

WEBPOISE INTERNSHIP PROJECT



PROJECT REPORT

Title : Event Affordability and Target Audience Segmentation

Submitted by :

Tharun G(1RV23CD405) 7th sem
B.E.CSE(Data science) RV College
of engineering
Contact : 8073911779
Email : tharung.cd23@rvce.edu.in

Name : Chandan Kumar K R(1RV23CS402)
7th sem
B.E.CSE RV College of engineering
Contact : 9380571828
Email : chandankumarkr.cs23@rvce.edu.in

Name : Prajwal N(1RV23CS412) 7th
sem
B.E.CSE RV College of engineering
Contact : 9663064308
Email : nprajwal.cs23@rvce.edu.in

Name : Harsha R B (1RV23CD401) 7th sem
B.E.CSE(Data science) RV College of
engineering
Contact : 8971996832
Email : harsharb.cd23@rvce.edu.in

Name : Somashekar A G (1RV23CS417) 7th sem
B.E.CSE RV College of engineering
Contact : 7338284871
Email : somashekarag.cs23@rvce.edu.in

ABSTRACT

This report details a comprehensive data analysis project focused on understanding **event affordability** and **target audience segmentation** for event organizers. Using a synthetic dataset of 1,000 samples containing demographic, economic, and event participation data, the project employed a multi-faceted methodology.

Key Findings and Methodology:

- **Feature Engineering** established the **Affordability Index** ($\frac{\text{Income}}{\text{Ticket Price}}$).
- **K-Means Clustering** identified four distinct customer segments, notably an **Affluent/Price-Insensitive** group (Cluster 2) with the highest average income (\$94,946) and 100% attendance, and a **Budget-Stressed** group (Cluster 3) facing high costs relative to income, resulting in the lowest attendance (77.4%).
- **Predictive Modeling** (Logistic Regression, Decision Tree) achieved an accuracy of up to **90.5%**.
- The analysis confirmed that the **Affordability Index** is the single **Most Important Factor** in predicting attendance.

Conclusion:

The primary recommendation is to adopt segment-specific pricing and marketing strategies. This includes focusing low-priced events (e.g., Festivals, \$20-\$60) on Low Budget segments and utilizing premium/VIP packages for High Budget, price-insensitive customers. Strategic pricing should aim for an Income/Ticket_Price ratio above 200 to mitigate attendance risk.

TABLE OF CONTENTS

Section	Title	Page
1.	INTRODUCTION	1
1.1.	Dataset Description	1
1.2.	Purpose	1
1.3.	Objectives	2
1.4.	Exploratory Data Analysis (EDA)	2
1.5.	Characteristic Relations	3
2.	METHODOLOGY	5
3.	ANALYSIS	7
3.1.	Cluster Analysis	7
4.	RESULT AND CONCLUSION	9
4.1.	Predictive Model Performance	9
4.1.1.	The feature importance analysis (Decision Tree) revealed:	9
4.2.	Conclusion and Strategic Recommendations	10

LIST OF FIGURES

Figure Number	Title	Page
Figure 1.1	Distribution of Key Demographic and Event Variables	2
Figure 1.2	Gender and Overall Event Attendance Distribution	3
Figure 1.3	The Relationship Between Income, Ticket Price, and Attendance	4
Figure 1.4	Event Type Popularity Across Different Age Groups	4
Figure 2.1	Elbow Method for Determining Optimal Number of Clusters (k)	5
Figure 4.1	Customer Segments (K-Means Clusters) based on Financial Metrics	8
Figure 4.2	Event Type Preferences by Cluster (%)	8
Figure 4.3	Classification Model Performance: Confusion Matrices	8
Figure 5.1	Event Type Preference by Budget Category	11

INTRODUCTION

1.1. Dataset Description

The analysis utilizes a **synthetic dataset** comprising 1,000 samples. This dataset simulates event participation, demographic, and economic data necessary for a comprehensive affordability analysis.

Variable	Data Type	Description
Age	Integer	Customer age (18 to 65 years)
Gender	Categorical	Male, Female, Other
Income	Integer	Annual income (\$20,000 to \$200,000)
Event_Type	Categorical	Concert, Sports, Festival, Seminar, Exhibition
Ticket_Price	Float	Price of the event ticket
Attendance	Binary	Target Variable (0=Did Not Attend, 1=Attended)

1.2. Purpose

The primary purpose of this project is to provide **data-driven strategic recommendations** for event organizers. By dissecting customer affordability and preference drivers, the goal is to enhance event targeting, optimize pricing, and improve overall attendance rates.

