HOW MUCH IS YOUR CONCERT TICKET REALLY WORTH?

YUSUF AKTAN

1. Abstract

Changes in technology and consumer spending habits have dramatically increased the importance of live events for the music industry over the past twenty years. One way tour promoters can measure consumer willingness to pay for concert tickets is by looking at the resale ticket market and analyzing how much markups differ from face value prices. This study collected data on event information and face value and resale ticket prices from Ticketmaster, Stubhub, and SeatGeek, and artist popularity information from Spotify in order to gauge the most important factors in determining markups. A variety of linear regression and classification machine learning methods were used to find an algorithm that could accurately predict resale markups. A LASSO linear regression algorithm was chosen with log-transformed ticket markups due to its accuracy and ease of interpretation. With an average markup of 158% in the United States, it appears promoters have significant opportunity to raise ticket prices, especially in the state of New York.

2. Introduction

Since the turn of the millenium, the music industry has been beset by online pirating, shrunken revenues, and feuds between artists and tech industry behemoths like Apple and Spotify. Revenue for recorded music has stabilized since 2015,[1] thanks to the mass adoption of paid music streaming services, but such a reversal is only part of the music industry's turnaround.

Instead, consumers have shifted away from album purchases and song downloads, and towards live events and days-long festivals. Industry experts cite two reasons for this disruption. First, at ten dollars a month for an Apple Music or Spotify subscription, recorded music is increasingly seen as a commodity that can be consumed anywhere, anytime, and with virtually no limits. Secondly, the spending habits of Millennials have repeatedly shown that they prefer shelling out money for experiences instead of possessions.[2]

---

[1] https://www.forbes.com/sites/hughmcintyre/2016/03/22/the-recorded-music-industry-actually-grew-in-terms-of-revenue-in-2015/#3b64b7787ea3

[2] https://www.billboard.com/articles/events/year-in-music-2016/7616524/concert-touring-business-2016

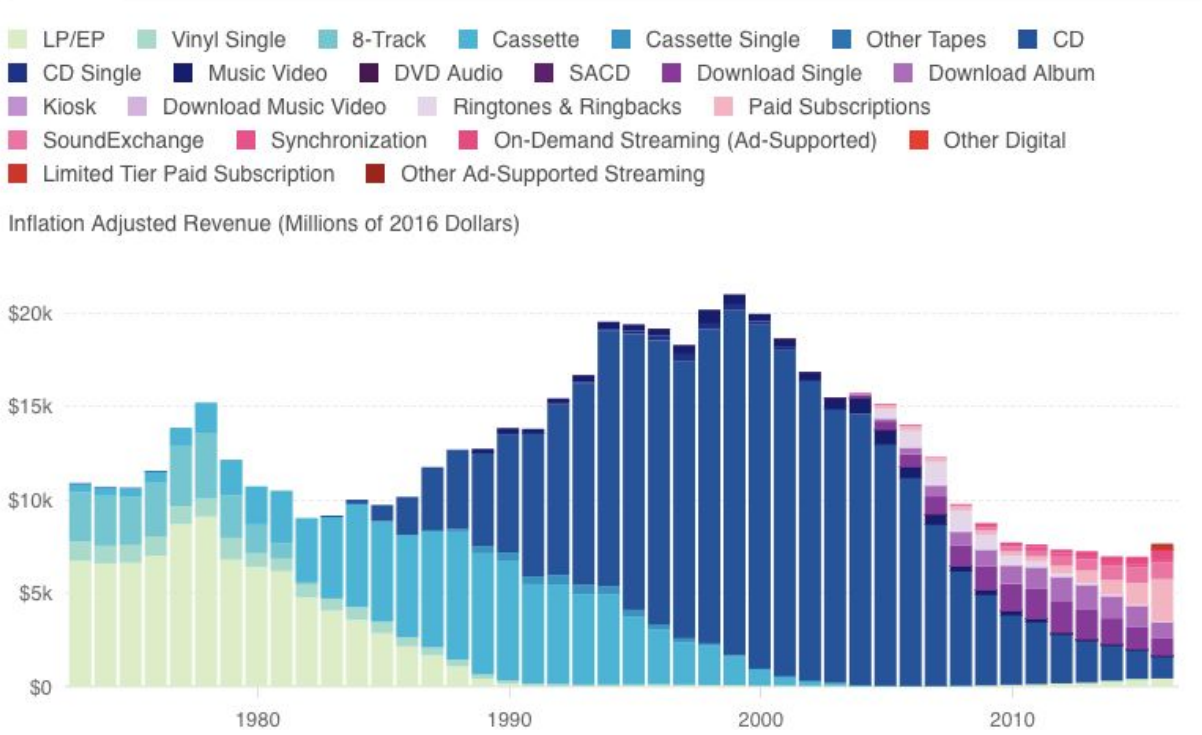


Figure 1. Recorded music revenue by year. Revenues collapsed in the new millennium but have since stabilized thanks to streaming services. Source: RIAA (Interactive version of graph available in footnote link)[3]

In 2016, Live Nation reported a 30% year over year growth in box office revenue and attendance figures. Other promoters have reported similar numbers, and estimated that live event revenues will grow 50% over the next ten years. While the tech giants fight over the future of recorded music, artists and tour promoters have found a bright spot in music festivals and live events.

