

WINE QUALITY

An endeavour to predict perfection...

THE DATA

- Collected at data.world [here](#)
- Details can be found in google datasets [here](#)
- No wrangling of the data was needed
- Contains a list of features and a quality rating
- Focusing on white wine
- Downsides: no real-world references, red wine is limited
- Only quality rating from 3-9 exist
- Bell shaped distribution on rating

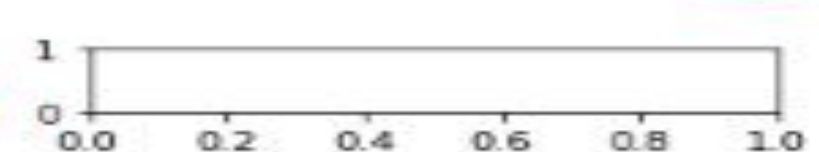
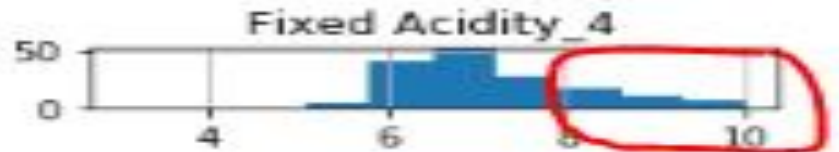
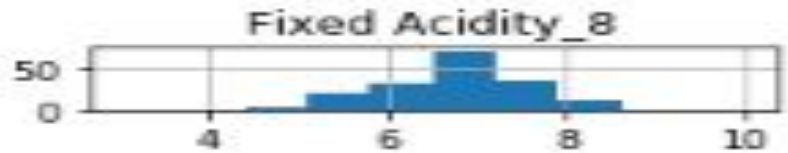
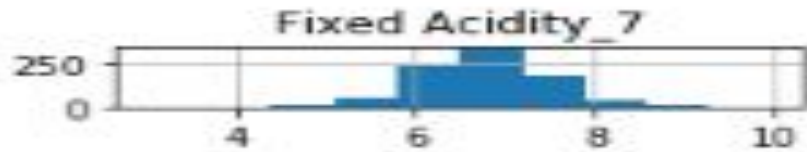
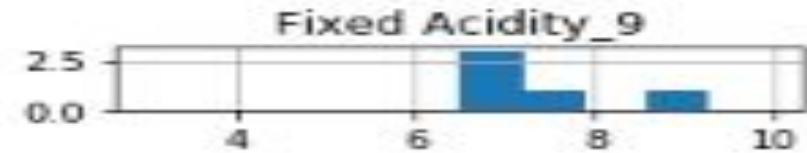
THE PROJECT

- Achieve a chemical understanding of good/bad wine
- Come up with a predictive model to determine wine quality
- Guide purchases from new vendors
- Possibly reduce purchase expense
- Current steps taken
 - Histogram analysis
 - Mean, median, value range analysis

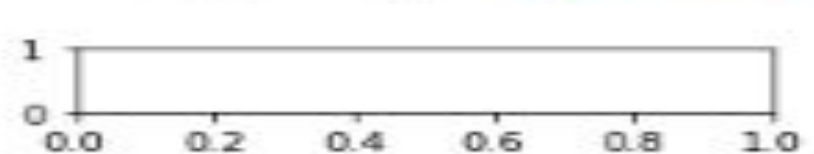
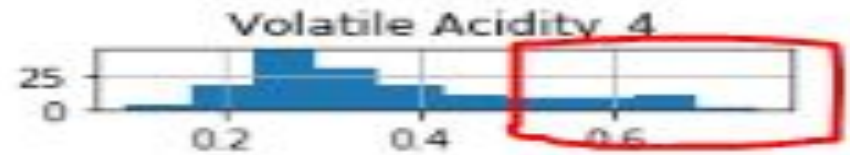
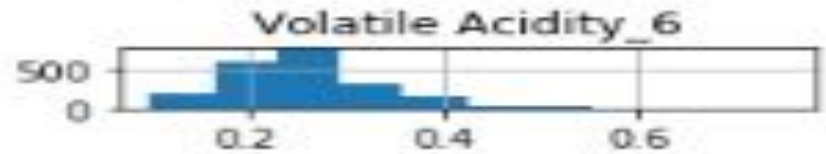
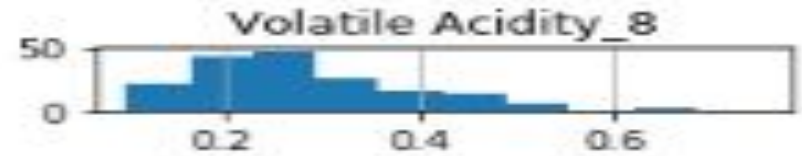
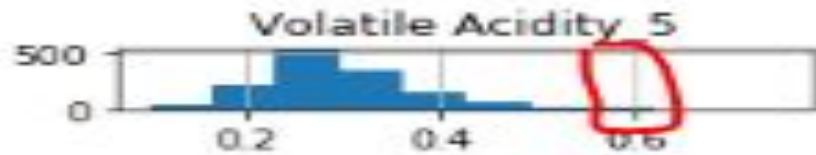
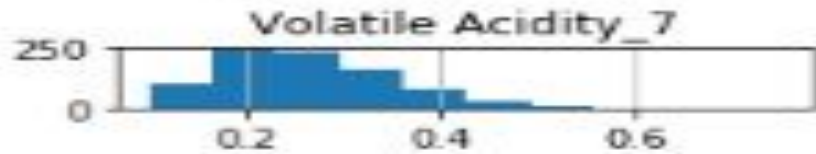
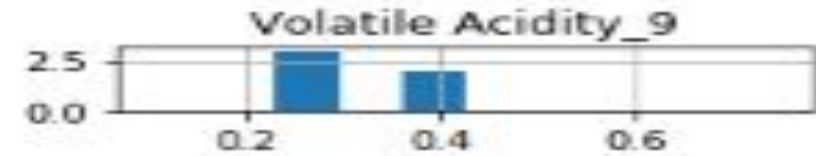
HISTOGRAM ANALYSIS

- Get a sense of the data; spot any obvious quality bias
- Findings
 - A greater 'fixed acidity' correlates to a lower quality wine.
 - A 'volatile acidity' greater than 0.5 correlates with lower quality wine.
 - Citric acid' levels above 0.6 or below 0.2 corresponds to a lower quality wine.
 - There seems to be no correlation of 'residual sugar' to the quality of wine.
 - A low 'chloride' content correlates to a higher quality wine.
 - Lower 'free sulfur dioxide' content correlates with lower quality wine.
 - Where 'total sulfur dioxide' less than 75, we see a high chance of low quality wine.
 - We see that a lower 'density' correlates to a better quality wine.
 - The higher the pH the better the chance it will be a high quality wine.
 - As for the 'sulphates' there was no noticeable correlation.
 - A higher 'alcohol' content correlates to a higher quality rating and vice versa.

