Rhythm, Risk and Regression

Predicting Ticket Sales and Reducing Risk in Live Performance Promotion

for Malmesbury Live Arts and
Imperial College Business School

July 2025

Prepared by
Neil Muttock

Executive Summary

This project explores the potential for using machine learning to predict ticket sales for live events promoted by Malmesbury Live Arts (MLA) - a grassroots arts organisation.

By analysing data from past events covering venue details, artist popularity, ticket pricing and seasonal trends, the aim is to support promoters in making smarter, less risky decisions when planning future shows.

Three machine learning approaches were tested: Linear Regression, Polynomial Logistic Regression and Decision Trees.

Each model was trained using a curated dataset built from ticketing records, event data and artist social media metrics. Of the three, a Linear Regression model performed best with the data available, offering the most accurate and consistent predictions with a relatively low margin of error.

To evaluate it's practical use, the model was tested on four upcoming, unseen events. The results aligned closely with outcomes from comparable past shows, suggesting the model could offer real value in planning and budgeting decisions.

While this is an early-stage tool, the outcomes are promising. With further development such as including more detailed marketing data, audience insights and automating predictions via a user interface, this model could become a reliable support system for promoters, reducing financial risk and increasing the chances of successful, well-attended events.

Prepared by
Neil Muttock

For
Malmesbury Live Arts and Imperial College Business School

Introduction

At it's core, gig promotion is about bums on seats. Whether it's music, comedy, theatre or anything in between, it only works by getting people to buy tickets for a live show. The Promoter is the person (or team) who makes it happen, handling everything behind the scenes: booking the venue, setting up deals with performers and getting the word out. But no matter how great the act is, success usually comes down to one thing: ticket sales. And it's the Promoter who takes the risk.

Once the show is booked, the pressure's on to sell. Promoters use social media, flyers, email lists and ads to build hype and reach the right crowd. For well-known acts, tickets can sell fast. But for smaller names, it can be really tough. And if enough tickets don't sell, the show might not even break even, let alone make a profit.

What adds to the pressure is that Promoters often have to pay upfront for the venue, marketing, and sometimes even the artist. If people don't show up, that money's lost. Even with great planning and a good team, things like bad weather or competing events can throw everything off. In the end ticket sales are what make or break a gig, turning all the hard work into a win, or a loss.

That's why I've decided to try to create a predictive model to help the promoters within a grassroots live-arts organisation to make smarter, data-informed decisions. By analysing ticket sales from past events, a good predictive model could give accurate income forecasts for future gigs, reducing financial risk, improving planning and ultimately giving promoters a better chance of filling the space in front of the stage and making their shows successful.

Context

“bringing high quality live arts to Malmesbury and including it in the heart of the community”

Malmesbury Live Arts is a Community Interest Company run by a team of volunteers all with experience in producing live arts events.

Together they run annual programmes of gigs across various live music genres, comedy and theatre from 4 venues in the town of Malmesbury in Wiltshire, UK.

Each member of the MLA team takes on the responsibility of a single promoter brand or music genre, programming the line up, hiring the venue, organising production and promoting the events.

Promoter Brands	Venues	Genres	Annual Events
10	4	12	13-26

The Aim

The venues used by MLA are small (under 250 capacity) and so are not attractive to outside promoters. With such small venues there is limited scope for ticket sales so every event is high risk and it takes careful planning in order to keep Malmesbury Live Arts sustainable.

The aim of this project is to see if a simple model can be developed, using MLA’s history of events and ticket sales data, to make predictions of ticket sales for each new event proposed when considering new line ups. This could help with planning, budgeting and financial risk management without the organisation being fully reliant on high risk hunches.

