand Weekly Movie Ticket Pricing Patterns

Written by:

Jamian Kortekaas

Contents in Brief

Introduction

Project Overview

The purpose of this project is to explore key insights and weekly trends at the Mission Grove Galaxy Theatre located in Riverside, CA. The primary objective is to examine and analyze the relationship between weekly movie screening insights and ticket pricing trends. On a technical level, this project applies academic concepts such as Descriptive Statistics and Econometrics, while also utilizing appropriate Business Applications. The tools used in this project include Excel, Data Analysis using RStudio, and Econometric Techniques.

Hypothesis

This analysis hypothesizes that time is the strongest predictor of final ticket prices. Time in this project is represented by three distinct variables which measure time in different units. These variables are grouped into their own Time-Related category which we will examine closely later in this project and are defined as follows:

- Screening Daypart → Whether the movie ticket was purchased as a 'Matinee' or 'Evening' ticket.
- Time Periods → Divides "Screening Daypart" into intervals based on the time of day the movie screening occurs.
- Day of the Week → What specific day the movie ticket was purchased.

DATA DESCRIPTION

Data Collection

Galaxy Theatres is a movie theater chain based in the Western United States and there are currently four movie theaters operating in the state of California. For this project, I chose the movie theater located in Riverside, a city with a robust and diverse movie theater market, due to its local proximity and competitiveness.

Data for this project was extracted directly from the company's website on the showtimes page and only contains information within a specific time frame for streamlining purposes. In this project, I chose to record various movie screening characteristics for one week, or to be specific, between December 9 and December 16 and information pertaining to this time frame can be found within the gray element at the top of the webpage. For each movie screening that appears on the webpage, there is a thumbnail, a title, an age rating, a screen time, bold text that may indicate seating arrangement or format, and details that may indicate other movie screening attributes. Each movie screening has several different tabs that indicate the screening showtime which may be grouped based off the movie format and a black progress indicator that specifies the percentage of auditorium seats purchased (this indicator gave unreliable numbers, so I chose to leave this out of the final document). Clicking on one of these tabs will open the movie theater's dedicated reservation system.

Galaxy Theatres Showtimes Page
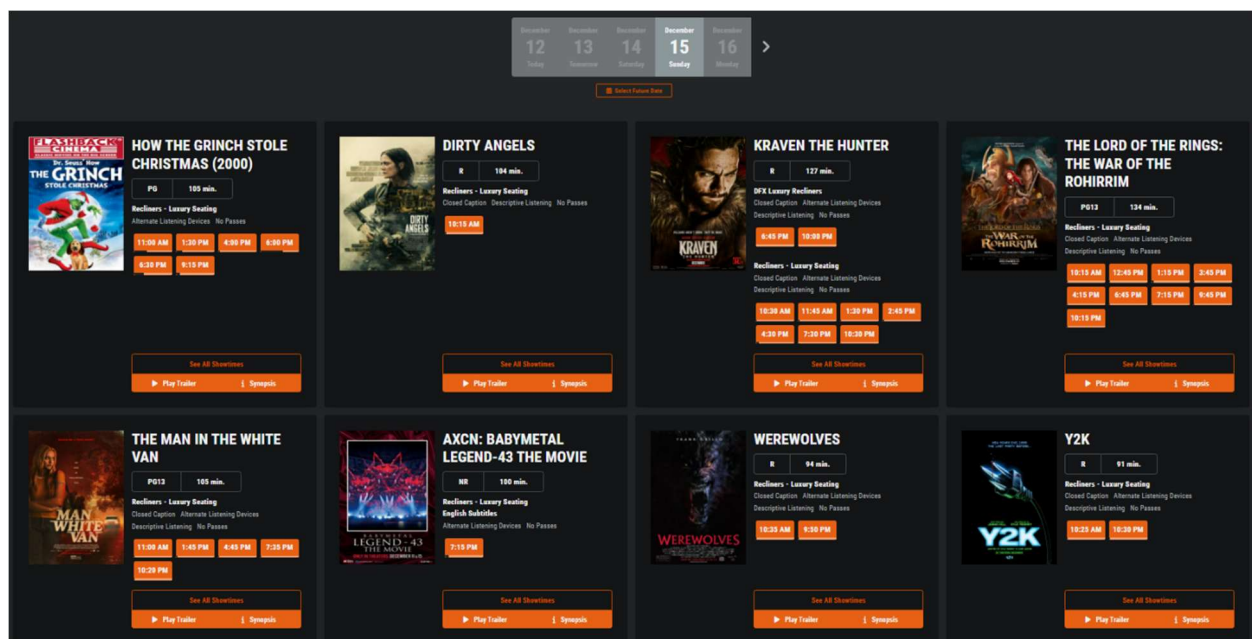
Not all the information I extracted from this website was accounted for in its entirety and so the usage of such data was at my discretion. The attributes for each movie screening entry I recorded were selected with the intention of capturing key characteristics that might influence ticket prices. These attributes include but are not limited to the title, age rating, format, movie type, screening date, and showtime.

Additional features for each entry can be found by accessing the dedicated reservation system. Within this webpage, users can select specific seats they wish to reserve as well as the type of ticket before finalizing their movie ticket purchase. The ticket type and the prices associated with each type were documented alongside the other attributes.

Galaxy Theatres Reservation System (Moana 2 - Monday 2, Dec. 16)

Data Anomalies

There were a few instances during the process of data collection in which unforeseen obstacles prevented the correct and accurate recording of either the Base or Final Ticket Price. These instances were recorded and addressed by designating specific entries within the 'Discounted Ticket' variable accordingly. Accounting for these discrepancies impacts not only multiple data visualizations but also the regression analysis in which this project is based on.

Elimination of these anomalies was deliberated but was ultimately deemed unnecessary as the inclusion of such data provided additional insight into the movie theater's pricing structure and preserved the integrity of the project's analysis. A list of these anomalies can be found below.

Data Anomaly Table

Anomalies	Description
Flashback Cinema Promo Coupon?	Frequent visits to website indicated there was a bug that failed to display promo coupon multiple times.
3D Moana 2 Charged Normal Price?	Thu_Moana23D_04 screening is charged regular matinee price instead of 3D matinee price.

Variable Classification

Full Variable Table

The full and complete Variable Table as indicated on the *Excel Document* can be found below.

Variable Name	Data Type	Description
Screening ID	Identifier	Unique ID assigned for each screening event.
Movie Short Form ID	Identifier	Abbreviated title for each movie screening.
Movie Screening Title	Identifier	Full Title for each movie screening.
Screening Daypart	Categorical	Segment of time within a day in which the screening event takes place.
Time Periods	Categorical	Divides Screening Daypart into finer intervals of time.
Screening Showtime	Continuous	Start time for each screening event (24-hour time format).
Genre	Categorical	Category of each movie based on shared subject matter.
Movie Type	Categorical	Specifies whether the movie has audio/subtitle features or not.
Day of the Week	Categorical	The Business Day in which the screening event takes place.
Format	Categorical	Indicates whether the movie screening is in a standard or premium format.
Ticket Type	Categorical	Segments of ticket prices that are based off of shared consumer characteristics.
Base Ticket Price	Continuous	The initial ticket price before any additional fees are added.
Online Fee	Continuous	Additional fee that's incurred when purchasing tickets online.
Final Ticket Price	Continuous	The total ticket price after any additional fees.
Purchase Method	Categorical	Indicates whether the movie ticket was purchased in-person or online.
Discounted Ticket	Binary	Indicates whether any discount was applied to the ticket or not.
Special Event Pricing	Binary	Indicates whether the screening event qualified for special event pricing or not.
Special Program	Categorical	Specifies which promotional campaign (if any) each movie was a part of.
Age Rating	Categorical	The movie's audience suitability label.

Omitted Variables

The following table showcases other variables that failed to appear on the final project and were subsequently omitted from the Full Variable Table for various reasons.

Omitted Variables	Description	Reasoning
Membership Pricing	Whether the ticket price was based on membership or loyalty program.	No Variability/Incomplete Data.
Occupancy Rate	Percentage of tickets sold for each movie screening.	Unreliable Metric.
Refundable Ticket	Whether the ticket is refundable.	Incomplete Data.
Advance Purchase	Whether the ticket was purchased in advance.	Incomplete Data.
Seat Type	Type of seating arrangements.	No Variability in Seating Type.

Variable Categorization

Variables from the Full Variable Table were further divided and grouped into categories based off shared characteristics for efficiency. The four primary categories can be found below:

A. Time-Related Characteristics
   - Screening Daypart → Matinee, Evening
   - Time Periods → Morning, Midday, Dusk, Night
   - Day of the Week → Monday 1, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday, Monday 2
B. Movie Characteristics
   - Genre → Action, Animation, Anime, Comedy, Concert, Documentary, Drama, Fantasy, Horror, Thriller
   - Format → Standard, 3D, DPX
   - Special Program → AXCN, Fathom Events, Flashback Cinema, None, Unique, Studio Ghibli
   - Age Rating → G, PG, PG13, R, NR, TBD
C. Consumer Characteristics
   - Ticket Type → Adult, Child, Senior, Admit One, FC Promo
   - Purchase Method → In-Person, Online
D. Continuous Variables
   - Base Ticket Price → Before Additional Fees/Effects
   - Final Ticket Price → After Additional Fees/Effects

Key Statistics

Ticket Pricing Summary

A statistics table for the Base Ticket Price is available below. This table can also be found on the associated *R Markdown Document* under the object: '*Base_TP_Stats*'.

| Min	Q1	Median	Mean	Q3	Max
0	12	12.5	13.49	15.5	20

A statistical table for the Final Ticket Price is available below. This table can also be found on the associated *R Markdown Document* under the object: *'Final_TP_Stats'.*

Min	Q1	Median	Mean	Q3	Max
0	12.5	14.49	14.74	16	22.49

Bar Charts Interpretation

Group 1: Time- Related Variables

Relationship Between Screening Dayparts and the Median Final Ticket Price
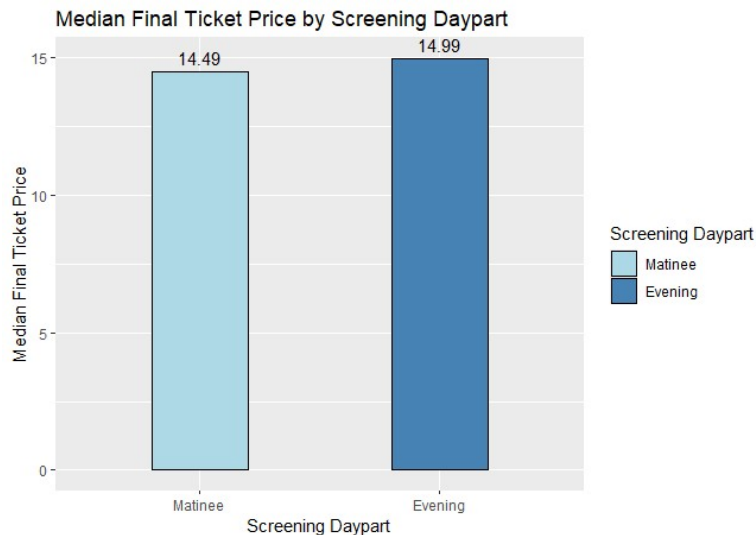

The Median Final Ticket Price for Matinee Tickets is $14.49, while the Median Final Ticket Price for Evening Tickets is $14.99. There is a $0.50 difference between both prices, suggesting that Matinee Ticket Prices tend to be less expensive and more desirable for the consumer than Evening Ticket Prices. Matinee Tickets can be purchased before 4 PM while Evening Tickets can be purchased after 4 PM.

Relationship Between Time Periods and the Median Final Ticket Price

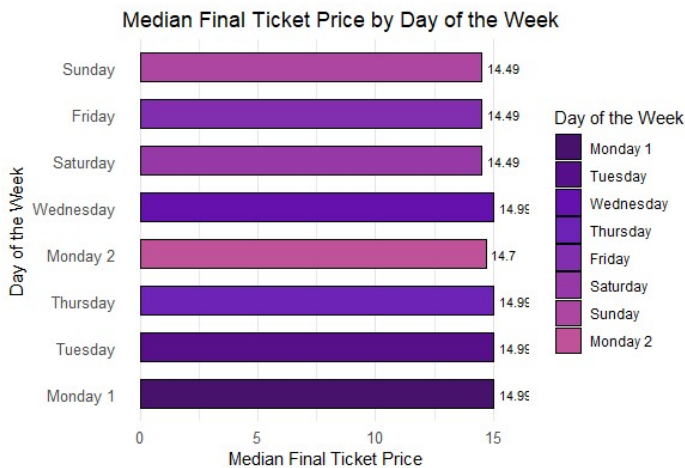The Median Final Ticket Price for both Morning and Midday Tickets is $14.49, while the Median Final Ticket Price for both Dusk and Night Tickets is $14.99. The lack of price variation between either Morning and Midday or Dusk and Night suggests that the 'Time Periods' variable isn't as strong of a predictor of Final Ticket Prices as I previously anticipated.

Median Final Ticket Price by Time Period
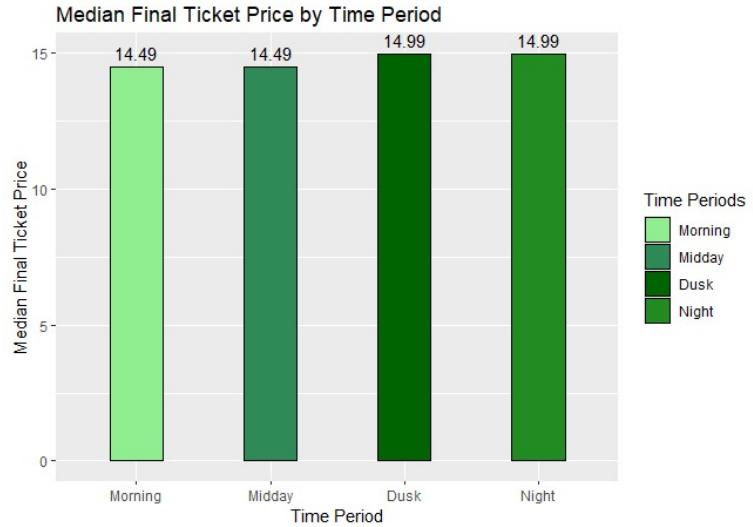
As previously stated in the Final Variable Table, the 'Time Periods' variable is a modified version of the 'Screening Daypart' variable whose sole intention is to identify the relationship between time and the Final Ticket Price by dividing prices based on a targeted time range. This variable was divided as follows:

Time Period	Time Range
Morning	10:00-11:59
Midday	12:00-15:59
Dusk	16:00-18:59
Night	19:00-22:59

Relationship Between Days of the Week and the Median Final Ticket Price

Median Final Ticket Price by Day of the Week

The Median Final Ticket Price from Monday 1 to Thursday held a consistent price of $14.99, while the Median Final Ticket Price from Friday to Sunday held a consistent price of $14.49. Monday 2 had the smallest Median Final Ticket Price at $14.70. The trend of ticket prices decreasing throughout the week suggests there may be a pricing strategy taking place based on consumer demand during the weekend (Fri-Sun) compared to the rest of the Business Week.

Group 2: Movie Characteristics

Relationship Between Genre and the Median Final Ticket Price

Genres such as Concert, Documentary, Animation, and Comedy exhibit the greatest price variation compared to other genres on the bar chart. This is likely because these genres coincide with special screening events and are subject to different pricing strategies.
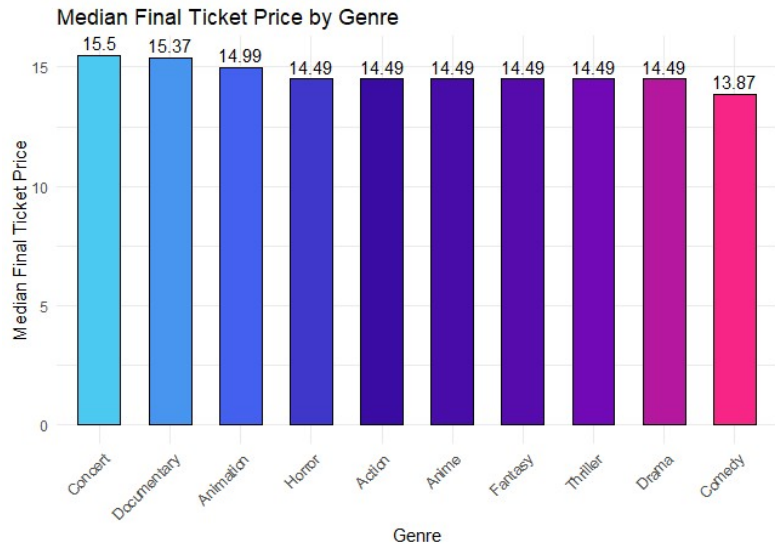
Relationship Between Format and the Median Final Ticket Price
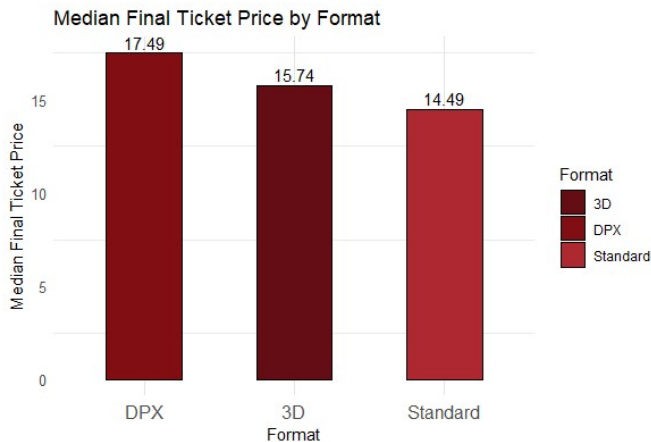
The 'Format' variable exhibits strong price variation among all types, and it suggests that the Format in which the customer watches movie screenings have a considerable impact on the Final Ticket Price. This movie theater location offers DPX, 3D, and Standard Formats with DPX movie screenings having the largest Median Final Ticket Price at $17.49.

Relationship Between Special Program and the Median Final Ticket Price

There is a considerable amount of price variation within the 'Special Program' variable, and it may suggest that the type of Special Program that a movie screening belongs to has an impact on the Final Ticket Price. Fathom Events has the highest Median Final Ticket Price at $18.24, while Flashback Cinema has the lowest at $5 (the low-ticket price might be due to an anomaly in the data).

Median Final Ticket Price by Special Program

Special Program
Flashback Cinema
None
AXCN
Studio Ghibli
Unique
Fathom Events

Fathom Events: 18.24
Unique: 17.49
Studio Ghibli: 16.24
None: 14.49
AXCN: 14.74
Flashback Cinema: 5

Median Final Ticket Price

Relationship Between Age Rating and the Median Final Ticket Price

Median Final Ticket Price by Age Rating

G: 16.24
R: 14.49
PG: 14.99
NR: 14.49
TBD: 14.49
PG13: 14.49

| Age Rating |
| --- |
| G |
| R |
| PG |
| NR |
| TBD |
| PG13 |

The only price variation that is present within the 'Age Rating' variable is when the movie screening is rated either G or PG. It is uncertain whether the Age Rating has a meaningful impact on the Final Ticket Price.

Group 3: Consumer Characteristics

<u>Relationship Between Ticket Type and the Median Final Ticket Price</u>

There is a considerable amount of price variation based on the 'Ticket Type' variable and suggests that the type of ticket the consumer qualifies to purchase has an impact on the Final Ticket Price. Adult Tickets have the highest Median Final Ticket Price at $16, while FC Promo has the lowest at $1.25 (the low-ticket price may be due to the same anomaly associated with Flashback Cinema.
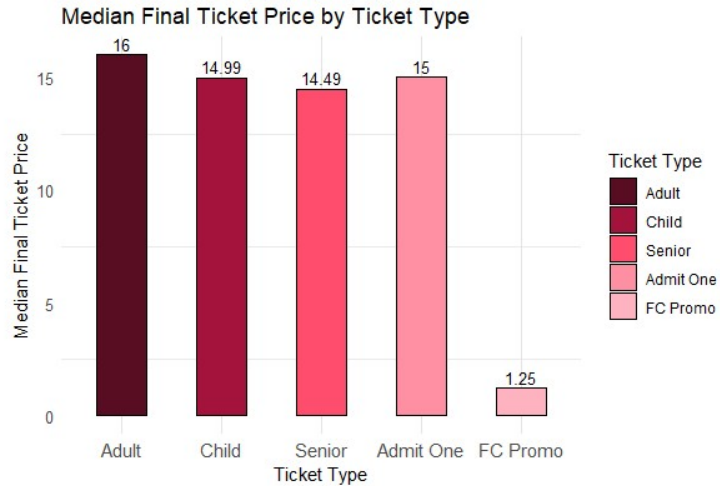
<u>Relationship Between Purchase Method and the Median Final Ticket Price</u>

The Median Final Ticket Price for tickets purchased Online is $14.99, while the Median Final Ticket Price for tickets purchased In-Person is $12.5. There is a $2.49 difference between both prices which is the same amount (Online Fee) added onto the Base Ticket Price when the consumer chooses to purchase tickets online. This has a profound impact on the Final Ticket Price.

Group 4: Continuous Variables

Since the Base Ticket Price and the Final Ticket Price are both continuous variables, I concluded that a scatterplot with a line of best fit would produce a more concise image of the relationship between both variables. A simple linear regression model is used to produce this line of best fit.

R Studio Output for Simple Linear Regression

```
Call:
lm(formula = Final_Ticket_Price ~ Base_Ticket_Price, data = MovieTheaterDataSet)

Residuals:
   Min     1Q Median     3Q    Max
-1.290 -1.255  1.145  1.235  2.744

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.28960    0.14522    8.88   <2e-16 ***
Base_Ticket_Price  0.99722    0.01064   93.76   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.251 on 3101 degrees of freedom
Multiple R-squared:  0.7392,    Adjusted R-squared:  0.7391
F-statistic:  8790 on 1 and 3101 DF,  p-value: < 2.2e-16
```
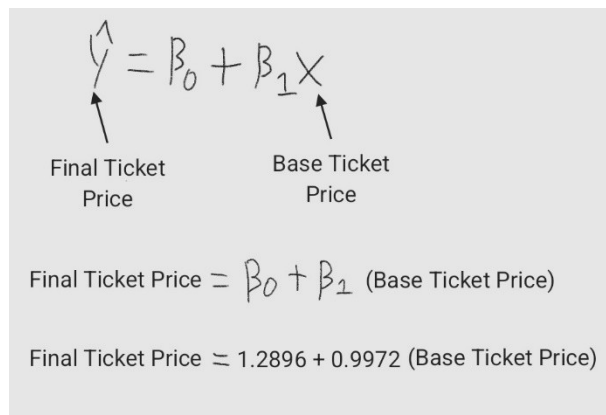
The intercept coefficient as detailed in the output is 1.2896, which represents the Final Ticket Price when the Base Ticket Price is equal to zero. The slope coefficient is 0.9972, which indicates that for every unit increase in the Base Ticket Price, the Final Ticket Price increases by $0.997. Both p values (indicated as $Pr(>|t|)$ in R) for both variables are less than 2e-16 which means that the relationship is statistically significant and not due to random chance. The model predicts there is high confidence that the Base Ticket Price has a strong relationship (perhaps even casual) with the Final Ticket Price as captured by the small standard errors and a t value of 93.76. With both the R^2 and Adjusted R^2 value being equal to 0.7392, approximately 73.92% of the variation in the Final Ticket Prices can be explained by Base Ticket Prices alone.

Simple Regression Equation

Both the intercept coefficient (1.2896) and the slope coefficient (0.9972) form the basis in which a Regression Equation can be created. Keep in mind that interpreting the intercept coefficient in a practical sense (since the Base Ticket Price cannot be $0) does not yield meaningful insights, but it remains necessary for model accuracy when analyzing the relationship between both the Base and Final Ticket Price.
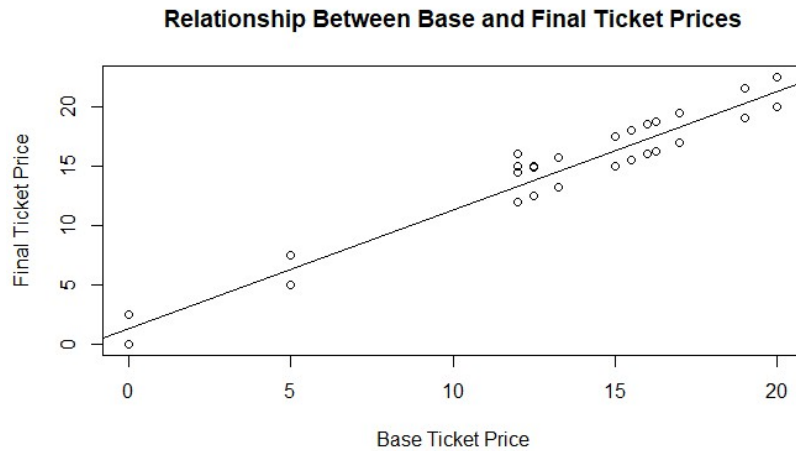
$$\hat{Y} = \beta_0 + \beta_1 X$$

Final Ticket Price   Base Ticket Price

Final Ticket Price $= \beta_0 + \beta_1$ (Base Ticket Price)

Final Ticket Price $= 1.2896 + 0.9972$ (Base Ticket Price)

Relationship Between Base and Final Ticket Prices

The correlation between both variables was also calculated. With a correlation coefficient of 0.86, there is a strong positive correlation between both variables, which means that when one variable increases, the other increases in the same direction. This scatterplot also shows a consistent spread of points both below and above the line of best fit indicating homoscedasticity.

While the relationship between the Base and Final Ticket Price remains strong in a simple linear regression model, refining the model to include additional factors from the dataset can help explain additional price variation and contribute to a more accurate understanding. The inclusion of such variables will be pronounced in the following chapter.

Grouped Bar Charts Interpretation

Day of the Week Price Variations Based on Ticket Type
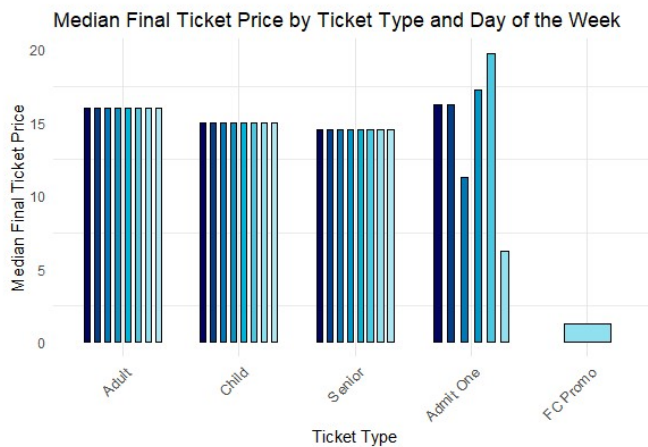
Grouped Bar Charts were used to explore potential price variation across multiple categorical variables. In this bar chart, we aimed to explore how the Final Ticket Price fluctuated throughout the week based on the type of Ticket the consumer purchased. Adult, Child, and Senior tickets exhibited uniform prices throughout the week, while Admit One and FC Promo contained substantial price variation.

<u>Ticket Type Price Variations</u>

Examining these price variations carefully shows the impact the Admit One and FC Promo ticket types have on the Final Ticket Price throughout the week and can reliably demonstrate the validity of adding the Ticket Type variable onto any future model.

Median Final Ticket Price: Admit One and FC Promo by Day
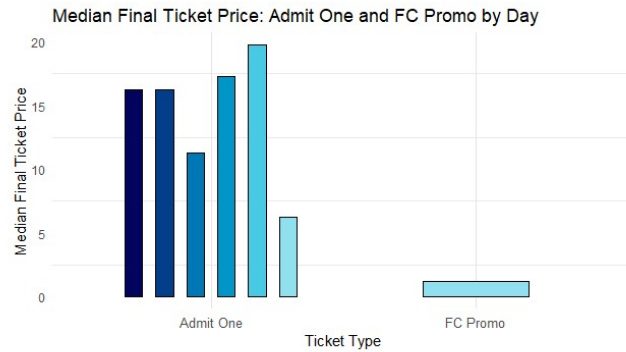
Day of the Week Price Variations Based on Format

Median Final Ticket Price by Format and Day of the Week
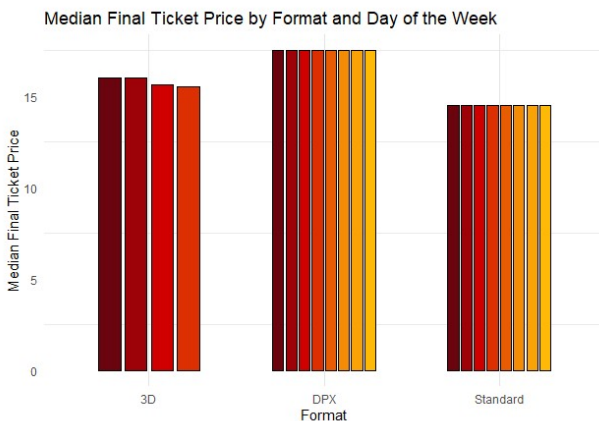
In this grouped bar chart, we observed how the Final Ticket Price fluctuated throughout the week based on the format of the movie screening. Both DPX and Standard format had the most uniform prices throughout the week, while the 3D format had the most price variation. This variability is more subtle than the previous grouped bar chart entailed because the 3D format is represented as a downward trend across the week before flatlining close to the weekend.

<u>Trend in 3D vs Non-3D Formats</u>

Trend in Median Final Ticket Price: 3D vs. Non-3D Formats

3D Format pricing trends were examined using a dummy variable in which 3D and Non-3D outcomes were captured. Between Monday 1 and Tuesday, the Final Ticket Price for 3D tickets was $16, then decreased by $0.4 on Wednesday. By Thursday, the Final Ticket Price was $15.50, without any price available for the rest of the week. This may closely reflect the demand for 3D movie screenings disappearing by the weekend as new movies start their release schedules.

Econometrics Model

Regression Variable Analysis

A simple linear regression model was produced in Chapter Two which captured the relationship between the Base and Final Ticket Price. The model happened to explain that approximately 73.92% of the variation in the Final Ticket Price is explained through just the Base Ticket Price alone. Meaningful insights reflect the significance of employing this model, but further refinement is needed to determine the impact other independent variables have on the Final Ticket Price (the primary dependent variable) as well as the remaining unexplained variation. Additional pricing factors, mostly categorical, were chosen and vetted to be included in the Final Regression Model.

Variables of Interest

The following list details independent variables within the dataset that are of strategic interest to the Final Regression Model:

- Screening Daypart → Ticket prices differ based on whether the movie screening occurs during the Matinee or Evening showing times.
- Genre → Genres may have different ticket prices based on market demand and market segmentation.
- Day of the Week → The day on which the ticket was purchased may impact its pricing due to demand fluctuations on the weekend.
- Format → Premium formats tend to have higher ticket prices than regular formats.
- Ticket Type → Price discrimination allows tickets to be charged based on different consumer categories.
- Base Ticket Price → The initial ticket price before additional charges or discounts can be applied.
- Purchase Method → Online ticket purchases incur a $2.49 Online Fee making them more expensive than In-Person ticket purchases.
- Discounted Ticket → Discounts for tickets may be applied which inadvertently reduces the ticket price.
- Special Event Pricing → Movie screenings may qualify for unique pricing models based on special programs or promotional events.
- Age Rating → Family-friendly movie screenings may have lower prices compared to age-restricted movie screenings which reflect differing audience demand.

