

Oduor Edwin

## Question 2: Time series Project

Import the required libraries

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

Read the csv file from your working directory

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

Date	EGAD	KUKZ	LIMIT	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 x 60 columns

Obtain the lowest price for Safaricom (SCOM)

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

: 36.5

Obtain the highest price for Safaricom (SCOM)

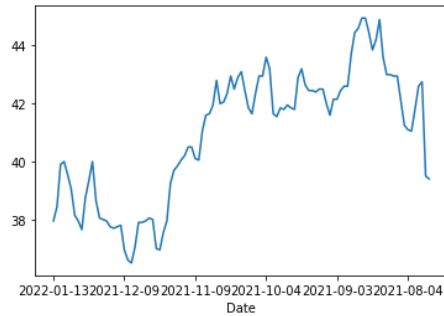
```
# highest price for Safaricom
df = pd.read_csv('cleaned_stock.csv')
max1 = df['SCOM'].max()
max1
```

44.95

## Plot a graph for Safaricom(SCOM) shares

```
df['SCOM'].plot()
```

<AxesSubplot:xlabel='Date'>



## Subsetting Agric sector

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

Date	EGAD	KUKZ	LIMIT	SASN	WTK
2022-01-13	12.90	385.0	320.0	22.20	130.00
2022-01-11	13.80	385.0	320.0	20.55	134.75
2022-01-07	13.80	420.0	320.0	21.25	132.00
2022-01-06	13.80	420.0	320.0	20.25	130.75
2022-01-05	12.85	420.0	320.0	19.95	130.75

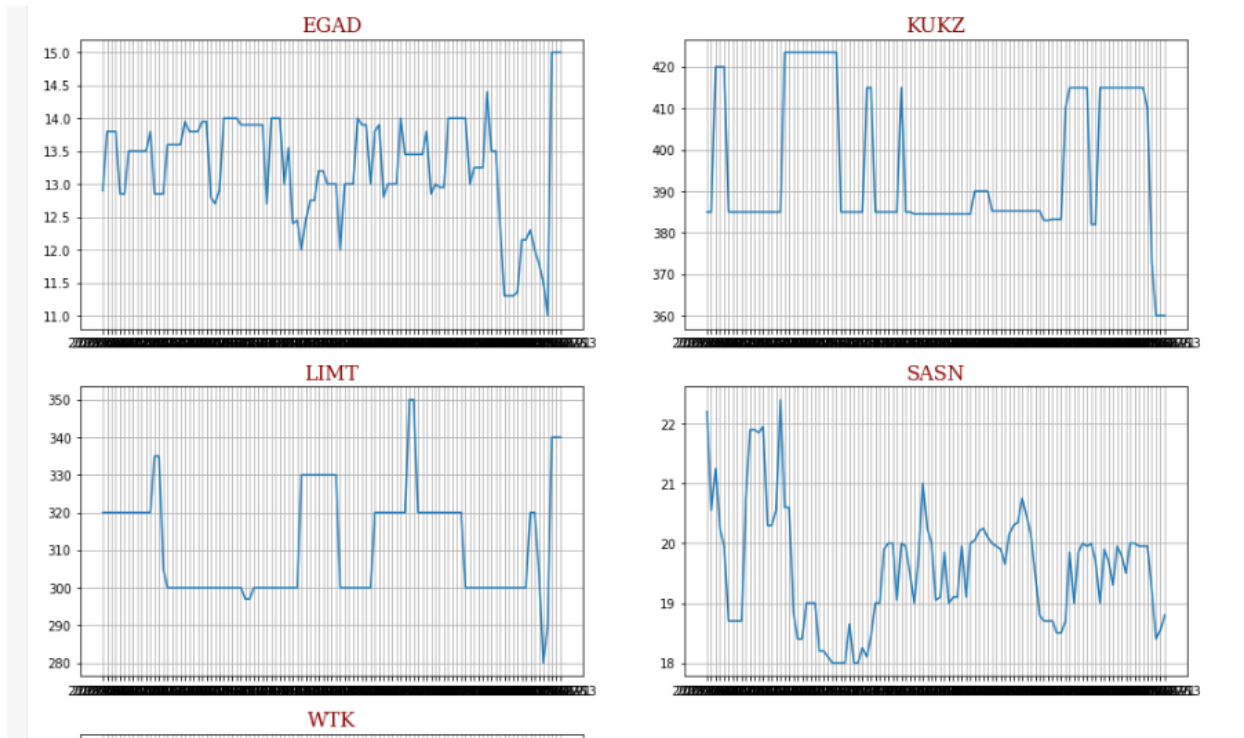
using matplotlib subplot to create subplots to fit all the sector stocks in one plot.

```
: #Using Matplotlib to come up with Subplots for Agric Sector
bank_cols = agric_df.columns

font = {'family': 'serif',
        'color': 'darkred',
        'weight': 'normal',
        'size': 16,
        }

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()
```



Use Seaborn to plot the correlation plot for your sector stocks.