[3] https://www.riaa.com/u-s-sales-database/

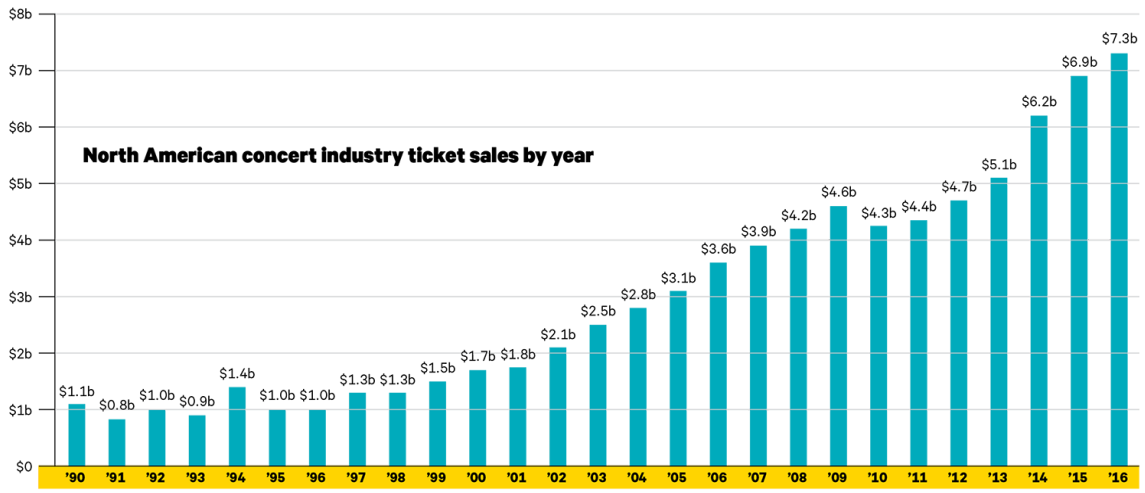


Figure 2. Concert ticket sales by year in billions. Since 2011, revenue has spiked expected to grow another 50% over the next decade. Source: Pollstar via *Variety*[4]

However, industry players must still tread carefully. Miscalculations over pricing and lineups have led to some high profile failures as well. In 2016, attendance at the music festival Bonnaroo decreased by 28,000 people from the year before after a price bump, setting an all time attendance low for the event.[5] Marc Gieger, head of the music division at William Morris notes that "All of the success is driven by the artist-fan connection, and it's the industry's job to make tickets accessible, price them correctly, make the experience in the venue first class and not block the ability for the fan and artist to interact.[6]" With so many live music options for consumers, it is important that promoters get the formula right.

As a concert and festival aficionado, I can personally attest to the chaos that the industry is experiencing. I've snagged tickets off of Facebook and Stubhub for events that I thought should have been sold out and marked up for double the original price, as well as found myself with quite ordinary tickets that everyone else wanted. Value is everything to many live event consumers, and one of the best indications of the value of an event ticket is how it fares on the resale market. High markups can suggest demand is greater than supply, or that promoters priced an event to low. On the other side, low, or even negative markups can indicate tepid demand for an artist or genre, or a lineup not worth the price. This study analyzes the resale concert ticket market as of

[4] http://variety.com/2017/music/features/live-nation-concert-business-1201979571/

[5] https://www.tennessean.com/story/money/2016/07/14/bonnaroos-ticket-sales-plummet-28000/87003736/

[6] https://www.billboard.com/articles/events/year-in-music-2016/7616524/concert-touring-business-2016

December 7 2017, for events occurring between December 2017 and May 2018, in order to determine what factors are most important in determining markups on the resale market and predict markups. Tour promoters such as Live Nation are suitable clients for this analysis to help them understand how to gage demand and pricing for events. Ticket flippers and live music consumers could also benefit from this study by identifying events that are under or overvalued.

2. Methods

This entire project was completed using Python 3.6 and supplementary packages. Review the requirements.txt file for a detailed list of packages and dependencies used.

2.1. Data Gathering

Data was gathered from four sources: Ticketmaster, SeatGeek, Stubhub, and Spotify via their publically accessible APIs. Information on events such as their dates, artists, minimum and maximum ticket prices, and ticket sale and presale dates was gathered from Ticketmaster. Minimum and maximum secondary market price data was then captured from Stubhub and SeatGeek.

