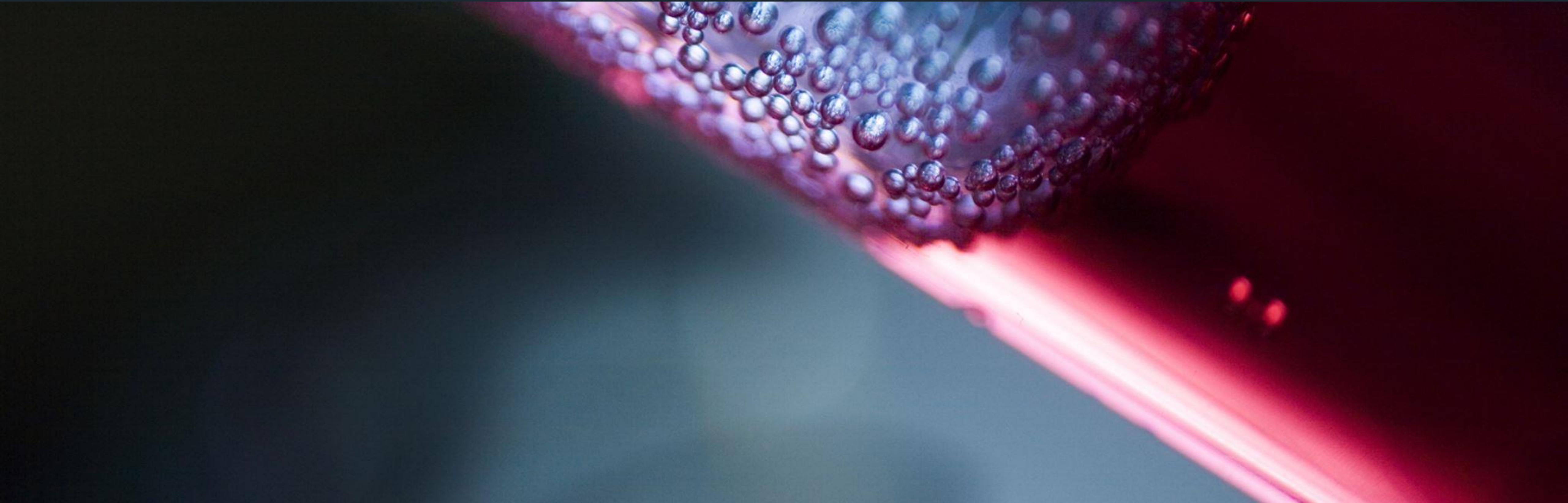
Predicting Optimal Wine Prices with Machine Learning

