Winery Project

By Ainemukama Ernest Jerry

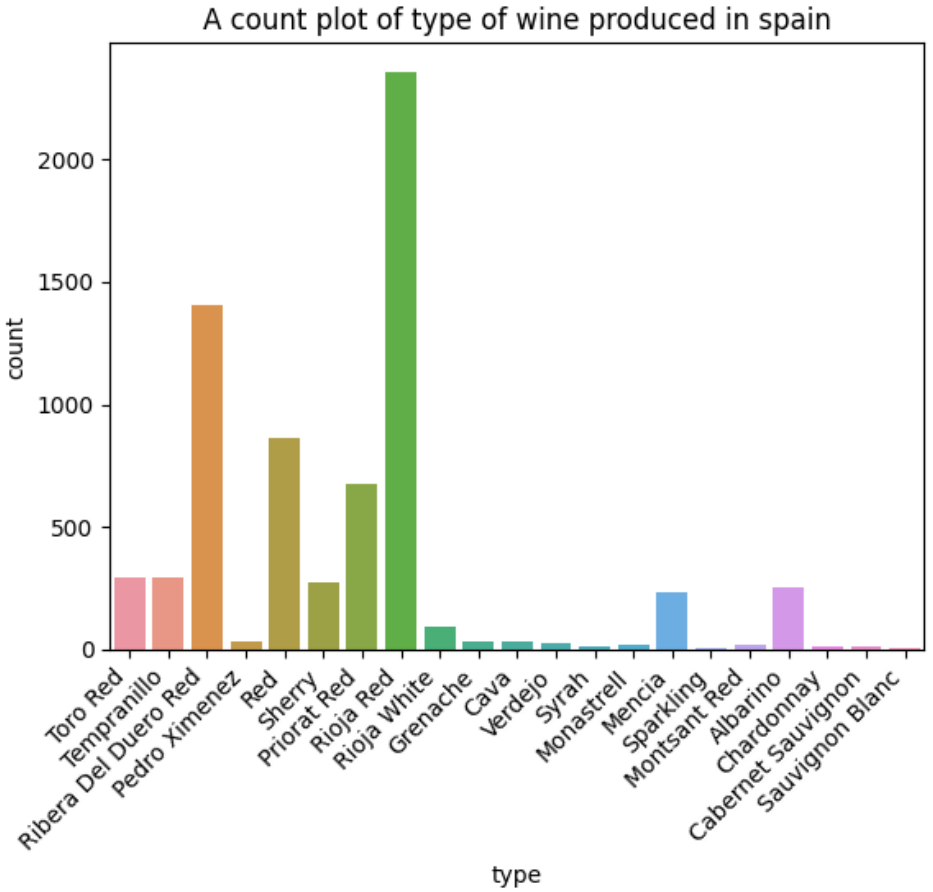
Who is your stakeholder, and what problem are you solving for them

- The wineries in Spain are our stakeholders and I was predicting the price of wine basing on factors like acidity, rating of wine, type of wine.
- Additionally learning how some of the above factors affect price of wine produced
    - Data dictionary

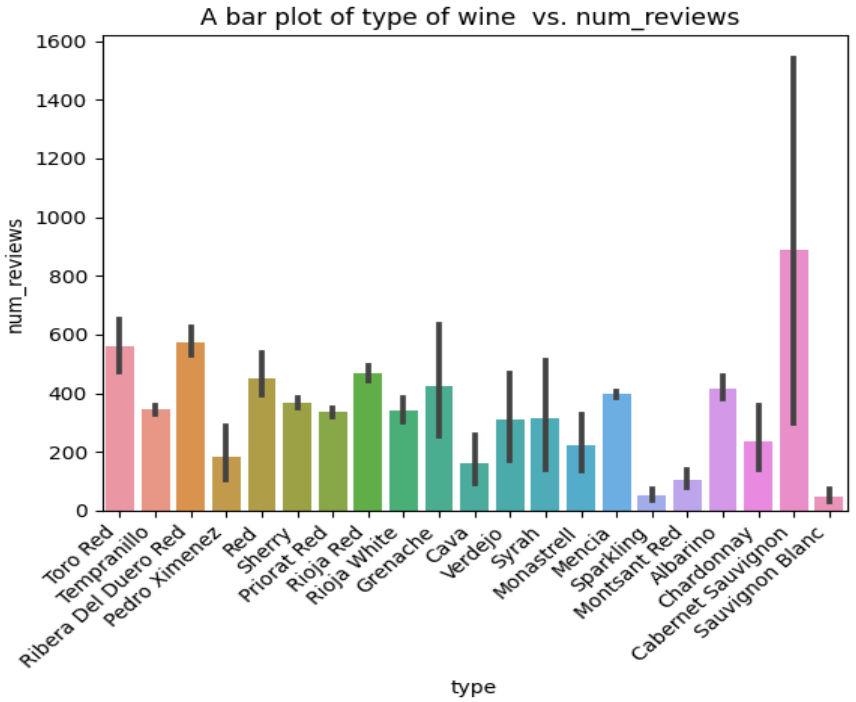
      see table here ➡

Column name	Description
Winery	Name of the winery
wine	The wine being produced
Rating	Rating of wine on scale of 1-5
Num_reviews	Number of reviews given to a wine by people
country	Country in which the winery is located
region	A place within the country
price	The price of wine produced by a winery
type	The type of wine produced
body	The nature of the bottle of wine
acidity	The level of acidity on a scale of 1-3

