

SFJAZZ Live Shows Ticket Sales Data Prediction

BUS 895 Research Project in Business

San Francisco State University

Renato Perez

Table of Contents

1.	Executive Summary	3
2.	Introduction.....	3
2.1.	Literature Review	4
3.	Methodology	6
3.1.	Data Collection and Variable Construction	6
3.2.	Descriptive Analysis	8
3.3.	Regression Analysis.....	10
4.	Discussion and Conclusion	11
5.	References	13
6.	Appendix.....	14

1. Executive Summary

This research investigates the factors influencing Live Shows Ticket Sales at SFJAZZ, focusing on the institution's digital marketing campaigns. The study aims to provide a methodology to merge the marketing campaign and ticket sales data through daily aggregation and an Ordinary Least Squares (OLS) regression model designed to prognosticate ticket sales. The available dataset limits the capability of capturing the data's variability due to customers' incomplete demographic information. Based on the available dataset, the study proposes several hypotheses, including the positive effect of digital marketing campaign spending on live show ticket sales, the tendency of San Francisco residents to spend more money on ticket sales than non-residents, and the positive impact of online purchases on ticket sales. The study also suggests a diminishing return effect on ad spending, meaning that beyond a certain point, increasing ad spending will not result in a proportional increase in ticket sales. The literature review highlights the impact of the pandemic on the music industry, the importance of digital platforms, and the role of musical taste in expressing identity and forming opinions about others.

2. Introduction

SFJAZZ is a cultural institution in San Francisco, California, promoting jazz music since 1983. It is a leading venue for jazz performance, education, and advocacy, supporting established and emerging artists. The SFJAZZ Center, a state-of-the-art performance space designed specifically for jazz concerts, opened in 2013. It is a hub for jazz enthusiasts, hosting various performances, festivals, and educational programs annually.

The research investigates the factors that influence Live Shows Ticket Sales at SFJAZZ, focusing on the institution's digital marketing campaigns. Additionally, the study considers the

potential impact of various purchase attributes, such as the volume of seats purchased in a single transaction, the ratio of buyers residing in San Francisco versus those outside the city, and the subscription status of customers to SFJAZZ ATHOME streaming service, among other variables.

This study aims to provide a methodology to merge the marketing campaign and ticket sales data through daily aggregation and an Ordinary Least Squares (OLS) regression model designed to prognosticate ticket sales. The marketing campaign data is also sourced from the Meta Ads platform, while the CRM data comes from Tessitura. However, incomplete demographic information of customers in the available dataset limits the capability of capturing the variability in the data.

Based on the available dataset, the study proposes several hypotheses. A) Digital marketing campaign spending positively affects live show ticket sales. B) San Francisco residents spend more money on ticket sales than non-residents. C) Online purchases positively affect ticket sales. D) There is a diminishing return effect on ad spending, meaning that beyond a certain point, increasing ad spending will not result in a proportional increase in ticket sales.

2.1. Literature Review

The work in (Denk, Burmester, Kandziora, & Clement, 2022) found that the pandemic has significantly impacted the music industry, with a decline in music consumption and spending. The study also found that the pandemic has led to a shift in music consumption patterns, with increased streaming service subscriptions and decreased physical music sales. The authors suggest that the music industry should adapt to these changes and focus on digital platforms to survive the pandemic.

Ticket sales and live performances make up 50% of total revenues that come from live music, while recorded music is the other major income stream for the industry, combining revenues from streaming, physical sales, and licensing royalties (University of Pennsylvania, 2020). The article examines the impact of COVID-19 on the music industry through three different lenses: Streaming, Touring, and Production. The article suggests that the music industry should adapt to these changes and focus on digital platforms to survive the pandemic.

Dr. Jason Rentfrow, from the University of Cambridge, conducted a research that found that people use their musical taste to express their identity and form opinions about others (University of Cambridge, 2023). San Francisco is known for its culture of freedom of expression. Its residents may enjoy jazz music and attending live shows.