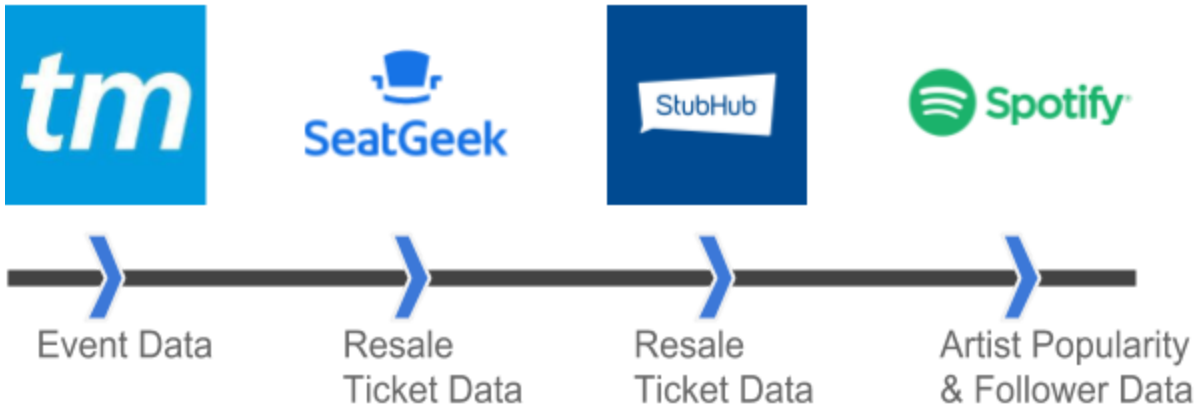


Figure 3. Data collection process, from initial event data and face value prices, to resale prices and information, and artist data from Spotify

Event listings from Ticketmaster, SeatGeek, and Stubhub were matched by joining them on date and venue name. SeatGeek's FuzzyWuzzy package was used to match events that were listed under variations of names for the same venue. Finally, a list of all performing artists was compiled and fed into the Spotify API, in order to collect the Spotify popularity index and number of followers for each artist.

2.2 Data Wrangling

The data cleaning process began by dropping duplicates, imputing missing values, and creating new features. Duplicated listings, and those for parking passes were removed, leaving about 5000 events in the dataset. In cases where Spotify information was missing, the values were imputed with the median for the genre. A new feature was created to note for a given event, how many artists were missing information on Spotify. Several new features were engineered from the data, such as length (in days) of a presale, number of days tickets have been on sale (as of December 7, 2017), the day of the week of an event, number of artists performing, average resale minimum and maximum prices (If an event was listed on both Stubhub and SeatGeek), listing source (To indicate if resale price data was from Stubhub, SeatGeek, or both), the average number of resale ticket listings, and minimum and maximum markups.

Once the the dataset was whole, the next steps were initial visualizations and outlier detection. Ticket prices and markups were found to follow a logarithmic distribution, which appeared roughly normal after a log transformation.

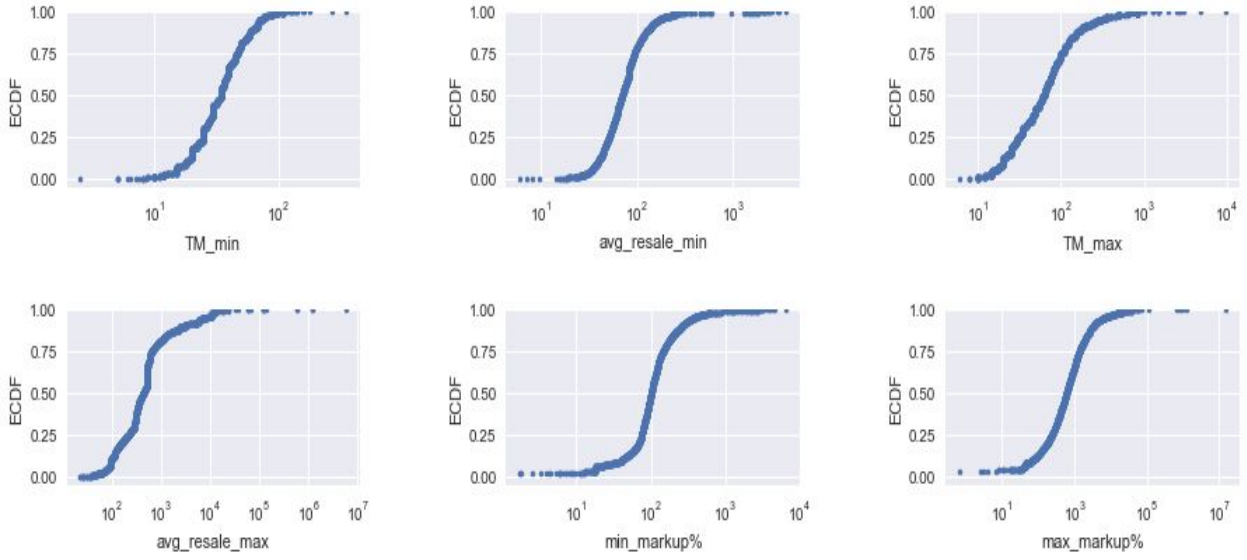


Figure 4. CDF plots of Ticketmaster ticket prices (as TM), resale prices, and markups as a percent of the Ticketmaster prices, with log-transformed Y scales.

The Tukey method was used to detect outlying ticket prices (3*IQR) for both minimum and maximum prices. The maximum ticket prices in particular had a great deal of variance, with over 12% of all events having maximum prices that the Tukey method

determined were outliers. I decided to focus the analysis on minimum ticket prices (comparing minimum face value prices from ticket prices to minimum listing prices on Stubhub and SeatGeek). Events in sparse categories with less than 50 samples were also reorganized. Events in states with less than 50 total events were dropped, and genres, subgenres, and promoters with less than 50 samples were recategorized as “Other”. All events occurring after May 2018 were dropped as well.

2.3 Data Exploration

Features were plotted for visual inspection of trends. The Pearson R and Spearman R values were computed for each of the continuous features against markup in order to evaluate the statistical significance of any correlations. ANOVA and Paired Tukey tests were used for categorical features to evaluate the significance of factors such as genre and state of events.

Features were also plotted against each other for intuitive visual analyses of their impact on markup.

2.4 Machine Learning

Machine learning was approached in several ways, from linear regression methods to predict actual markup values, to creating equal-sized bins and experimenting with classification methods from logistic regression to random forests and neural nets. After basic data preprocessing, the various methods described were experimented with using SciKit Learn, and Keras and Tensorflow.

Data was preprocessed by one-hot-encoding categorical features such as venue state, promoter, ticket source and day of the week. The final preprocessed dataset consisted of 74 features and about 3700 events. Continuous features were then normalized on a -1 to 1 scale with standard deviation of 1, and divided into independent and dependent variables. A copy of the dependent variable was made and log transformed, with all zero and negative values replaced with .01. Data was split into training and test sets. Models were evaluated using 5-fold cross validation.

Various linear regression methods were used, using both the standard dependent variable array as well as the log transformed one to compare performance.

Model	R^2 Mean	R^2 Std	MSE Mean	MSE Std	MAE Mean	MAE Std	Coefficients Count

Standard	0.1895	0.0941	84555.4864	22199.238489	136.241328	3.967828	74.0
ElasticNet	0.1615	0.0587	87752.3323	22725.8311	134.3831	4.2068	73.0
Lasso	0.1909	0.0893	84493.9095	22226.6698	135.8466	3.9690	65.0
Ada Boosting	0.0079	0.5919	107593.5526	80899.2701	260.6001	96.3962	NaN
Gradient Boosting	0.6355	0.1722	39072.1826	23130.5875	71.8796	5.2809	NaN
Random Forest	0.5875	0.1970	43085.1353	23121.9424	71.9831	5.1459	NaN
Standard Log	0.1672	0.0463	1.7560	0.4081	0.6939	0.0382	74.0
ElasticNet Log	0.1672	0.0463	1.7560	0.4081	0.6939	0.0382	74.0
Lasso Log	0.1407	0.0304	1.8068	0.3875	0.6741	0.0356	14.0
Ada Boosting Log	-0.5701	0.5715	3.2501	1.2222	1.2413	0.3313	NaN
Gradient Boosting Log	0.2930	0.1233	1.4650	0.2773	0.5762	0.0231	NaN
Random Forest Log	0.2831	0.1302	1.4965	0.3489	0.5465	0.0287	NaN

Table 1. Linear regression performance metrics on the training dataset.

Classification methods were also evaluated after transforming the dependent variable data into three equal sized bins.

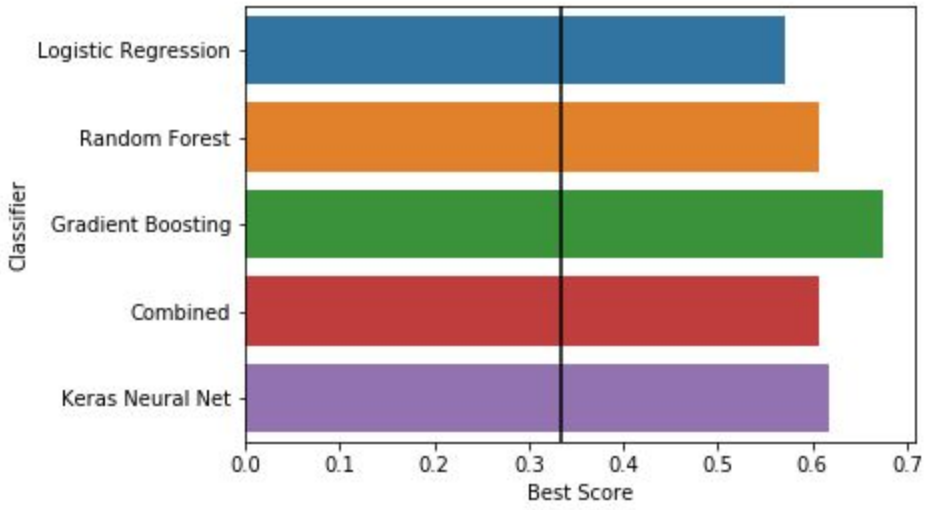


Figure 5. Accuracy scores of various classification methods after hyperparameter tuning on the test set. “Combined” is a voting classifier comprised of random forest and gradient boosting.

