



# **Predicting Vinyl Record Prices**

**Metis Fall Online 2020 Project 2**

**Bethany Baumann**



# The Vinyl Revival

- Vinyl sales have outperformed CD sales for the first time since 1986
  - \$225 million / year market for vinyl \*
- Used vinyl collecting and trading is increasing in popularity
  - 40 million for sale on Discogs.com

\* Recording Industry Association of America



# Data Collected by Web Scrapping

**Discogs**<sup>®</sup>

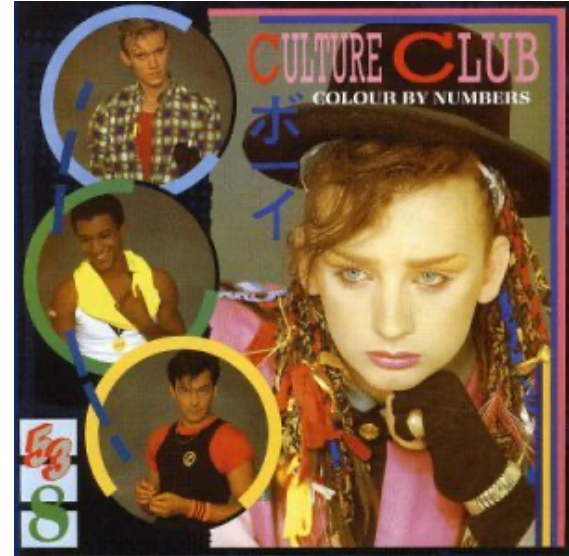


**Spotify**<sup>®</sup>





# The Target: Median Sale Price



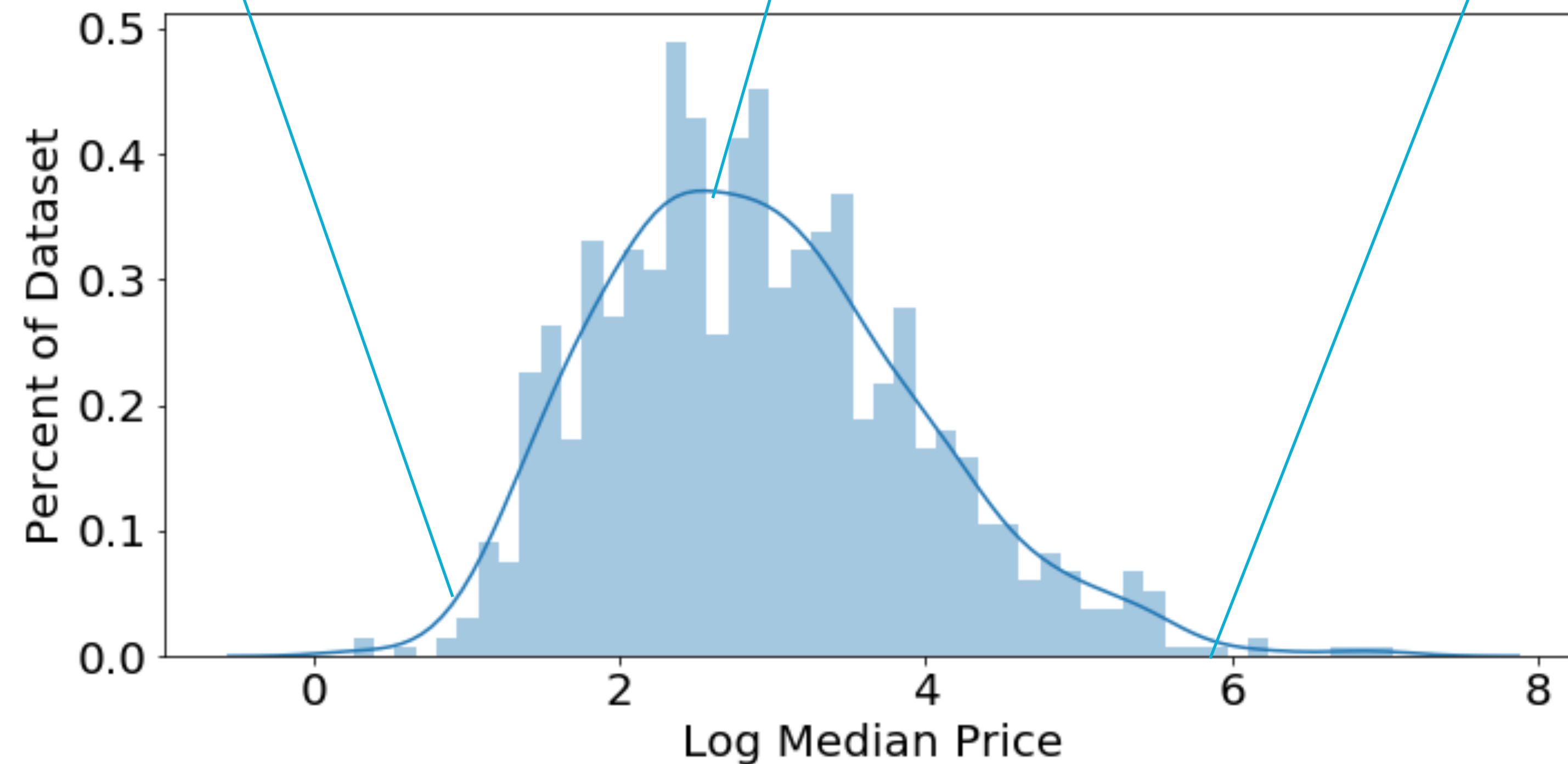
Culture Club  
Color by  
Numbers  
\$2.73



