MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

Mercado is a company that produces different varieties of whiskies, with Mercado Red and Black label being the best performing brand on the market. The Company is interested in producing another brand known as Mercado gold using the chemical composition of these two performing brands as a guide. As such the company want her team of analysis to analyze these two brand using data from their physicochemical properties and advise the production and brand manager on the potential new brand.

For the purpose of the task, the team will use two different dataset from the two performing brand i.e. the Mercado red label and Mercado black label. The dataset is grouped into two categories: red label wine and black label wine. With the following information; acidity, volatility level, citric level, sugar concentration, chloride, sulfur, sum of sulfur, density level, PH, sulfhate quantity, alcohol level, quality, group and quality label. The CSV was loaded into python and analyzed. I am going to analyze each dataset separately i.e. red label dataset and black label data set after the two dataset was combined and a more rigorous machine learning analysis was performed. It is important to state that the dataset used is a synthetic data generated using the University of California Irvine wine dataset as a guide.

Mercado Red Label Analysis.

Red wine dataframe output.

	acidity	volatility level	citric level	sugar concentration	chloride	sulfur	sum of sulfur	density level	PH	sulphate quantity	alcohol level	quality
100	8.39	0.1793	0.269	16.50	0.039	45.0	167.0	0.9375	3.40	0.69	6.6	7
101	8.09	0.0493	0.219	12.60	0.033	57.0	137.0	0.9362	3.80	0.66	6.9	7
102	6.99	0.1393	0.139	15.10	0.029	48.0	139.0	0.9370	3.74	0.64	6.9	7
103	8.49	0.2343	0.299	21.90	0.038	37.0	145.0	0.9400	3.39	0.71	6.5	7
104	8.39	0.1793	0.269	16.50	0.039	45.0	167.0	0.9375	3.40	0.69	6.6	7
105	8.29	0.0593	0.219	17.40	0.030	27.0	84.0	0.9374	3.60	0.60	6.7	7
106	8.09	0.0493	0.219	12.60	0.033	57.0	137.0	0.9362	3.80	0.66	6.9	6
107	8.09	0.1593	0.249	19.50	0.019	53.0	146.0	0.9390	3.56	0.84	6.6	6
108	8.09	0.1593	0.249	19.50	0.019	53.0	146.0	0.9390	3.56	0.84	6.6	6
109	7.89	0.2593	0.179	4.30	0.030	30.0	162.0	0.9327	3.67	0.85	7.8	6
110	7.49	0.0993	0.439	11.50	0.061	57.0	138.0	0.9359	3.29	0.64	6.3	6
111	8.19	0.1993	0.359	21.75	0.031	38.0	230.0	0.9400	3.44	0.77	6.4	6
112	8.19	0.2393	0.399	16.30	0.035	61.0	170.0	0.9382	3.41	0.72	6.7	6
113	7.69	0.3393	0.239	12.20	0.028	22.0	125.0	0.9368	3.62	0.76	6.6	6
114	7.69	0.3393	0.239	12.20	0.028	22.0	125.0	0.9368	3.62	0.76	6.6	6
115	6.49	0.4143	-0.101	4.50	0.044	1.0	78.0	0.9340	4.03	0.65	7.2	6
116	6.99	0.2393	0.139	6.30	0.020	18.0	118.0	0.9314	3.71	0.69	8.8	6
117	7.99	0.0693	0.299	4.70	0.014	9.0	60.0	0.9311	3.59	0.67	9.3	9
118	8.19	0.2393	0.399	16.30	0.035	61.0	170.0	0.9382	3.41	0.72	6.7	9
119	8.29	0.2493	0.379	16.30	0.039	50.0	171.0	0.9382	3.44	0.75	6.7	9
120	6.89	0.2893	-0.061	8.70	0.025	14.0	62.0	0.9334	3.62	0.76	7.7	9
121	8.79	0.1693	0.219	15.20	0.033	35.0	113.0	0.9384	3.41	0.79	6.3	9
122	8.39	0.0893	0.209	9.85	0.038	24.0	106.0	0.9352	3.69	0.59	7.2	9
123	7.89	0.1193	0.179	8.00	0.037	7.0	121.0	0.9352	3.69	0.61	6.6	9
124	7.39	0.0593	0.369	4.60	0.071	33.0	133.0	0.9328	3.61	0.61	7.3	9
125	7.69	0.1193	0.259	4.10	0.005	56.0	118.0	0.9312	3.67	0.73	8.5	9
126	8.39	0.3193	0.129	10.00	0.012	22.0	101.0	0.9340	3.54	0.67	8.0	9
127	7.49	0.1693	0.219	10.60	0.017	41.0	178.0	0.9358	3.85	0.79	7.2	9
128	7.09	0.2293	0.459	5.80	0.023	40.0	154.0	0.9324	3.70	0.82	8.4	9
129	7.09	0.2293	0.459	5.70	0.025	39.0	159.0	0.9324	3.71	0.82	8.4	9
130	6.69	0.1893	0.149	13.40	-0.001	0.0	32.0	0.9340	3.79	0.62	8.1	9
131	7.49	0.1693	0.219	10.60	0.017	41.0	178.0	0.9358	3.85	0.79	7.2	7
132	7.49	0.3543	0.299	16.10	0.017	52.0	216.0	0.9379	3.63	0.82	6.5	7
133	7.59	0.1693	0.169	18.80	0.014	39.0	163.0	0.9382	3.64	0.76	6.7	7
134	7.79	0.1993	0.119	11.10	0.013	48.0	178.0	0.9361	3.59	0.77	6.4	7
135	7.69	0.1993	0.209	18.70	0.015	37.0	154.0	0.9379	3.66	0.81	7.1	7
136	9.19	0.1593	0.299	4.20	0.006	29.0	96.0	0.9320	3.52	0.63	8.2	7
137	8.09	0.2993	0.569	13.50	0.024	42.0	130.0	0.9375	3.56	0.69	6.2	7
138	7.79	0.1193	0.259	4.90	0.014	23.0	71.0	0.9317	3.55	0.79	8.3	7
139	9.09	0.2093	0.289	4.90	0.008	11.0	54.0	0.9323	3.63	0.77	9.3	8
140	7.29	0.2393	0.239	5.20	0.024	13.0	52.0	0.9327	3.70	0.68	7.7	8

The above table shows the entries for the red wine dataset between the 100th and 138th rows:
red_wine_sample = df_red.iloc[100:140]

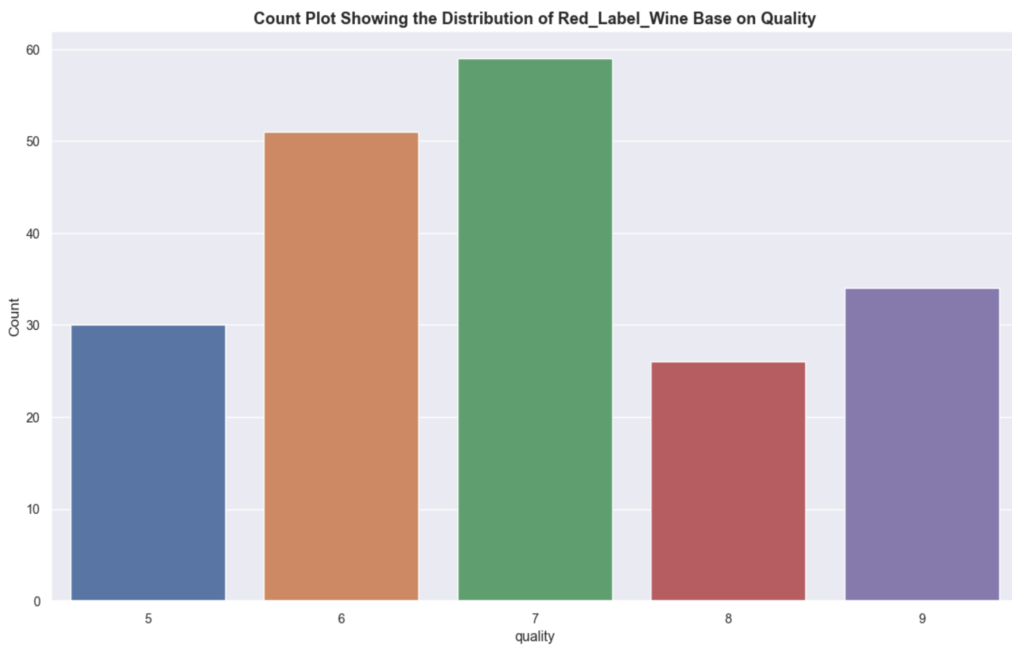
MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

Count	acidity	volatility level	citric level	sugar concentration	chloride	sulfur	Sum of silfur	Density level	PH	sulphate quantity	alcohol level	quality
0 200.0	200.00000	200.00000	200.00000	200.00000	200.00000	200.00000	200.00000	200.00000	200.000	200.00000	200.00000	200.00000
Mean	7.922500	0.21460	0.249250	9.801750	0.031060	31.242500	121.072500	0.934586	3.58210	0.717250	7.545000	6.915000
Std	0.710717	0.10117	0.126059	5.4587010	0.025505	16.784249	44.092821	0.002831	0.14295	0.095058	1.137473	1.29077
Min	6.190000	0.04930	0.101000	3.800000	0.001000	3.0000000	22.0000	0.92920	3.2700	0.520000	6.100000	5.000000
25%	7.465000	0.15930	0.179000	4.50000	0.019000	19.000000	86.750000	0.932375	3.49750	0.640000	6.600000	6.000000
50%	7.890000	0.19680	0.249000.1	8.300000	0.027000	31.000000	124.000000	0.9346500.	3.58000	0.710000	7.200000	7.000000
75%	8.290000	0.24930	0.309000	13.50000	0.34000	43.000000	148.250000	0.936925	3.68000	0.780000	8.300000	8.000000
Max	10.790000	0.59930	0.569000	25.00000	0.1790000	74.000000	230.000000	0.941000	4.09000	1.090000	10.300000	9.000000

The table above indicates that each column of the Red label win dataset has the same number of entries, 200, which is shown in the row count, and also the values for descriptive statistics.

Count Plot Showing the Distribution of Red_Label_Wine Base on Quality"

From the figure below we can see that the red_label_wine is graded on the scale of 5 to 9 as such the graph below show the Distribution of Red_Label_Wine Base on Quality. From the figure below most of the red_label_wine have a quality score of 7 followed by 6 and 9 respectively while a few wine can been seen in the quality score of 8 and 5 respectively.



In the next figure I used pairplot to showing likely Combination of Columns in Red_Label_Dataset

The preceding scattered plots below shows the likely pair of columns. The graph shows some positive correlation between some of the columns and a negative correlation with others, however a heat map in the subsequent graphs will clearly show which columns in the dataset are positively correlated vice versa.

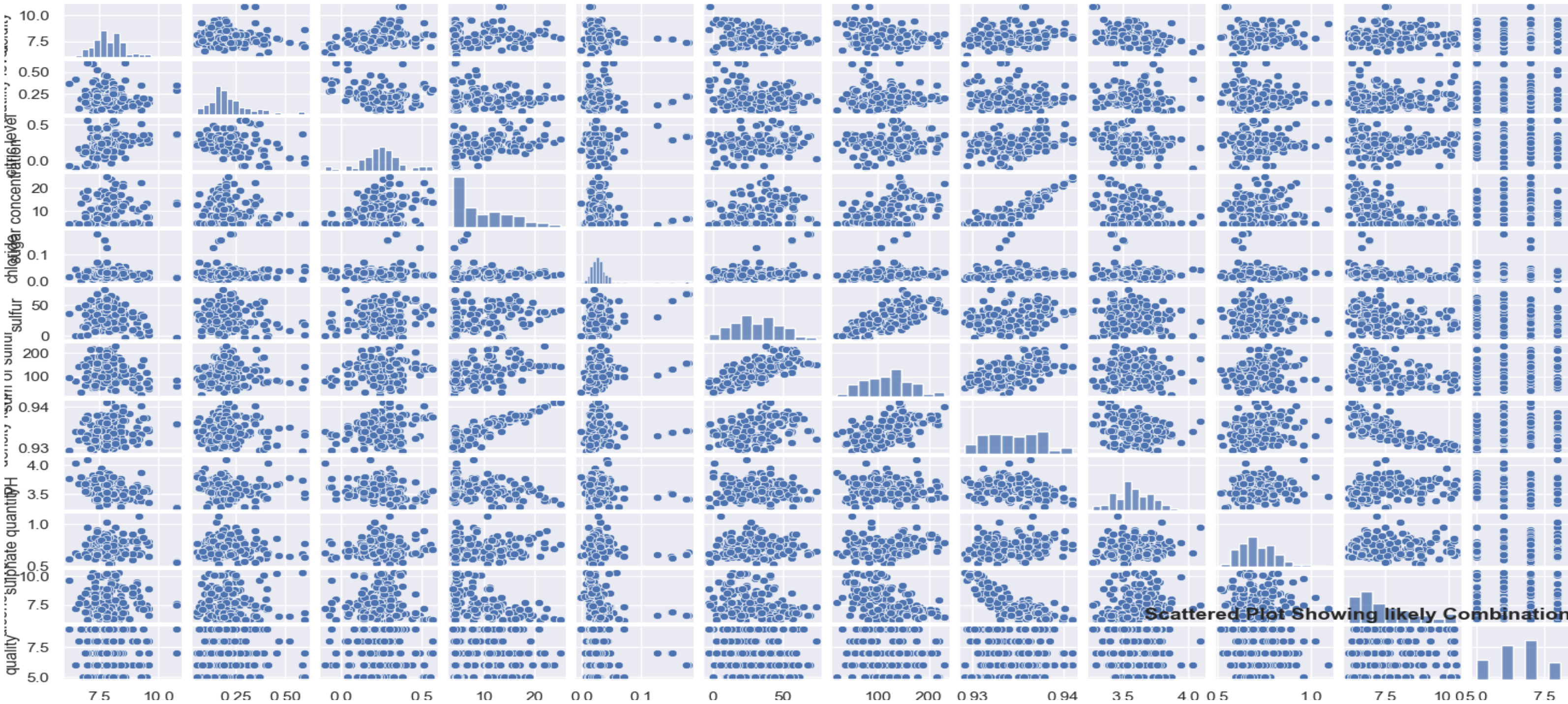
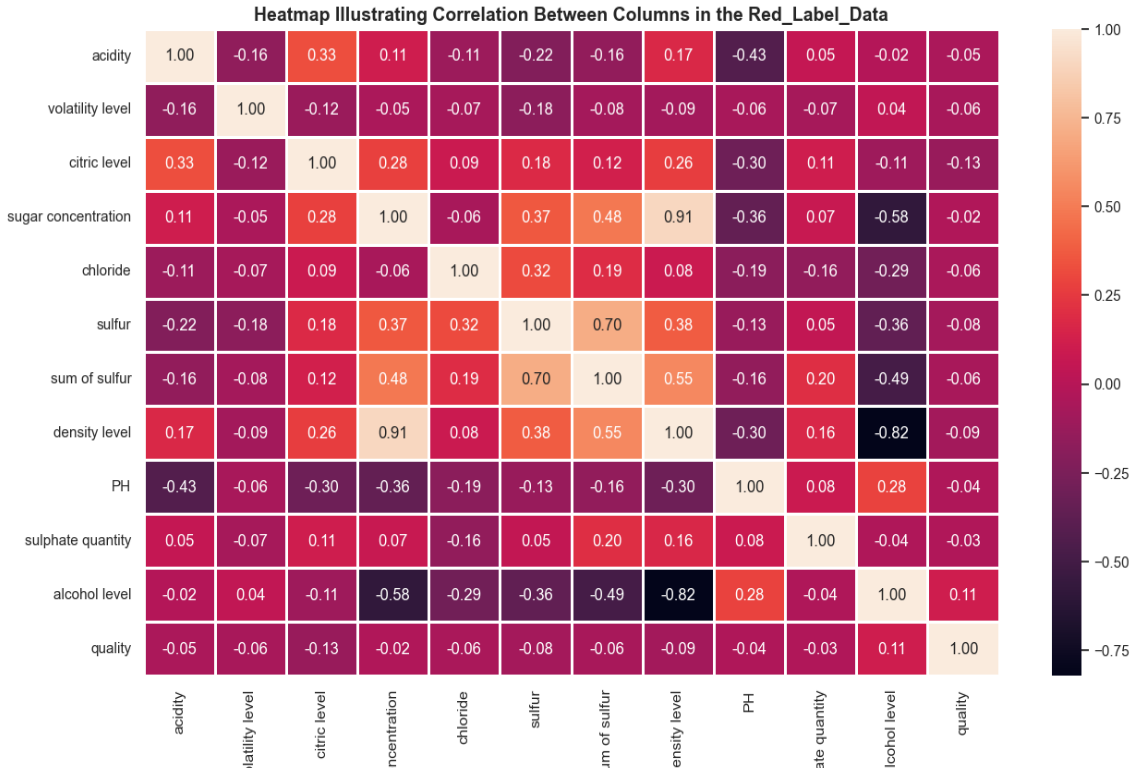