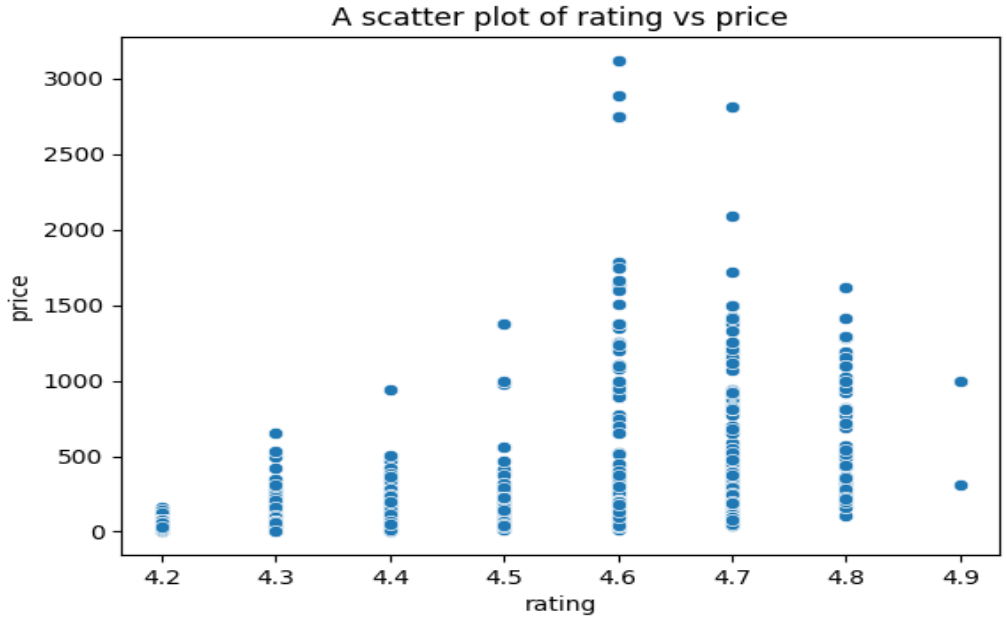
Visuals That Demonstrate Key Findings Of Interest



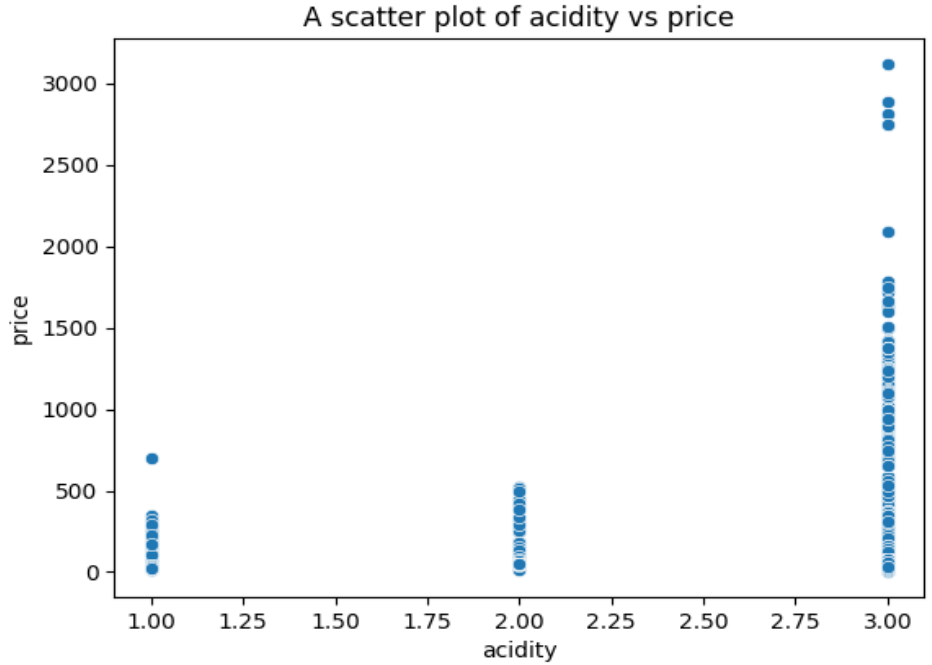
- From the visualization on the left we can see that rioja wine was the most produced type of wine by wineries in Spain
- But on our right we see that although rioja red type of wine was most produced in Spain, cabernet sauvignon wine received the most reviews



Visuals That Demonstrate Key Findings Of Interest



- From the scatter plot on our left we see that wine that received rating between 4.6 and 4.8 attracted a very good price
- From the scatter plot on my right we see that wine with an acidity level of 3 attracted a very price



Now lets talk about our model

After implementation of three models, I went for a random forest model as our data required a regression type of model for out data after introducing a new column called ‘nature’ which described wine as either good wine or bad wine

Strengths

- The model scored highly on the test data with an r-squared score of 0.748
- The mean squared error reduced from 6,149.574 to 6,138.832 on the test data after applying new parameters to the model
- The root mean squared error reduced from 78.419 to 78.351 after applying new parameter to the model on the test data

weakness

- The mean absolute error increased from 16.070 to 18.327 after applying new parameters to the model on the test data

Final recommendations

- I recommend producing wine of a favorable acidity that can attract premium prices for wine produced
- I recommend better awareness methods for the wines as we saw in slide two the type of wine that was most produced had the lowest number of reviews and yet the type of wine that was among the least produced had the highest number of reviews
- A winery should seek to produce unique wine which is in a small amount of supply to attract premium prices other than produce lots of bottles of poor quality that attract bad review and low prices