

Predicting Vintage Vinyl Resale Prices



Agenda



MOTIVATION



DATA
& FEATURES



EDA AND MODEL
SELECTION



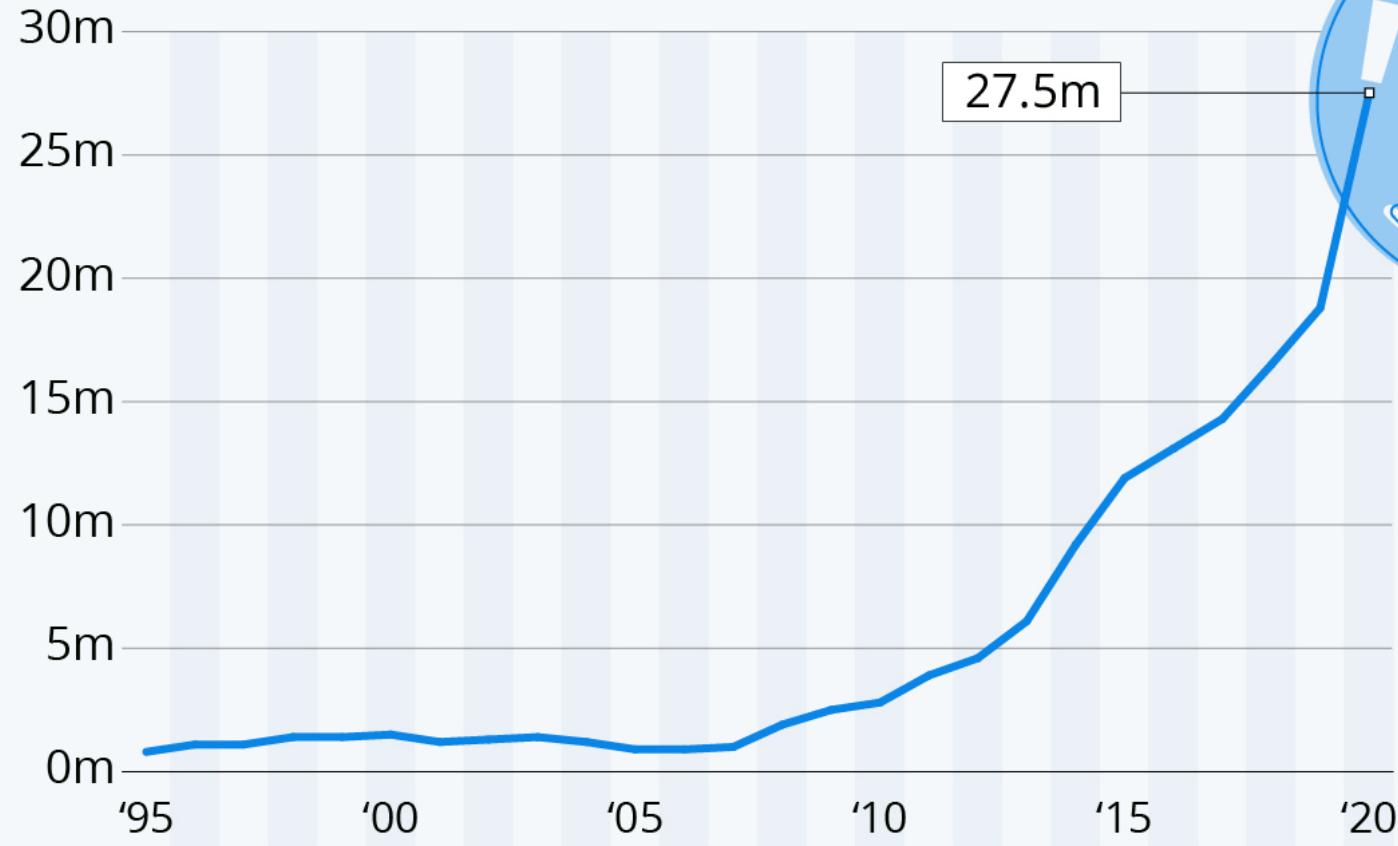
FINDINGS

Motivation

- **10%** increase in vinyl sales every year for past 15 years
- **26%** of physical sales of recorded music (new records)
- New vinyl prices increased by **490%** between 2007 and 2017

