WINE AND THE INFLUENCE OF CLIMATE

Katherine Hickok

March 02, 2022

THE PLAN

BACKGROUND

PROBLEM STATEMENT

DATASETS

EDA and MODELLING

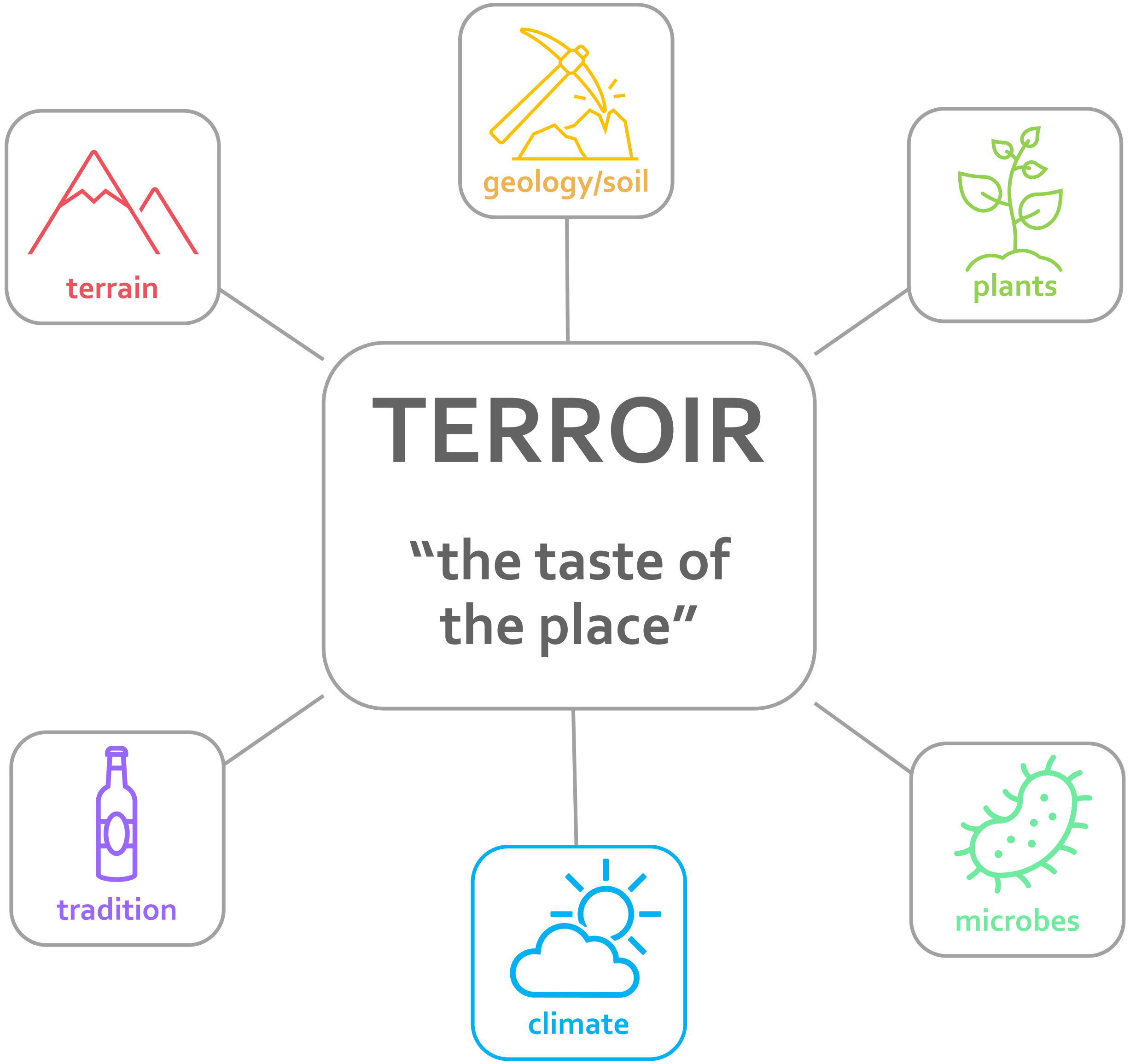
WHAT’S NEXT

“Hardly did it appear, than from my mouth it passed into my heart.”