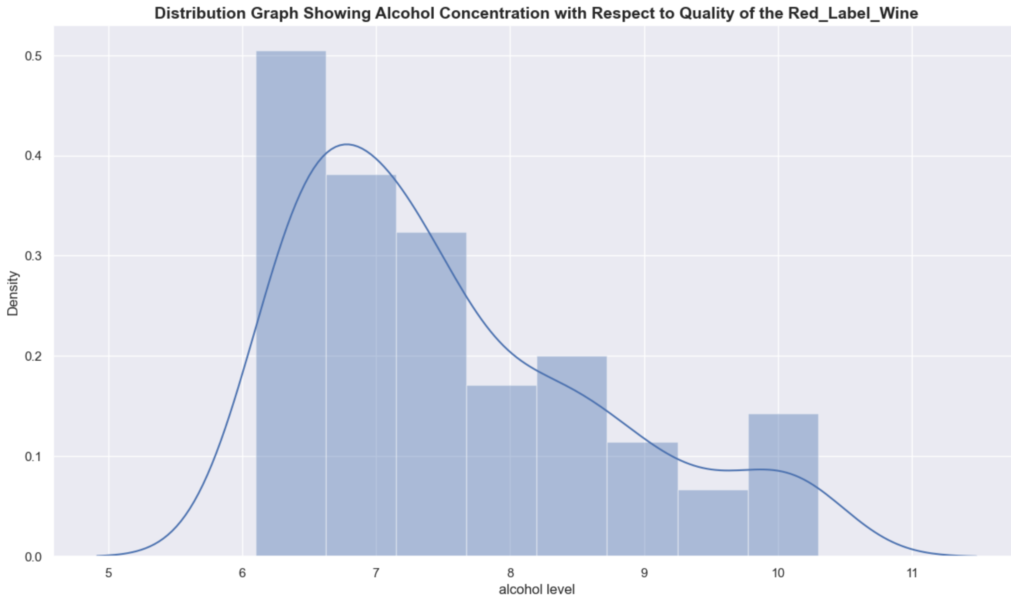
Figure below *shows* correlation between various columns. Let look at the quality column, the quality column has a positive correlation with only alcohol but negative correlation with sulfates quantity, sugar concentration, citric level, and acidity. Alcohol has a weak positive correlation with the quality of the red wine. Alcohol has a weak positive correlation with the pH value, negative correlation with citric level and density level have positive correlation with acidity. PH has a negative correlation with density level, acidity, citric level, but positive correlation with sulfates quantity.

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT



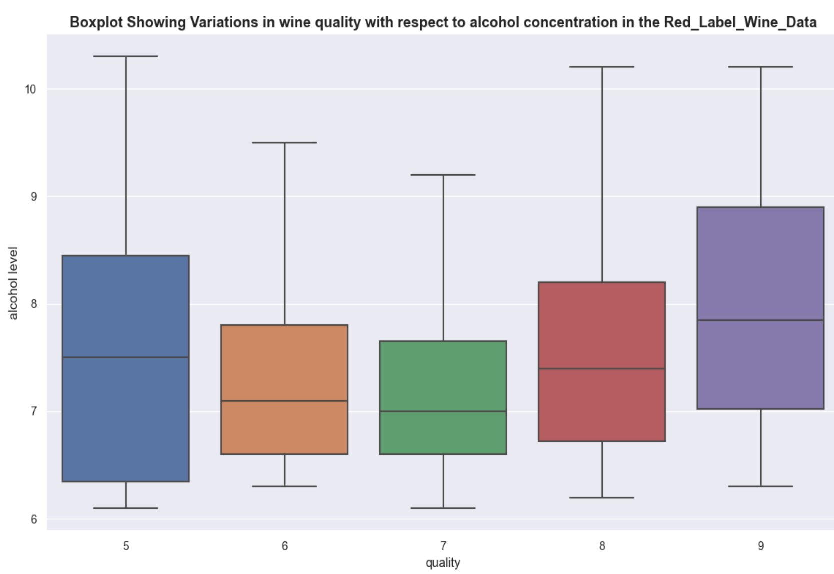
Using sns.distplot the figure below shows that alcohol distribution is positively skewed with the quality of the red label wine. It means that as the quality of the red wine increases, the distribution of alcohol content tends to have a longer tail on the right side (higher values). In other words, higher-quality red wines are more likely to have higher alcohol content.

This positive skewness suggests that there may be a positive relationship between alcohol content and wine quality. From the previous figure you can see that alcohol has a positive correlation value of 0.11 with quality.

