WALCHAND INSTITUTE OF TECHNOLOGY, SOLAPUR

INFORMATION TECHNOLOGY

2020-21 SEMESTER –II

Elective 1 : Data Science

Name : Shivani Suresh Kulkarni

Roll No. : 28

Data Science Assignment 2

Title : pandas dataframe, Exploratory_Data_Analysis and Data Visulization

Program :

```
import pandas as pd

import numpy as np

import os

import matplotlib.pyplot as plt


data_csv=pd.read_csv("C:\\Users\\Admin\\Documents\\college\\Toyota.csv",
index_col=0, na_values=["??","????"])

print(data_csv)


#shallow copy

sam=data_csv.copy(deep=False)

#deep copy

sam=data_csv.copy(deep=True)
```

```
#get index
print(data_csv.index)
#get column
print(data_csv.columns)
#get size
print("size = ",data_csv.size)
#get shape
print('shape = ',data_csv.shape)
#memory usage
print('Memory usage : \n', data_csv.memory_usage())
#array dimentions
print('Array Dimensions : ', data_csv.ndim)
#first n rows
print('First n rows : \n', data_csv.head(6))
#last n rows
print('Last n now : \n', data_csv.tail(6))

print(data_csv.at[4,'FuelType'])
print(data_csv.iat[4,6])
print(data_csv.loc[:,'FuelType'])

#checking data types
print(data_csv.dtypes)
#selectng
print(data_csv.select_dtypes(exclude=[object]))
#summary of dataframe
print(data_csv.info)
```

```
#unique elements
print(np.unique(data_csv['KM']))
print(np.unique(data_csv['HP']))
print(np.unique(data_csv['MetColor']))
print(np.unique(data_csv['Automatic']))
print(np.unique(data_csv['Doors']))

#converting variables dtype
data_csv['MetColor']=data_csv['MetColor'].astype('object')
data_csv['Automatic']=data_csv['Automatic'].astype('object')

#Category vs object dtype
print(data_csv['FuelType'].nbytes)
print(data_csv['FuelType'].astype('category').nbytes)

#rechecking dtypes
print(data_csv.info())

#detect missing value
print(data_csv.isnull().sum())
```

OUTPUT :

Untitled6

localhost:8888/notebooks/Untitled6.ipynb?kernel_name=python3

jupyter Untitled6 Last Checkpoint: 8 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

In [1]:
```
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
```

In [6]:
```
data_csv=pd.read_csv("C:\\Users\\Admin\\Documents\\college\\Toyota.csv", index_col=0, na_values=["??","????"])
print(data_csv)
```

```
       Price  Age       KM FuelType     HP  MetColor  Automatic    CC  Doors  \
0      13500 23.0  46986.0   Diesel   90.0       1.0          0  2000  three
1      13750 23.0  72937.0   Diesel   90.0       1.0          0  2000      3
2      13950 24.0  41711.0   Diesel   90.0       NaN          0  2000      3
3      14950 26.0  48000.0   Diesel   90.0       0.0          0  2000      3
4      13750 30.0  38500.0   Diesel   90.0       0.0          0  2000      3
5      12950 32.0  61000.0   Diesel   90.0       0.0          0  2000      3
6      16900 27.0      NaN   Diesel    NaN       NaN          0  2000      3
7      18600 30.0  75889.0      NaN   90.0       1.0          0  2000      3
8      21500 27.0  19700.0   Petrol  192.0       0.0          0  1800      3
9      12950 23.0  71138.0   Diesel    NaN       NaN          0  1900      3
10     20950 25.0  31461.0   Petrol  192.0       0.0          0  1800      3
11     19950 22.0  43610.0   Petrol  192.0       0.0          0  1800      3
12     19600 25.0  32189.0   Petrol  192.0       0.0          0  1800      3
13     21500 31.0  23000.0   Petrol  192.0       1.0          0  1800      3
14     22500 32.0  34131.0   Petrol  192.0       1.0          0  1800      3
15     22000 28.0  18739.0   Petrol    NaN       0.0          0  1800      3
16     22750 30.0  34000.0   Petrol  192.0       1.0          0  1800      3
17     17950 24.0  21716.0   Petrol  110.0       1.0          0  1600      3
18     16750 24.0  25563.0   Petrol  110.0       0.0          0  1600      3
```