-- Abbe de Challieu, 1715

https://www.cejavineyards.com/Wine-Quotes

https://giphy.com/gifs/teamcoco-conan25-wine-tasting-napa-valley-UuwMQQksoKtIMa7L3N



TERROIR

“the taste of the place”

terrain

geology/soil

plants

tradition

climate

microbes

https://unsplash.com/photos/sj4M_cwUGBo https://www.allrecipes.com/article/what-is-terroir/ https://www.afoodieworld.com/tersina/7592-rewriting-wine-101-terroir

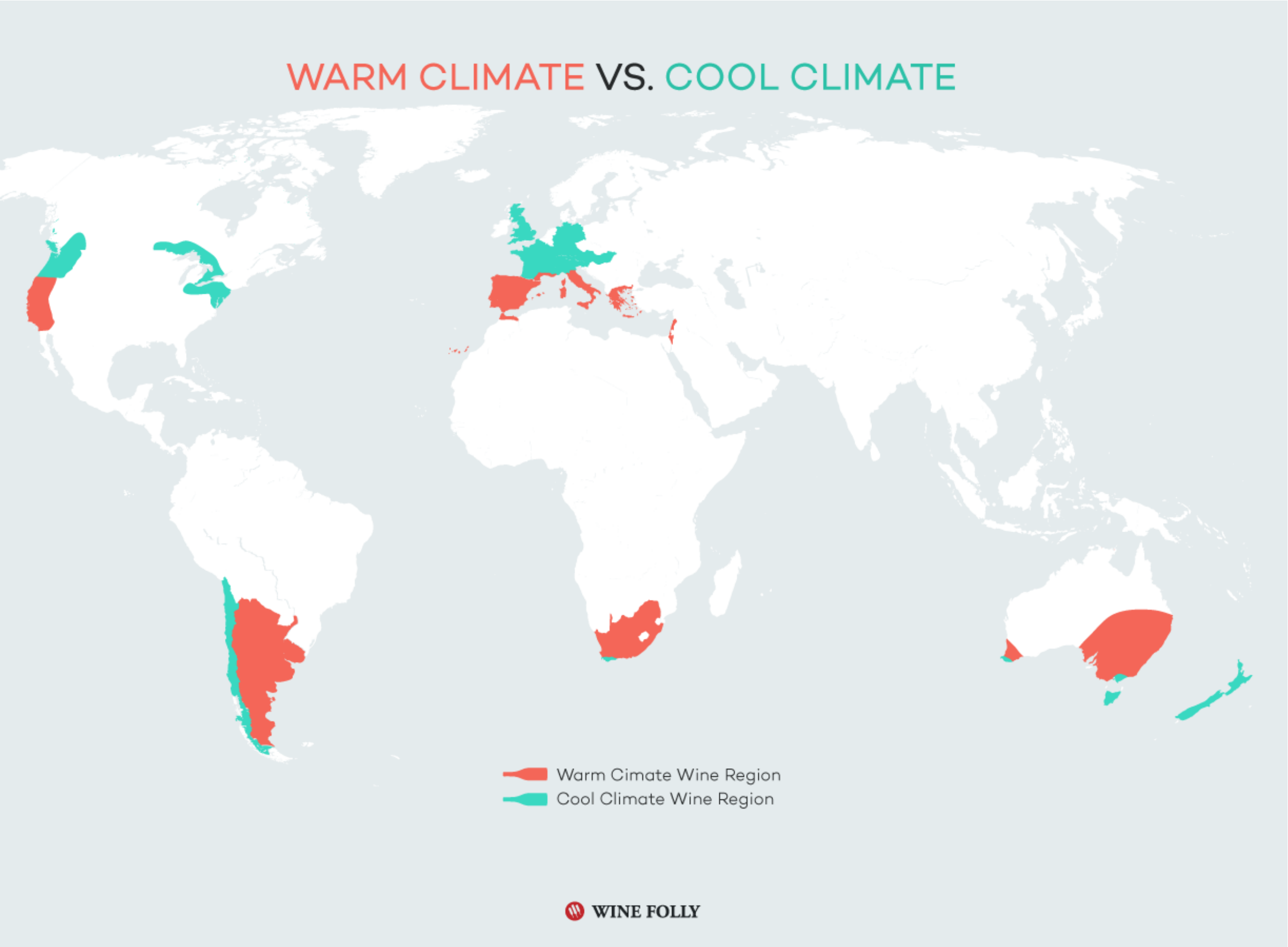
WARM

Sweet

Less acidity

Higher alcohol

zinfandel, grenache, syrah



COOL

Tart

More acidity

Less Alcohol

riesling, pinot noir, sauvignon blanc

*** Don’t forget microclimates!**

https://www.winegame.com/blog/2019/2/15/coolvswarm

https://winefolly.com/tips/warm-climate-vs-cool-climate-wine/

CAN WE PREDICT WINE QUALITY BETTER THAN THE BASELINE?

