# 

# **AD3301-DATA EXPLORATION AND VISUALIZATION**

# **MINI PROJECT**

# **TIME SERIES ANALYSIS FOR UNIVARIATE,BIVATIATE,MULTIVARIATE VARIABLE**

# Time series analysis for univariate, bivariate and multivariate variable

# UNIVARIATE ANALYSIS:

# importing dataset

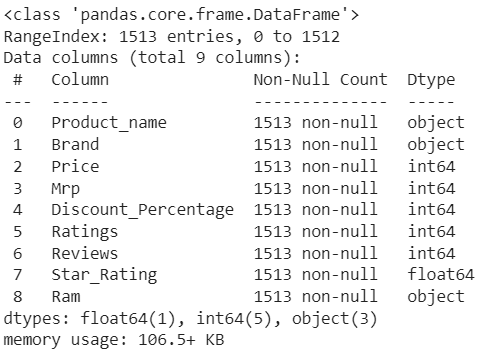
import pandas as pd

import numpy as np

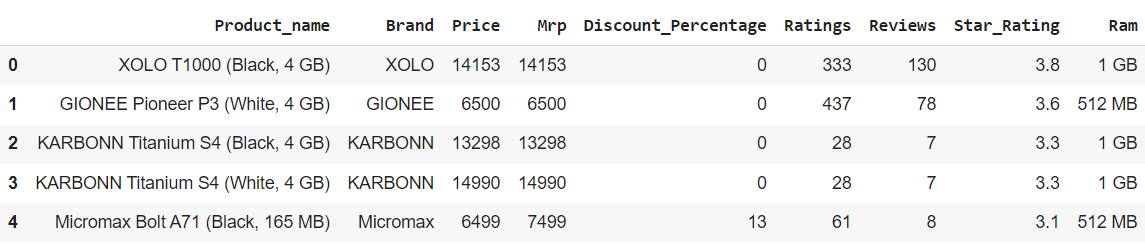
import matplotlib.pyplot as plt

df=pd.read\_csv('/content/smartphone.csv')

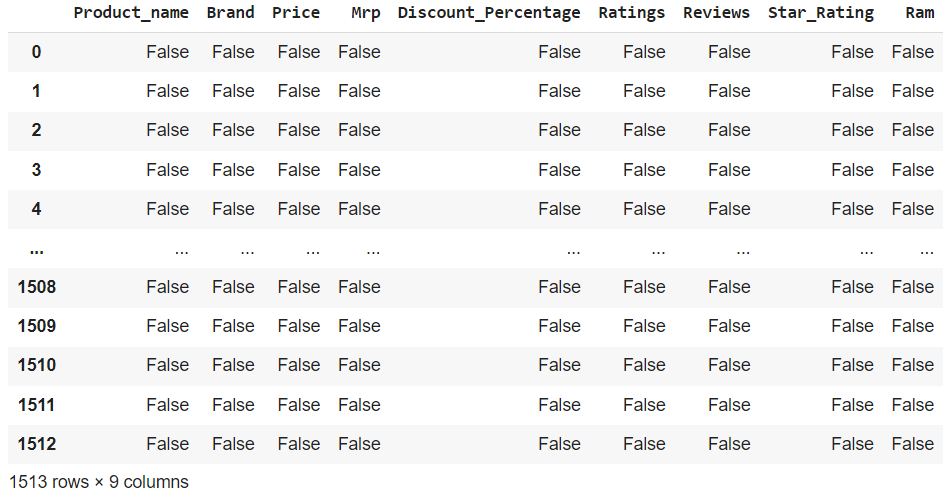
df.info()



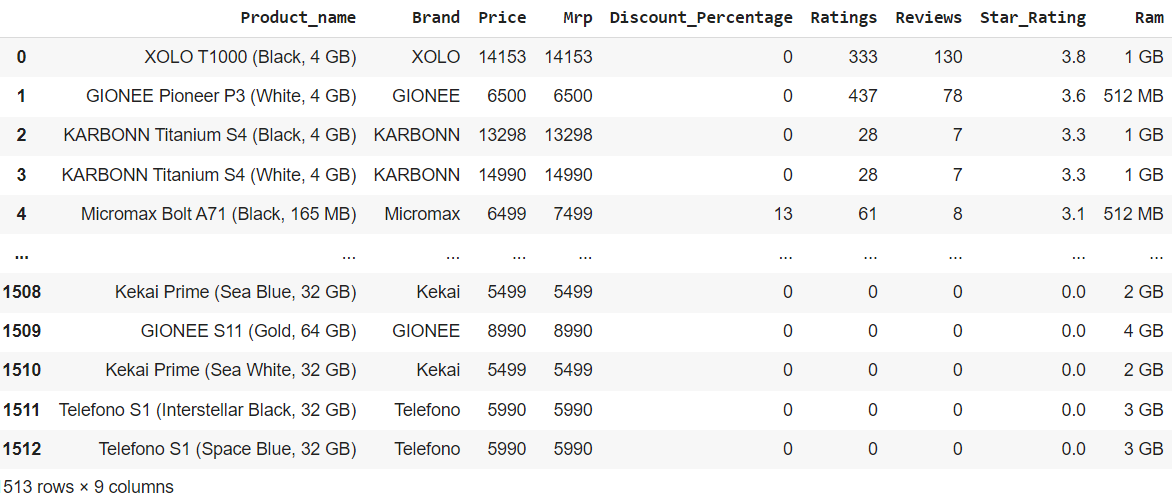
df.head(5)



df.isnull()



df.dropna()