3. Results

3.1 Data Story

The relationships between categorical and continuous features and markup were analyzed both visually, and using statistical tests to determine their significance. Since markup data, and some of the features followed a logarithmic distribution, Spearman R could be more indicative of a relationship than Pearson R.

Feature	Pearson R	Pearson R p-value	Pearson R Null Hypothesis	Spearman R	Spearman R p-value	Spearman R Null Hypothesis
Ticketmaster Minimum Price	0.011576	4.781360e-01	Accept	-0.438978	8.478316e-177	Reject

Average Ticket Listings/Event	-0.094	0.0000	Reject	-0.199	0.0000	Reject
Average Artist Spotify Followers/Event	0.049	0.0022	Reject	-0.092	0.0000	Reject
Average Artist Spotify Popularity Index/Event	0.049	0.0023	Reject	-0.091	0.0000	Reject
Length of Presale (In Days)	-0.077	0.000002	Reject	-0.18	0.0000	Reject
Number of Days on Sale	-0.045	0.0051	Reject	-0.167	0.0000	Reject
Days until Show	-0.073	0.000006	Reject	0.013	0.4222	Accept
Number of Performing Artists/Event	0.01	0.5094	Accept	0.0103	0.5254	Accept

Table 2. Pearson R and Spearman R correlations for continuous features and markup. The Ticketmaster minimum price, number of days until a show, and the number of artists per event do not have statistically significant correlations at the .05 alpha level.

I found it quite interesting that the Ticketmaster minimum price features is only statistically significant under the Spearman R calculation. There are slight negative correlations between ticket markup and average number of tickets listed, presale length, and days on sale. Interestingly, different methods of calculating correlation yielded negative and positive correlations between Spotify popularity/followers and markup. The number of days until a show also had opposing positive and negative correlations, and may not be statistically significant enough to have an influence. The number of artists in a show does not seem correlated to markup. Pop music drove the highest markups, and Latin music yielded the lowest (By subgenre, which I found to be more relevant than the genre listed by Ticketmaster).

Figure 6. Markups by Event Day of the Week. There isn't a clear trend between markup and day of the week, or weekday versus weekend.

There also interestingly wasn't a clear visual pattern between day of the week and markup, though I would have intuitively expected weekend events to have higher markups. Additionally, events with large tour promoters had much lower markups than those that were self-promoted by the venue, or promoted by smaller companies like Masquerade, which was also surprising as large promoters like Live Nation tend to work with top headlining artists.

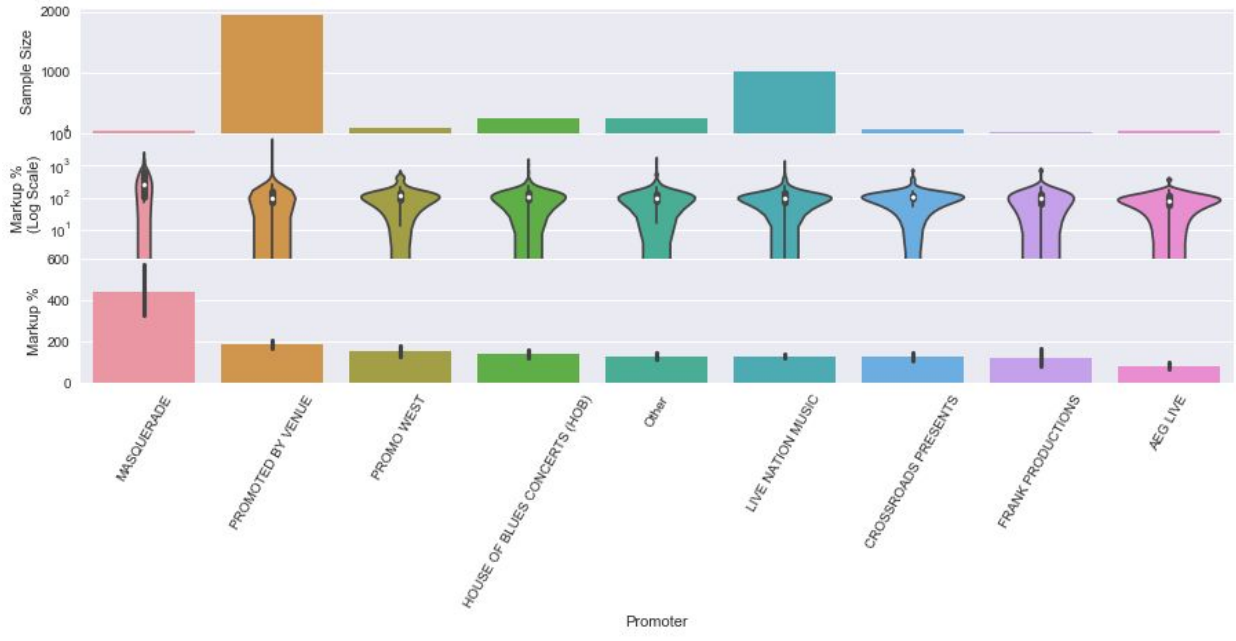


Figure 7. Ticket Markups by Promoter. Large promoter companies like Live Nation and AEG Live had much lower markups compared to events by Masquerade, and venue-promoted events.