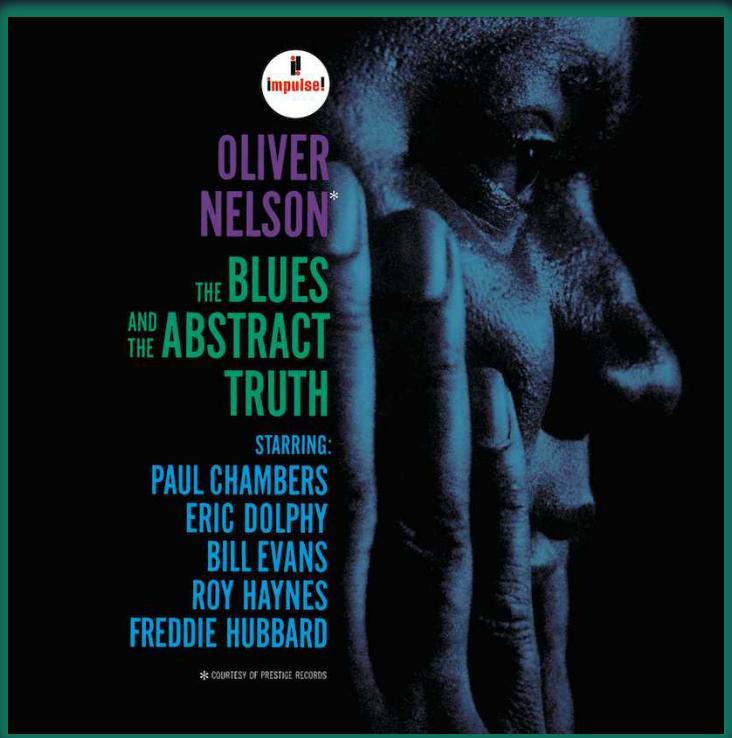
The Vinyl Comeback Continues

Vinyl LP unit sales in the United States from 1995 to 2020

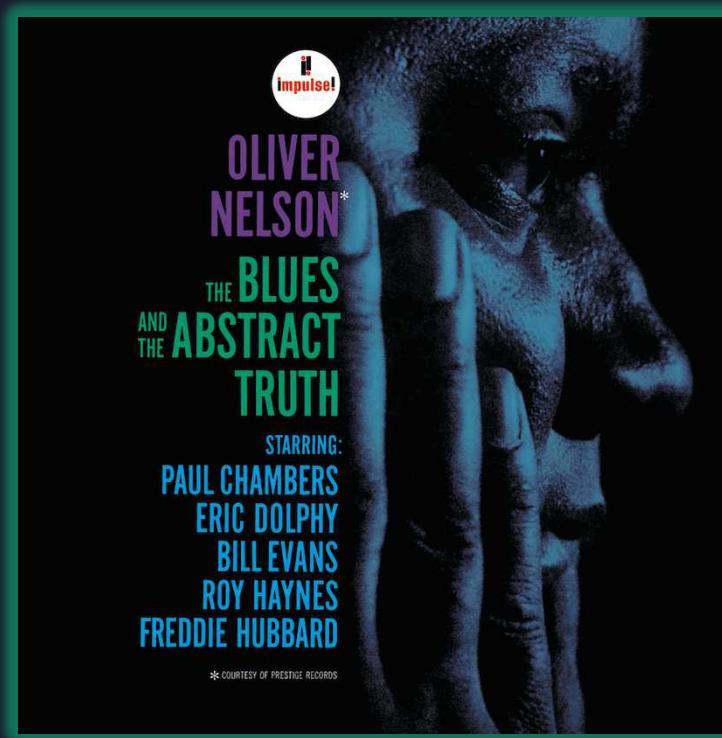


Sources: MRC Data, Nielsen Music, Billboard

Motivation



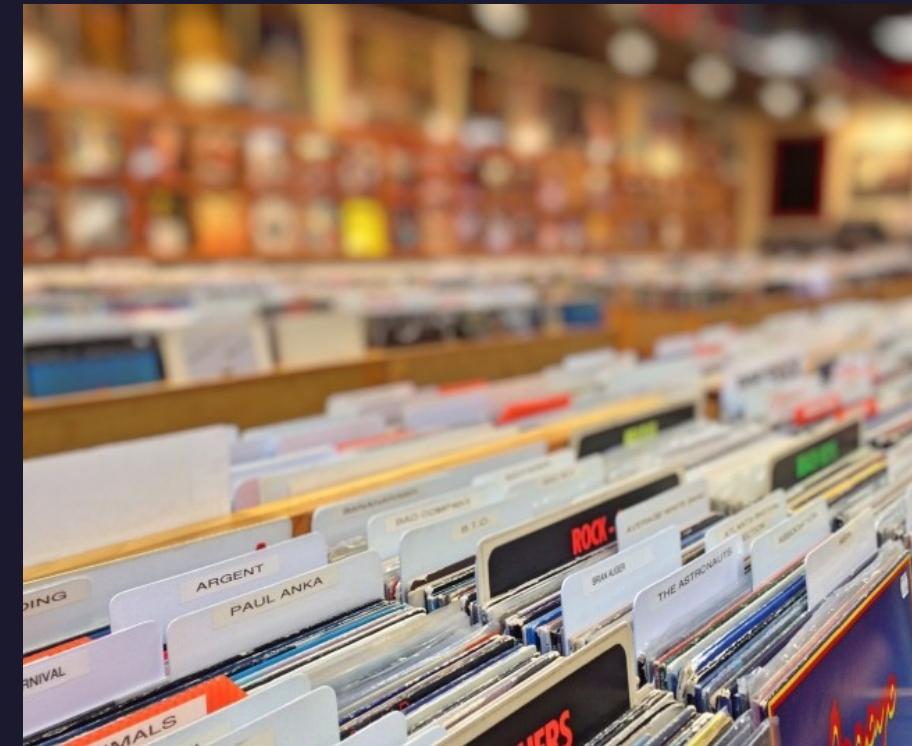
Record Shop A = **\$25**



Record Shop B = **\$45**

Goal

- Help **buyers** and **sellers** estimate a **target price** for Vintage Jazz Vinyl
- Better understand important **features** that impact the **resale price**



