

# Getting Started with Python

## Time Series Project

Author : Bradley Daudi

Date: 26/3/2022

## Resources

1. [Clean Dataset \(some-link\)](#)
2. [Submission Portal \(https://dasclab.uonbi.ac.ke/analytics/projects\)](https://dasclab.uonbi.ac.ke/analytics/projects)

If you are having problems please refer to this document:

3. [Time Series Notebook \(https://dasclab.uonbi.ac.ke/dstraining/time-series-data.html\)](https://dasclab.uonbi.ac.ke/dstraining/time-series-data.html)

## Instructions

Import all the libraries listed in the first cell. Make sure all modules are installed.

Use the provided data set to answer the following:

1. a) What is the lowest price for Safaricom (SCOM) b) What was the date when Safaricom had the lowest price?
2. a) What is the highest price Safaricom stock reached in the data b) What was the date when Safaricom stock recorded the highest price?
3. Create a line plot for Safaricom stock and verify if the information provided above is indeed correct.
4. Select **one** of the sectors provided (agric, comm, bank, const, energy, insur, invest, manu)
  - a) Use **pandas** to create a subset containing all the rows of the dataframe and only companies in your selected sector. Rename this dataframe to the **sector\_name\_df**
  - b) Using the subset for the sector, use **matplotlib** subplot to create subplots to fit all the sector stocks in one plot. One row can have a maximum of 3 charts.
  - c) Using your sector DataFrame use the `corr()` DataFrame method to come up with a correlogram. Create a DataFrame for these correlations
  - d) Use **Seaborn** to plot the **correlation plot** for your sector stocks.

### Key performance Metrics:

- Go an extra step to produce charts that are visually appealing
- Ensure all the plots have a Title
- Ensure all plots have x labels and y labels where applicable
- Your plots should be clearly visible. Change the size of your plot to a comfortable width and height.
- Save all your plots

In [1]:

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

Ensure that you have the *clean\_stock\_prices.csv* file in your working directory

In [2]:

```
os.listdir()
```

Out[2]:

```
['.ipynb_checkpoints',  
'01 High.mp3',  
'01-11-2020-041924The 5 AM Club - Robin Sharma.pdf',  
'2018 GTEE APPLICATION FORM (4).pdf',  
'227-Intern-Statistician.pdf',  
'61N96kYP7dL.jpg',  
'ACTIVATION.pdf',  
'ADMISSION LETTER (539).pdf',  
'BitTorrent.exe',  
'Book3.xlsx',  
'BraveBrowserSetup.exe',  
'Californication 2007 Season 3 Complete 720p AMZN WEBRip x264 [i_c]',  
'Californication 2007 Season 5 Complete 720p AMZN WEBRip x264 [i_c]',  
'Californication 2007 Season 6 Complete 720p AMZN WEBRip x264 [i_c]',  
'Californication 2007 Season 7 Complete 720p AMZN WEBRip x264 [i_c]',  
'Californication Season 1 - COMPLETE 720p HDTV x264 [MKV,AC3,5.1] Ehhhh',  
'Californication Season 2 - COMPLETE 720p HDTV x264 [MKV,AC3,5.1] Ehhhh',  
'Californication Season 4 - COMPLETE 720p HDTV x264 [MKV,AC3,5.1] Ehhhh',  
'chriska-adventur-2022.zip',  
'ChromeSetup.exe',  
'clean_stock_prices.csv',  
'Cover letter- Joan Ngigi.pdf',  
'Cv dau latest.pdf',  
'CV-Joan Ngigi.PDF',  
'DashboardSetup.exe',  
'DAUDI_OTIENO_Elements_report (1).pdf',  
'DAUDI_OTIENO_Elements_report.pdf',  
'desktop.ini',  
'download.jfif',  
'dtb v9.8.bin',  
'epm_free_install_20211226.28161.exe',  
'epm_suite_freeB18.exe',  
'Everything Is Fucked - Mark Manson.pdf',  
'expressvpn_windows_10.21.0.9_release.exe',  
'GazettedLabslist.pdf',  
'General Experience_F.docx',  
'HDDScan.zip',  
'IMG-1628.jpg',  
'IMG-1629.jpg',  
'IMG-20210704-WA0010.jpg',  
'IMG-20211229-WA0004.jpg',  
'IMG-20211229-WA0005.jpg',  
'IMG-20211229-WA0006.jpg',  
'IMG-20211229-WA0007.jpg',  
'IMG-20211229-WA0008.jpg',  
'IMG-20211229-WA0009.jpg',  
'Joan wangui -attachment letter and admission letter.pdf',  
'ks4.021.3.10.391en_25092.exe',  
'Mo+Dao+Zu+Shi.crx',  
'MyProductList.csv',  
'mysql pic.jpg',  
'mysql-installer-web-community-8.0.28.0.msi',  
'newton raptison derivation.pdf',  
'new_daily_prices.csv',  
'ORPPJOBAPPLICATIONFORM.pdf',
```

```
'PCC-DET6Z7D7-POLICE CLEARANCE CERTIFICATE P0.pdf',
'pdfcoffee.com_rollo-tomassi-the-rational-male-4-pdf-free.pdf',
'pexels-aleksandar-pasaric-325185.jpg',
'pexels-maxime-francis-2246476.jpg',
'Pictures - Shortcut.lnk',
'project-time-series-workbook (1).ipynb',
'project-time-series-workbook.ipynb',
'python-3.10.2-amd64 (1).exe',
'python-3.10.2-amd64.exe',
'R-4.0.0',
'R-4.0.0.tar.gz',
'rufus-3.17.exe',
'SAINT_JHN_-_The_Best_Part_Of_Life.mp3',
'Sample_letters_and_resume.pdf',
'student_copy_pandas_workbook.ipynb',
'student_workbook_stocks.ipynb',
'sublime_text_build_4126_x64_setup.exe',
'The Big Book Of Small House Designs 75 Award Winning Plans For Your Dream
House 1250 Square Feet Or Less',
'The.Boys.2019.S01.COMPLETE.720p.AMZN.WEBRip.x264-GalaxyTV[TGx]',
'The.Boys.S02.COMPLETE.720p.AMZN.WEBRip.x264-GalaxyTV[TGx]',
'The.Kings.Man.2021.HDRip.XviD.B4ND1T69',
'top-10-brands.png',
'TOR-FOR-KELCOP-VACANCIES.pdf',
'Unconfirmed_803033.crdownload',
'vehicle_data (1).csv',
'vehicle_data.csv',
'vehicle_dataset_project (1).ipynb',
'vehicle_dataset_project.ipynb',
'video.mp4',
'WD Backup_1.9.7435.38388.zip',
'WindowsPCHealthCheckSetup (1).msi',
'WindowsPCHealthCheckSetup.msi',
'winrar-x64-602.exe',
'_Getintopc.com_Adobe_Lightroom_Classic_2022_v11.0.0.10.rar',
'_igetintopc.com_Adobe_Premiere_Pro_2022_x64 (1).rar']
```

If you can see the *clean\_stock\_prices.csv* as an output in the above cell, read the data into a DataFrame using pandas

In [3]:

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

Out[3]:

	EGAD	KUKZ	LIMT	SASN	WTK	CGEN	ABSA	BKG	DTK	EQTY	...	BAT	CAR
<b>Date</b>													
<b>2022-01-13</b>	12.90	385.0	320.0	22.20	130.00	54.00	11.80	30.00	59.00	49.55	...	440.0	10.8
<b>2022-01-11</b>	13.80	385.0	320.0	20.55	134.75	44.75	11.90	30.75	59.50	52.00	...	445.0	10.8
<b>2022-01-07</b>	13.80	420.0	320.0	21.25	132.00	37.05	11.80	29.05	60.00	53.00	...	442.0	10.9
<b>2022-01-06</b>	13.80	420.0	320.0	20.25	130.75	33.70	11.80	29.30	60.00	53.00	...	442.0	10.9
<b>2022-01-05</b>	12.85	420.0	320.0	19.95	130.75	30.60	11.75	29.50	59.75	53.00	...	442.0	10.9

5 rows × 60 columns

## Use this part to answer questions 1, 2 and 3

In [5]:

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

Out[5]:

36.5

In [6]:

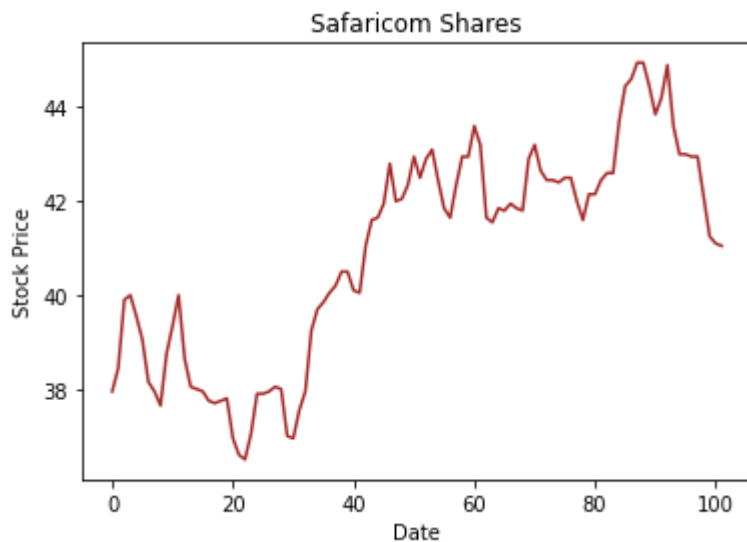
```
# highest price for Safaricom
df = pd.read_csv('clean_stock_prices.csv')
max1 = df['SCOM'].max()
max1
```

Out[6]:

44.95

In [10]:

```
# 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="brown")
plt.show()
```



## Use this part to answer question 4

In [12]:

```
# agricultural companies
agric = ['EGAD', 'KUKZ', 'LIMT', 'SASN', 'WTK']
```

To subset a sector simply use the **slice** notation. For example if I choose the Insurance sector, i will use the **insur** list

In [13]:

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

Out[13]:

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

In [ ]:

In [18]:

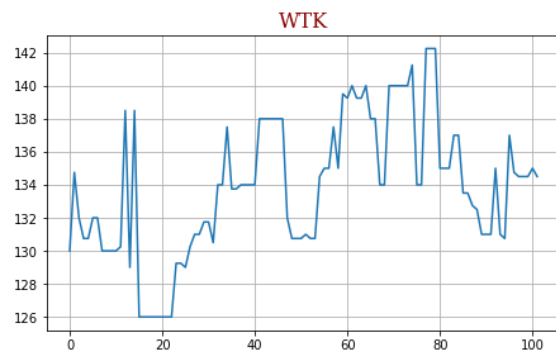
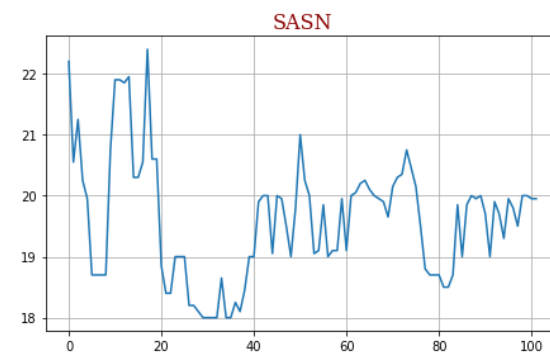
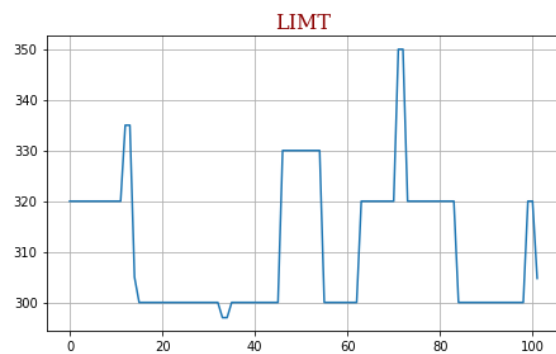
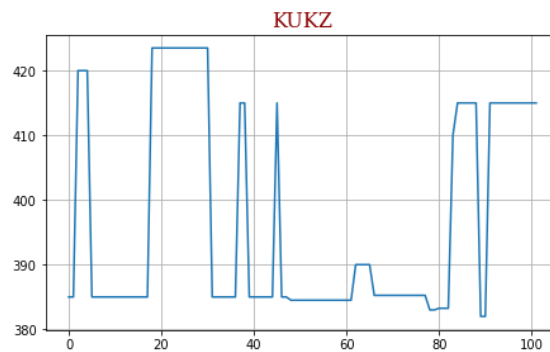
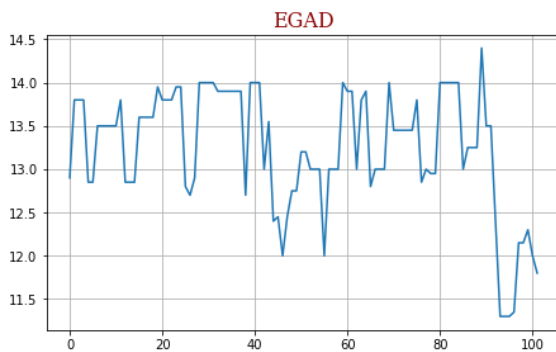
```
#Using Matplot to creat subplots
```

```
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()
```



