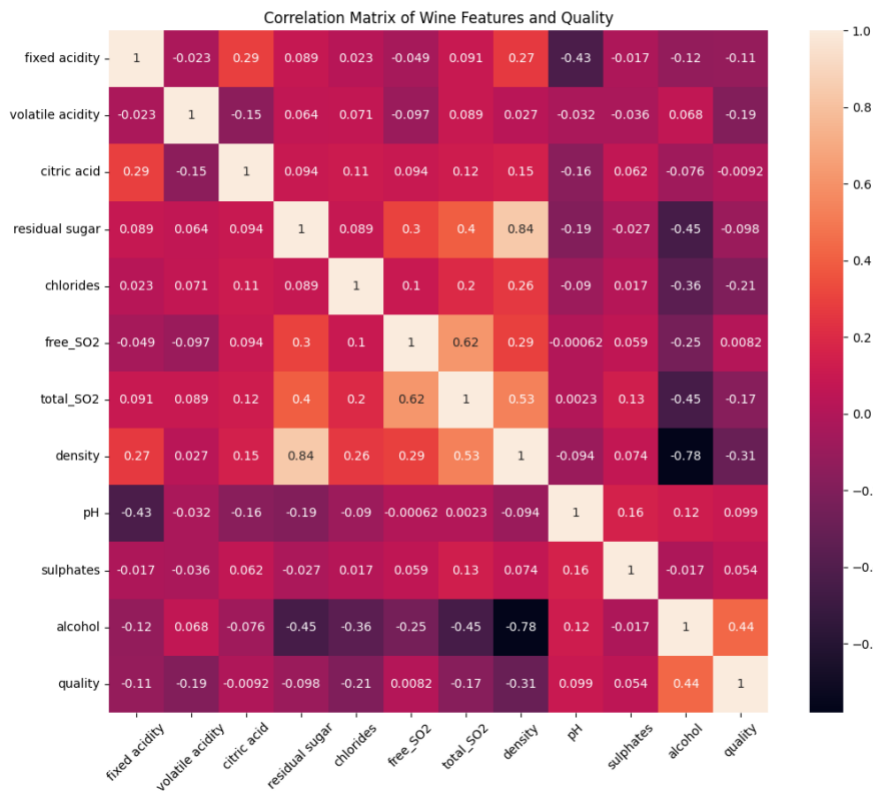


<https://github.com/Rucha1796/CS131/tree/master/a4>

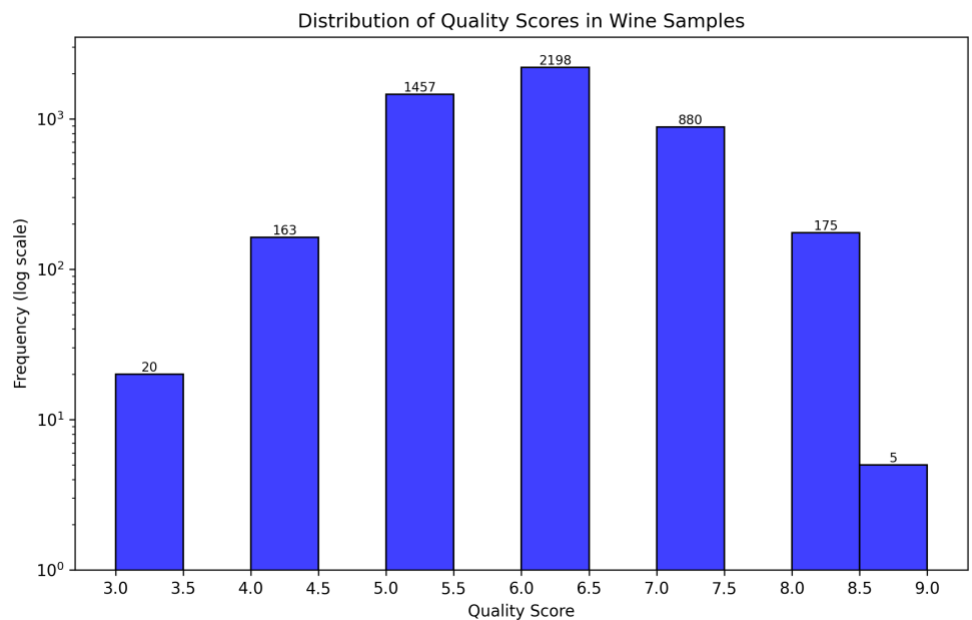
Task 1: For each feature and the target variables, obtain the mean, standard deviation, min, max, and 25/50/75% percentiles. (Show a table.)

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free_SO2	total_SO2	density	pH	sulphates	alcohol	quality
count	4898.0	4898.0	4898.0	4898.0	4898.0	4898.0	4898.0	4898.0	4898.0	4898.0	4898.0	4898.0
mean	6.85	0.28	0.33	6.39	0.05	35.31	138.36	0.99	3.19	0.49	10.51	5.88
std	0.84	0.1	0.12	5.07	0.02	17.01	42.5	0.0	0.15	0.11	1.23	0.89
min	3.8	0.08	0.0	0.6	0.01	2.0	9.0	0.99	2.72	0.22	8.0	3.0
25%	6.3	0.21	0.27	1.7	0.04	23.0	108.0	0.99	3.09	0.41	9.5	5.0
50%	6.8	0.26	0.32	5.2	0.04	34.0	134.0	0.99	3.18	0.47	10.4	6.0
75%	7.3	0.32	0.39	9.9	0.05	46.0	167.0	1.0	3.28	0.55	11.4	6.0
max	14.2	1.1	1.66	65.8	0.35	289.0	440.0	1.04	3.82	1.08	14.2	9.0

