Refactor: Wine Quality Analysis

In this exercise, you'll refactor code that analyzes a wine quality dataset taken from the UCI Machine Learning Repository here (https://archive.ics.uci.edu/ml/datasets/wine+quality). Each row contains data on a wine sample, including several physicochemical properties gathered from tests, as well as a quality rating evaluated by wine experts.

The code in this notebook first renames the columns of the dataset and then calculates some statistics on how some features may be related to quality ratings. Can you refactor this code to make it more clean and modular?

In [21]:
```
import pandas as pd
df = pd.read_csv('winequality-red.csv', sep=';')
df.head()
```

Out[21]:

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

Renaming Columns

You want to replace the spaces in the column labels with underscores to be able to reference columns with dot notation. Here's one way you could've done it.

In [22]:
```
new_df = df.rename(columns={'fixed acidity': 'fixed_acidity',
                            'volatile acidity': 'volatile_acidity',
                            'citric acid': 'citric_acid',
                            'residual sugar': 'residual_sugar',
                            'free sulfur dioxide': 'free_sulfur_dioxide',
                            'total sulfur dioxide': 'total_sulfur_dioxide'
                           })
new_df.head()
```

Out[22]: