



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 3

Collecting the concert links for each artist

Step 2

Collecting artists

Step 4

Collecting the tickets of each concert

WEB SCRAPED DATASET

Artist

Individual artist or band

Data

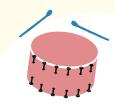
Venue, city, state, date, and time

Price

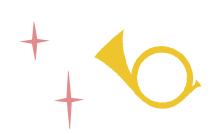
Ticket price in US Dollars

Level

Section and row







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

•	Ar	tist
---	----	------

• Time

Median Salary

Level

Day

Average Salary

Venue

Month

Price

• City

Year

• Price Class

State

VISUALIZATIONS

Moving to Tableau...



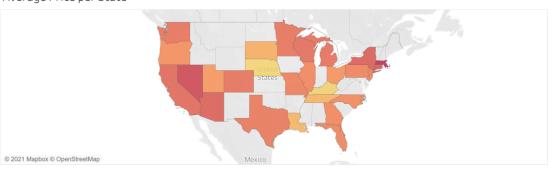
Concert Tickets



State and City

State All

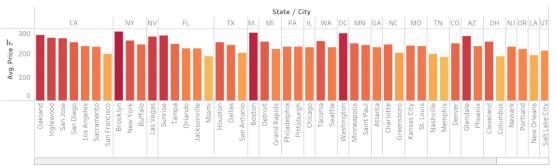
Average Price per State



Avg. Price 171.6

288.3

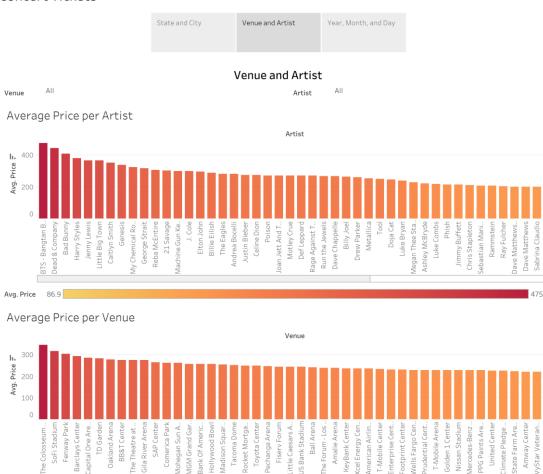
Average Price per State and City



Avg. Price 171.6

291.9

Concert Tickets



Concert Tickets



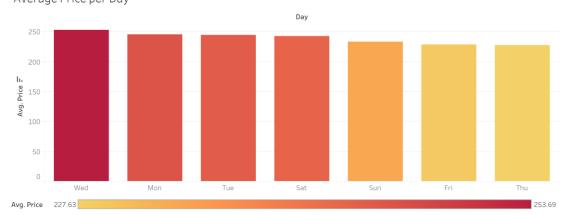
Year, Month, and Day



Average Price per Year and Month



Average Price per Day

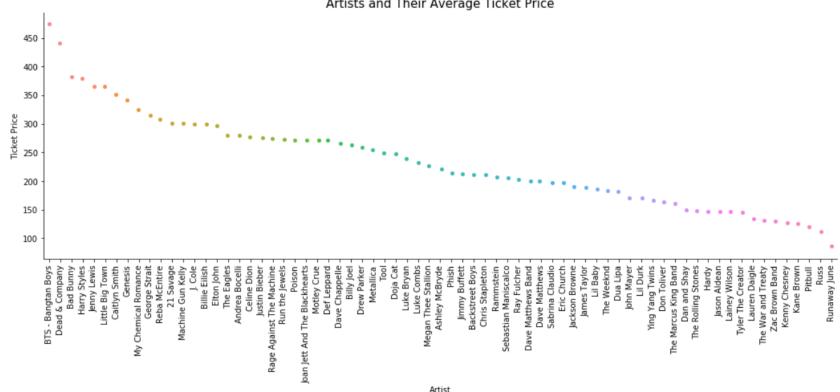






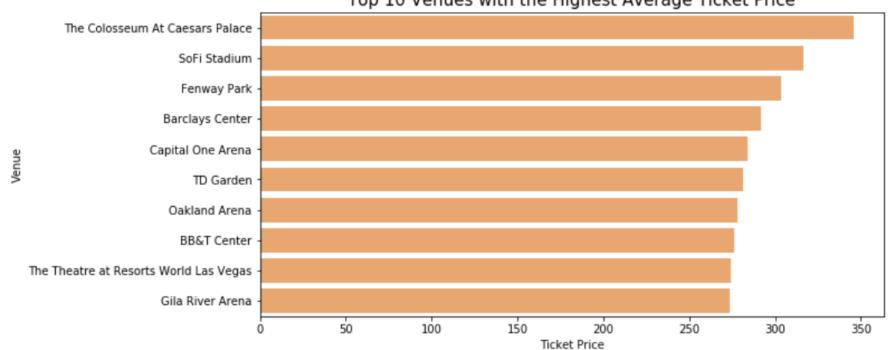
ARTISTS AND THEIR AVERAGE TICKET PRICE





Top 10 Venues

Top 10 Venues with the Highest Average Ticket Price





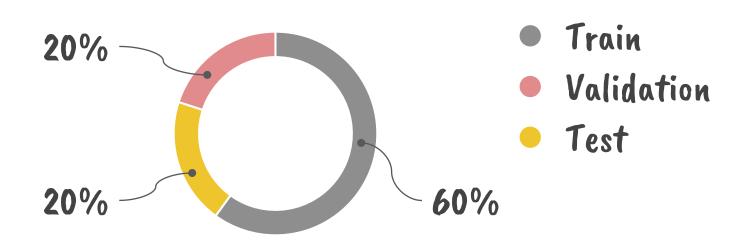


CORRELATION BETWEEN THE FEATURES

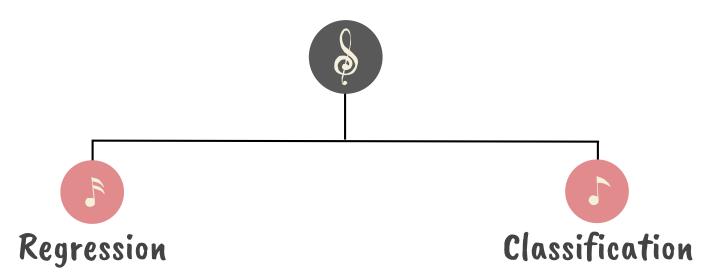
artist -	1	0.067	-0.17	0.011	0.025	0.012	-0.044	-0.11	0.16	-0.0064	0.025	0.038	-0.15
level -	0.067	1	-0.19	-0.079	0.055	-0.0088	0.12	-0.046	0.046	-0.048	0.041	0.055	-0.16
price	-0.17	-0.19	1	0.027	-0.03	-0.037	0.079	0.0063	0.037	-0.031	0.11	0.11	0.81
venue -	0.011	-0.079	0.027	1	-0.24	0.037	0.033	-0.013	-0.01	-0.033	-0.038	-0.013	0.027
city -	0.025	0.055	-0.03	-0.24	1	0.061	0.1	-0.0037	-0.027	-0.0034	0.19	0.15	-0.042
state -	0.012	-0.0088	-0.037	0.037	0.061	1	-0.0085	0.024	-0.033	-0.0077	-0.17	-0.17	-0.029
time -	-0.044	0.12	0.079	0.033	0.1	-0.0085	1	-0.024	-0.096	-0.17	-0.027	-0.0032	0.033
month -	-0.11	-0.046	0.0063	-0.013	-0.0037	0.024	-0.024	1	-0.87	0.054	-0.0039	-0.0009	-0.00019
year -	0.16	0.046	0.037	-0.01	-0.027	-0.033	-0.096	-0.87	1	-0.078	0.088	0.084	0.042
day -	-0.0064	-0.048	-0.031	-0.033	-0.0034	-0.0077	-0.17	0.054	-0.078	1	-0.11	-0.11	-0.026
median_salary -	0.025	0.041	0.11	-0.038	0.19	-0.17	-0.027	-0.0039	0.088	-0.11	1	0.94	0.098
avg_salary -	0.038	0.055	0.11	-0.013	0.15	-0.17	-0.0032	-0.0009	0.084	-0.11	0.94	1	0.088
price_class -	-0.15	-0.16	0.81	0.027	-0.042	-0.029	0.033	-0.00019	0.042	-0.026	0.098	0.088	1
'	artist -	level -	price -	venue -	aity -	state -	time -	month -	year -	- ф	lian_salary -	avg_salary -	price_class -

- 0.8

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		
Ridge		
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		
Polynomial Features		
Decision Tree		
Ada Boost		
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		
K Neighbors		
Bagging		
Decision Tree		
Ada Boost	0.864294	0.863967
Random Forest		

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?

