

```
In [ ]: #Red Wine Quality Prediction Project

In [ ]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')

In [ ]: import pandas as pd
df=pd.read_csv('winequality-red.csv')
df

In [ ]: #Exploratory Data Analysis.

In [ ]: df.shape

In [ ]: df.dtypes

In [ ]: df.columns

In [ ]: df.info()

In [ ]: df.quality.unique()

In [ ]: #Summary Statistics

In [ ]: df.describe()

In [ ]: df.quality.unique()

In [ ]: df.quality.value_counts()

In [ ]: # Data Visualizations

In [ ]: sns.heatmap(df.isnull())

In [ ]: # To check correlation

In [ ]: dfcor=df.corr()
dfcor

In [ ]: sns.heatmap(dfcor)# eexample of multivariate analysis.

In [ ]: plt.figure(figsize=(10,6))
sns.heatmap(dfcor,cmap='YlOrRd_r',annot=True)

In [ ]: # Plotting outliers
# Univariate analysis
```

```
In [ ]: df.columns

In [ ]: df['fixed acidity'].plot.box()

In [ ]: df['citric acid'].plot.box()

In [ ]: df['alcohol'].plot.box()

In [ ]: df['free sulfur dioxide'].plot.box()

In [ ]: df['total sulfur dioxide'].plot.box()

In [ ]: df.shape

In [ ]: collist=df.columns.values
ncol=12
nrows=10

In [ ]: plt.figure(figsize=(ncol,5*ncol))
for i in range(1, len(collist)):
    plt.subplot(nrows,ncol,i+1)
    sns.boxplot(df[collist[i]],color='green',orient='v')
    plt.tight_layout()

In [ ]: sns.distplot(df['density'])

In [ ]: sns.distplot(df['citric acid'])

In [ ]: #plt.figure (figure=(5*totalcol,5*totalcol))
plt.figure(figsize=(16,16))
for i in range(0,len(collist)):
    plt.subplot(nrows,ncol,i+1)
    sns.distplot(df[collist[i]])

In [ ]: # bivariate analysis
plt.scatter(df['pH'],df['quality'])

In [ ]: sns.pairplot(df)

In [ ]: plt.scatter(df['volatile acidity'],df['quality'])
plt.show()

In [ ]: # Data Cleaning

In [ ]: df.drop('volatile acidity',axis=1,inplace=True)

In [ ]: df.head()

In [ ]: df.shape

In [ ]: #Removing the outliers
```

```
In [ ]: from scipy.stats import zscore  
z=np.abs(zscore(df))  
z
```

```
In [1]: threshold=3  
print(np.where(z>3))
```

```
-----  
NameError                                Traceback (most recent call last)  
Cell In[1], line 2  
      1 threshold=3  
----> 2 print(np.where(z>3))  
  
NameError: name 'np' is not defined
```