Data Collection and Features



Data Source: **Discogs**

Genre: Jazz

Price: \$20+

Release Year: 1950-1970

Features

Album

**Record
Label**

Album
Ratings

Artist

Price

Seller
Rating

Wants

Haves

Release
Date

**Media
Grade**

Sleeve
Grade

Notes

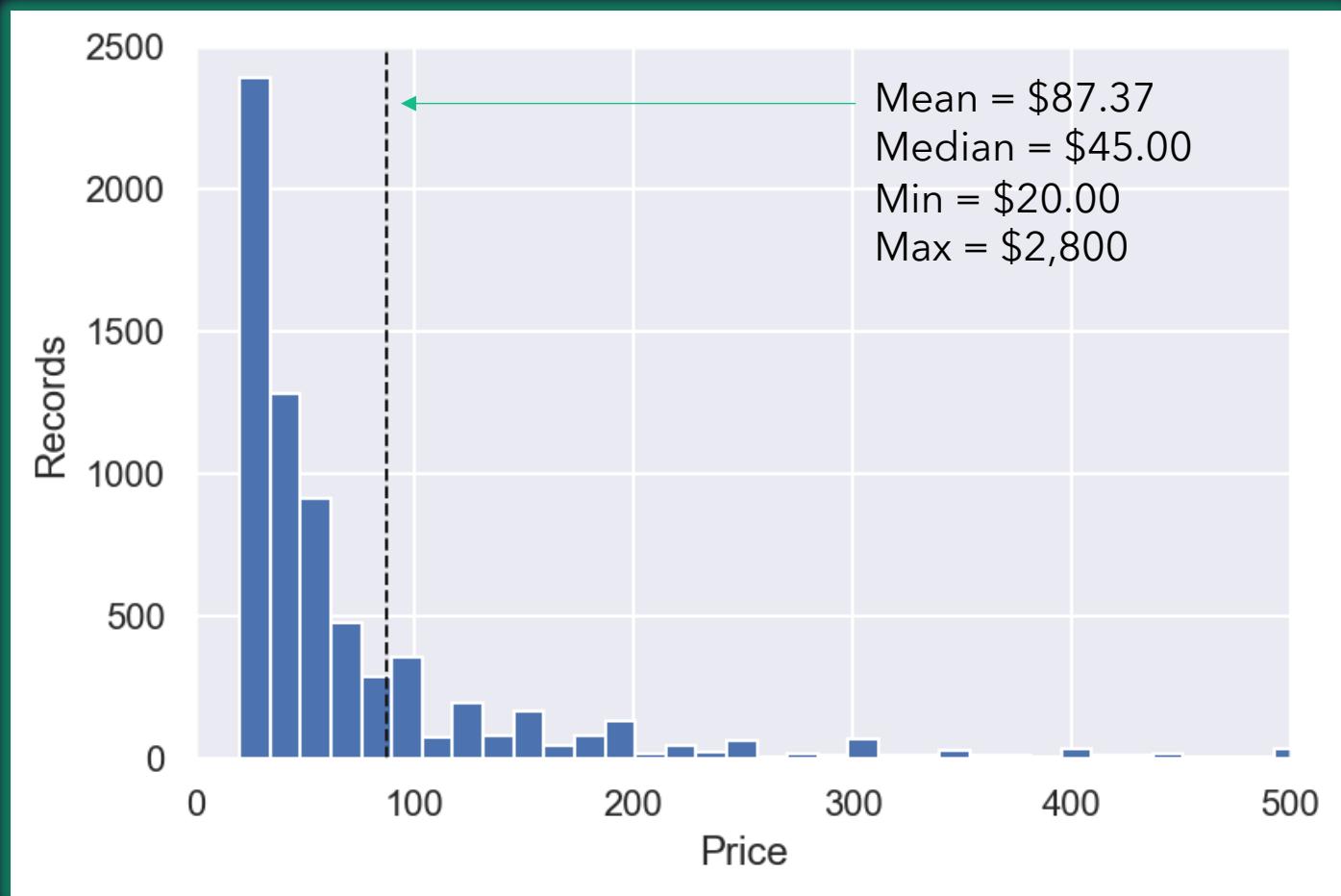