Variables of Concern

There are some variables within the dataset that have little to no significance and face obstacles that impact their relevance to the Final Regression Model. Most variables tend to have properties which are incompatible with the overall Model, while others require deeper analysis and diagnosis. Nevertheless, these variables of concern are represented in the table below.

Variables of Concern	Description
Screening ID	Identifier variable with little to no predictive power. Irrelevant.
Movie Short Form ID	Identifier variable with little to no predictive power. Irrelevant.
Movie Screening Title	Identifier variable with little to no predictive power. Irrelevant.
Time Periods	Multicollinearity with Screening Daypart. Non-Significant
Screening Showtime	Identifier variable with little to no predictive power. Irrelevant.
Movie Type	Identifier variable with little to no predictive power. Irrelevant.
Online Fee	Impact on Final Ticket Price can be explained through Purchase Method instead. Redundant.
Special Program	Limited Sample Size and no consistent price variation. Non-Significant.

Both the Time Periods and Special Program variables faced extensive scrutiny before eventually being excluded from the Final Regression Model. For example, since Time Periods is constructed from the Screening Daypart variable, it is fair to suggest that there is not only a strong relationship between the two, but a state of multicollinearity, which occurs when two predictor variables are highly correlated. The correlation coefficient, calculated within the *R Markdown Document,* was found to be sufficiently high (0.89) above the threshold for multicollinearity (0.7), indicating a strong correlation between Time Periods and Screening Daypart. This strong correlation justified the elimination of Time Periods from the broader model. Similar calculations were made to determine the validity of the Special Program variable, which included testing for multicollinearity with respect to Special Event Pricing, as both captured aspects of special pricing strategies. While there was a weak correlation between the two (0.29), what ultimately led to the exclusion of Special Program was a limited sample size among various select programs. A low sample size meant that the price variation recorded was both inconsistent and unreliable, and that its elimination would ultimately improve the Final Regression Model's accuracy.

Interaction Effects

Alongside numerous variables of interest, the impact of interaction terms, which measure the simultaneous effects two independent variables have on the final output, was also tested in the service of determining whether certain variable combinations may significantly influence the Final Ticket Price beyond individual effects. Only statistically significant variable combinations are present within this section, though a larger list of interaction terms were tested and documented in the associated *R Markdown Document*. These interaction terms were computed using ANOVA interactions tests and tools available in RStudio. Despite an initial knowledge gap in ANOVA testing, meaningful interactions were identified and incorporated into the Final Regression Model.

Interaction Between Screening Daypart and Format

```
## Exploring Screening_Daypart:Format

anova_interaction2 <- aov(Final_Ticket_Price ~ Screening_Daypart * Format,
                          data = MovieTheaterDataSet)
summary(anova_interaction2)
```

```
##                          Df Sum Sq Mean Sq F value Pr(>F)
## Screening_Daypart         1    901   901.2  193.18 <2e-16 ***
## Format                    2   3215  1607.3  344.53 <2e-16 ***
## Screening_Daypart:Format  2     42    20.8    4.46 0.0116 *
## Residuals              3097  14448     4.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# The interaction between both variables is statistically significant.
# Include in the final regression model.
```

Not only does the interaction term display statistical significance ($p = 0.0116 < 0.05$) but also contributes meaningful price variation, as indicated by its F-Value (4.46), to the Final Ticket Price. This suggests that the impact of Format on the Final Ticket Price varies across Screening Dayparts, or in other words, some formats may show larger price differences between Matinee and Evening showings than other formats.

Interaction Between Genre and Screening Daypart

This interaction term is statistically significant ($p = 0.0224 < 0.05$) but has a weaker interaction effect (F = 2.235) due to both Screening Daypart having a dominant effect on the Final Ticket Price as well as Genre having multiple categories, which spreads variation across different groups. Matinee and Evening tickets show larger price difference across different genres despite weaker interaction effects.

```
## Exploring Genre:Screening_Daypart

anova_interaction4 <- aov(Final_Ticket_Price ~ Genre * Screening_Daypart,
                          data = MovieTheaterDataSet)
summary(anova_interaction4)
```

```
##                          Df Sum Sq Mean Sq F value Pr(>F)
## Genre                     9   1151   127.9  23.905 <2e-16 ***
## Screening_Daypart         1    854   854.4 159.650 <2e-16 ***
## Genre:Screening_Daypart   8     96    12.0   2.235 0.0224 *
## Residuals              3084  16504     5.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# The interaction between both variables is statistically significant.
# Include in the final regression model.
```

Interaction Between Special Event Pricing and Day of the Week

```
## Exploring Special_Event_Pricing:Day_of_the_Week

anova_interaction5 <- aov(Final_Ticket_Price ~ Special_Event_Pricing * Day_of_the_Week,
                    data = MovieTheaterDataSet)
summary(anova_interaction5)


##                                    Df Sum Sq Mean Sq F value  Pr(>F)
## Special_Event_Pricing               1   1278  1278.5 255.718 < 2e-16 ***
## Day_of_the_Week                     7    117    16.7   3.333 0.00153 **
## Special_Event_Pricing:Day_of_the_Week  5   1767   353.3  70.675 < 2e-16 ***
## Residuals                        3089  15444     5.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


# The interaction between both variables is statistically significant.
# Include in the final regression model.
```

This interaction term, while statistically significant ($< 2e\text{-}16 < 0.05$), has an F-Value (70.675) that is higher than Day of the Week alone, indicating price variation that is driven primarily through Special Event Pricing with less uniformity across the week. In business terms, some days may have higher ticket prices for Special Events than others.

Parametric Statistical tests

After evaluating both variables of interest and interaction effects, parametric statistical tests were conducted to determine each variable's predictive power in influencing the Final Ticket Price. While familiar with using t-statistics academically, Analysis of Variance (also known as ANOVA) was chosen instead to account for the needs of the dataset. ANOVA tests are particularly useful in assessing the significance and predictive power of variables with multiple groups—rather than the two that are usually allowed in a standard t-test—and can be used to analyze the mean differences across such groups within variables like Genre, Format, or Ticket Type, to name a few. Despite the initial knowledge gap in ANOVA statistical tests, both single-factor and multi-factor ANOVA tests were conducted to classify primary predictor variables and retain them in the Final Regression Model.

A single-factor ANOVA test was conducted to analyze the differences between groups based on each independent variable. Tests results were recorded in the associated *R Markdown Document* under the object 'SingelFactorANOVAOutput' in detail. A table displaying each variable's individual tests results, including each F-Value, P Value, and predictive power can be found below.

Variable Name	F-Value	P Value	Predictive Power
Screening Daypart	157.9	<2e-16	Moderate
Genre	22.67	<2e-16	Moderate
Day of the Week	5.132	8.11e-06	Marginal
Format	326.8	<2e-16	Strong
Ticket Type	476	<2e-16	Strong
Base Ticket Price	8790	<2e-16	Dominant

Purchase Method	1065	<2e-16	Dominant
Discounted Ticket	285.1	<2e-16	Strong
Special Event Pricing	228.8	<2e-16	Strong
Age Rating	6.988	1.68e-06	Marginal

Key takeaways can be made when observing this table. The first is that virtually all variables, apart from Day of the Week and Age Rating, have high statistical significance. Every P Value in this table completes the threshold for statistical significance ($p < 0.05$). The second is that the Base Ticket Price and Purchase Method, both with exceedingly high F Values (8790 and 1065, respectively), commands the most variation in the Final Ticket Price and are the most dominant predictors of pricing. Ticket Type, Format, Discounted Ticket, and Special Event Pricing explain the second highest group of variation in the Final Ticket Price as each have considerably high F Values compared to the rest of the table. The third takeaway is that both Day of the Week and Age Rating, with their comparatively low statistical significance, only exhibit marginal predictive power and can only explain small amounts of price variation in the Final Ticket Price. The final takeaway is that these test results directly challenge the hypothesis of this project. Time (which is represented by Screening Daypart and Day of the Week) is not the most dominant predictor of price like previously hypothesized. Instead, the Purchase Method and Ticket Type seem to be the strongest determinants of pricing, excluding the Base Ticket Price as the baseline. That means that the characteristics of the consumer explain most of the variation in the Final Ticket Price. While the consumer has control over whether they purchase tickets Online or In-Person, they have less control over the Ticket Type in which they are grouped in. Keep in mind that the results from this ANOVA test only account for independent effects and do not include combined effects which will be analyzed using a different parametric statistical test.

A multi-factor ANOVA test was conducted to assess how the combined effects of multiple independent variables and the potential interactions between them may influence or explain the variation in the Final Ticket Price. ANOVA tools in RStudio were used to conduct this test.

Note: While conducting this test in RStudio, the Special Event Pricing variable was omitted from the multi-factor ANOVA test results. It is unclear what caused this to occur, but assuming this happened with good judgment, the variable was not accounted for in the analysis.

Multi-Factor ANOVA Test

```
# Multi-Factor ANOVA tests can be used to examine
# combined effects variables have on the Final_Ticket_Price.
anova_formula <- Final_Ticket_Price ~ Screening_Daypart + Genre + Day_of_the_Week +
  Format + Ticket_Type + Base_Ticket_Price + Purchase_Method +
  Discounted_Ticket + Special_Event_Pricing + Age_Rating

multi_factor_anova <- aov(anova_formula, data = MovieTheaterDataSet)

summary(multi_factor_anova)
```

```
##                     Df Sum Sq Mean Sq   F value   Pr(>F)
## Screening_Daypart    1    901     901 3.006e+04 < 2e-16 ***
## Genre                9   1105     123 4.093e+03 < 2e-16 ***
## Day_of_the_Week      7    142      20 6.770e+02 < 2e-16 ***
## Format               2   2707    1354 4.514e+04 < 2e-16 ***
## Ticket_Type          4   6553    1638 5.464e+04 < 2e-16 ***
## Base_Ticket_Price    1   2351    2351 7.841e+04 < 2e-16 ***
## Purchase_Method      1   4754    4754 1.585e+05 < 2e-16 ***
## Discounted_Ticket    1      0       0 6.600e-02 0.79697
## Age_Rating           4      0       0 3.810e+00 0.00429 **
## Residuals         3072     92       0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Key takeaways can be made when observing these test results. First, recall from the single-factor ANOVA test that the two strongest predictors of the Final Ticket Price were the Base Ticket Price and the Purchase Method. In a multi-factor ANOVA test, the order of these variables is flipped, and the Purchase Method now has the highest F-Value (158,500) compared to the Base Ticket Price (78,410). This indicates that the decision to purchase tickets Online of In-Person has a profound impact on the Final Ticket Price when accounting for multiple variables simultaneously. Second, recall from the previous test how Day of the Week and Age Rating explained the least amount of variation when accounting for variables independent of each other. In this ANOVA test, however, Day of the Week exhibits moderate predictive power while Age Rating has a Marginal influence. What this means in a business context is that pricing strategies are more likely to fluctuate across the week than they are based off age restrictions. Third, when accounting for the combined effects of all variables, the Discounted Ticket has both a low F-Value (0.066) and non-significant P-Value (0.79697 > 0.05), indicating that its predictive power is Insignificant. Lastly, these test results once again reaffirm the notion that variables concerned with time don't have as great of an impact on ticket pricing as once hypothesized, even accounting for multiple variables simultaneously. This takeaway has become a signal for recognizing whether time-related variables will contribute meaningfully to the Final Regression Model or not. A table displaying each variable's predictive power can be found below.

Multi-Factor ANOVA Test Results

Variable Name	Predictive Power
Screening Daypart	Strong
Genre	Moderate
Day of the Week	Moderate
Format	Strong
Ticket Type	Strong
Base Ticket Price	Dominant
Purchase Method	Dominant
Discounted Ticket	Insignificant
Age Rating	Marginal

Final Regression Model

Final List of Variables

Below is the full list of variables included in the Final Regression Model, designated by variable and group type:

A. Continuous Variables
  - Base Ticket Price
B. Categorical Variables
  1. Time-Related Characteristics
     - Screening Daypart
     - Day of the Week
  2. Movie Characteristics
     - Genre
     - Format
     - Age Rating
  3. Consumer Characteristics
     - Ticket Type
     - Purchase Method
C. Binary Variables
  - Special Event Pricing
  - Discounted Ticket
D. Interaction Terms
  - Screening Daypart x Format
  - Genre x Screening Daypart
  - Special Event Pricing x Day of the Week
E. Dependent Variable
  - Final Ticket Price

Final Regression Equation

Below is the final multi-linear regression equation without any numeric values in written form:

$$\hat{Y} = \beta_0 + \underbrace{\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}_{} + \ldots$$

$\uparrow$ Final Ticket Price

$\uparrow$ Independent Variables

Final Ticket Price $= \beta_0 + \beta_1$ (Screening Daypart) $+ \beta_2$ (Genre) $+ \beta_3$ (Day of the Week)

$+ \beta_4$ (Format) $+ \beta_5$ (Ticket Type) $+ \beta_6$ (Base Ticket Price) $+ \beta_7$ (Purchase Method)

$+ \beta_8$ (Special Event Pricing) $+ \beta_9$ (Discounted Ticket) $+ \beta_{10}$ (Age Rating) $+$

β_{11} (Screening Daypart:Format) $+ \beta_{12}$ (Genre:Screening Daypart) $+$

β_{13} (Special Event Pricing:Day of the Week)

$$\beta_0 = \text{Intercept}$$

$$\beta_1 - \beta_{10} = \text{Independent Variables}$$

$$\beta_{11} - \beta_{13} = \text{Interaction Terms}$$

Here is the same Final Regression Model expressed in the associated *R Markdown Document:*

```
## Call:
## lm(formula = Final_Ticket_Price ~ Screening_Daypart + Genre +
##     Day_of_the_Week + Format + Ticket_Type + Base_Ticket_Price +
##     Purchase_Method + Special_Event_Pricing + Discounted_Ticket +
##     Age_Rating + Screening_Daypart:Format + Genre:Screening_Daypart +
##     Special_Event_Pricing:Day_of_the_Week, data = MovieTheaterDataSet)
```

CHAPTER FOUR: RESULTS

Regression Output

Below is the Regression Output, which includes mostly the main variable effects, generated from the Final Regression Equation created from the previous chapter:

```
Residuals:
    Min      1Q   Median       3Q      Max
-0.3971  -0.0156  -0.0015   0.0128   3.5980