In [33]:

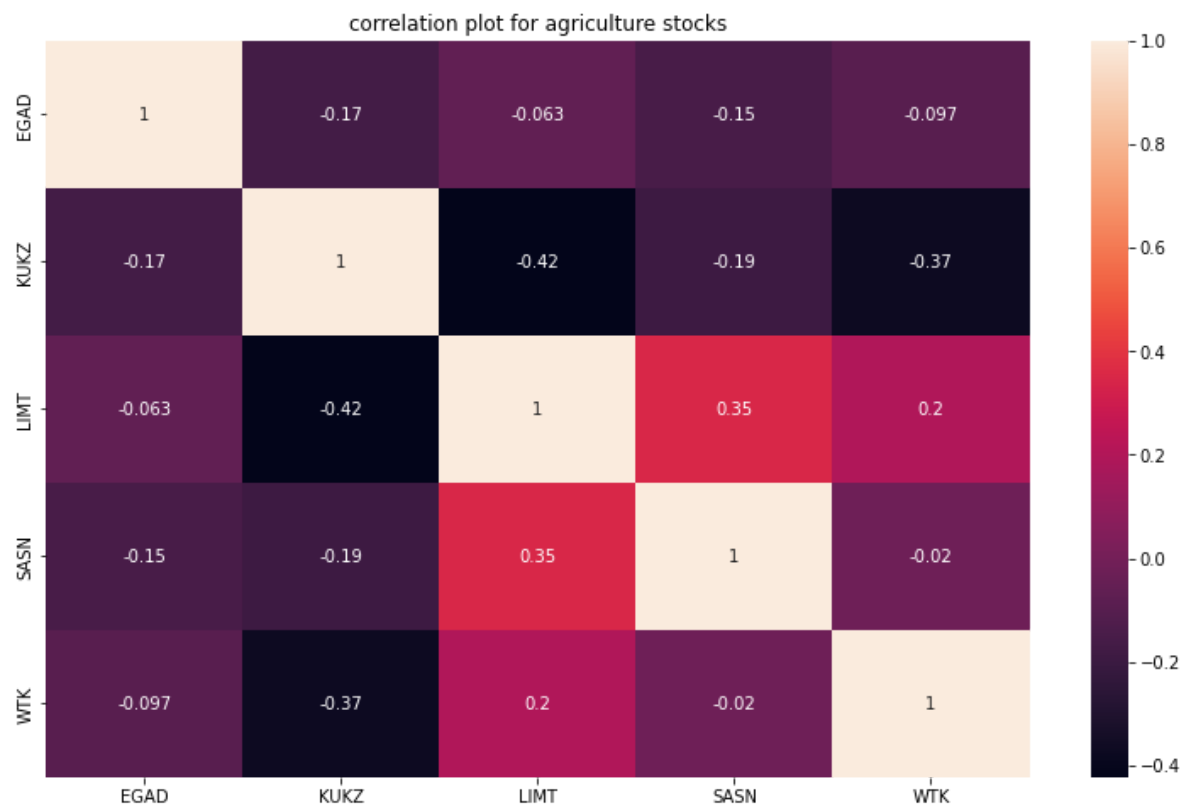
```
#correlation for Agric
```

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



In [37]:

```
#Using Seaborn
import seaborn as sns
plt.figure(figsize=(13,8))
sns.heatmap(corr_df,annot=True)
plt.title('correlation plot for agriculture stocks')
fig = plt.gcf()
fig.savefig('correlation plot for agriculture stocks.png')
plt.show()
```



In [ ]:

# Getting Started with Python

## Vehicle Analysis Project

Author : Bradley Daudi

Date: 26/3/2022

## Resources

1. [Vehicle Dataset \(https://dasclab.uonbi.ac.ke/dstraining/vehicle\\_data.csv\)](https://dasclab.uonbi.ac.ke/dstraining/vehicle_data.csv)
2. [Submission Portal \(https://dasclab.uonbi.ac.ke/analytics/projects\)](https://dasclab.uonbi.ac.ke/analytics/projects)

If you are having problems please refer to this document:

3. [Data Analysis with Python Pandas Notebook \(https://dasclab.uonbi.ac.ke/dstraining/data-analysis-with-python-pandas.html\)](https://dasclab.uonbi.ac.ke/dstraining/data-analysis-with-python-pandas.html)

## Instructions

Import all the libraries listed in the first cell. Make sure all modules are installed.

Use the provided data set to answer the following:

**Use pandas to come up with:**

1. The titles and prices of **10** Cars with highest price
2. The titles and prices of 5 Buses & Microbuses with highest price
3. The titles and prices of 5 Trucks & Trailers with highest price

## Plotting

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

### Key performance Metrics:

- Ensure all the plots have a Title
- Ensure all plots have x labels and y labels where applicable
- Your plots should be clearly visible. Change the size of your plot to a comfortable width and height.
- Save all your plots

In [63]:

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

In [64]:

```
os.listdir()
```

Out[64]:

```
['.ipynb_checkpoints',  
'01 High.mp3',  
'01-11-2020-041924The 5 AM Club - Robin Sharma.pdf',  
'2018 GTEE APPLICATION FORM (4).pdf',  
'227-Intern-Statistician.pdf',  
'61N96kYP7dL.jpg',  
'ACTIVATION.pdf',  
'ADMISSION LETTER (539).pdf',  
'BitTorrent.exe',  
'Book3.xlsx',  
'BraveBrowserSetup.exe',  
'Californication 2007 Season 3 Complete 720p AMZN WEBRip x264 [i_c]',  
'Californication 2007 Season 5 Complete 720p AMZN WEBRip x264 [i_c]',  
'Californication 2007 Season 6 Complete 720p AMZN WEBRip x264 [i_c]',  
'Californication 2007 Season 7 Complete 720p AMZN WEBRip x264 [i_c]',  
'Californication Season 1 - COMPLETE 720p HDTV x264 [MKV,AC3,5.1] Ehhhh',  
'Californication Season 2 - COMPLETE 720p HDTV x264 [MKV,AC3,5.1] Ehhhh',  
'Californication Season 4 - COMPLETE 720p HDTV x264 [MKV,AC3,5.1] Ehhhh',  
'chriska-adventur-2022.zip',  
'ChromeSetup.exe',  
'Cover letter- Joan Ngigi.pdf',  
'Cv dau latest.pdf',  
'CV-Joan Ngigi.PDF',  
'DashboardSetup.exe',  
'DAUDI_OTIENO_Elements_report (1).pdf',  
'DAUDI_OTIENO_Elements_report.pdf',  
'desktop.ini',  
'download.jfif',  
'dtb v9.8.bin',  
'epm_free_install_20211226.28161.exe',  
'epm_suite_freeB18.exe',  
'Everything Is Fucked - Mark Manson.pdf',  
'expressvpn_windows_10.21.0.9_release.exe',  
'GazettedLabslist.pdf',  
'General Experience_F.docx',  
'HDDScan.zip',  
'IMG-1628.jpg',  
'IMG-1629.jpg',  
'IMG-20210704-WA0010.jpg',  
'IMG-20211229-WA0004.jpg',  
'IMG-20211229-WA0005.jpg',  
'IMG-20211229-WA0006.jpg',  
'IMG-20211229-WA0007.jpg',  
'IMG-20211229-WA0008.jpg',  
'IMG-20211229-WA0009.jpg',  
'Joan wangui -attachment letter and admission letter.pdf',  
'ks4.021.3.10.391en_25092.exe',  
'Mo+Dao+Zu+Shi.crx',  
'MyProductList.csv',  
'mysql pic.jpg',  
'mysql-installer-web-community-8.0.28.0.msi',  
'newton raptison derivation.pdf',  
'new_daily_prices.csv',  
'ORPPJOBAPPLICATIONFORM.pdf',
```

```
'PCC-DET6Z7D7-POLICE CLEARANCE CERTIFICATE P0.pdf',  
'pdfcoffee.com_rollo-tomassi-the-rational-male-4-pdf-free.pdf',  
'pexels-aleksandar-pasaric-325185.jpg',  
'pexels-maxime-francis-2246476.jpg',  
'Pictures - Shortcut.lnk',  
'python-3.10.2-amd64 (1).exe',  
'python-3.10.2-amd64.exe',  
'R-4.0.0',  
'R-4.0.0.tar.gz',  
'rufus-3.17.exe',  
'SAINT_JHN_-_The_Best_Part_Of_Life.mp3',  
'Sample_letters_and_resume.pdf',  
'student_copy_pandas_workbook.ipynb',  
'student_workbook_stocks.ipynb',  
'sublime_text_build_4126_x64_setup.exe',  
'The Big Book Of Small House Designs 75 Award Winning Plans For Your Dream House 1250 Square Feet Or Less',  
'The.Boys.2019.S01.COMPLETE.720p.AMZN.WEBRip.x264-GalaxyTV[TGx]',  
'The.Boys.S02.COMPLETE.720p.AMZN.WEBRip.x264-GalaxyTV[TGx]',  
'The.Kings.Man.2021.HDRip.XviD.B4ND1T69',  
'TOR-FOR-KELCOP-VACANCIES.pdf',  
'Unconfirmed 803033.crdownload',  
'vehicle_data (1).csv',  
'vehicle_data.csv',  
'vehicle_dataset_project (1).ipynb',  
'vehicle_dataset_project.ipynb',  
'video.mp4',  
'WD Backup_1.9.7435.38388.zip',  
'WindowsPCHealthCheckSetup (1).msi',  
'WindowsPCHealthCheckSetup.msi',  
'winrar-x64-602.exe',  
'_Getintopc.com_Adobe_Lightroom_Classic_2022_v11.0.0.10.rar',  
'_igetintopc.com_Adobe_Premiere_Pro_2022_x64 (1).rar']
```

