EXERCISE 3

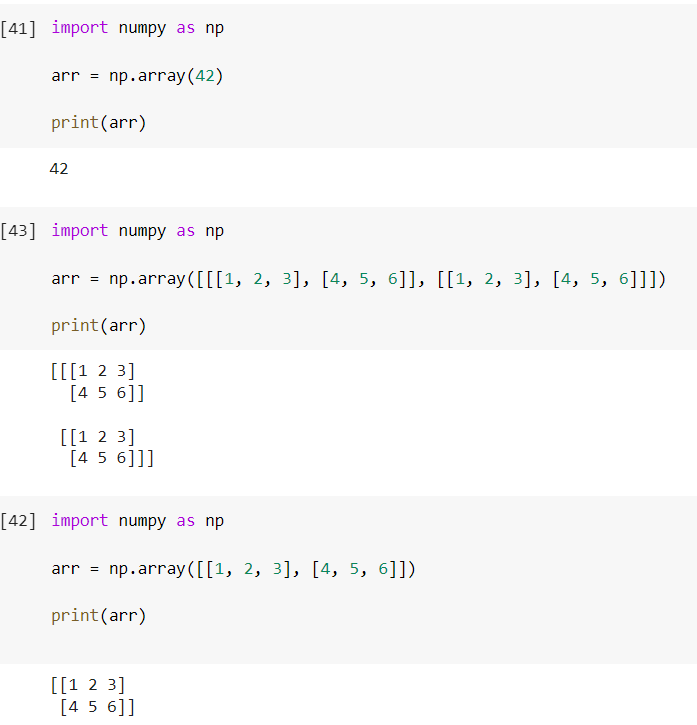
Colab link

https://colab.research.google.com/drive/1yBn\_5jrzxLEApsW4zlflfHtQa1HZGqfx?usp=sharing

Aim:

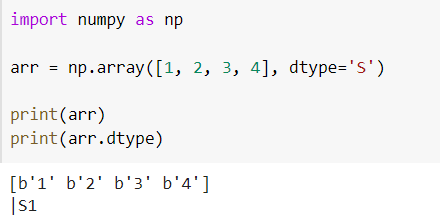
To work with numpy arrays, pandas dataframe and basic plots using matplotlib.

# Numpy in a Array









import pandas as pd

# Uploading data set

from google.colab import files

uploaded = files.upload()

import pandas as pd

import io

df = pd.read\_csv(io.BytesIO(uploaded['Car.csv']))

print(df)

## My Car dataset

Data Description:

MPG - Miles per Gallon

CYL - Cylinder capacity of car

DISP - Displacement of a car

HP - Horse power

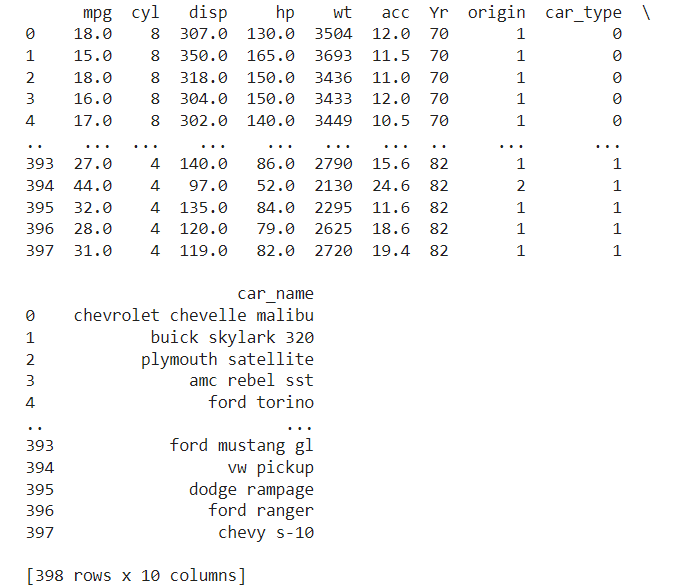
WT - Weight of a car

ACC - Acceleration of a car

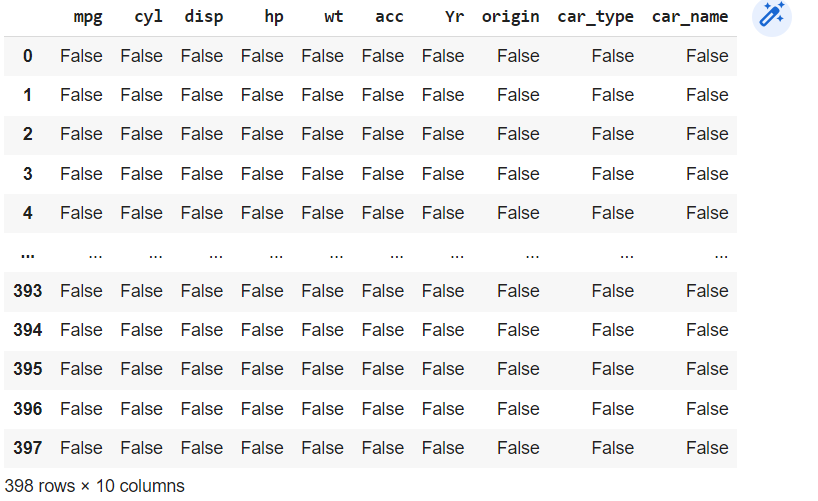
YR - Year of car manufactured

ORIGIN - 1.America 2.Asia 3.Europe

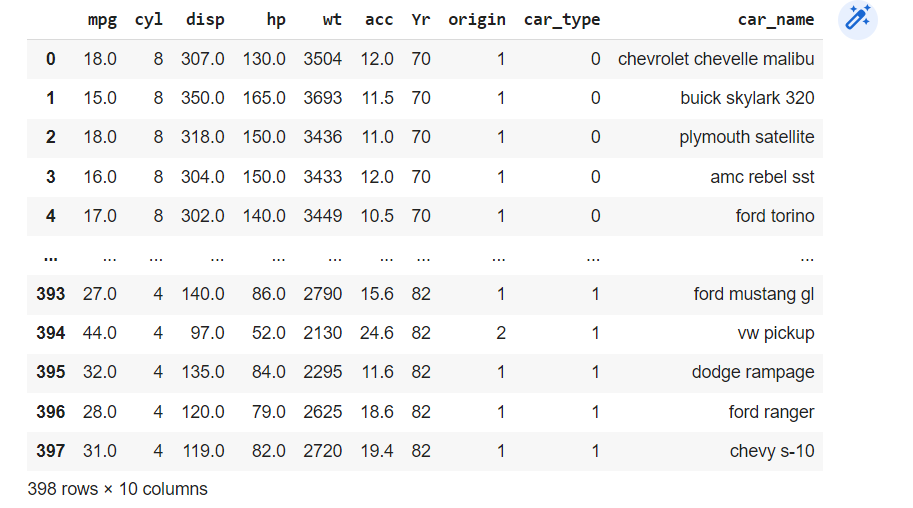
CAR\_TYPE - Fuel type of a car - 1.Petrol 2.Diesel



df.isnull()



df.fillna(0)



# Pandas in Dataset

import pandas as pd

nba = pd.read\_csv("Car.csv")

type(nba)

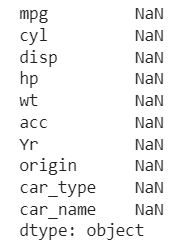
len(nba)



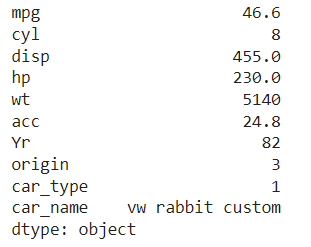
nba.shape()



df[df["hp"] == 1].mean()

