

Problem Datasets:

Dataset Overview: The datasets contain information about the chemical properties of red and white wines, along with a quality rating based on sensory testing. Your goal is to use **data visualization** techniques (Matplotlib & Seaborn) to explore and present insights about the relationship between chemical features and the quality of wine.

You are tasked with generating several visualizations to identify trends, correlations, and key insights in the data. Use appropriate visualizations like histograms, boxplots, heatmaps, and scatter plots to better understand the data.

Columns in the Dataset:

- `fixed_acidity`: The amount of fixed acidity in the wine.
- `volatile_acidity`: The amount of volatile acidity in the wine.
- `citric_acid`: The amount of citric acid in the wine.
- `residual_sugar`: The amount of residual sugar in the wine.
- `chlorides`: The amount of chlorides in the wine.
- `free_sulfur_dioxide`: The amount of free sulfur dioxide in the wine.
- `total_sulfur_dioxide`: The amount of total sulfur dioxide in the wine.
- `density`: The density of the wine.
- `ph`: pH level of the wine.
- `sulphates`: The amount of sulphates in the wine.
- `alcohol`: Alcohol content of the wine.
- `quality`: Quality rating of the wine (integer from 0 to 10).

Additional Dataset Links:

- [Red Wine Quality Dataset \(CSV\) and White Wine Quality Dataset \(CSV\)](#)

Summary:

The tasks above focus on using Matplotlib and Seaborn to explore, analyze, and visualize the relationships between various chemical properties of wine and the quality rating. These visualizations will help you uncover trends, correlations, and insights into what influences wine quality. After completing the tasks, you should be able to identify key features and their impact on wine quality, and present your findings with clear visual evidence.

NOTE:USE ONLY RED WINE DATASET

Task 1: Basic Distribution of Numerical Variables

- **Goal:** Visualize the distribution of key numerical features.
- **Task:**
 - Use histograms to display the distribution of the following numerical features: `fixed_acidity`, `volatile_acidity`, `residual_sugar`, `chlorides`, `free_sulfur_dioxide`, `total_sulfur_dioxide`, `density`, `ph`, `sulphates`, and `alcohol`.
 - What patterns or trends can you observe from the distributions? Are there any features with skewed distributions or outliers?

Task 2: Correlation Heatmap

- **Goal:** Visualize relationships between numerical features.
 - **Task:**
 - Create a **correlation heatmap** to visualize the correlations between the numerical variables.
 - Highlight and discuss any strong positive or negative correlations, especially those that could influence wine quality. Which variables seem to be most closely related?
-

Task 3: Wine Quality vs Alcohol Content

- **Goal:** Explore the relationship between wine quality and alcohol content.
 - **Task:**
 - Create a **boxplot** that shows the distribution of `alcohol` across different `quality` ratings.
 - What trends or patterns do you observe? Does higher alcohol content generally correlate with higher wine quality?
-

Task 4: Quality Distribution

- **Goal:** Understand the distribution of wine quality scores.
 - **Task:**
 - Use a **countplot** to visualize how wine quality is distributed (i.e., how often each quality score appears in the dataset).
 - Which quality rating appears most frequently? Discuss the percentage of wines that have this common quality.
-

Task 5: Visualizing Distribution of Key Variables with Boxplots and Histograms

- **Goal:** Explore relationships between wine quality and several key chemical features.
 - **Task:**
 - Visualize the distribution of key variables (`alcohol`, `sulphates`, `citric_acid`, and `fixed_acidity`) with boxplots and histograms.
 - What do you observe about how wine quality is related to these features? Are there any patterns that stand out? **note:** Create a subplot showing boxplots and histograms for each of the selected features (`alcohol`, `sulphates`, `citric_acid`, and `fixed_acidity`) with a title describing the feature being visualized.
-

Task 6: Scatter Plot for Fixed Acidity vs Alcohol Content

- **Goal:** Investigate the relationship between acidity and alcohol content.
 - **Task:**
 - Create **scatter plots** to analyze the relationship between `fixed_acidity` and `alcohol`, as well as `volatile_acidity` and `alcohol`.
 - Do you observe any trends or patterns? How do these variables appear to correlate with each other?
-

Task 7: Impact of Sulfur Dioxide on Wine Quality

- **Goal:** Explore the impact of sulfur dioxide levels on wine quality.
- **Task:**
 - Create **scatter plots** to show the relationship between `free_sulfur_dioxide` and `quality`, and between `total_sulfur_dioxide` and `quality`.

- How does the presence of sulfur dioxide seem to influence the quality rating of wine?

Advanced Task :

Task 8: Boxplot for Quality vs Various Features

- **Goal:** Explore the impact of several features on wine quality.
- **Task:**
 - For each of the following features (`fixed_acidity`, `residual_sugar`, `sulphates`, `citric_acid`, `density`), create a **boxplot** showing how the feature values vary across different wine quality ratings.
 - Which features show the most variation across quality ratings? Are there clear patterns that indicate which features influence wine quality?

Conclusion Question:

1. Based on the boxplots, **what do you observe about the relationship between wine quality and the following features:**
 - Fixed Acidity
 - Residual Sugar
 - Sulphates
 - Citric Acid
 - Density
2. **Which feature seems to have the most noticeable impact on wine quality?** Explain briefly.

BY Pranjal Gajbhiye(AIE)

Happy Learning....

In []: