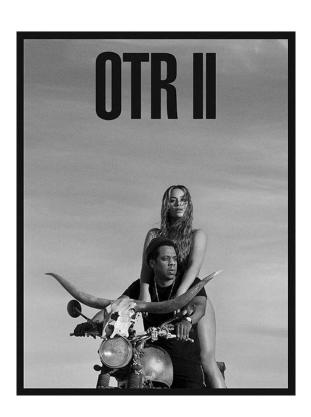


Modeling the concert ticket market and Predicting high value revenue opportunities



















Fully Digital ⇒ Fully Automated

Centralized Information (Ticketmaster & Stubhub)

Ease of Access (robust APIs)

Fully Digital ⇒ Fully Automated



Centralized Information (Ticketmaster & Stubhub)

Ease of Access (robust APIs)

Fully Digital ⇒ Fully Automated

Centralized Information (Ticketmaster & Stubhub)



Ease of Access (robust APIs)

Fully Digital ⇒ Fully Automated

Centralized Information (Ticketmaster & Stubhub)

Ease of Access (robust APIs)



Formulating My Frustrations



Research Question

Can you predict which events will have the greatest resale opportunity on the secondary market?

Metrics of Analysis

Return - % difference between wholesale and market price (mean resale price - wholesale price) / wholesale price

Risk - DOD % change in ticket quantities (tomorrow's ticket count - today's ticket count) / today's ticket count

Risk * Return - Unified metric for algorithmic simplification

Formulating My Frustrations

Research Question

Can you predict which events will have the greatest resale opportunity on the secondary market?



Metrics of Analysis

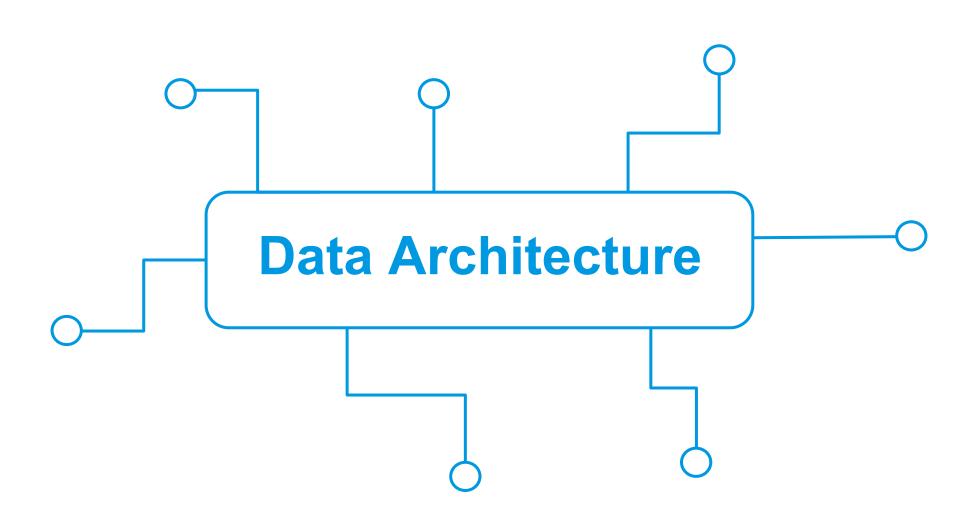
Return - % difference between wholesale and market price (mean resale price - wholesale price) / wholesale price

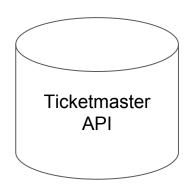
Risk - DOD % change in ticket quantities (tomorrow's ticket count - today's ticket count) / today's ticket count

Risk * Return - Unified metric for algorithmic simplification

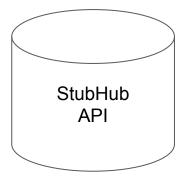
Presentation Structure

- 1. Data Architecture
- 2. Feature E&E (Engineering & Exploration)
- 3. Modeling

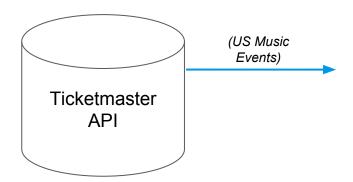




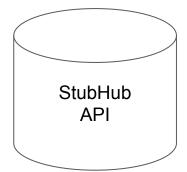
Static Variables



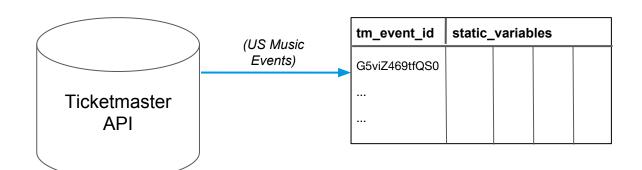
Time Dependent Variables



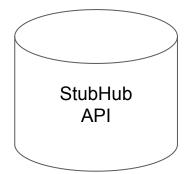
Static Variables



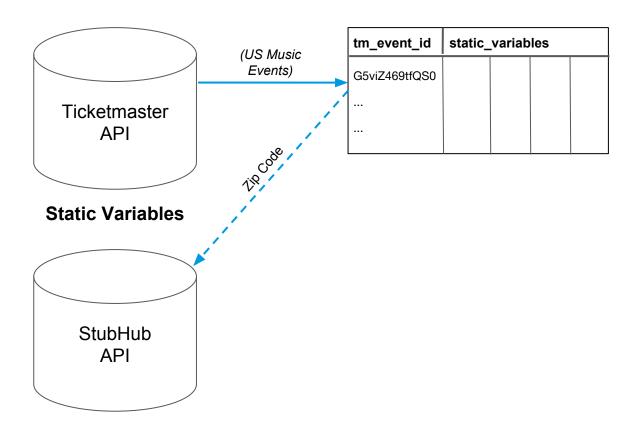
Time Dependent Variables



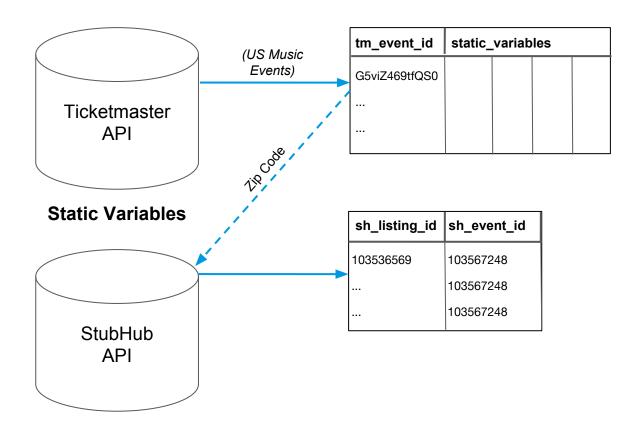
Static Variables



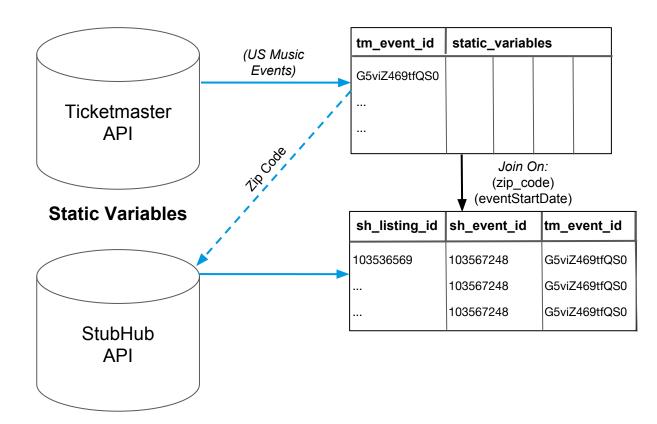
Time Dependent Variables



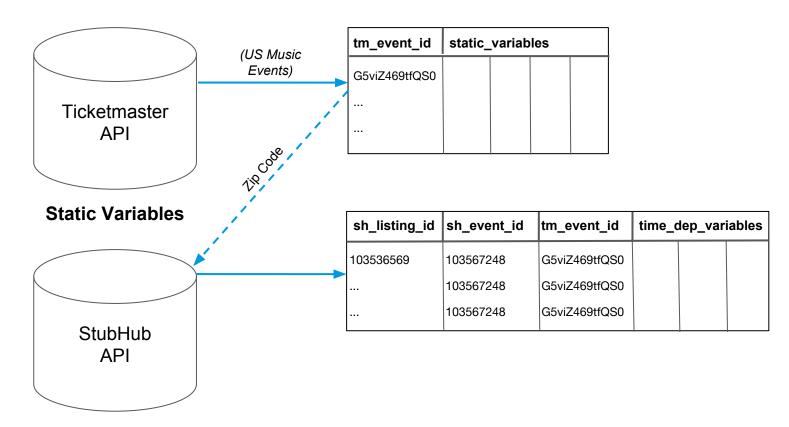
Time Dependent Variables



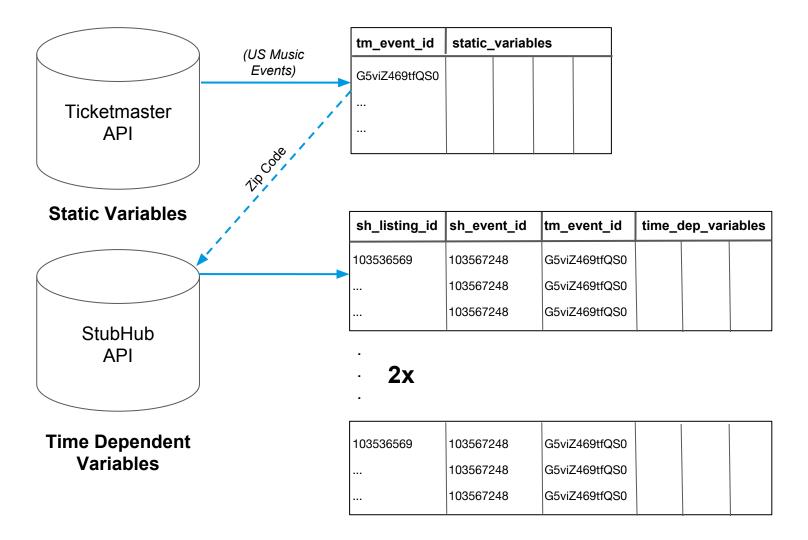
Time Dependent Variables

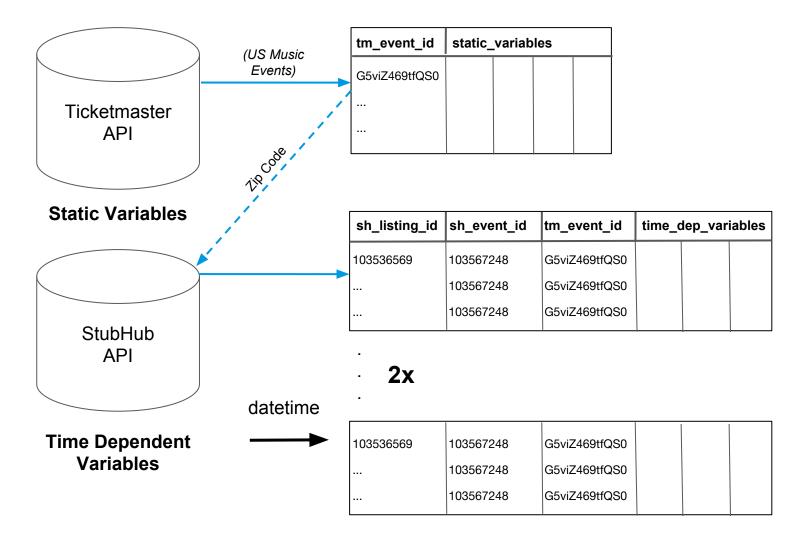


