Data Description

🍷 Red Wine Quality Dataset

Divya Ramanujam

This dataset is all about red wine 🍷. Each row tells us the details about one sample of red wine – like how much sugar it has, how acidic it is, and how strong the alcohol content is. There's also a **quality score** given to each wine, which is our target — we want to **predict how good a wine is** using all the other information.

📋 What's in the Dataset?

There are **1599 wine samples**, and for each sample, we have:

Feature	What it Means
`fixed acidity`	Type of acid that stays in wine (affects taste and stability)
`volatile acidity`	Acid that can turn into a gas – too much gives a vinegar smell
`citric acid`	Adds freshness and flavor
`residual sugar`	Sugar left after making the wine
`chlorides`	Salt in the wine
`free sulfur dioxide`	Keeps wine from going bad

`total sulfur dioxide`	Total amount of sulfur preservative
`density`	Thickness of the wine – related to sugar and alcohol
`pH`	How acidic or basic the wine is
`sulphates`	Helps protect the wine – acts like a preservative
`alcohol`	Alcohol percentage
`quality`	A score from 0 to 10 given to each wine (this is what we want to predict!)

🧼 What Cleaning Did We Do?

1. **Checked for Missing Data**:
   ✅ There were no missing values – great start!
2. **Looked at Each Feature (Univariate Analysis)**:
   - We checked how each column looked (its range, average, etc.).
3. **Fixed Outliers (Extreme Values)**:
   - For columns with weird or very large values, we fixed them using a method called IQR.
4. **Reduced Skewed Data**:
   - Some columns like `sugar`, `chlorides`, and `sulphates` were not evenly spread out.
   - We used a simple math trick called `log transform` to make them more normal.

🎯 What Are We Trying to Predict?

We are focusing on the `quality` column.

- It's a number between **0 and 10**, showing how good a wine is.
- Later on, we can turn this into:
    - "Good wine" (quality ≥ 6)
    - "Not-so-good wine" (quality < 6)

This makes it a **classification problem** (e.g., good vs not good wine).

📈 What Did We Learn from the Data?

- Wines with **more alcohol** usually get higher ratings 🍷.
- **Too much volatile acidity** (vinegar smell) lowers the rating.
- **Sulphates** and **citric acid** also help improve wine quality.

We showed this using a colorful **heatmap** that compares all the features.