

Github LINK : git clone <https://github.com/moonc/CS131.git>

- 1) feature and the target variables, obtain the mean, standard deviation, min, max, and 25/50/75% percentiles

```
[ ] features.describe()
```

	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	total_sulfur_dioxide	density	pH	sulphates	alcohol
count	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000
mean	6.854788	0.278241	0.334192	6.391415	0.045772	35.308085	138.360657	0.994027	3.188267	0.489847	10.514267
std	0.843868	0.100795	0.121020	5.072058	0.021848	17.007137	42.498065	0.002991	0.151001	0.114126	1.230621
min	3.800000	0.080000	0.000000	0.600000	0.009000	2.000000	9.000000	0.987110	2.720000	0.220000	8.000000
25%	6.300000	0.210000	0.270000	1.700000	0.036000	23.000000	108.000000	0.991723	3.090000	0.410000	9.500000
50%	6.800000	0.260000	0.320000	5.200000	0.043000	34.000000	134.000000	0.993740	3.180000	0.470000	10.400000
75%	7.300000	0.320000	0.390000	9.900000	0.050000	46.000000	167.000000	0.996100	3.280000	0.550000	11.400000
max	14.200000	1.100000	1.660000	65.800000	0.346000	289.000000	440.000000	1.038980	3.820000	1.080000	14.200000

```
targets.describe()
```

	quality
count	4898.000000
mean	5.877909
std	0.885639
min	3.000000
25%	5.000000
50%	6.000000
75%	6.000000
max	9.000000

- 2) Heatmaps for Correlation Matrix [Illustrations are B & C Respectively]

```
[95] feature_correlation = features.corr()
feature_correlation
```

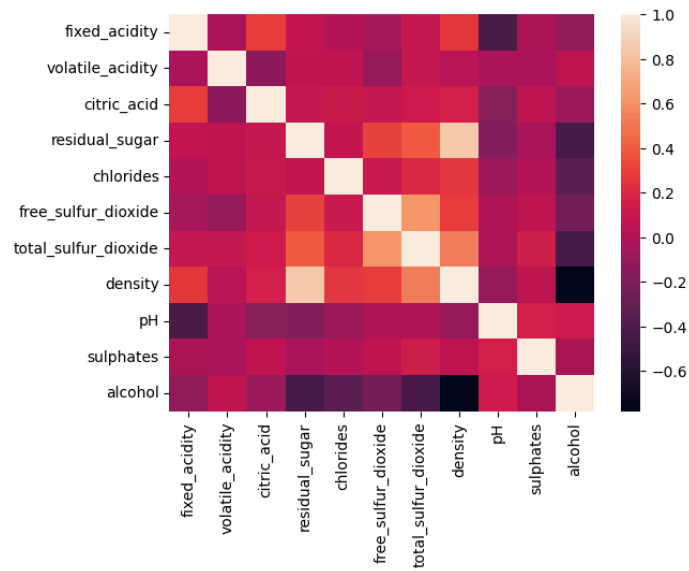
	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	total_sulfur_dioxide	density	pH	sulphates	alcohol
fixed_acidity	1.000000	-0.022697	0.289181	0.089021	0.023086	-0.049396	0.091070	0.265331	-0.425858	-0.017143	-0.120881
volatile_acidity	-0.022697	1.000000	-0.149472	0.064286	0.070512	-0.097012	0.089261	0.027114	-0.031915	-0.035728	0.067718
citric_acid	0.289181	-0.149472	1.000000	0.094212	0.114364	0.094077	0.121131	0.149503	-0.163748	0.062331	-0.075729
residual_sugar	0.089021	0.064286	0.094212	1.000000	0.088685	0.299098	0.401439	0.838966	-0.194133	-0.026664	-0.450631
chlorides	0.023086	0.070512	0.114364	0.088685	1.000000	0.101392	0.198910	0.257211	-0.090439	0.016763	-0.360189
free_sulfur_dioxide	-0.049396	-0.097012	0.094077	0.299098	0.101392	1.000000	0.615501	0.294210	-0.000618	0.059217	-0.250104
total_sulfur_dioxide	0.091070	0.089261	0.121131	0.401439	0.198910	0.615501	1.000000	0.529881	-0.093591	0.134562	-0.448892
density	0.265331	0.027114	0.149503	0.838966	0.257211	0.294210	0.529881	1.000000	-0.093591	0.074493	-0.780138
pH	-0.425858	-0.031915	-0.163748	-0.194133	-0.090439	-0.000618	0.002321	-0.093591	1.000000	0.155951	0.121432
sulphates	-0.017143	-0.035728	0.062331	-0.026664	0.016763	0.059217	0.134562	0.074493	0.155951	1.000000	-0.017433
alcohol	-0.120881	0.067718	-0.075729	-0.450631	-0.360189	-0.250104	-0.448892	-0.780138	0.121432	-0.017433	1.000000

Next steps: [Generate code with feature_correlation](#) [View recommended plots](#) [New interactive sheet](#)

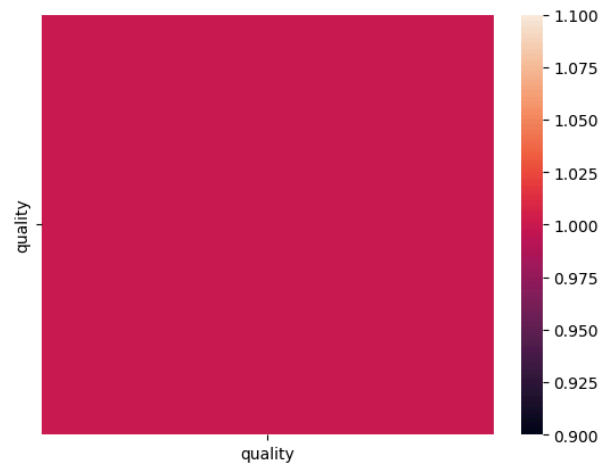
```
[96] target_correlation = targets.corr()
target_correlation
```

	quality
quality	1.0

a)