Sugarhill Gang  
Rapper's  
Delight  
\$17.50

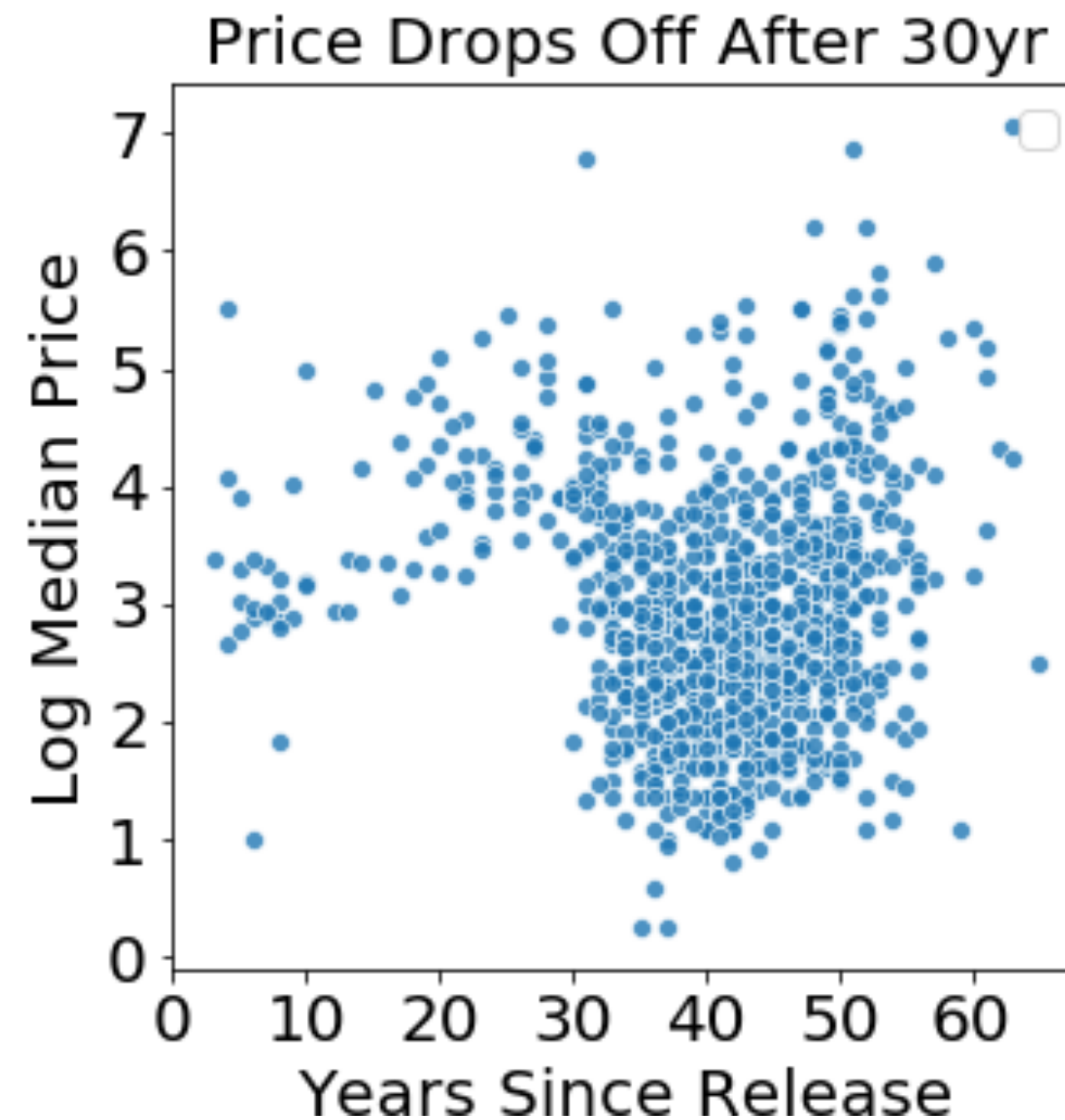


The Beatles  
Please  
Please Me  
\$369.32



