PROJECT 1

#Importing the necessary libraries

```
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

#Reading the csv file

```
df = pd.read_csv('vehicle_data (2).csv')
df.head(10)
```

#sorting values of ten cars in vehicle dataset with price in desceding order

```
df[df['category']=='Cars'].sort_values(by='price', ascending=False).head(10).loc[:,['category','price']]
```

	category	price
22	Cars	14500000
265	Cars	9500000
224	Cars	9000000
156	Cars	8799999
249	Cars	8199999
53	Cars	6500000
0	Cars	6500000
241	Cars	6300000
177	Cars	6150000
8	Cars	5800000

#sorting values of ten Buses & Microbuses in vehicle dataset with price in desceding order

```
df[df['category']=='Buses & Microbuses'].sort_values(by='price', ascending=False).head(5).loc[:,['category','price']]
```

	category	price
148	Buses & Microbuses	11200000
221	Buses & Microbuses	5200000
174	Buses & Microbuses	4900000
211	Buses & Microbuses	4300000
268	Buses & Microbuses	3800000

#sorting values of ten Trucks & Trailers in vehicle dataset with price in desceding order

```
df[df['category']=='Trucks & Trailers'].sort_values(by='price',
ascending=False).head(5).loc[:,['category','price']]
```

	category	price
195	Trucks & Trailers	7500000
222	Trucks & Trailers	6000000
103	Trucks & Trailers	5100000
176	Trucks & Trailers	4300000
62	Trucks & Trailers	3650000

 # grab the top 10

```
df['region'].value_counts()[:10]
```

```
Out[13]: Mombasa CBD        92
         Mvita              28
         Nairobi Central    27
         Kilimani           23
         Lavington          16
         Ridgeways          15
         Tudor              13
         Karen               8
         Langata             7
         Nairobi             6
         Name: region, dtype: int64
```

 # make it a variable

```
top_10 = df['region'].value_counts()[:10]
```

#Using matplotlib

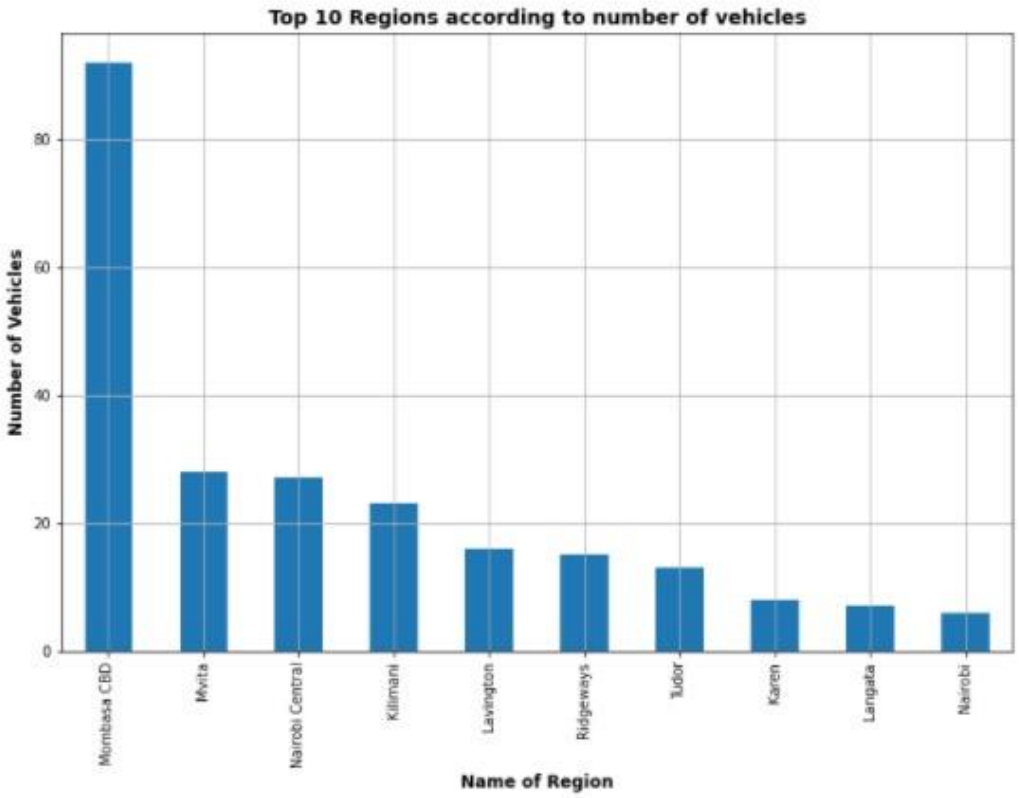
```
plt.figure(figsize=(12,8)) plt.title("Top 10 Regions according to number of vehicles",
fontsize=14, fontweight='bold') top_10.plot.bar() plt.xlabel('Name of Region',fontsize=12,
fontweight='bold') plt.ylabel('Number of Vehicles',fontsize=12, fontweight='bold')
plt.grid()
```

 # save the plot to file