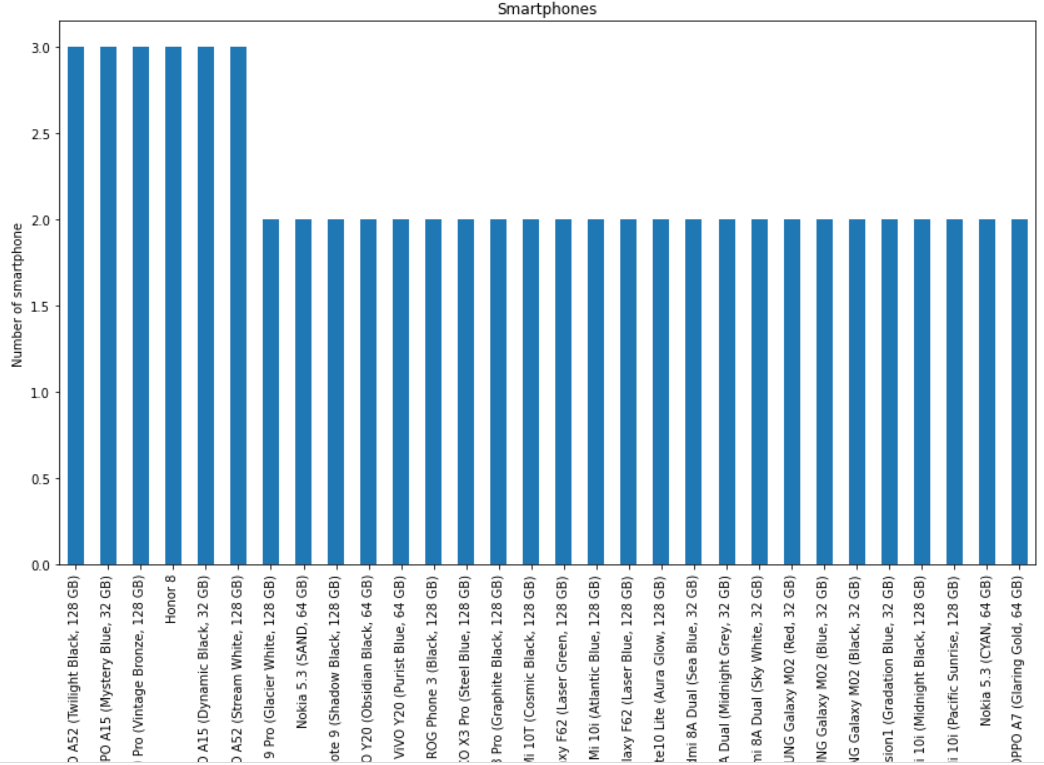


df.Product\_name.value\_counts().nlargest(30).plot(kind='bar', figsize=(14,8))

plt.title("Smartphones")

plt.ylabel('Number of smartphone')

plt.xlabel('Name of the smartphone')



INSIGHT:

Oppo, Redmi and Honor are the most used smartphones.

# DESCRIPTIVE STATISTICS

#standard variance of dataset using std() function

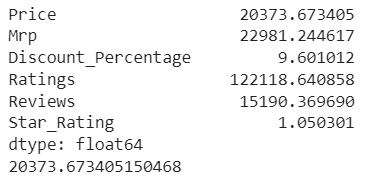
std\_dev =df.std()

print(std\_dev)

# standard variance of the specific column

sv\_Price=df.loc[:,"Price"].std()

print(sv\_Price)



# variance of dataset using var() function

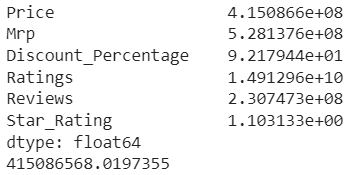
variance=df.var()

print(variance)

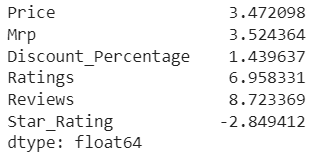
# variance of the specific column

var\_Price=df.loc[:,"Price"].var()

print(var\_Price)



df.skew()

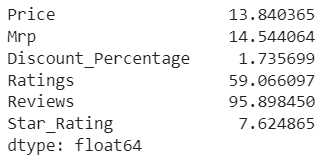


INSIGHT:

Star\_Rating has negative skew

kurtosis =df.kurt()

print(kurtosis)



INSIGHT:

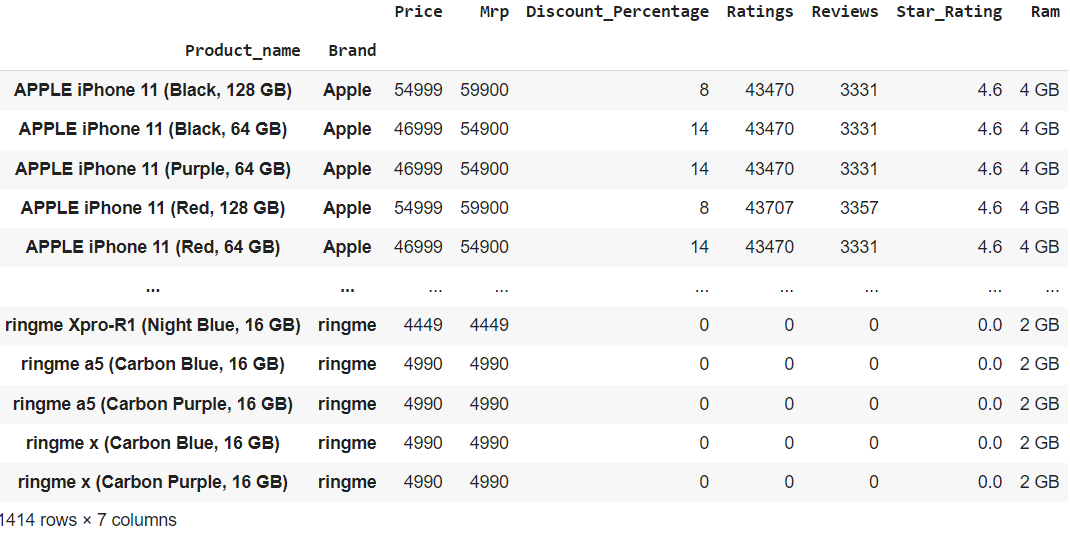
Platykurtic-Discount\_percentage

All others have leptokurtic.

# GROUPING

double\_grouping = df.groupby(["Product\_name","Brand"])

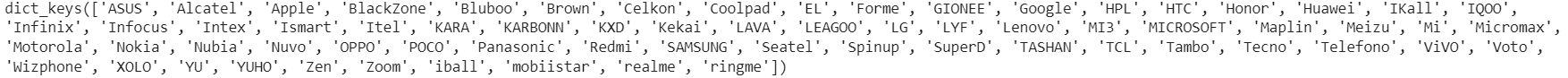
double\_grouping.first()



INSIGHT:

All the numerical datas are grouped together.

df.groupby('Brand').groups.keys()

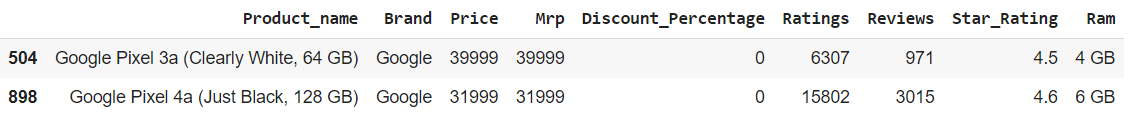


INSIGHT:

All the brands of smartphones are listed as dictionary

style = df.groupby('Brand')

style.get\_group("Google")

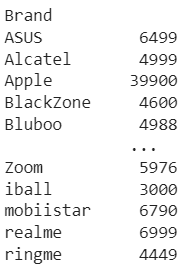


INSIGHT:

All the products of google along with their price, mrp, discount, reviews, star ratings and ram are shown.

style['Mrp'].max()

style['Mrp'].min()



INSIGHT:

ASUS has the highest market value and ringme has the least.

style.mean()



INSIGHT:

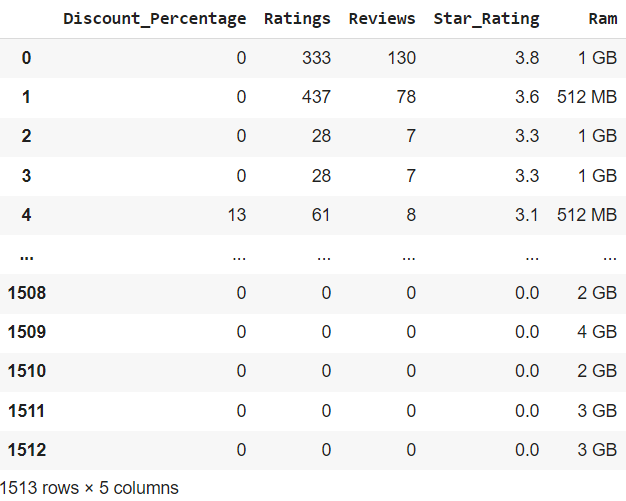
Mean of numerical column in each group is shown.

# Data aggregation

# new dataframe that consist length,width,height,curb-weight and price

new\_dataset = df.filter(["Discount\_Percentage","Ratings","Reviews","Star\_Rating","Ram"],axis=1)

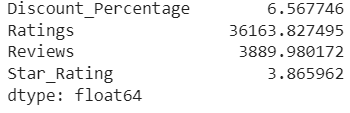
new\_dataset



INSIGHT:

Numerical variables are grouped together.