According to a research brief by Kevin F. McCarthy from RAND, the performing arts industry is undergoing a fundamental shift (McCarthy, 2001). While the commercial recorded and broadcast performing arts industry is growing increasingly concentrated globally, live performances are proliferating locally, typically in very small organizations with low operating budgets and a mix of paid and unpaid performers and staff. At the same time, a few very large nonprofit and commercial organizations are growing larger and staging ever more elaborate productions. Midsized nonprofit organizations, on the other hand, are facing the greatest difficulty in attracting enough revenues to cover their costs.

The work in (Yilmaz, Easley, & Ferguson, 2023) discusses how sports organizations can leverage analytics and technology to improve their revenue streams. The article also discusses state-of-the-art technologies and data-related issues, presents new pricing strategies, and suggests topics for future research. The implementation of data analytics should allow organizations to

capitalize on a much richer understanding of customer behavior, enabling them to offer individualized experiences and more precisely targeted strategies.

The article "Subscription and Market Research: Factors Influencing Subscription and Single-Ticket Purchases at Performing Arts Organizations" by Joanne Scheff explores the factors that impact ticket sales at performing arts organizations (Scheff, 1999). The study conducted on four major San Francisco organizations found that increasing ticket prices reduces the subscriber base and the frequency of single-ticket buyers. People tend to select less expensive seating locations as ticket prices increase. The study recommends that performing arts managers should analyze the earned revenue needed to meet internal costs while understanding the market viability of their ticket price increases.

3. Methodology

3.1. Data Collection and Variable Construction

The data used for this analysis comes from SFJAZZ and includes live show ticket sales, membership history, and digital marketing campaign data from 2020 to date. Several steps were taken in the data engineering process to construct the dataset to be fitted using an OLS linear regression.

First, the digital marketing campaign data was cleaned and aggregated daily, with a one-month lag applied to account for the time delay between ad spend and its effect on ticket sales. For example, results from digital marketing expenditure on February 1st, 2023, would be reflected as January 1st, 2023.

Next, the membership history data was filtered to include only SFJAZZ ATHOME subscribers and then joined with the ticket sales data to add information on subscription status

and level. The dataset was further refined by identifying purchases made by San Francisco residents, filtering out in-house sales (like giveaways or courtesies) and sales with values less than 1. Then, we calculated the number of seats sold per purchase. In the next step, we joined the two datasets: ad spending data and the ticket sales with membership data. Finally, we aggregated the resulting data daily.