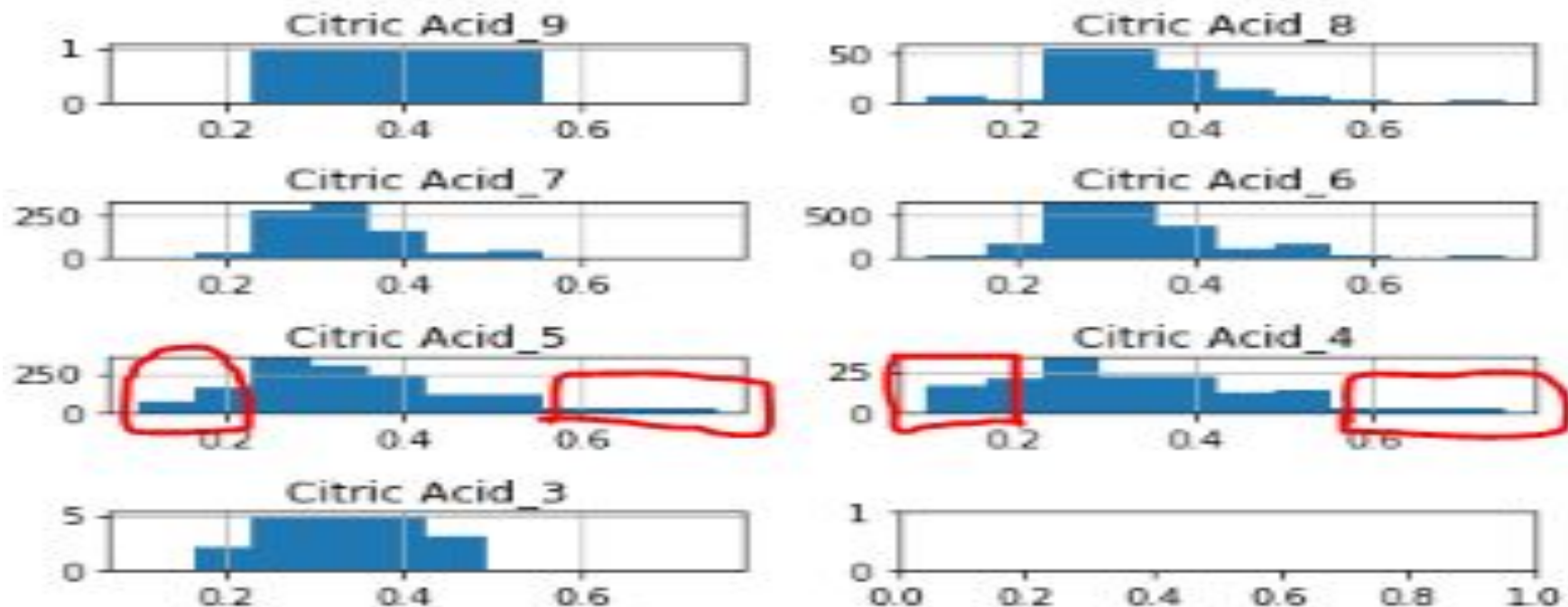
A GREATER '**FIXED ACIDITY**' CORRELATES TO A LOWER QUALITY WINE



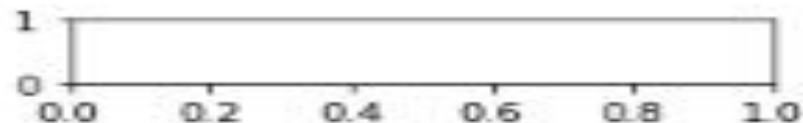
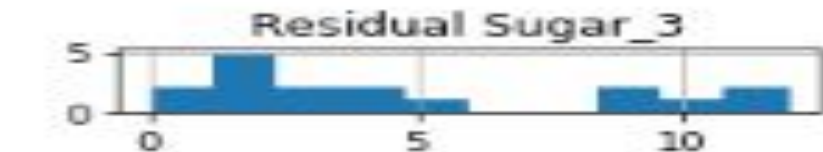
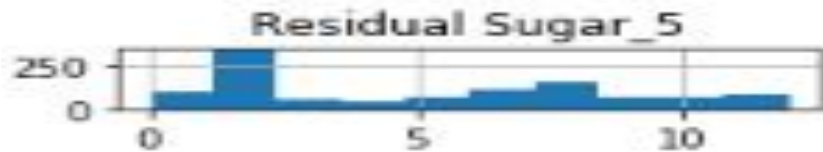
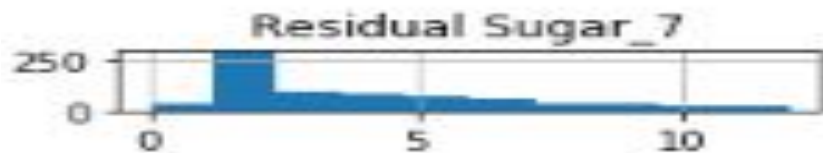
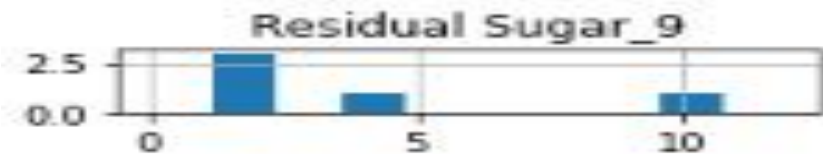
'VOLATILE ACIDITY' GREATER THAN 0.5 CORRELATES WITH LOWER QUALITY WINE.



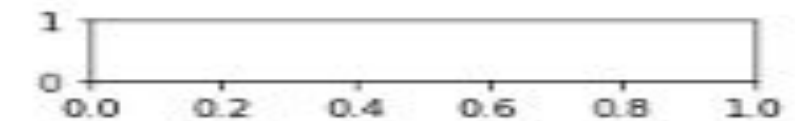
'CITRIC ACID' LEVELS ABOVE 0.6 OR BELOW 0.2 CORRESPONDS TO A LOWER QUALITY WINE



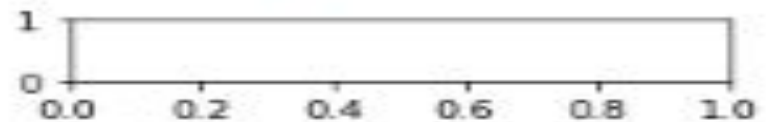
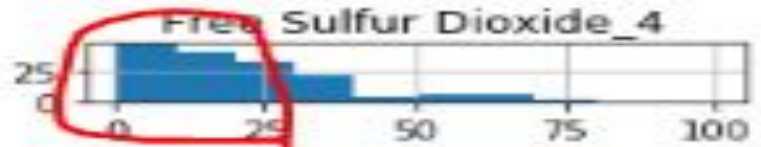
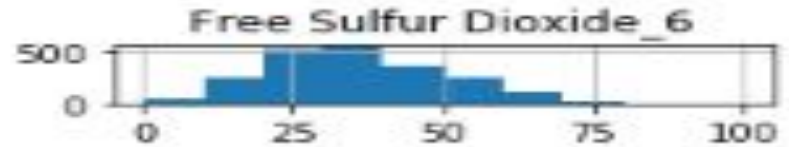
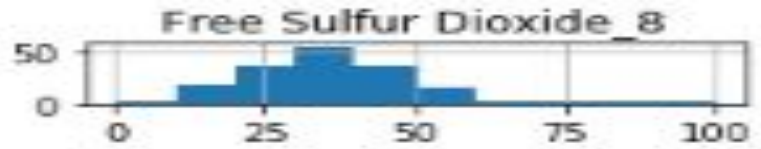
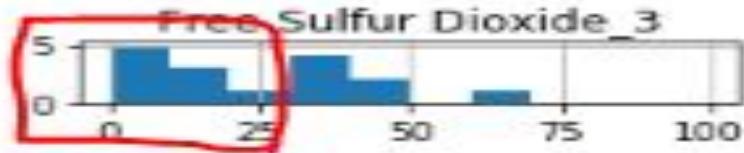
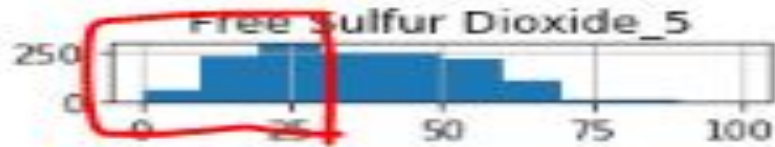
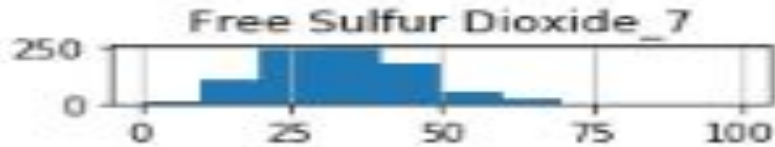
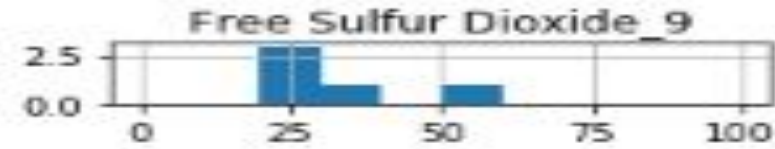
THERE SEEMS TO BE NO CORRELATION OF 'RESIDUAL SUGAR' TO THE QUALITY OF WINE.