In [7]:
```
#shallow copy
sam=data_csv.copy(deep=False)
#deep copy
sam=data_csv.copy(deep=True)
```

Type here to search

ENG INTL 09:57 18-06-2021

Untitled6

localhost:8888/notebooks/Untitled6.ipynb?kernel_name=python3

jupyter Untitled6 Last Checkpoint: 8 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

In [1]:
```
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
```

In [6]:
```
data_csv=pd.read_csv("C:\\Users\\Admin\\Documents\\college\\Toyota.csv", index_col=0, na_values=["??","????"])
print(data_csv)
```

```
       Weight
0        1165
1        1165
2        1165
3        1165
4        1170
5        1170
6        1245
7        1245
8        1185
9        1105
10       1185
11       1185
12       1185
13       1185
14       1185
15       1185
16       1185
17       1105
18       1065
```

In [7]:
```
#shallow copy
sam=data_csv.copy(deep=False)
#deep copy
sam=data_csv.copy(deep=True)
```

Type here to search

ENG INTL 09:57 18-06-2021

Untitled6

localhost:8888/notebooks/Untitled6.ipynb?kernel_name=python3

jupyter Untitled6 Last Checkpoint: 8 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

In [1]:
```
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
```

In [6]:
```
data_csv=pd.read_csv("C:\\Users\\Admin\\Documents\\college\\Toyota.csv", index_col=0, na_values=["??","????"])
print(data_csv)
```

```
1419    1050
1420    1075
1421    1045
1422    1050
1423    1015
1424    1015
1425    1000
1426    1080
1427    1045
1428    1015
1429    1065
1430    1015
1431    1025
1432    1015
1433    1015
1434    1015
1435    1114

[1436 rows x 10 columns]
```

In [7]:
```
#shallow copy
sam=data_csv.copy(deep=False)
#deep copy
sam=data_csv.copy(deep=True)
```

Type here to search

ENG INTL 09:57 18-06-2021

Untitled6

localhost:8888/notebooks/Untitled6.ipynb?kernel_name=python3

jupyter Untitled6 Last Checkpoint: 8 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

In [7]:
```
#shallow copy
sam=data_csv.copy(deep=False)
#deep copy
sam=data_csv.copy(deep=True)
```

In [17]:
```
#get index
print(data_csv.index)
#get column
print(data_csv.columns)
#get size
print("size = ",data_csv.size)
#get shape
print('shape = ',data_csv.shape)
#memory usage
print('Memory usage : \n', data_csv.memory_usage())
#array dimentions
print('Array Dimensions : ', data_csv.ndim)
#first n rows
print('First n rows : \n', data_csv.head(6))
#last n rows
print('Last n now : \n', data_csv.tail(6))
```

```
Int64Index([   0,    1,    2,    3,    4,    5,    6,    7,    8,    9,
            ...
            1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435],
           dtype='int64', length=1436)
Index(['Price', 'Age', 'KM', 'FuelType', 'HP', 'MetColor', 'Automatic', 'CC',
       'Doors', 'Weight'],
      dtype='object')
size =  14360
shape =  (1436, 10)
Memory usage :
 Index        11488
Price        11488
Age          11488
```