Figure 1. Data engineering part 1

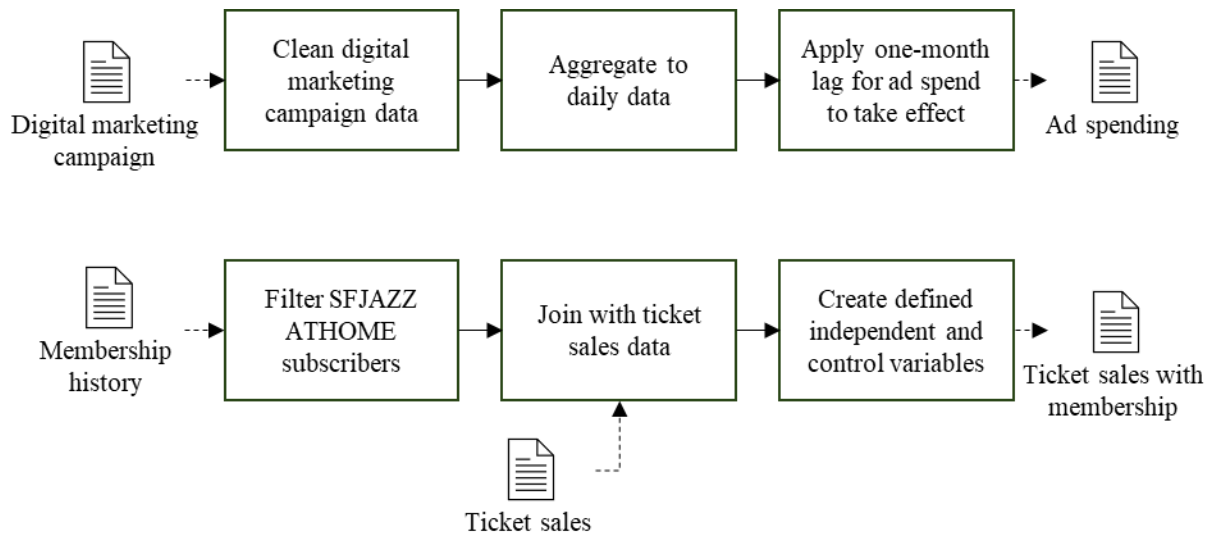


Figure 2. Data engineering part 2

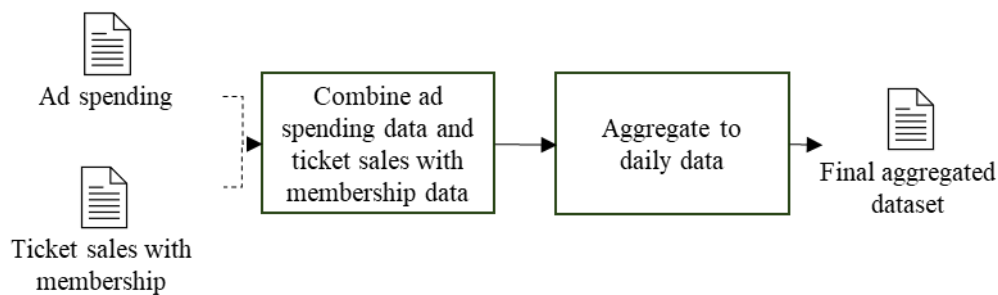


Figure 1 and Figure 2 show a flow short summarizing the steps previously detailed. The variables assembled for analysis are as follows:

- Ad Spend: (ad spend) Indicates the daily expenditure on marketing campaigns within the Meta Ads platform. The variable Ad spend2 (ad spend^2) aims to investigate potential diminishing returns from increased ad spending on Live Shows Ticket Sales.
- Subscription Level: (subscription level_n) Encoded as dummy variables, this denotes the number of purchases made for each SFJAZZ ATHOME subscription level per day.
- The percentage of online purchases ($\% \text{ online purchases} / \text{total}$) is the ratio of online purchases to the total number of purchases in a day.
- The percentage of multiple seat purchases ($\% \text{ multiple seats purchase} / \text{total}$) is the ratio of purchases involving more than one seat to the total number of daily purchases.
- The percentage of SF residents' purchases ($\% \text{ of customers living in SF}$) represents the proportion of San Francisco residents' daily purchases.
- Control Variables:
 - Quarter (quarter_n): Categorized as dummy variables, indicating the quarter in which the purchases were made.
 - Weekday or Weekend (weekday or weekend): Binary-coded variable, where '0' denotes purchases made on weekdays and '1' denotes purchases made on weekends.

3.2. Descriptive Analysis

The following in Table 1 summarizes statistics that include the control, independent, and dependent variables. One important observation to note is the frequency distribution of the subscription-level variables. The subscription levels *Corporate Annual Digital*, *Corporate One Month Digital*, and *Three Month Digital* make up a tiny portion of the total distribution. Therefore, we consider them as outliers and exclude them from further analysis. Additionally, we

have combined the subscription levels *Monthly Digital* and *One Month Digital* into a single variable called *Monthly Digital*.