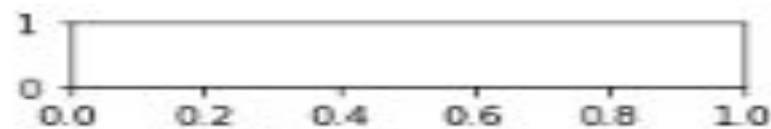
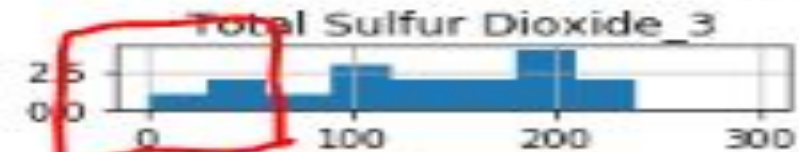
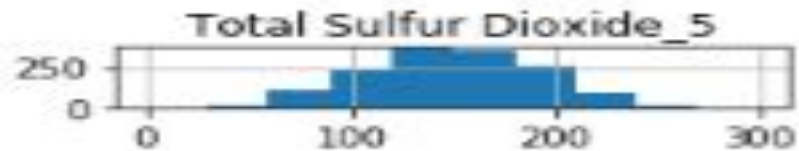
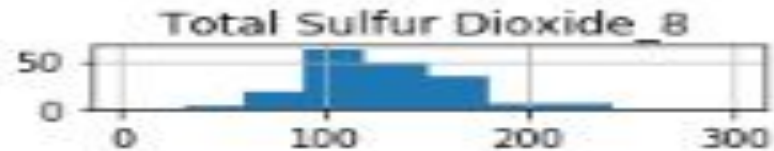
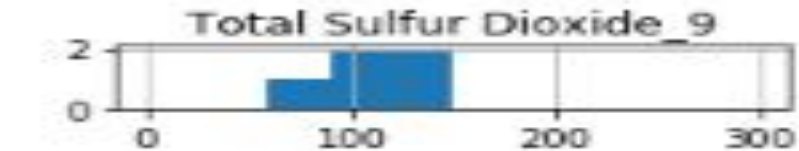
LOW 'CHLORIDE' CONTENT CORRELATES TO A HIGHER QUALITY WINE



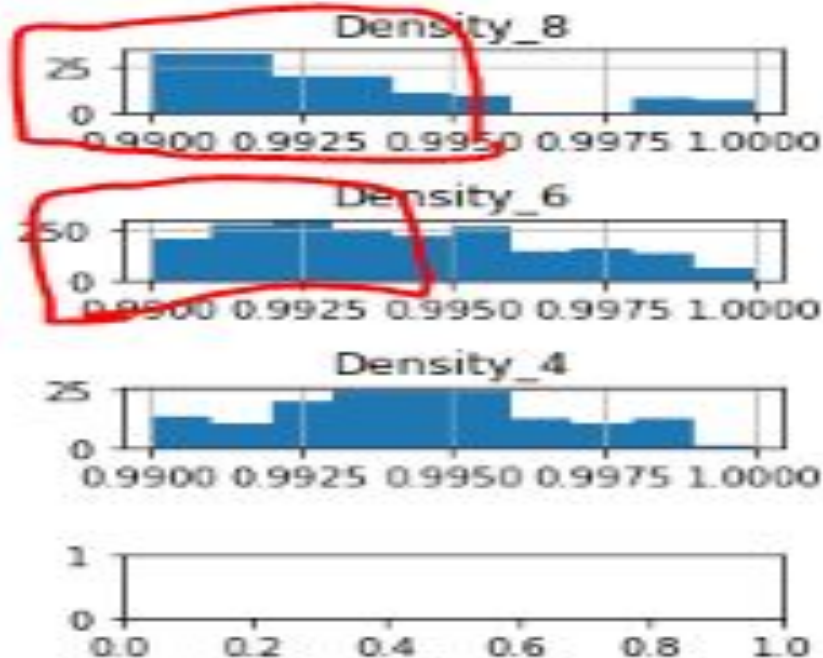
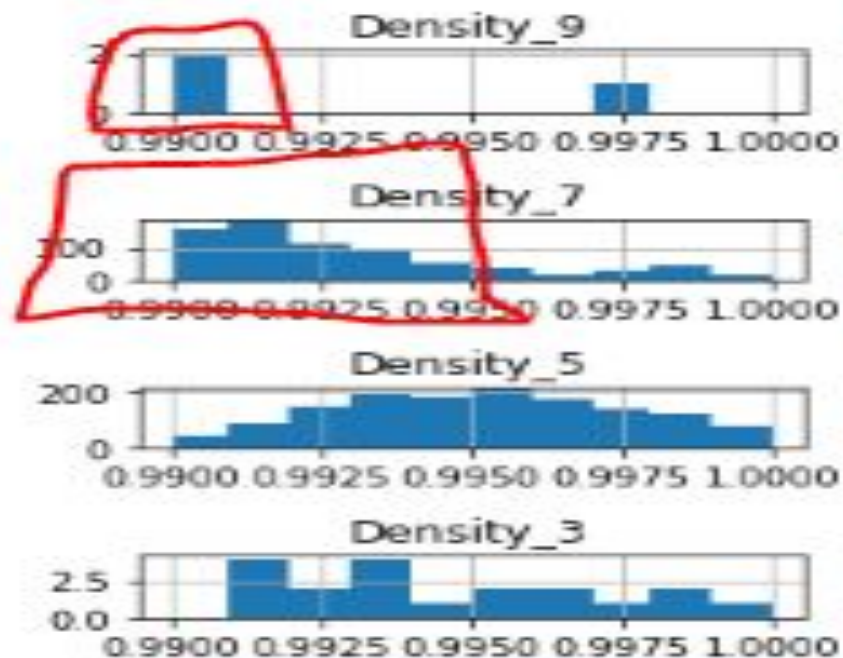
LOWER 'FREE SULFUR DIOXIDE' CONTENT CORRELATES WITH LOWER QUALITY WINE