~15
Features

~8,000
Records

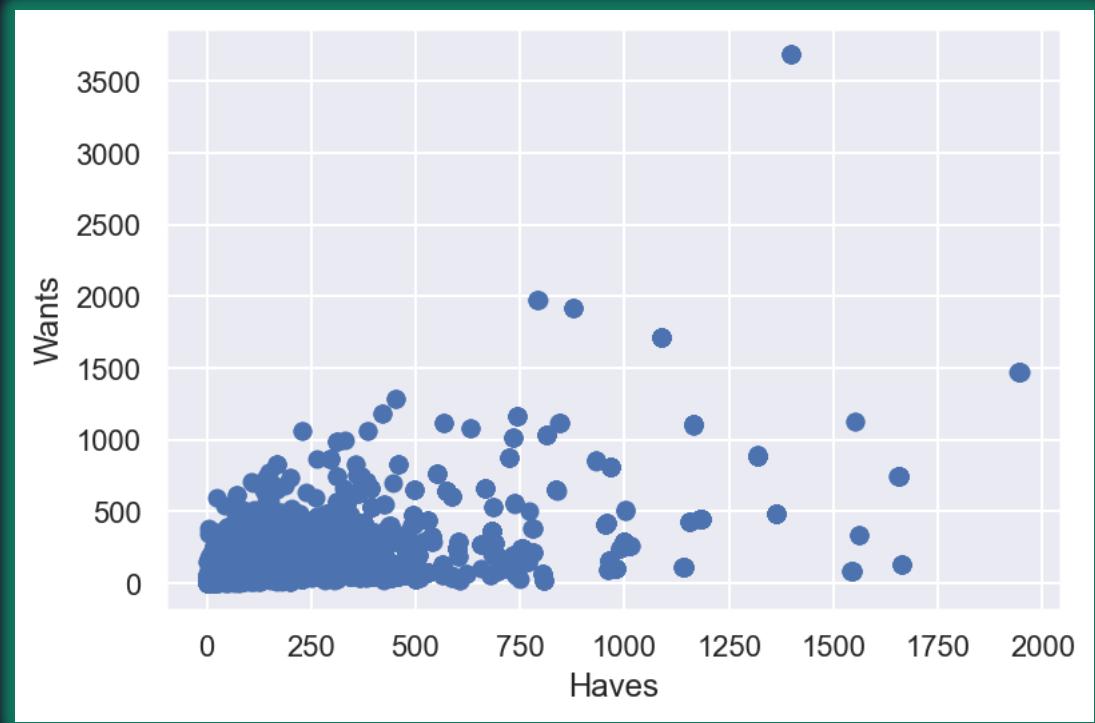
EDA & Model Selection



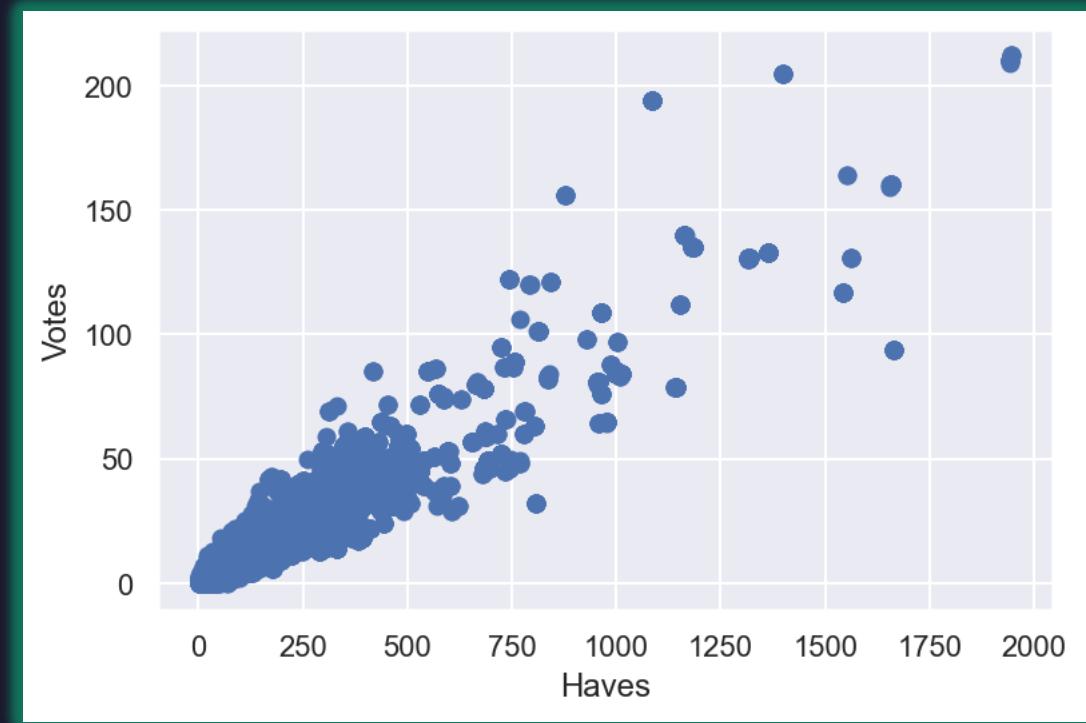
EDA - Resale Price



EDA - Feature Collinearity



Haves by Wants



Haves by Number of Ratings

Feature Selection (OLS)

