Demographic and Time-Dependent Factors to Houston Grand Opera Donor Conversions

The Operagoers: Yurie Han, Gal Kadmon, Amy Lam, Yi Pan, Junyao Ren, Janet Fu, & Dr. Xinjie Lan | Rice University & HGO

Introduction & Objectives

The Houston Grand Opera is a premier opera house bringing performances to thousands of viewers each year.

Donations drive roughly 80% of HGO's annual revenue, with only 20% being accounted for by ticket sales and subscriptions.

Goal	How	Impact
Help HGO convert and retain customers into donors	Identify key short and long term factors that influence donations	Enable HGO to strategize effectively, maximizing revenue

Dataset

We combined data about HGO customers from **4 different channels** into one dataset, united by each customer's unique ID number. The final dataset contains **1500 rows** (1 row = 1 customer) and **46 columns**, with data collected between 2010-2024.

- Surveys
- Ticketing
- Donations
- Marketing

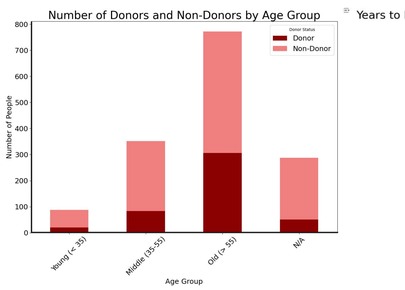
Final Dataset

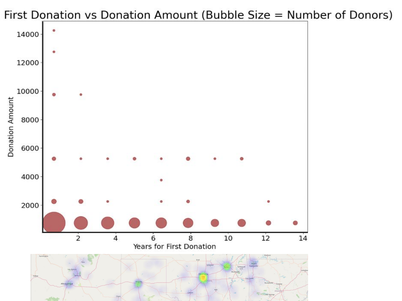
- 460 donors & 1039 non-donors
- Customer demographics
- Customer feedback
- Attendance history
- Donation dates and amounts

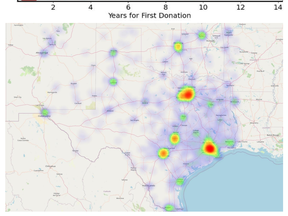
Data Science Pipeline

Data Wrangling & Data Cleaning	Exploratory Data Analysis	Modeling	Validation & Interpretation
Merge datasets & engineer new features	Identify features to help select input variables	Model data by machine learning models (survival analysis)	Check the accuracy and performance of the models

Data Exploration: HGO Donors



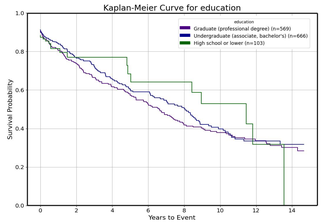


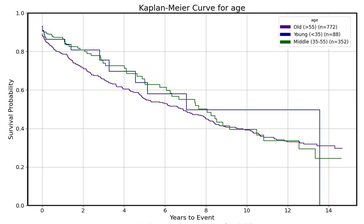


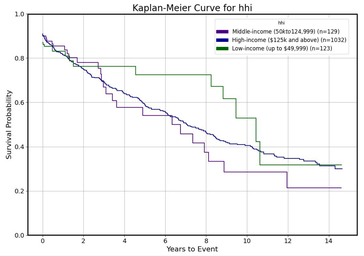
Donor trends from data exploration:

- **Older patrons** contribute more donations than other age groups
- Most donations are made within the **first year** after attending a show
- Donors primarily come from **Houston** and big cities in Texas

Data Exploration: Impact of Customer Features on Time-to-Donation







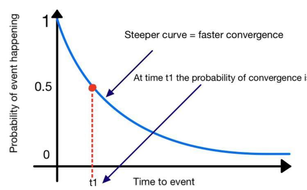
Donor trends from data exploration

- **High-income** patrons (125K+) are most likely to make donations over time
- **Older** (age 55+) patrons are most likely to make donations over time
- **Highly educated** (bachelor's or graduate degree) patrons are most likely to make donations over time

Methodology

Utilizing **survival analysis** tools to predict time-to-event based on certain features

- Kaplan-Meier (KM) Curve: Used to show the probability of an event happening over time for different groups.
- Cox Regression: Identifies how specific factors affect the timing of an event while accounting for other variables.
- Random Survival Forest: Uses machine learning to predict survival outcomes and identify important factors influencing the time to an event.



