WINE WITNESS

A DATA-DRIVEN APPROACH TO ITEM PROMOTION & MERCHANT ONBOARDING AT VIVINO.COM

Dmitri Levonian

April 2020

VIVINO

VIVINO STORY

- Founded in 2010 in Denmark as an app that recognized photos of wine labels
- Raised $56 m to date from VCs, currently Round C
- 43 million registered users
- 55 million reviews

- In 2017, launched a marketplace for buying and selling wines
- Collects 10-15% commission

TODAY

- Available when accessing from the U.S. :
  - 380,000 wines reviewed
  - 14,400 wines sold ← ***scraped***

KEY QUESTIONS

1. How to identify wines to be promoted to customers as great price/quality?
2. Given that 96% of wines are not available for sale, which ones to onboard?

WINES IN STOCK: 4% OF TOTAL

La Crema

Willamette Valley Pinot Noir

Red wine from Willamette Valley · United States

4.0 ★★★★☆ 4533 ratings Add to Wishlist

Among top 2% of all wines in the world (2015 Vintage)

$22.98

Price is per bottle

12

2017

Add to cart

You are saving 20% on every bottle by buying a case of 12.

You get FREE shipping on orders over $99 with Yiannis Wine Shop

Estimated between Fri, May 01 and Tue, May 05

Sold by Yiannis Wine Shop

Show all buying options

WINES OUT OF STOCK: 96% OF TOTAL

Château du Mont

Graves Rouge

Red wine from Graves · France

3.9 ★★★★☆ 171 ratings Add to Wishlist

Among top 5% of all wines in the world (N.V. Vintage)

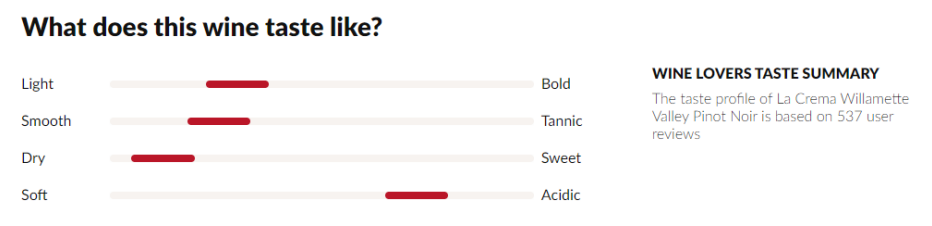
Vivino doesn't sell this wine.