	coef	std err	t	P> t	[0.025	0.975]
xnotes	0.2741	4.211	0.065	0.948	-7.982	8.530
recyear	-9.6688	0.563	-17.171	0.000	-10.773	-8.565
rating	7.6184	1.966	3.874	0.000	3.763	11.474
votes	-1.5395	0.352	-4.376	0.000	-2.229	-0.850
haves	-0.0245	0.030	-0.818	0.413	-0.083	0.034
wants	0.3253	0.017	19.592	0.000	0.293	0.358
seller_rating	373.8893	36.966	10.114	0.000	301.418	446.360
media_cond_num	36.8863	2.517	14.655	0.000	31.952	41.821
sleeve_cond_num	4.5317	2.210	2.051	0.040	0.199	8.864
reclabel2_Blue Note	106.3759	9.768	10.890	0.000	87.225	125.526
reclabel2_Columbia	8.5879	10.768	0.798	0.425	-12.522	29.698
reclabel2_Impulse!	15.2091	12.592	1.208	0.227	-9.477	39.895
reclabel2_Other	8.4935	9.625	0.882	0.378	-10.377	27.364
reclabel2_Prestige	43.6859	10.370	4.213	0.000	23.355	64.017
reclabel2_Riverside Records	3.9393	11.343	0.347	0.728	-18.298	26.177
reclabel2_Vee Jay Records	16.1354	19.349	0.834	0.404	-21.799	54.070
reclabel2_Verve Records	-4.1191	13.479	-0.306	0.760	-30.545	22.307

Features Reconsidered

Notes: Keywords, mostly noise

Haves: Collinear with votes

Seller Rating: Majority of ratings were at or near 100%

Features

Album

**Record
Label**

Album
Ratings

Artist

Price

Seller
Rating

Wants

Haves

Release
Date

**Media
Grade**

Sleeve
Grade

Notes

~15
Features

~8,000
Records

Features Selection

**Record
Label**

Album
Ratings

Price

Release
Date

**Media
Grade**

Wants

Sleeve
Grade

Haves

Created Dummies
for **Record Label**

Created a **Wants to
Haves** ratio as
weight

Modeling Methods

Type	R^2	MAE	RMSE
OLS	0.233	\$59.15	\$112.16
Linear Regression Poly, Int, Stand	0.462	\$50.68	\$110.72
Lasso CV, Poly, Int, Stand	0.464	\$48.58	\$128.29
Ridge CV, Poly, Int, Stand	0.469	\$48.70	\$123.59
TEST SET Ridge CV, Poly, Int, Stand	0.377	\$50.80	\$110.24

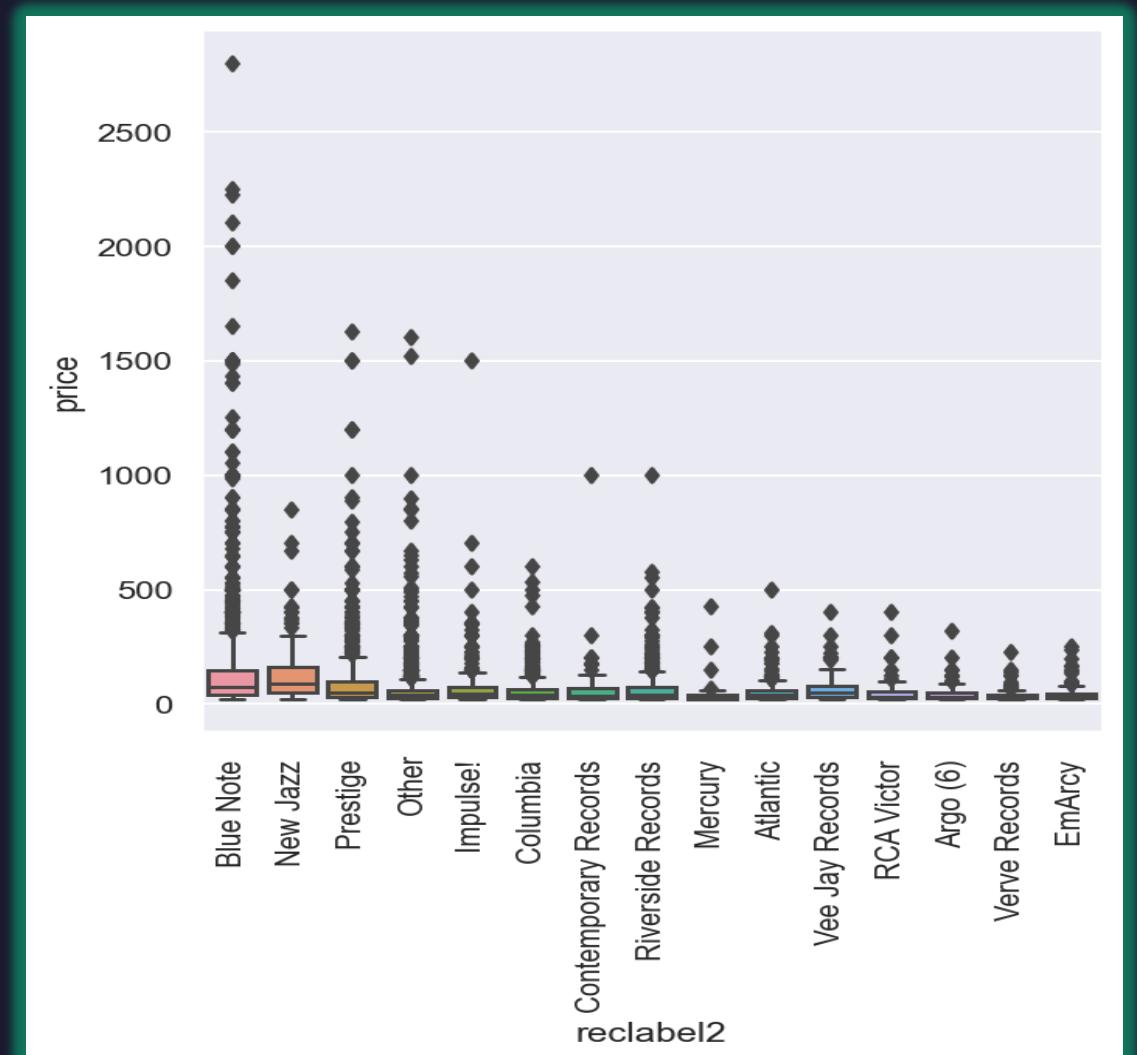
Key Features Based on Model Coefficients