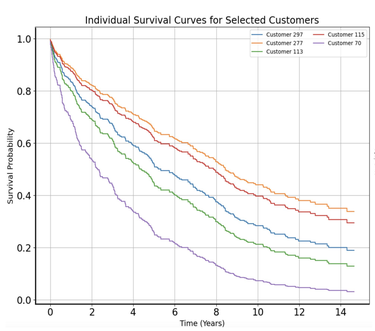
$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}.$$

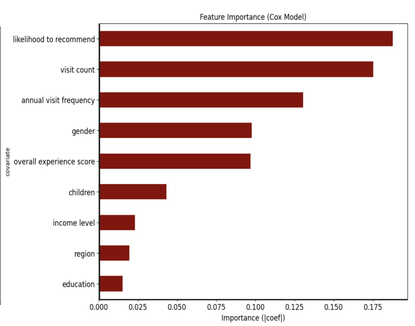
Hazard function (rate at which event is happening at time t)

$$S(t) = P(T > t) = 1 - P(T \le t) = 1 - F(t)$$

Survival function (probability that the event does not occur before time t)

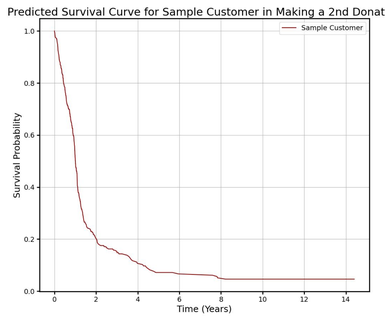
Survival Analysis: Making a 1st Donation





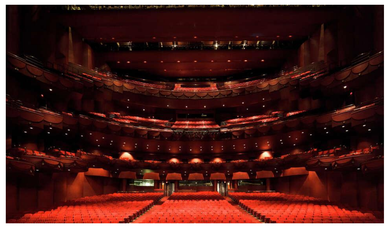
- For new customers, a survival curve can be created, estimating time to first donation
- Key features which highly influence the probability to donate over time could be found based on the 'coef'

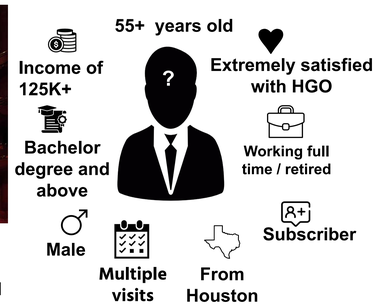
Modeling Donation Likelihood Through Survival Analysis: 1st to 2nd Donation



- Next step: investigate factors that influence donor loyalty and repeat donations
- Model performance: **Concordance index = 61%.**
- Feature selection algorithm optimizes model performance
- Predicts likelihood of second donation for any sample of customer features
- Identifies ideal donor profile by maximizing hazard ratios, or minimizing survival time
- Offers both predictive capabilities and strategic outreach insights for loyalty

Conclusion & Discussion





Short-term marketing towards HGO's ideal potential donor:

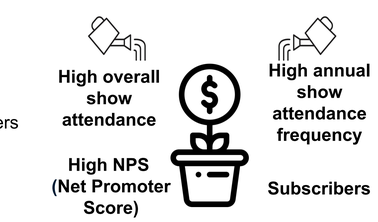
- Targeting key demographic features

Long term strategic planning:

- Investing in key factors that influence donations

Our recommendations to HGO:

- Tailored email marketing towards customers with certain features
- Focusing efforts on boosting subscriber perks
- Encouraging word of mouth promotion



References

- [1] Kuhajda, D. "Using survival analysis to evaluate medical equipment battery life," Biomedical Instrumentation & Technology, vol. 50, no. 3, pp. 184–189, 2016. [Online]. Available: https://array.aami.org/doi/epdf/10.2345/0899-8205-50.3.184
- [2] Jung, S.-H., Lee, H. Y., and Chow, S.-C. "Statistical methods for conditional survival analysis," 2018.[Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6195126/