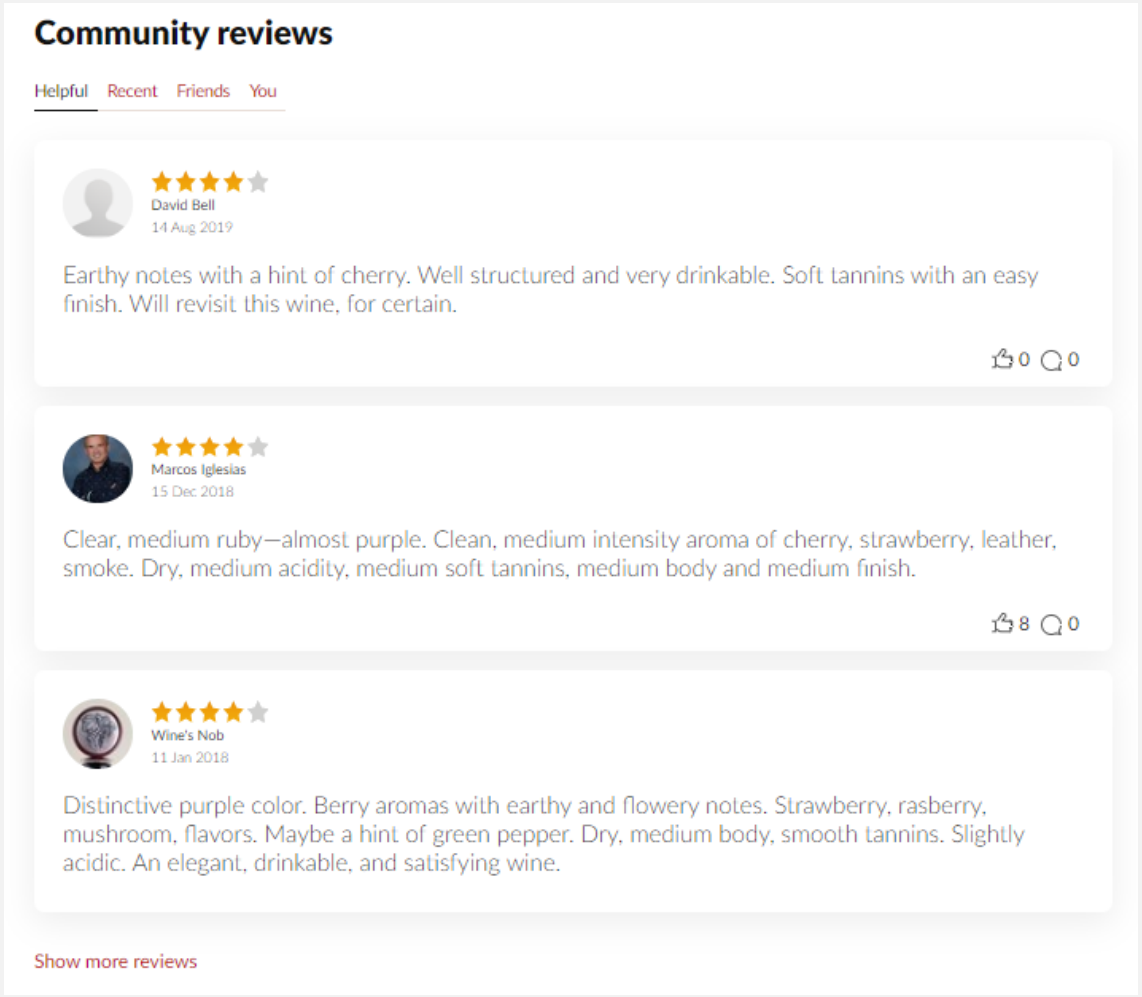
ⓘ Also, we searched other online shops, but couldn't find it for sale anywhere.

TOTAL OF 18 FEATURES PER WINE

18 fields scraped per wine item:

Numeric	Categorical	Descriptive
Price	Winery	Wine ID
# of ratings	Vintage	URL
Avg rating	Wine type	Review 1
Light-Bold	Wine style	Review 2
Smooth-Tannic	Country	Review 3
Dry-Sweet	Region	Image link
Soft-Acidic		



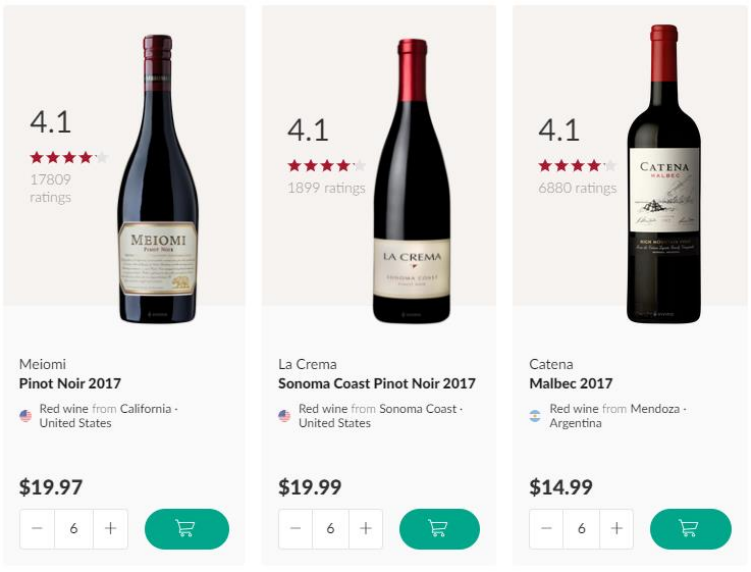


DATA-DRIVEN APPROACH

Q1. CULTIVATE DEMAND

- Once the merchant lists the wine at Vivino at a certain price, which ones to promote to customers as great price/quality?
- How accurately can we predict the 'true' price given the scraped features?