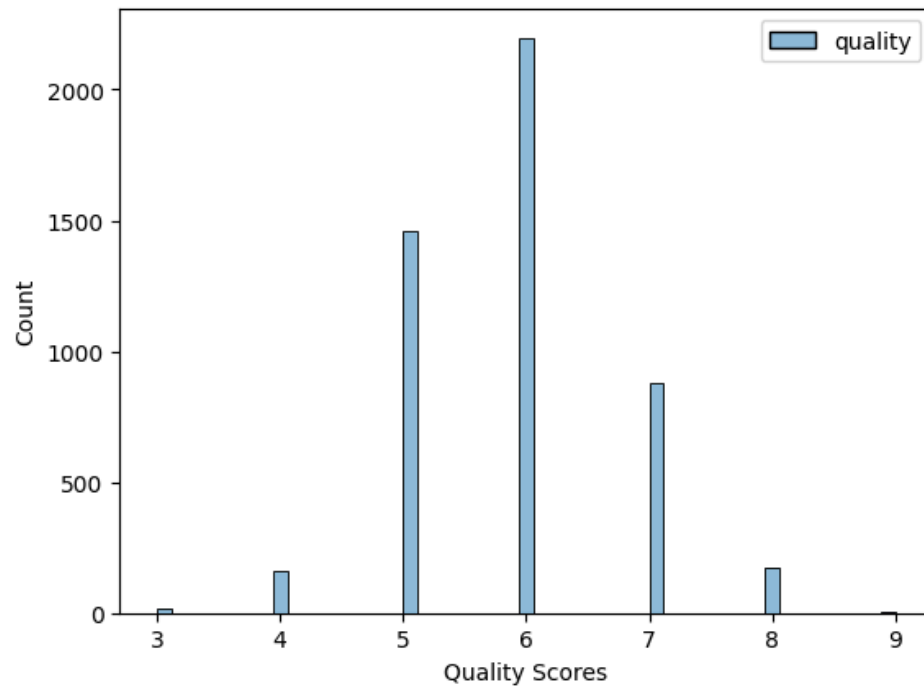


b)



c)

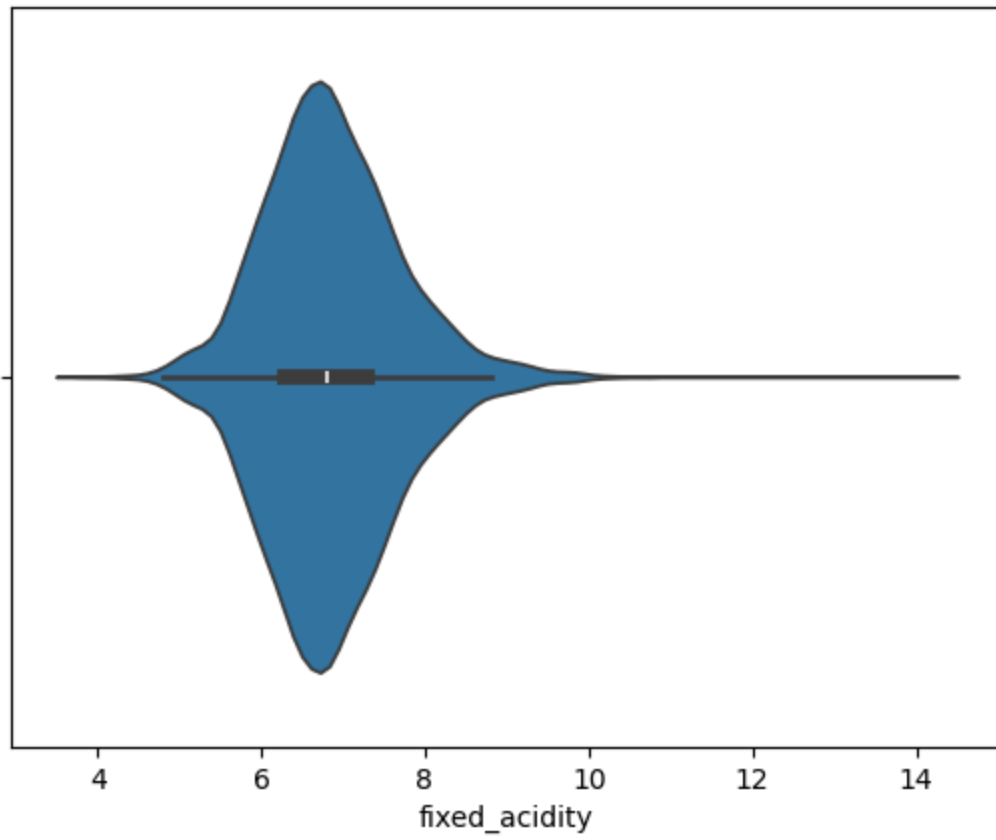
3) Plot a histogram of the target value (quality).