1.3. Objectives

The core objectives of the analysis are to:

- **Identify** which customer segments can afford premium, mid-range, or budget events.
- **Determine** the preferences of different audience segments for specific event types.
- **Develop** predictive models to forecast event attendance.
- **Suggest** effective event targeting strategies grounded in affordability and preferences.

1.4. Exploratory Data Analysis (EDA)

The initial exploration confirmed the dataset's structure and quality, with no missing values.

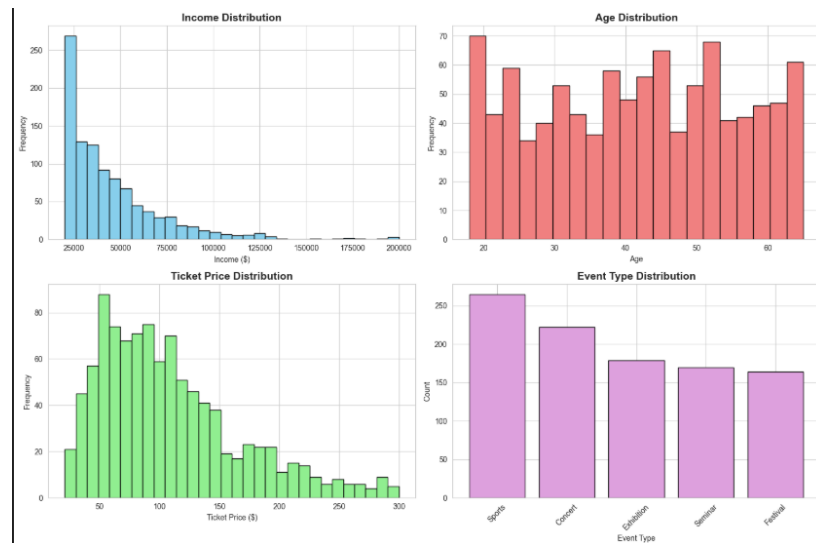


Figure 1.1: Distribution of Key Demographic and Event Variables

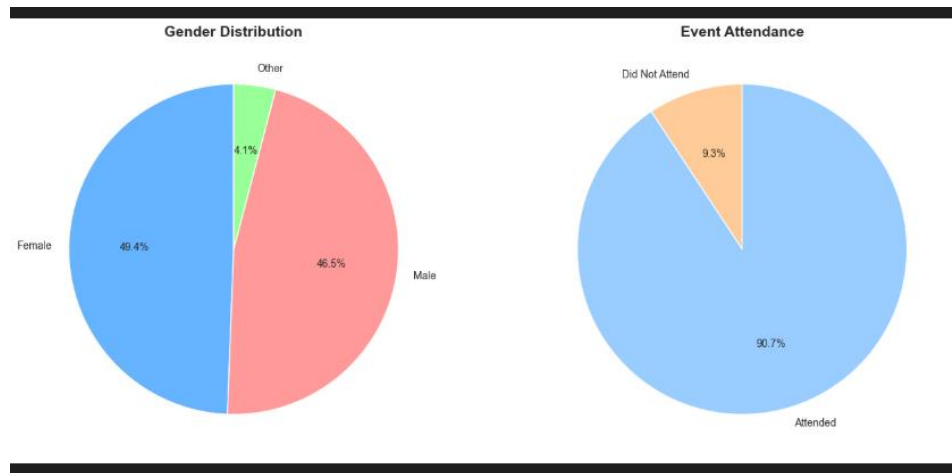


Figure 1.2: Gender and Overall Event Attendance Distribution

- **Mean Income:** \$44,816.
- **Mean Ticket Price:** \$108.59.
- **Attendance Distribution:** The dataset shows a high overall attendance rate of **90.7%** (907 attended, 93 did not attend).
- **Event Type Distribution:** **Sports** events were the most frequent (265), followed by **Concerts** (222).

1.5. Characteristic Relations

Initial visualizations highlighted key relationships (See Figure 3):

- **Income vs. Price:** Attendance is clearly correlated with the consumer's income level relative to the ticket price. This relationship is formalized in the **Affordability Index**.