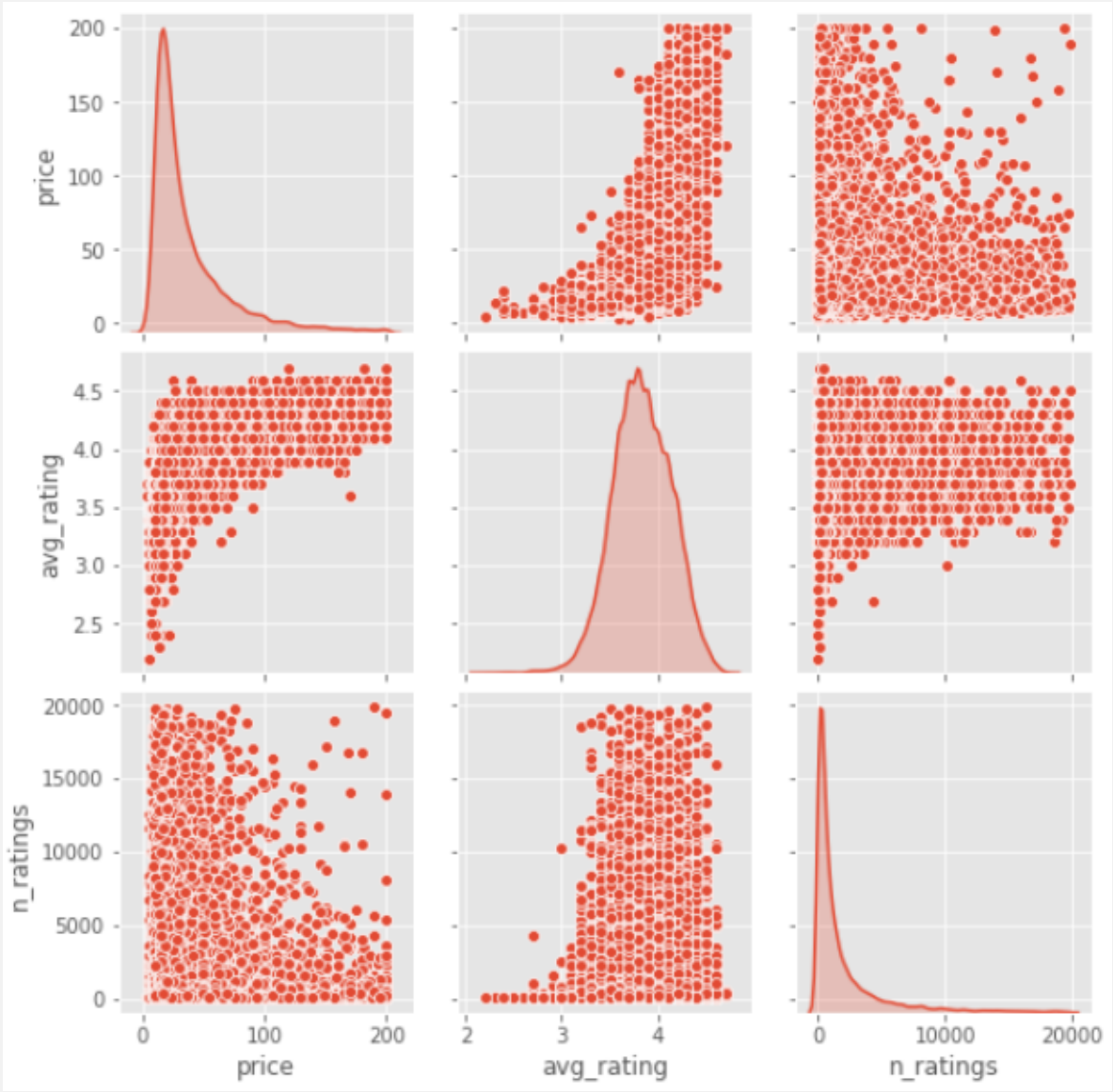
Are these good deals?



Q2. EXPAND PRODUCT OFFERING

- From the 96% of wine items not currently sold by Vivino, which ones should be prioritized for onboarding?
  - Less expensive ones, go for the volume? Rare finds rated 4.8 stars, go for the price?
  - U.S. wines or imported?
- Assumption: potential demand is proportional to the number of ratings
- This may be an oversimplification, will need actual sales data to verify

Q1: CAN PRICE BE INFERRED?