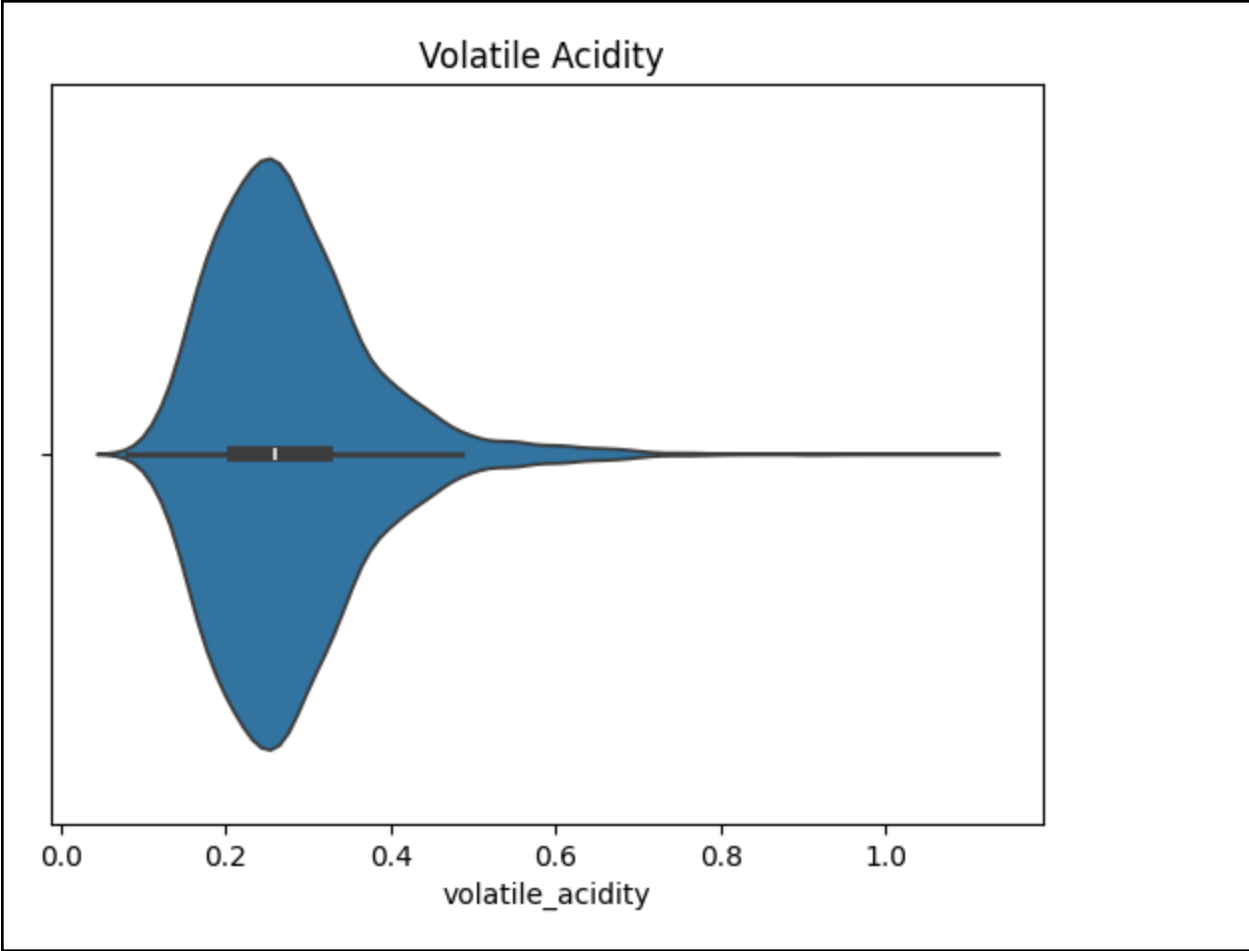


a)

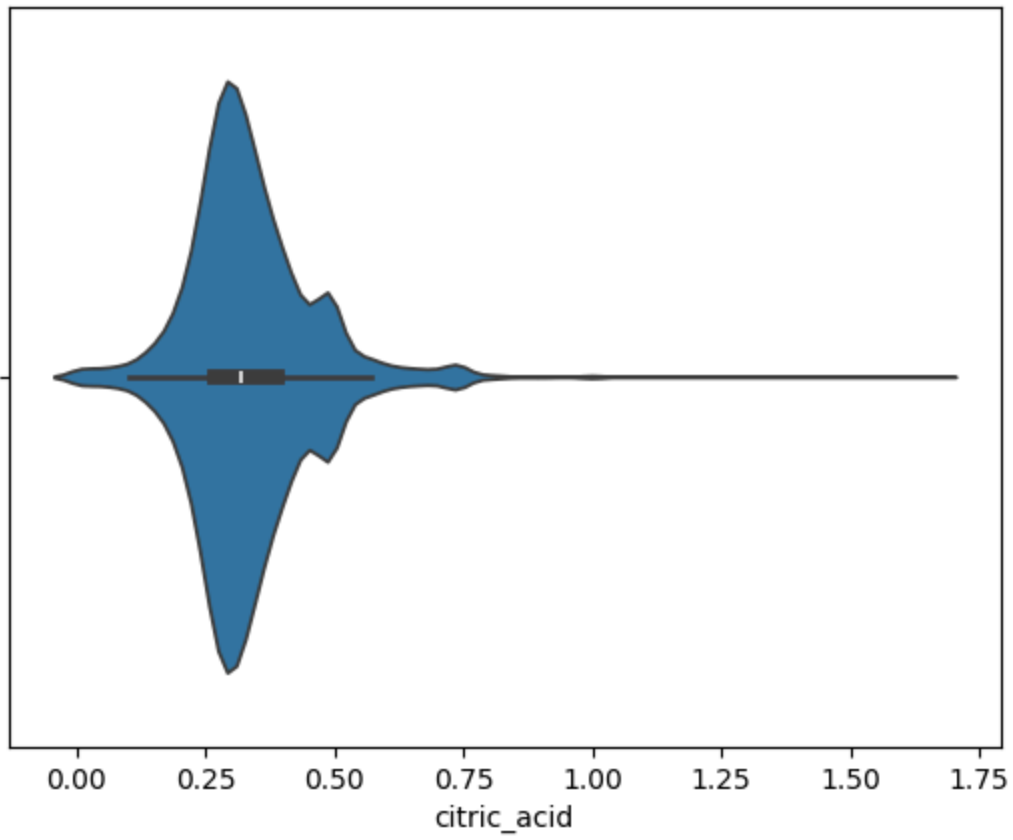
4) Violin Plots of White Wine (on next page)