```
fig = plt.gcf()
fig.savefig('top-5-regions.png')
```

plt.show()

```
plt.show()
```



PROJECT 2

import os
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns

 # read in the necessary file ('clean_stock_prices.csv')

```
# read in the necessary file ('clean_stock_prices.csv')
df = pd.read_csv('clean_stock_prices.csv',index_col=0)
df.head()
```

Date	EGAD	KUKZ	LIMT	SASN	WTK	CGEN	ABSA	BKG	DTK	EQTY	...	BAT	CARB	EABL	EVRD	FTGH	ORCH	MSC	UNGA	SCOM	FAHR
2022-01-13	12.90	385.0	320.0	22.20	130.00	54.00	11.80	30.00	59.00	49.55	...	440.0	10.80	151.50	0.96	1.34	10.4	0.27	27.10	37.95	6.52
2022-01-11	13.80	385.0	320.0	20.55	134.75	44.75	11.90	30.75	59.50	52.00	...	445.0	10.85	161.00	0.88	1.31	10.4	0.27	27.65	38.45	6.56
2022-01-07	13.80	420.0	320.0	21.25	132.00	37.05	11.80	29.05	60.00	53.00	...	442.0	10.90	164.75	0.94	1.30	10.4	0.27	27.65	39.90	6.38
2022-01-06	13.80	420.0	320.0	20.25	130.75	33.70	11.80	29.30	60.00	53.00	...	442.0	10.90	160.75	0.99	1.29	10.4	0.27	27.65	40.00	6.40
2022-01-05	12.85	420.0	320.0	19.95	130.75	30.60	11.75	29.50	59.75	53.00	...	442.0	10.90	163.75	0.99	1.26	10.4	0.27	27.65	39.55	6.02

5 rows × 60 columns

```
df.tail()
```

Date	EGAD	KUKZ	LIMT	SASN	WTK	CGEN	ABSA	BKG	DTK	EQTY	...	BAT	CARB	EABL	EVRD	FTGH	ORCH	MSC	UNGA	SCOM	FAHR
2021-08-09	12.15	415.0	300.00	19.50	134.5	35.0	9.80	32.40	65.75	50.25	...	445.5	12.25	179.25	0.96	1.32	10.4	0.27	31.0	42.95	6.78
2021-08-06	12.15	415.0	300.00	20.00	134.5	35.0	9.80	32.40	65.75	50.00	...	454.0	12.25	179.00	0.98	1.32	10.4	0.27	31.0	42.10	6.98
2021-08-05	12.30	415.0	320.00	20.00	134.5	35.0	9.82	31.85	65.00	49.40	...	450.0	12.20	178.50	0.98	1.31	10.4	0.27	31.0	41.25	6.80
2021-08-04	12.00	415.0	320.00	19.95	135.0	35.0	9.76	29.75	64.00	49.10	...	455.0	12.00	179.75	0.98	1.30	10.4	0.27	31.0	41.10	6.92
2021-08-03	11.80	415.0	304.75	19.95	134.5	35.0	9.82	29.50	65.00	49.00	...	450.0	12.00	180.00	0.98	1.31	10.4	0.27	28.3	41.05	6.94

5 rows × 60 columns

```
# Lowest price for Safaricom
df = pd.read_csv('clean_stock_prices.csv')
min1 = df['SCOM'].min()
min1
```

```
36.5
```

```
# highest price for Safaricom
df = pd.read_csv('clean_stock_prices.csv')
max1 = df['SCOM'].max()
max1
```

```
44.95
```

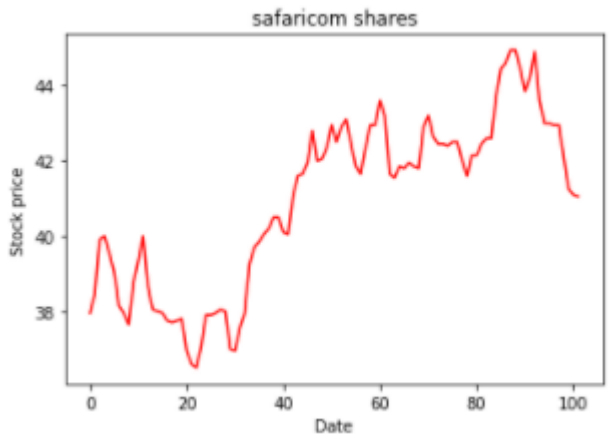
 # Plot SCOM to confirm above observations

```
plt.title('safaricom shares')

# set the xlabel
plt.xlabel("Date")

# set the ylabel
plt.ylabel("Stock price")

plt.plot('SCOM',data=df, color="red")
plt,show()
```



#creating subsets containing all data

