Getting Started with Python

Vehicle Analysis Project

Author : PURITY KAMAU

Date: 17/3/2022

Resources

1. #### Vehicle Dataset

2. #### Submission Portal

If you are having problems please refer to this document:

1. #### Data Analysis with Python Pandas Notebook

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 [74]:
```
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
os.listdir()
```

Out[2]:
```
['.ipynb_checkpoints',
 'cleaned_stock.csv',
 'clean_stock_data.csv',
 'clean_stock_prices.csv',
 'desktop.ini',
 'microsoft.microsoftskydrive_8wekyb3d8bbwe!App',
 'new_daily_prices.csv',
 'project-time-series-workbook.ipynb',
 'R-4.1.3-win.exe',
 'receipt.pdf',
 'RStudio-2022.02.0-443.exe',
 'student_copy_pandas_workbook.ipynb',
 'student_workbook_stocks-Copy1.ipynb',
 'student_workbook_stocks.ipynb',
 'student_workbook_stocks.py',
 'Telegram Desktop',
 'Untitled.ipynb',
 'Untitled1.ipynb',
 'vehicle_data (1).csv',
 'vehicle_data (2).csv',
 'vehicle_data (3).csv',
 'vehicle_data (4).csv',
 'vehicle_data (5).csv',
 'vehicle_data (6).csv',
 'vehicle_data (7).csv',
 'vehicle_data.csv',
 'vehicle_dataset_project.ipynb']
```

vehicle_data.csv should be listed in your output from the above cell

In [75]:
```
df = pd.read_csv('vehicle_data.csv')
df.head()
```

Out[75]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom	mileage
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
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	Demio	2014.0	92000.0
2	Clean NV300 Caravan 2014 Model Dielsel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used	Nissan	Nissan	NaN	Caravan (Urvan)	2014.0	180000.0
3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	Crown	2014.0	75000.0
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0	58000.0

As an example I have shown the top 10 most expensive vehicles that are in _parent*region* Mombasa

In [48]:
```
# filter only rows that have Mombasa as their region
df['parent_region'] == 'Mombasa'
```

Out[48]:
```
0       True
1      False
2      False
3      False
4       True
       ...
295     True
296     True
297     True
298    False
299     True
Name: parent_region, Length: 300, dtype: bool
```

In [49]:
```
mask = df['parent_region'] == 'Mombasa'
```

In [50]:
```
df
```

Out[50]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom
0	Toyota Land Cruiser Prado 2016 Black	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults	Toyota	Black	Land Cruiser Prado	2016.0
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	Demio	2014.0
2	Clean NV300 Caravan 2014 Model Dielsel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used	Nissan	Nissan	NaN	Caravan (Urvan)	2014.0
3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	Crown	2014.0
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0
...	...	...	...	...	...	...	...	...	...	...
295	Toyota Allion 2008 Silver	Cars	Ganjoni	Mombasa	Kenyan Used	No faults	Toyota	Silver	Allion	2008.0
296	Mitsubishi Hd	Trucks & Trailers	Kisauni	Mombasa	Used	Used	Mitsubishi	NaN	NaN	2006.0
297	Mitsubishi Fuso Refrigerated	Trucks & Trailers	Kisauni	Mombasa	Used	Used	Mitsubishi	NaN	Canter	2014.0
298	Toyota Ractis 2009 Black	Cars	Ridgeways	Nairobi	Kenyan Used	First owner, No faults, Original parts	Toyota	Black	Ractis	2009.0

299	Subaru Forester 2015 Matt Black	Cars	Mombasa CBD	Mombasa	Foreign Used	Unpainted, Original parts, First registration	Subaru	Matt Black	Forester	2015.0	

300 rows × 18 columns

In [51]:
```
# all the rows in the dataFrame that have parent_region Mombasa
mombasa_df = df[mask].copy()
mombasa_df.head()
```

Out[51]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom	mileage
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
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0	58000.0
5	Mitsubishi Delica 2013 White	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults, Unpainted	Mitsubishi	White	Delica	2013.0	88000.0
6	New Toyota Premio 2013 Red	Cars	Mvita	Mombasa	Brand New	No faults, First registration	Toyota	Red	Premio	2013.0	45000.0
7	Toyota Sienta 2014 1.5 AWD Gray	Cars	Ganjoni	Mombasa	Foreign Used	No faults	Toyota	Gray	Sienta	2014.0	51000.0