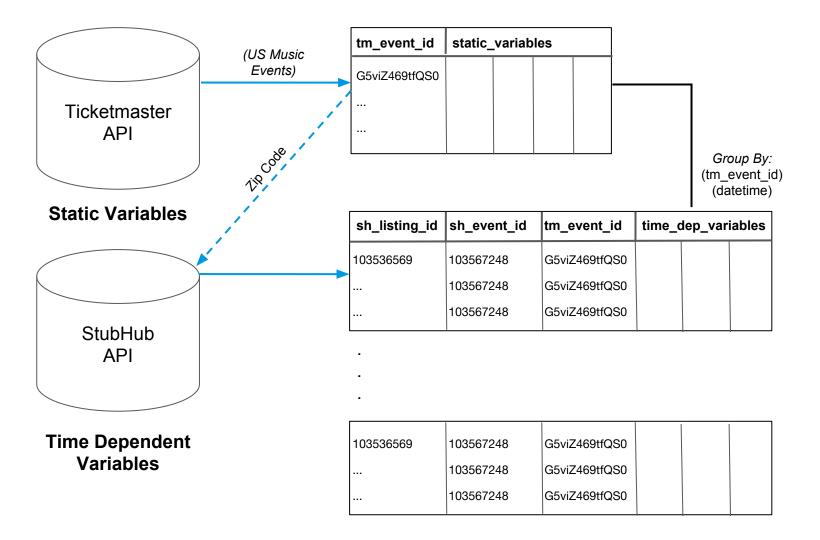
Time Dependent Variables

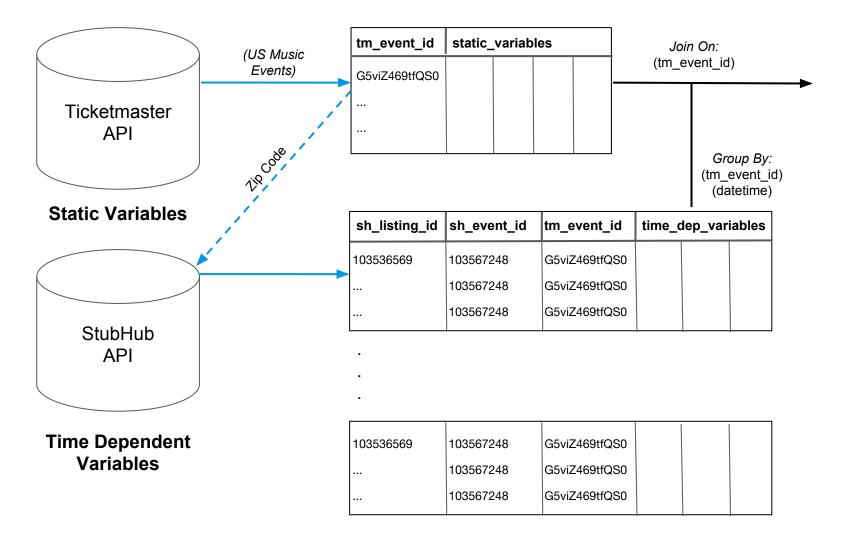


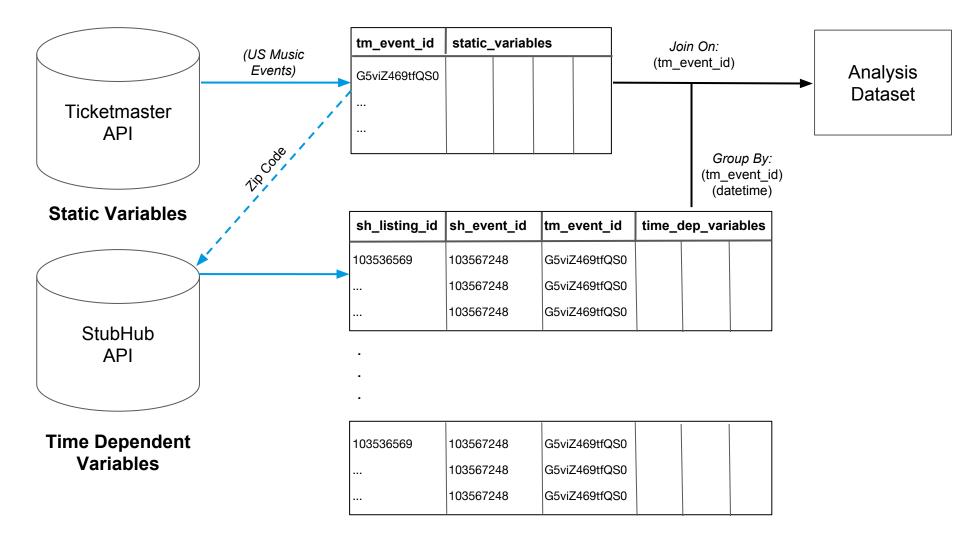
Time Dependent Variables

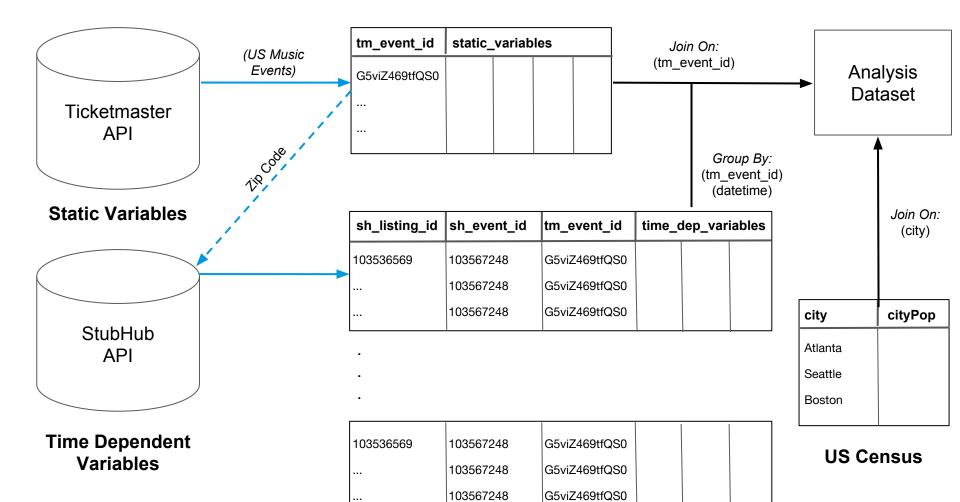












Data Limitations



Manual Pulls on Local Machine

- Time ranges are inexact
- Duplicates and strange values due to local server / connection malfunctions

Imperfect Joins

- Could not join TM and SH databases by <u>pricing zone</u> ⇒ analysis limited to <u>event</u>
- Most events joined well along zip code & event start date, but some problematic instances (what happens in Vegas stays out of my analysis)