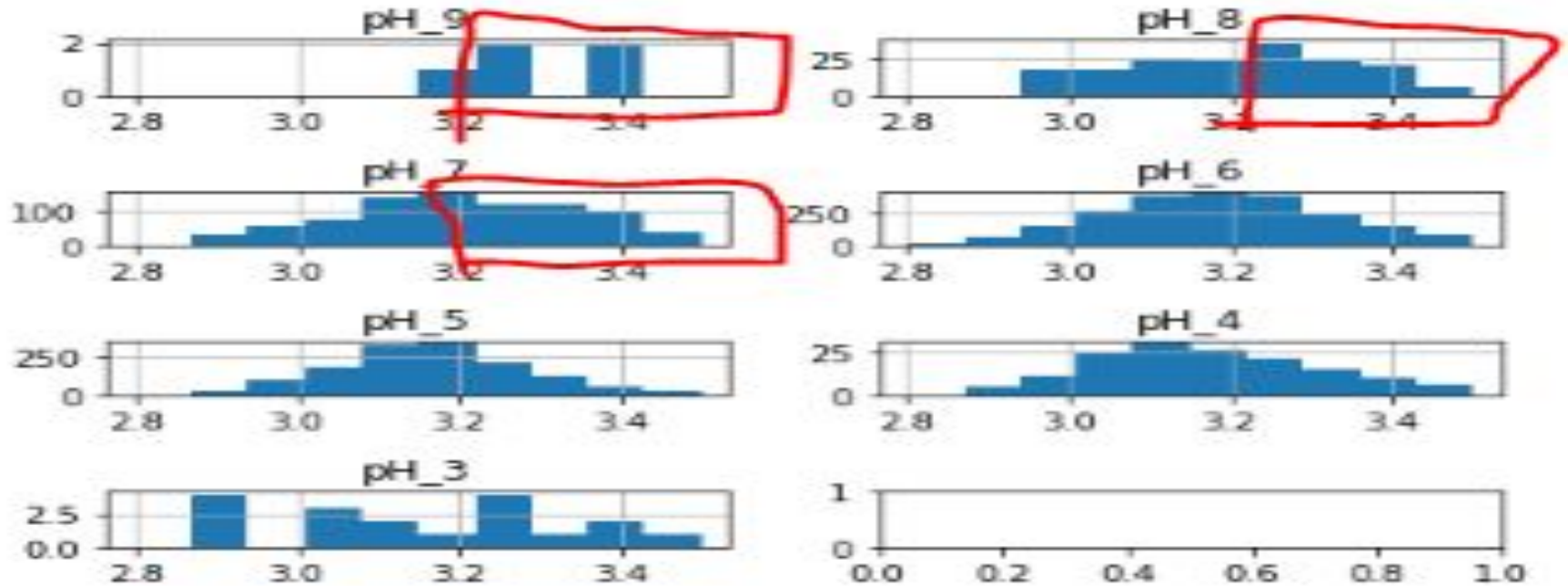
'TOTAL SULFUR DIOXIDE' LESS THAN 75, WE SEE A HIGH CHANCE OF LOW QUALITY WINE



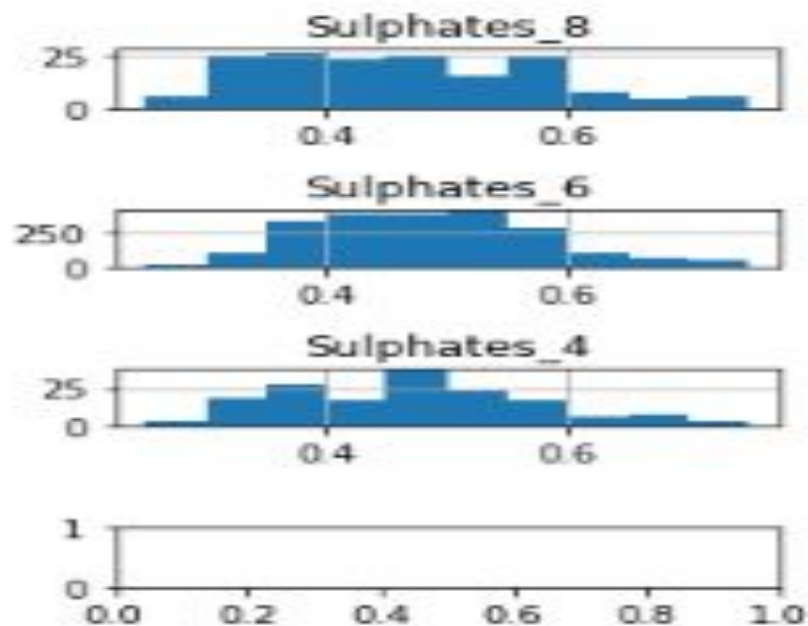
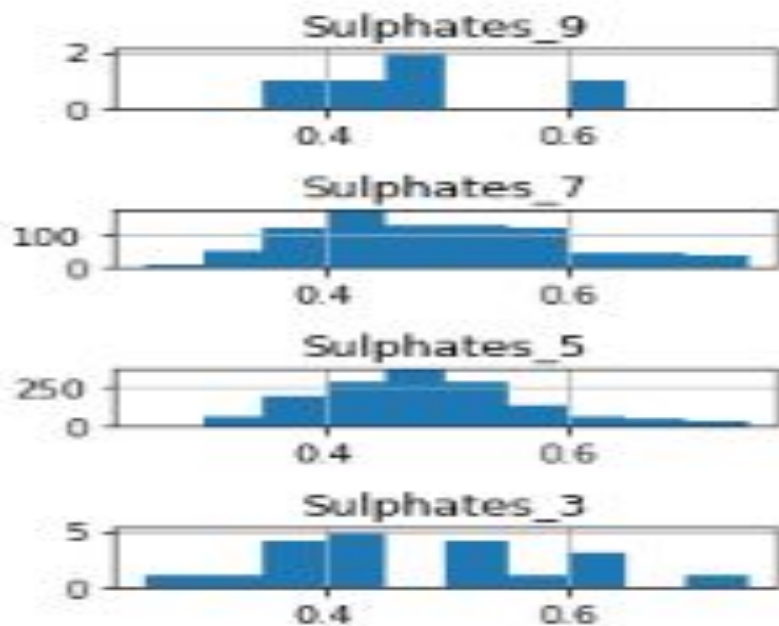
LOWER 'DENSITY' CORRELATES TO A BETTER QUALITY WINE



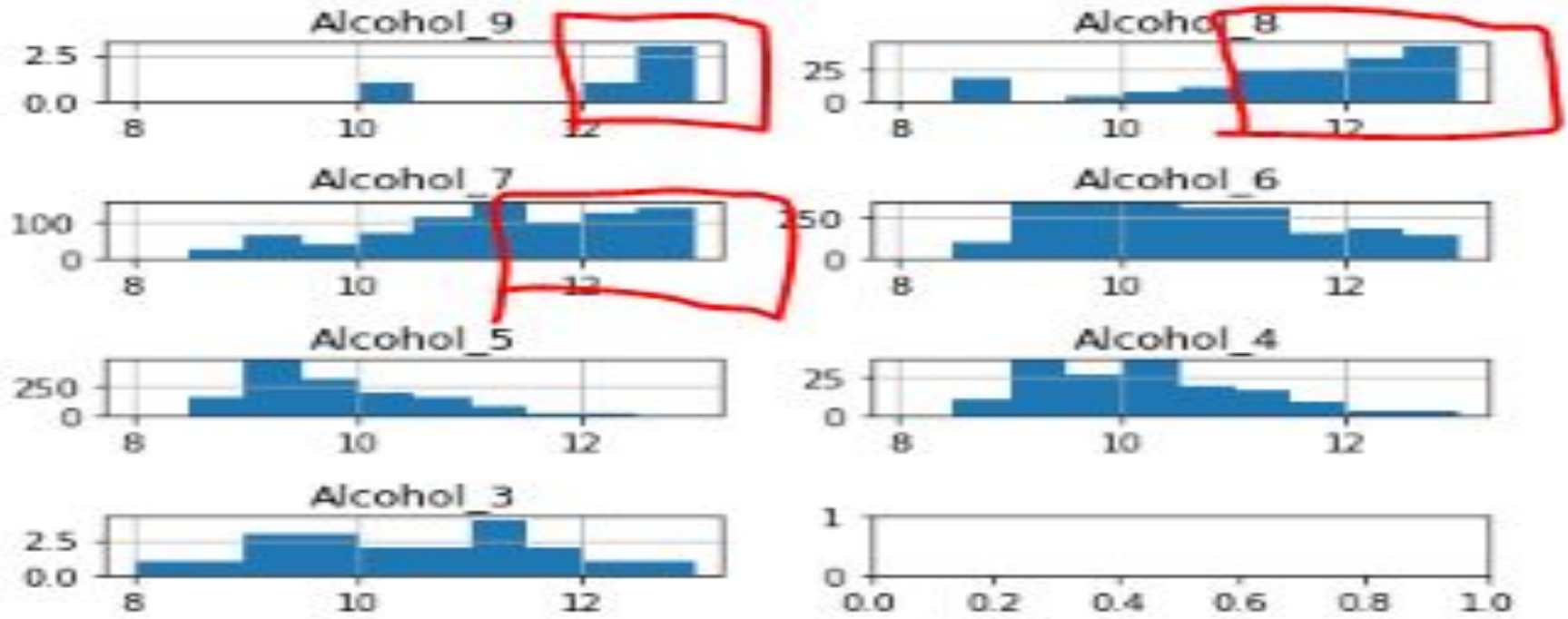
THE HIGHER THE **pH** THE BETTER THE CHANCE IT WILL BE A HIGH QUALITY WINE



AS FOR THE 'SULPHATES' THERE WAS NO NOTICEABLE CORRELATION



A HIGHER 'ALCOHOL' CONTENT CORRELATES TO A HIGHER QUALITY RATING AND VICE VERSA

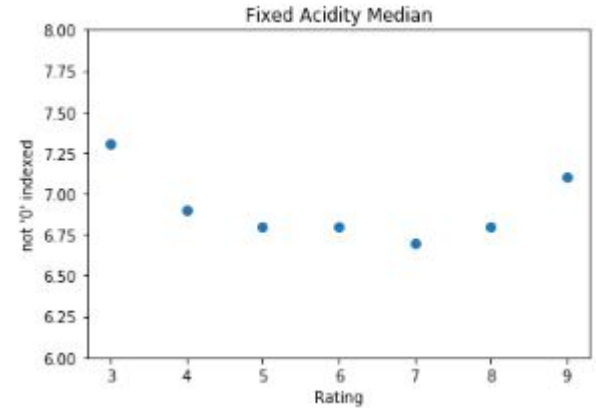
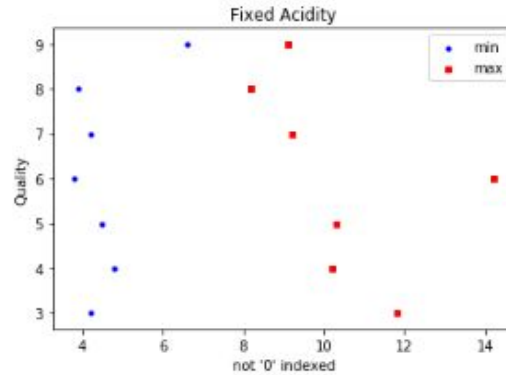
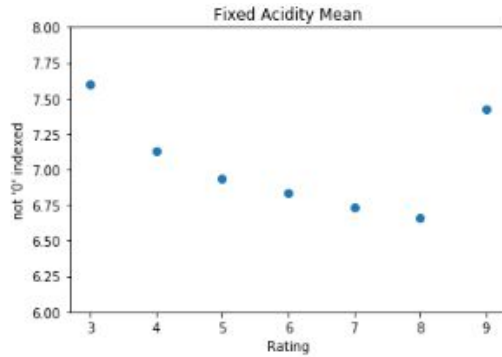


MOVING ON: MEAN, MEDIAN AND VALUE RANGE ANALYSIS

- Findings Summary

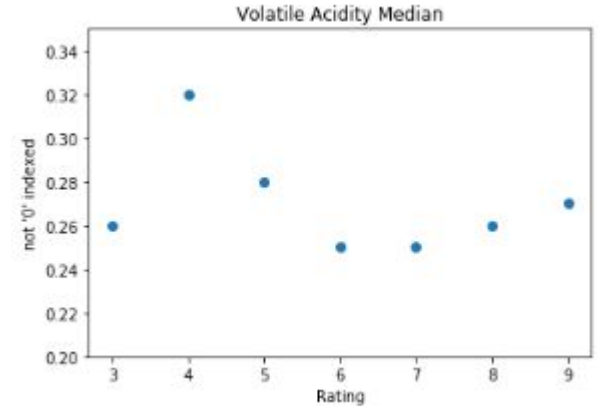
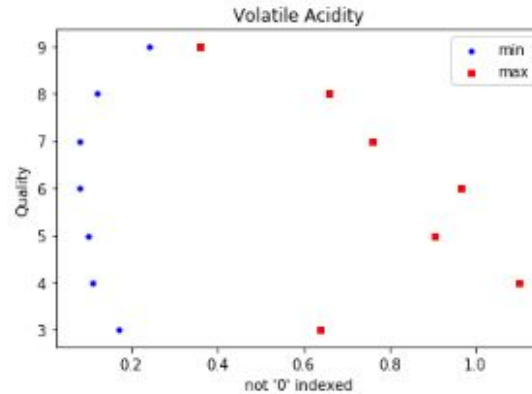
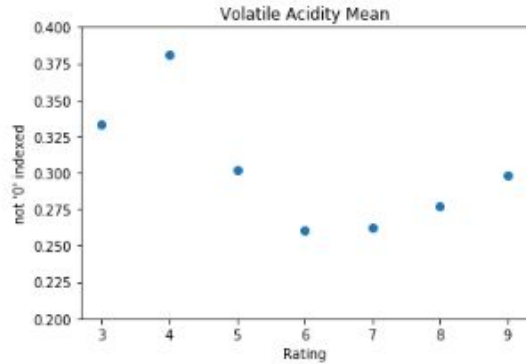
- A greater 'fixed acidity' correlates to a lower quality wine.
- A 'volatile acidity' greater than 0.5 correlates with lower quality wine.
- 'Citric acid' levels above 0.6 or below 0.2 corresponds to a lower quality wine.
- Lower 'free sulfur dioxide' content correlates with lower quality wine.
- Where 'total sulfur dioxide' less than 75, we see a high chance of low quality wine.
- A low 'chloride' content correlates to a higher quality wine.
- We see that a lower 'density' correlates to a better quality wine.
- The higher the 'pH' the better the chance it will be a high quality wine.
- A higher 'alcohol' content correlates to a higher quality rating and vice versa.
- There seems to be no correlation with 'residual sugar' on the quality of wine.
- As for the 'sulphates' there was no noticeable correlation.

FIXED ACIDITY



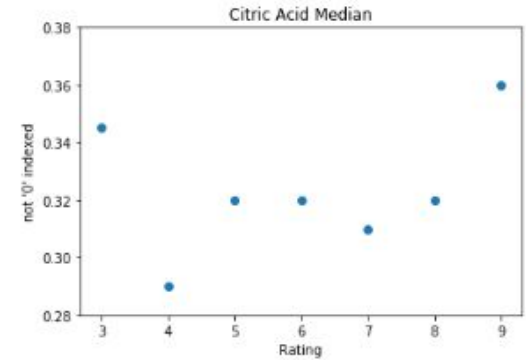
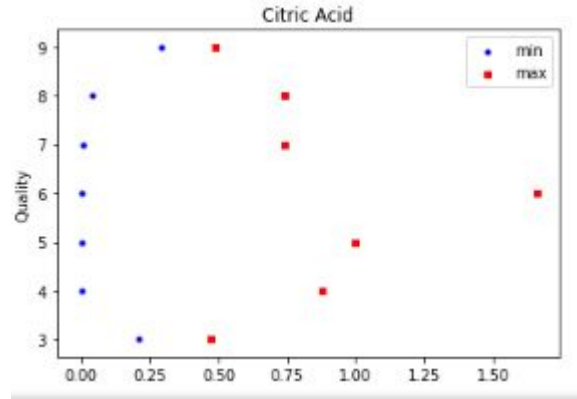
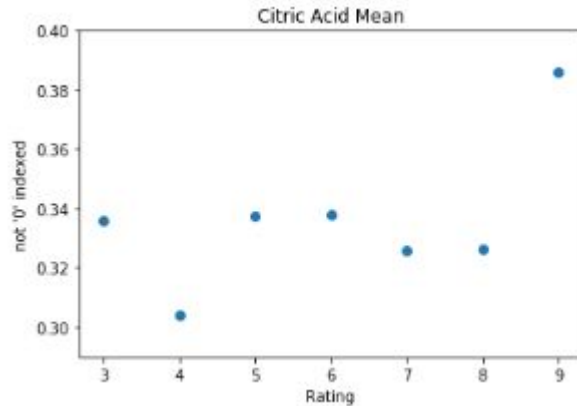
- Original Hypothesis: “A greater ‘fixed acidity’ correlates to a lower quality wine.”
- Mean and median analysis confirms the original hypothesis.
- Value range analysis presents a new hypothesis: ‘fixed acidity’ above 9.2 correlates to low quality wine.

VOLATILE ACIDITY



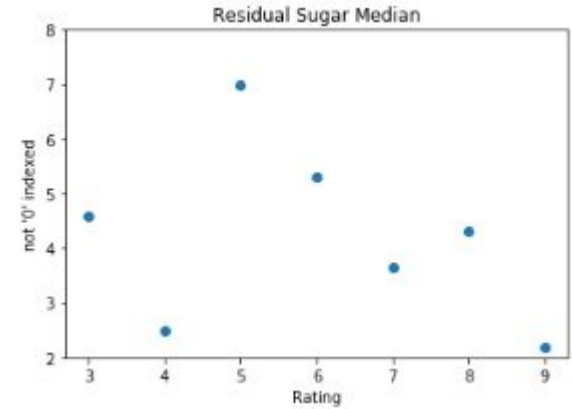
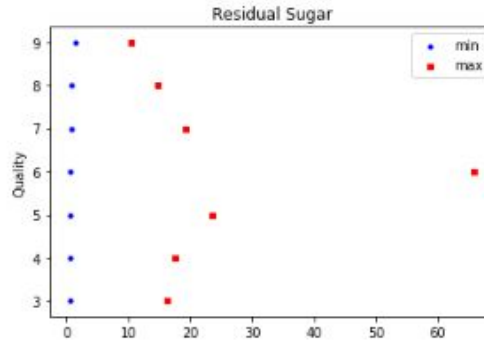
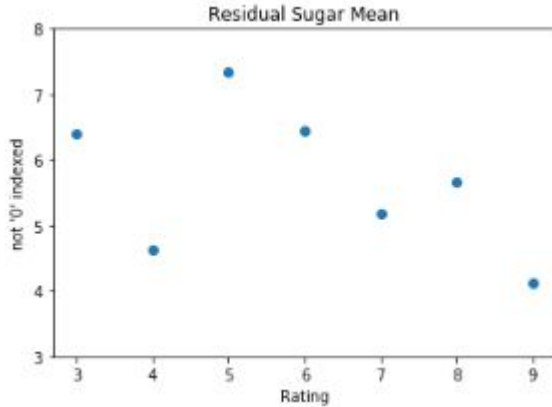
- Original Hypothesis: “A ‘volatile acidity’ greater than 0.5 correlates with lower quality wine.”
- Mean and median analysis shows higher values for lower quality wine.
- Value range analysis presents a new hypothesis: high chance that ‘volatile acidity’ greater than 0.8 corresponds to low-mid or low quality wine.
- Original hypothesis proved to be an overgeneralization.