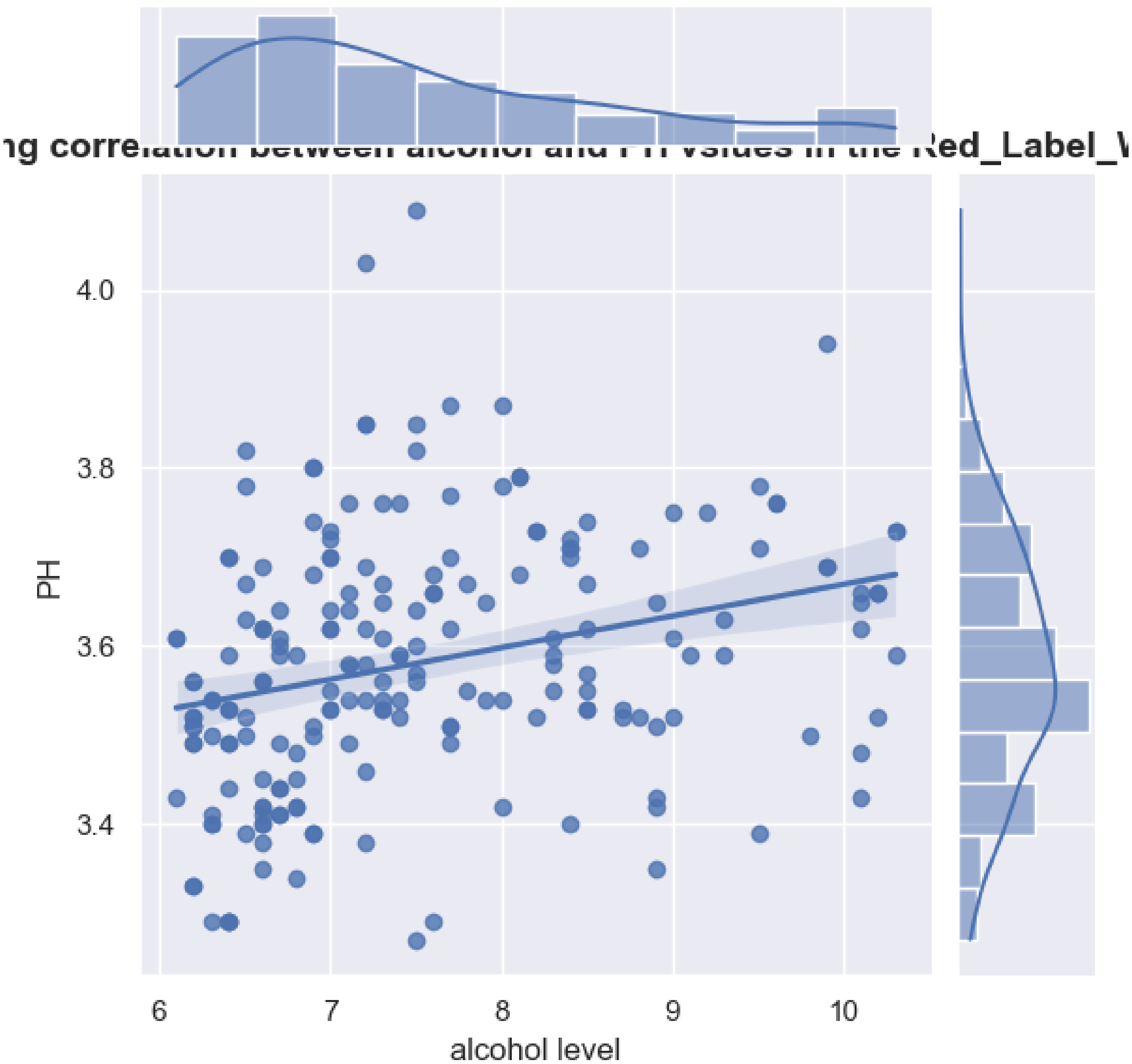


Alcohol versus quality

Using boxplot from seaborn the next figure shows how the quality of red label wine varies with respect to alcohol level. It seems that as the quality of red label wine increases, so does the Alcohol concentration. Thus the higher the alcohol concentration
, the higher the quality of the red label wine.



Now let's look at how the alcohol column and pH levels relate to one another. Based on the previous plots, we already know they have a weakly positive association. Let's now validate the conclusions from this section: Alcohol and pH levels have a weakly positive correlation, as demonstrated in the graph below. Additionally, the screenshot shows the regression line, which shows how they are correlated.



MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

Mercado Black Label Analysis

Black wine dataframe output.

	acidity	volatility level	citric level	sugar concentration	chloride	sulfur	sum of sulfur	density level	PH	sulphate quantity	alcohol level	quality
100	13.8	0.760	0.32	2.25	0.504	17	53	0.9972	3.91	0.713	10.25	7
101	13.3	0.650	0.32	2.05	0.495	14	25	0.9959	3.82	0.663	10.45	7
102	13.6	0.695	0.20	2.05	0.500	19	38	0.9972	3.81	0.693	9.05	7
103	13.6	0.725	0.24	2.25	0.497	18	68	0.9967	3.80	0.613	9.25	6
104	12.7	0.640	0.26	2.35	0.490	11	39	0.9960	3.84	0.583	9.45	6
105	13.6	0.725	0.24	2.25	0.497	18	68	0.9967	3.80	0.613	9.25	6
106	13.3	0.560	0.70	1.85	0.887	24	72	0.9973	3.59	1.413	9.35	6
107	11.7	0.780	0.33	1.85	0.508	21	67	0.9969	3.97	0.893	9.35	6
108	13.5	0.480	0.55	2.65	0.511	24	83	0.9976	3.88	0.903	9.65	7
109	13.6	0.935	0.54	2.15	0.542	43	156	0.9969	3.72	0.793	9.35	6
110	13.3	0.710	0.21	1.95	0.524	18	50	0.9964	3.70	1.033	9.55	6
111	13.9	0.770	0.11	2.35	0.504	17	111	0.9964	3.66	0.763	9.85	6
112	13.9	0.750	0.12	2.35	0.505	20	114	0.9964	3.66	0.763	9.85	6
113	15.6	0.460	0.46	2.45	0.500	28	49	0.9988	3.83	0.773	9.75	7
114	13.3	0.710	0.21	1.95	0.524	18	50	0.9964	3.70	1.033	9.55	6
115	14.9	0.550	0.33	2.35	0.510	19	65	0.9966	3.58	0.733	10.55	7
116	13.8	0.690	0.30	2.05	0.497	17	43	0.9978	3.90	0.713	10.05	7
117	13.3	0.710	0.14	2.15	0.502	13	31	0.9970	3.88	0.603	9.45	7
118	14.3	0.700	0.06	2.35	0.539	20	59	0.9962	3.72	0.703	10.95	7
119	12.5	0.840	0.10	1.95	0.517	28	92	0.9959	3.85	0.643	9.25	7
120	12.8	1.220	0.11	1.85	0.598	16	92	0.9962	3.81	0.673	9.05	6
121	14.3	0.700	0.06	2.35	0.539	20	59	0.9962	3.72	0.703	10.95	7
122	12.8	0.845	0.02	2.65	0.495	9	16	0.9980	4.00	0.623	9.25	6
123	13.5	0.860	0.02	2.75	0.500	17	37	0.9976	3.95	0.633	9.55	6
124	13.3	0.650	0.19	1.75	0.502	27	105	0.9960	3.90	0.583	9.55	6
125	14.5	0.770	0.06	2.05	0.566	33	93	0.9984	3.67	0.803	9.45	6
126	13.7	1.480	0.02	1.85	0.501	9	15	0.9964	4.04	0.593	10.95	6
127	13.6	1.480	0.02	1.95	0.502	9	15	0.9964	4.05	0.583	10.95	6
128	13.5	0.740	0.18	1.95	0.485	9	19	0.9962	3.93	1.023	10.55	8
129	11.6	0.530	0.17	1.95	0.492	12	22	0.9955	3.93	0.673	9.45	6
130	13.5	0.895	0.58	2.15	0.538	36	137	0.9968	3.75	0.763	9.45	6
131	11.1	0.650	0.11	2.45	0.469	23	102	0.9937	4.14	0.733	13.05	6
132	11.1	0.650	0.11	2.45	0.469	23	102	0.9937	4.14	0.733	13.05	6
133	12.1	0.650	0.03	1.65	0.480	23	29	0.9952	3.91	0.683	9.85	7
134	13.4	1.190	0.07	2.35	0.504	19	32	0.9959	3.73	0.653	9.95	7
135	13.9	0.895	0.13	2.05	0.510	22	66	0.9965	3.70	0.923	9.65	6
136	13.8	0.865	0.17	1.95	0.509	16	55	0.9968	3.74	0.873	9.55	6
137	12.7	0.565	0.38	2.15	0.501	19	48	0.9972	3.99	0.743	9.25	6
138	13.3	0.710	0.21	2.25	0.501	21	108	0.9962	3.84	0.643	9.55	6
139	13.3	0.710	0.21	2.15	0.501	23	111	0.9962	3.83	0.643	9.55	6
140	13.9	0.895	0.13	2.05	0.510	22	66	0.9965	3.70	0.923	9.65	6

The above table shows the entries for the black label wine dataset between the 100th and 140th rows: red_wine_sample = df_red.iloc[100:140]

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

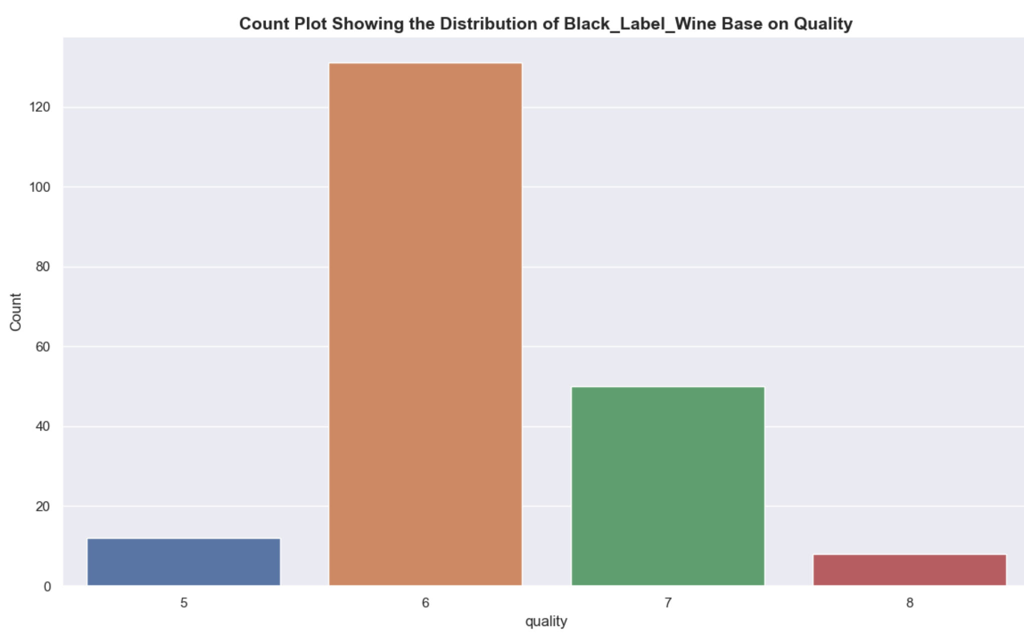
The Descriptive output of the Dataset

```
      acidity  volatility level  citric level  sugar concentration    chloride      sulfur  sum of sulfur  density level          PH  sulphate quantity  alcohol level     quality
count 201.000000        201.000000    201.000000           201.000000  201.000000  201.000000     201.000000     201.000000  201.000000         201.000000     201.000000  201.000000
mean   13.103980          0.723085      0.236269             2.513184    0.518682   21.577114      63.348259       0.996626    3.843980           0.783000       9.898756    6.268657
std     1.021021          0.178068      0.182539             1.216311    0.072378    9.382179      38.088950       0.001175    0.158196           0.266919       0.838398    0.630446
min    10.100000          0.370000      0.020000             1.350000    0.465000    9.000000      11.000000       0.991600    3.250000           0.433000       9.050000    5.000000
25%    12.600000          0.600000      0.090000             1.950000    0.493000   15.000000      32.000000       0.996200    3.740000           0.643000       9.450000    6.000000
50%    13.200000          0.710000      0.200000             2.150000    0.501000   20.000000      55.000000       0.996800    3.850000           0.693000       9.550000    6.000000
75%    13.600000          0.810000      0.330000             2.450000    0.512000   25.000000      92.000000       0.997200    3.930000           0.823000      10.150000    7.000000
max    17.000000          1.480000      1.020000            10.850000    1.030000   58.000000     156.000000       0.999600    4.410000           2.103000      14.050000    8.000000
black wine describe:           acidity  volatility level  citric level  sugar concentration  ...          PH  sulphate quantity  alcohol level     quality
count 201.000000        201.000000    201.000000           201.000000  ...  201.000000         201.000000     201.000000  201.000000
mean   13.103980          0.723085      0.236269             2.513184  ...    3.843980           0.783000       9.898756    6.268657
std     1.021021          0.178068      0.182539             1.216311  ...    0.158196           0.266919       0.838398    0.630446
min    10.100000          0.370000      0.020000             1.350000  ...    3.250000           0.433000       9.050000    5.000000
25%    12.600000          0.600000      0.090000             1.950000  ...    3.740000           0.643000       9.450000    6.000000
50%    13.200000          0.710000      0.200000             2.150000  ...    3.850000           0.693000       9.550000    6.000000
75%    13.600000          0.810000      0.330000             2.450000  ...    3.930000           0.823000      10.150000    7.000000
max    17.000000          1.480000      1.020000            10.850000  ...    4.410000           2.103000      14.050000    8.000000

[8 rows x 12 columns]
```

The table above is the output of the pd.describe() method, indicates that each column of the Black Label win dataset has the same number of entries, 201, which is shown in the row count, and also the values for descriptive statistics.

Count Plot Showing the Distribution of Black_Label_Wine Base on Quality"



From the figure above we can see that the black_label_wine is graded on the scale of 5 to 8 as such the graph below show the Distribution of black_Label_Wine Base on Quality. From the figure below most of the black_label_wine have a quality score of 6 followed by 7 followed by 5 and 8 respectively.

Image showing correlation between columns in the Black Label Dataset

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

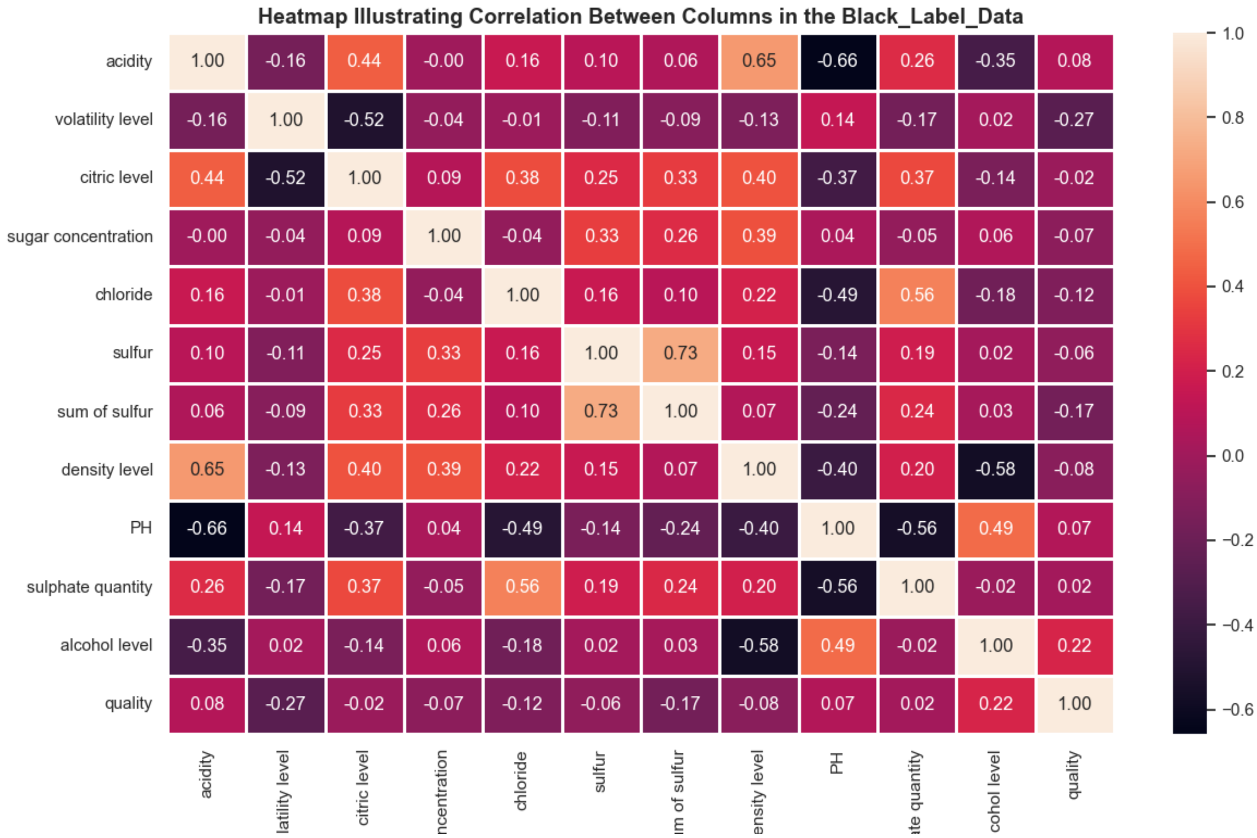
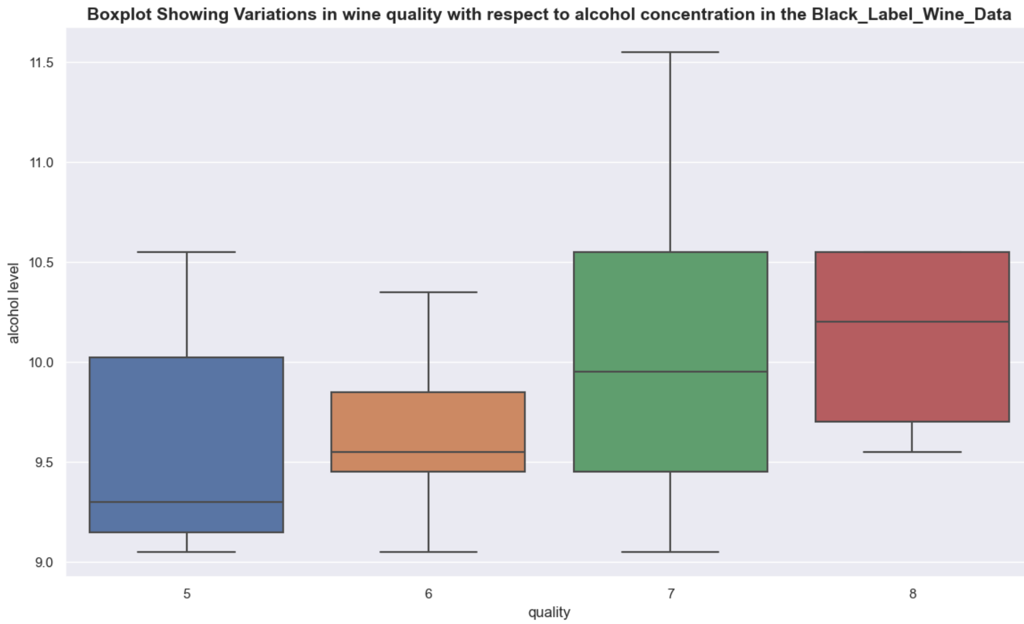


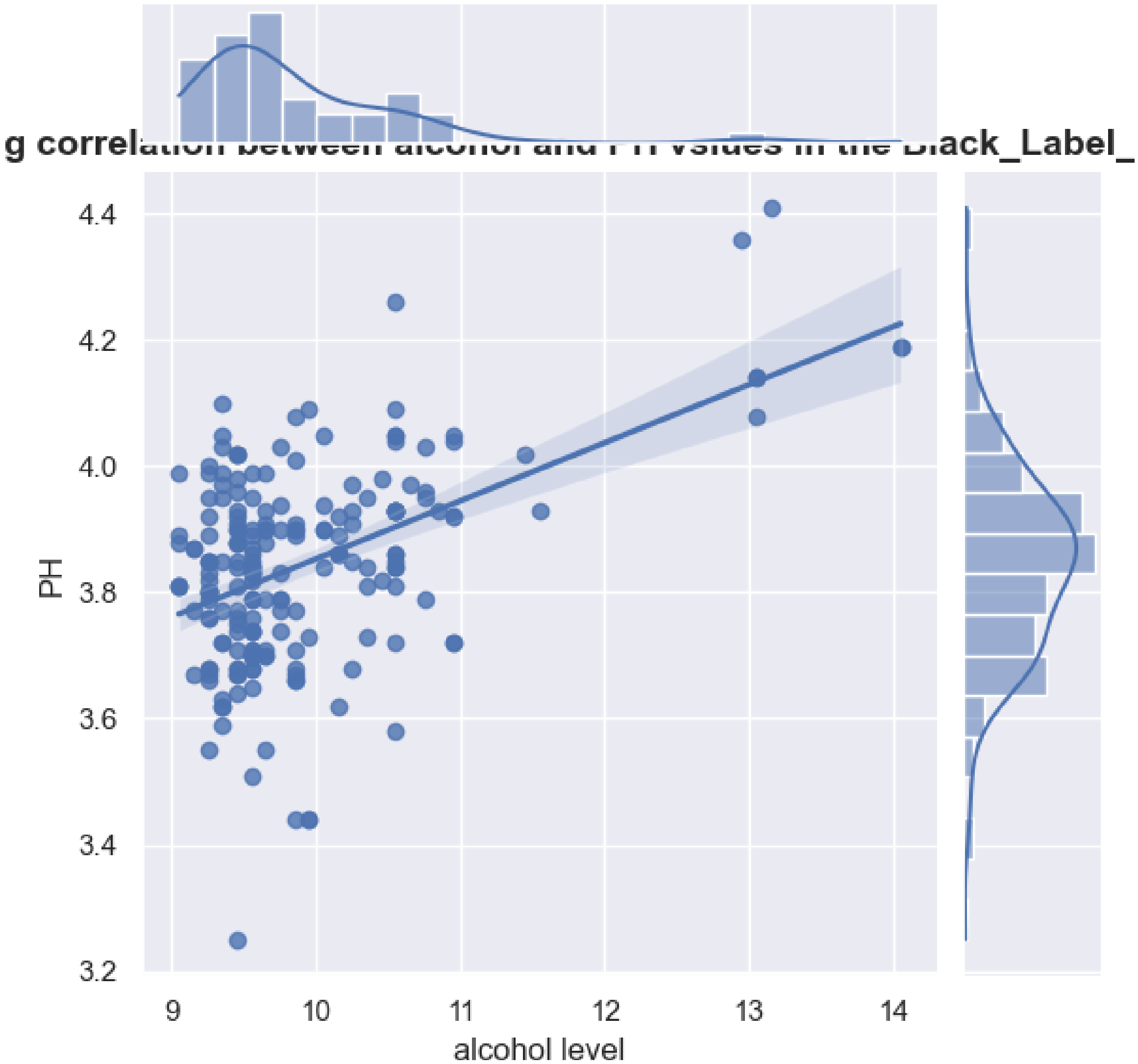
Figure above *shows* correlation between various columns. Let look at the quality column, the quality column has a positive correlation with alcohol, sulfhate, PH and acidity but negative correlation with density level, sugar concentration, citric level, and acidity. Alcohol has a weak positive correlation with the quality of the red wine. Alcohol has a weak positive correlation with the pH value, negative correlation with citric level and density level have positive correlation with acidity. PH has a negative correlation with density level, acidity, citric level, but positive correlation with alcohol level.

Alcohol versus quality



Using boxplot from seaborn the figure above shows how the quality of black label wine varies with respect to alcohol level. It seems that as the quality of red label wine increases, so does the Alcohol concentration. Thus the higher the alcohol concentration, the higher the quality of the black label wine.

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT



Let's now examine the relationship between the pH levels and the alcohol column in the black label dataset. We already know they have a weakly positive correlation from the earlier graphs. Now let's verify the findings from this section: The graph below shows the weakly positive association between alcohol and pH levels. Furthermore, the regression line illustrating their correlation is displayed in the screenshot above.

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

Converting into a categorical column

In the next stage I concatenated the two dataframe (Black wine and Red wine) but first let first convert the numerical values (quality column) into categorical values (quality_label column) and add a price column to the datasets before concatenating the two dataframe. Let give a breakdown of the parameters we used to create the categorical column and added the price

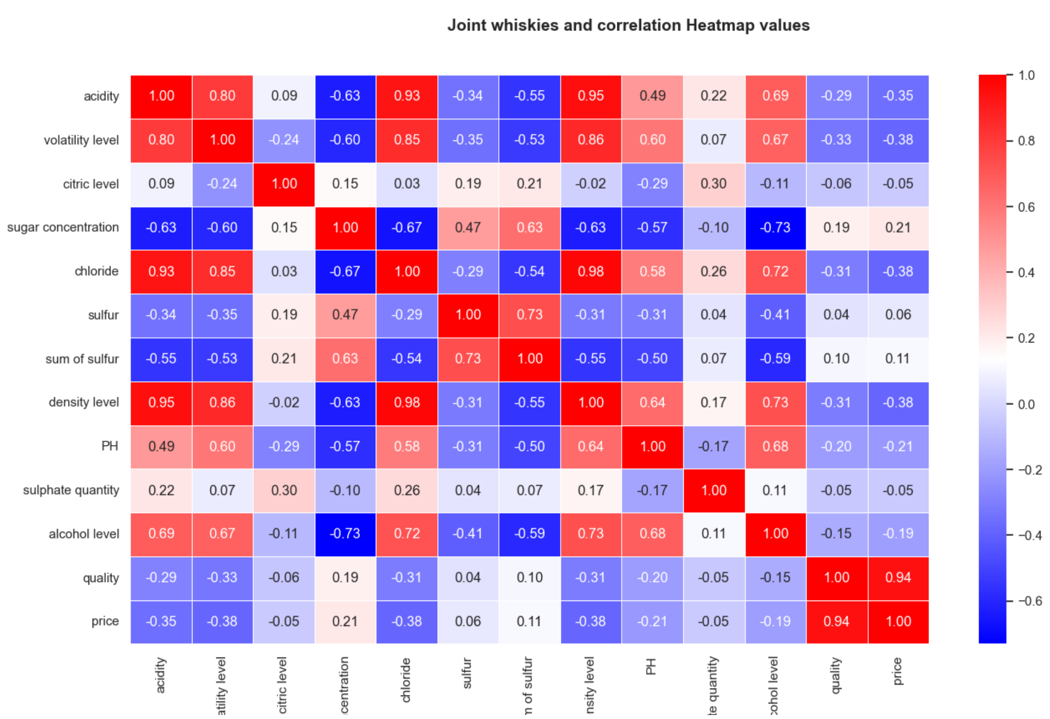
```
#parameters used in changing numerical column to categorical column
low_value= "<=5"
medium="<=7"
High=">7"
df_black['quality_label'] = df_black['quality'].apply(lambda x: 'low' if x < 7 else ('medium' if x == 7 else 'high'))
df_red['quality_label'] = df_red['quality'].apply(lambda x: 'low' if x < 7 else ('medium' if x == 7 else 'high'))
# Adding price column
def map_quality_to_price(label):
    if label=='low':
        return 10
    elif label == 'medium':
        return 15
    elif label == 'high':
        return 20
    else:
         return None
df_black['price']=df_black['label'].apply(map_quality_to_price)
df_red['price']=df_red['label'].apply(map_quality_to_price)
```

After converting the numerical column into categorical column and adding a price column in the two dataframes, they were concatenated and the screenshot below shows the result of the concatenation.

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

	acidity	volatility level	citric level	sugar concentration	chloride	sulfur	sum of sulfur	density level	PH	sulphate quantity	alcohol level	quality	group	label	price
0	8.09	0.3593	0.509	14.80	0.024	47.0	130.0	0.9374	3.51	0.700	6.20	5	red label	low	10
1	7.79	0.1493	0.209	4.40	0.032	27.0	89.0	0.9329	3.79	1.020	8.10	8	red label	high	20
2	12.40	0.7550	0.140	10.85	0.493	46.0	86.0	0.9993	3.96	0.623	9.45	7	black Label	medium	15
3	9.09	0.1993	0.309	4.45	0.012	4.0	38.0	0.9308	3.39	0.810	9.50	6	red label	low	10
4	13.20	0.6400	0.280	2.05	0.482	15.0	34.0	0.9966	3.90	0.743	9.65	6	black Label	low	10
5	11.80	0.4500	0.500	1.95	0.489	24.0	64.0	0.9959	3.95	0.883	10.35	7	black Label	medium	15
6	7.49	0.2093	0.179	11.50	0.026	47.0	185.0	0.9362	3.49	0.790	6.40	7	red label	medium	15
7	10.50	1.1700	0.060	1.55	0.465	47.0	88.0	0.9938	4.26	0.583	10.55	5	black Label	low	10
8	7.59	0.1993	0.309	4.30	0.031	9.0	117.0	0.9351	3.82	0.720	7.50	5	red label	low	10
9	13.70	1.4800	0.020	1.85	0.501	9.0	15.0	0.9964	4.04	0.593	10.95	6	black Label	low	10
10	13.00	0.6500	0.380	6.25	0.491	23.0	105.0	0.9978	3.86	0.903	10.55	6	black Label	low	10
11	7.39	0.1493	0.459	17.50	0.034	20.0	134.0	0.9380	3.38	0.650	6.60	9	red label	high	20
12	15.70	0.5700	0.590	3.55	0.490	10.0	13.0	0.9971	3.55	0.733	9.65	6	black Label	low	10
13	13.20	0.8400	0.240	2.05	0.504	24.0	97.0	0.9961	3.82	0.583	9.55	6	black Label	low	10
14	11.10	0.6500	0.110	2.45	0.469	23.0	102.0	0.9937	4.14	0.733	13.05	6	black Label	low	10
15	13.00	0.6400	0.220	2.75	0.752	14.0	17.0	0.9968	3.72	1.003	10.55	7	black Label	medium	15
16	7.79	0.1293	0.489	3.90	0.126	31.0	107.0	0.9330	3.45	0.630	6.60	7	red label	medium	15
17	8.09	0.1093	0.259	4.40	0.022	24.0	62.0	0.9298	3.66	0.620	10.20	8	red label	high	20
18	9.29	0.0693	0.239	4.10	0.021	0.0	22.0	0.9334	3.87	0.650	7.70	5	red label	low	10
19	7.99	0.1993	0.259	23.70	0.024	38.0	145.0	0.9410	3.40	0.700	6.30	6	red label	low	10
20	12.30	0.9350	0.020	2.55	0.524	20.0	33.0	0.9966	4.03	0.653	10.75	7	black Label	medium	15

