

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from docx import Document
from docx.shared import Inches
```

Wine Quality Analysis

1. Load the dataset

```
In [ ]: # Load the dataset
wine_data = pd.read_csv("/mnt/data/winequality-white.csv", delimiter=";")
wine_data.head()
```

2. Data Cleaning and Preparation

```
In [ ]: # Check for missing values in the dataset
missing_values = wine_data.isnull().sum()
missing_values
```

3. Data Exploration

```
In [ ]: # Set the style for seaborn
sns.set_style("whitegrid")

# Plot the distribution of wine quality scores
plt.figure(figsize=(10, 6))
sns.countplot(x='quality', data=wine_data)
plt.title("Distribution of Wine Quality Scores")
plt.xlabel("Wine Quality")
plt.ylabel("Number of Samples")
plt.show()
```

4. Feature Analysis

```
In [ ]: # Plotting boxplots for each feature against wine quality
features = wine_data.columns[:-1] # Excluding 'quality' from the features

plt.figure(figsize=(20, 15))
for i, feature in enumerate(features, 1):
    plt.subplot(3, 4, i)
    sns.boxplot(x='quality', y=feature, data=wine_data)
    plt.title(f"Distribution of {feature} by Wine Quality")
    plt.xlabel("Wine Quality")
    plt.ylabel(feature)
    plt.tight_layout()

plt.show()
```

5. Correlation Analysis

```
In [ ]: # Compute the correlation matrix
correlation_matrix = wine_data.corr()

# Extract the correlation values with the 'quality' column
quality_correlation = correlation_matrix["quality"].sort_values(ascending=False)

quality_correlation
```