# Types of Data Collected

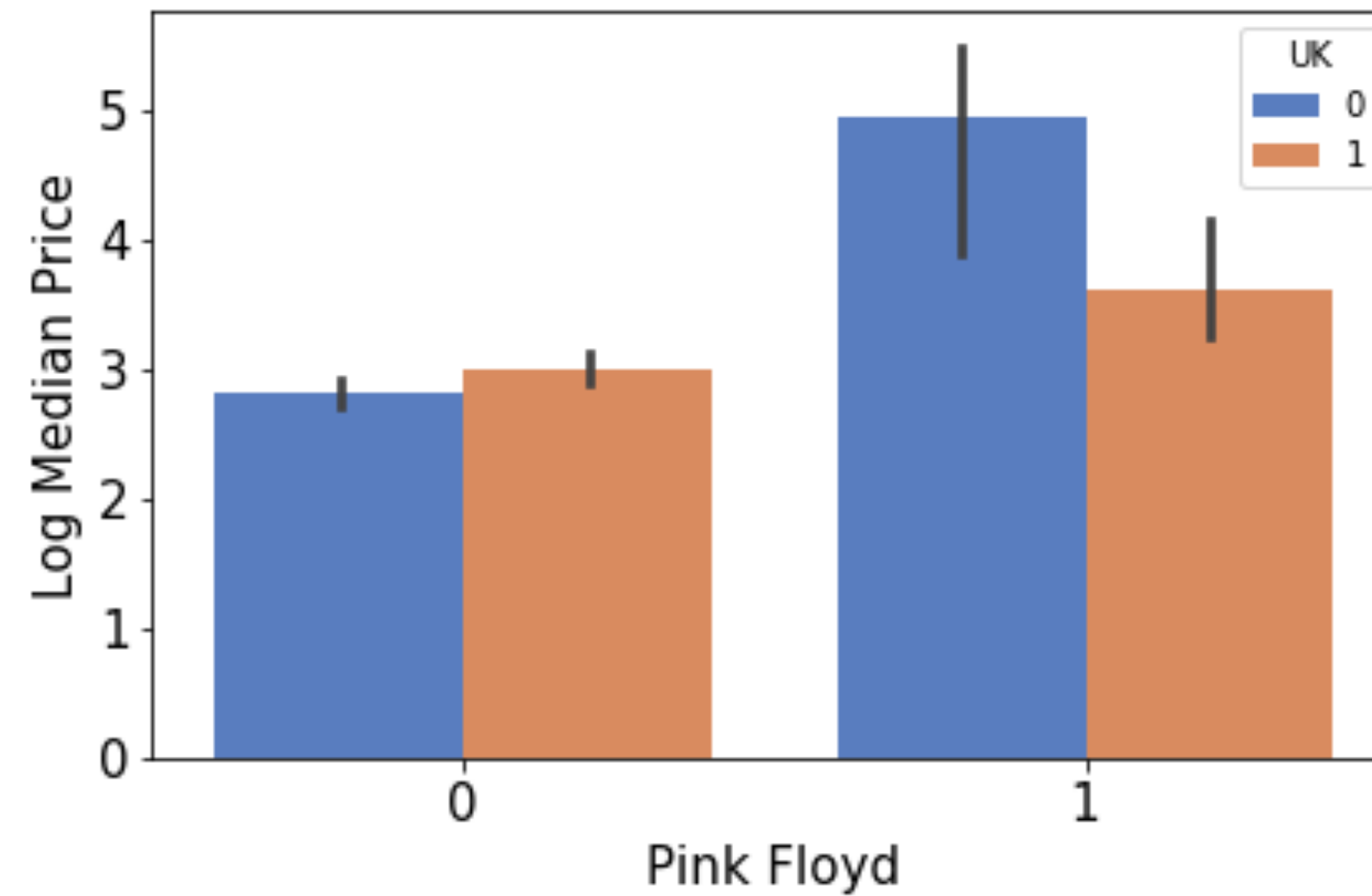
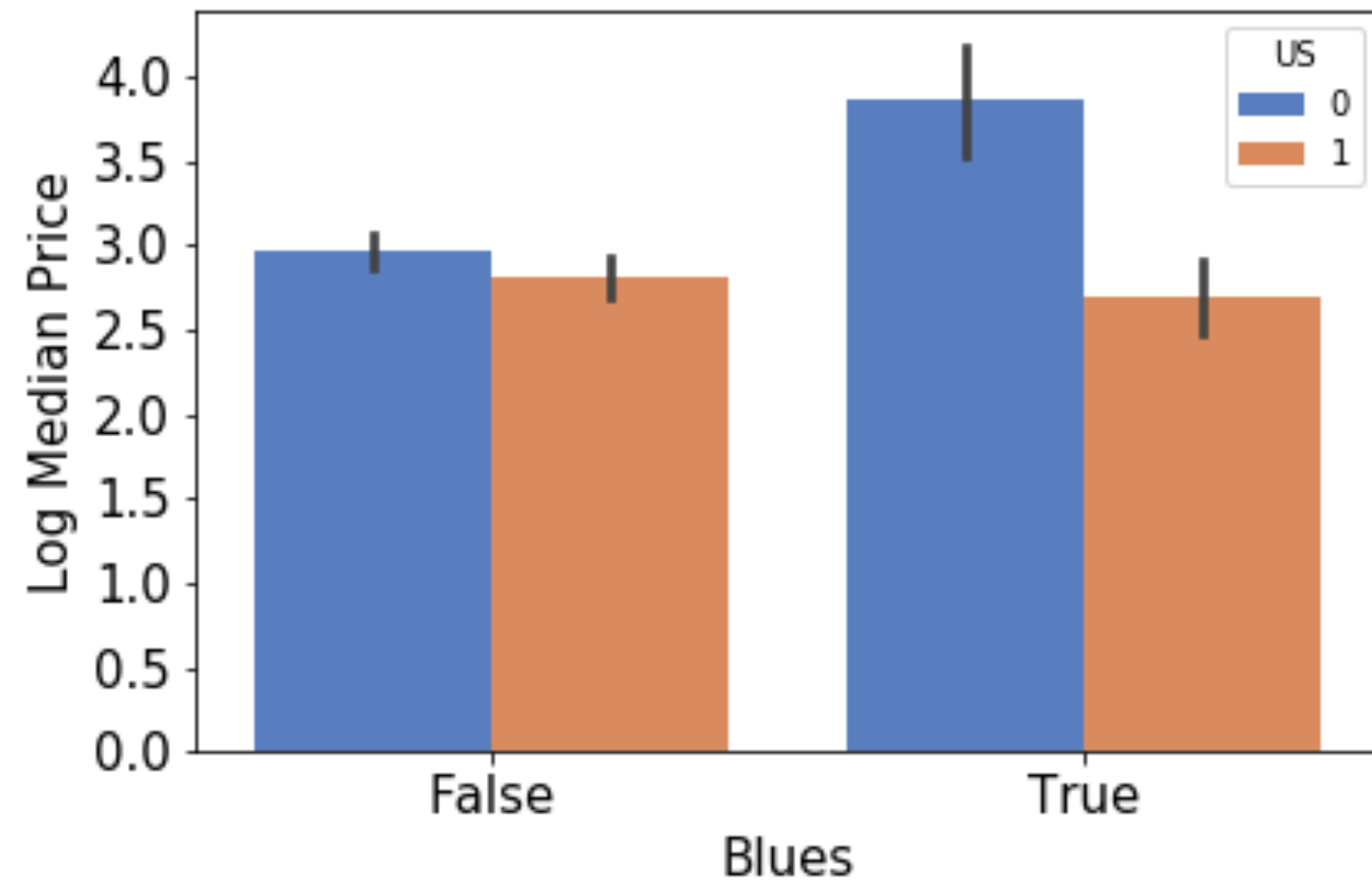
- Discogs user rating
- Number of songs
- Number of repressings
- Genre
- Artist
- Label
- Release country
- Spotify artist monthly listeners
- Years since album release
- Years after first artist album