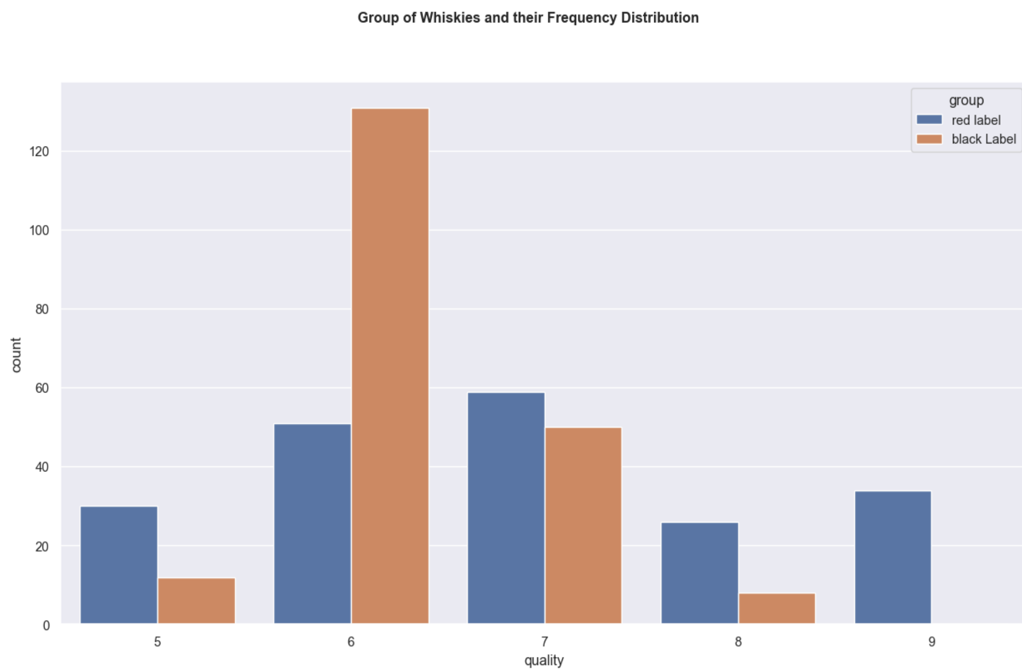
The image below is a multivariate analysis on the combined dataframe.



MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

The figure above *shows* correlation between various columns. Let look at the quality column, the quality column has a strong positive correlation with price and sugar concentration, but negative correlation with density level, citric level, and acidity. Alcohol on the other hand has a strong positive correlation with the PH, chloride and density level but a weak negative correlation with quality but strong negative correlation with sugar concentration and sum of sulfur in the combined data frame (Join whiskies)

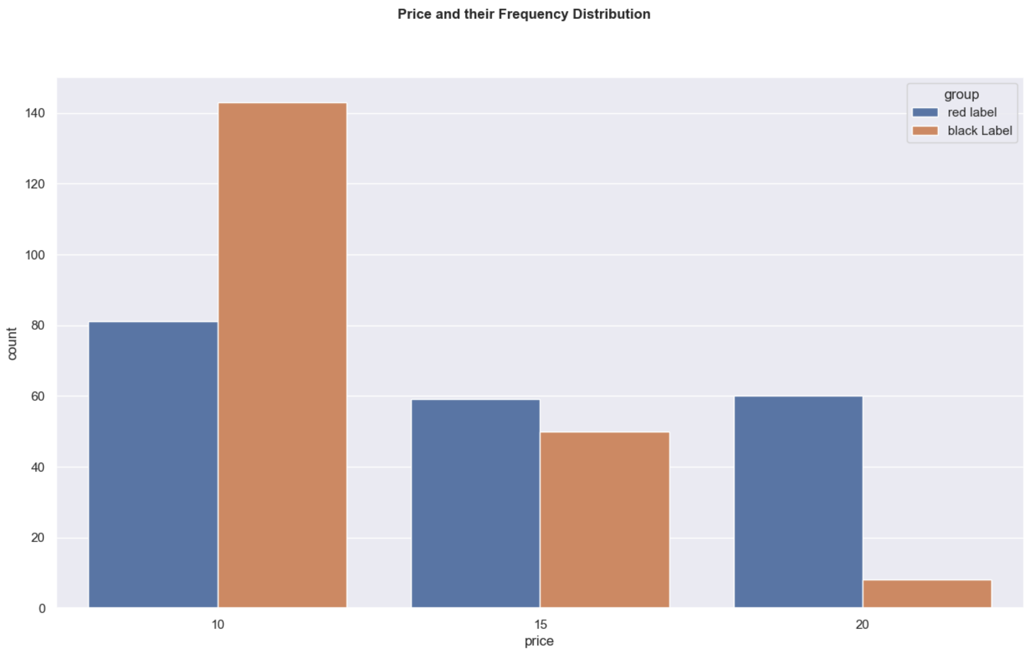
Categorizations of group of whiskies based on the quality



The accompanying figure displays the frequency distributions for various wine group based on quality categorization [5, 6, 7, 8, and 9] across a beautiful count plot. For end users, it is a more understandable illustration. From the graph, about 130 of the black wine belong to the quality group of 6, while around 50 are counted in the quality group of 7 followed by 10 and around 5 for 5 and 8 respectively. With no record for quality group of 9 for black wine. On the other hands, the red wine has more count for quality group 7 followed 6 with about 50 count. Around 35 of the red wine are within the quality mark of 9 followed by 5 and 8 respectively.

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

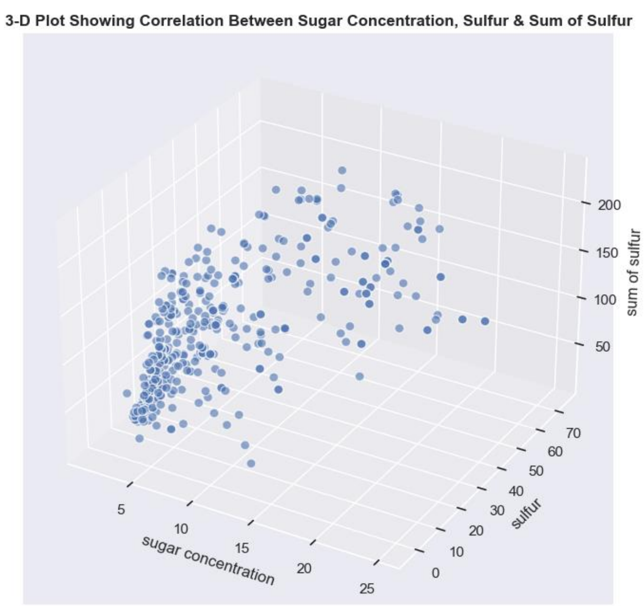
Categorizations of group of whiskies based on the Prices



The accompanying figure above displays the frequency distributions for various wine prices [10, 15, and 20] across a beautiful count plot. For end users, it is a more understandable illustration, from the figure more of the black label wine about 150 are in the 10 dollar price column followed by 15 dollars and finally 20 USD while on the other hands, majority of the red label wine about 80 are priced at 10 USD followed by 10 and 15 USD respectively.

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT

Three Dimensional visualization of three variables (sum of sulfur, sugar concentration and sulfur)



Using the Matplotlib library to generate the 3D dimension visualization, the figure above illustrate that the three variables show a positive correlation with respect to one another.

Model development and evaluation

The final part of the work dealt into model development specifically logistic regression and evaluation. Using model to classify or predict a variable. First using label encoding I encode the quality label column i.e. from categorical variable to numeric then split the dataset into a training set and test set. The data was grouped into x, y i.e. X representing the predictor in this case, **X_classification** consists of all features except the 'alcohol level', why the Y represent the outcome or classify In this case, **classification** represents the label column. After this the model was initialized and trained the logistic model was fixed on the split variables for prediction of the Y variable. After this the accuracy score and the mean of square error was computed. The following screenshot is the result.

	Actual	Predicted
285	1	1
281	1	1
33	1	1
211	1	1
93	1	1
84	3	3
390	3	3
94	1	1
225	1	1
126	1	1
9	1	1
362	2	2
56	2	2
72	1	1
132	3	3
42	1	1
255	1	1
277	1	1
232	2	3
208	1	1
77	3	2
15	2	2
392	2	2
349	2	2
0	1	1

Prediction score/accuracy: **0.9917355371900827**

A prediction score of 0.9917 suggests that the model's accuracy is approximately 99.17%.

Mean Squared Error**: 0.008264462809917356**

A Mean Squared Error of 0.0083 suggests that, on average, the squared difference between the model's predictions and the actual values is approximately 0.0083. A Lower values indicating a high performance.

THANK YOU!

MERCADO WINE MARCHING LEARNING ANALSIS PROJECT