**Blue Note
Records**

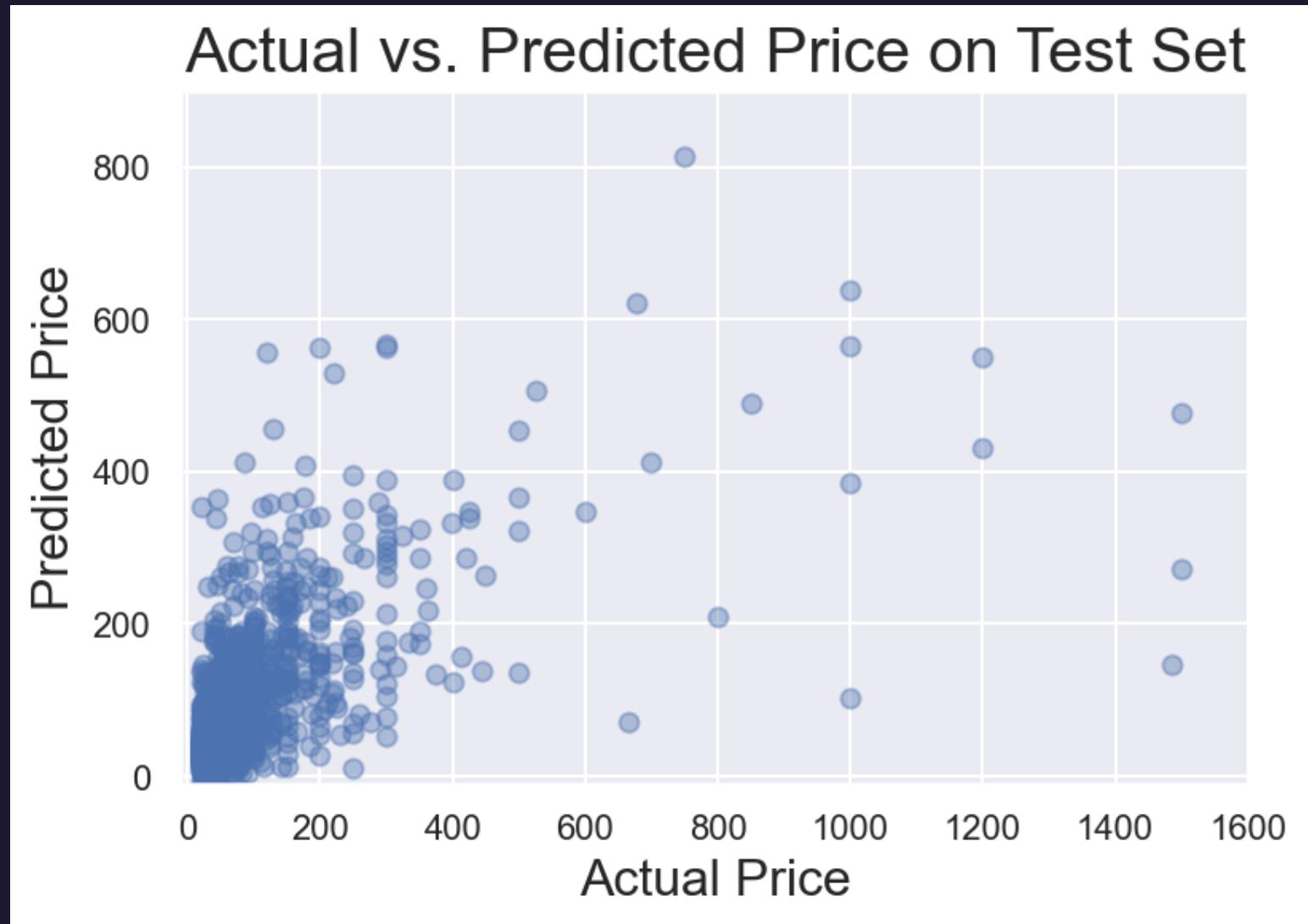
**Number of
Wants**

Seller Notes not