Data Limitations

Manual Pulls on Local Machine

- Time ranges are inexact
- Duplicates and strange values due to local server / connection malfunctions

- Could not join TM and SH databases by <u>pricing zone</u> ⇒ analysis limited to <u>event</u>
- Most events joined well along zip code & event start date, but some problematic instances (what happens in Vegas stays out of my analysis)

Data Limitations

Manual Pulls on Local Machine

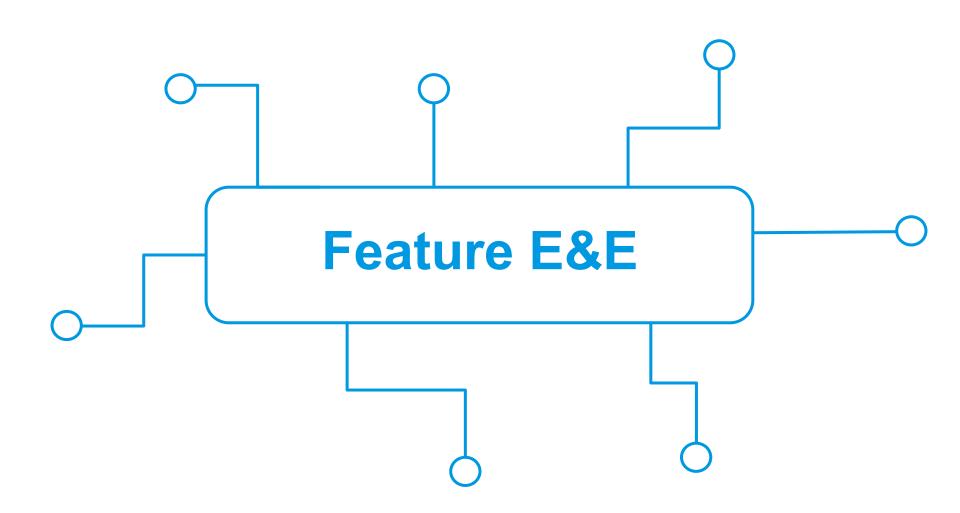
- Time ranges are inexact
- Duplicates and strange values due to local server / connection malfunctions

Imperfect Joins

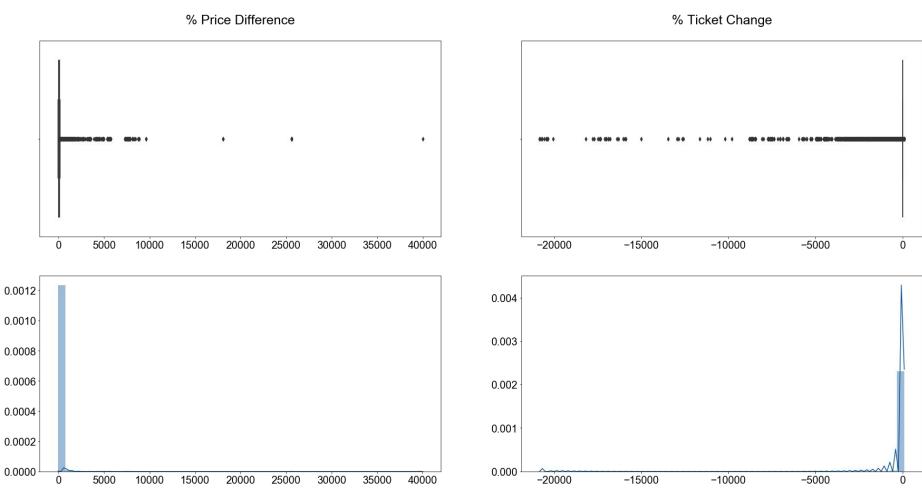
- Could not join TM and SH databases by <u>pricing zone</u> ⇒ analysis limited to <u>event</u>
- Most events joined well along zip code & event start date, but some problematic instances (what happens in Vegas stays out of my analysis)

BUT...