WINE ENTHUSIAST®

Wine Reviews Data

Dataset from scrape wine reviews

SamuelMcGuire • updated a month ago (Version 1)

- Kaggle dataset
- over 323,000 wines
- rating, price, alcohol content, varietal, appellation, etc.

GHCN Daily

NOAA

Global Historical Climatology Network Daily Weather Data

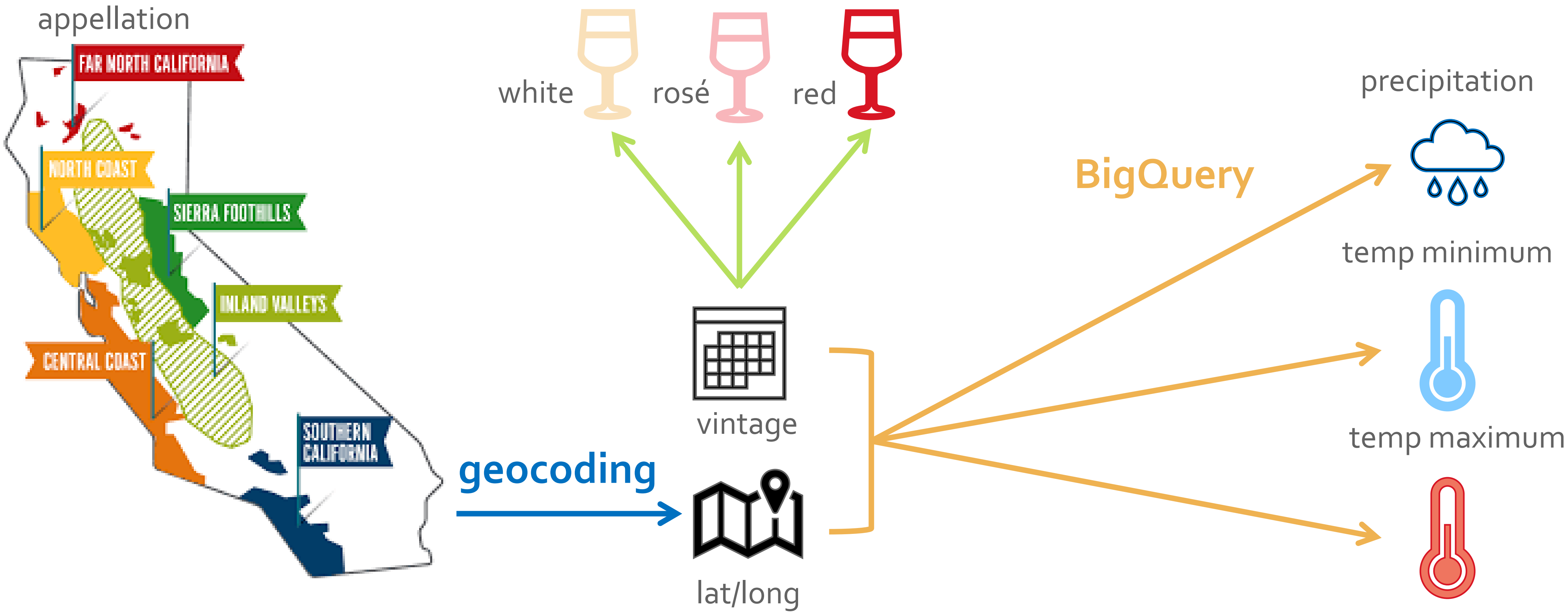
- weather station
- precipitation (mm)
- maximum temperature (ºC)
- minimum temperature (ºC)

Google BigQuery

https://www.kaggle.com/samuelmcguire/wine-reviews-data

https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily

HOW DO WE DETERMINE CLIMATE?