Type here to search

ENG INTL 09:58 18-06-2021

Untitled6

localhost:8888/notebooks/Untitled6.ipynb?kernel_name=python3

jupyter Untitled6 Last Checkpoint: 8 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

```
Int64Index([   0,    1,    2,    3,    4,    5,    6,    7,    8,    9,
            ...
            1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435],
           dtype='int64', length=1436)
Index(['Price', 'Age', 'KM', 'FuelType', 'HP', 'MetColor', 'Automatic', 'CC',
       'Doors', 'Weight'],
      dtype='object')
size =  14360
shape =  (1436, 10)
Memory usage :
 Index        11488
Price        11488
Age          11488
KM           11488
FuelType     11488
HP           11488
MetColor     11488
Automatic    11488
CC           11488
Doors        11488
Weight       11488
dtype: int64
Array Dimensions :  2
First n rows :
    Price   Age       KM FuelType    HP  MetColor  Automatic    CC  Doors  \
0  13500  23.0  46986.0   Diesel  90.0       1.0          0  2000  three
1  13750  23.0  72937.0   Diesel  90.0       1.0          0  2000      3
2  13950  24.0  41711.0   Diesel  90.0       NaN          0  2000      3
3  14950  26.0  48000.0   Diesel  90.0       0.0          0  2000      3
4  13750  30.0  38500.0   Diesel  90.0       0.0          0  2000      3
5  12950  32.0  61000.0   Diesel  90.0       0.0          0  2000      3

   Weight
0    1165
1    1165
```

Type here to search

Untitled6

localhost:8888/notebooks/Untitled6.ipynb?kernel_name=python3

jupyter Untitled6 Last Checkpoint: 8 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

```
   Weight
0    1165
1    1165
2    1165
3    1165
4    1170
5    1170
Last n now :
       Price   Age       KM FuelType     HP  MetColor  Automatic    CC Doors  \
1430   8450  80.0  23000.0   Petrol   86.0       0.0          0  1300     3
1431   7500   NaN  20544.0   Petrol   86.0       1.0          0  1300     3
1432  10845  72.0      NaN   Petrol   86.0       0.0          0  1300     3
1433   8500   NaN  17016.0   Petrol   86.0       0.0          0  1300     3
1434   7250  70.0      NaN      NaN   86.0       1.0          0  1300     3
1435   6950  76.0      1.0   Petrol  110.0       0.0          0  1600     5

      Weight
1430    1015
1431    1025
1432    1015
1433    1015
1434    1015
1435    1114
```

In [18]:
```
print(data_csv.at[4,'FuelType'])
print(data_csv.iat[4,6])
print(data_csv.loc[:,'FuelType'])
```

```
Diesel
0
0       Diesel
1       Diesel
2       Diesel
3       Diesel
4       Diesel
```

Type here to search

Untitled6

localhost:8888/notebooks/Untitled6.ipynb?kernel_name=python3

jupyter Untitled6 Last Checkpoint: 8 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

In [18]:
```
print(data_csv.at[4,'FuelType'])
print(data_csv.iat[4,6])
print(data_csv.loc[:,'FuelType'])
```

```
Diesel
0
0       Diesel
1       Diesel
2       Diesel
3       Diesel
4       Diesel
5       Diesel
6       Diesel
7          NaN
8       Petrol
9       Diesel
10      Petrol
11      Petrol
12      Petrol
13      Petrol
14      Petrol
15      Petrol
16      Petrol
17      Petrol
18      Petrol
19      Petrol
20      Petrol
21         NaN
22      Petrol
23      Petrol
24      Petrol
25      Petrol
26         NaN
27      Petrol
```