Table 1. Descriptive statistics

	count	mean	min	max	median	std
ad spend	846	359.22	8.39	2425.19	250.77	378.81
ad spend^2	846	272368.20	70.39	5881547.00	62883.10	630418.50
live shows ticket sales	846	17367.54	5	768739.50	14410.59	38168.20
membership Corporate Annual Digital	846	0.00	0	1	0	0.05
membership Corporate One Month Digital	846	0.03	0	2	0	0.17
membership Monthly Digital	846	42.30	0	1126	22	86.17
membership Three Month Digital Membership	846	0.32	0	10	0	0.91
% online purchases / total	846	91.36	55.72	100	92.92	9.11
% multiple seats purchase / total	846	54.68	0	100	68.51	32.58
% customers living in SF / total	846	7.68	0	71.43	4.80	9.02
quarter 2	846	0.21	0	1	0	0.41
quarter 3	846	0.32	0	1	0	0.47
quarter 4	846	0.26	0	1	0	0.44
weekday or weekend	846	0.29	0	1	0	0.45

Base condition for subscription level variable (all subscription levels = 0) is subscription level 'Annual Digital'.

Base condition for quarter variable (all quarters = 0) is 'quarter 1'.

Rows marked in red denote variables to be removed from the analysis.

After analyzing Table 2, it was found that the highest absolute correlation value was 0.71. This value is below the commonly identified threshold of 0.8 for identifying multicollinearity issues. Therefore, it can be concluded that further inferential analysis using ordinary least-squares regression is appropriate.

Table 2. Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ad spend								
membership Monthly Digital	-0.033							
% online purchases / total	-0.171	0.121						
% multiple seats purchase / total	0.316	-0.232	-0.492					
% customers living in SF / total	0.039	0.103	-0.084	0.073				
quarter 2	-0.138	0.075	0.098	-0.233	-0.055			
quarter 3	0.117	-0.063	-0.155	0.319	0.015	-0.356		
quarter 4	0.038	-0.002	0.011	0.107	0.077	-0.308	-0.406	
weekday or weekend	-0.026	-0.021	-0.041	-0.038	-0.028	0.002	-0.003	-0.001

3.3. Regression Analysis

The OLS regression model aims to identify and quantify the impact of various factors on Live Shows Ticket Sales at SFJAZZ. The model utilizes multiple independent variables, including ad spend, different membership levels, percentages of online purchases, multiple seat purchases, customers living in San Francisco, quarterly variations, and purchases made on weekdays or weekends, to predict the dependent variable, which is Live Shows Ticket Sales.

As seen in Table 3, the regression output provides insights into the relationship between the dependent variable (live shows ticket sales) and each independent variable. Below is a breakdown of the coefficients, standard errors, t-values, and p-values associated with each predictor:

- Ad spend: This variable's p-value (0.703) indicates insignificance in predicting ticket sales.
- Membership Monthly Digital: This variable significantly positively impacts ticket sales. A one-unit increase in this membership level is associated with a considerable increase in Live Shows Ticket Sales.
- % online purchases/total: With a p-value of 0.346, this variable might not be statistically significant in predicting ticket sales.
- % multiple seats purchase/total: This variable significantly positively impacts ticket sales. When the percentage of multiple-seat purchases increases, it significantly boosts ticket sales.
- % customers living in SF / total: This variable's coefficient suggests a negative impact on ticket sales, but its p-value (0.864) implies it might not significantly affect sales.

- Quarters 2, 3, and 4 represent different quarters. However, they appear non-significant as their p-values are above the conventional threshold.
- Weekday or weekend: Its p-value (0.192) indicates a lack of statistical significance in predicting sales.

Table 3. Regression Analysis

Model 1	
Dependent Variable	live shows ticket sales (USD)
n	846
F	14.350 (0.000)***
R-squared	0.134
Adj. R-squared	0.124
ad spend	495.142 (0.352)
membership Monthly Digital	7534.639 (0.000)***
% online purchases / total	1341.813 (0.173)
% multiple seats purchase / total	13750.000 (0.000)***
% customers living in SF / total	-213.443 (0.432)
quarter 2	1067.962 (0.390)
quarter 3	1403.407 (0.352)
quarter 4	-1218.763 (0.373)
weekday or weekend	-3557.221 (0.096)
const	18030.000 (0.000)***
coefficient and/or standard errors shown as 0 have near-zero values on the order of 1E-06 or less. *p ≤ .05; **p ≤ .01; ***p ≤ .005	

4. Discussion and Conclusion

The model is significant since the F statistic's p-value is less than 0.05. In other words, the model performs better than a constant (horizontal line). Also, the model explains 13.4% of the variability in the data.