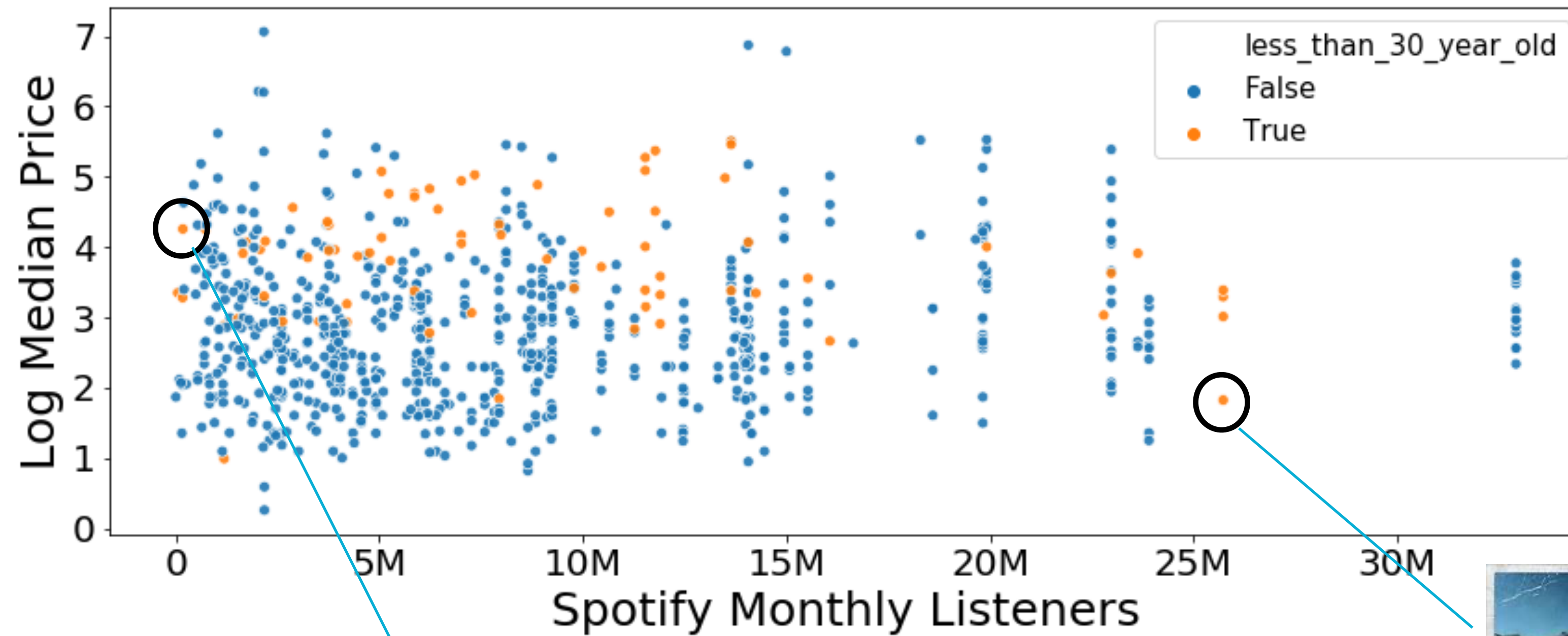
# Feature Engineering

Rarer combinations are more valuable



# Feature Engineering

Newer albums are worth