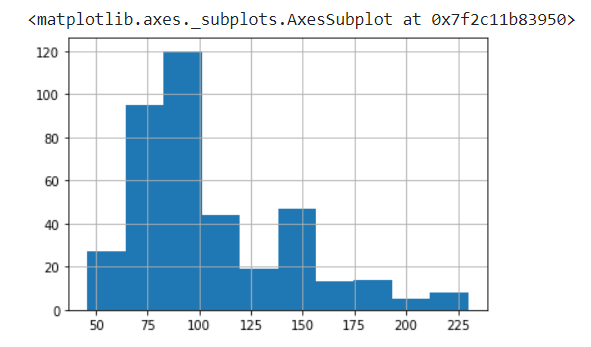


df.apply(np.max)



# Visualization in Dataset

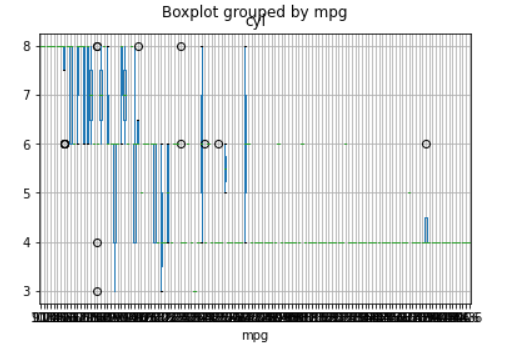
# df['hp'].hist()



INSIGHTS:

HP IS VARYING FROM 10 TO 120

df.boxplot(by ='mpg',column=['cyl'])

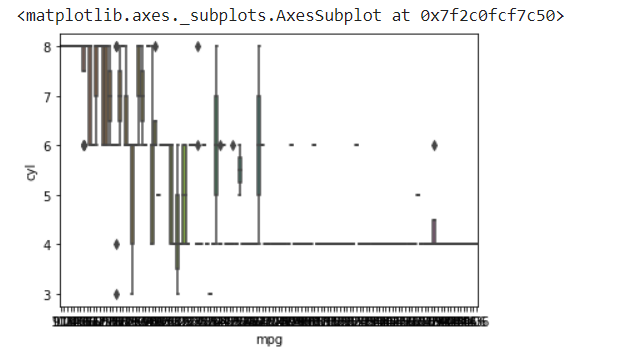


INSIGHTS:

IT GIVES US THE VARIOUS OUTLIERS IN MPG AND CYLINDER COLUMN.

import seaborn as sns

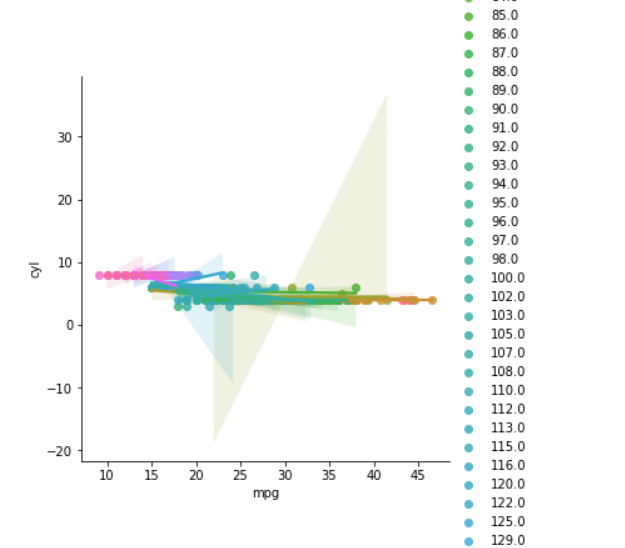
sns.boxplot(data = df, x = 'mpg', y = 'cyl')



INSIGHTS:

BOXPLOT EXECUTED USING SEABORN LIBRARY.

sns.lmplot(x='mpg', y='cyl',data=df,hue='hp')



df.boxplot(column = 'acc', by = 'wt')