CITRIC ACID



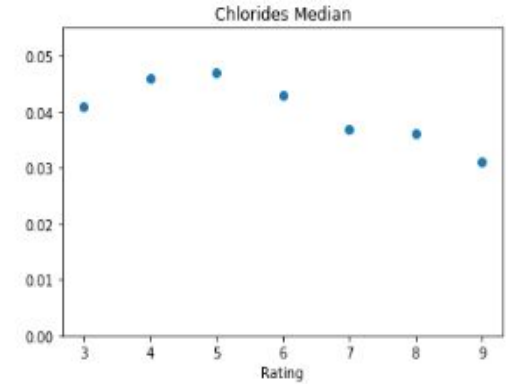
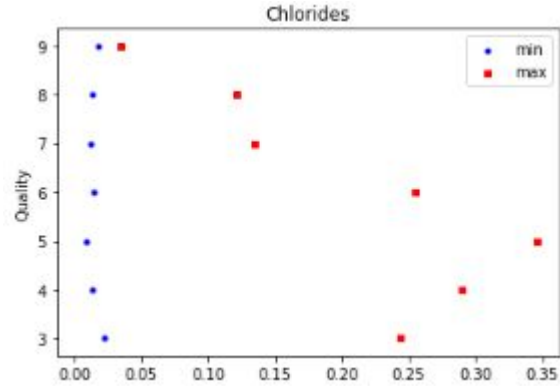
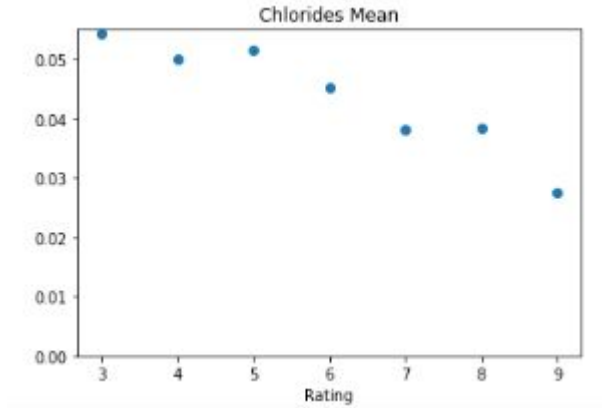
- Original Hypothesis: “‘Citric acid’ levels above 0.6 or below 0.2 corresponds to a lower quality wine.”
- Mean and median analysis shows noticeably higher values for highest quality wine.
- Value range analysis clarifies the original hypothesis: ‘Citric acid’ above 0.75 or below 0.29 is a 100% chance of being mid to low quality wine.

RESIDUAL SUGAR



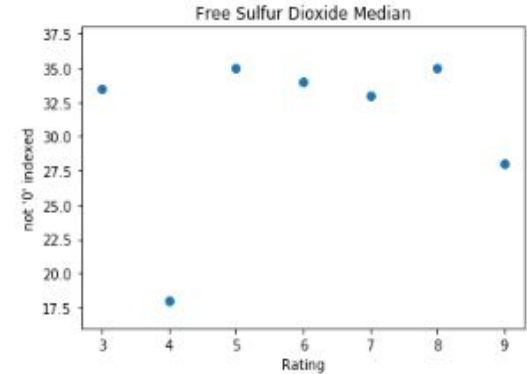
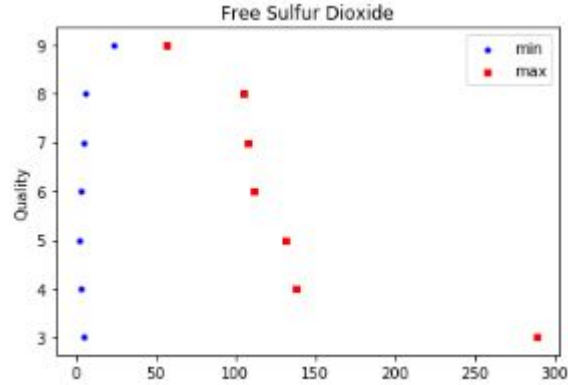
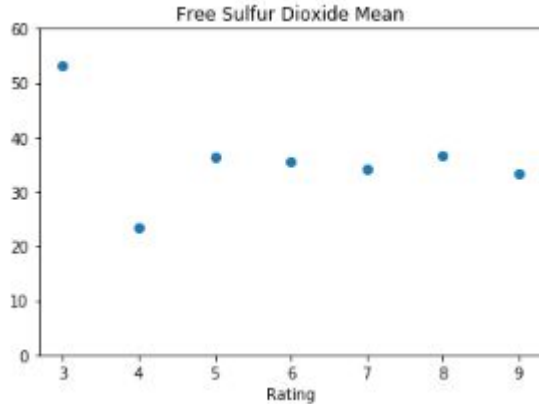
- Original Hypothesis: “There seems to be no correlation with ‘residual sugar’ on the quality of wine.”
- Mean and median analysis shows on average, residual sugar is lower for higher quality wine.
- Value range analysis presents a new hypothesis: ‘Residual sugar’ above 19 correlates to mid or low quality wine.

CHLORIDE



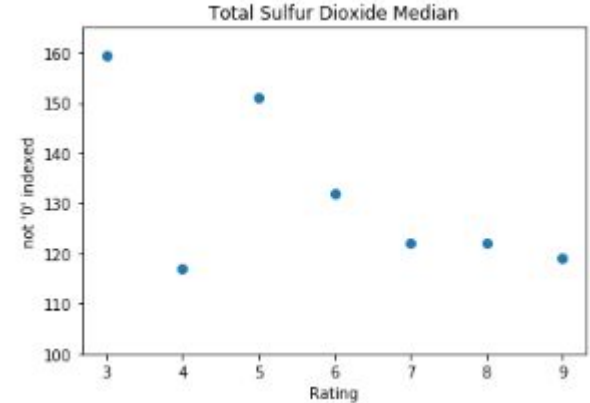
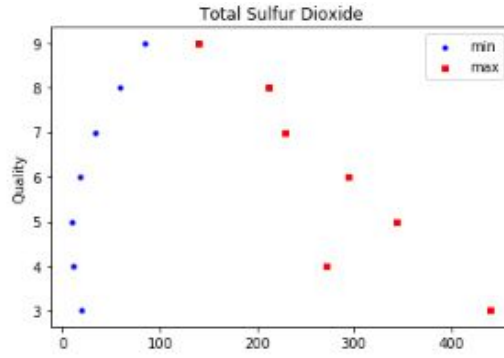
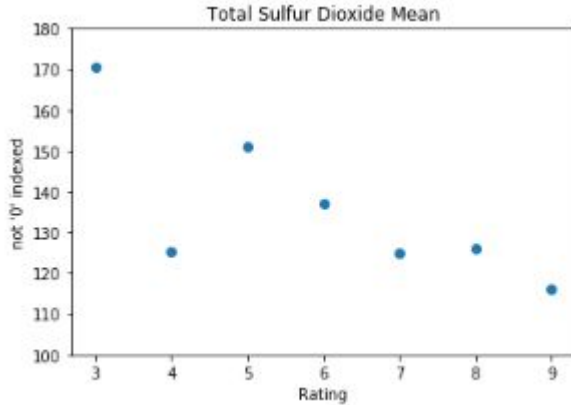
- Original Hypothesis: “A low ‘chloride’ content correlates to a higher quality wine.”
- Mean and median analysis confirms the original hypothesis.
- Value range analysis presents a new hypothesis: 100% chance if ‘chloride’ levels are above 0.135 it corresponds to a mid or low quality wine.