Regarding the independent variables, membership Monthly Digital is significant. An interpretation may be that the streaming service is an exposure platform where customers

discover new artists or subgenres. An actionable recommendation would be to have a free tier for the streaming service. That way, the exposure to new customers could be improved.

We have noticed that a considerable percentage of customers purchase multiple tickets for live shows. This suggests that people are more likely to attend events with companions, such as family or friends, rather than alone. Offering discounts for bulk ticket purchases may be beneficial to encourage more sales.

We have discovered that the amount of money spent on marketing campaigns is relatively small in the sales of live show tickets. Table 4 in the Appendix shows no diminishing return effect on ad spend. Similarly, online purchases have little effect on live show ticket sales. Additionally, whether a customer lives in San Francisco or not does not significantly affect live show ticket sales. Therefore, the impact on ticket sales may be due to the artist and not necessarily the location.

Based on the results, there is no seasonal effect on ticket sales. In other words, the control variables, such as whether a purchase is made on a weekday or weekend or during any quarter, do not impact the amount of money generated by live shows.

Notably, the model used in this study only accounts for 13.4% of the variation in live show ticket sales. This suggests that there may be other important factors that were not included in the model (Statology, 2019). For future research, we recommend considering the inclusion of ad spending data from Google Ads, as well as metadata related to artists or genres. Additionally, this study proposes using a daily long panel time series approach for data aggregation, which may benefit from further revision.

5. References

- Denk, J., Burmester, A., Kandziora, M., & Clement, M. (2022). *The impact of COVID-19 on music consumption and music spending*. Plos one.
- McCarthy, K. F. (2001). The Performing Arts Trends and Their Implications. *Research Briefs*.
- Scheff, J. (1999). Factors Influencing Subscription and Single-Ticket Purchases at Performing Arts Organizations. *International Journal of Arts Management*, 16-27.
- Statology. (2019, March 20). *How to Read and Interpret a Regression Table*. Retrieved from Statology: <https://www.statology.org/read-interpret-regression-table/>
- University of Cambridge. (2023, August 21). *You Are What You Listen To*. Retrieved from Research: <https://www.cam.ac.uk/research/news/you-are-what-you-listen-to>
- University of Pensilvania. (2020, August 4). *COVID-19's Impact on Music: An Analysis of the Industry Post-Lockdown*. Retrieved from WQHS Radio: <https://wqhs.upenn.edu/covid-19s-impact-on-music-an-analysis-of-the-industry-post-lockdown/>
- Yilmaz, Ö., Easley, R., & Ferguson, M. (2023). The future of sports ticketing: Technologies, data, and new strategies. *Journal of Revenue and Pricing Management*, 219-230.

6. Appendix

Table 4 - Model 2 with $ad\ spend^2$

Model 2		
Dependent Variable	live shows ticket sales (USD)	
n	846	
F	14.33	(0.000)***
R-squared	0.134	
Adj. R-squared	0.124	
ad spend ²	25.3495	(0.492)
membership Monthly Digital	7555.2628	(0.000)***
% online purchases / total	1331.2584	(0.175)
% multiple seats purchase / total	1.39E+04	(0.000)***
% customers living in SF / total	-209.2765	(0.434)
quarter 2	973.8253	(0.400)
quarter 3	1383.8978	(0.355)
quarter 4	-1249.5071	(0.370)
weekday or weekend	-3571.7077	(0.095)
const	18070.000	(0.000)***

coefficient and/or standard errors shown as 0 have near-zero values on the order of 1E-06 or less. * $p \leq .05$; ** $p \leq .01$; *** $p \leq .005$