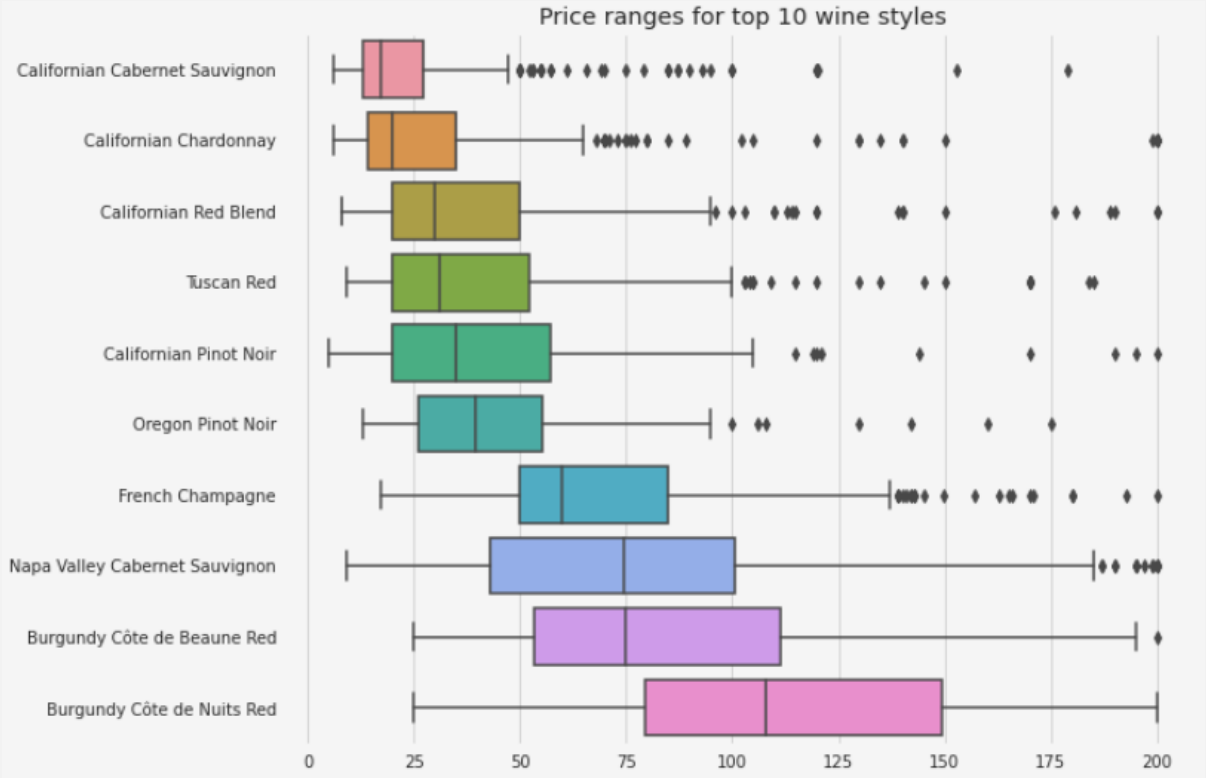
- Prices correlate with the average rating (67% correlation)
- However, there is still lots of variability – a great 4.2-star wine can be easily priced at $20 and at $200
- Number of ratings correlate slightly negatively with the price, and the relationship looks like the typical inverse price/quantity curve
- NB: non-linear!