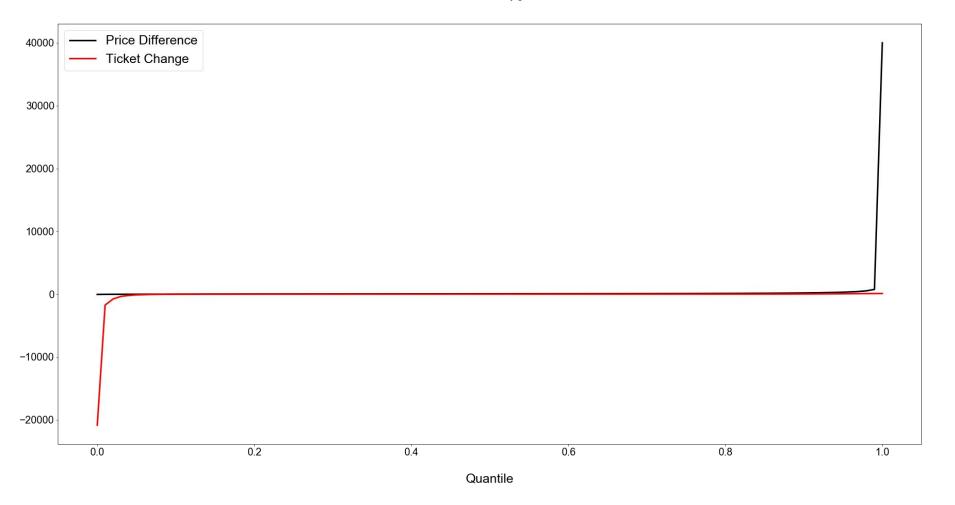
4700 Events
10 Instances of Time



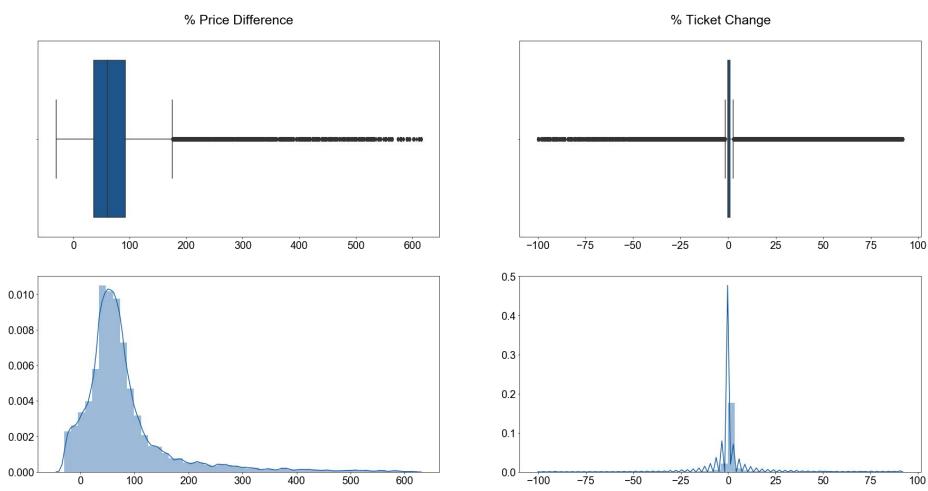
Metric Distributions



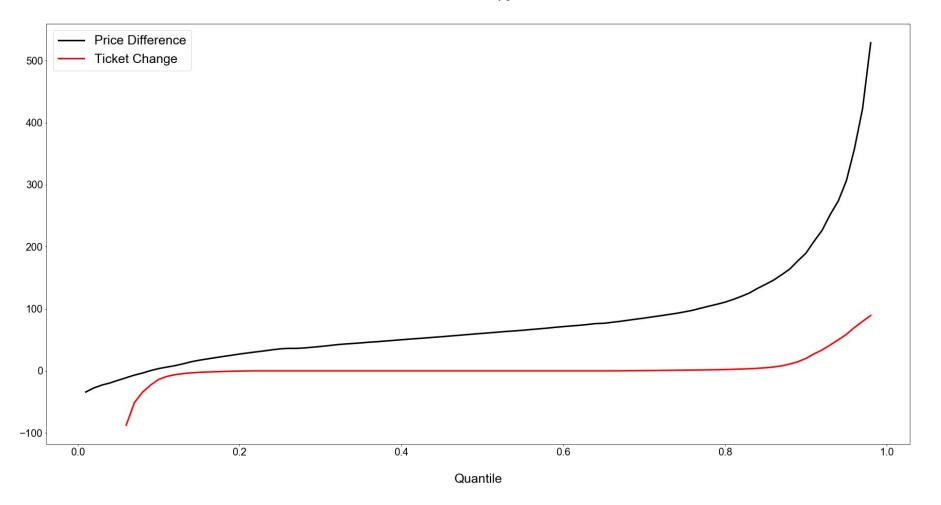
Distribution of 1% Quantiles



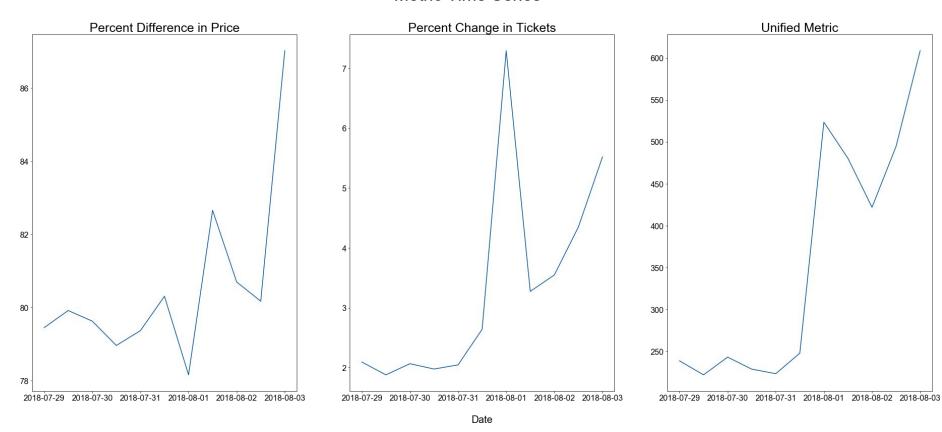
Metric Distributions



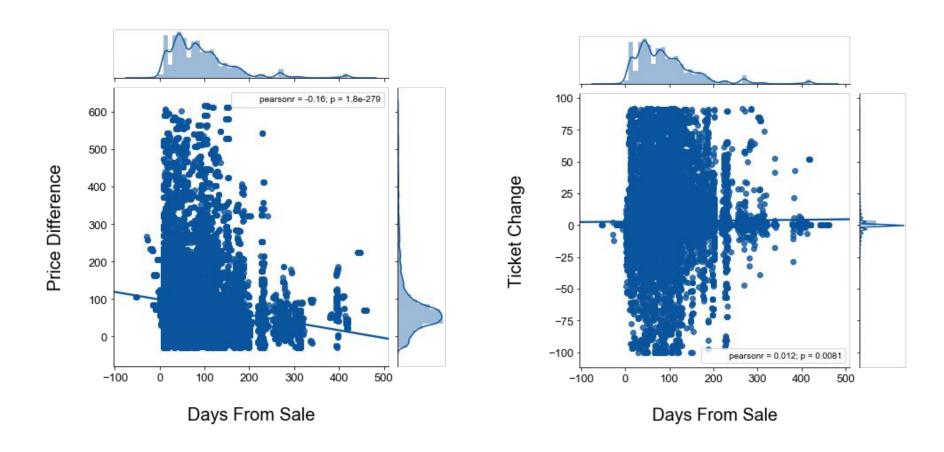
Distribution of 1% Quantiles



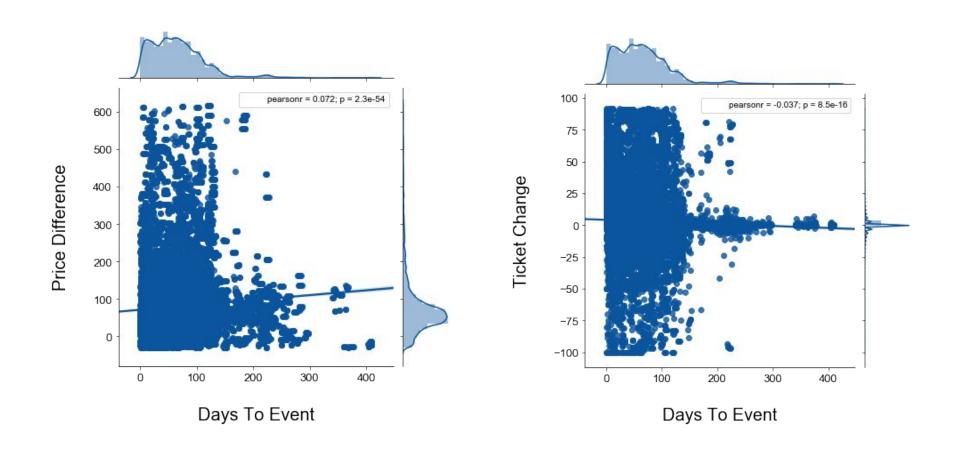
Metric Time Series



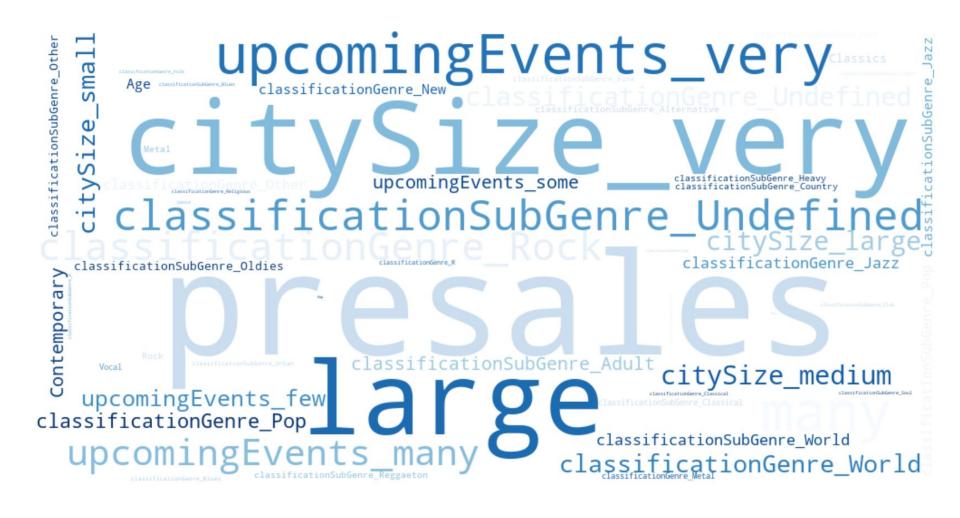
Metrics Since Sale

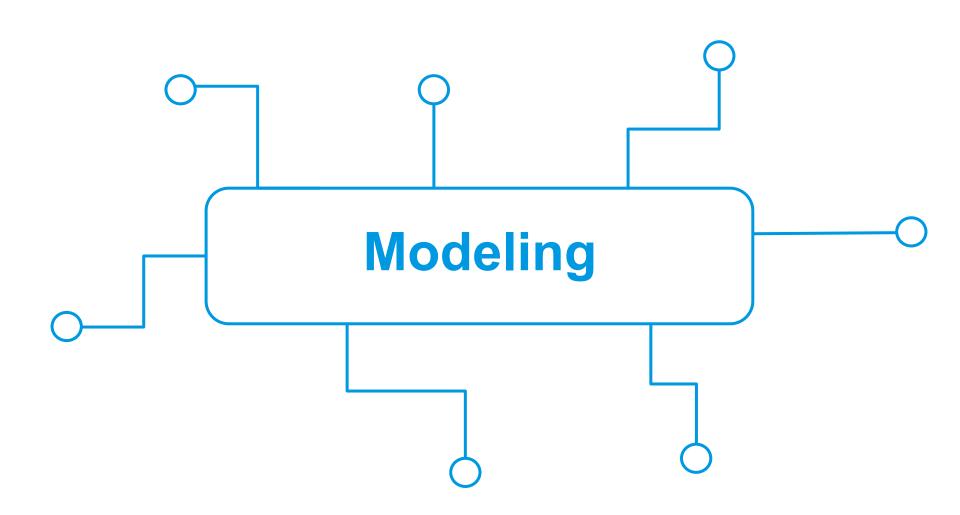


Metrics Leading up to Event



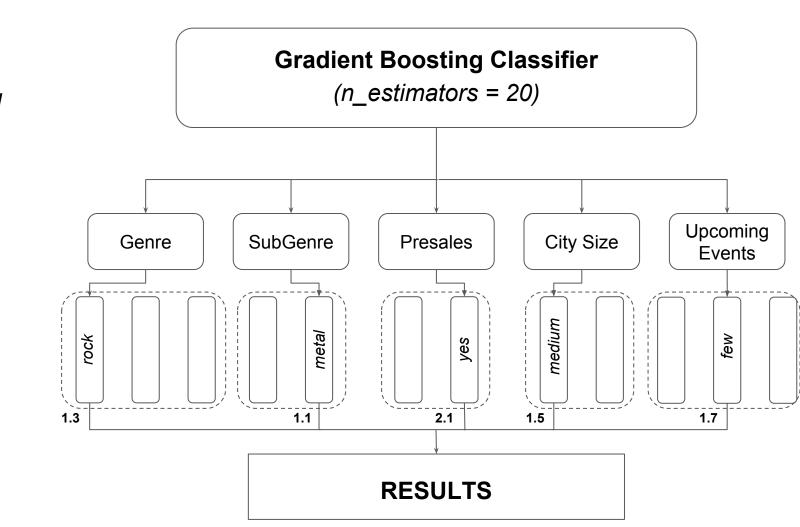






Supervised Learning

Layer 2

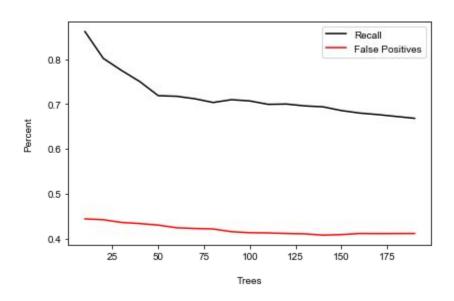