less if streamed more



Prodigy  
The Fat of the Land  
\$70.70

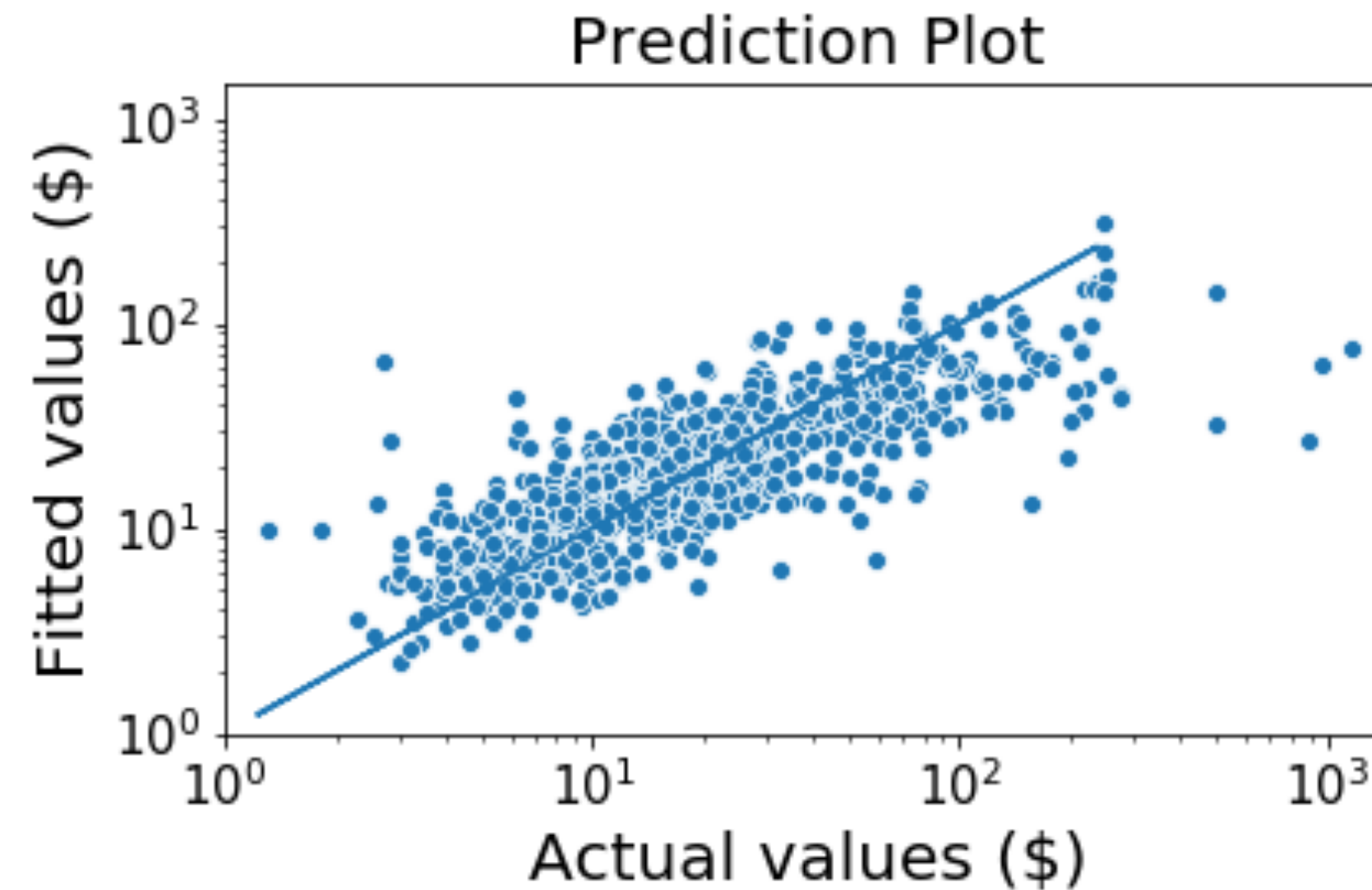
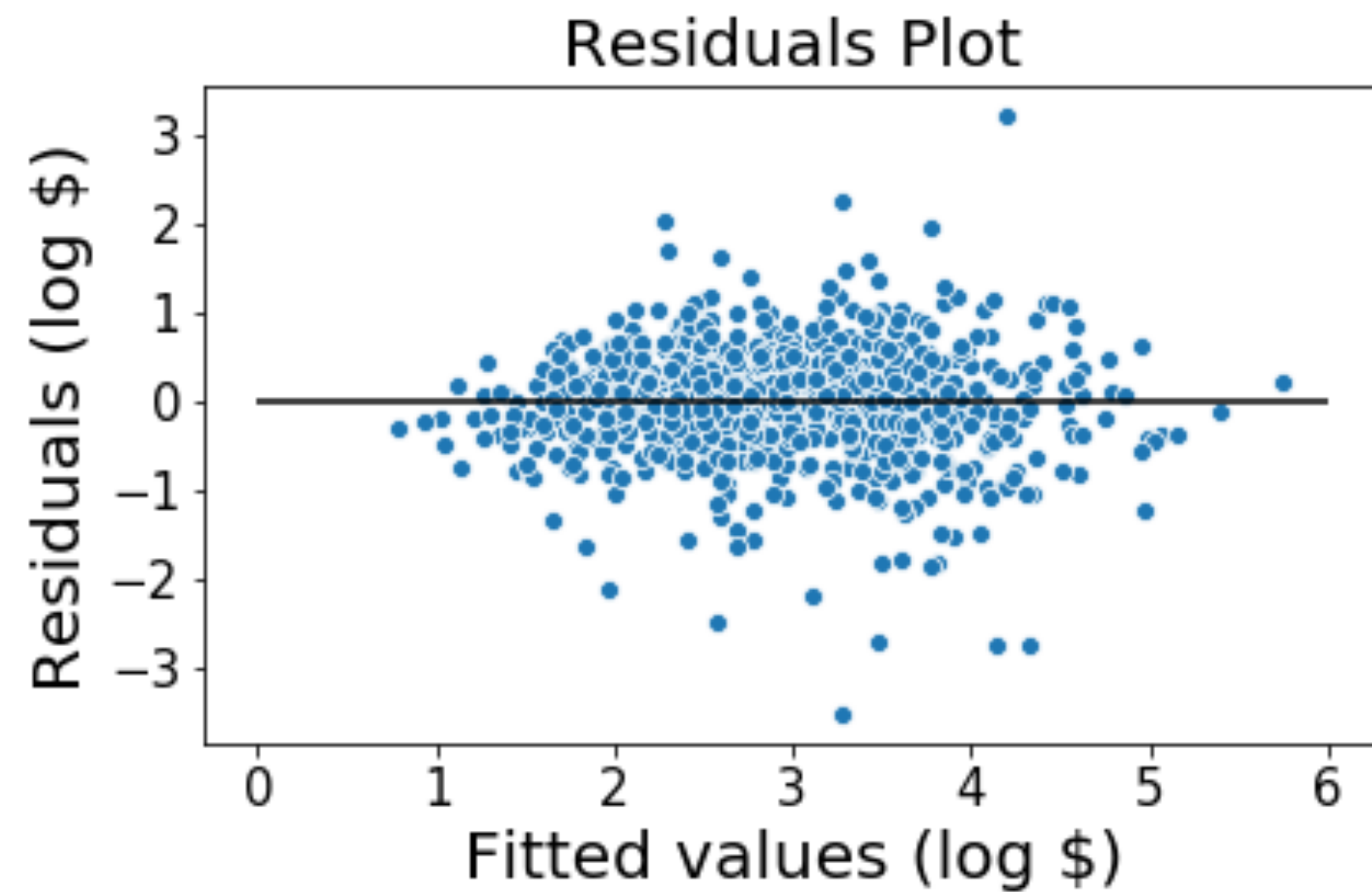