Coefficients: (5 not defined because of singularities)
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 3.849e-01  9.865e-02   3.902 9.75e-05 ***
Screening_DaypartEvening    3.385e-02  5.104e-02   0.663 0.507272
GenreDocumentary           -2.205e-01  9.485e-02  -2.324 0.020175 *
GenreAnimation             -2.251e-01  6.661e-02  -3.379 0.000737 ***
GenreHorror                -2.353e-01  6.988e-02  -3.367 0.000768 ***
GenreAction                -2.303e-01  6.516e-02  -3.534 0.000415 ***
GenreAnime                 -1.735e-01  1.236e-01  -1.404 0.160535
GenreFantasy               -2.205e-01  6.588e-02  -3.346 0.000828 ***
GenreThriller              -1.429e-01  6.373e-02  -2.243 0.024993 *
GenreDrama                 -2.242e-01  6.876e-02  -3.261 0.001124 **
GenreComedy                -2.373e-01  6.010e-02  -3.949 8.04e-05 ***
Day_of_the_WeekTuesday      3.786e-05  1.279e-02   0.003 0.997637
Day_of_the_WeekWednesday    3.577e-03  1.293e-02   0.277 0.782095
Day_of_the_WeekThursday     2.030e-02  1.286e-02   1.578 0.114655
Day_of_the_WeekFriday       1.401e-02  1.264e-02   1.108 0.267807
Day_of_the_WeekSaturday     1.413e-02  1.274e-02   1.109 0.267532
Day_of_the_WeekSunday       1.411e-02  1.288e-02   1.095 0.273564
Day_of_the_WeekMonday 2     2.347e-02  1.317e-02   1.782 0.074861 .
FormatDPX                   7.091e-03  2.900e-02   0.245 0.806836
FormatStandard             -2.228e-02  2.659e-02  -0.838 0.402133
Ticket_TypeChild           -1.859e-02  1.251e-02  -1.487 0.137216
Ticket_TypeSenior          -6.241e-03  1.347e-02  -0.463 0.643250
Ticket_TypeAdmit One       -1.012e-02  1.085e-01  -0.093 0.925739
Ticket_TypeFC Promo        -3.469e-02  1.501e-01  -0.231 0.817201
Base_Ticket_Price           9.888e-01  4.134e-03 239.168  < 2e-16 ***
Purchase_MethodOnline       2.476e+00  6.107e-03 405.386  < 2e-16 ***
Special_Event_PricingYes          NA         NA      NA       NA
Discounted_TicketYes       -3.138e-02  7.386e-02  -0.425 0.670908
Age_RatingR                 1.608e-02  1.230e-02   1.308 0.191128
Age_RatingPG                7.524e-03  1.327e-02   0.567 0.570814
Age_RatingNR               -2.074e-01  3.070e-02  -6.756 1.70e-11 ***
Age_RatingTBD              -3.885e-02  1.146e-01  -0.339 0.734637
Age_RatingPG13                    NA         NA      NA       NA
```

The Regression Output had extensive information that could not be displayed within one single image. Therefore, a second image is shown that captures mostly the interaction effects and other key metrics vital to interpreting the model:

```
Screening_DaypartEvening:FormatDPX                   -1.495e-02  4.209e-02  -0.355 0.722428
Screening_DaypartEvening:FormatStandard              -1.814e-02  3.946e-02  -0.460 0.645757
Screening_DaypartEvening:GenreDocumentary             1.343e-01  1.443e-01   0.930 0.352257
Screening_DaypartEvening:GenreAnimation              -5.088e-03  3.374e-02  -0.151 0.880123
Screening_DaypartEvening:GenreHorror                  3.694e-01  4.538e-02   8.139 5.72e-16 ***
Screening_DaypartEvening:GenreAction                 -4.432e-03  3.337e-02  -0.133 0.894328
Screening_DaypartEvening:GenreAnime                  -5.448e-03  5.063e-02  -0.108 0.914306
Screening_DaypartEvening:GenreFantasy                -8.878e-03  3.314e-02  -0.268 0.788819
Screening_DaypartEvening:GenreThriller               -8.561e-02  4.346e-02  -1.970 0.048971 *
Screening_DaypartEvening:GenreDrama                  -4.686e-03  5.435e-02  -0.086 0.931295
Screening_DaypartEvening:GenreComedy                        NA        NA      NA       NA
Day_of_the_WeekTuesday:Special_Event_PricingYes      -1.428e-02  1.578e-01  -0.091 0.927873
Day_of_the_WeekWednesday:Special_Event_PricingYes    -3.206e-02  1.109e-01  -0.289 0.772468
Day_of_the_WeekThursday:Special_Event_PricingYes     -1.823e-01  1.322e-01  -1.379 0.168116
Day_of_the_WeekFriday:Special_Event_PricingYes             NA        NA      NA       NA
Day_of_the_WeekSaturday:Special_Event_PricingYes     -2.132e-02  1.456e-01  -0.146 0.883615
Day_of_the_WeekSunday:Special_Event_PricingYes       -8.419e-02  1.247e-01  -0.675 0.499632
Day_of_the_WeekMonday 2:Special_Event_PricingYes           NA        NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1701 on 3057 degrees of freedom
Multiple R-squared:  0.9952,    Adjusted R-squared:  0.9952
F-statistic: 1.422e+04 on 45 and 3057 DF,  p-value: < 2.2e-16
```

Two key takeaways that can be taken from the Regression Output without regard to any concrete interpretation: The first is that the inclusion of reference categories for all main effects can have a profound impact on the way in which each variable is interpreted, therefore it is paramount to take any calculated metric with a grain of salt until further analysis. For example, metrics associated with the main effect 'Screening_Daypart:Evening' are relative to the main effect 'Screening_Daypart:Matinee' despite the latter not being present in the output. Reference categories may be explored in detail later in this chapter. The second is that main effects that produce 'NA' metrics are automatically excluded from the model whether due to an unreliable reference category, multicollinearity, insufficient variability, or a simple lack of observations. These circumstances may be examined on a limited basis later in this chapter.

A byproduct of this Regression Output is the frequency in which various main effects are considered non-significant or lacking meaningful insights. A table consisting of statistically significant main effects was created because of this byproduct and contains columns that indicate the direction of the coefficient, Std. Error Precision, T value Confidence, Predictive Power (measured by t value), and P value Significance. The table can be found below.

Variable	Coefficient Direction	Std. Error Precision	T Value Confidence	Predictive Power	P Value Significance
Intercept	Positive (0.3849)	Moderate (0.09865)	Strong Effect (3.902)	Strong	High (9.75e-05)
Genre:Documentary	Negative (-0.2205)	Moderate (0.09485)	Strong Effect (-2.324)	Marginal	Moderate (0.020175)
Genre:Animation	Negative (-0.2251)	Moderate (0.06661)	Strong Effect (-3.379)	Strong	High (0.000737)

Genre:Horror	Negative (-0.2353)	Moderate (0.0688)	Strong Effect (-3.367)	Strong	High (0.000768)
Genre:Action	Negative (-0.2303)	Moderate (0.06516)	Strong Effect (-3.534)	Strong	High (0.000415)
Genre:Fantasy	Negative (-0.2205)	Moderate (0.06588)	Strong Effect (-3.346)	Strong	High (0.000828)
Genre:Thriller	Negative (-0.1429)	Moderate (0.06373)	Strong Effect (-2.243)	Marginal	Moderate (0.024993)
Genre:Drama	Negative (-0.2242)	Moderate (0.06876)	Strong Effect (-3.261)	Strong	High (0.001124)
Genre:Comedy	Negative (-0.2373)	Moderate (0.06010)	Strong Effect (-3.949)	Strong	High (8.04e-05)
Base Ticket Price	Positive (0.9888)	High (0.004134)	Strong Effect (239.386)	Dominant	High (<2e-16)
Purchase Method:Online	Positive (2.476)	High (0.006107)	Strong Effect (405.386))	Dominant	High (<2e-16)
Age Rating:NR	Negative (-0.2074)	High (0.03070)	Strong Effect (-6.756)	Strong	High (1.70e-11)
Screening Daypart:Evening x GenreHorror	Positive (0.3694)	High (0.04538)	Strong Effect (8.139)	Strong	High (5.72e-16)
Screening Daypart:Evening x GenreThriller	Positive (-0.0856)	High (0.04346)	Weak Effect (-1.970)	Weak	Moderate (0.048971)

The reliability and performance of the Final Regression Model can be measured using metrics located at the bottom of the output. The Residual Standard Error for this Model (0.1701) is supposed to represent the difference between predicted ticket prices and observed ticket prices. This difference, which is around $0.17 per ticket, captures variation in the Final Ticket Price that cannot be explained by the model alone, and it underscores how mostly accurate the model is at predicting ticket prices. The coefficient of determination, denoted as R-Squared or Adjusted R-Squared, measures the proportion of variation in the Final Ticket Price that is explained by independent variables in the dataset, and can be used to determine the model fit. Both the R-Squared and Adjusted R-Squared values are equal to 0.992, which indicates that 99.2% of variation in the dependent variable can be explained by the model. This exceedingly high percentage suggests that there may be variables within the model capture more predictive power than other variables (i.e. Base Ticket Price, Purchase Method, etc.). Keep in mind that this high percentage could also indicate overfitting, which occurs when the model fits too closely to the dataset and can impact the reliability of the model. Since the Adjusted R-Squared is the same as R-Squared, it is safe to assume all predictor variables contribute meaningfully to the model and the likelihood of overfitting is low. The F-Statistic (1.44e+04), measures the statistical significance of the model and determines whether there are some factors that influence ticket prices more than others. The high F-Statistics means there are greater differences in the variation of ticket prices that can be attributed to different independent variables. The low p-value at the end of the output (<2.2e-16) reinforces this conclusion and confirms that the results of the model are not due to random chance.

Interpretation

Findings from the Regression Output were analyzed and dissected based on each variable and interaction term's relationship to the regression model and dependent variable. These insights led to the documentation of reference categories for each main effect, the interpretation of general insights within the model's results, and the illumination of real-world business implications for pricing strategies. Additionally, Data Visualizations were also generated to support findings, though their analysis will materialize later in this chapter. Each variable and interaction term, regardless of statistical significance, was interpreted below.

<u>Screening Daypart</u>

- Reference Category → Screening Daypart (Matinee)
- Results from the regression output indicate that the Screening Daypart (Evening) is not statistically significant, which means that there is hardly any meaningful difference between Matinee and Evening ticket prices.
- Flipping the reference category only reaffirms the lack of statistical significance.
- Interaction effects between Screening Daypart and Format do not yield interesting insights since there is no statistical significance for any of the associated main effects.
- On the contrary, interaction effects between Screening Daypart and Genre (which exhibits some level of statistical significance) reveal certain genres may have different pricing structures in the Evening.
- Business Implications → While the movie theater doesn't apply uniform price increases for Evening tickets, it may be motivated by the demand of specific genres to adjust ticket prices accordingly.

<u>Genre</u>

- Reference Category → Genre (Concert)
- Concert ticket prices tend to be more expensive than other genres, which is reaffirmed by the fact that many genres have negative coefficients in the output.
- Most genres except Documentary and Thriller have Strong predictive power on the Final Ticket Price.
- The interaction effect between Genre and Screening Daypart suggests that Horror films that take place during the Evening receive a $0.37 unit price increase, while Thriller films during the Evening receive a $0.08 unit price decrease.
- Business Implications → The movie theater recognizes there is a high demand for Concert films compared to other genres and is charging them accordingly.

<u>Day of the Week</u>

- Reference Category → Day of the Week (Monday 1)
- Results from the regression output indicate that Day of the Week is not statistically significant across most of the business week, meaning there is little price variation and weak predictive power.
- The interaction effect between Day of the Week and Special Event Pricing yields inconsistent statistical significance across its components suggesting a lack of a strong relationship between the two factors.
- Business Implications → Customers pay roughly the same price for tickets regardless of the day of the week.

<u>Format</u>

- Reference Category → Format (3D)
- Results from the regression output indicate that both DPX and Standard formats are not statistically significant, meaning there was little price variation compared to 3D Formats.
- Recall from earlier chapters how the grouped bar chart associated with Formats exhibited relatively strong price variation compared to the regression output. The reasoning for this contrast is likely due to the price variations not being strong enough to hold predictive power when accounting for other factors in the model.
- The interaction effect between Format and Screening Daypart are also not statistically significant, at least in this model, suggesting that ticket prices do not vary meaningfully between Screening Dayparts across different formats.
- Business Implications → The model suggests that format-based pricing is mostly fixed for this movie theater, and that there is little variation based on showtimes.

<u>Ticket Type</u>

- Reference Category → Ticket Type (Adult)
- All ticket types compared to the reference category are not statistically significant in the model.
- Adult ticket prices tend to be more expensive than other ticket types, which is reaffirmed by the fact that many ticket types have negative coefficients in the output.
- The grouped bar chart for Ticket Type created in previous chapters showed considerable price variation compared to regression results, and it is likely due to the variable lacking key predictive power when accounting for other factors in the model.

- Business Implications → Ticket Type does not meaningfully influence ticket prices since general ticket pricing is mostly fixed, but ticket types that are occasionally promotional (i.e. Admit One and FC Promo) may exhibit profound price differences not particularly observed in the model.

<u>Base Ticket Price</u>

- Reference Category → Continuous Variable
- Results from the regression output indicate that the Base Ticket Price is one of the strongest predictors of the Final Ticket Price within the model since the variable possesses an extremely high t value, high statistical significance, and Dominant predictive power.
- There is a very strong relationship between the Base Ticket Price and the Final Ticket Price as specified by the simple linear regression plot created from previous chapters.
- A $1 increase in the Base Ticket Price increases the Final Ticket Price by $0.98.
- The Base Ticket Price explains most of the price variation in the Final Ticker Price.
- Business Implications → The Base Ticket Price serves as the baseline ticket price that this movie theater charges before additional factors are applied.

<u>Purchase Method</u>

- Reference Category → Purchase Method (In-Person)
- Results from the regression output indicate that the Purchase Method (Online), relative to the reference category, is the strongest predictor of the Final Ticket Price within the model as it not only boasts an extremely high t value with Dominant predictive power but explains most price variation within the model with high statistical significance.
- The decision to purchase tickets Online rather than In-Person results in a $2.47 increase in the Final Ticket Price, which aligns with the $2.49 Online Fee applied at checkout.
- Business Implications → Consumers who opt to purchase tickets Online face substantial price increases compared to other factors. This movie theater aims to capitalize on consumer willingness to pay for ease of access with the goal of capturing additional revenue from Online transactions. Whether this pricing strategy proves a reliable source of revenue is dependent on consumer behavior patterns.

Special Event Pricing

- Reference Category → Special Event Pricing (No)
- Special Event Pricing (Yes) was dropped from the model due to lack of sufficient data or collinearity and is represented in the regression output with NA values.
- Special Event Pricing with respect to the reference category, is an inconsistent predictor of the Final Ticket Price.
- The interaction effect between Special Event Pricing and Day of the Week produced main effects that had either NA or not statistically significant values, which means that special event price differences are not tied to specific days of the week.
- Business Implications → Based on this model, this movie theater is not keen on adjusting ticket prices for special events alone nor does it see itself charging different ticket prices for special events on different days of the week.

Discounted Ticket

- Reference Category → Discounted Ticket (No)
- Results from the regression output indicate that Discounted Ticket (Yes) does not have statistical significance, nor does it explain a meaningful amount of price variation in the Final Ticket Price, which means it is not a strong predictor of ticket prices.
- The particularly low sample size in Discounted Ticket is likely the reason why the model fails to capture meaningful insights from this variable.
- Business Implications → This movie theater likely only applies discounts very rarely to most ticket purchases and does not currently see value in employing discount-based strategies (excluding membership benefits).

Age Rating

- Reference Category → Age Rating (G)
- Results from the regression output indicate that most age restrictions, apart from NR and PG13, are not statistically significant and do not have Strong predictive power over the model, compared to G-rated films.
- The only statistically significant main effect is Age Rating (NR), whose direction and coefficient are negative, and who possess Strong predictive power over the model.
- For every one unit increase in Age Rating (NR), the Final Ticket Price decreases $0.20, suggesting that the main effect is cheaper than the reference category.

- Age Rating (PG13) is the only main effect that produced NA results, which means it was effectively dropped from the regression model due to collinearity or lack of sufficient data.
- Business Implications → Age restrictive pricing alone is not appealing to the movie theater since there are little price differences between age ratings nor is it in their best interest to schedule more NR-rated films (i.e. limited releases, independent films) as ticket prices for these films tend to reflect lower demand.

Screening Daypart and Format

- Reference Category → Screening Daypart (Evening) and Format (3D)
- Results from the regression output indicate that the interaction effect between Screening Daypart and Format are mostly not statistically significant relative to the combined main effect of 3D Evening tickets.
- The model failed to capture a relationship between Screening Daypart and Format that meaningfully influenced ticket prices.
- Business Implications → Based on this model, this movie theater does not currently recognize any value in adjusting ticket prices based on the combined effects of Screening Daypart and Format since format-based pricing remains fixed regardless of screening showtime.

Genre and Screening Daypart

- Reference Category → Screening Daypart (Evening) and Genre (Concert)
- Results from the regression output indicate that the interaction effect between Screening Daypart and Genre is statistically significant for certain genres during the Evening compared to Concert films during the Matinee.
- Screening Daypart (Evening) and Genre (Comedy) is the only main effect that was ultimately dropped from the model.
- The interaction effect between Genre and Screening Daypart suggests that Horror films that take place during the Evening receive a $0.37 unit price increase, while Thriller films during the Evening receive a $0.08 unit price decrease.
- Horror films likely have higher demand during the Evening while Thriller films have lower demand which is reflected in their ticket price.
- Business Implications → This movie theater is more likely to offer extra Horror film screenings during the Evening compared to other genres in hopes of optimizing various revenue streams.

<u>Special Event Pricing and Day of the Week</u>

- Reference Category → Special Event Pricing (No) and Day of the Week (Monday1)
- Results from the regression output indicate that the interaction effect between Special Event Pricing and Day of the Week produced inconsistencies as several main effects were either dropped entirely from the model or lacked statistical significance, which would suggest that special event price differences are not tied to specific days of the week at least in this model.
- Business Implications → This movie theater does not see value in charging different ticket prices for special events across specific days of the week as it favors a more flexible pricing strategy that adjusts prices on a case-by-case basis instead.

Visualizing the Regression Model

Numerous data visualizations were produced to evaluate the assumptions, accuracy, and performance of the Regression Model. These visualizations are mostly used to address the fundamental limitations of the model and how a future model may be generated to solve these issues. Plots are expressed within the associated *R Markdown Document.*

<u>Distribution of Final Ticket Prices</u>

Distribution of Final Ticket Prices
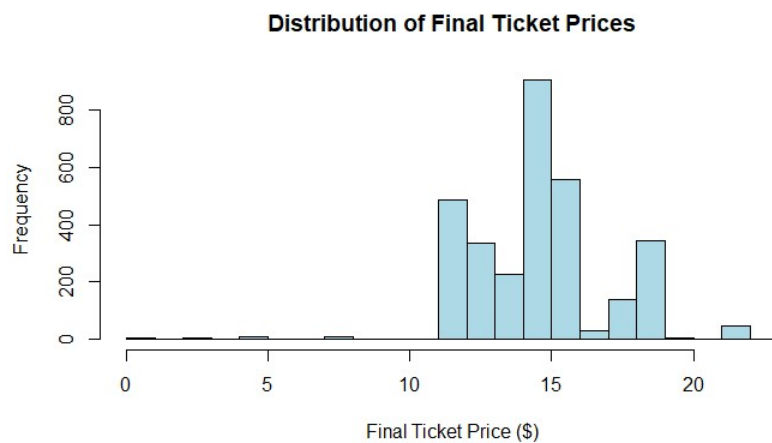
The distribution of Final Ticket Prices is left-skewed where ticket prices cluster toward higher values, while lower priced tickets are generally considered outliers. This suggests that certain factors that have high predictive power may be skewing the distribution closer to the right. Data from the output should be interpreted cautiously as this may indicate potential bias.

Residuals vs Fitted Plot

Since most residuals hover close to 0, apart from a few outliers, it is safe to assume the model is relatively balanced and accurate, though there may be some deviations that require further analysis. Data also does not form any distinct shape validating the assumption of homoscedasticity, where the variance of residuals remains constant across all levels of predicted values.
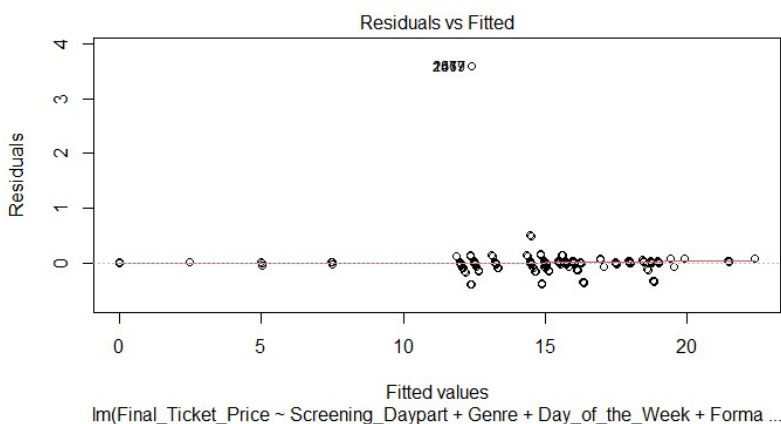

Quantile-Quantile (QQ) Plot
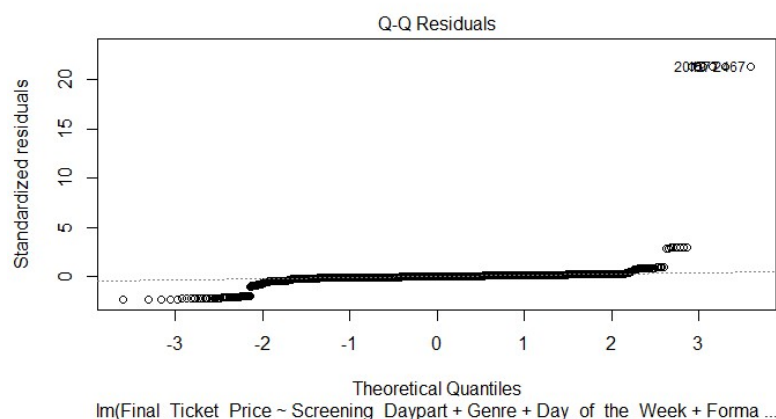
Most points fall along the dotted line of the QQ plot, suggesting that the residuals are mostly normally distributed. The lower and higher quantiles of ticket prices fall either below or above 0 and represent the model failing to capture extreme ticket prices, potentially impacting the interpretation of certain predictor coefficients, likely those that are related to extreme ticket prices.

So far, the Regression Model performs well for most ticket prices, however, it fails to capture meaningful insights for a considerable number of other predictors, either underpredicting or overpredicting the most extreme ticket prices. This is reflected in the results of the regression output where numerous variables are either statistically non-significant or whose predictive power is minimal compared to the strongest predictors.

<u>Predicted vs Actual Ticket Prices</u>

The model accurately predicts most ticket prices, compared to the actual observed prices, reinforcing both the high variation explained by the model (99.2%) as well as the model's reliability in assessing predictive power. The few outliers in the plot may indicate overestimation in certain ticket prices and highlight the potential need for refining the model.
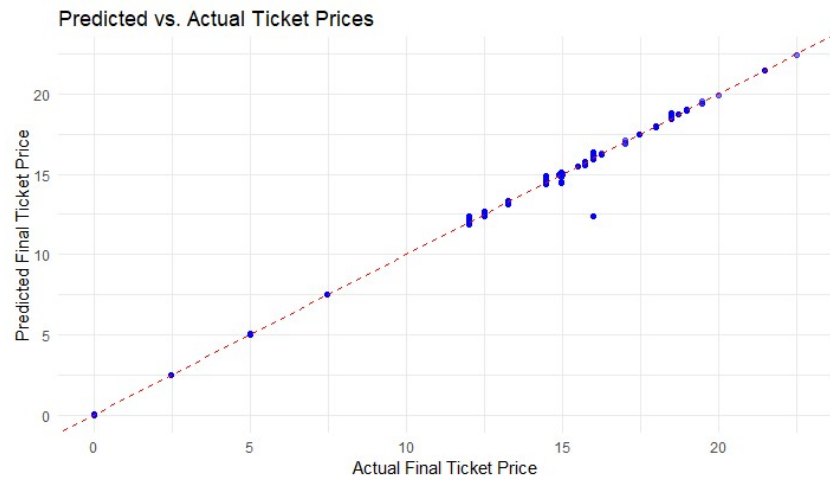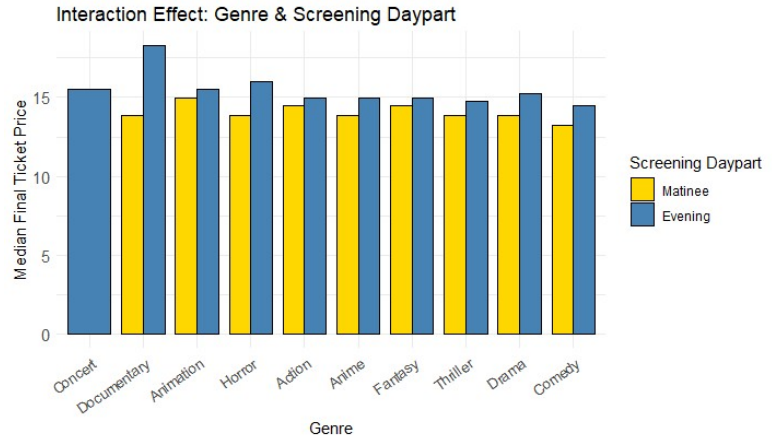
Interaction Plots

The last group of data visualizations produced for this Regression Model were the Interaction Plots, which were intended to illustrate how the interaction effects of two variables influenced Final Ticket Prices. Insights were also extracted by comparing plot findings with results from the regression output. Each interaction effect was interpreted visually below.

<u>Screening Daypart and Format Plot</u>
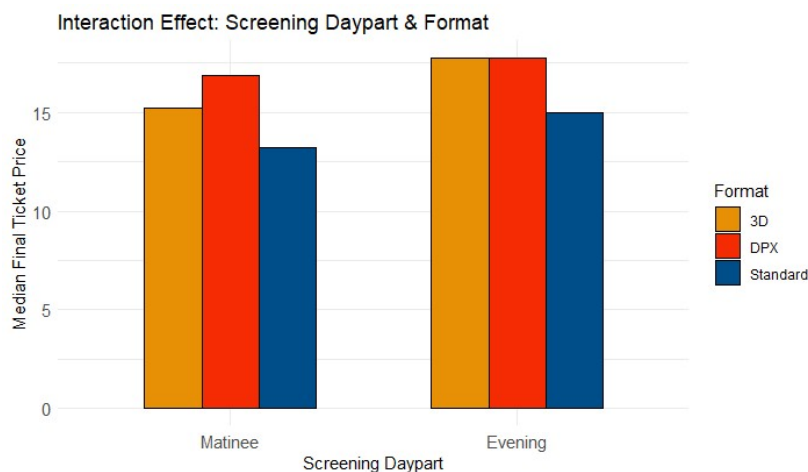
While the regression model did not detect a statistically significant interaction between Screening Daypart and Format, the price variation displayed in the plot suggests that subtle price differences exist but are not large or consistent enough to be fully captured by the model. Testing a different reference category may capture these differences more clearly.

Genre and Screening Daypart Plot

The regression model only captured statistically significant price effects for Horror and Thriller films which contrast with the interaction plot showing price changes for all genres, minus Concert films. Testing a new reference category is recommended as the reference category suffers from missing data (Evening prices for Concert films do not exist), impacting the model's ability to estimate the interaction effect accurately.

Special Event Pricing and Day of the Week Plot
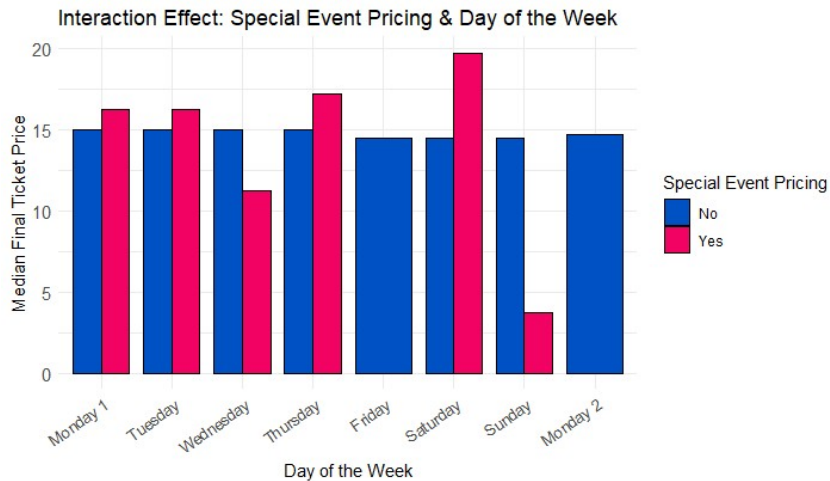
Both the interaction plot and the regression results suffer from a lack of consistent pricing patterns, though the plot suggests that price variation does exist. The model failed to capture the magnitude of this variation due to highly irregular data for Special Event Pricing, suggesting that price adjustments are made on a case-by-case basis rather than predictably applied.

Chapter Five: Discussion

Insights