To get the highest price I will use the `nlargest` function

In [52]:
```
# top 10 vehicles with highest price
mombasa_df.nlargest(10,'price')
```

Out[52]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom
22	Lexus RX 2016 Black	Cars	Mombasa CBD	Mombasa	Foreign Used	No faults	Lexus	Black	RX	2016.0
224	Toyota Hilux 2016 Black	Cars	Mombasa CBD	Mombasa	Foreign Used	First registration	Toyota	Black	Hilux	2016.0
0	Toyota Land Cruiser Prado 2016 Black	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults	Toyota	Black	Land Cruiser Prado	2016.0
53	Toyota Land Cruiser Prado 2015	Cars	Mvita	Mombasa	Foreign Used	No faults	Toyota	Brown	Land Cruiser Prado	2015.0

	2.7 VVT-i Brown									
241	BMW X5 2015 White	Cars	Mombasa CBD	Mombasa	Foreign Used	First registration	BMW	White	X5	2015.0
8	BMW X4 2015 xDrive35i Black	Cars	Mombasa CBD	Mombasa	Foreign Used	No faults	BMW	Black	X4	2015.0
13	BMW 520i 2014 Black	Cars	Mombasa CBD	Mombasa	Foreign Used	First registration	BMW	Black	520i	2014.0
73	Toyota Land Cruiser Prado 2014 2.7 VVT-i Gold	Cars	Mombasa CBD	Mombasa	Foreign Used	First owner, First registration	Toyota	Gold	Land Cruiser Prado	2014.0
220	Volkswagen Touareg 2015 TDI Executive AWD 4MOT...	Cars	Tudor	Mombasa	Foreign Used	No faults	Volkswagen	Black	Touareg	2015.0
276	Lexus NX 2014 Red	Cars	Mombasa CBD	Mombasa	Foreign Used	No faults	Lexus	Red	NX	2014.0

To get only the `titles`

In [53]:
```
# top 10 vehicles with highest price
mombasa_df.nlargest(10,'price')[['title','category','price']]
```

Out[53]:

	title	category	price
22	Lexus RX 2016 Black	Cars	14500000
224	Toyota Hilux 2016 Black	Cars	9000000
0	Toyota Land Cruiser Prado 2016 Black	Cars	6500000
53	Toyota Land Cruiser Prado 2015 2.7 VVT-i Brown	Cars	6500000
241	BMW X5 2015 White	Cars	6300000
8	BMW X4 2015 xDrive35i Black	Cars	5800000
13	BMW 520i 2014 Black	Cars	5700000
73	Toyota Land Cruiser Prado 2014 2.7 VVT-i Gold	Cars	5600000
220	Volkswagen Touareg 2015 TDI Executive AWD 4MOT...	Cars	5500000
276	Lexus NX 2014 Red	Cars	5400000

In [97]:
```
mombasa_df
```

Out[97]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom
0	Toyota Land Cruiser Prado 2016 Black	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults	Toyota	Black	Land Cruiser Prado	2016.0

4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0
5	Mitsubishi Delica 2013 White	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults, Unpainted	Mitsubishi	White	Delica	2013.0
6	New Toyota Premio 2013 Red	Cars	Mvita	Mombasa	Brand New	No faults, First registration	Toyota	Red	Premio	2013.0
7	Toyota Sienta 2014 1.5 AWD Gray	Cars	Ganjoni	Mombasa	Foreign Used	No faults	Toyota	Gray	Sienta	2014.0
...	...	...	...	...	...	...	...	...	...	...
294	Subaru Outback 2014 White	Cars	Mvita	Mombasa	Foreign Used	No faults	Subaru	White	Outback	2014.0
295	Toyota Allion 2008 Silver	Cars	Ganjoni	Mombasa	Kenyan Used	No faults	Toyota	Silver	Allion	2008.0
296	Mitsubishi Hd	Trucks & Trailers	Kisauni	Mombasa	Used	Used	Mitsubishi	NaN	NaN	2006.0
297	Mitsubishi Fuso Refrigerated	Trucks & Trailers	Kisauni	Mombasa	Used	Used	Mitsubishi	NaN	Canter	2014.0
299	Subaru Forester 2015 Matt Black	Cars	Mombasa CBD	Mombasa	Foreign Used	Unpainted, Original parts, First registration	Subaru	Matt Black	Forester	2015.0