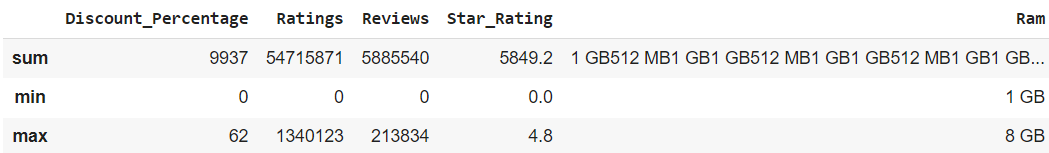
new\_dataset.agg("mean", axis="rows")



INSIGHT:

Average star ratings given is 3.865962

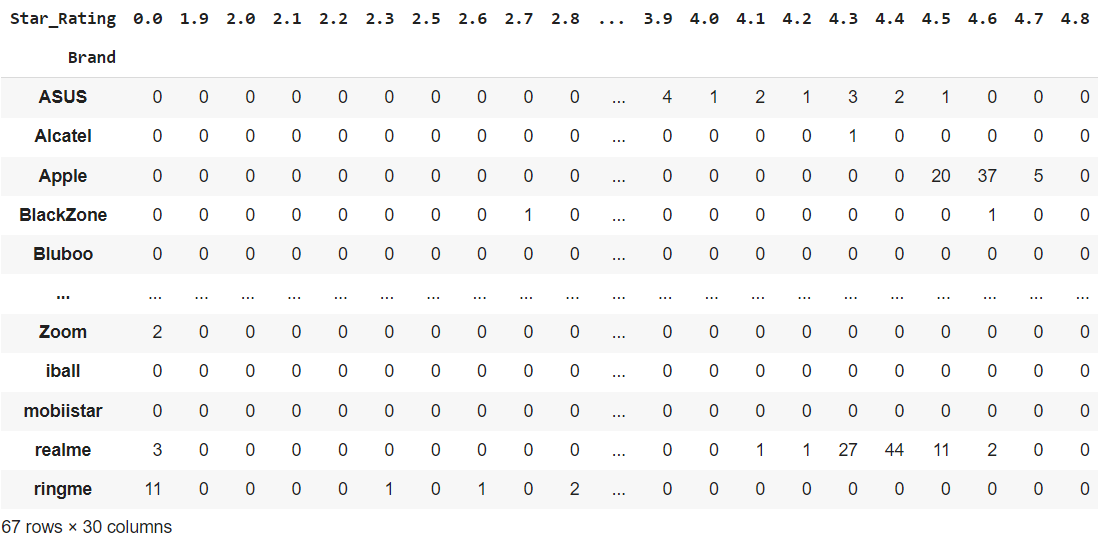
new\_dataset.agg(['sum', 'min','max'])



INSIGHT:

Minimum RAM speed of smartphone is 1GB and maximum is 8GB.

pd.crosstab(df["Brand"], df["Star\_Rating"])



INSIGHT:

ASUS is having more than 4 star rating which could be seen from crosstable.

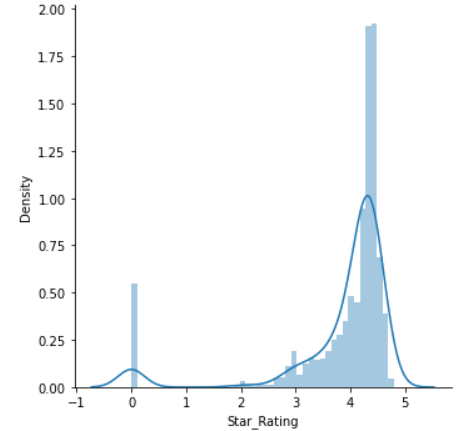
# CORRELATION

## UNIVARIATE ANALYSIS

#distribution plot

import seaborn as sns

sns.FacetGrid(df,size=5).map(sns.distplot,"Star\_Rating").add\_legend()



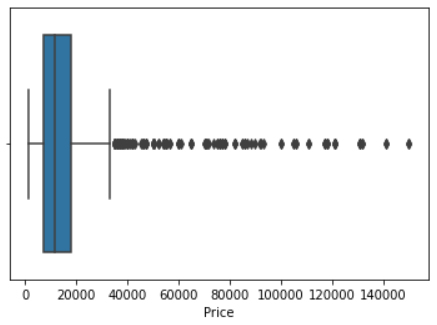
INSIGHT:

Maximum star rating is around 4.5.

#boxplot for albedo

sns.boxplot(x="Price",data=df)

plt.show()

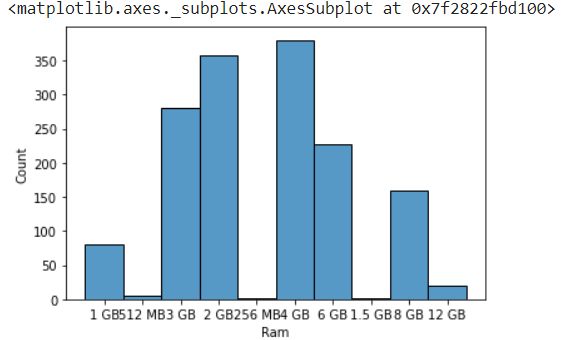


INSIGHT:

Maximum smartphones have the price of 20000.