- A lot of variability in the wine prices comes from perceived quality

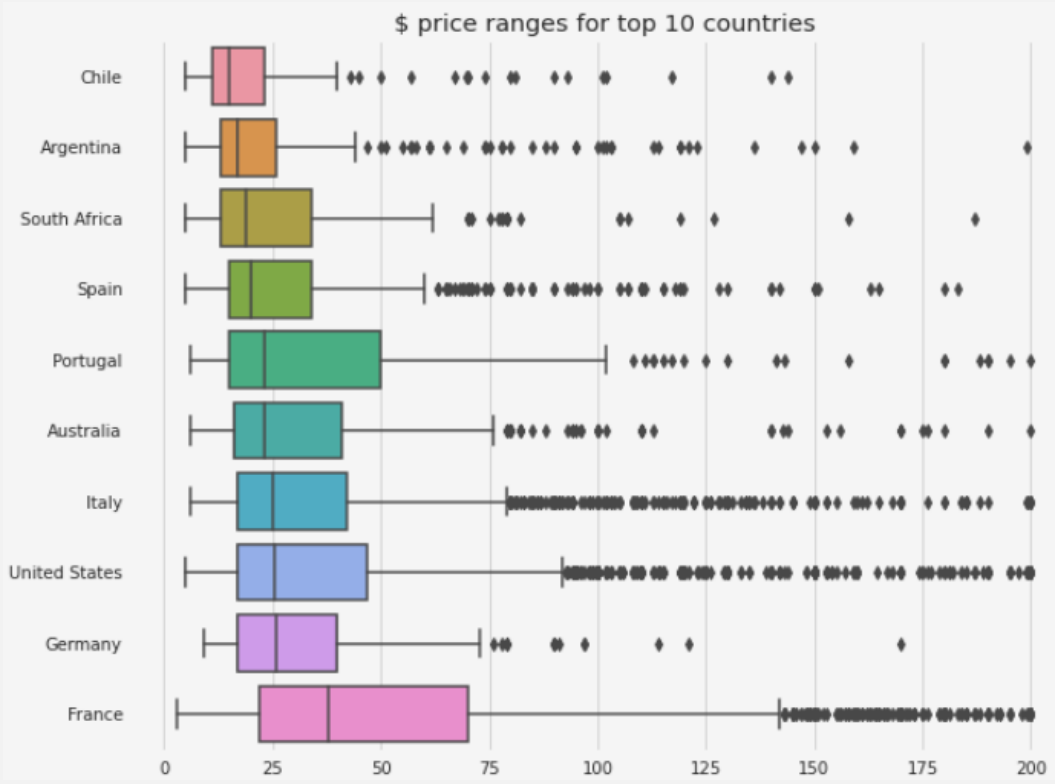
Q1: CAN PRICE BE INFERRED?

- Winery and vintage certainly influence price but were removed from the features b/c they are too granular: e.g. less than 3 wine items per winery, which will definitely overfit
- Wine styles and countries of origin clearly have explanatory power:

Price ranges for top-10 wine styles (varieties)