Fixed Acidity

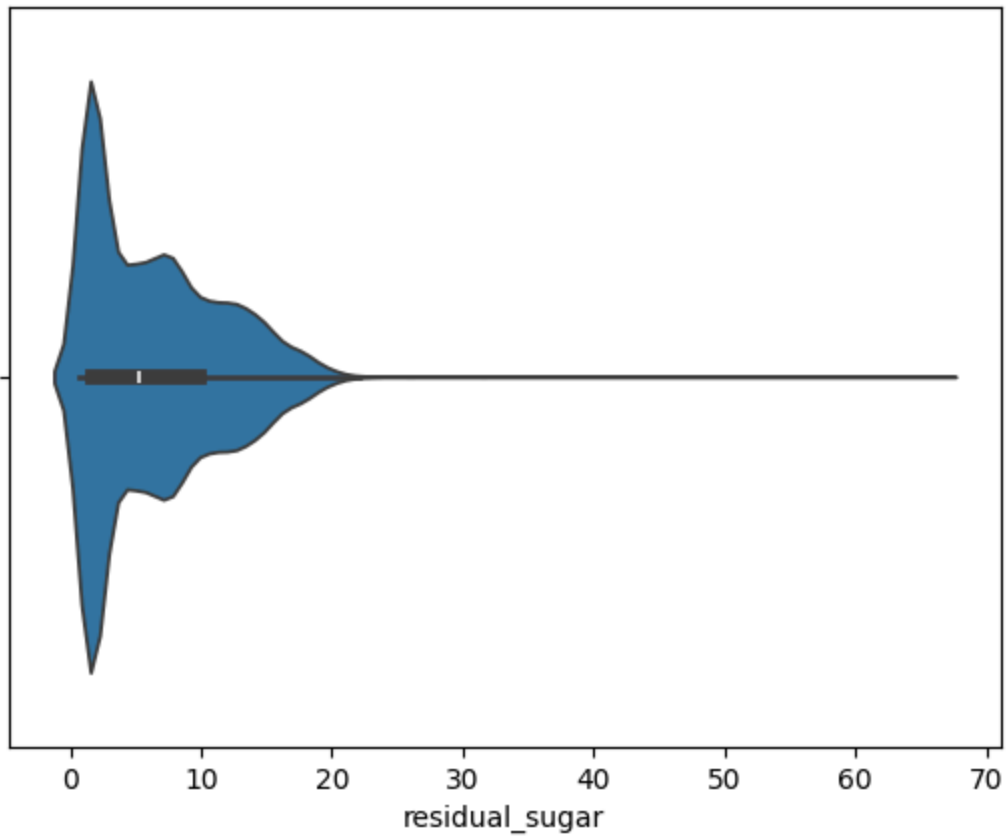




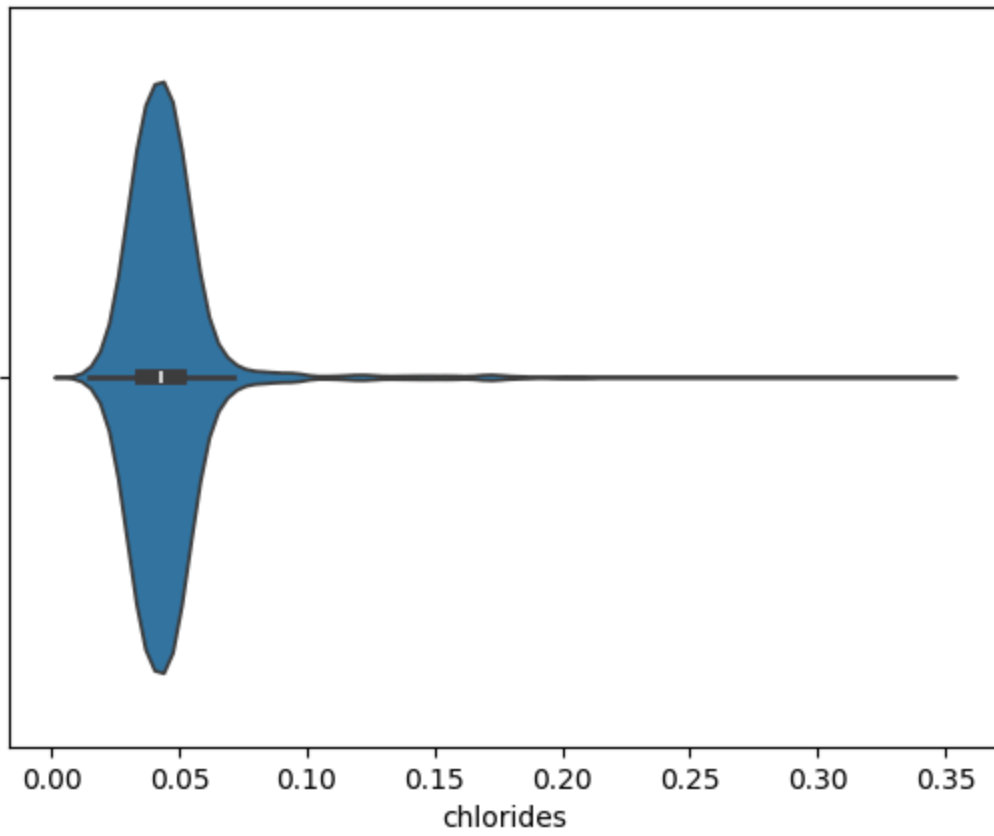
Citric Acid



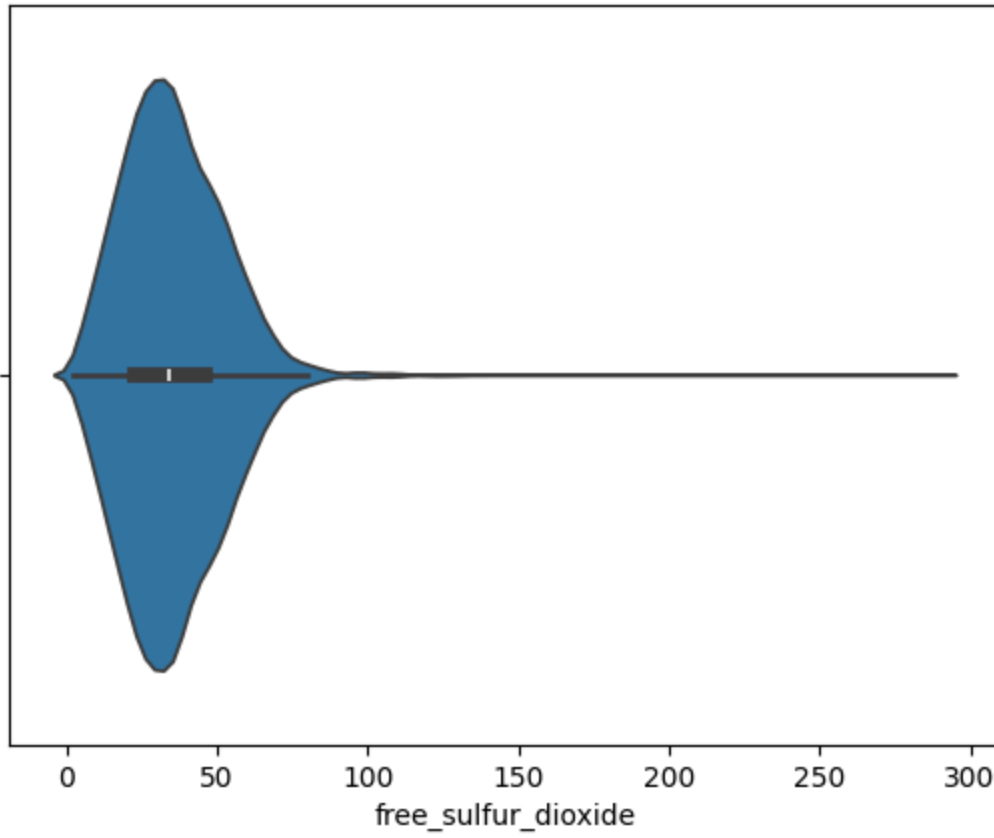