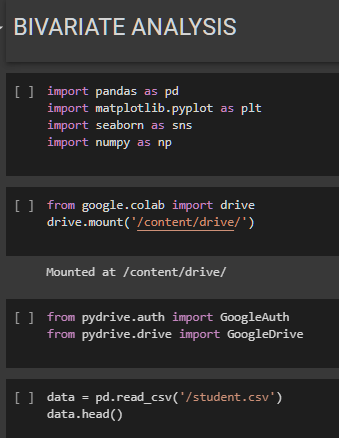
import seaborn as sns

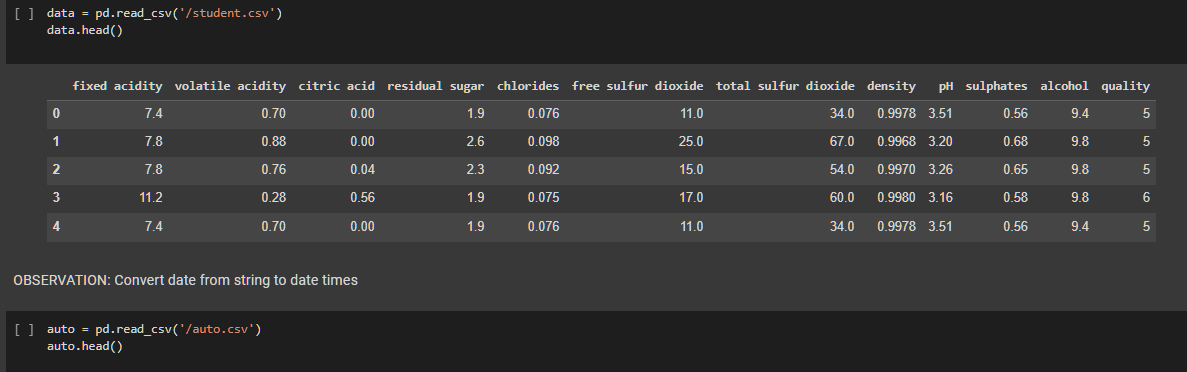
sns.histplot(x='Ram', data=df)

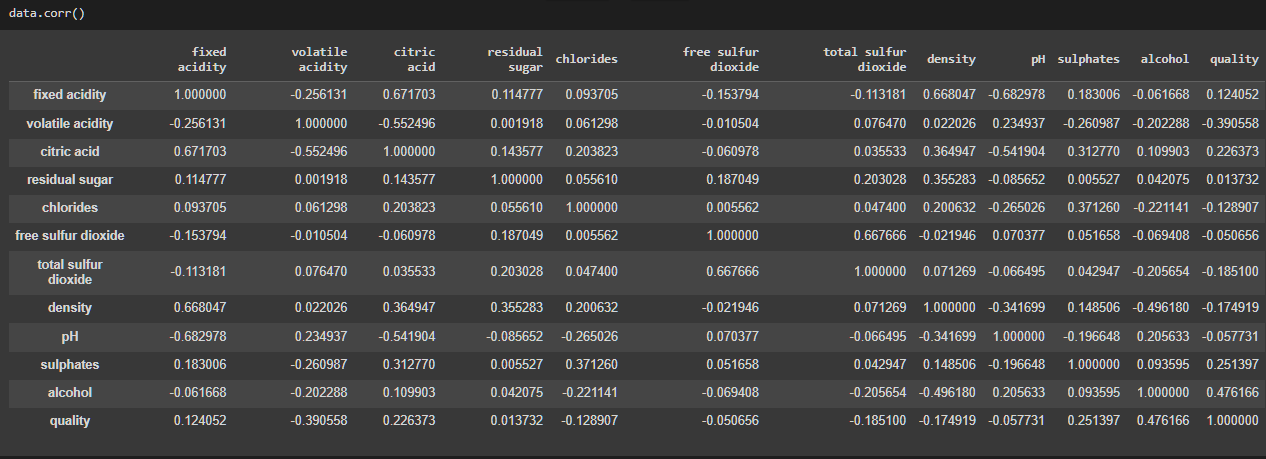


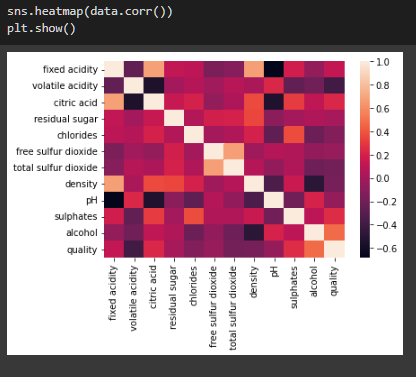
INSIGHT:

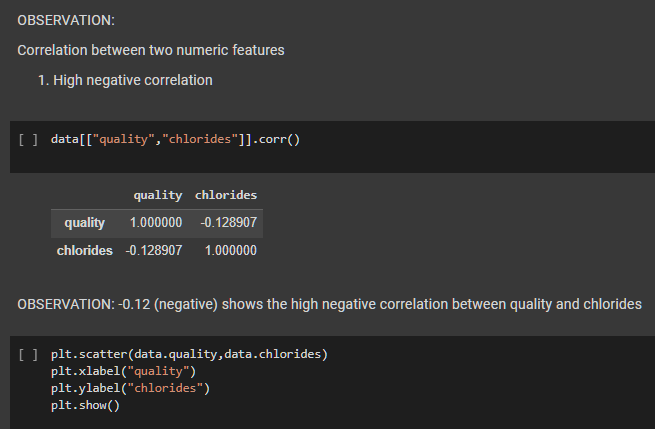
Maximum smartphones have the RAM speed of 4GB.