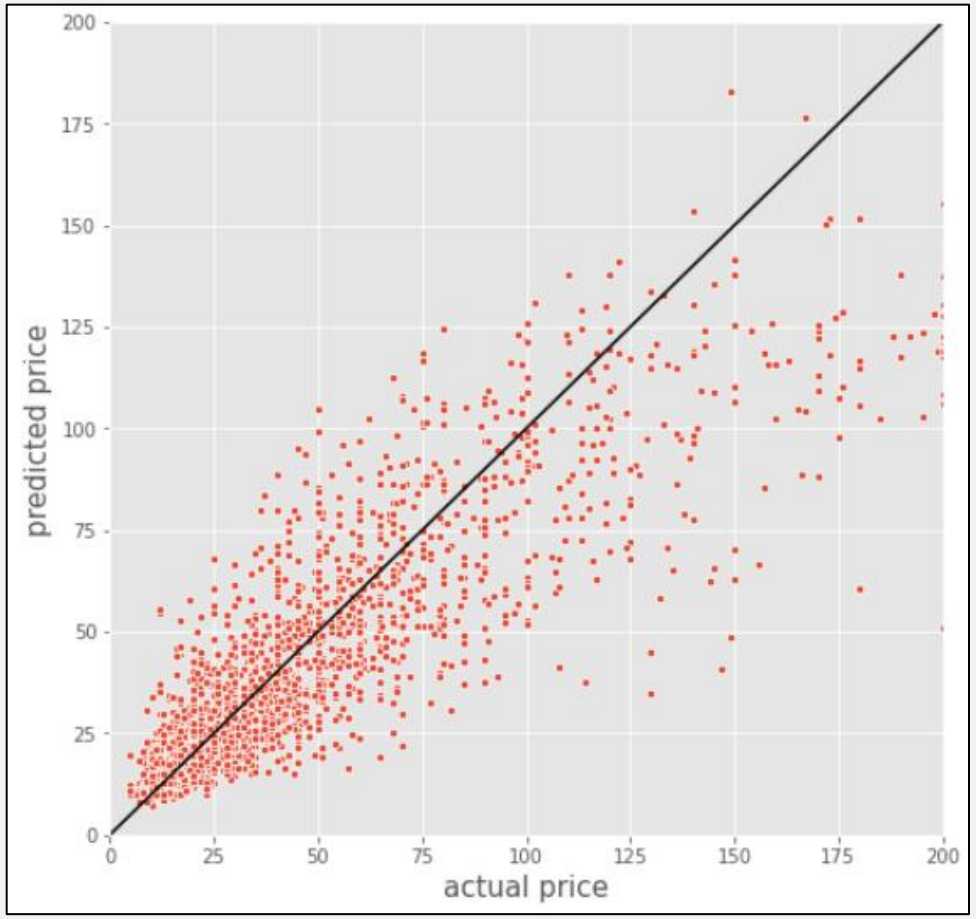


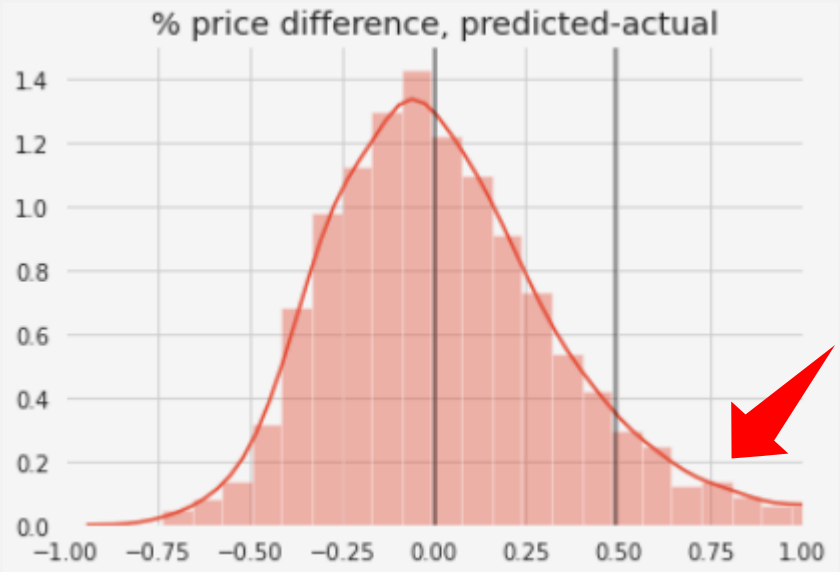
Price ranges for top-10 countries



Q1: HOW ACCURATELY?

- The best model is a simple fully-connected DNN with just one hidden layer
- All features are important! Discarding any set (e.g. country/region) weakens explanatory power
- The model can predict the price with Mean Absolute Error of about $10, RSquared~75%