FREE SULFUR DIOXIDE



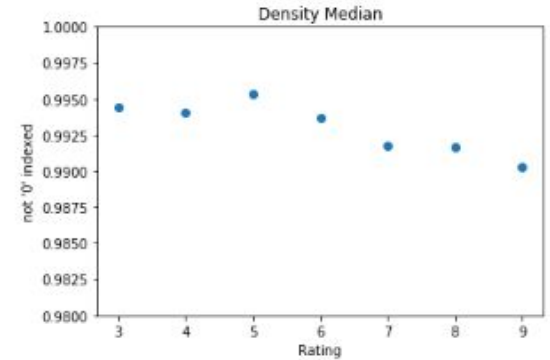
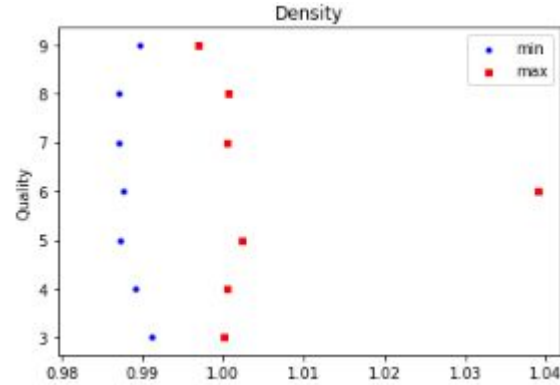
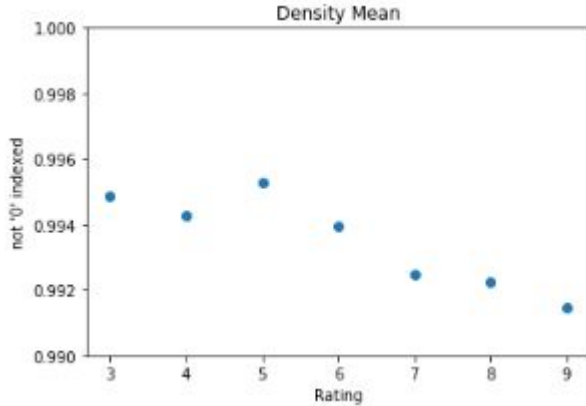
- Original Hypothesis: “Lower 'free sulfur dioxide' content correlates with lower quality wine.”
- Mean and median analysis: 'free sulfur dioxide' is constant on average, except for low or high quality wine
- Value range analysis presents a new hypothesis: A ‘free sulfur dioxide’ content greater than 108 corresponds 100% to a mid or low quality wine. A ‘free sulfur dioxide’ content greater than 131 corresponds 100% to a low quality wine.

TOTAL SULFUR DIOXIDE



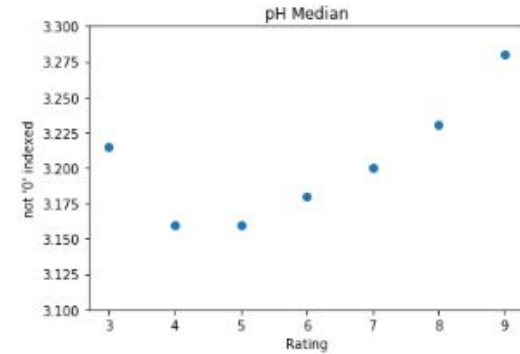
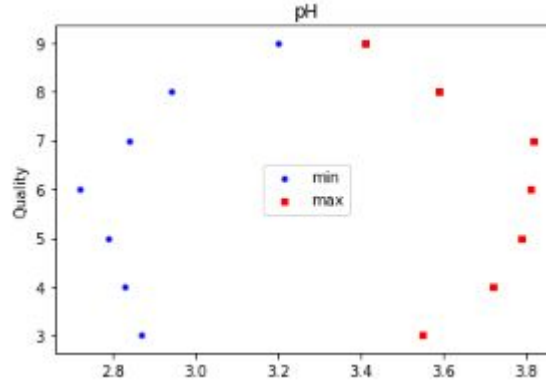
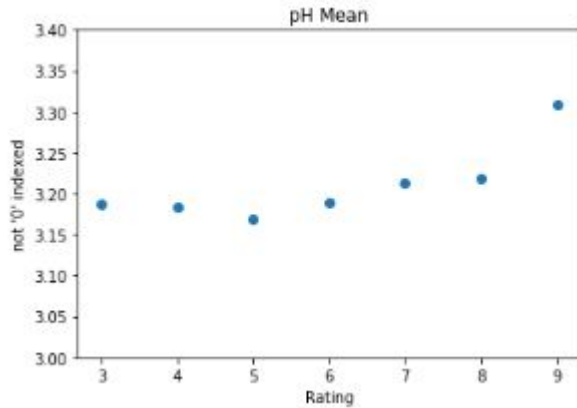
- Original Hypothesis: “Where 'total sulfur dioxide' less than 75, we see a high chance of low quality wine.”
- Mean and median analysis tends downwards from low to high quality wine.
- Value range analysis presents a new hypothesis: A ‘total sulfur dioxide’ content greater than 229 or less than 34 corresponds 100% of the time to a mid or low quality wine.
- Original hypothesis was an overgeneralization.

DENSITY



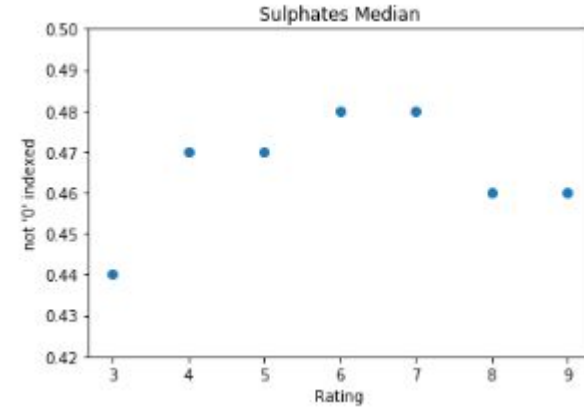
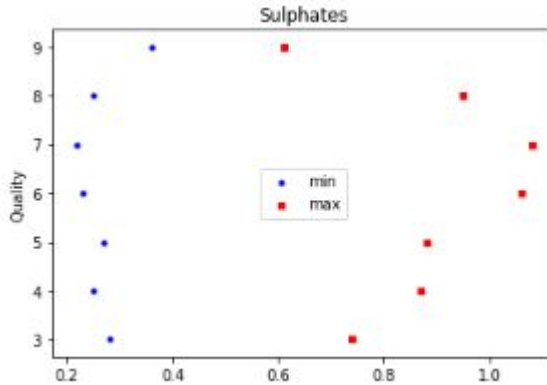
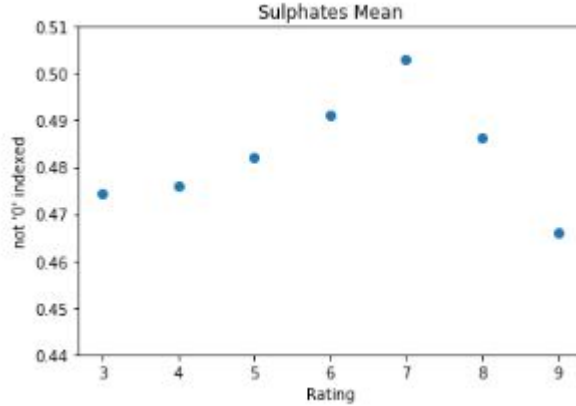
- Original Hypothesis: “We see that a lower ‘density’ correlates to a better quality wine.”
- Mean and median tends downwards from low to high quality wine. This confirms the original hypothesis.
- Value range analysis: A ‘density’ greater than 1.00040 corresponds most of the time to a mid or low quality wine.