Type here to search

Untitled6

localhost:8888/notebooks/Untitled6.ipynb?kernel_name=python3

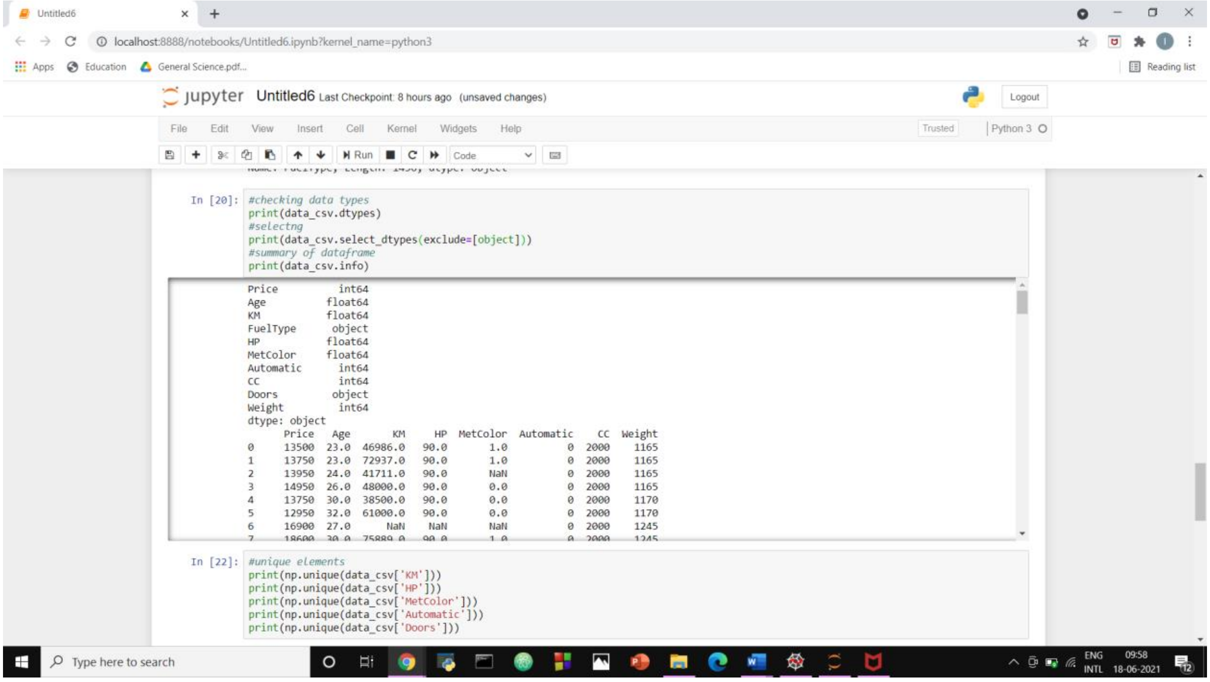
jupyter Untitled6 Last Checkpoint: 8 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

```
29         NaN
           ...
1406    Petrol
1407    Petrol
1408    Petrol
1409    Petrol
1410    Petrol
1411    Petrol
1412    Petrol
1413    Petrol
1414    Petrol
1415    Petrol
1416    Petrol
1417    Petrol
1418    Petrol
1419    Petrol
1420       NaN
1421    Petrol
1422       NaN
1423    Petrol
1424    Petrol
1425    Petrol
1426    Petrol
1427    Petrol
1428    Petrol
1429    Petrol
1430    Petrol
1431    Petrol
1432    Petrol
1433    Petrol
1434       NaN
1435    Petrol
Name: FuelType, Length: 1436, dtype: object
```

In [20]:
```
#checking data types
print(data_csv.dtypes)
```

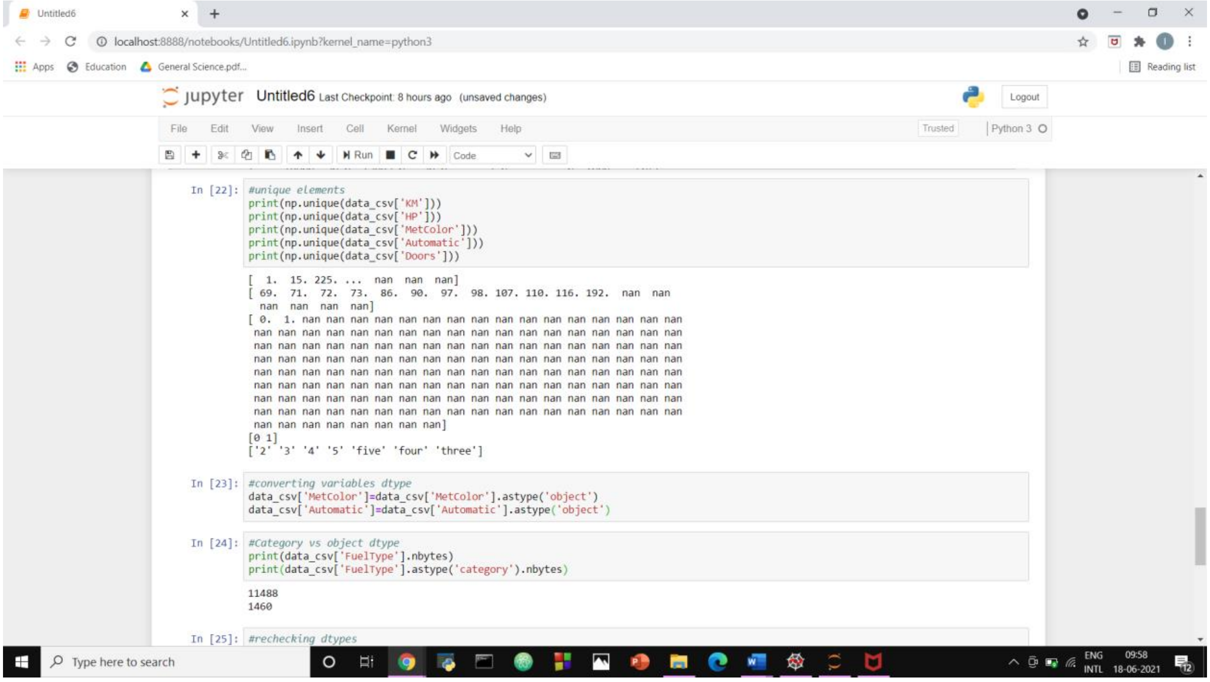
Type here to search



```
In [20]: #checking data types
         print(data_csv.dtypes)
         #selectng
         print(data_csv.select_dtypes(exclude=[object]))
         #summary of dataframe
         print(data_csv.info)
```

```
Price          int64
Age          float64
KM           float64
FuelType      object
HP           float64
MetColor     float64
Automatic      int64
CC             int64
Doors         object
Weight         int64
dtype: object
      Price   Age       KM    HP  MetColor  Automatic    CC  Weight
0     13500  23.0  46986.0  90.0       1.0          0  2000    1165
1     13750  23.0  72937.0  90.0       1.0          0  2000    1165
2     13950  24.0  41711.0  90.0       NaN          0  2000    1165
3     14950  26.0  48000.0  90.0       0.0          0  2000    1165
4     13750  30.0  38500.0  90.0       0.0          0  2000    1170
5     12950  32.0  61000.0  90.0       0.0          0  2000    1170
6     16900  27.0      NaN   NaN       NaN          0  2000    1245
7     18600  30.0  75889.0  90.0       1.0          0  2000    1245
```

```
In [22]: #unique elements
         print(np.unique(data_csv['KM']))
         print(np.unique(data_csv['HP']))
         print(np.unique(data_csv['MetColor']))
         print(np.unique(data_csv['Automatic']))
         print(np.unique(data_csv['Doors']))
```



```
In [22]: #unique elements
         print(np.unique(data_csv['KM']))
         print(np.unique(data_csv['HP']))
         print(np.unique(data_csv['MetColor']))
         print(np.unique(data_csv['Automatic']))
         print(np.unique(data_csv['Doors']))
```