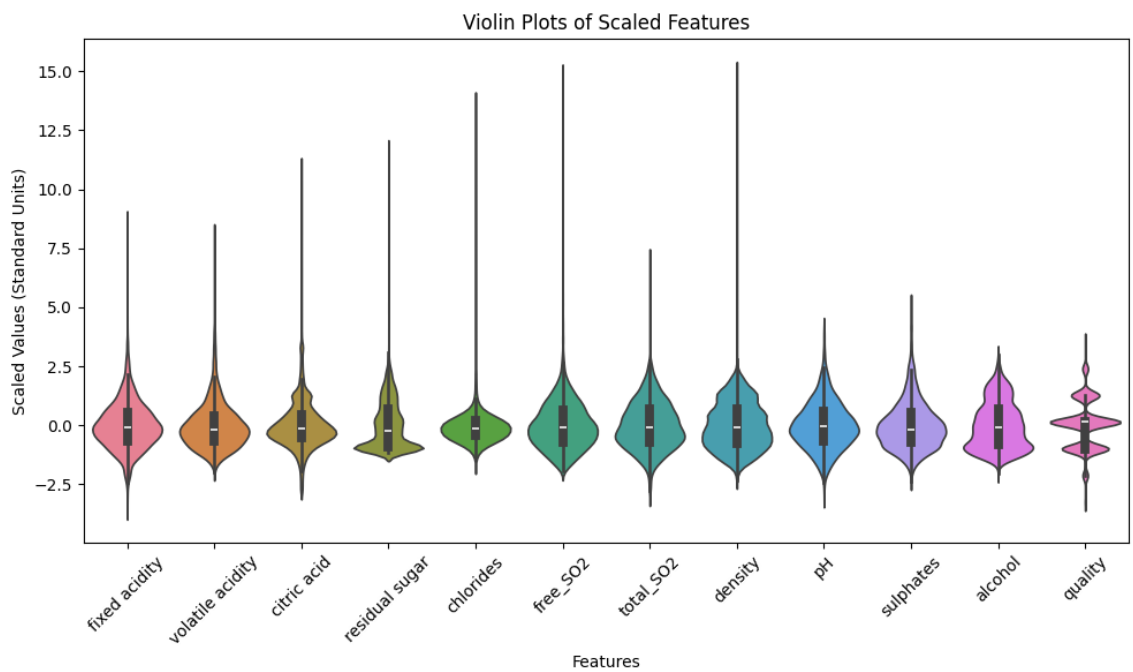
Task 2. Illustrate the correlation matrix of all features and the target



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



Task 4. Illustrate the violin plots of all features. Make sure that each plot is visible.



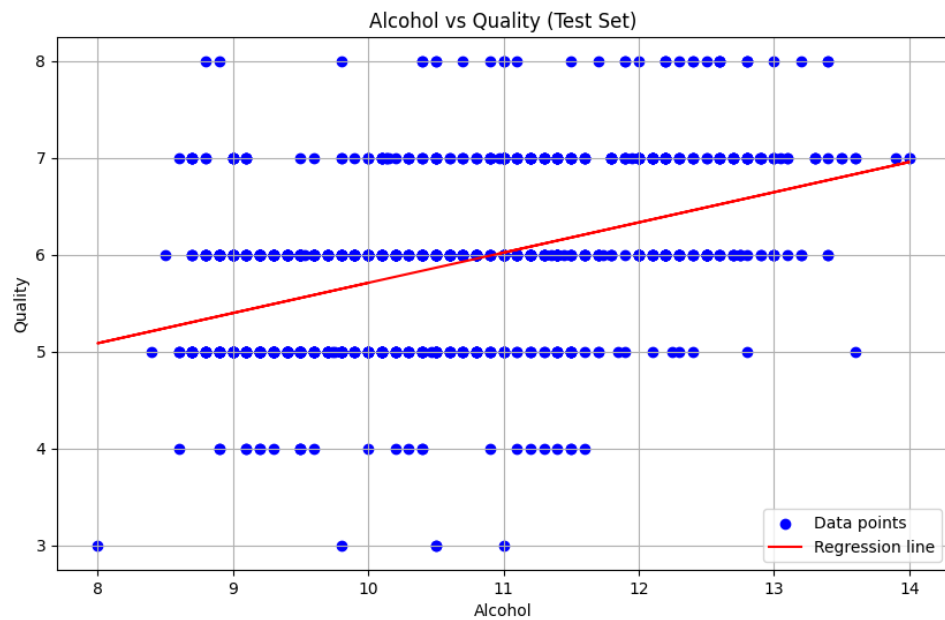
Task 5. Train a linear regression model with a single input variable (alcohol). Plot all data points ($x=\text{alcohol}$; $y=\text{quality}$) and the obtained regression model. Show the intercept, coefficient, RMSE, and R^2 .

$b_0 = 2.594634883462978$

$b_1 = 0.3118068153786768$

RMSE: 0.7859719307331404

$R^2: 0.20235854927544705$



Task 6. Train a linear regression model with input variables (volatile_acidity, alcohol, and density). Show the intercept, coefficient, RMSE, and R^2 .

Intercept (b_0): -37.49257502316353

Coefficients (b_1): {'volatile acidity': -2.128325298480065, 'alcohol': 0.39908482916666943, 'density': 40.00172502326192}

RMSE: 0.7652181198384326

R^2 : 0.24392629832553359

Linear regression Target(Quality) vs Input Variable (test set)