```
agric_df = df.loc[:,'EGAD':'WTK'].copy()
agric_df.head()
```

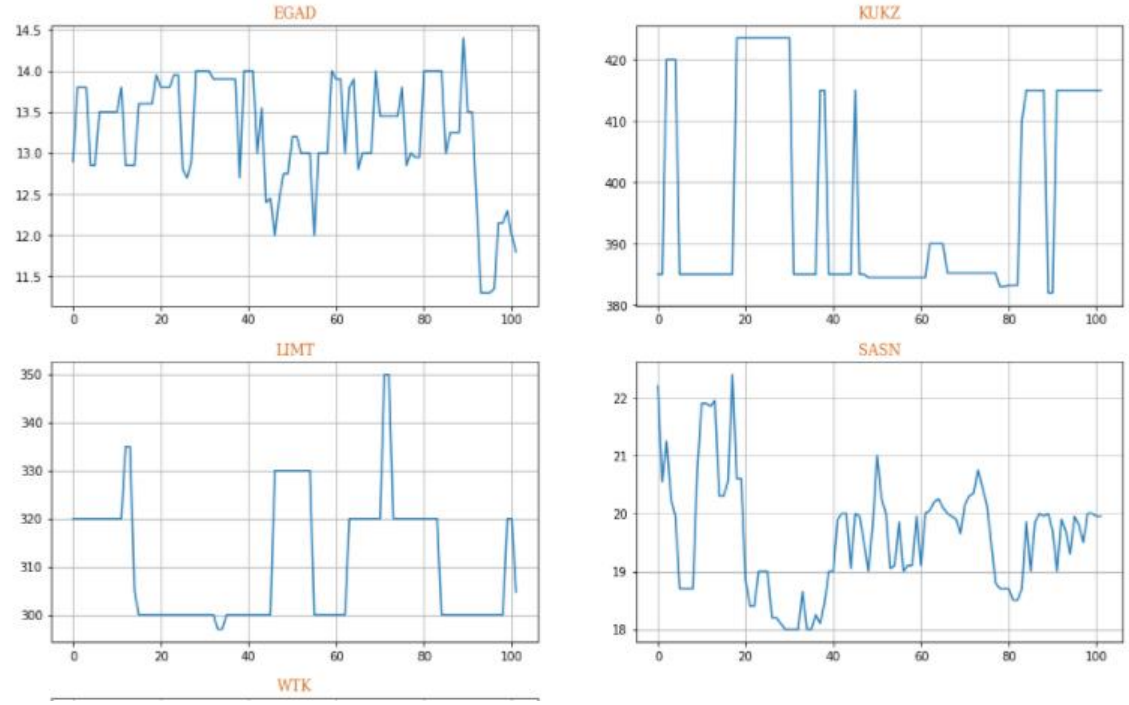
	EGAD	KUKZ	LIMT	SASN	WTK
0	12.90	385.0	320.0	22.20	130.00
1	13.80	385.0	320.0	20.55	134.75
2	13.80	420.0	320.0	21.25	132.00
3	13.80	420.0	320.0	20.25	130.75
4	12.85	420.0	320.0	19.95	130.75

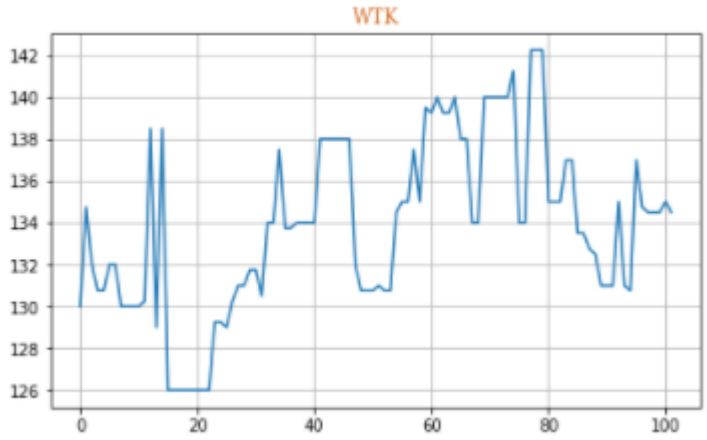
#using matplotlib to create subplots

```
bank_cols = agric_df.columns

font = {'family':'serif',
      'color': 'chocolate',
      'weight': 'normal'
      }
for idx,bank in enumerate (bank_cols,start=1):
   plt.subplot(6,2,idx)
   plt.title(bank,fontdict=font)
   plt.grid()
   plt.plot(bank,data=df)

fig = plt.gcf()
fig.set_size_inches(16,30)
plt.show()
```





#correlation for Agric

corr_df = agric_df.corr(method="pearson")

#Using Seaborn

plt.figure(figsize=(13,8))
sns.heatmap(corr_df, annot=True, cmap='RdYlGn')
plt.figure()