Kendrick Lamar  
good kid, m.A.A.d city  
\$6.21 (single)



# Model Selection

A lasso regression was fit on the training data



Average error: \$18.90



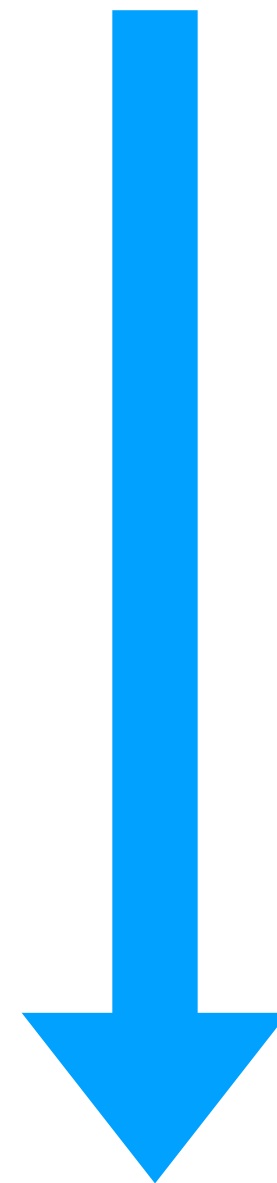
# Strongest predictors of price

## Positive effect:



- User rating
- Number of repressings
- Less than 30yr old
- Pink Floyd
- Years since release

## Negative effect:



- Years released after artist's first album
- Years since release X number of repressings
- Pop Rock
- Blues X US release