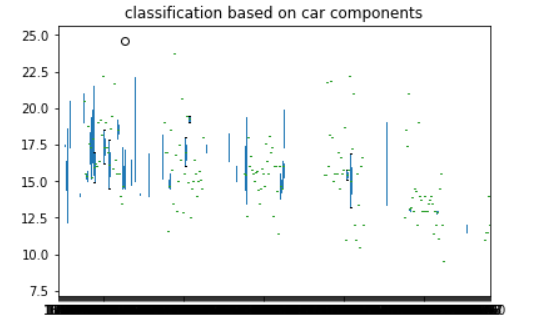
plt.suptitle('') # remove the automatic title

plt.title('classification based on car components')   # add a title

plt.ylabel('')   # add a y-axis title

plt.xlabel('')   # remove the x-axis title

plt.grid(None)



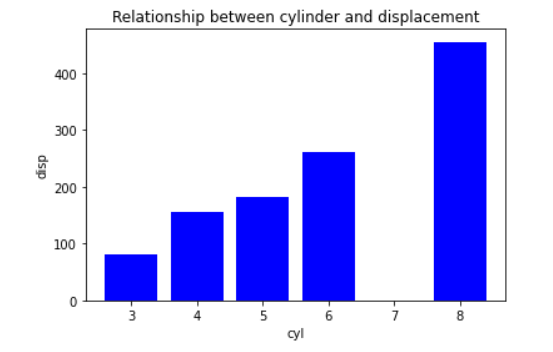
plt.bar("cyl", "disp", data = df, color = "blue")

plt.xlabel("cyl")

plt.ylabel("disp")

plt.title("Relationship between cylinder and displacement")

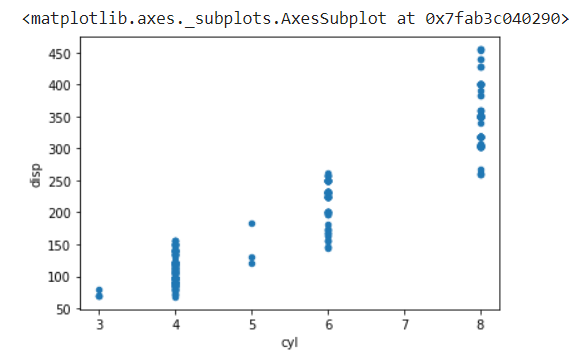
plt.show()



INSIGHTS:

THIS GIVES US THE RELATIONSHIP BETWEEN CYLINDER AND DISPLACEMENT.

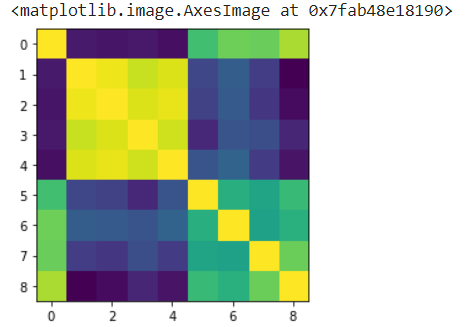
df.plot.scatter(x = 'cyl', y = 'disp')



INSIGHTS:

THERE IS NO VALUES IN 7.

plt.imshow(df.corr())



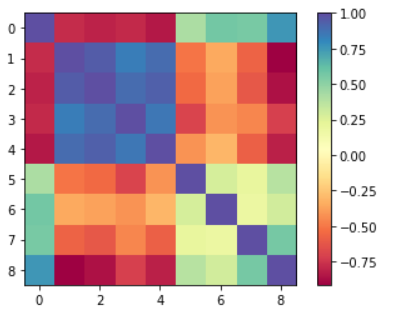
INSIGHTS:

THIS GIVES US THE COORELATION.

plt.imshow(df.corr(), cmap="Spectral")

plt.colorbar()

plt.show()



INSIGHTS:

THIS IS THE COLOUR BAR OF COORELATION.

Result:

Thus working with numpy arrays, pandas dataframe and basic plots using matplotlib were successfully executed.