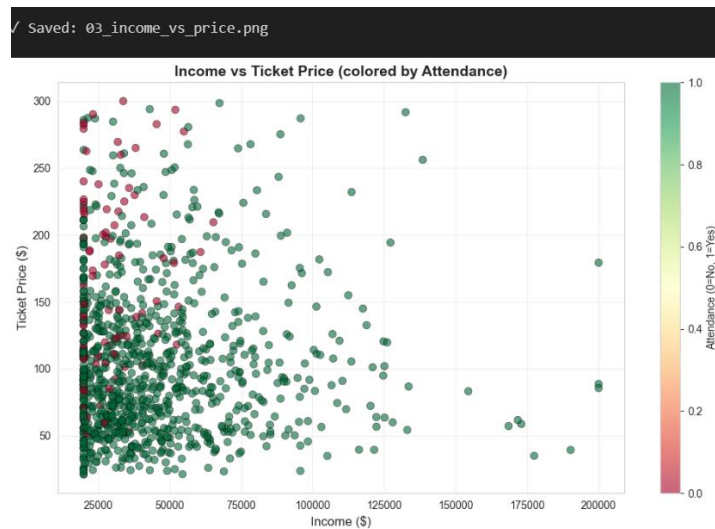


Figure 1.3: The Relationship Between Income, Ticket Price, and Attendance

- **Event Type vs. Age:** There is a visible correlation between age and preferred event type:

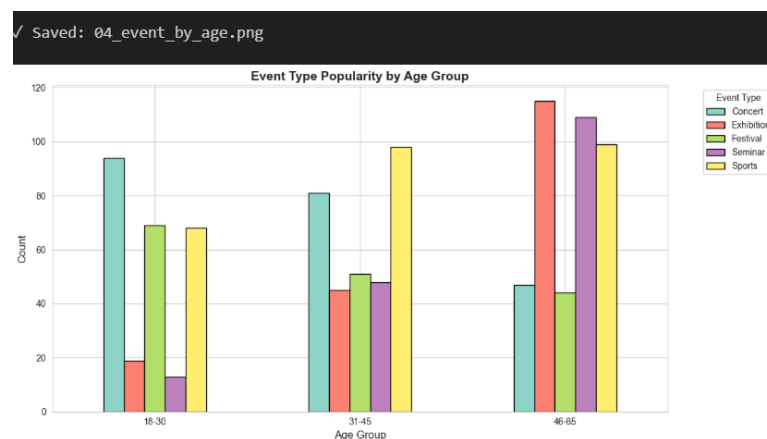


Figure 1.4: Event Type Popularity Across Different Age Groups

- Younger segments (**18-30**) prefer **Concerts** and **Festivals**.
- Older segments (**46-65**) prefer **Seminars** and **Exhibitions**.
- Middle-aged segments (**31-45**) show a mixed preference, slightly favoring **Sports**.

METHODOLOGY

The project employed a standard data science methodology combining feature engineering, unsupervised learning for segmentation, and supervised learning for prediction.

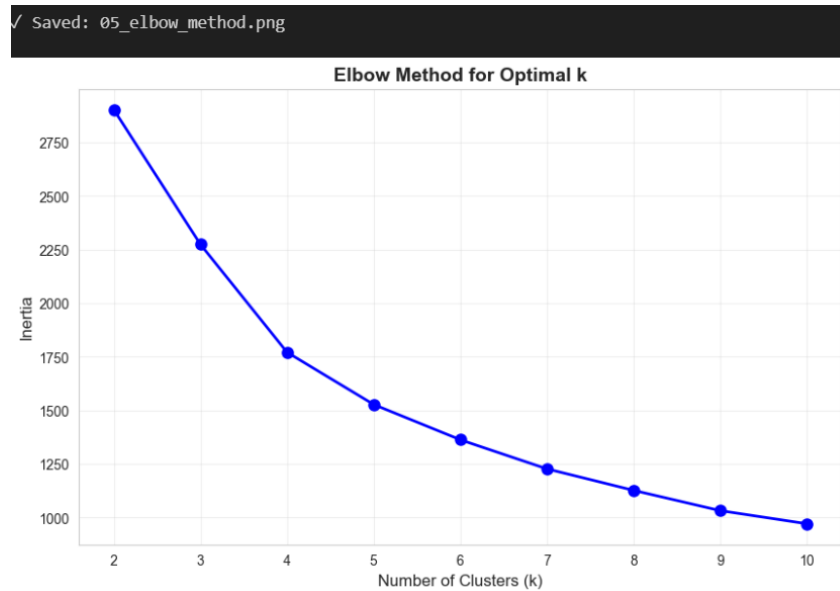


Figure 2.1: Elbow Method for Determining Optimal Number of Clusters (k)

Phase	Technique	Purpose
Feature Engineering	Creation of Affordability Index (Income / Ticket Price).	Quantify price sensitivity and its impact on attendance.
	Creation of Budget Categories (Low, Medium, High).	Segment customers based on income terciles.
Segmentation	K-Means Clustering.	Group customers into distinct, actionable segments based on financial and demographic data.
Prediction	Logistic Regression and Decision Tree Classifier.	Predict the likelihood of attendance and identify the primary drivers.