**vehicle\_data.csv should be listed in your output from the above cell**

In [65]:

```
df = pd.read_csv('vehicle_data (1).csv')
df.head(10)
```

Out[65]:

	title	category	region	parent_region	condition	attrs	brand	color	r
0	Toyota Land Cruiser Prado 2016 Black	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults	Toyota	Black	C
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	I
2	Clean NV300 Caravan 2014 Model Dielsel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used	Nissan	Nissan	NaN	Ca (L
3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	(
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	
5	Mitsubishi Delica 2013 White	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults, Unpainted	Mitsubishi	White	
6	New Toyota Premio 2013 Red	Cars	Mvita	Mombasa	Brand New	No faults, First registration	Toyota	Red	F
7	Toyota Sienta 2014 1.5 AWD Gray	Cars	Ganjoni	Mombasa	Foreign Used	No faults	Toyota	Gray	;
8	BMW X4 2015 xDrive35i Black	Cars	Mombasa CBD	Mombasa	Foreign Used	No faults	BMW	Black	
9	Mitsubishi Outlander 2015 White	Cars	Lavington	Nairobi	Foreign Used	Unpainted, Original parts, No faults	Mitsubishi	White	Out

In [70]:

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

Out[70]:

	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

In [73]:

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

Out[73]:

	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

In [75]:

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

Out[75]:

	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

In [77]:

```
df['brand'].value_counts()
```

Out[77]:

Toyota	82
Nissan	33
Mitsubishi	32
Mazda	26
Subaru	22
Volkswagen	21
Isuzu	19
Honda	17
BMW	17
Mercedes-Benz	9
Suzuki	5
Lexus	4
Tata	2
Volvo	2
Ashok Leyland	2
Land Rover	1
Shacman	1
Other	1
Hyundai	1
Daihatsu	1
Audi	1

Name: brand, dtype: int64

In [78]:

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

Out[78]:

Toyota	82
Nissan	33
Mitsubishi	32
Mazda	26
Subaru	22
Volkswagen	21
Isuzu	19
Honda	17
BMW	17
Mercedes-Benz	9

Name: brand, dtype: int64

In [79]:

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

In [80]:

```
plt.figure(figsize=(12,8))
plt.title("Top 10 brands according to dataset", fontsize=14, fontweight='bold')
top_10.plot.bar()
plt.xlabel('Name of brand',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-10-brands.png')

plt.show()
```