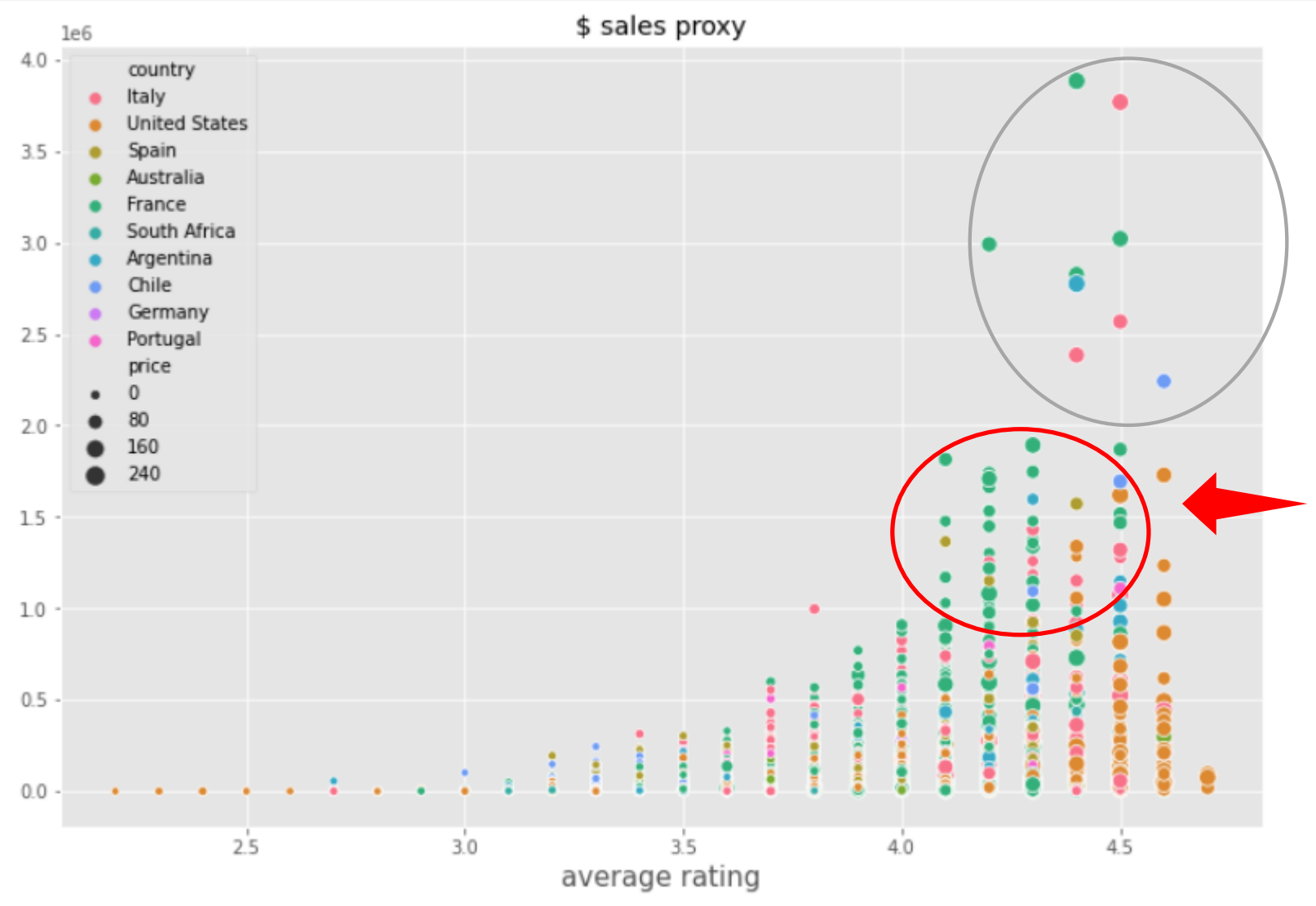
10% staged for human inspection

Why are these wines priced at say $60 when the model predicts $90+?

Candidates for promotion

	ols	ridge	relu	dnn
rsquared_train	7.083861e-01	0.698801	0.623952	0.744241
rsquared_test	-2.345877e+20	0.658696	0.612480	0.756403
mae_train	1.206122e+01	12.509818	12.145876	9.923902
mae_test	4.403333e+10	13.668473	12.880139	10.243977

Q2: WHICH WINERIES TO ONBOARD?



outliers, ignore

Sweet spot

100s of wines including those not currently sold by Vivino

Higher side of the spectrum, rated 4.1-4.5 stars

French & Italian wines, $60+ retail

CONCLUSIONS

- The model trained on limited features explains about 75% of variability in the wine prices
- For ~10% of wines, the model predicts a 1.5x higher price than the actual:
  - Flag such items for human verification
  - If verified, promote and recommend as good deals!
- For less expensive wines of $15-25, the error of $10 is too large. Need larger dataset + more predictive features

- Prioritize resources to list highly rated (4.2-4.5), more expensive wines
- Focus on France and Italy
- NB: a key assumption on proportionality of number of ratings and bottles sold – needs to be verified on actual sales data

NEXT STEPS

- Deploy NLP to analyze customer reviews, produce sentiment analysis, 'flavor vectors'
- Integrate network data into wine suggestions (influential users' opinion)
- Integrate exogenous data – social media, competitors, wine experts

- Wine is a $300 b global industry, $70 b in the U.S. Less than 5% is e-commerce
- Vivino will need to rethink itself as a data company to become the 'Amazon of wine'

- Likely already underway:
- *"Vivino's machine learning algorithms look at the scanning and rating behavior of each individual user to determine wines they are most likely to be interested in (scans) and love (ratings)"*

Heini Zachariassen, founder of Vivino