# Model Evaluation on Test Set

Average error: \$17.82

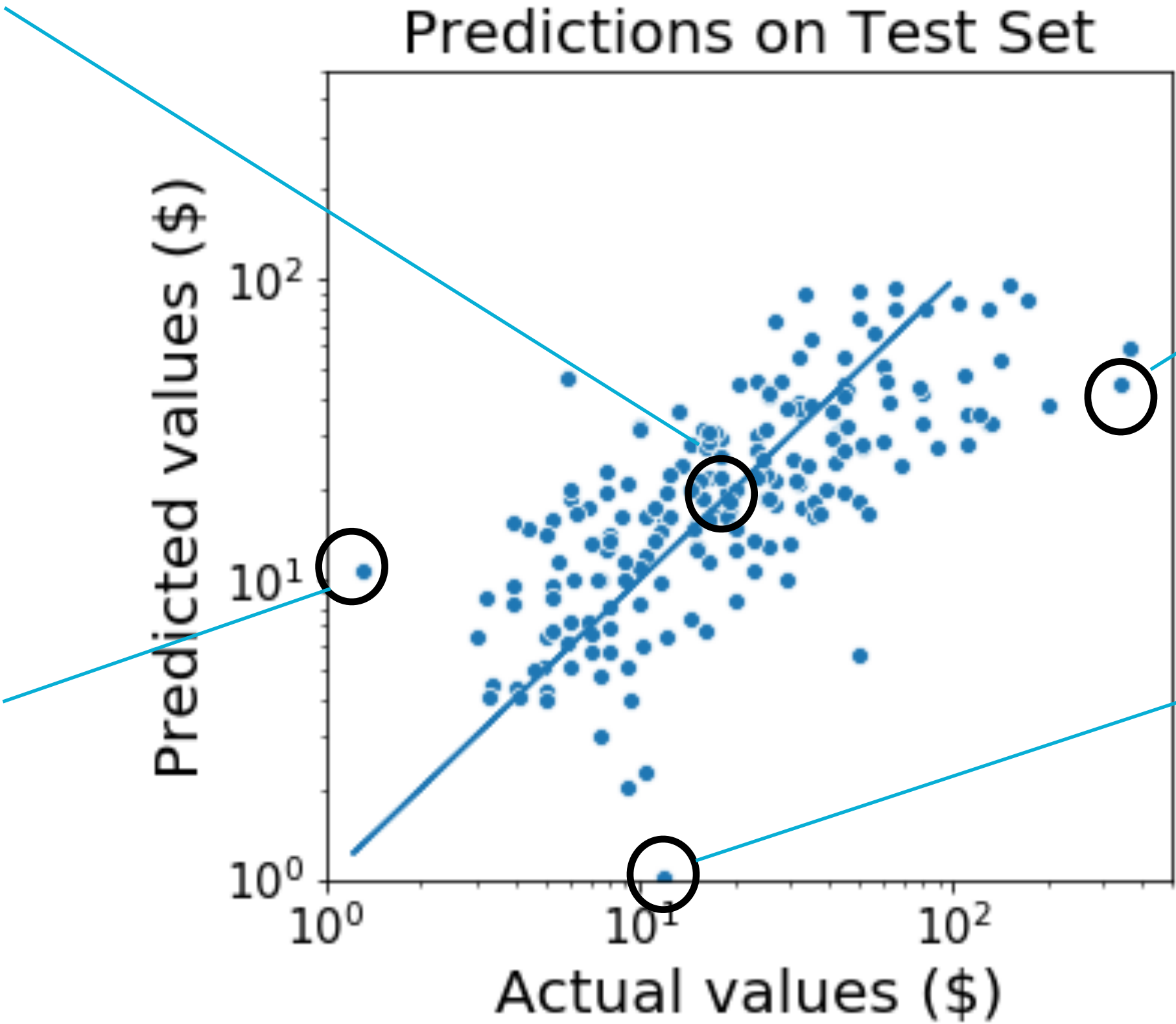
Pink Floyd  
Piper at the Gates of Dawn  
\$338.80  
predicted: \$45.36



Prince 1999  
\$19.98  
predicted: \$20.15



Paul Hardcastle  
19  
\$1.29  
predicted: \$10.80



Simon and Garfunkel  
Sounds of Silence  
\$11.99  
predicted: \$1.02



# Model vs. Subject-Matter Expert

Predicting price on 50 test set cases

Expert mean error: **\$23.59**

Model mean error: **\$14.88**





# Take Aways

- Basic album info can account for some variance in album prices, but the error rate is still high
- A larger training set may improve the model
- Album prices may be difficult to predict because of unique, difficult to model, features of each record

