DETERMINING QUALITY RATING OF RED WINE

Prepared for IBM Data Science Certificate

Kathryn Dark

2020

Business Problem and Data

Restaurants, vineyards and winemakers want to know if the quality rating of red wine can be determined from eleven different measures of said wine.

Features of Red Wine:

|	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

Data Analysis

Correlation coefficients for eleven variables of red wine compared to quality rating

```
:   alcohol                  0.476166
    sulphates                0.251397
    citric acid              0.226373
    fixed acidity            0.124052
    residual sugar           0.013732
    free sulfur dioxide     -0.050656
    pH                      -0.057731
    chlorides               -0.128907
    density                 -0.174919
    total sulfur dioxide    -0.185100
    volatile acidity        -0.390558
    Name: quality, dtype: float64
```
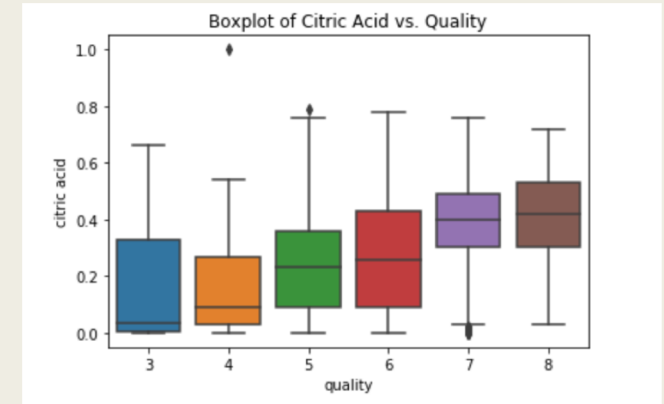
Alcohol, Volatile Acidity, Sulphates and Citric Acid have strongest correlation
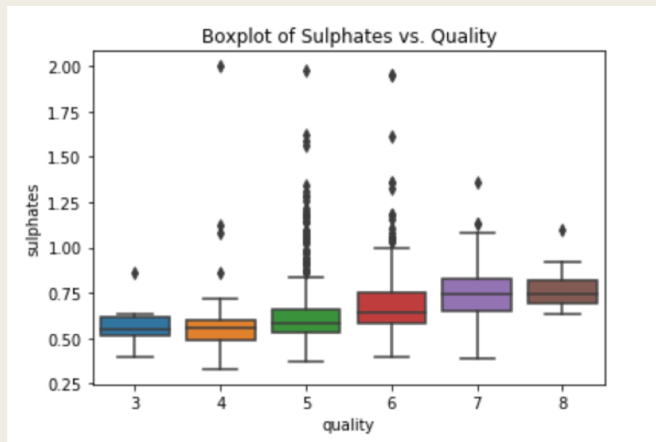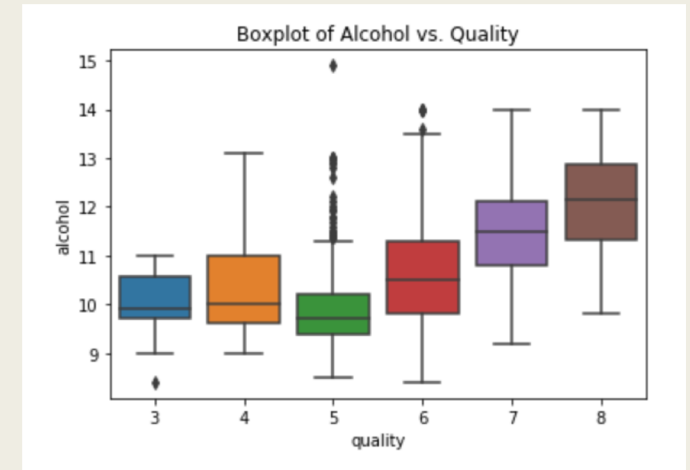
Data Visualization
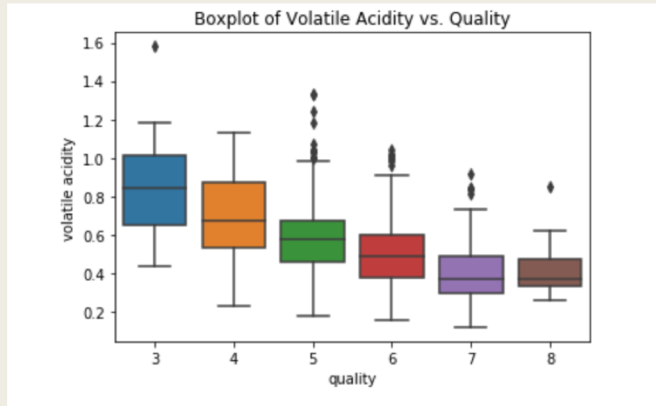
Negative Correlation
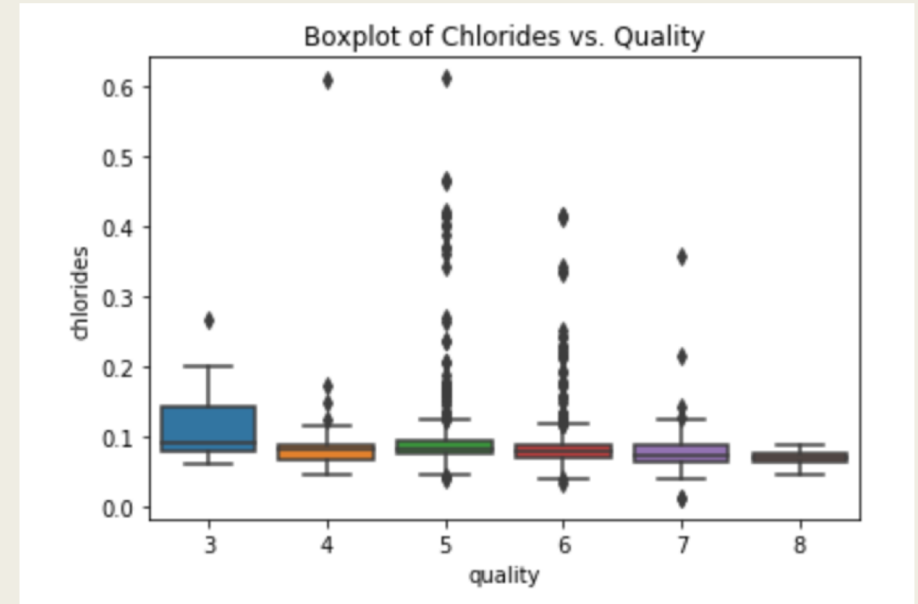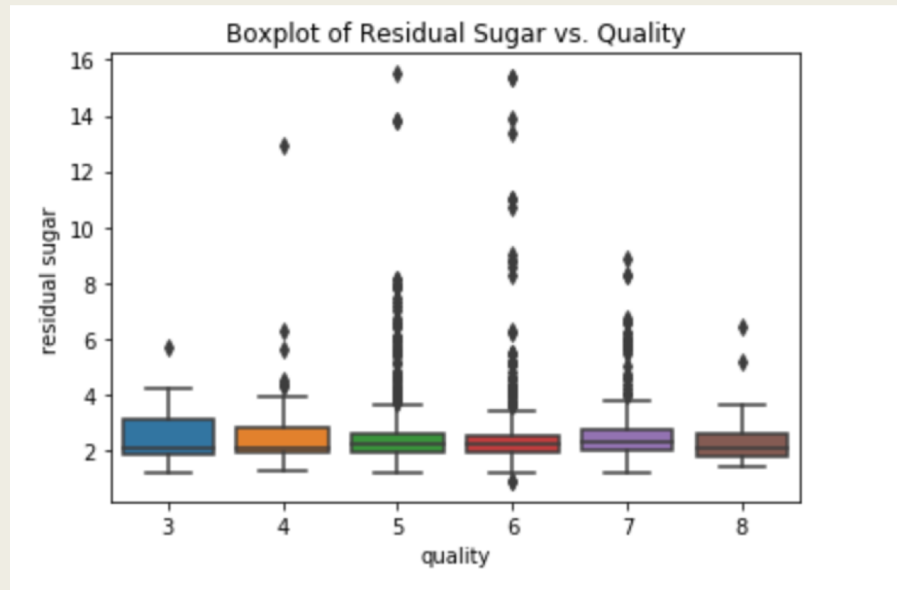

Positive Correlation

Positive Correlation, Many Outliers

Positive Correlation

Boxplots Show the Outliers

Machine Learning Algorithms

- **K-Nearest Neighbor (KNN)**
  - *Accuracy: 89.1% with k=1*
- **Logistic Regression**
  - *Accuracy: 87.5%*

Conclusion

- These four measured features of red wine have the strongest impact on its quality rating:

  - *Alcohol*

  - *Volatile Acidity*

  - *Sulphates*

  - *Citric Acid*

- The KNN Machine Learning Algorithm with k=1 should be used to determine the quality of red wine based on the eleven measured features.