ANALYSIS

3.1. Cluster Analysis

The K-Means algorithm identified four distinct customer clusters.

Cluster	Description	Key Characteristic	Event Preferences (%)
Cluster 0	Mature Middle-Income	Avg. Age 53.1 , Avg. Income \$36.4K, High Attendance (94.5%)	Exhibition (32.8), Sports (30.5), Concert (12.9)
Cluster 1	Young Middle-Income	Avg. Age 28.4 , Avg. Income \$37.0K, High Attendance (91.1%)	Sports (33.6), Concert (28.9), Festival (24.1)
Cluster 2	Affluent/Price-Insensitive	Avg. Income \$94.9K , Highest Affordability (1555.2) , 100% Attendance	Sports (31.4), Festival (28.9), Exhibition (23.1)
Cluster 3	Budget-Stressed	Avg. Ticket Price \$204.18 , Lowest Affordability (212.1) , Lowest Attendance (77.4%)	Seminar (62.6%) , Concert (33.3), Sports (4.1)

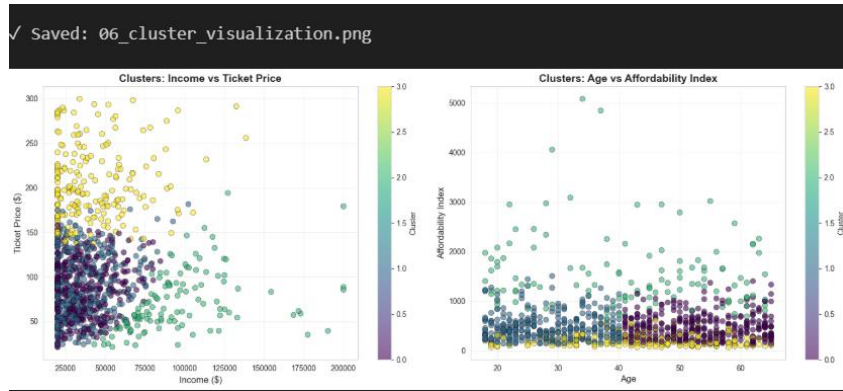


Figure 4.1: Customer Segments (K-Means Clusters) based on Financial Metrics

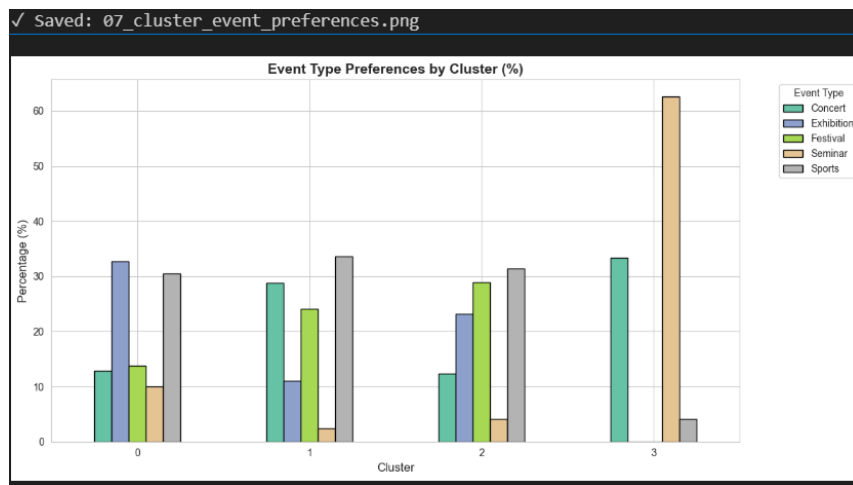


Figure 4.1: Customer Segments (K-Means Clusters) based on Financial Metrics

- **Observation:** Cluster 3 has the lowest attendance, suggesting that events like Seminars, which are typically high-priced, are being purchased by customers whose income-to-price ratio is low, indicating a financial strain that may lead to non-attendance or last-minute cancellation.