Residual Sugar



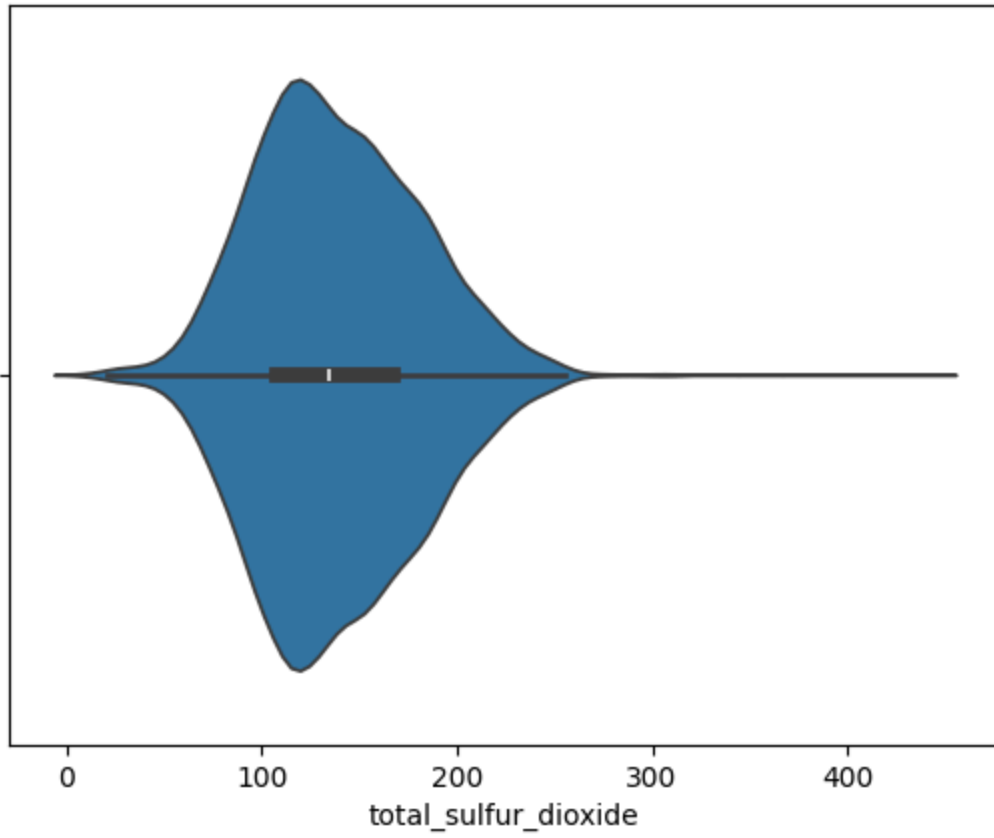
Chlorides

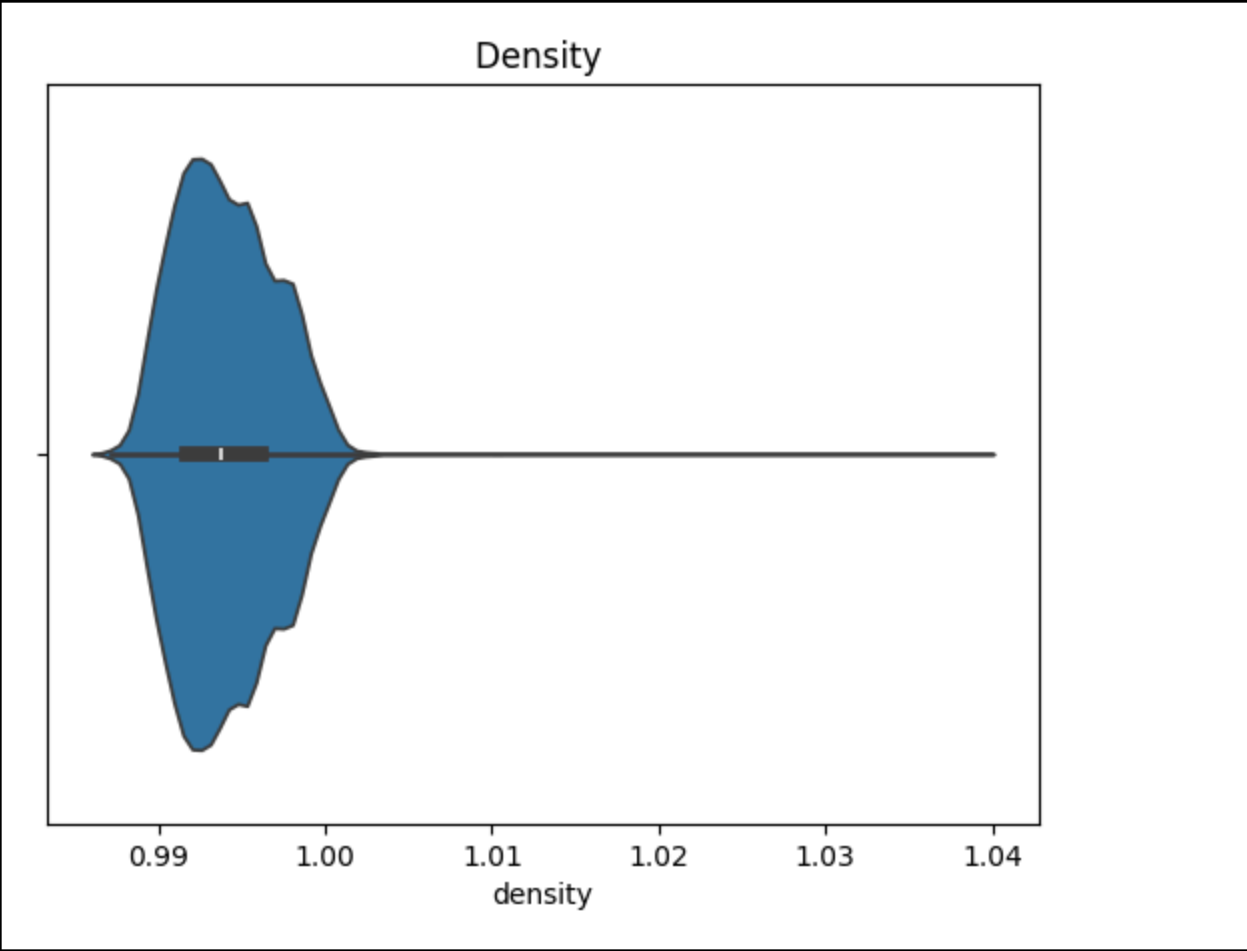


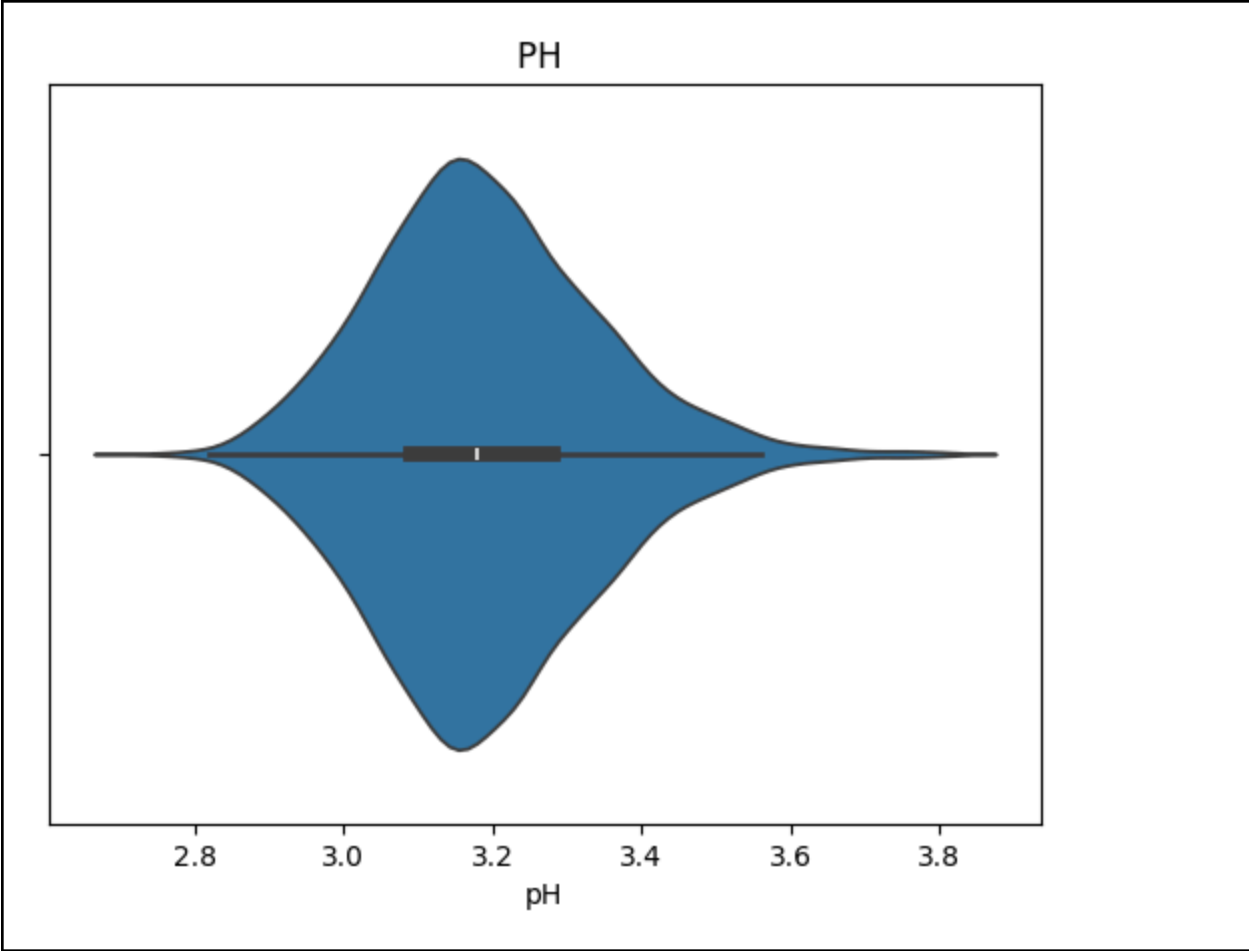
Free Sulfur Dioxide