At the beginning of this project, we hypothesized that time was the strongest predictor of final ticket prices. Time was represented by three distinct categories, two of which were included in the Final Regression Model. Though data graphics like bar charts and grouped charts indicated that some level of price variation was apparent across multiple time-related categories, these price differences were not large or statistically significant enough to be fully captured by the model. Nor did the single-factor or multi-factor ANOVA tests from earlier chapters designate either time-related variable as having Dominant predictive power over final ticket prices. The results from this project ultimately challenged the previous hypothesis and constituted the emergence of more fixed pricing factors, such as Base Ticket Price and Purchase Method, serving as the primary drivers of final ticket prices.

The most robust predictors of final ticket prices were indicated by their predictive power in the Regression Model. Both the Base Ticket Price and Purchase Method were the dominant factors influencing ticket prices and could explain most of the price variation within the model with an exceedingly high statistical significance. Base Ticket Price is the baseline component for pricing strategies applied by the target movie theater and serves as the foundation in which all additional pricing factors are applied. The Purchase Method introduces a fixed cost variation, whereby customers can opt to purchase movie tickets either In-Person at the box office at no additional charge or through the Online reservation system incurring a fixed Online Fee. Customers who choose to purchase tickets Online may have a higher willingness for ease of access features compared to those who purchase tickets In-Person and are less responsive to the additional charge. Regardless of whether consumers exbibit ease of access willingness or are simply indifferent to their method of purchase, the Online Fee represents a considerable revenue stream for the movie theater to collect as well as a tool to observe consumer behavior.

The third major predictor of final ticket prices is the intercept which yields strong predictive power over the model, and it represents the ticket price when all independent variables inside the model are equal to zero or at their reference levels. While the Base Ticket Price represents the baseline and primary driver of price differences, the intercept captures fixed pricing structures that exist independent of fluctuations in other variables. Foundational pricing structures are crucial in the movie theater's ability to observe operational costs while also establishing reliable streams of revenue for business objectives. With consistent fixed pricing, consumers can anticipate stable ticket prices which reduces any level of uncertainty the consumer might have upon each ticket purchase or visit to the movie theater.

Genre and Age Rating represent special cases in the Regression Model where certain groups exhibited strong predictive power and statistical significance while others did not. Given the fact that both factors are categorized as attributes of movie screenings, there is reason to believe that this movie theater selectively applies pricing strategies based on movie attributes like demand expectations for specific genres and audience demographics. Most film genres, excluding Documentary and Thriller, consistently influence ticket prices relative to Concert films. This relationship suggests that Concert movie screenings are treated as a premium pricing tailored to high demand expectations compared to every other genre. Similar to Genre, most Age Rating groups apart from NR and PG13, are neither statistically significant nor yield strong predictive power compared to G-rated films. In fact, the PG13 group was automatically dropped from the model altogether due to producing NA values. This leaves NR as the only age rating that has strong predictive power in the model. NR-rated films tend to be limited releases or exclusive films that cater to niche audience demographics, usually ending up cheaper than their mainstream counterparts, and reflect the movie theater's desire to strategically attract niche audiences that have lower demand. It could also reflect a commitment to diversify film offerings to a wider audience, making niche screenings more affordable for the average consumer, while also encouraging habitual movie attendance and fostering brand loyalty.

Documentary and Thriller film screenings are the only main effects in this model that exhibit marginal predictive power and moderate statistical significance which means their ticket prices remain relatively stable compared to other genres as price differences are typically smaller. The lack of price variation indicates frequent price adjustments seem unnecessary for Documentary and Thriller films as they tend to attract smaller, more predictable audiences compared to other screenings in high demand, thus eliminating the need for erratic pricing strategies. Stable pricing strategies in this domain enable the movie theater to generate a reliable source of revenue and may foster habitual movie attendance for consumers seeking affordable pricing. Fans of Documentary or Thriller films benefit from this stability and are more likely to consider brand loyalty.

Multiple interaction effects were tested in earlier ANOVA tests, however, only three of them made it to the final regression model: Screening Daypart and Format; Screening Daypart and Genre; Special Event Pricing and Day of the Week. Findings from the regression results showed that only the interaction between Screening Daypart and Genre produced meaningful insights as other interaction terms required additional testing and analysis. The interaction between both variables when Horror films occur during Evening showtimes yielded strong predictive power and pricing premiums. It indicates that the target movie theater expects stronger demand for Horror films in the afternoon and capitalizes on it through price increases. In contrast, Thriller movies during the Evening have the weakest predictive power in the entire model and signal a reduction in ticket prices. Thriller films do not independently produce stable pricing adjustments and neither do the interaction between Screening Daypart and Genre. The inconsistency between the main effects for this interaction term suggests that pricing adjustments are likely made on a

case-by-case basis instead of having a uniform pricing strategy. Horror films are seen as more monetarily attractive than Thriller films both during the Evening and may induce higher ticket prices for the average consumer.

The Final Regression Model is highly effective at predicting ticket prices, as evidenced by its high R-Squared value (99.2%), and indicates that most of the variation in final ticket prices can be explained by all predictors in the regression model. It's vital to be cautious over the possibility of encountering overfitting in the model due to the very high R-Squared value, which can impact the model's ability to generate new data beyond the dataset's scope. There may also be a tendency to weight specific predictors like Base Ticket Price and Purchase Method more than other predictors in the model, which runs the risk of overpowering the influence that other variables have on ticket pricing. Additionally, Other data and analytical limitations may decrease the reliability of the regression model and deduce the interpretations to its simplest form. Nevertheless, insights from the regression model serve as the foundation for understanding weekly pricing patterns and can be leveraged to refine pricing strategies, observe consumer behavior, and optimize monetary decisions to achieve business objectives.

Limitations

This project encountered a mirage of setbacks and limitations that prevented the peak of a full and comprehensive analysis of weekly ticket pricing patterns from being reached. Various challenges ranging from data collection, model design, and statistical analysis impeded the reliability, precision, and accuracy of the project's findings. These constraints were expanded upon below.

Data Limitations

The scope of this project only captured information about movie theater ticket prices over the course of one week. Naturally, this limited timeframe meant that the dataset suffered from some variables having smaller sample sizes than if the timeframe was longer. The advent of a small sample size prevented some variables from being accounted for entirely within the regression model despite their insights having some level of business relevance. If one were to replicate this project, the data collected would ideally focus on a larger timeframe (i.e. monthly, seasonal) and would be able to capture additional pricing trends that weren't previously recorded.

During the process of data collection, a handful of observations were categorized as anomalies within the dataset. Irregular data was mostly due to website bugs and instances that were entirely out of our control but were ultimately left in the analysis. Data anomalies produced ticket prices that were outside the normal range of prices—also known as outliers—and impacted the distribution of ticket prices. In fact, outliers were the main reason for opting to use Median instead of Mean as a statistical metric for summarizing central tendencies in ticket prices, because the Median is not as sensitive to extreme values as the Mean typically is.

Numerous variables also suffered from a lack of meaningful variation in their observations, challenging the project's ability in fully understanding pricing trends. The scope of the project likely impaired the extent to which price differences in certain factors could be observed as the timeframe failed to produce enough data. These inconsistencies were both recorded and visualized through various data graphics within this project and highlighted the difficulties in determining whether patterns were influenced by fixed or fluid pricing structures. Additionally, the regression model failed to detect substantial predictive power for the affected variables, diminishing or in some cases expelling their statistical effects on the model. Increasing the scope of the project or eliminating the specific variable altogether from the model are two solutions that address these discrepancies.

The process of data collection was conducted by passively recording ticketing information as they appeared on the movie theater's website. This approach may have introduced potential selection bias in the project as the dataset only captured statistics without randomness and may not reflect the true distribution of movie screening features taken across the selected timeframe.

As a result, other key pricing trends failed to be fully represented in the statistical analysis, which ultimately limits the broader applicability of the project's results onto long-term pricing strategies. Other forms of bias may include omitted variable bias, wherein external factors that may be relevant to the project are not accounted for, such as seasonality, promotional discounts associated with membership participation, and other consumer-driven factors, to name a few. Mitigating any potential bias would require expanding the scope of this project, incorporating a longer timeframe to account for additional external factors, and experimenting with different sampling approaches to capture a more accurate representation of the data. Future versions of this analysis would enhance the accuracy of any statistical model and constitute a deeper understanding of ticket pricing patterns.

Model Limitations

The Final Regression Model for this project was heavily influenced by the Base Ticket Price and Purchase Method as these factors demonstrated dominant predictive power and statistical significance. This constituted a major concern for the model as these variables overpowered the effects of smaller pricing factors while explaining most of the price variation. In fact, these variables are likely the reason why the regression model exhibited such a high R-Squared value (99.2%), which, while indicated strong predictive power for the model and its factors, constituted grave concerns about overfitting. This phenomenon occurs when the model is too reliant on the trained dataset and struggles to predict accurate data beyond what's already been observed. Overfitting risks both the accuracy and reliability of the regression model, and it also reinforces the fact that some variables simply have higher explanatory power than others. In pursuit of capturing more nuanced pricing trends, the imbalance in the regression model can be corrected by adjusting the model structure to be less dependent on dominant predictors so that the pricing effects from smaller variables can be examined more meaningfully.

Background statistical procedures within RStudio, though not operating intentionally or with any malice, influenced the model-building process through variable selection and exclusion. Certain variables from the dataset were automatically dropped from the regression model due to multicollinearity concerns, lack of statistical significance, or insufficient variation, and likely prevented the full capture of all main effects and their impact on ticket prices. Some interaction terms and their main effects also failed to produce meaningful insights for the model to fully seize unto. While these exclusions helped streamline the model, they may have inadvertently omitted insightful relationships between other variables. It may be worthwhile to reintroduce these omitted variables for further assessment so long as other model limitations are addressed.

The model-building process was also influenced by statistical procedures aimed at designating reference categories for each variable, which subsequently shaped the interpretation of regression results. Each reference category chosen served as the baseline against which other main effects

were compared, meaning that the estimated coefficients reflected values relative to that specific category. Reference category selection introduced challenges within the regression model, particularly when selected baselines failed to represent broader pricing patterns or lacked sufficient price variation. These challenges skewed the interpretability of certain variables, making it difficult to assess their relevance to the model and subsequently their influence on ticket pricing patterns. Extensive testing with alternative reference categories is needed to yield different insights, highlighting the magnitude of selecting a valid baseline in regression modeling.

Analytical Limitations

One of the most difficult limitations faced was the incessant learning gaps that constrained the heights of interpretation. Traditionally, t-statistics were the most intuitive parametric statistical test due to familiarity in academic settings, however they were insufficient in tackling the needs of the dataset. ANOVA statistical tests were conducted instead, to capitalize on its prowess in handling categorical variables with multiple levels, though it required brief independent study of surface level ANOVA concepts to complete. The ANOVA tests in this project, while their results should be considered carefully, laid the foundation for modeling the final regression equation and proved to be a vital tool in assessing the predictive power of various factors.

The issue of knowledge gaps repeated itself throughout the course of this project, particularly in the realm of visualizing the regression model. Most of the regression diagnostic plots generated to measure model performance—such as the Residuals vs. Fitted Plot, QQ Plot, and Predicted vs Actual Plot—required additional self-study to properly interpret without error. This lack of experience with such diagnostics may have impacted the depth of understanding the regression model though fortunately not as insidious. Concepts such as heteroskedasticity, homoskedasticity, and multicollinearity presented themselves initially in an academic setting, though as the analysis progressed, their applications appeared more complex than previously anticipated. Despite interpretation difficulties, efforts were made to address potential weaknesses and assumptions in the model without compromising on core features of this project. Future projects will obviously benefit from a stronger foundation of understanding these tools and concepts to improve the reliability of any regression analysis.

A consequence of conducting regression diagnostics was the divulgence of distribution and outlier challenges constraining the model's analysis. For example, the histogram plot revealed a left-skewed distribution with numerous outliers showing ticket prices being substantially lower than other ticket prices. Had these outliers been removed from the model entirely, ticket prices would have appeared more normally distributed instead. The decision to keep these outliers initially depended on the preservation of data integrity and the fact that the median, the preferred measure of central tendency, was indifferent to the inclusion of extreme values, however their

presence may have inadvertently affected the reliability of the regression results by reducing the statistical strength of specific variables. An additional complication was that certain statistical tests that were used to assess variable predictive power, though were not a part of the final product, were dependent on the condition of normality. These statistical tests were promptly removed from the project due to a high level of education needed to be used with full confidence, and the insights for ticket price distribution were deliberately simplified. Advanced statical methods may be used in the future to capture the nuance of non-normal distributions and the outliers that constrain these distributions may be removed.

Deliberate steps were taken to balance the simplicity and complexity of the regression model by testing specific variables for multicollinearity or homoscedasticity. The exclusion of certain factors due to multicollinearity issues may have unintentionally removed variables that could have provided meaningful insights into ticket pricing patterns, meaning that relationships between pricing factors may have not been fully captured. This issue limits the explanatory power of the regression model. In fact, within the model itself there were instances where certain predictors may have indicated multicollinearity but were not removed due to inconclusiveness or out of preserving vital pricing factors. This would also limit the explanatory power of the regression model, though not as urgently. This limitation highlights the importance of balancing model interpretability and variable retention to ensure pricing-relevant factors remain while mitigating regression model distortion.

Applications

Over the duration of this project, we've explored key insights and weekly ticket pricing trends at Mission Grove Galaxy Theatre, examining the relationship between movie screening characteristics and ticket pricing patterns. Through statistical analysis and regression modeling, we've uncovered meaningful insights into ticket pricing structures that can be broadly implemented to make key business decisions. Business objectives at this movie theater include optimizing revenue, retaining loyal customers, and ultimately creating unique value in the market that is both competitive and innovative. Transitioning from analysis to application, we will explore how insights can be leveraged to achieve strategic business autonomy, forecasting, and long-term pricing refinement. Keep in mind that the practical applications of these insights must rely on two key assumptions: First, they assume that the regression results accurately reflect real-world movie theater dynamics beyond the confines of the dataset. Second, they assume that the ticket prices recorded are a measure of consumer demand, and that if this assumption holds, then the estimated coefficients may also serve as an indicator of demand. The business applications of this project's findings may be found below.

Leveraging Consumer Behavior Patterns

Key consumer behavior patterns can be extracted from this project, and it is imperative that we recognize them so that they may be leveraged to shape future ticket pricing strategies. There is a distinction that can be made between behaviors that are controllable by the consumer and those that may be determined by the business itself. The regression results have indicated that controllable behaviors like what Purchase Method the consumer chooses to use exhibit dominant predictive power over the final ticket price. Likewise, uncontrollable factors such as the Base Ticket Price also have dominant predictive power and are predetermined by the movie theater before any additional factors can be included. This distinction between both is made because it highlights the extent to which consumers can influence their own price sensitivity versus adapting to an already established pricing structure and can impact the extent to which the movie theater can adjust ticket prices to influence consumer behaviors. Factors and behaviors that are predetermined by the business may reflect its competitive position in the market and an adaptable strategy.

Other key behavioral patterns may indicate consumer willingness to pay for different movie-going experiences. For example, consumers who opt to purchase tickets online accept the price of accessibility and convenience, while in-person consumers are averse to these additional charges. Similarly, there are certain genres and age ratings associated with different movie screenings that may cater to niche audiences and their pricing levels may reflect their level of demand. Concert films tend to command higher ticket prices because of a perceived higher demand, while NR-rated films are typically cheaper than other films once again because of the

demand perception. These different pricing levels are indicative of the larger business strategy at play, one that segments its own customer base based on their willingness to pay for specific movie preferences. Market segmentation allows the movie theater to cater to both premium-seeking and price-averse customers without compromising the commitment to profitability and value. If consumer behaviors drastically change, whether it is based on factors like the purchase method, genre, or age rating, the movie theater is obliged to align their pricing strategy with shifting demand patterns and evolving consumer perceptions. They could even be placed in a position which warrants the creation of additional consumer segments that are poised to maintain the movie theater's long-term business strategy.