Supervised Learning

Gradient Boosting Classifier

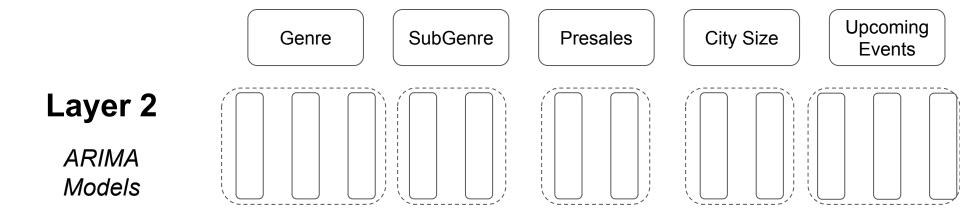
Supervised Learning

Gradient Boosting Classifier



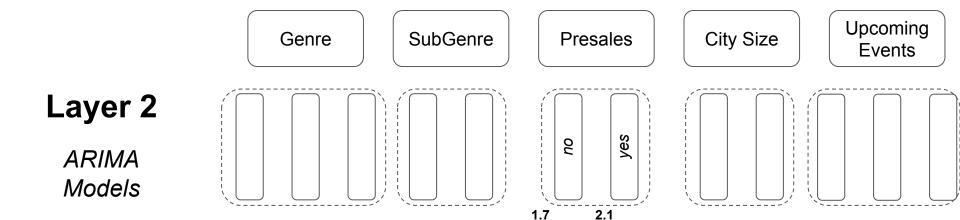
Supervised Learning

Gradient Boosting Classifier



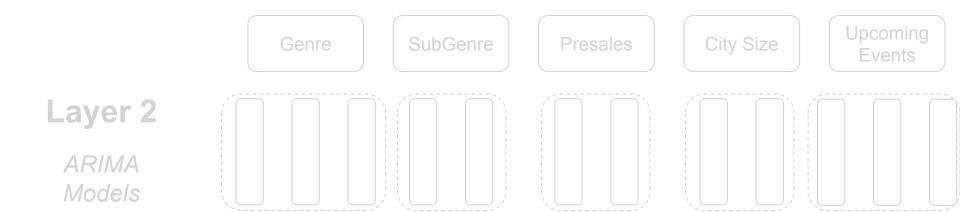
Supervised Learning

Gradient Boosting Classifier



Supervised Learning

Gradient Boosting Classifier

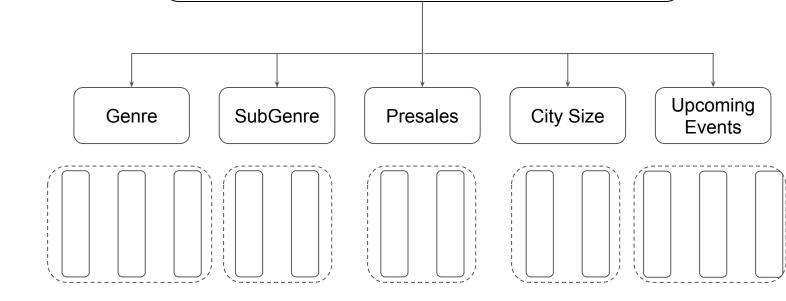


Supervised Learning



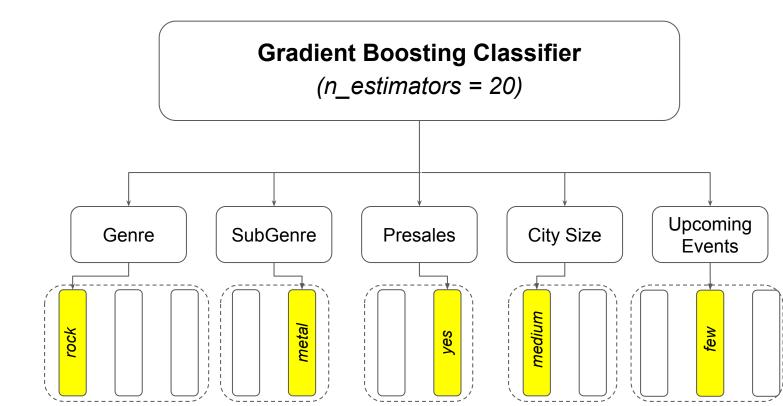
 $(n_{estimators} = 20)$





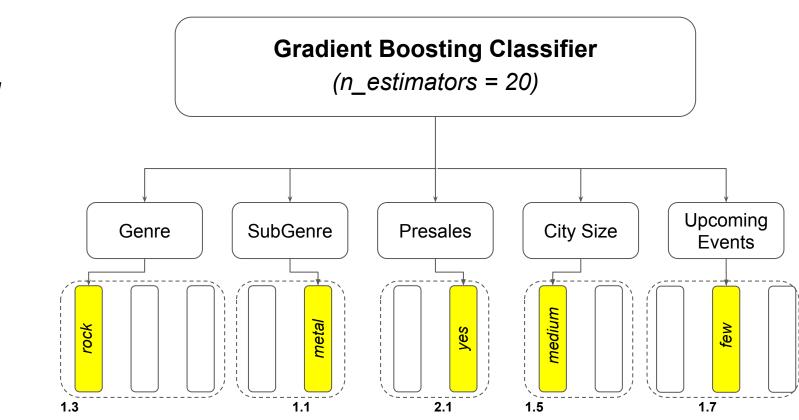
Supervised Learning





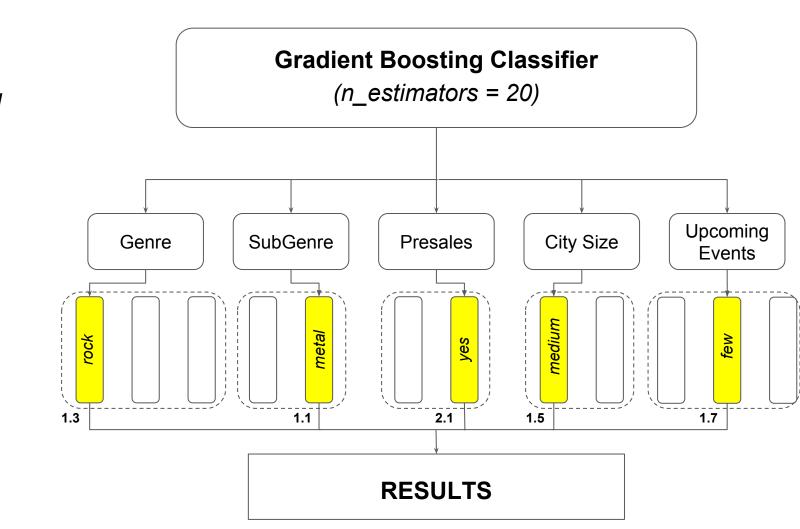
Supervised Learning



