Concert Ticket Price Prediction

Sarah Alabdulwahab & Asma Althakafi

Razorgator Tickets

Online ticket **reselling** platform for sports, theater and concert tickets, and vacation packages for sporting events.

Goal

Predict the price of concert tickets in USA

Web Scraping Process

Step 1

Beautiful Soup & Selenium

Step 2

Collecting artists

Step 3

Collecting the concert links for each artist

Step 4

Collecting the tickets of each concert

Web Scraped Dataset

Artist

Individual artist or band

Data

Venue, city, state, date, and time

Level

Section and row

Price

Ticket price in US Dollars

180,094 Tickets

Took **THIRTEEN** Hours!!!

Additional Dataset

Average and median salaries of each state in USA

- From Wikipedia

Cleaning

Drop duplicated data

Remove festivals from artist

Unify the level feature

Remove outliers

Feature Engineering

Extracting the venue, city, state, date, and time

Extracting the day, month, and year from date

Adding a price “class” feature:

- 0 for cheap, 1 for expensive

After Cleaning & Feature Engineering

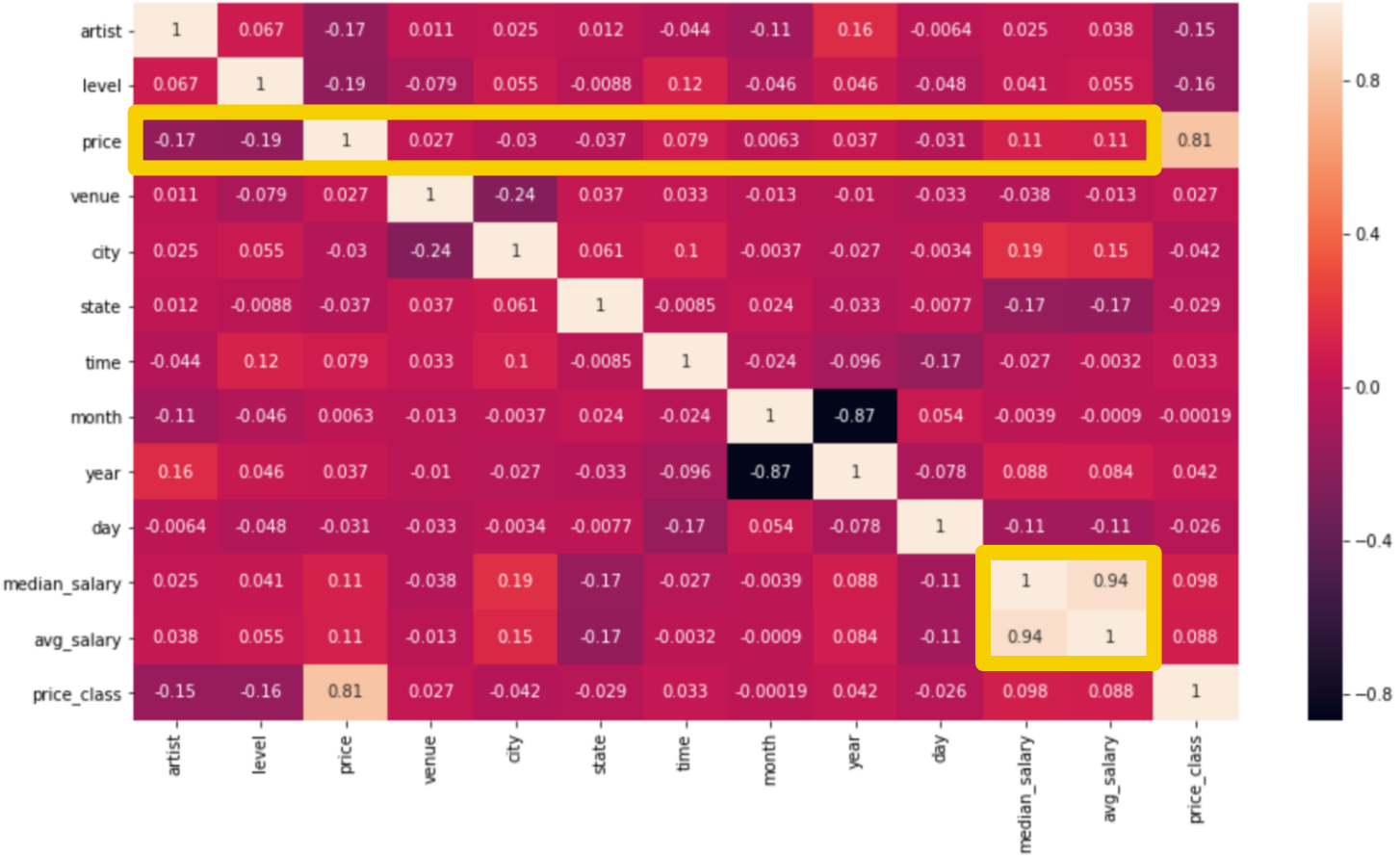
13 Features and 54,234 tickets

- Artist
- Level
- Venue
- City
- State
- Time
- Day
- Month
- Year
- Median Salary
- Average Salary
- Price
- Price Class

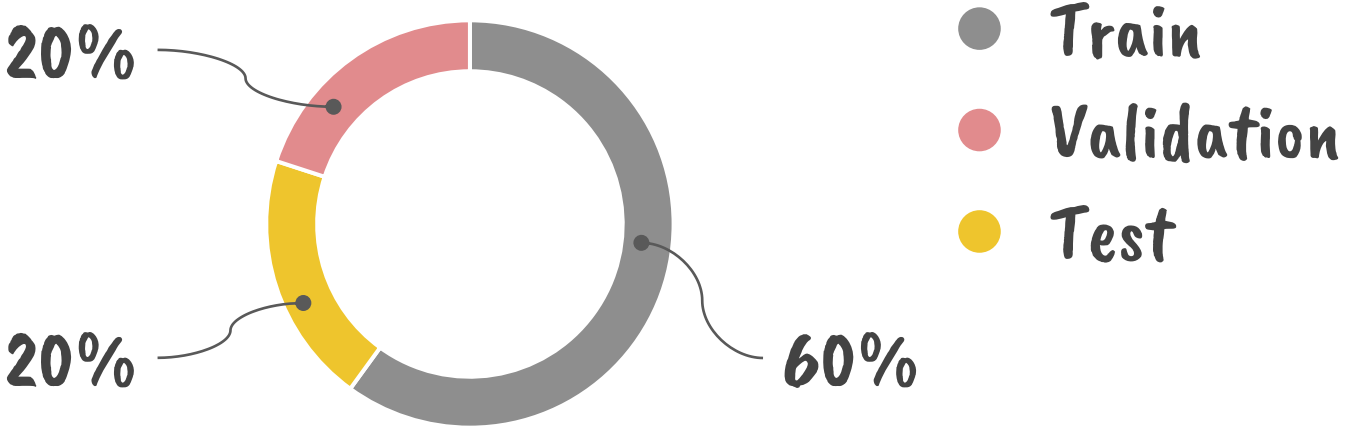
Visualizations

Moving to Tableau…

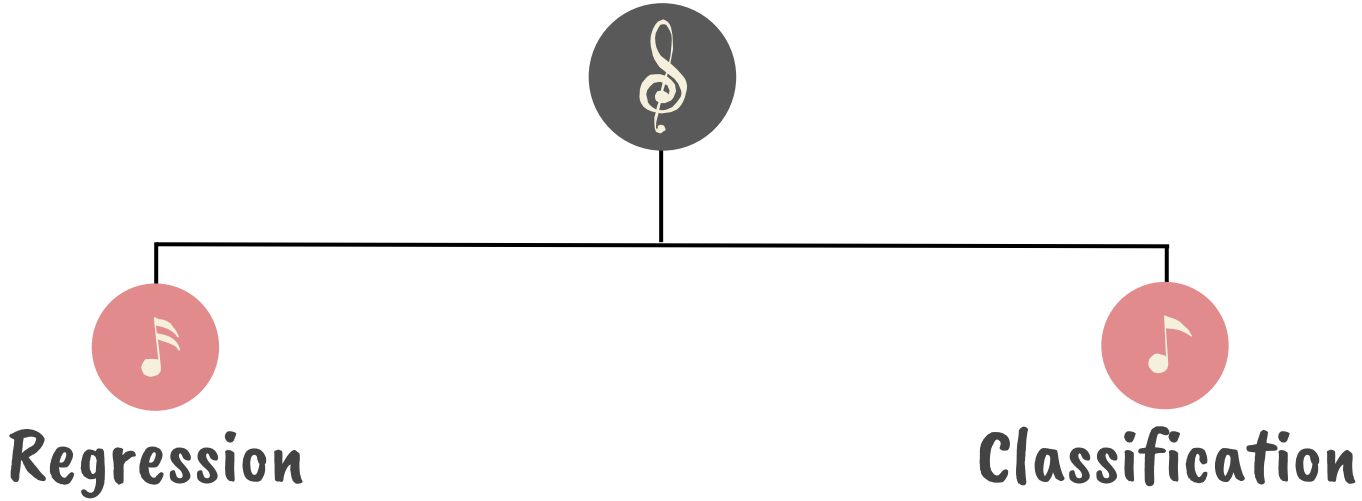
Correlation Between the Features



Spliting The Data



Modeling



Linear Regression Results

R Squared	MAE
0.103485	93.756173

Very bad…

Regularization Results

	R Squared	MAE
Lasso	0.102097	92.454616
Ridge	0.102108	92.453806
Elastic Net	0.102110	92.453659

Regularization Results

	R Squared	MAE
Lasso	0.102097	92.454616
Ridge	0.102108	92.453806
Elastic Net	0.102110	92.453659

Regression Results

	R Squared	MAE
Linear Regression	0.107720	94.688636
Polynomial Features	0.421972	69.884868
Decision Tree	0.741464	44.595658
Ada Boost	0.709937	49.418598
Random Forest	0.745291	44.589948

Regression Results

	R Squared	MAE
Linear Regression	0.107720	94.688636
Polynomial Features	0.421972	69.884868
Decision Tree	0.741464	44.595658
Ada Boost	0.709937	49.418598
Random Forest	0.745291	44.589948

Random Forest Results

On validation set:

R Squared	MAE
0.745291	44.589948

On test set:

R Squared	MAE
0.762172	42.128088

Classification Results

	Accuracy	F1
Logistic Regression	0.633631	0.602560
K Neighbors	0.849359	0.848807
Bagging	0.859132	0.858819
Decision Tree	0.863833	0.863353
Ada Boost	0.864294	0.863967
Random Forest	0.860791	0.860482

Classification Results

	Accuracy	F1
Logistic Regression	0.633631	0.602560
K Neighbors	0.849359	0.848807
Bagging	0.859132	0.858819
Decision Tree	0.863833	0.863353
Ada Boost	0.864294	0.863967
Random Forest	0.860791	0.860482

Ada Boost Results

On validation set:

Accuracy	F1
0.864294	0.863967

On test set:

Accuracy	F1
0.867520	0.867361

Conclusion

- Linear Regression was not suitable for this data
- The best regression model is Random Forest
- The best classification model is Ada Boost

THANK YOU!!

Any questions?