147 rows × 18 columns

In [79]:
```
df = pd.read_csv('vehicle_data.csv')
df.head()
```

Out[79]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom	mileage
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
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	Demio	2014.0	92000.0
2	Clean NV300 Caravan 2014 Model Dielsel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used	Nissan	Nissan	NaN	Caravan (Urvan)	2014.0	180000.0

3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	Crown	2014.0	75000.0
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0	58000.0

In [80]:
```
df=df[df['category']=='Cars']
df.nlargest(5,'price')[['title','category','price']]
```

Out[80]:

	title	category	price
22	Lexus RX 2016 Black	Cars	14500000
265	New Hyundai Palisade 2021 White	Cars	9500000
224	Toyota Hilux 2016 Black	Cars	9000000
156	Toyota Land Cruiser 2010 4.6 V8 ZX Black	Cars	8799999
249	Toyota Land Cruiser 2014 4.6 V8 ZX Black	Cars	8199999

In [86]:
```
df = pd.read_csv('vehicle_data.csv')
df.head()
```

Out[86]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom	mileage
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
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	Demio	2014.0	92000.0
2	Clean NV300 Caravan 2014 Model Dielsel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used	Nissan	Nissan	NaN	Caravan (Urvan)	2014.0	180000.0
3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	Crown	2014.0	75000.0
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0	58000.0

In [82]:
```
df=df[df['category']=='Buses & Microbuses']
df.nlargest(5,'price')[['title','category','price']]
```

Out[82]:

	title	category	price
148	Mazda Bongo	Buses & Microbuses	11200000

221	Selling Buses In Mombasa Town	Buses & Microbuses	5200000
174	Roller Coaster	Buses & Microbuses	4900000
211	Toyota Coaster 2014 White	Buses & Microbuses	4300000
268	Toyota Hiace 2015 White	Buses & Microbuses	3800000

In [84]:
```
df = pd.read_csv('vehicle_data.csv')
df.head()
```

Out[84]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom	mileage
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
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	Demio	2014.0	92000.0
2	Clean NV300 Caravan 2014 Model Dielsel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used	Nissan	Nissan	NaN	Caravan (Urvan)	2014.0	180000.0
3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	Crown	2014.0	75000.0
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0	58000.0

In [85]:
```
df=df[df['category']=='Trucks & Trailers']
df.nlargest(5,'price')[['title','category','price']]
```

Out[85]:

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

The above output is what the question is asking for. So take a screenshot.

In [ ]:
```

```

Plotting

I will demonstrate how to solve the plotting challenge using the following question:

Use **matplotlib** to come up with a plot indicating the **top 5 regions** that we have in the vehicle_dataset

In [10]:
```
# get number of rows with same region
df['region'].value_counts()
```

Out[10]:
```
Mombasa CBD                     92
Mvita                           28
Nairobi Central                 27
Kilimani                        23
Lavington                       16
Ridgeways                       15
Tudor                           13
Karen                            8
Langata                          7
Nairobi                          6
Kileleshwa                       4
Ganjoni                          4
Kasarani                         3
Nyali                            3
Embakasi                         3
Thome                            3
Kisauni                          3
Parklands/Highridge              3
Donholm                          3
Mombasa Road                     2
Nakuru Town West                 2
Westlands                        2
Runda                            2
Kikuyu                           2
Ruiru                            2
Mbaraki                          2
Nakuru Town East                 2
Makadara                         1
Ziwa la Ngombe                   1
Shanzu                           1
Imara Daima                      1
Majengo                          1
Kiambaa                          1
Ngara                            1
Ngong                            1
Kiambu / Kiambu                  1
Machakos Town                    1
Woodley/Kenyatta Golf Course     1
Thika                            1
Changamwe                        1
Kisumu Central                   1
Ruai                             1
Eldoret CBD                      1
Syokimau                         1
Kerugoya                         1
Municipality                     1
Roysambu                         1
Name: region, dtype: int64
```

In [11]:
```
# grab the top 5
df['region'].value_counts()[:5]
```

Out[11]:
```
Mombasa CBD        92
Mvita              28
Nairobi Central    27
Kilimani           23
```

```
Lavington          16
Name: region, dtype: int64
```