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

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

<Figure size 720x360 with 0 Axes>



<Figure size 720x360 with 0 Axes>

## Question 1. Vehicle Dataset

Import the necessary libraries

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

Read the required CSV file from your working directory

```
#Listing the first 5 vehicles from the vehicle dataset
```

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

	title	category	region	parent_region	condition	attrs	brand	color	model	yom	mileage	body_type	fuel	drive_train	trans	seat
0	Toyota Land Cruiser Prado 2016 Black	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults	Toyota	Black	Land Cruiser Prado	2016.0	87000.0	NaN	NaN	NaN	Automatic	NaN
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	Demio	2014.0	92000.0	NaN	NaN	NaN	Automatic	NaN
2	Clean NV300 Caravan 2014 Model Diesel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used		Nissan	Nissan	NaN	Caravan (Urvan)	2014.0	180000.0	NaN	NaN	NaN	NaN
3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	Crown	2014.0	75000.0	NaN	NaN	NaN	Automatic	NaN
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0	58000.0	NaN	NaN	NaN	Automatic	NaN

The titles and prices of 10 Cars with highest price

```
#Titles and prices of 10 cars with the highest price
```

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

	title	price
22	Lexus RX 2016 Black	14500000
265	New Hyundai Palisade 2021 White	9500000
224	Toyota Hilux 2016 Black	9000000
156	Toyota Land Cruiser 2010 4.6 V8 ZX Black	8799999
249	Toyota Land Cruiser 2014 4.6 V8 ZX Black	8199999
53	Toyota Land Cruiser Prado 2015 2.7 VVT-i Brown	6500000
0	Toyota Land Cruiser Prado 2016 Black	6500000
241	BMW X5 2015 White	6300000
177	Toyota Land Cruiser Prado 2014 Blue	6150000
8	BMW X4 2015 xDrive35i Black	5800000

The titles and prices of 5 Buses & Microbuses with highest price

```
#Titles and prices of 5 Buses and microbuses with the highest price.  
df[df['category']=='Buses & Microbuses'].sort_values(by='price', ascending=False).head(5).loc[:,['title','price']]
```

	title	price
148	Mazda Bongo	11200000
221	Selling Buses In Mombasa Town	5200000
174	Roller Coaster	4900000
211	Toyota Coaster 2014 White	4300000
268	Toyota Hiace 2015 White	3800000

The titles and prices of 5 Trucks & Trailers with highest price

```
#Titles and prices of 5 trucks and trailers with highest price  
df[df['category']=='Trucks & Trailers'].sort_values(by='price', ascending=False).head(5).loc[:,['title','price']]
```

	title	price
195	Mercedes-Benz Actros	7500000
222	Tata Signa LPK-1618 Tipper 10 Ton	6000000
103	Shacman F2000 Tipper	5100000
176	Isuzu Forward 7 Tonne Freezer	4300000
62	Isuzu Elf,Year 2015 Manual	3650000

Regions with the highest number of vehicles from the vehicle dataset

```
: #Regions with the highest number of vehicles  
df['region'].value_counts()[:10]
```

```
: 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
```

Create a variable for regions with the highest number of vehicles

```
: #Creating a variable for top 10 regions with the highest number of vehicles  
top_10 = df['region'].value_counts()[:10]
```

Use matplotlib to come up with a plot indicating the top 10 brands that we have in the vehicle\_dataset

```
#Using Matplotlib to come up with a graph of top 10 regions with the highest number of vehicles
```

```
plt.figure(figsize=(12,8))
plt.title("Top 10 regions with the highest number of vehicle")
fontsize = 14
fontweight = 'bold'
top_10.plot.bar()
plt.xlabel("Name of Region",fontsize = 12,fontweight = 'bold')
plt.ylabel("Number of vehicle",fontsize = 12, fontweight = 'bold')
plt.grid()
```