Name: Gabriel Pantoja



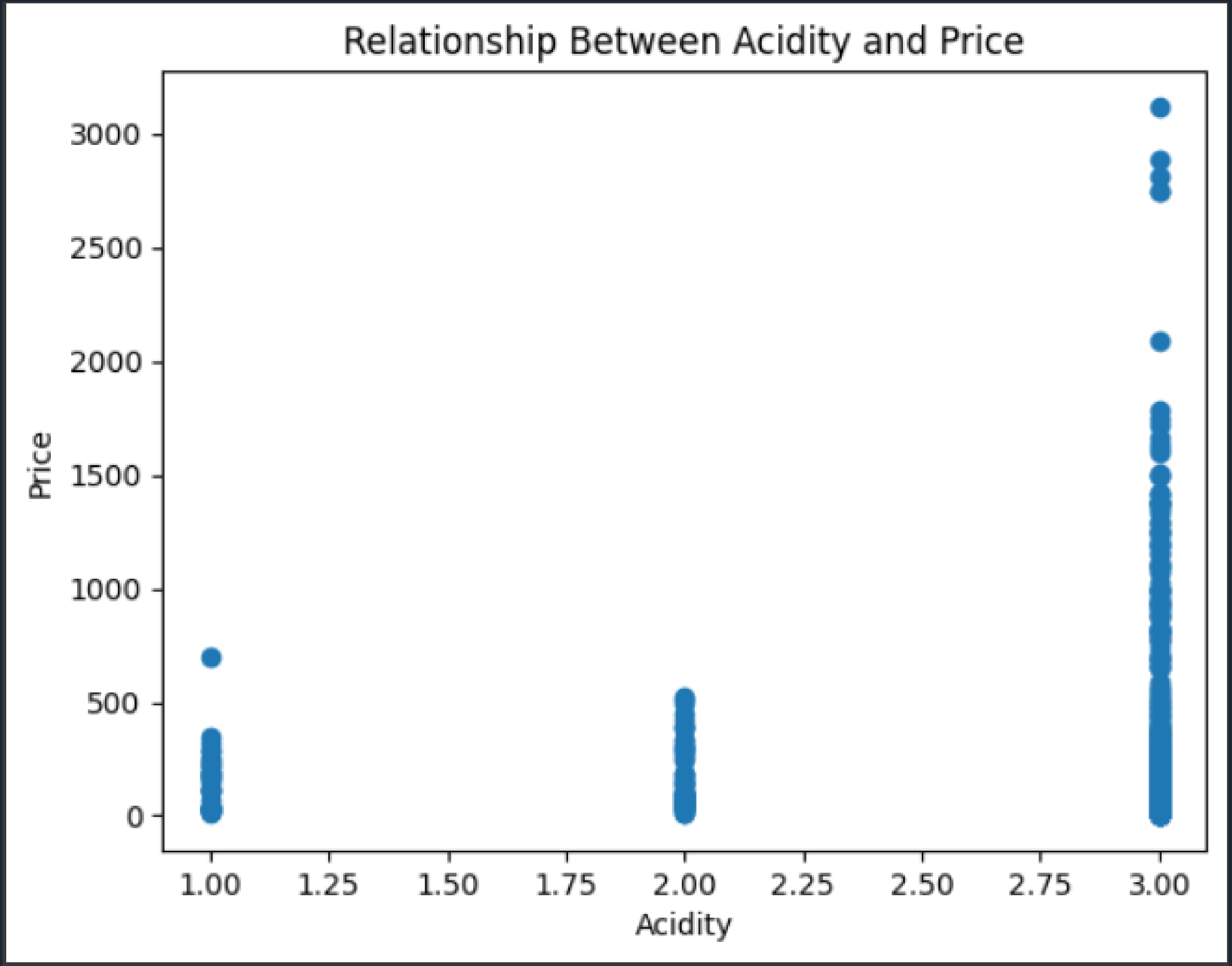
Data Overview

	winery	wine	year	rating	num_reviews	country	region	price	type	body	acidity
0	Teso La Monja	Tinto	2013	4.9	58	Espana	Toro	995.00	Toro Red	5.0	3.0
1	Artadi	Vina El Pison	2018	4.9	31	Espana	Vino de Espana	313.50	Tempranillo	4.0	2.0
2	Vega Sicilia	Unico	2009	4.8	1793	Espana	Ribera del Duero	324.95	Ribera Del Duero Red	5.0	3.0
3	Vega Sicilia	Unico	1999	4.8	1705	Espana	Ribera del Duero	692.96	Ribera Del Duero Red	5.0	3.0
4	Vega Sicilia	Unico	1996	4.8	1309	Espana	Ribera del Duero	778.06	Ribera Del Duero Red	5.0	3.0

- Data source: Public dataset from Kaggle
- Includes measurements like:
  - Winery
  - Wine
  - Year
  - Rating
  - Num_reviews
  - Country
  - Region
  - Price
  - Type
  - Body
  - Acidity
- Also includes wine quality rating by experts on scale of 0 (worst) to 5(best)
- 7500 entries
- 11 features