helpful

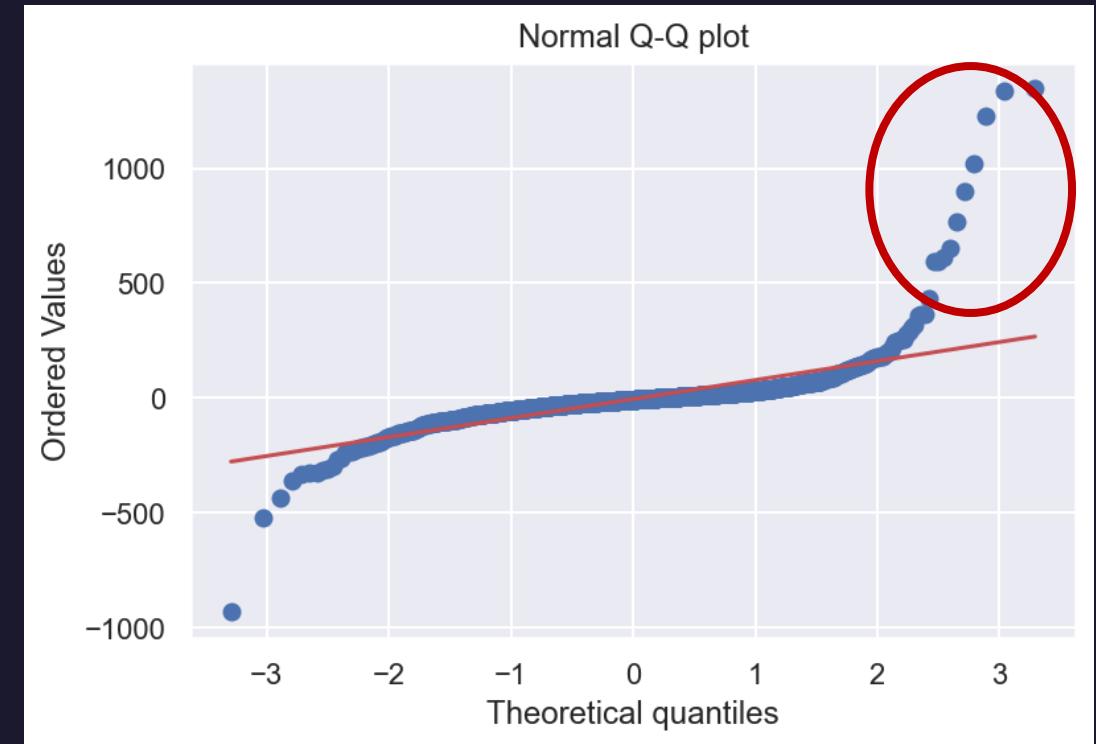
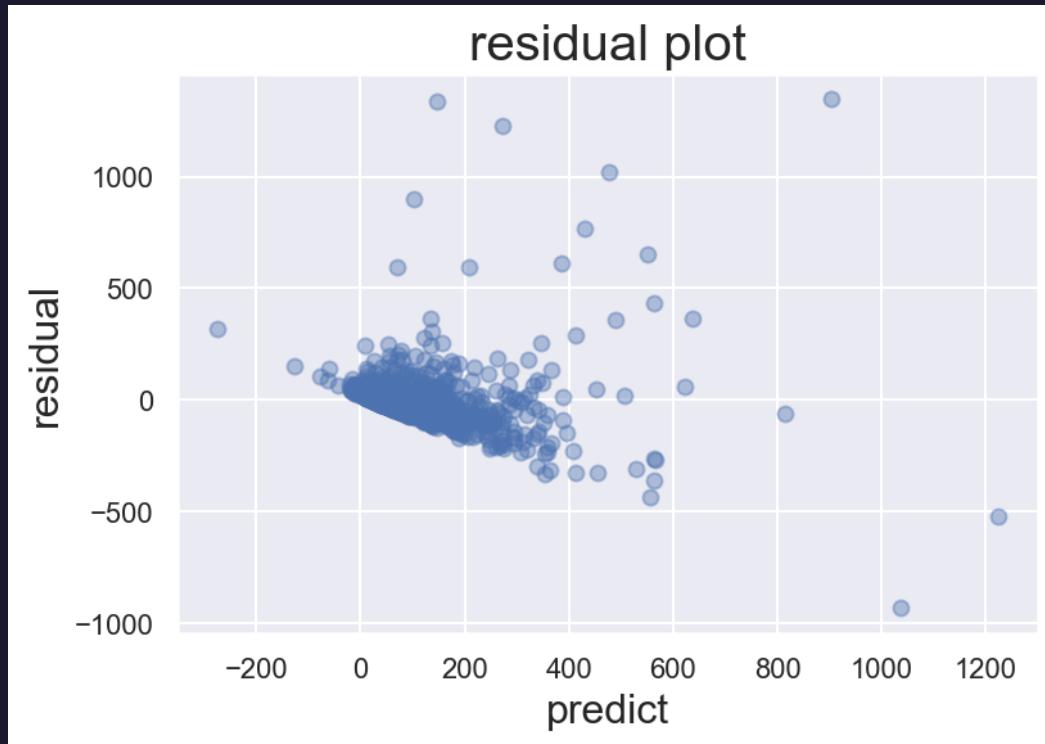
Media Grade