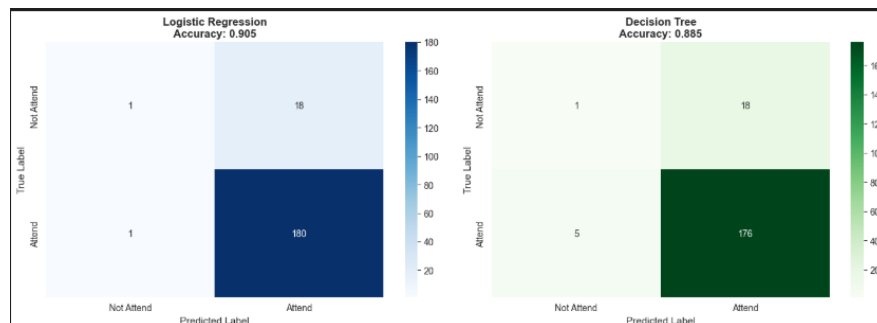


Figure 4.3: Classification Model Performance: Confusion Matrices

RESULT AND CONCLUSION

4.1. Predictive Model Performance

The Logistic Regression model was slightly more accurate (90.5%) than the Decision Tree (88.5%). However, the **Decision Tree's feature importance** provided the clearest insights into attendance drivers.

5.1.1. The feature importance analysis (Decision Tree) revealed:

- **Primary Factor: Affordability_Index** (57% Importance).
- **Secondary Factor: Income** (22% Importance).
- **Tertiary Factor: Age** (13% Importance).
- Gender_Encoded and Event_Type_Encoded had negligible direct importance, suggesting their influence is captured indirectly through pricing and age correlation

4.2. Conclusion and Strategic Recommendations

The analysis confirms that financial affordability (the ratio of income to price) is the single most important factor determining event attendance. Successful event targeting requires aligning pricing and event type with specific budget segments and age groups.

Segment	Strategy	Pricing Range
Low Budget (\$0-\$40K)	Volume Focus: Use group discounts and early bird pricing. Target with low-priced events like Festivals and Community Events .	\$20 - \$60
Medium Budget (~\$40K-\$60K)	Loyalty Focus: Use tiered pricing and package deals. Target with Concerts, Sports Events, and Exhibitions .	\$50 - \$120
High Budget (\$60K+)	Premium Focus: Offer exclusive experiences, VIP packages , and networking opportunities. Target with Premium Seminars and VIP Concerts .	\$100 - \$300+

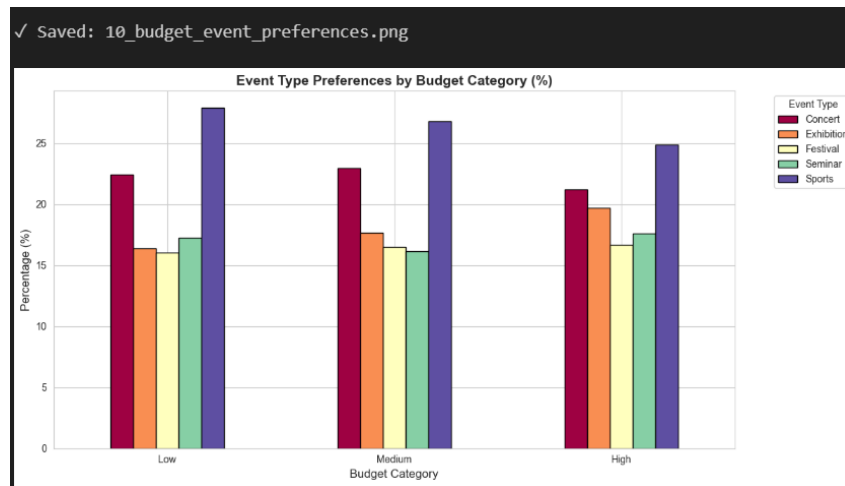


Figure 5.1: Event Type Preference by Budget Category

- **Pricing Optimization:** An Affordability Index between 200 and 500 appears to be optimal for maximizing attendance. Prices resulting in a ratio below 100 should be avoided due to the high risk of non-attendance.
- **Operational:** Implement dynamic pricing based on these demand segments and use the predictive models to forecast attendance and optimize inventory.