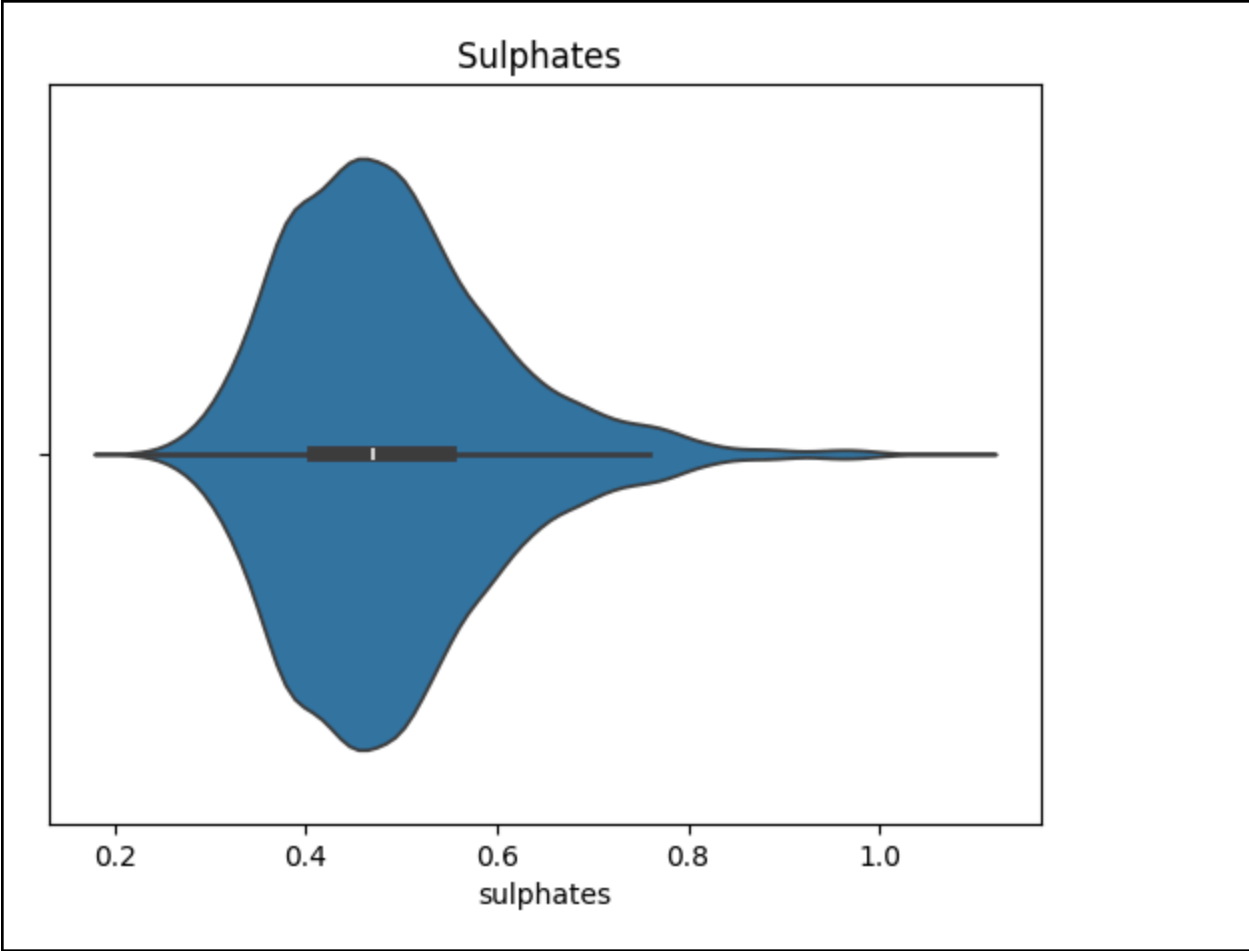


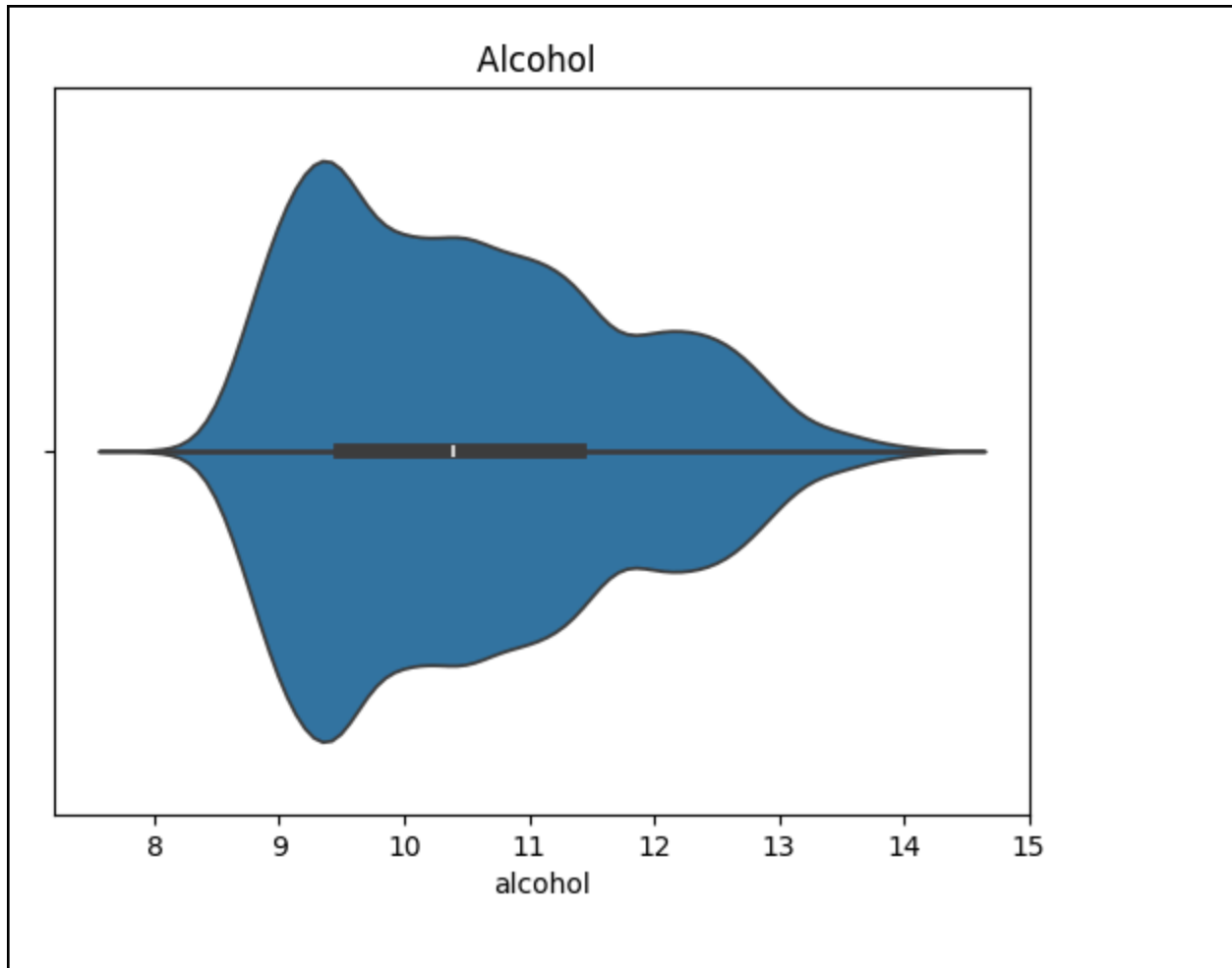
Total Sulfur Dioxide











5) Single Input Variable Linear Regression Model

- a) *Intercept* 2.582009399160653
- b) *Coefficient* 0.3134693019133235
- c) *RMSE* 0.7971284628752668
- d) R^2 0.18972533274950443

6) Multiple Input Variable Linear Regression Model

- a) *Intercept* -36.498611308344806
- b) *Coefficient* -2.0721159353648115
- c) *RMSE* 0.7684872688163132
- d) R^2 0.24690635620080081