appellation

geocoding

white

rosé

red

vintage

lat/long

BigQuery

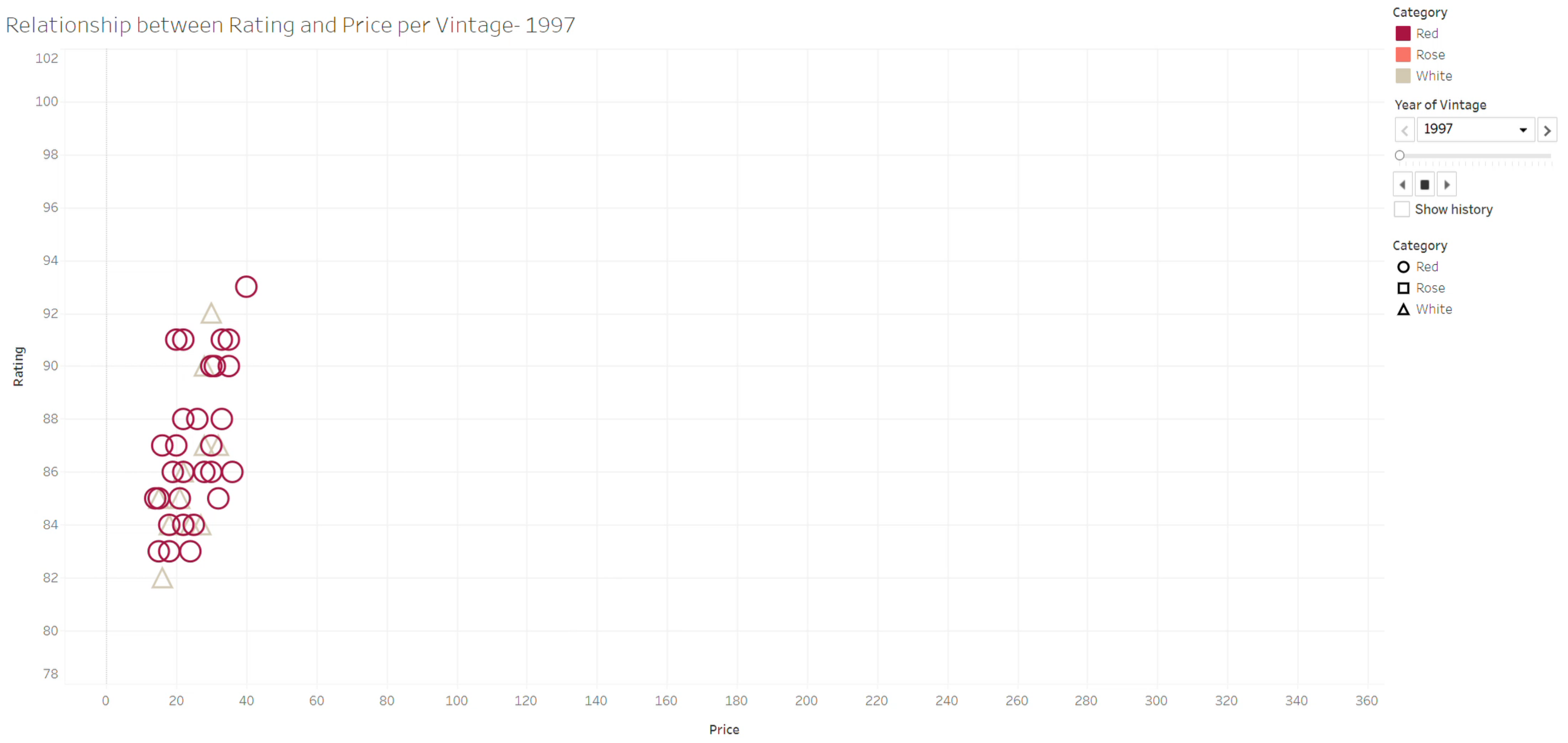
precipitation

temp minimum

temp maximum

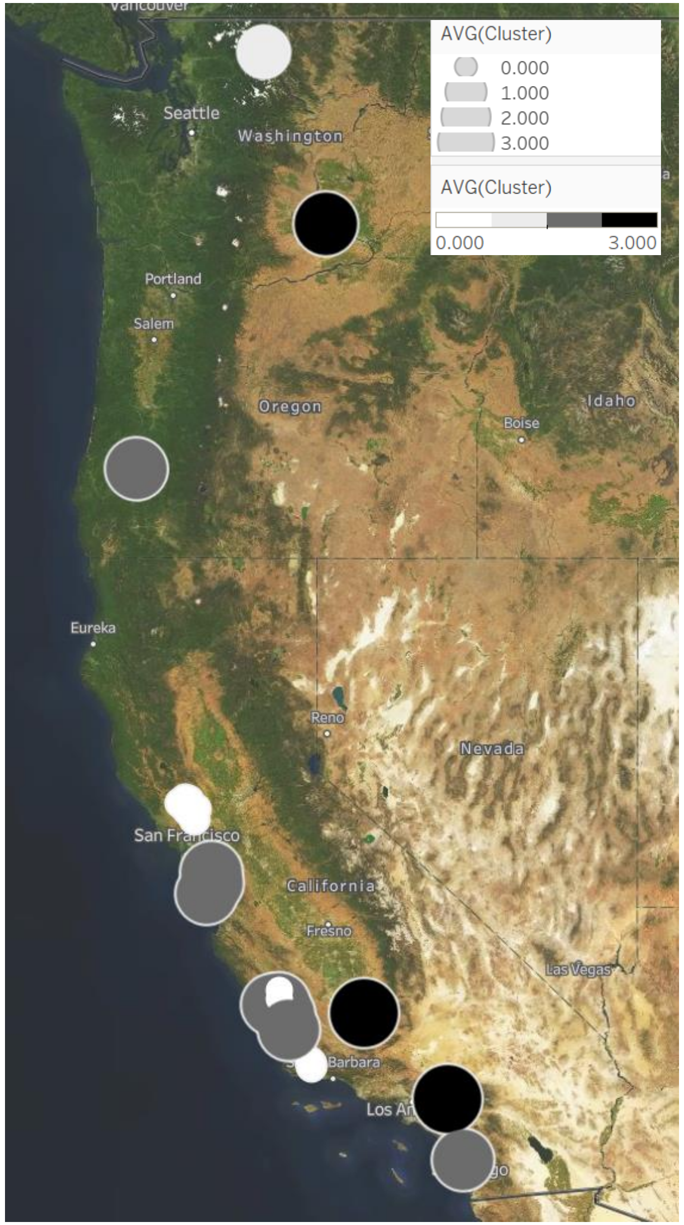
https://discovercaliforniawines.com/discover-california/

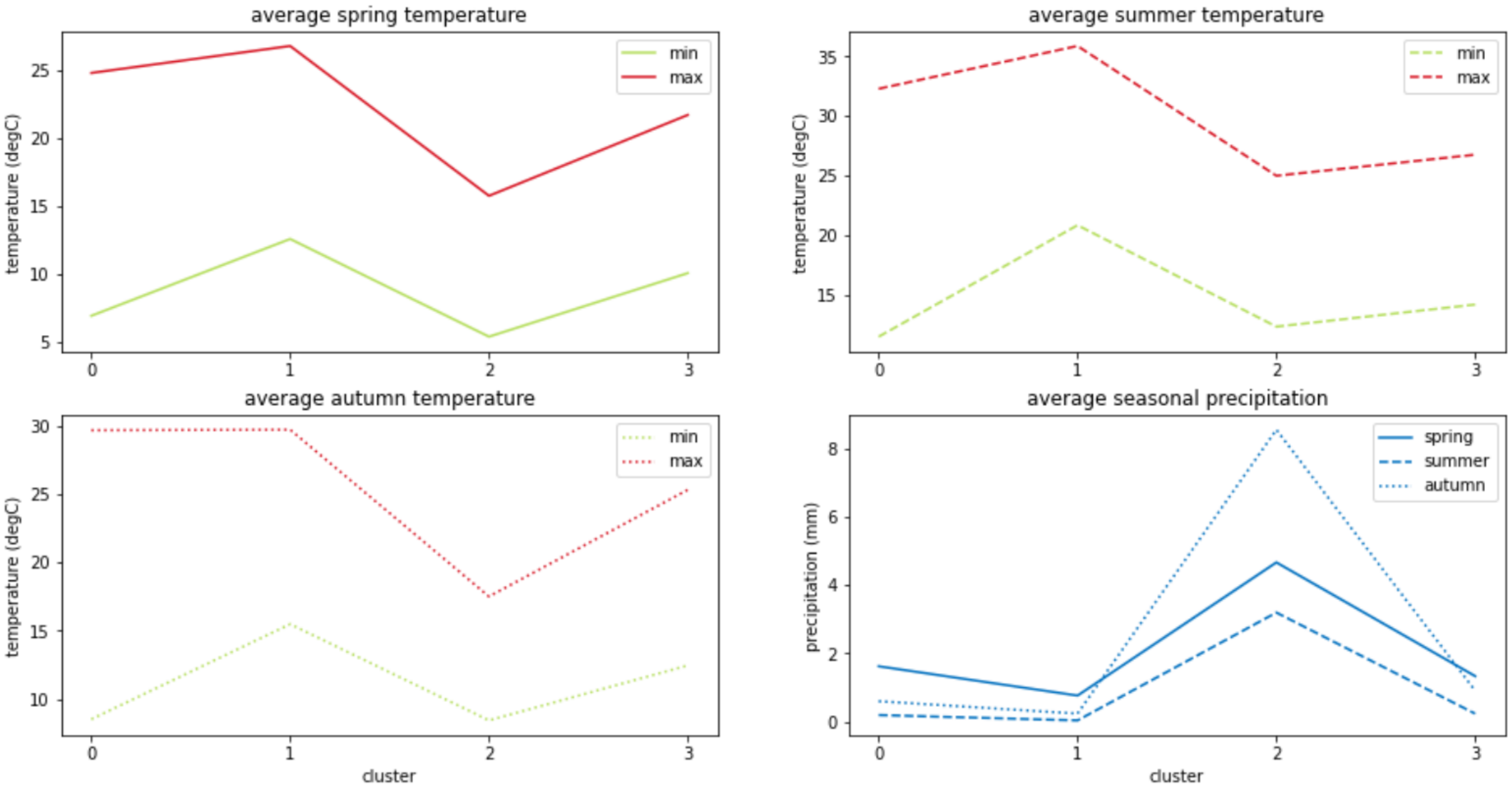
PRICE = QUALITY, RIGHT?



Year of Vintage: 1997

https://www.brightcellars.com/blog/is-expensive-wine-worth-it/





cluster 0- greatest temperature range
cluster 1- highest temperature min and max, cheapest
cluster 2- more precipitation, higher latitude
cluster 3- average

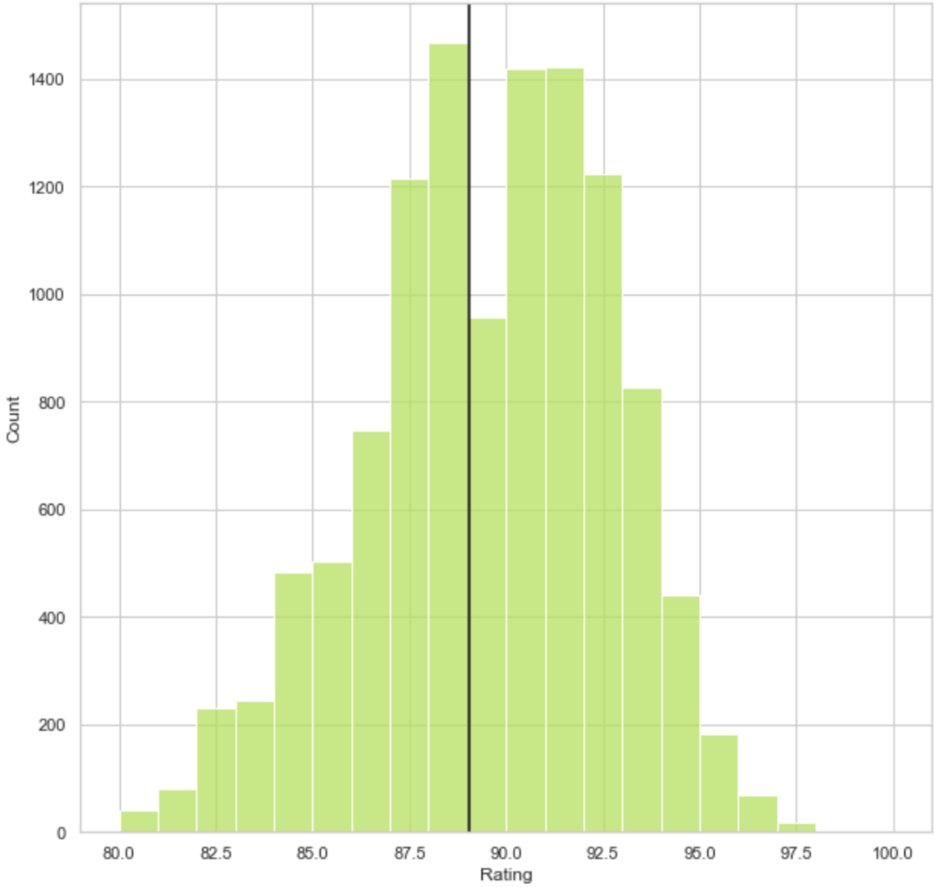
WHAT DO THE MODELS SAY?

likelihood of predicting most common value: 0.1269

MODEL	DESCRIPTION	TRAIN R^2	TEST R^2
LinReg, StandardScalar	**only monthly climate data, rating**	**0.1846**	**0.1675**
LinReg, StandardScalar	dummied varietal , category + all numeric data	0.3901	0.3734
GridSearch, Lasso, StandardScalar	found best Lasso parameters	0.39	0.3738
PolyFeatures, Lasso, Standard Scalar	2nd order polynomial features, overfit	0.4841	0.3842

FEATURE TRANSFORMERS?

mean= 89.04



TRANSFORMER	DESCRIPTION	TRAIN R²	TEST R²
StandardScalar	mean = 0, std dev= 1, data already normal	0.39	0.3738
PowerTransformer	deals with heteroskedastic data, changes to normal distribution	0.4016	0.394
QuantileTransformer	**converts distribution to normal or uniform**	**0.3983**	**0.3991**

https://www.analyticsvidhya.com/blog/2020/07/types-of-feature-transformation-and-scaling/

WHAT DO THE MODELS SAY?

likelihood of predicting most common value: 0.1269

MODEL	DESCRIPTION	TRAIN R^2	TEST R^2
LinReg, StandardScalar	**only monthly climate data, rating**	**0.1846**	**0.1675**
LinReg, StandardScalar	dummied varietal , category + all numeric data	0.3901	0.3734
GridSearch, Lasso, StandardScalar	found best Lasso parameters	0.39	0.3738
PolyFeatures, Lasso, Standard Scalar	2nd order polynomial features, overfit	0.4841	0.3842
SVR, QuantileTransformer	**The best!**	**0.4244**	**0.404**

WHAT DO THE MODELS THINK IS IMPORANT?

LASSO MODEL COEFFICIENTS

POSITIVE CORRELATION

- price: 1.157
- vintage: 0.937
- Oct average max temp: 0.788
- Sept average max temp: 0.776
- Oct min temp: 0.685

NEGATIVE CORRELATION

- Oct average min temp: -0.807
- Aug average max temp: -0.668
- May average max temp: -0.361
- June average min temp: -0.326
- May min temp: -0.257

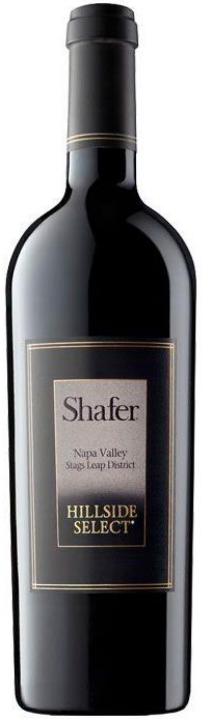
CLIMATE IS JUST THE BEGINNING

- Get more weather data including sunlight, wind, extreme weather events
- Use wines from all over the world
- Join with an NLP of word usage in reviews to match trends in cool vs warm climate descriptors
- Include other terroir variables like geology
- Focus on varietal or appellation or winery, get more microscale

https://unsplash.com/photos/IhOamKjNWwI/

OK, BUT WHAT'S GOOD?

best red



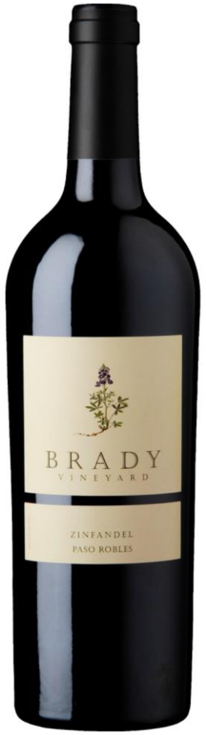
best white

best bang for your buck

recent vintage

>= 95 rating

< $30



www.wine.com

www.vivino.com/