pH



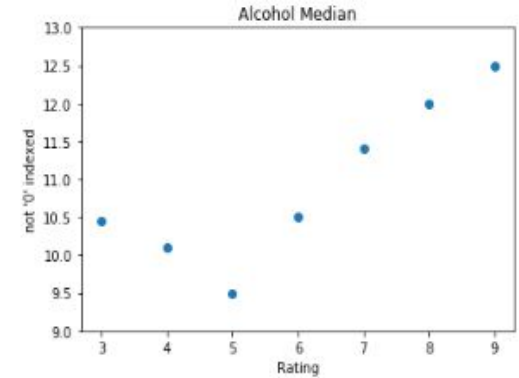
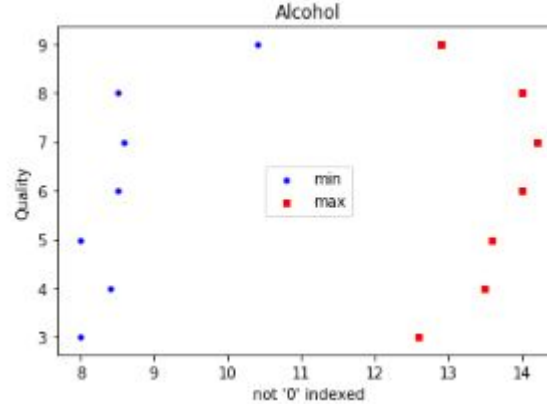
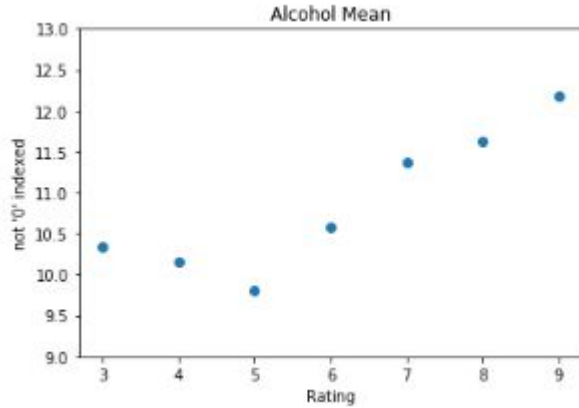
- Original Hypothesis: “The higher the ‘pH’ the better the chance it will be a high quality wine.”
- Mean and median ‘pH’ trends up from low to high quality wine. This confirms the original hypothesis.
- Value range analysis: The lower bound of ‘pH’ increases with the quality of wine. ‘pH’ levels below 2.72 correspond to lower quality wine.
- Need to check for outliers.

SULPHATES



- Original Hypothesis: “As for the ‘sulphates’ there is no noticeable correlation.”
- Mean analysis: Mid to high-mid level wine has a high mean value of sulphates.
- Median analysis: Mid level wine has a high median value of sulphates.
- Value range analysis: Sulphates level of mid level wine ranges from less than 0.28 and greater than 3.55, most of the time.

ALCOHOL



- Original Hypothesis: “A higher ‘alcohol’ content correlates to a higher quality rating and vice versa.”
- Mean analysis and median analysis confirms the original hypothesis.
- Value range analysis confirms the original hypothesis: If the ‘alcohol’ content is less than 8.6 it is a 100% that the wine is of mid or low quality.

MMMR ANALYSIS RESULTS

- If the 'alcohol' content is less than 8.6 it is a 100% that the wine is of mid or low quality.
- A 'density' greater than 1.00040 corresponds most of the time to a mid or low quality wine.
- A 'total sulfur dioxide' content greater than 229 or less than 34 corresponds 100% of the time to a mid or low quality wine.
- A 'free sulfur dioxide' content greater than 108 corresponds 100% to a mid or low quality wine. A 'free sulfur dioxide' content greater than 131 corresponds 100% to a low quality wine.
- 100% chance if 'chloride' levels are above 0.135 it corresponds to a mid or low quality wine.
- 'Residual sugar' above 19 correlates to mid or low quality wine.
- 'Citric acid' above 0.75 or below 0.29 is a 100% chance of being mid to low quality wine.

EXPLORATORY TOPICS

- high chance that 'volatile acidity' greater than 0.8 corresponds to low-mid or low quality wine.
- 'fixed acidity' above 9.2 correlates to low quality wine.
- Sulphates level of mid level wine ranges from less than 0.28 and greater than 3.55, most of the time.
- 'pH' levels below 2.72 correspond to lower quality wine. Need to check for outliers.

CONFIDENCE INTERVALS

- Previous analysis showed absolute conditions
- Following shows 90% confidence intervals for all features and a brief summary of findings
- Bucketed by high, mid, low quality
- High Quality: ratings 9, 8, 7
- Mid Quality: ratings 6, 5
- Low Quality: ratings 3, 4

ALCOHOL MEAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 10.77 - 12.07, mean = 11.42
- Mid Quality: 9.73 - 10.88, mean = 10.27
- Low Quality: 9.67 - 10.74, mean = 10.17
- Alcohol < 8.6 corresponds to low or mid quality wine
- Strong positive correlation of alcohol and quality

DENSITY MEDIAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 0.99059 - 0.9935, mean = 0.99185
- Mid Quality: 0.9925 - 0.99634, mean = 0.99438
- Low Quality: 0.9927 - 0.99583, mean = 0.99421
- Density > 1.0006 corresponds to mid quality wine
- Density < 0.9872 corresponds to high quality wine
- Strong indicator for high quality wine

TOTAL SULFUR DIOXIDE MEDIAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 104.5 - 144.5, mean = 122.59
- Mid Quality: 115.5 - 172.0, mean = 140.83
- Low Quality: 93.5 - 175.0, mean = 124.62
- Total sulfur dioxide < 34 and > 229 corresponds to a mid or low quality wine
- Total sulfur dioxide > 344 corresponds to a low quality wine
- Overall strong indicator potential, will need to test this further

FREE SULFUR DIOXIDE MEDIAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 27.0 - 42.0, mean = 33.42
- Mid Quality: 25.5 - 47.5, mean = 34.63
- Low Quality: 11.5 - 34.0, mean = 19.56
- Free sulfur dioxide > 108 corresponds to a mid or low level wine
- Free sulfur dioxide > 131 corresponds to a low quality wine
- Absolute findings are strong. Need further testing for surety

CHLORIDE MEAN 90% CONFIDENCE INTERVALS AND ABSOLUTE ANALYSIS

- High Quality: 0.0332 - 0.045, mean = 0.03817
- Mid Quality: 0.0397 - 0.0646, mean = 0.04771
- Low Quality: 0.0411 - 0.0733, mean = 0.05059
- Chloride content > 0.135 corresponds to mid or low quality wine
- Chloride content < 0.012 corresponds to mid quality wine
- Strong indicator for high quality wine

RESIDUAL SUGAR MEDIAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 2.15 - 8.18, mean = 4.04
- Mid Quality: 3.0 - 10.9, mean = 5.95
- Low Quality: 1.55 - 8.4, mean = 3.45
- Residual sugar < 0.8 corresponds to mid or low quality wine
- Residual sugar > 17.55 corresponds to a mid or high quality wine
- Not a strong indicator

CITRIC ACID MEAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 0.288 - 0.371, mean = 0.326
- Mid Quality: 0.281 - 0.414, mean = 0.338
- Low Quality: 0.237 - 0.399, mean = 0.308
- Citric acid level > 0.74 corresponds to a mid or low quality wine
- Citric acid < 0.29 corresponds to all quality ratings 7 and below
- Not a strong indicator

VOLATILE ACIDITY MEAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 0.222 - 0.32, mean = 0.265
- Mid Quality: 0.234 - 0.335, mean = 0.277
- Low Quality: 0.301 - 0.48, mean = 0.376
- Volatile acidity > 0.76 corresponds to mid or low quality wine
- Volatile acidity < 0.11 corresponds to mid or high quality wine
- Strong indicator for low quality wine

FIXED ACIDITY MEAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 6.33 – 7.13, mean = 6.724
- Mid Quality: 6.46 – 7.34, mean = 6.876
- Low Quality: 6.62 – 7.84, mean = 7.181
- Fixed acidity > 9.2 corresponds to low quality wine
- Potential strong indicator for low quality wine

SULPHATES MEAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 0.44 - 0.58, mean = 0.5
- Mid Quality: 0.44 - 0.55, mean = 0.49
- Low Quality: 0.42 - 0.54, mean = 0.48
- Sulphates < 0.25 or > 0.87 corresponds to mid or high quality wine
- Sulphates < 0.23 or > 1.06 corresponds to a high quality wine
- Absolute measures potentially a strong indicator

PH MEAN 90% CONFIDENCE INTERVALS AND SUMMARY

- High Quality: 3.14 - 3.3, mean = 3.22
- Mid Quality: 3.11 - 3.26, mean = 3.18
- Low Quality: 3.1 - 3.27, mean = 3.18
- Not a strong indicator

SUMMARY

- Strong indicators: volatile acidity, chloride, alcohol, density
- Needs more testing: sulphates, fixed acidity, free sulfur dioxide, total sulfur dioxide
- Weak indicators: pH, citric acid, residual sugar