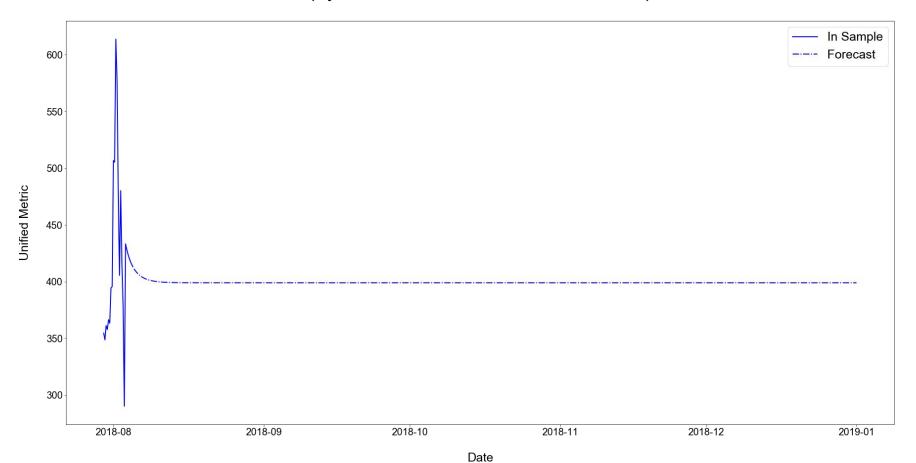


Supervised Learning

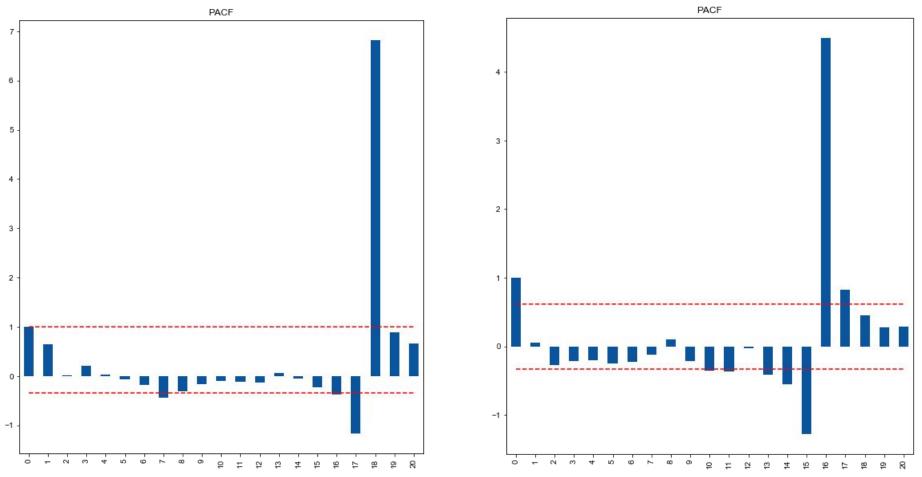
Layer 2



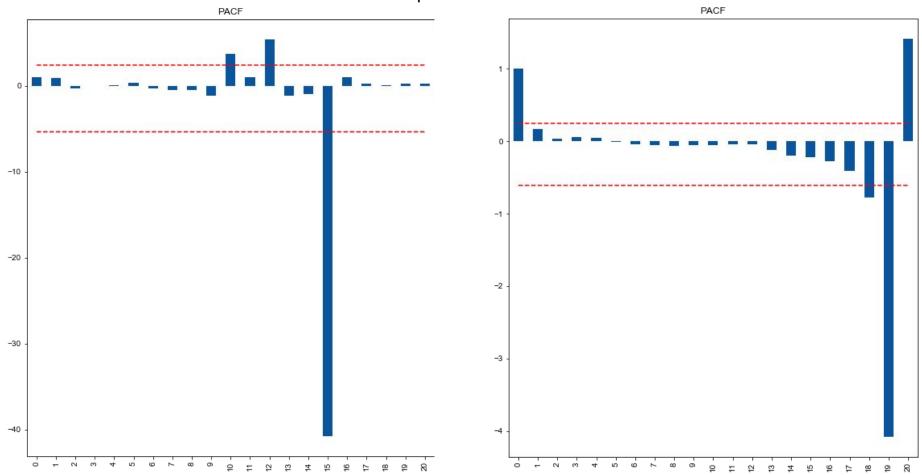
Sample Plot (Tyler Childers at the Louisville Palace)



Sample PACF Plots



Sample PACF Plots



Artist	# of Performers	Genre	Venue	City
Ario Guthrie	1	Folk	NYCB Theatre at Westbury	Westbury
Jackie Green	1	Folk	The Fillmore	San Francisco
I'm With Her	1	Folk	Southern Theatre	Colombus
Tyler Childers	1	Folk	The Louisville Palace	Louisville
Tyler Childers	1	Folk	Varsity Theater	Minneapolis
Rebeloution	3	Reggae	Neal Blaisdell Arena	Honolulu
Ario Guthrie	1	Folk	Michigan Theater	Ann Arbor
Ziggy Marley	4	Reggae	Skyline Stage at the Mann	Philadelphia
The Paper Kites	1	Folk	Paradise Rock Club	Boston



Data Architecture is SO IMPORTANT

Big Cities & Rare Events

Presales Work

Data Architecture is SO IMPORTANT



Big Cities & Rare Events

Presales Work

Data Architecture is SO IMPORTANT

Big Cities & Rare Events



Presales Work

Data Architecture is SO IMPORTANT

Big Cities & Rare Events

Presales Work





To the Cloud!

Seasonal ARIMA

Automated Model Fitting

To the Cloud!



Seasonal ARIMA

Automated Model Fitting

To the Cloud!

Seasonal ARIMA



Automated Model Fitting

To the Cloud!

Seasonal ARIMA

Automated Model Fitting



Fin