

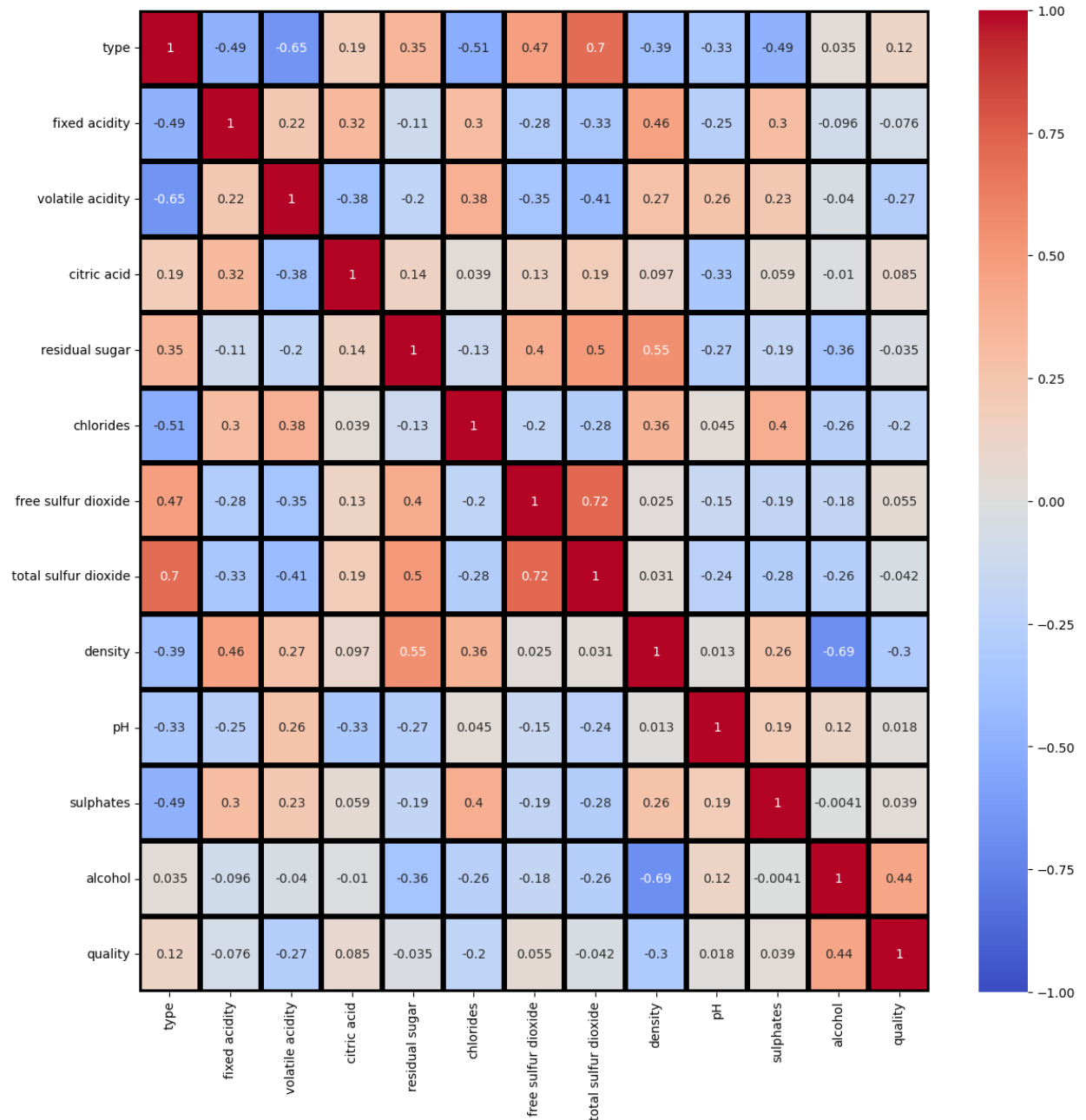
Отчет

Загружен файл с данными. Проведен анализ.

В данных есть пропуски, но их доля – 0,5% от всех данных.

Для анализа строки с пропусками можно удалить

В данных наблюдается корреляция по признакам, что видно по корреляционной матрице:



Произведена выборка тренировочных данных и тестовых из соотношения 80/20
Данные обработаны на разных моделях (Логическая регрессия, линейная, дерево, классификация “Случайный лес”). Результаты внесены в таблицу:

	model	train_score	test_score	comment
0	LogisticRegression	0.471760	0.457850	
1	LinearRegression	0.284295	0.346475	
2	DecisionTreeRegressor(max_depth=6)	0.399798	0.332253	
3	RandomForestClassifier(random_state=4)	1.000000	0.657386	
4	RandomForestClassifier(random_state=4, max_dep...	1.000000	0.666667	
5	RandomForestClassifier(random_state=4, max_dep...	0.815667	0.621810	

Наилучший результат получен на модели RandomForestClassifier с параметрами: random_state=4, max_depth = 20

Конечно, над улучшением результата можно еще работать:

- можно провести более глубокий анализ признаков, их корреляцию
- для регрессии попробовать нормализовать данные
- произвести вариацию переменных моделей
- попробовать построить модель на нейросетях

и т.д.