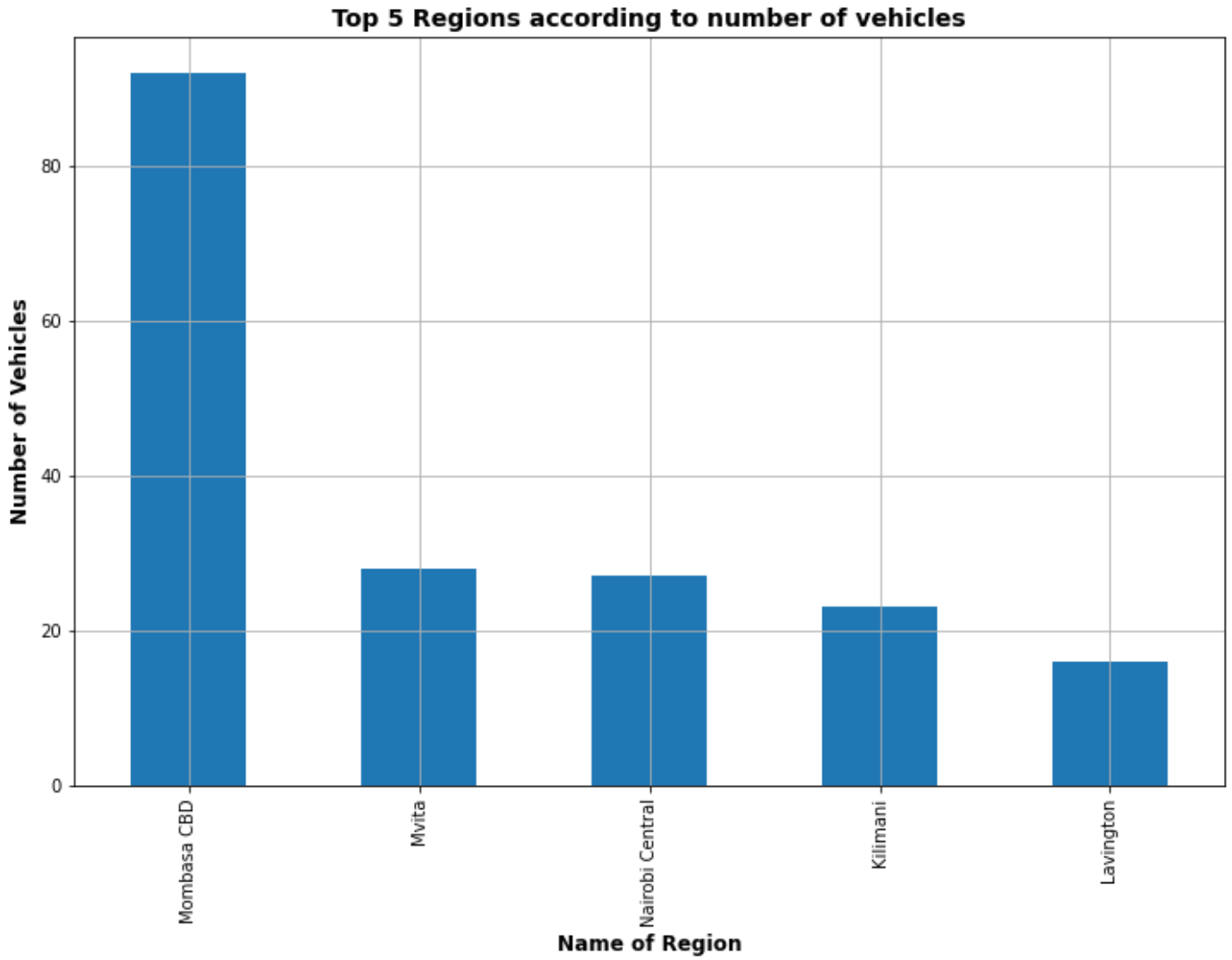
In [12]:
```
# make it a variable
top_5 = df['region'].value_counts()[:5]
```

Now to create a bar plot of the top 5 regions

In [13]:
```
plt.figure(figsize=(12,8))
plt.title("Top 5 Regions according to number of vehicles", fontsize=14, fontweight='bold
top_5.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()
```



In [87]:
```
df = pd.read_csv('vehicle_data.csv')
df.head()
```

Out[87]:

	title	category	region	parent_region	condition	attrs	brand	color	model	yom	mileage
0	Toyota Land Cruiser Prado	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults	Toyota	Black	Land Cruiser Prado	2016.0	87000.0

	2016 Black											
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	Demio	2014.0	92000.0
2	Clean NV300 Caravan 2014 Model Dielsel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used	Nissan	Nissan	NaN	Caravan (Urvan)	2014.0	180000.0
3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	Crown	2014.0	75000.0
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit	2014.0	58000.0

