About Wine

Wine is One of the most complex beverages in the world ,with a rich history and a variety of flovours that can be inflenced by numerous factors. In this notebook, we will explore a dataset of wine characteristics to uncover interesting insights and potentially build a predictive model. If you find this notebook useful, please consider upvoting it.

In [23]:
```
import warnings
warnings.filterwarnings('ignore')
```

In [24]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

In [25]:
```
df = pd.read_csv("Wine_Dataset.csv")
```

In [26]:
```
df.head()
```

Out[26]:

	Malic acid	Ashe	Alcalinity of ashe	Magnesium	Total phenols	Flavanoidse	Nonflavanoid phenols	Proanthocyanins	Color intensity	OD280	OD31	Proline	Al
0	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	
2	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	
3	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45	
4	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93	

Now,let's check for any missing values in dataset.

In [27]:
```
df.isnull().sum()
```

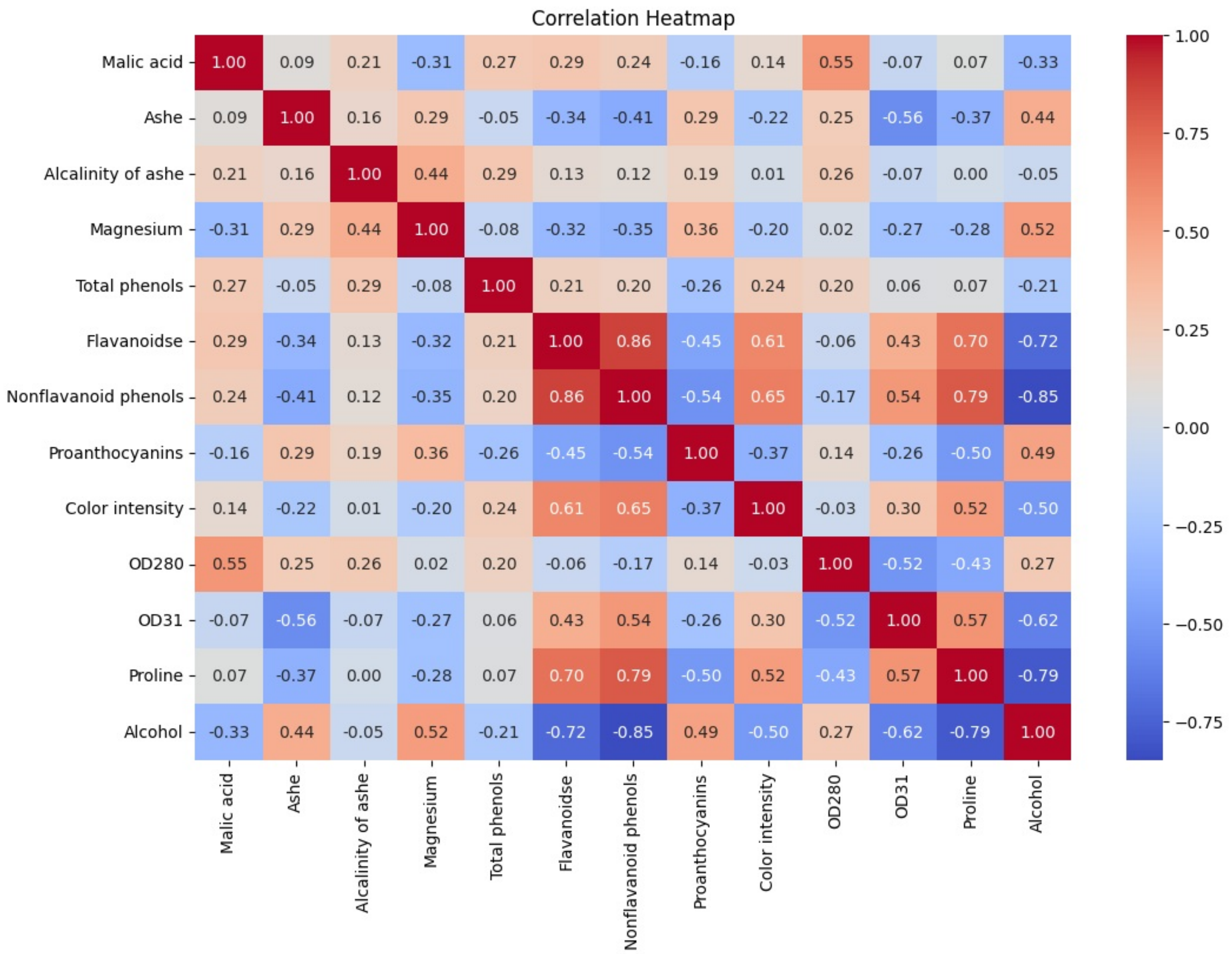
Out[27]:
```
Malic acid              0
Ashe                    0
Alcalinity of ashe      0
Magnesium               0
Total phenols           0
Flavanoidse             0
Nonflavanoid phenols    0
Proanthocyanins         0
Color intensity         0
OD280                   0
OD31                    0
Proline                 0
Alcohol                 0
dtype: int64
```

Correlation Heatmap

Let's start by visualizing the correlation batween diferent features in the dataset

In [28]:
```
# Select only numeric columns for correlation heatmap
numeric_df = df.select_dtypes(include=[np.number])

# Plot the correlation heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```



Pairplot

A pairplot can help us visualize the relationships between different features.

In [ ]:
```
# Plot pairplot
sns.pairplot(numeric_df)
plt.show()
```

Feature Distribution

Let's visualize the distribution of each feature to understand their individual charactrestics

In [ ]:
```
# Plot distribution of each feature
numeric_df.hist(bins=15, figsize=(15, 10), layout=(4, 4))
plt.tight_layout()
plt.show()
```

Building a Predictive Model

Based on the dataset, it seems plausible to predict the type of wine based on its characteristics. Let's build a Random Forest Classifier to see how well we can predict the wine type.

In [ ]:
```
# Define features and target variable
X = numeric_df.drop('Alcohol', axis=1)  # Assuming 'Alcohol' is the target variable
y = numeric_df['Alcohol']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Initialize and train the Random Forest Classifier
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model
```

```
accuracy = accuracy_score(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)
class_report = classification_report(y_test, y_pred)

accuracy, conf_matrix, class_report
```

Conclusion and Future Work

In this notebook, we explored the wine dataset, visualized the relationships between different features, and built a predictive model to classify wine types based on their characteristics. The Random Forest Classifier provided us with an accuracy score, which gives us an idea of how well the model performs.

For future analysis, we could consider:

1. Tuning the hyperparameters of the Random Forest model to improve accuracy.
2. Tuning the hyperparameters of the Random Forest model to improve accuracy.
3. Investigating the impact of each feature on the prediction using feature importance.