The connection between Spotify data, subgenre, and markup is also murky, as latin music was the most popular Spotify, but commanded the lowest markups.

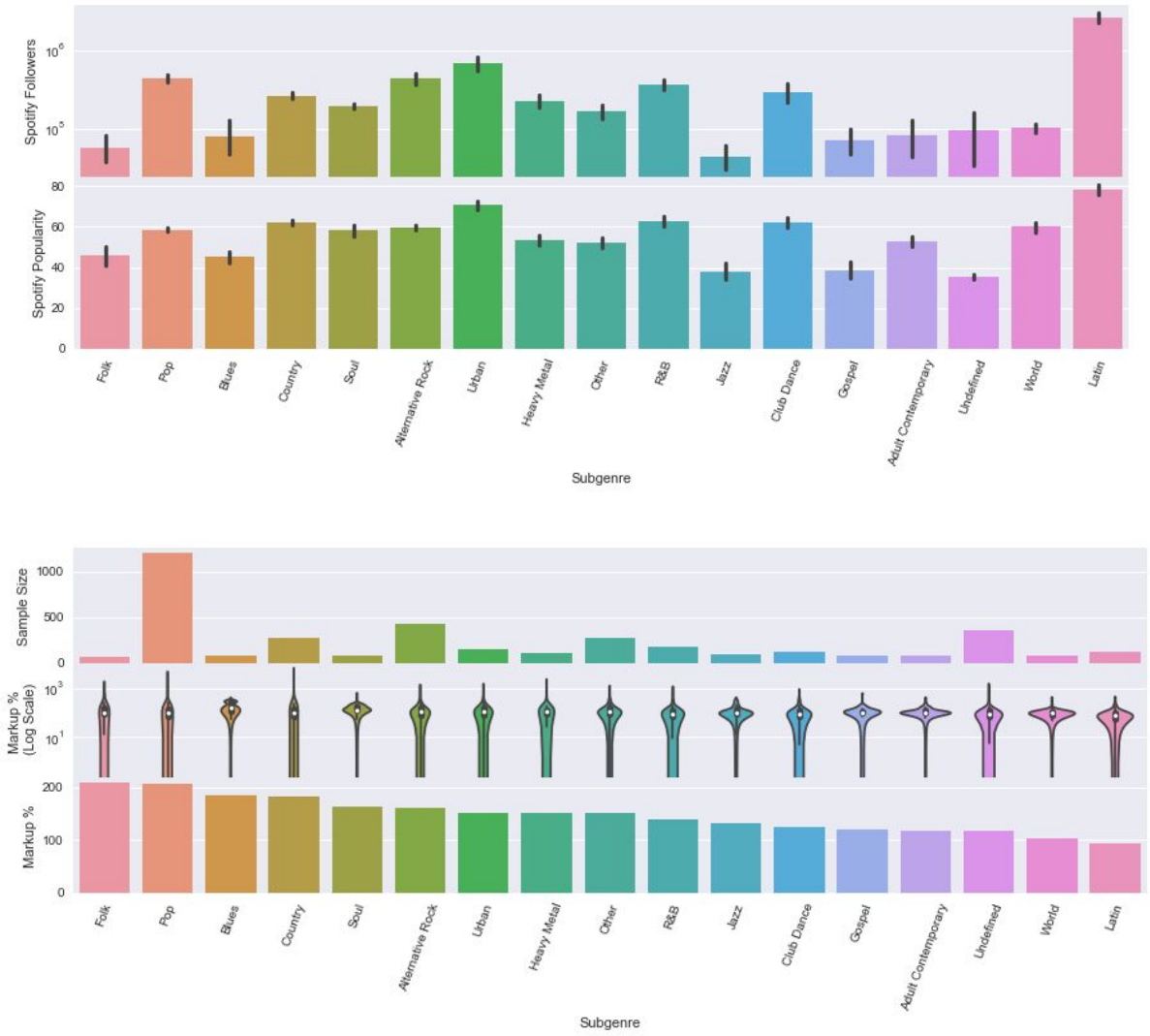


Figure 8-9. Ticket markups by subgenre & Artist Popularity Index/Followers on Spotify by Subgenre. Pop and folk events had double the markup percentage on the resale market, compared to latin and world events.

Events with resale tickets only on Stubhub had much higher markups than those with tickets on SeatGeek or both platforms. This suggests that ticket flippers would best off concentrating on Stubhub tickets, while consumers should check SeatGeek for lowest prices. Unsurprisingly, tickets listed across both platforms had lower markups, indicating greater supply and availability keeps markups down.

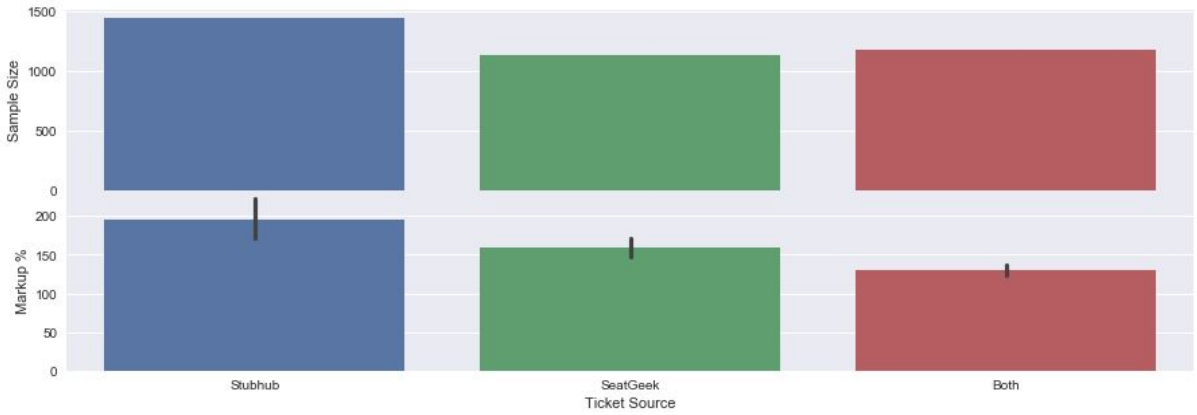
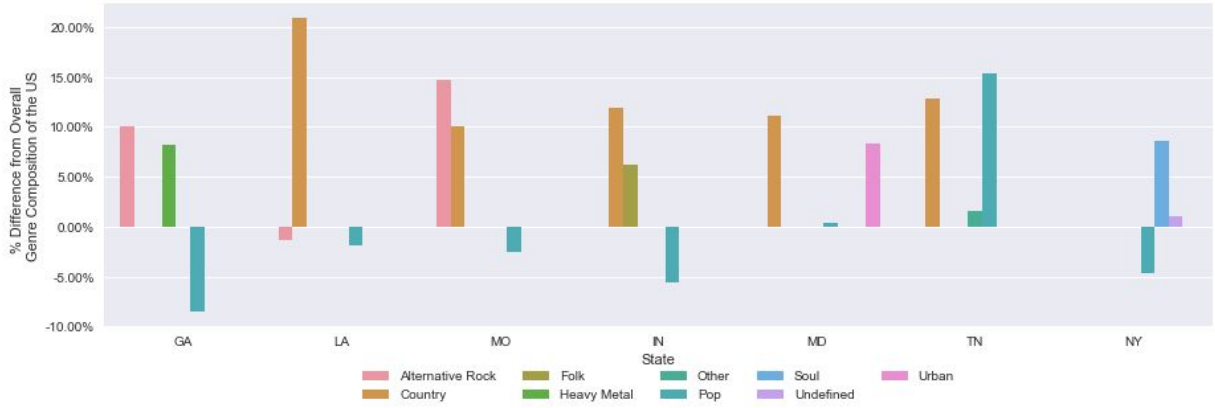


Figure 10. Markups by Source of Ticket. Events with tickets listed on both Stubhub and SeatGeek had lower markups, suggesting a greater supply of listed tickets drives down prices.

New York and Nevada had the highest sample sizes of events but were on opposite ends of the markup spectrum. Georgia and Louisiana had higher markups than California and Illinois which I found to be surprising. However, blues and folk music events had high markups, and are most likely concentrated in southern states. In addition, the tour promoter Masquerade is based in Georgia and focuses in events in southern states, which may be connected to the high markups in those states. However, there doesn't seem to be a clearly identifiable connection to factors such as cost of living, or supply of events per state.



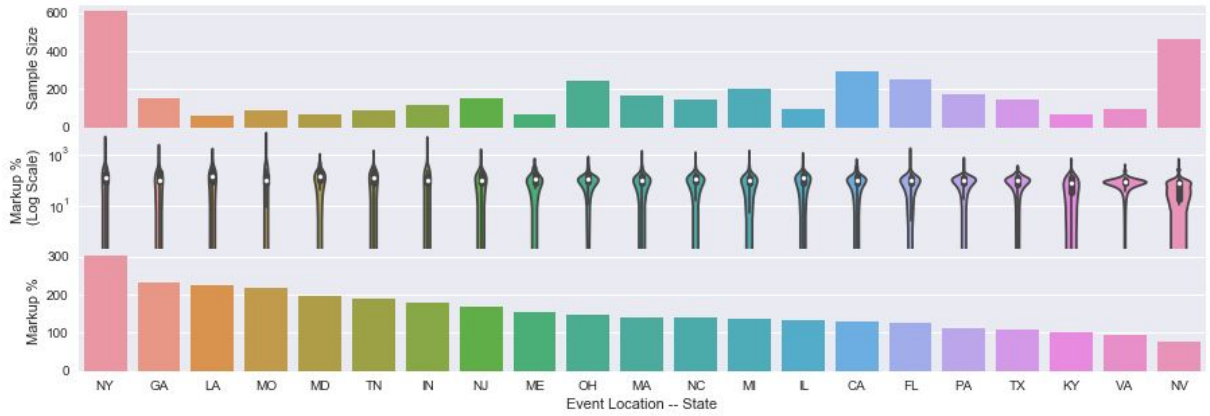


Figure 11-12. Top 3 Subgenres by State, & Subgenres Markup by Event State. Markups don't appear correlated to cost of living or event size. There could be a correlation between popular subgenres and markup by state, and promoter and markup by state. States with high overall ticket markups also had greater concentrations of events in genres with high markups, like Country, Folk, and Soul

Overall, events with small tour promoters tended to have higher markups, while larger supplies of resale tickets reduced markups. Folk, blues, pop, and country events drive the highest markups.

3.2 Machine Learning Interpretation

Because the target client of this analysis is tour promoters or ticket flippers, ease of interpretability of the machine learning algorithm is paramount. The Lasso linear regression model with log transformed Y best fits this business case and is selected for interpreting machine learning results. The coefficients of the standardized features are in the table below

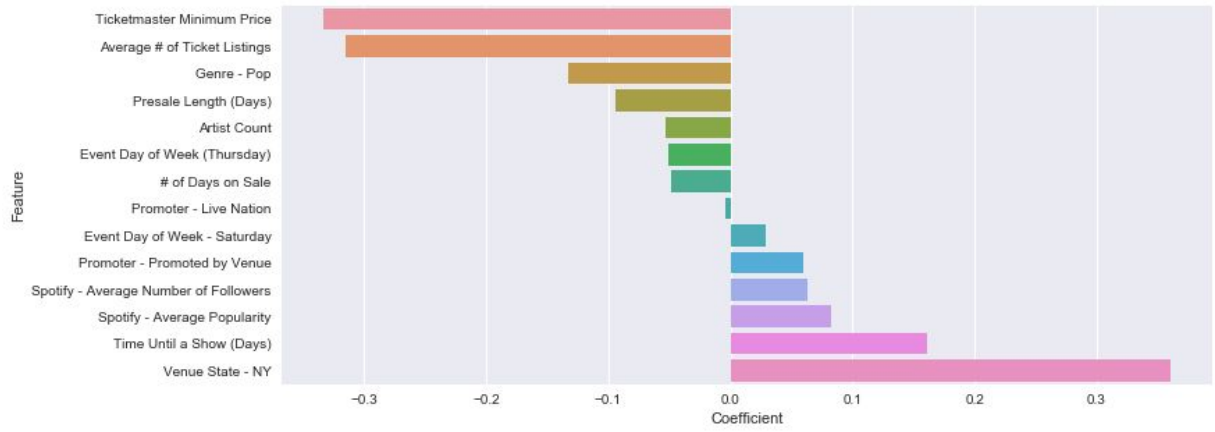


Figure 13. Coefficients of LASSO-Selected Features, on a standardized scale with log-transformed Y

Features were standardized as part of pre-processing, therefore the coefficients cannot be interpreted as is. However, they indicate which features had the largest impacts on markups on a standardized scale, and whether the effect was negative or positive. Steps were taken to minimize collinearity between features, nevertheless, it is possible that collinearity affected the interpretability of the regression coefficients.

Looking at the features chosen and coefficients, several conclusions can be made on the factors affecting resale markups. Higher face value prices indicated smaller markups, as did longer presales and pop events. Spotify popularity for an artist increased markup, as did the number of days until a show and events in the state of New York. It also appears that Live Nation events tended to have lower markups, perhaps as a big company it has more resources to accurately price events.

4. Discussion

An event ticket on average had a 158% markup over its original price. This suggests tour promoters have opportunities to boost prices and capture value from the resale market. A 158% ROI also demonstrates a significant opportunity for price arbitrage by ticket flippers, although it is important to note that this study only analyzed the listed prices of tickets, with no indication of whether a tickets listed on resale platforms were actually sold. In particular, the data suggests that events in New York, and for folk, country, blues, and pop music could be priced higher. Larger promoters like Live Nation and AEG seem to do a better job of pricing events than individual venues or smaller promotion companies. Masquerade in particular appears to have an opportunity to raise ticket prices.

This study compiled its list of events from Ticketmaster, which tends to sell tickets for large-scale events. Adding data from other platforms such as Ticketfly, AXS, or databases like Pollstar and SongKick would allow for a more comprehensive analysis, but do not have easily accessible APIs. It would be interesting to pull data over the course of several months to increase sample size as well as analyze how ticket prices change over time. Another consideration to include in future studies would be venue capacity, if the data could be gathered, to better understand the supply and demand dynamics for individual events and venues.