Results



Residual and QQ plots



Future Improvements



Future Improvements

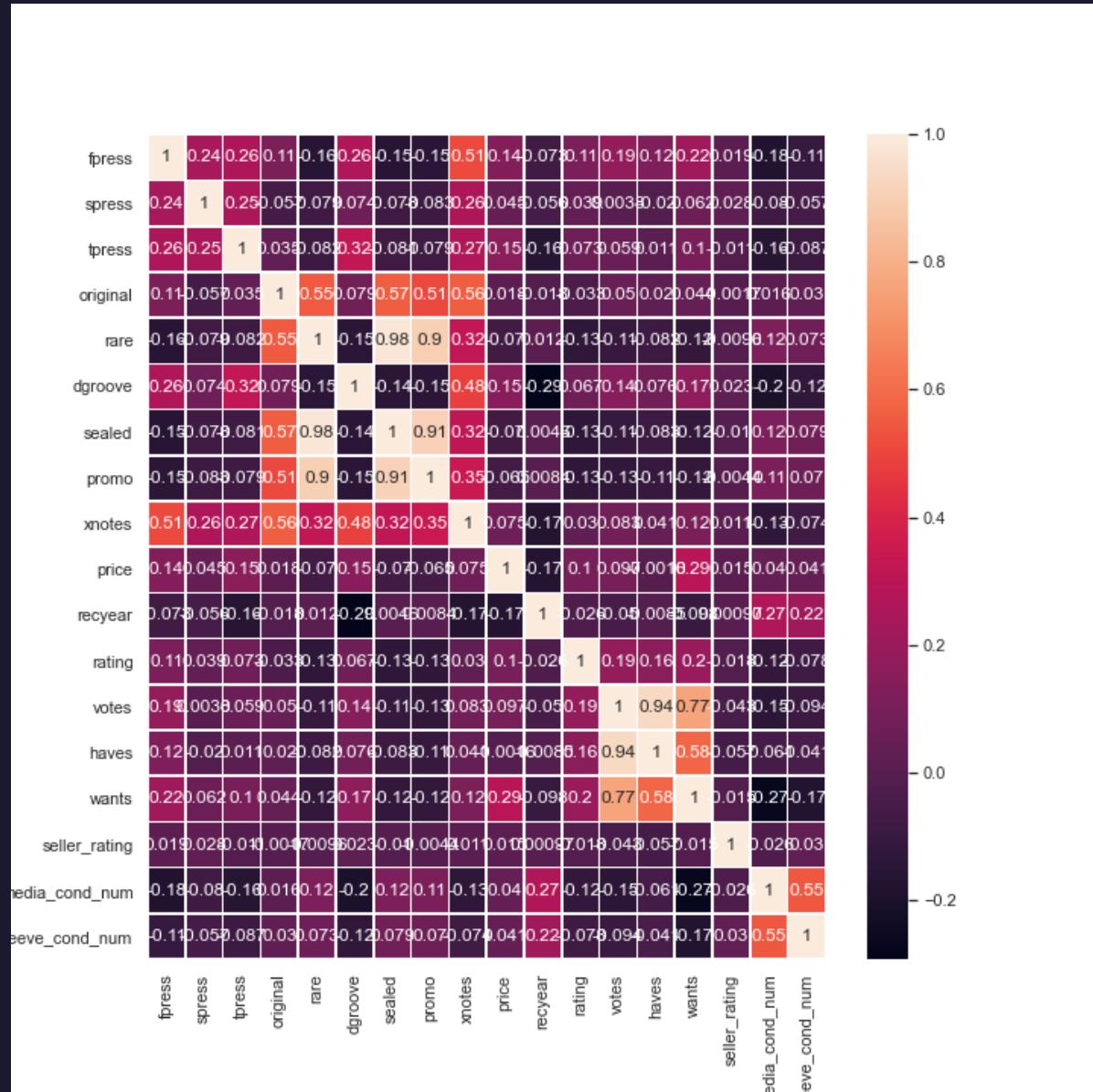
- Transform Price to a log scale
- Search more key-words in Notes section
- Add data from e-bay
- Weight albums by how frequent in market

Thank you!



- Appendix

Correlation Heatmap



Pair Plots