The consumer base for this movie theater can be divided into the two broadest categories: Casual versus Frequent moviegoers. Casual moviegoers tend to be more price sensitive per transaction, selective in their movie choices, and less likely to join loyalty programs offered by the movie theater. They are also more likely to seek premium movie experiences like blockbuster releases, special events, or high-demand screenings and have a higher willingness to pay for occasional premium experiences rather than consistently engaging in the theater's offerings. Frequent moviegoers, on the other hand, exhibit strong brand loyalty and are less price sensitive in the long run as they are more likely to take advantage of cost-saving opportunities like membership programs, discounts, or promotional campaigns. These consumers are also more likely to purchase tickets for more niche films and respond warmly to value-based pricing strategies. This movie theater happens to offer a token-economy-based loyalty program which is meant to encourage repeat purchases, customer retention, and brand loyalty integration all without succumbing to any major membership fees. The loyalty program serves as a strategic tool for price differentiation, allowing the movie theater to balance lucrative offerings with consistent pricing structures in the pursuit of optimizing revenue opportunities.

For this movie theater, opportunities exist for pricing strategy refinement that can reliably leverage the behavioral insights obtained in this project. Understanding the distinction between controllable and uncontrollable consumer behaviors, for example, enables the movie theater to implement a business strategy that can take advantage of major consumer preferences, such as the choice in purchase method, movie screening features (i.e. genre, age rating, etc.), or even membership status. If trends in consumer behavior favor certain movie characteristics over others, the pricing strategy must reflect these preferences while maintaining profitability. Fixed and variable costs, like operating and labor expenses, are factors that the consumer cannot control and must be managed by the movie theater to retain key margins.

Forecasting Movie Theater Trends

Formal forecasting techniques cannot be utilized to predict movie theater trends due to academic knowledge constraints but also due to its inclusion considered to be mostly unnecessary. Keep in mind that the model only captures a limited scope of statistical variation, and so long-term forecasting models would fail to reliably forecast expected movie theater trends. As an alternative, we can forecast pricing trends within the confines of the model itself, highlighting short-term consumer behaviors and movie screening characteristics. Additionally, shifts in consumer behavior, whether due to new screening releases or evolving consumer preferences, introduces a layer of complexity that the model regrettably fails to address. The regression model relies on a static dataset captured without regard to real-time fluctuations in consumer demand.

To account for model limitations, short-term forecasts focused on consistent patterns tied to specific variables rather than relying on a broad swath of variables that assume consistency. Screening Daypart and Purchase Method are variables that show repeatable weekly patterns instead of volatility. Evening ticket prices are consistently higher than Matinee ticket prices throughout the week suggesting that most consumers tend to prefer later movie screening showtimes rather than earlier. Furthermore, if we assume that the movie theater segments their consumer base into Casual and Frequent moviegoers, Casual moviegoers—who are more likely to attend later showtimes—typically face higher ticket premiums. In contrast, Frequent moviegoers may be more likely to attend screenings during off-peak hours—they may be eager to leverage their membership for reduced costs—and so are exposed to lower ticket premiums. This segment of consumer's eagerness to utilize membership benefits enables the theater to implement targeted retention strategies that encourage consistent revenue streams. We can expect this movie theater to repeatedly implement promotional material that targets consumers during off-peak hours on a weekly basis. As the moviegoing experience transitions to the digital age, consumers have a higher willingness for ease of access compared to previous generations and are more likely to purchase their tickets online. Price differences between online and in-person tickets are consistently fixed at $2.49, an online fee that is less likely to change on a week-to-week basis. What is likely to change, however, is how the movie theater chooses to engage with its loyal consumer base. A dedicated mobile app may be promoted with the implicit purpose of encouraging repeat online purchases, and these promotions may fluctuate based on weekly market conditions.

Format is a variable that is tied to contextual variability as its consistency depends on the specific film being shown and its anticipated demand. High-demand titles, for example, may be offered in premium formats such as DPX or 3D which typically have higher ticket premiums across the week. In the case of 3D screenings, however, demand tends to taper off after the weekend as screenings are either reduced or eliminated entirely. The reason for this might have to do with the genre of the film itself since G-rated films that are also shown in 3D attract younger audiences who are more likely to attend during the weekend than the regular business week. Excluding 3D

movie screenings, Format does not produce any meaningful price variation in ticket pricing trends and so from a business perspective, does not require a weekly dynamic pricing strategy and can remain fixed without jeopardizing consistent revenue streams.

Stable variables reduce uncertainty in business applications, increase reliability in forecast assessment, and prove to be a vital factor in developing effective pricing structures for the movie theater. Regrettably, most of the variables within the regression model display some level of price variation and so there is a level of volatility movie theaters must navigate to reduce their uncertainty. Major movie theater chain brands develop rigorous pricing algorithms that incorporate static data and real-time movie theater trends to respond dynamically to changes in the movie theater's business model. Independent or smaller movie theater chains can still identify and apply stable variables as a gateway towards more reliable revenue planning, and in this case, basic forecasting tools are enough to make effective pricing decisions in the short term.

Cost-Oriented Pricing Strategies

Cost-oriented pricing strategies revolve around the relationship between operational costs and pricing decision-making, with the sole goal of balancing these costs with financial objectives. In the context of this project, formal cost-plus or target-cost pricing methods could not be leveraged appropriately due to the project's limited scope and incomplete information. The lack of access to key input costs (whether variable or fixed) renders these pricing methods infeasible and only available through movie theater backend data channels. Conceptually, however, we can say that the movie theater is driven to price its offerings strategically to sustain a markup that is above its operational costs. If the share of revenue allocated towards profit (margin) is known and the movie theater has constructed a baseline price per unit, they can target lower operational costs in the long run with the primary goal of maximizing business profitability. While these ratios are largely hidden from the average observer, available pricing data can reflect how cost-oriented pricing pressures influence consumer pricing decisions.

Two of the most predominant pricing frameworks clearly in practice are cost-based price discrimination and peak-loud pricing. Cost-based price discrimination occurs when the cost of doing business varies between different elements of the business itself. For this movie theater, there are real price differences for movie screenings across varying customer characteristics, movie formats, or purchase methods. Second-degree price discrimination, concerning itself with offering different prices based on the quantity or version of the product, is evident in how the theater charges varying prices based on which purchase method or format the customer chose to purchase the movie ticket in. Allowing customers to self-select into pricing tiers is a core element of second-degree price discrimination. On the other hand, Third-degree price discrimination, which deals with segmenting groups of consumers based on observable characteristics, is identified through the theater's strategy of charging prices based on movie ticket type (Adult,

Child, Senior) or membership participation (data for this is less known but assume there is a cost reduction). The price differences between these select groups will also reflect in the movie theater's operational costs, where they perhaps might incur an additional cost for offering a premium format or more convenient purchasing method (Online services may cost more for the movie theater to maintain). Additional operational costs can also vary based on the time and day of the business week and may require peak-load pricing to address. This framework involves raising prices during high-demand periods to offset higher operational costs. This movie theater offers lower Matinee prices and higher Evening prices to account for varying consumer demand. While they don't typically charge higher prices during the weekend, they may instead offer additional movie screenings, especially those that come in premium formats. These dominant pricing strategies enable the movie theater to have predictable pricing structures contingent on predictable pricing data while still optimizing business profitability and efficiency. In support of these frameworks, other pricing tools like product differentiation and product line pricing serve as effective tools in segmenting offerings based on consumer willingness to pay. The former aligns with second-degree price discrimination and allows the movie theater to maximize consumer surplus whilst charging varying prices for different products. The latter, which sets different prices for various products within the same product line, enables the movie theater to segment consumers based on their perceived value of the product and willingness to pay.

Despite cost-oriented pricing strategies relying heavily on hidden operational costs, they are the most practical and simplistic foundation in reforming business strategy and efficiency. Perhaps the only recommendation that can be made with respect to the movie theater's costs, is to create additional segments of consumers to extract revenue from. They could potentially create an additional ticket type that groups customers that have a Student ID and charge this group a reduced fee to induce more demand. Whether this is truly needed relies on having additional information as well as estimating expected demand. They could also realistically explore raising prices during the weekend to offset additional costs, but also to account for an increase in demand. This would allow the movie theater to keep a constant amount of high-demand movie screenings but could have the unintended effect of decreasing the amount of niche film offerings. Ultimately, a trade off would have to be made between the current business model and an updated one that prioritizes maximizing weekend earnings over weekday revenue.

Value-Based Pricing Strategies

Value-based pricing emphasizes setting prices based on the customer's perceived value of the product or service rather than the firm's operational costs like in cost-oriented pricing. This approach shifts the focus from internal factors like input costs to external factors such as customer perception or behavioral science and is well suited for making educated assessments with less information compared to the limitations of other pricing methods. In the context of this project, the movie theater manifests its value perception dynamically in numerous forms,

offering unique moviegoing experiences for a variety of different consumers, each that values their experience uniquely and produces varying willingness to pay. They may additionally use behavioral and psychological tactics designed to frame their products more attractively to their audience. As long-term growth becomes a vital business objective, the movie theater will focus heavily on customer retainment strategies that utilize core value principles. Once the movie theater can leverage its own value perception, whether it's through market segmentation, psychological pricing, or by managing long-term value, it can set movie theater prices accordingly.

Value-based pricing concerns itself with market segmentation in its broadest form and it is a prerequisite to enacting third degree price discrimination without regard to any cost (this is different from cost-based price discrimination which factors in the cost). Since the movie theater does not know the true willingness to pay for all its customers, it must divide its audience into distinct groups based on observable traits, needs, or behaviors. Once it has segmented its audience, it can then determine the appropriate price to charge each distinct group (this would be third degree price discrimination). For example, this movie theater offers three distinct formats (Standard, 3D, DPX) and charges different prices for each according to the perceived value each group of customer decides. 3D and DPX movie tickets are typically priced higher because not only do these movies formats cost more to operate and maintain, but they are also tied to the assumption that customers who value these formats more are also more likely to pay a higher premium than other groups. Through third degree price discrimination, the movie theater is capable of extracting revenue from all types of customers even when they have a disadvantage in information. Another way to navigate market segmentation is through the concept of price fences which are conditions imposed by the company to prevent customers from jumping between pricing tiers to obtain lower prices. Whilst price fences don't exist for movie formats or screening showtimes, they are present in ticket types where customers are required to provide identification to qualify for Student or Senior discounts. This movie theater has also implemented price fences in their promotional campaigns and membership benefits, selectively offering discounts based on the status of the customer and their willingness to pay. While price fences generally protect the lower end of the pricing structure, the concept of Trading Up typically protects earnings from the upper end, as it encourages customers to purchase higher-end offerings based on the perceived incremental enhancement of benefits. Movie theaters may promote membership participation to its customers by offering incremental benefits to customers willing to pay more, thereby generating higher revenue for the movie theater from capturing a larger pool of consumers. Trading Up is also a tool used in encouraging repeat purchases or retaining loyal customers in the long run, a phenomenon that we will observe later.

Beyond intuitive market segmentation efforts, value-based pricing will employ a myriad of psychological tactics to nudge consumers toward favorable purchasing habits, even by altering their perception of value and price. Perhaps the most apparent pricing tactic employed by movie theaters—and really the most common tactic in business today—is charm pricing, which

leverages the left digit bias to set prices just below a round number to take advantage of consumers perceiving the price lower than it really is. Apart from a few outlier prices, movie ticket prices for this movie theater typically ended in either .49 or .99 and serve as an indicator in the trade-off between bargains for the consumer and higher revenue for the movie theater. It is safe to assume that by employing this pricing tactic, the movie theater is aiming to target price-sensitive and bargain-seeking customers in hopes of attracting more revenue on the lower end of the price structure. Other pricing tactics will generally utilize mental shortcuts to influence the customer's perception of the price, though these may be less observable in this analysis due to the collection of data being conducted primarily from online sources. Anchoring and price primacy describe the mechanisms that shape the consumer's value perception, wherein the former refers to high-priced products being presented first to make lower-priced products more appealing to price-sensitive buyers, and the latter refers to the sequence in which consumers are shown either the price or the product first. Movie theaters mostly apply these tactics alongside promotional material either on their website, mobile app, or in the lobby of the theater itself. For example, since the movie theater offers premium movie formats, they could intentionally market higher-priced movie screenings without first presenting the price—this is the movie theater using the price primacy tactic—and increase the likelihood of movie tickets being purchased. Though this movie theater does not clearly have any observable framing effects (these are mostly presented alongside membership benefits), the effects of the pain of paying can still shape how gains and losses are perceived by consumers. This phenomenon occurs when there are negative emotions associated with consumer's parting from their money when purchasing a good or service. The most apparent use of this tactic comes from the movie theater's surcharge associated with purchasing movie tickets online. Customers that are price-sensitive but who also value convenience may experience the pain of paying when purchasing their movie tickets online. There is a myriad of other pricing tactics in addition to those already mentioned, including mental accounting, bundling, or visual cues, that may also be employed by this movie theater, though for simplicity we will refrain from expanding more.

So far, we've concerned ourselves with the concepts of market segmentation and various psychological pricing tactics, which have been leveraged to capture insights from incomplete information. While these concepts have a stronger relevance in the short run of the business operation, the movie theater must consider strategies that favor their growth and profitability in the long run. Customer Involvement, which refers to the level of emotional investment the customer makes before purchasing a good or service, can be used as a tool in reaching long-term objectives. In the context of a movie theater, customers who value their moviegoing experience are considered to have high Involvement and may plan their movie screenings in advance, be more susceptible to promotional campaigns, and more likely to participate in loyalty programs or memberships. Their high involvement in the ticket purchasing process translates into an opportunity for the movie theater to cultivate customer relationships that foster brand loyalty in the long run, usually through customer retention strategies, but more on that later. Excluding concessions and merchandise purchases, there is a varying level of consumption visibility, which

is the degree to which spending habits are noticeable by others, since it is dependent on the social context of the moviegoing experience. Consumers that exhibit high consumer visibility usually attend in larger groups than those with lower visibility and so are much more likely to purchase goods that signal taste or status. Movie theaters are more likely to capitalize on higher consumer visibility by selectively promoting premium offerings to consumers in larger groups (these may range from families, friends, or couples). One of the premium screenings offered at a movie theater is a private screening event, which are typically reserved for larger, non-traditional groups (corporate meetings, parties, or gaming sessions), and can end up charging cheaper prices per seat. Since these private events face higher consumption visibility, the movie theater may offer special deals on bundled concessions or merchandise to groups that value signaling taste or status. While behavioral nudges like involvement and consumer visibility can shape the long-term customer perception of value, incentive curves provide a framework that aligns consumer behavior with structured business goals in the form of a system that rewards higher spending and repeat purchases with incremental benefits. This movie theater introduces incentive curves in the form of a membership system or more specifically a loyalty program which, while free to sign up for, typically requires customers to meet a spending goal to qualify for various benefits (these programs are referred to as token economies). When a customer signs up for this membership, they are both choosing to participate in an incentive curve, but they are also providing a consistent stream of revenue to the movie theater. The more customers that the movie theater attracts to its loyalty program, the larger stream of consistent revenue that the business receives, thereby increasing both its profitability and growth in the long run. If we recall from earlier, the concept of Trading Up is functionally like an upward-sloping incentive curve, as they both encourage consumers to spend more by offering higher value or benefits.

The process of price refinement is a considerable challenge, especially since value perception is difficult to measure without advanced analytical tools like surveys or conjoint analysis, and so we may have to rely on already established consumer insights to make minor pricing adjustments with the goal of increasing ticket sales. This movie theater has the option of revisiting core market segmentation strategies to generate completely new groups of consumers and charge these groups based on newfound shared characteristics. Since larger groups have higher revenue potential, the movie theater should promote bundled ticket deals during high-demand time periods (perhaps a bundled ticket deal targeted toward families should only be offered for a limited time; a bundled ticket deal targeted toward families would have to replace cheap seasonal tickets) and could even align concessions or merchandise offerings to increase customer value. Additionally, movie theaters must also prioritize trading up to shape existing consumer perceptions and encourage upward mobility within the general price structure (only price fences should have the ability to separate general pricing from membership pricing). The more valuable the premium offerings are to the average consumer, the higher their willingness to pay will be, which is a benefit for both parties. Since effective psychological pricing requires robust market research methods, we'll focus on alternatives to basic charm pricing. The movie theater should consider finding a price that balances fairness with quality, and their best method of doing so