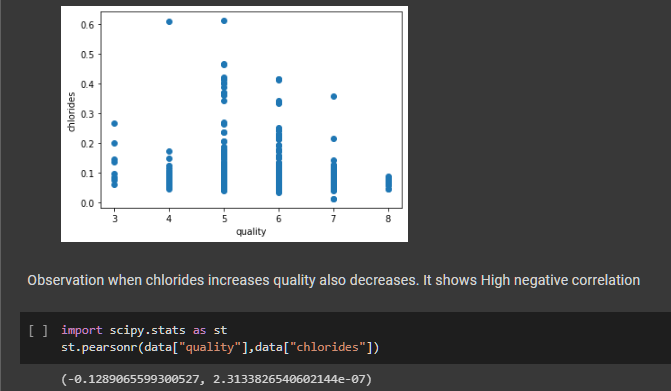


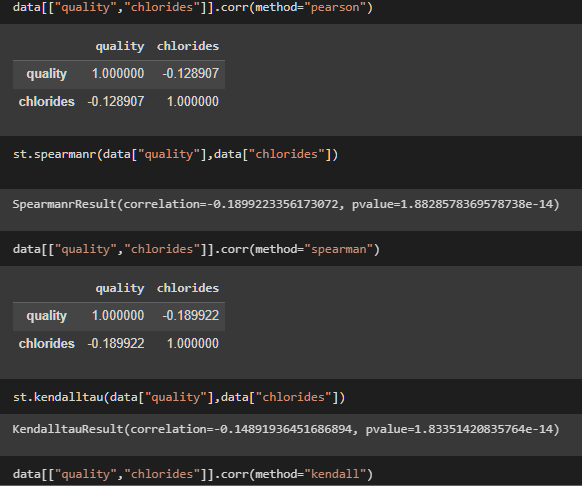


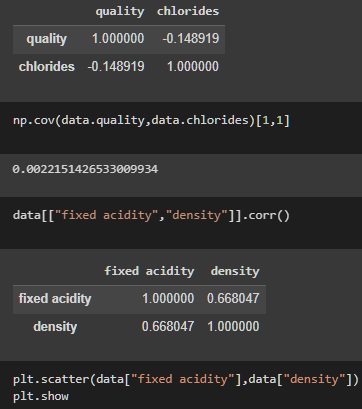


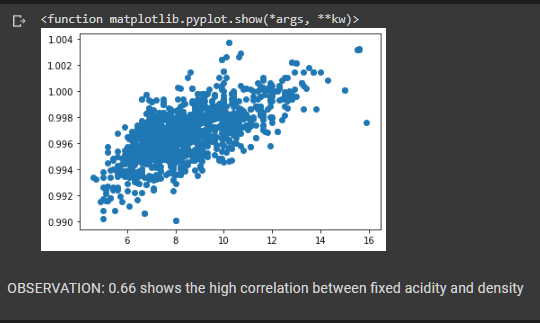


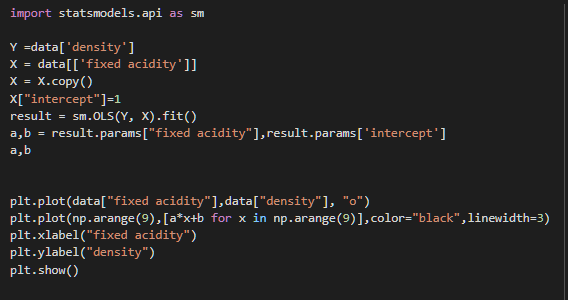


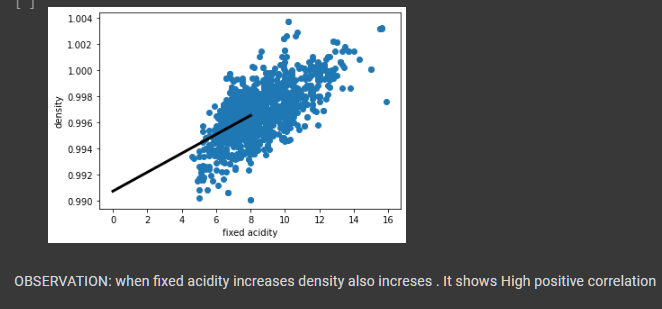


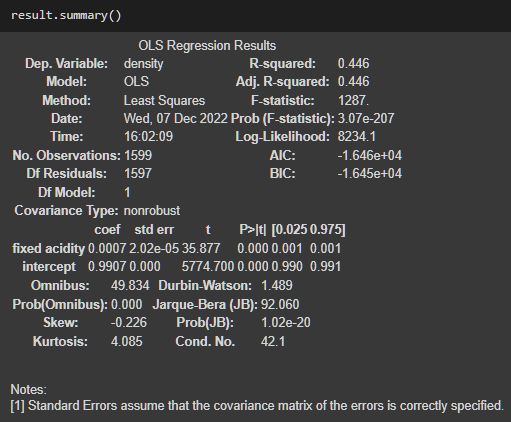