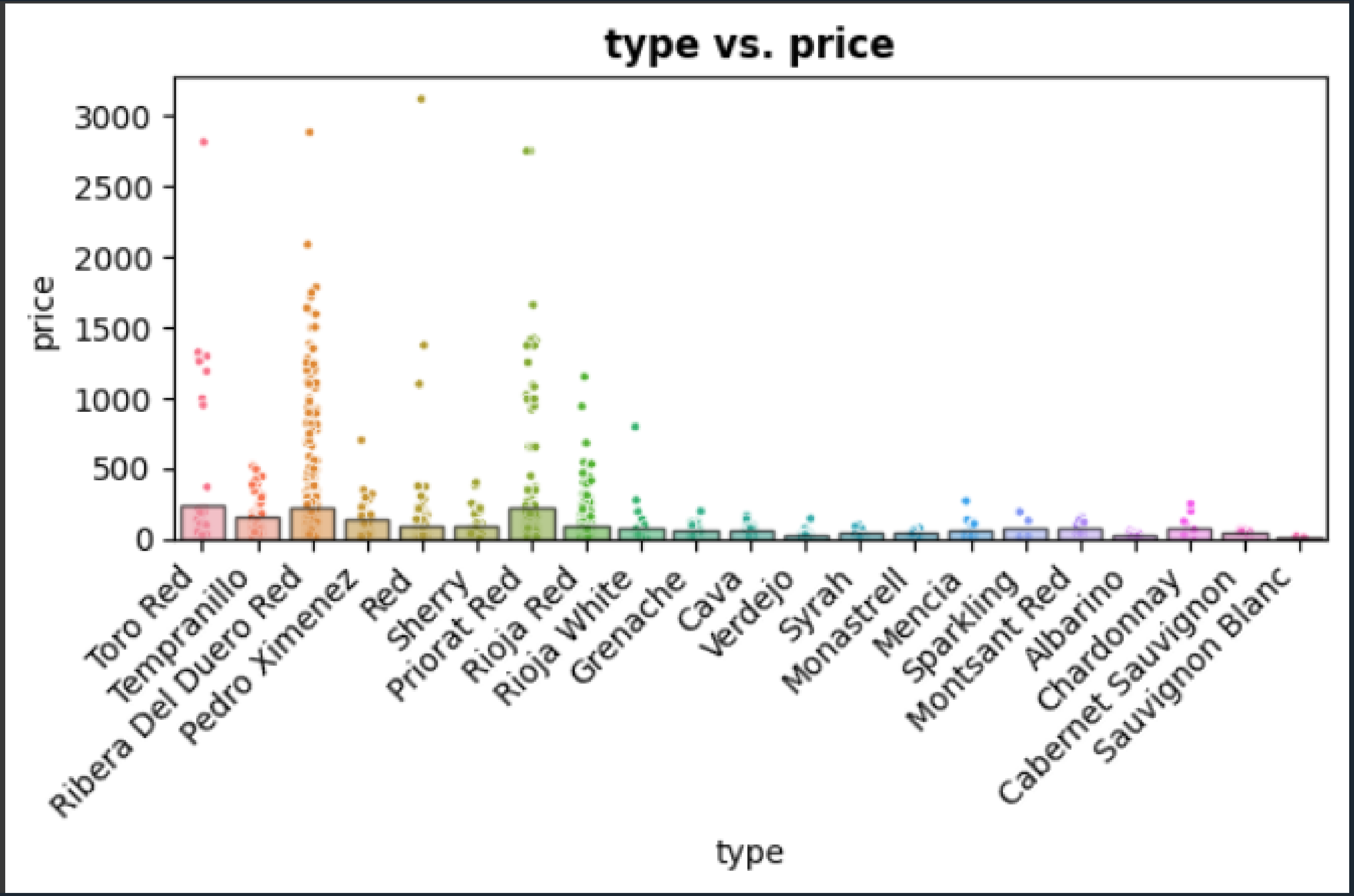
1st Visual

- This scatterplot shows the relationship between acidity levels and price for the wines in the dataset. Each point represents a single wine sample. There is a clear positive correlation between acidity and price.



2nd Visual

- This bar chart displays the mean price for each wine type in the dataset. The height of each bar represents the average price of wines of that type. There are noticeable differences in average price between various wine types.



strengths and limitation

- Strengths:
- The model incorporates key attributes like variety, ratings, and alcohol content that influence price.
- With more data and tuning, the model has potential to provide highly accurate price forecasts.
- Automated ML predictions can help determine optimal pricing for new wines entering the portfolio.
- The model helps standardize pricing across a large inventory of diverse wines.

- Limitations:
- Currently limited data size leads to overfitting and reduces real-world applicability.
- Model lacks robustness with only ~2000 training examples and a few features. More data needed.

Final Recommendation

1

Gather more data to improve model training and avoid overfitting.

2

Try more advanced models - Beyond basic linear and tree models, explore advanced models like XGBoost, neural networks, or ensembles which may produce better predictions.