Methods

Building, training and testing the models

For interest, the code behind the development of all the models examined in this section is available for review in the GitHub repository here:

https://github.com/creology/rhythm-risk-regression

The Data

Sources

Ticketing Platform (Eventbrite)
for all tickets issued over the last 4 years - tickets sold online, tickets sold on the door and most complimentary guest tickets

Website Content Management System
for all events staged over the past 4 years with event dates, artists, venues and ticket prices

Social media sites
for artist reach and engagement metrics

Email Marketing Platform
for additional customer details

Enrichment

Artist social media metrics
added to each event according to the Headline artists performing

Genres
added to each artist to help with categorisation

Cleansing

Unusual events (eg workshops, jam sessions)
excluded from all data sets so as not to skew the analysis and focus on live perfomance

Complimentary tickets
all removed as they do not contribute to budget

Average ticket price paid
calculated and used rather than the face value of tickets, to account for different ticket types at different prices and discount offers

Ticket Orders **3,727**	Time Period **4 years**
Event Features **10**	Event Attendees **1,840**
Events **78**	Artists **162**

Strategy

All data was extracted from the relevant sources and imported into a relational SQL database via a reusable SQL import pipeline. This allows for new data to be reimported into the database as it grows over time and so all data manipulation can automated with a set of saved SQL commands to prepare the data for analysis.

Various entities were joined and where needed, aggregated to create a single dataset for model training to ensure the data was clean, consistent and structured in a way that supported analysis and model development.

The Features

In the effort to predict ticket sales for events, there are a lot of moving parts like who the artist is, what kind of event it is, where and when it's happening, how much tickets cost and how much promotion is going to happen. These factors can affect ticket sales in all sorts of ways, sometimes in ways that aren't immediately obvious.

For this first attempt at building a predictive model, the data that's readily available was identified. Whilst there's definitely room to build a more advanced model in the future with richer, more detailed inputs, this first version focuses on what we have to hand. The features considered for use in training the models include:

Venue features

Capacity

Format
(standing or seated)

Artist features

Genre

Facebook Likes

Instagram Followers

YouTube Subscribers

Spotify Monthly Listeners

Event features

Promoter Brand

Month of the year

Average Ticket Price Paid

Initial finds

Preliminary modelling quickly revealed a relationship between venue capacity and ticket price, indicating that ticket prices tend to increase with the size of the venue.

This relationship arises because artists capable of attracting larger audiences are generally also able to command higher ticket prices and perform in larger venues. Consequently, including both venue capacity and average ticket price in the analysis was considered redundant.

Therefore, for the purposes of this task, venue capacity was excluded. In practice, accurately predicting ticket sales would be invaluable information when selecting an appropriately sized venue for any given artist.

The Models

Three different types of model were trained with MLA's data with the results and justifications for each model shown below. Each model went through a number of iterations to fine tune and optimise each one as much as possible. Here, the results from the most optimal (ie the least errors in prediction) of each model are shown.

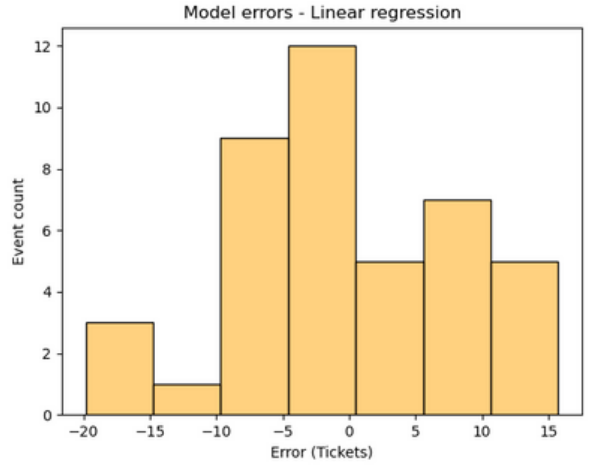


Fig. 1 - Linear regression model errors

METHOD 1 - LINEAR REGRESSION

Produces a straightforward formula that considers how each feature influences sales. Serves as a benchmark for more complex models.

	Tickets
Average Prediction Error	**7**
Max Over Prediction	**16**
Max Under Prediction	**-20**

MOST ACCURATE MODEL

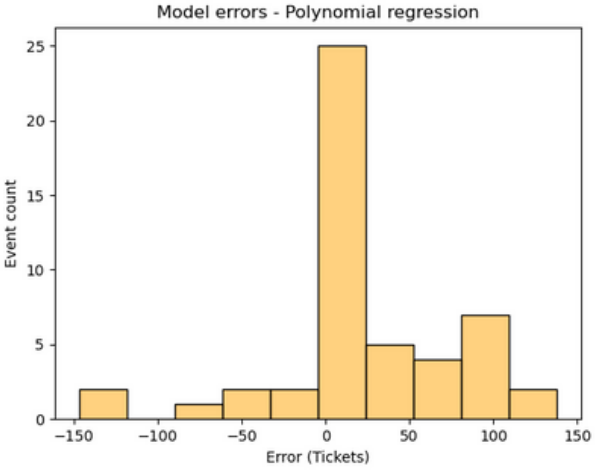


Fig. 2 - Polynomial regression model errors

METHOD 2 - POLYNOMIAL LOGISTIC REGRESSION

Captures complex interactions and can uncover relationships missed by a linear model.

	Tickets
Average Prediction Error	**39**
Max Over Prediction	**147**
Max Under Prediction	**-168**

MAYBE OVERLY COMPLEX FOR NOW

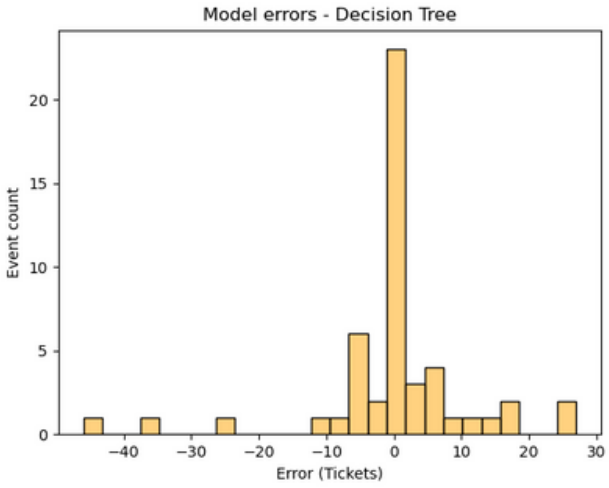


Fig. 3 - Decision tree model errors

METHOD 3 - DECISION TREES

Works well with a mix of numbers and categories and understands that certain combinations lead to different levels of sales.

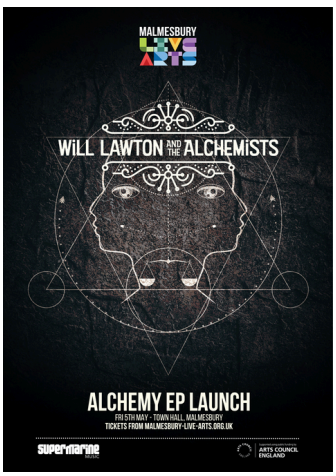
	Tickets
Average Prediction Error	**12**
Max Prediction Over	**27**
Max Prediction Under	**-46**

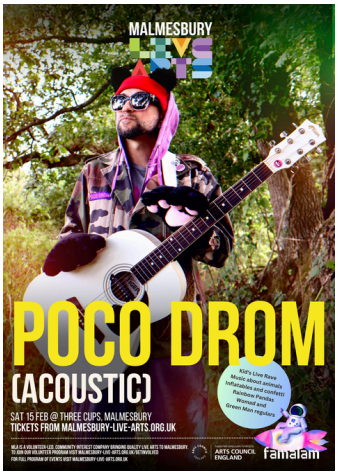
SHOWS POTENTIAL, NEEDS REFINEMENT

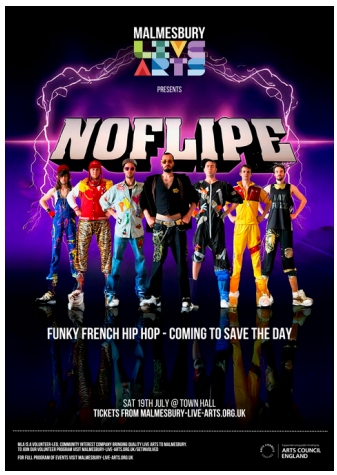
Results

Of the three models evaluated, the Linear Regression model (despite being the simplest) demonstrated the highest level of accuracy. To help assess it's potential usefulness in practice, the model was tested on four proposed events which it had not previously seen. This was done to evaluate whether the ticket sales predictions it generated appeared reasonable.

Each prediction was then compared to actual sales figures from similar past events to provide a benchmark for how accurate and reliable the model's forecasts are likely to be.







EVENT PROPOSAL

Genre: Acoustic
Brand: MAS
Month: March
Avg Price: £12.50
Performance: Evening
Format: Seated
Facebook Likes: 742
Insta Followers: 116
YouTube Subs: 331
Spotify Listeners: 12

MODEL PREDICTION
95 ticket sales

BENCHMARK
102 ticket sales

ASSESSMENT
REASONABLE

EVENT PROPOSAL

Genre: Comedy
Brand: Gloryhole
Month: December
Avg Price: £14
Performance: Evening
Format: Seated
Facebook Likes: 0
Insta Followers: 0
YouTube Subs: 0
Spotify Listeners: 0

MODEL PREDICTION
76 ticket sales

BENCHMARK
79 ticket sales

ASSESSMENT
REASONABLE

EVENT PROPOSAL

Genre: Kids' Music
Brand: Famalam
Month: September
Avg Price: £5
Performance: Matinee
Format: Standing
Facebook Likes: 5,700
Insta Followers: 2,534
YouTube Subs: 1,070
Spotify Listeners: 154

MODEL PREDICTION
51 ticket sales

BENCHMARK
58 ticket sales

ASSESSMENT
REASONABLE

EVENT PROPOSAL

Genre: Party Hip Hop
Brand: MLA
Month: June
Avg Price: £15
Performance: Evening
Format: Standing
Facebook Likes: 3,800
Insta Followers: 1,220
YouTube Subs: 1,080
Spotify Listeners: 434

MODEL PREDICTION
117 ticket sales

BENCHMARK
147 ticket sales

ASSESSMENT
REASONABLE
(and notoriously unpredictable!)

Conclusion

This project showed that even a simple predictive model can provide reasonable predictions of ticket sales, with future improvements focused on adding more features and making the tool accessible for everyday use by promoters.

Next Steps

- **Further refine the Decision Tree model** to see if accuracy can be improved compared to the Linear Regression model.
- **Incorporate additional external features** such as weather data, marketing activity or local competition to help improve model accuracy.
- **Investigate customer loyalty** and see if this has an effect on the accuracy of predictions.
- **Deploy the best model** to a simple interface that promoters can access and use when planning events.

This project set out to explore whether a simple predictive model could be developed to support promoters within Malmesbury Live Arts in making informed decisions about future events.

Using readily available data from ticketing platforms, artist metrics and internal event records, three models were developed and tested.

Of these, a Linear Regression model emerged as the most accurate, offering predictions with a relatively low margin of error. This suggests that even a basic model, when built on clean and structured data, can provide reasonable and meaningful insights into expected ticket sales.

While the results are promising, this should be viewed as a starting point rather than a final solution. To increase the model's accuracy and relevance over time, further refinement and enrichment of the dataset is needed.

By continuing to build on this foundation, MLA can reduce risk and improve planning to help maximise the success of it's grassroots live events.