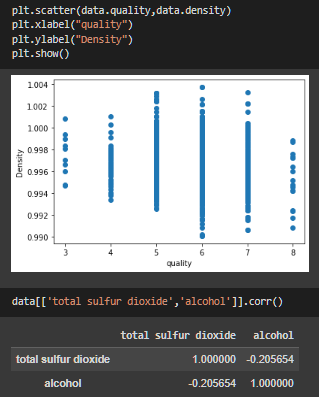


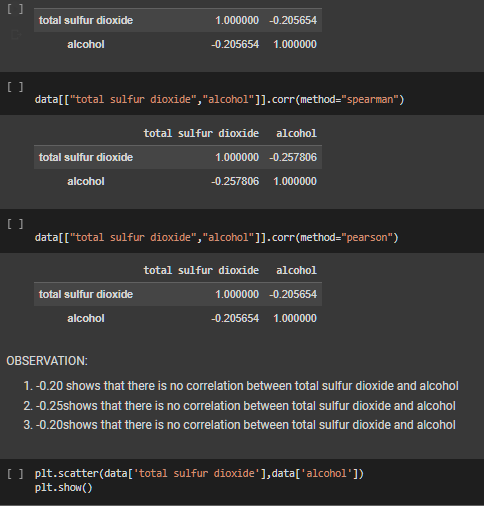


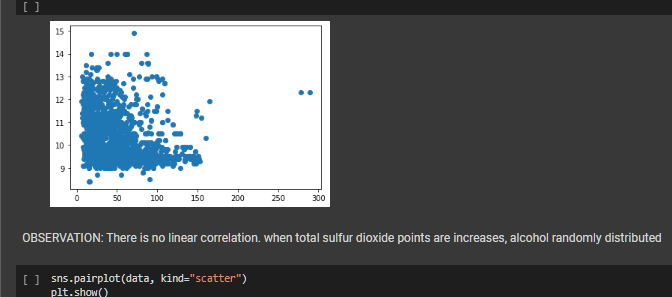


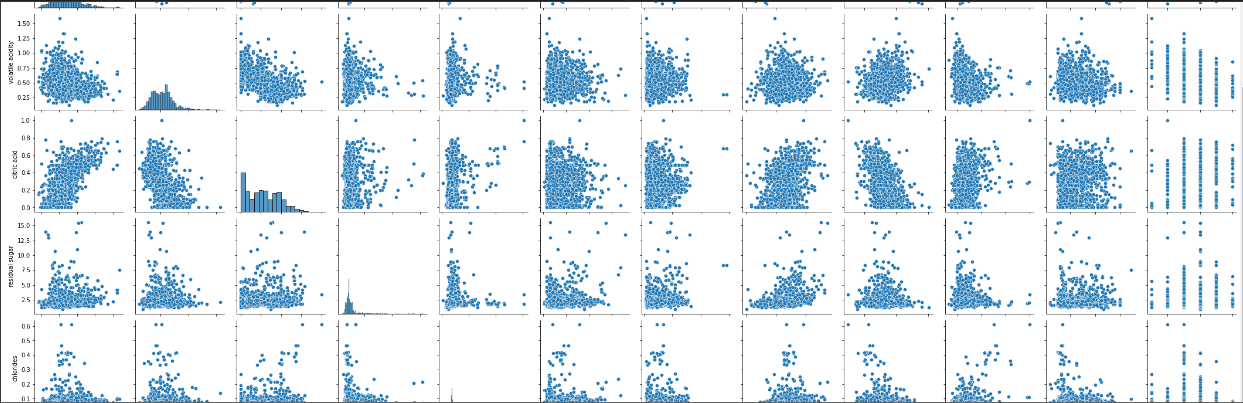








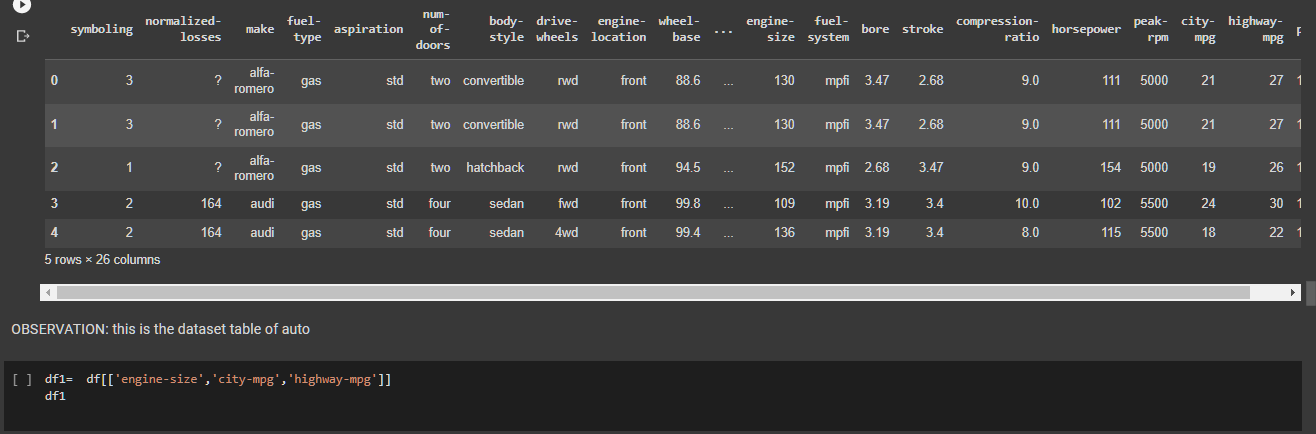


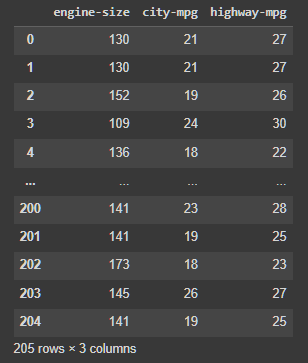


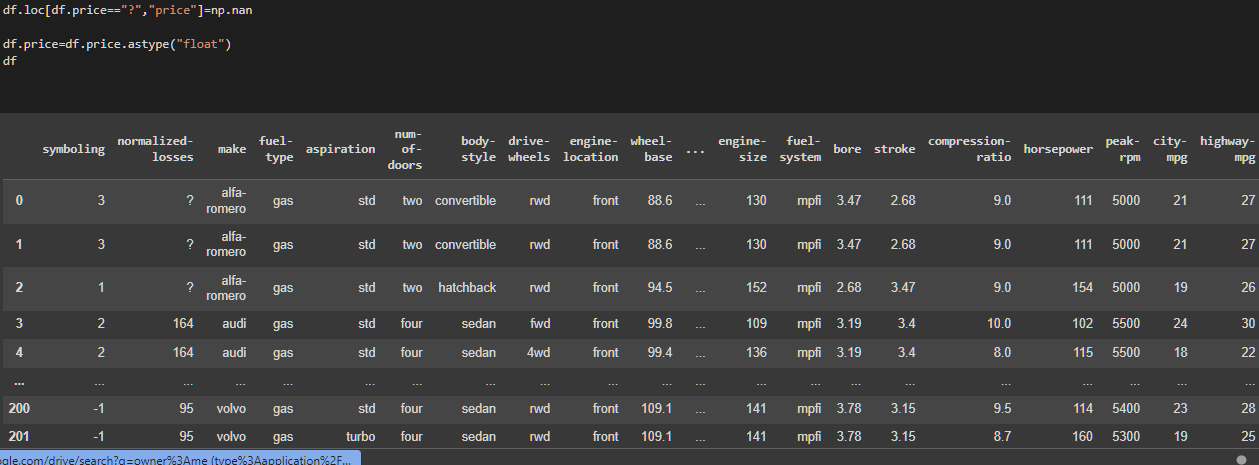


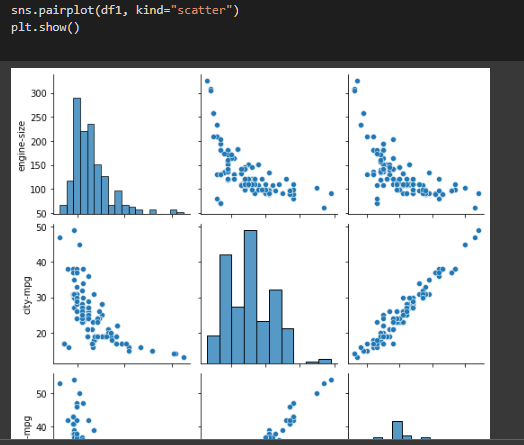
OBSERVATION: this is the scatter plot for my student dataset

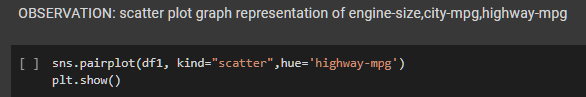


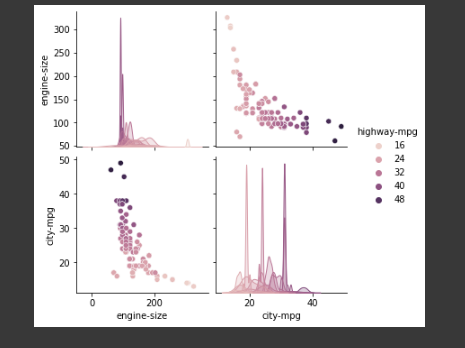












OBSERVATION: this is the scatter plot graph representation of engine-size and city-mpg then the colour represent the difference between highway-mpg