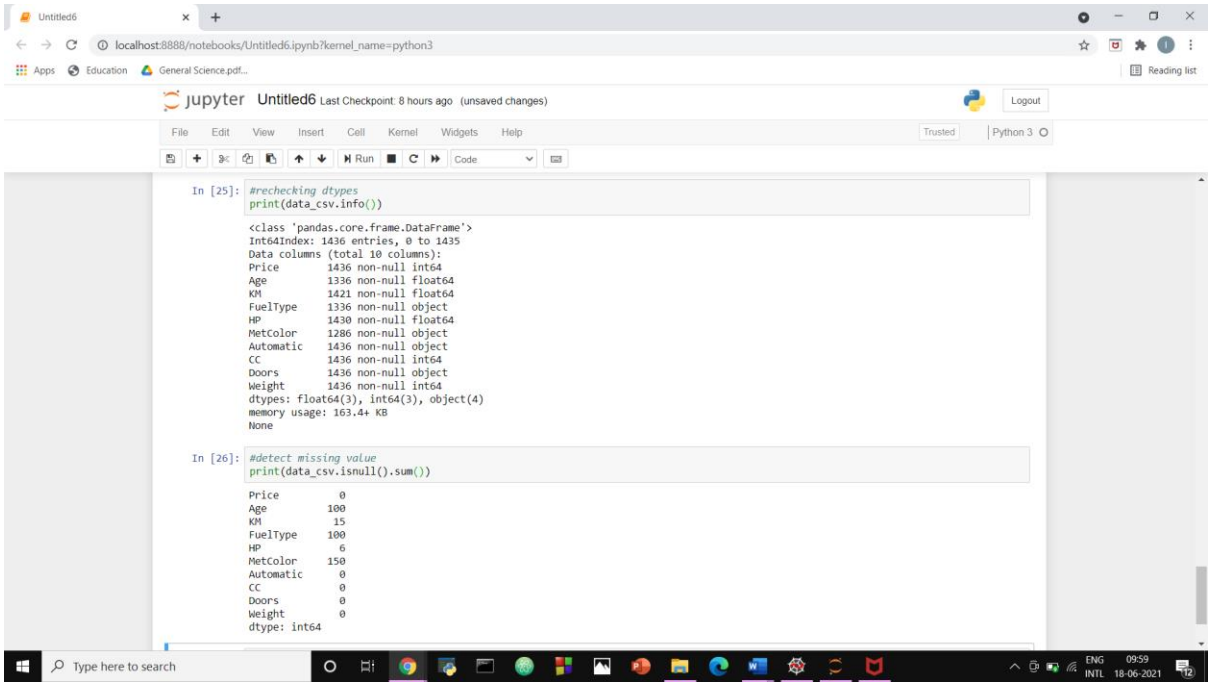
```
[  1.   15. 225. ...  nan  nan  nan]
[ 69.  71.  72.  73.  86.  90.  97.  98. 107. 110. 116. 192.  nan  nan
  nan  nan  nan  nan]
[ 0.  1. nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan]
[0 1]
['2' '3' '4' '5' 'five' 'four' 'three']
```

```
In [23]: #converting variables dtype
         data_csv['MetColor']=data_csv['MetColor'].astype('object')
         data_csv['Automatic']=data_csv['Automatic'].astype('object')
```

```
In [24]: #Category vs object dtype
         print(data_csv['FuelType'].nbytes)
         print(data_csv['FuelType'].astype('category').nbytes)
```

```
11488
1460
```

```
In [25]: #rechecking dtypes
```



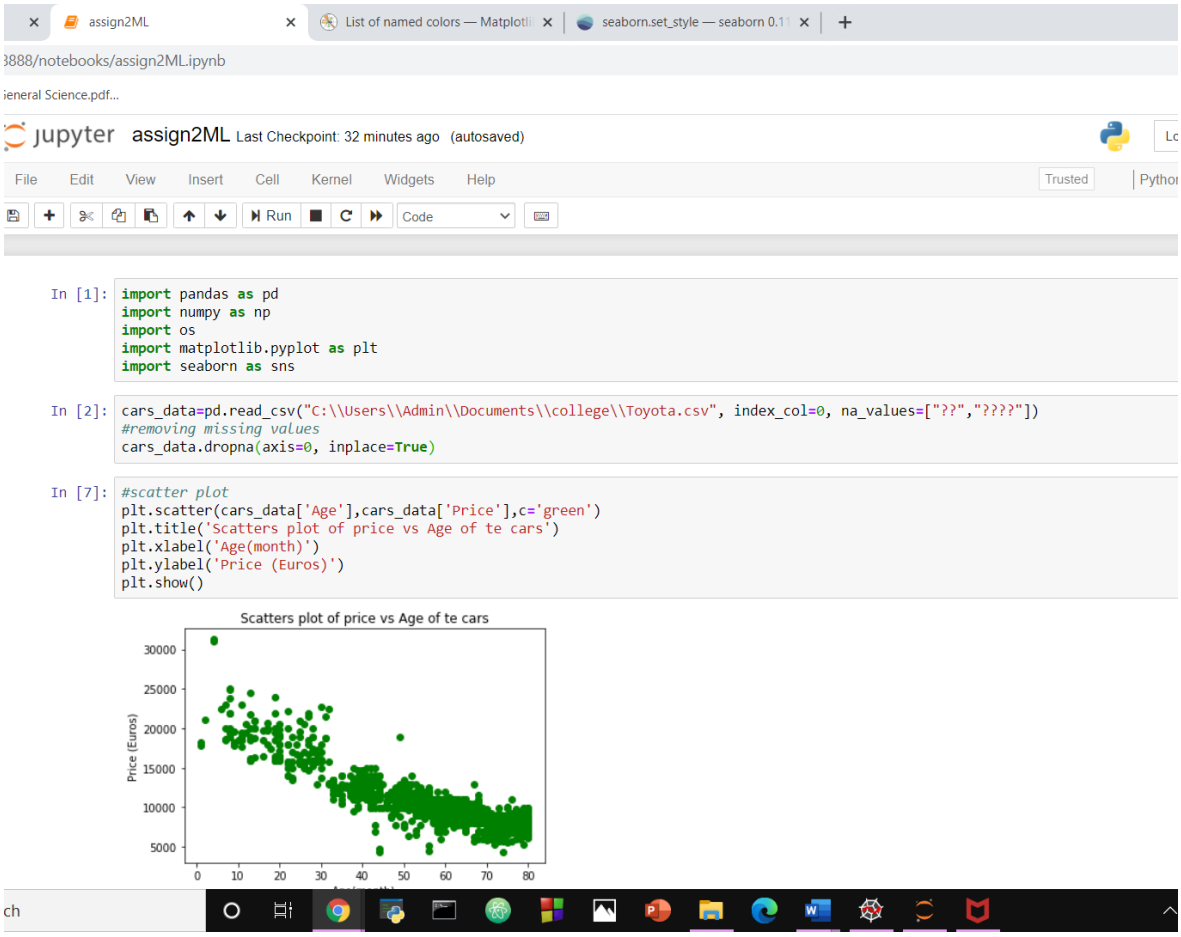
Program : (visualization)

```
import pandas as pd

import numpy as np

import os

import matplotlib.pyplot as plt

import seaborn as sns


cars_data=pd.read_csv("C:\\Users\\Admin\\Documents\\college\\Toyota.csv",
index_col=0, na_values=["??","????"])

#removing missing values

cars_data.dropna(axis=0, inplace=True)


#scatter plot
```

```
plt.scatter(cars_data['Age'],cars_data['Price'],c='green')
plt.title('Scatters plot of price vs Age of te cars')
plt.xlabel('Age(month)')
plt.ylabel('Price (Euros)')
plt.show()

#histogram
sns.set(style="white")
plt.hist(cars_data['KM'])
plt.hist(cars_data['KM'],color='darkviolet',edgecolor='white', bins=5)
plt.title('Histogram of kilometer')
plt.xlabel('Kilometer')
plt.ylabel('Frequency')
plt.show()

#barplot
counts=[979, 120, 12]
fuelType=('Petrol', 'Diesel', 'CNG')
index= np.arange(len(fuelType))
plt.bar(index, counts,color=['crimson','dodgerblue','navy'])
plt.title('Bar plot of fuel types')
plt.xlabel('Fuel Types')
plt.ylabel('Frequency')
plt.show()

#scatter plot with seaborn
sns.set(style="darkgrid")
sns.regplot(
```

```
    data=cars_data,
    x="Age", y="Price"
)
```

Output :



assign2ML | List of named colors — Matplotli | seaborn.set_style — seaborn 0.11

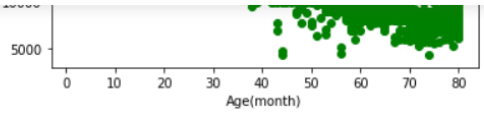
host:8888/notebooks/assign2ML.ipynb

General Science.pdf...

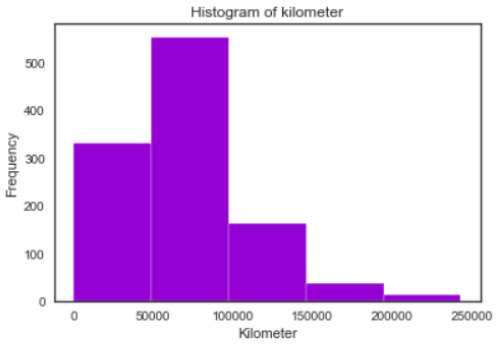
jupyter assign2ML Last Checkpoint: 33 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code



In [30]:
```
#histogram
sns.set(style="white")
plt.hist(cars_data['KM'])
plt.hist(cars_data['KM'],color='darkviolet',edgecolor='white', bins=5)
plt.title('Histogram of kilometer')
plt.xlabel('Kilometer')
plt.ylabel('Frequency')
plt.show()
```



In [29]:
```
#barplot
counts=[979, 120, 12]
fuelType=('Petrol', 'Diesel', 'CNG')
```

search

assign2ML | List of named colors — Matplotli | seaborn.set_style — seaborn 0.11

ost:8888/notebooks/assign2ML.ipynb

General Science.pdf...

jupyter assign2ML Last Checkpoint: 33 minutes ago (autosaved)

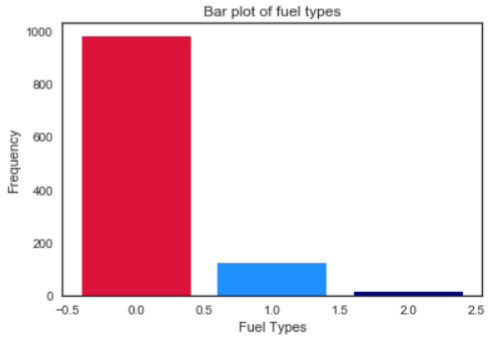
File Edit View Insert Cell Kernel Widgets Help

Run Code

0 50000 100000 150000 200000 250000

Kilometer

In [29]:
```
#barplot
counts=[979, 120, 12]
fuelType=('Petrol', 'Diesel', 'CNG')
index= np.arange(len(fuelType))
plt.bar(index, counts,color=['crimson','dodgerblue','navy'])
plt.title('Bar plot of fuel types')
plt.xlabel('Fuel Types')
plt.ylabel('Frequency')
plt.show()
```



In [22]:
```
#scatter plot with seaborn
sns.set(style="darkgrid")
#sns.regplot(x=cars_data['Age'],y=cars_data['Price'])
#sns.regplot(x=cars_data['Age'],y=cars_data['Price'],fit_reg=False)
#sns.regplot(x=cars_data['Age'],y=cars_data['Price'],marker"*",fit_reg=False)
```

earch

assign2ML | List of named colors — Matplotli | seaborn.set_style — seaborn 0.11

:8888/notebooks/assign2ML.ipynb

General Science.pdf...

jupyter assign2ML Last Checkpoint: 33 minutes ago (autosaved)

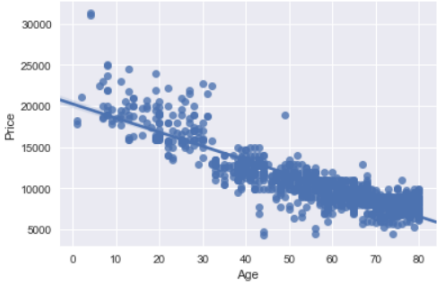
File Edit View Insert Cell Kernel Widgets Help Trusted

Run Code

Fuel Types

In [22]:
```
#scatter plot with seaborn
sns.set(style="darkgrid")
#sns.regplot(x=cars_data['Age'],y=cars_data['Price'])
#sns.regplot(x=cars_data['Age'],y=cars_data['Price'],fit_reg=False)
#sns.regplot(x=cars_data['Age'],y=cars_data['Price'],marker"*",fit_reg=False)

sns.regplot(
    data=cars_data,
    x="Age", y="Price"
   # hue="smoker", style="smoker", size="size",
)
```

Out[22]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x21b2d5df668>
```



In [ ]:

rch