while maximizing revenue comes from odd pricing. Odd pricing may be used to appear fairer to consumers who are particularly price-conscious and can still communicate prestige better than charm pricing can. The movie theater should aspire to charge prices that end in .65 or .75 for movie screenings that may have more premium offerings than normal (3D and DPX screenings are a perfect opportunity to capitalize on this). Lastly, the pricing concept that is likely to pay the highest dividends for the movie theater if they choose to apply it more aggressively is consumption visibility. We have already explored how high consumption visibility encourages more concession and merchandise sales, but this movie theater fails to attractively market their special private events more seriously. By advertising various private event opportunities for different segments of groups, the movie theater will succeed in not only increasing ticket sales but increasing sales for other goods and services as well. Besides creating new customer segments for capturing new sources of revenue and visually altering the price themselves, capitalizing on the concept of consumption visibility remains the most realistic opportunity for the movie theater to engage in pricing refinement.

Competitive Pricing Strategies

Unlike cost-oriented or value-based pricing, which may focus on internal cost pressure or customer value perception, competitive pricing concerns itself with a firm's position within the broader market and how price positioning can be leveraged to influence broader market share. For this project, the target movie theater operates within proximity to two other competitors, both of which reside within the city limits of Moreno Valley and may adjust their offerings and ticket prices as customers within the area continuously evaluate their next best alternative. Since data for rival movie theaters was not collected in this analysis, inferences must be made based on observable pricing patterns that capture each theater's market positioning. These inferences offer insight into evaluating whether the movie theater views itself as a price leader, follower, or niche player within the competitive landscape. Patterns may be examined through the competitive pricing framework, strategic price moves and signaling behavior, and product lifecycle considerations alongside more sophisticated mechanisms.

Market dynamics can be measured effectively by leveraging the competitive pricing framework, which segments relative market share into four distinct categories: Price leadership, Price followership, Transparent pricing, and Profit optimization. If the firm is a price leader, they are a larger player in a concentrated market and have market power to list higher prices than their competitors. They will also possess the ability to influence the market through steep discounts among other strategies. A price follower, on the other hand, may be a smaller player in a concentrated market, pricing their products just below the price leader and small enough to avoid provoking threatening price maneuvers. Their discount moves are more stable and predictable compared to price leaders. Both price leaders and followers reflect the concentration of the top players within the market and their overall market share (usually more than 50%). Firms with a

market share that is lower than 50% will fall either in Transparent pricing or Profit optimization and are likely to perform better in a fragmented market compared to a concentrated. Firms engaged in transparent pricing are large players in a fragmented market, offering predictable and stable prices without deploying excessive discounting tactics, and their primary goals are to establish consumer trust. If a firm is concerned with maximizing revenue, but they do not possess high market share, they will engage in Profit optimization. These firms are smaller players in a fragmented market and typically price their products based on price elasticity and bargain-seekers since they have a higher cost advantage. Firms must price their products to both optimize their profitability and maintain or increase their market share in the long run. In the context of this project, we can use the competitive pricing framework to assess whether the Galaxy Theatre falls under a specific segment of market share or not and can also determine how much market share each specific competitor holds. As noted previously, the target movie theater has two rival theaters within its geographic proximity: Harkins Theatre Moreno Valley 16 and Regency Theatre Towngate 8. Using personal anecdotal evidence, we can infer that each movie theater within this market appeals to a specific share of consumers who prefer a unique moviegoing experience, and as such, are differentiated based on features that are unique to each location. Both Galaxy Theatre and Harkins Theatre are reputable theater chain brands, while Regency Theatre is mostly recognized for offering bargain product offerings. Based on brand recognition alone, it is safe to assume that all players, except one, possess a local market share that is higher than 50%. With Regency Theatre Towngate 8 being the sole movie theater with less than 50% market share, their best-fit segment within the framework of competitive pricing is Profit optimization since their product offerings and ticket prices are aggressively lower than their rivals. Movie screenings are reserved for commercial films only and there may be little to no premium screenings for even high-demand movies. They face an uphill struggle not only to increase their market share but to simultaneously maintain it whilst offering cheaper products to price-conscious buyers. The next movie theater in this market that we will observe is Harkins Theatre Moreno Valley 16. Harkins Theatre is a relatively larger and more established theater chain brand compared to the target movie theater, and so they are more likely to possess a more rigid pricing structure that offers higher baseline customer value (these are offers streamlined across all Harkins locations). Unfortunately, this specific movie theater is not only dated and desperate for repairs but is located within a less than desirable mall with declining traffic. These negative factors keep this movie theater from maximizing its true revenue potential and reflect upon the competitive pricing framework more as Price followership rather than leadership. So long as external factors continue to affect its bottom-line, Harkins Theatre places itself as a price follower, pricing its ticket prices just below the price leader, and likely yields diminishing leverage over the market as its share of the market slowly shrinks. Long-term customers who favor more stable prices and offers are more likely to choose Harkins Theatre instead of the other players within the market. If consumers favor neither a bargain nor an average moviegoing experience, they may have a higher willingness to pay for ticket prices set by the price leader, which in this market, is the movie theater of interest for this project. Galaxy Theatre is

remarkably adept at generating high demand for its offerings despite them having higher ticket prices than their rivals, and it can do so because of its attention to several value-based pricing strategies that we have already explored. If this movie theater wishes to maintain its competitive edge, it doesn't have to make major price maneuvers as it can simply offer promotional material unique to its location. The movie theater market demonstrates moderate concentration, with two players holding the highest market share, and the last player eager to capitalize on the weaknesses of the price follower to gain market share.

Competitive Pricing Framework

This is the Competitive Pricing Framework as introduced in the Darden School of Business Pricing Certification, an academic point of reference applied to this project.
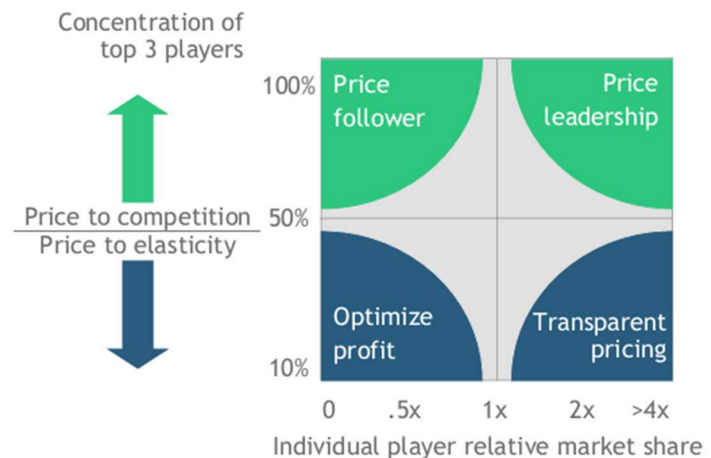
Beyond the competitive pricing framework, the most general pricing model that a firm may use relative to its market are signaling games (related to the study of Game Theory). There are various types of signaling games that firms may use, but for this analysis, we shall only focus on one type: Sequential Games. In a sequential pricing game, one firm will conduct the first move—whether through strategic pricing maneuvers or other signals—while other players observe and decide whether to make maneuvers of their own or ignore it. If one is the price leader and makes the first move, that move will have significant power in influencing the rest of the market, thereby setting a reference point for the competition. These games don't often provoke an immediate reaction but may subtly reinforce the rules of the market over time. For the target movie theater, it maintains its market position as the price leader by holding immense information asymmetry over both itself and its competitors, applying this information with the goal of subtly influencing the competition. If they know the true value of their product and produce a signal that reflects this, not only will they influence their competitor's behaviors (likely intentionally) but they may also appear more credible in the eyes of general consumers, therefore reinforcing the movie theater's high prices. For example, since Galaxy Theatre participates in third party movie programs, their competitor, Harkins Theatre, is more likely to participate in a similar program with the goal of incrementally increasing their demand. This would be an example of a signal that's intended for long-term growth influencing other players within the market. In contrast, the decision to participate in a third-party movie program (Flashback Cinema, etc.) doesn't shape Regency Theatre's business strategy because the bargain

movie theater is unlikely to sustain the program without sacrificing its goal of profit optimization. Information asymmetry and price leadership may exacerbate market conditions and reveal a competitive landscape that is more reactive than previously recognized.

While the competitive pricing framework and sequential pricing games illustrate external competitive dynamics within the market, the product lifecycle model captures the internal evolution of pricing decisions and product maturity over time. The product lifecycle has four phases with varying pricing strategies for each: Development, Growth, Maturity, and Decline. Products within the Development phase are subject to being priced based on buyer price-sensitivity and are marketed either through trial promotions, direct sales, or distribution channels. When the product reaches the Growth phase, the cost and benefits of the product become more apparent and so do the competition. Products are priced either to differentiate from the rest of the competition or to focus on high margins with low volume in the short term (skim pricing, a strategy that gradually reduces the price over time, may be useful for this exact purpose). Products may also be susceptible to penetration pricing, the opposite of skim pricing, which establishes a price in a highly competitive market and is subject to gradual price increases over time. Pricing strategies in the Growth phase are dependent on the volume of segments the firm wishes to focus on. The Maturity phase is the most detailed phase in the product lifecycle model, and it aims to combat an increased sense of price-sensitivity as well as higher competition. The success of the product is likely to impact the firm's competitive position within the market and will result in profit margins becoming more difficult to maintain over the long run. Pricing strategies within this phase focus on improving the efficiency of either the product, costs of the product, or producing additional product lines (through unbundling, the utilization of costs, expanding the product line, or reevaluating distribution channels). Products in the Decline phase undergo declining demand challenges and subsequently must be restructured to account for these deficiencies. Products in this phase may face Retrenchment, which is when the firm withdraws from the weakest segments and defends the remaining profitable segments, or Harvesting processes, where the firm prepares to leave the industry altogether alongside maximizing its revenue. Regardless of which phase the product is in, the firm must acknowledge its position within the product lifecycle and price accordingly. In the operation of a movie theater, the product lifecycle is reflected strongly in the theatrical run of individual films. For example, blockbuster films with high demand expectations release somewhere between the Development and Growth phase as the movie theater shifts to prioritize a high volume of consumers that are segmented based on expected demand for different formats, and as a result tend to be subject to penetrative pricing. Once the blockbuster enters the Maturity phase, demand for additional movie screenings, including premium showtimes, slows down prompting the movie theater to adjust its movie screening showtimes accordingly. Finally, a blockbuster film in the Decline phase may undergo the process of Retrenchment which would reduce a significant portion of movie screenings available whilst only offering high-demand screenings in the evening. Once the process of Retrenchment is complete, the blockbuster film will be phased out entirely so that the movie theater can shift its focus to promoting the next blockbuster film. Understanding the

product lifecycle enables the movie theater to respond decisively to customer demands by applying pricing strategies that maintain a competitive edge over the market.

All competitive pricing strategies that we've explored, including the competitive pricing framework, signaling games, and the product lifecycle model, reinforce the notion that the target movie theater operates from a position of strength within the local movie theater market. As a price leader, it has set the standard for ticket pricing while simultaneously producing a premium moviegoing experience for consumers that its competitors would find difficult and costly to replicate. There are little strategic moves the movie theater must make to maintain a competitive edge over its competitors, and so operational refinement would largely consist of doubling down on business strengths through steep discounting and attractive promotional material to highlight already existing value perceptions. Choosing to employ these business maneuvers occasionally will preserve the movie theater's dominance over the local theater market in the long run for years to come.

Final Reflections and Future Analysis

I started this project in early December of 2024, a few days after choosing to set aside my previous project which contained overlapping subject material. The decision to work on an entirely new project was driven by the need for clearer direction, stronger technical grounding, and a timeline that aligned better with my career goals. It would also allow me to narrow the scope of this project without getting too ambitious and focus my efforts on delivering actionable insights extracted from both technical and theoretical perspectives. I approached this project with a renewed sense of focus, and the success of this analysis allowed me to explore pricing patterns in greater detail and with greater confidence. The movie theater industry was chosen as the topic of concern due to a personal interest and desire to apply academic concepts and newly acquired pricing tools to real-world applications. The Galaxy Theatre located in Riverside presented an opportunity to analyze a real-world local player in the movie theater industry, where insights into their pricing structure, consumer base, and position within the market were scrutinized heavily. Data-based tools such as Excel and RStudio were applied with the purpose of collecting local movie theater data and generating visual outputs (summary tables, bar charts, and regression models), that offered insights into pricing behavior across multiple variables. Statistical tools in the form of regression analysis were used to quantify the relationship between ticket prices and key variables, and it was ultimately leveraged to test the main hypothesis which proposed that time—in any form—would be the strongest predictor of movie theater ticket prices.

Through statistical analysis and regression modeling, numerous key insights and takeaways emerged which enhanced my understanding of how movie theater factors influenced weekly pricing trends. The regression model created for this project revealed the Base Ticket Price and the Purchase Method as key variables with the strongest predictive power over the Final Ticket

Price and could explain most of the price variation within the model. This result reflects the significance of the Base Ticket Price being the foundation of price setting and highlights how the Purchase Method, which differs between Online and In-Person, delivers predictable pricing adjustments over time. Other key variables within the model, such as Genre and Age Rating, produced results that indicated variation in how specific groups exhibited predictive power over the model, supporting the notion that consumer behavioral patterns and their impact on weekly ticket prices were mostly context-driven and less predictable than other factors. Addressing Time in its broadest context required nuance in the form of three distinct variables, but only the interaction between Screening Daypart and Genre produced any meaningful insights within the model. Still, even this interaction effect was highly context-dependent, suggesting that time-based variables are not uniform across weekly ticket pricing trends. Regression results directly challenged the initial hypothesis of the project, prompting a deeper evaluation of the initial model and the weight of each predictor. Intense scrutiny allowed me to conclude that although the chance of overfitting was low, the excessive weight of both the Base Ticket Price and the Purchase Method constituted a refinement of the regression model, a process that I felt was best reserved for a future analysis. Nevertheless, this statistical analysis offered a concise lens into the technicalities of price setting within a local movie theater context, demonstrating the significance in leveraging regression modeling to inform pricing decisions and consumer insights.

As this project neared its competition, progress on it was abruptly suspended due to experiencing an unplanned personal emergency. Work did not resume until late June of 2025, when this external setback began to gradually fade away, and the final direction of the project could once again take shape. As external circumstances improved, internal challenges that have already been discussed continued to permeate throughout the structure of the project. Among these are data limitations related to the scope of the dataset, outliers and anomalies recorded during data collection, a lack of meaningful price variation in some instances, and other statistical discrepancies. As mentioned before, regression modeling in this project faced major challenges in balancing complexity with interpretability, and by deciding to include most variables in the model it led to a minority of variables weighing more than others, which may constitute a grave distortion in interpreting the main hypothesis. Perhaps the most persistent challenge this project faced came from incessant learning gaps in statistical analysis, which required constant clarity and input from revisiting core concepts. Despite these setbacks, the persistent troubleshooting had the positive effect of strengthening my technical and analytical skills, and my understanding of key pricing concepts was consolidated and reinforced.

After evaluating numerous external and internal challenges, it is imperative that we look ahead into a potential future where these setbacks are addressed and accounted for. A future analysis would benefit from a larger scope and a cleaner dataset, with the goal of collecting data across seasonal ticket pricing trends with a minimal number of outliers present, and there should be efforts to resolve statistical discrepancies like selection bias before proceeding to the modeling phase. Once the regression modeling phase is reached, there should be a greater emphasis on

testing variables extensively for model compatibility, and perhaps even testing multiple regression models simultaneously. The output generated from these models must be formatted properly within RStudio which will require additional practice with statistical tools to achieve. A future analysis yields the strongest results when learning gaps in statistical analysis are managed without major interference. It is paramount that I revisit key academic concepts like econometrics and microeconomics, practice formal statistical techniques through Excel and RStudio, expand my existing understanding of pricing theory, and explore how these skills can be leveraged to produce a future analysis with actionable insights.

The purpose of this project was to demonstrate my working knowledge of pricing theory and the application of various statistical tools at my disposal. It has laid a foundation for me to achieve long-term career growth within the field of pricing analytics. From extracting raw data and constructing regression models based on the dataset, to interpreting regression results through lens of business application, each phase of this project has contributed to my ability to extract real-world insights. While numerous challenges and setbacks were often frustrating to face, I persisted in finalizing the project and used each obstacle as an opportunity to think retrospectively about future refinement. Ultimately, this experience has reaffirmed my passion for pricing strategy, taught me tangible steps that can be taken to improve my data analysis process, and has served as a measure for my readiness in contributing meaningfully within the field of pricing analytics.

APPENDIX: REFERENCES

APPENDIX: SUPPLEMENTARY MATERIALS

The following files are available in the GitHub repository and support the results in this project:

- *Movie Theater Insights Data Analysis - R Markdown (rmd)* → Contains all statistical tests, regression models, and visual plots generated using RStudio.
- *Movie Theater Insights Data Analysis - R Markdown Document (pdf/html)* → Contains all knitted output generated from the R Markdown document for easy viewing.
- *Weekly Movie Theater Characteristics from Galaxy Theatres (xlsx)* → Includes all raw data extracted from the Galaxy Theatres Website; a slightly modified version of this document was used for statistical analysis.
- *Movie Theater Insights General Excel Workbook (xlsx)* → Includes multiple tables that contain variable classification notes or regression modeling output; used for organizing and keeping track of each variable.
- *Visual Plots Used in Project (png/pdf)* → Includes grouped bar charts, interaction plots, and outputs from regression modeling.

Access the Repository Here:

https://github.com/jkshortcheese46/Weekly-Movie-Theater-Insights-Project