In [88]:
```
df['brand'].value_counts()[:10]
```

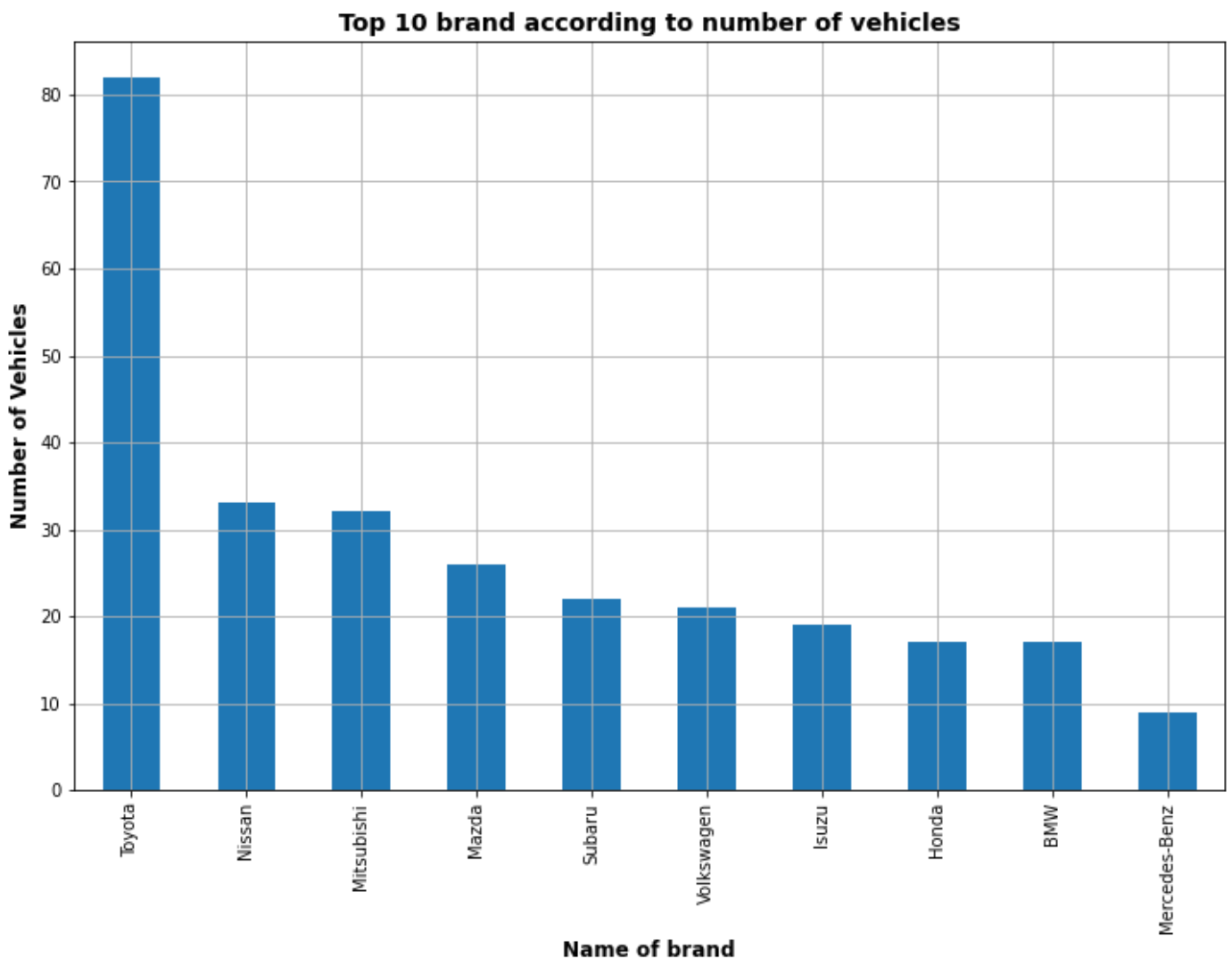
Out[88]:
```
Toyota           82
Nissan           33
Mitsubishi       32
Mazda            26
Subaru           22
Volkswagen       21
Isuzu            19
Honda            17
BMW              17
Mercedes-Benz     9
Name: brand, dtype: int64
```

In [89]:
```
top_10 = df['brand'].value_counts()[:10]
```

In [90]:
```
plt.figure(figsize=(12,8))
plt.title("Top 10 brand according to number of vehicles", 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-regions.png')

plt.show()
```

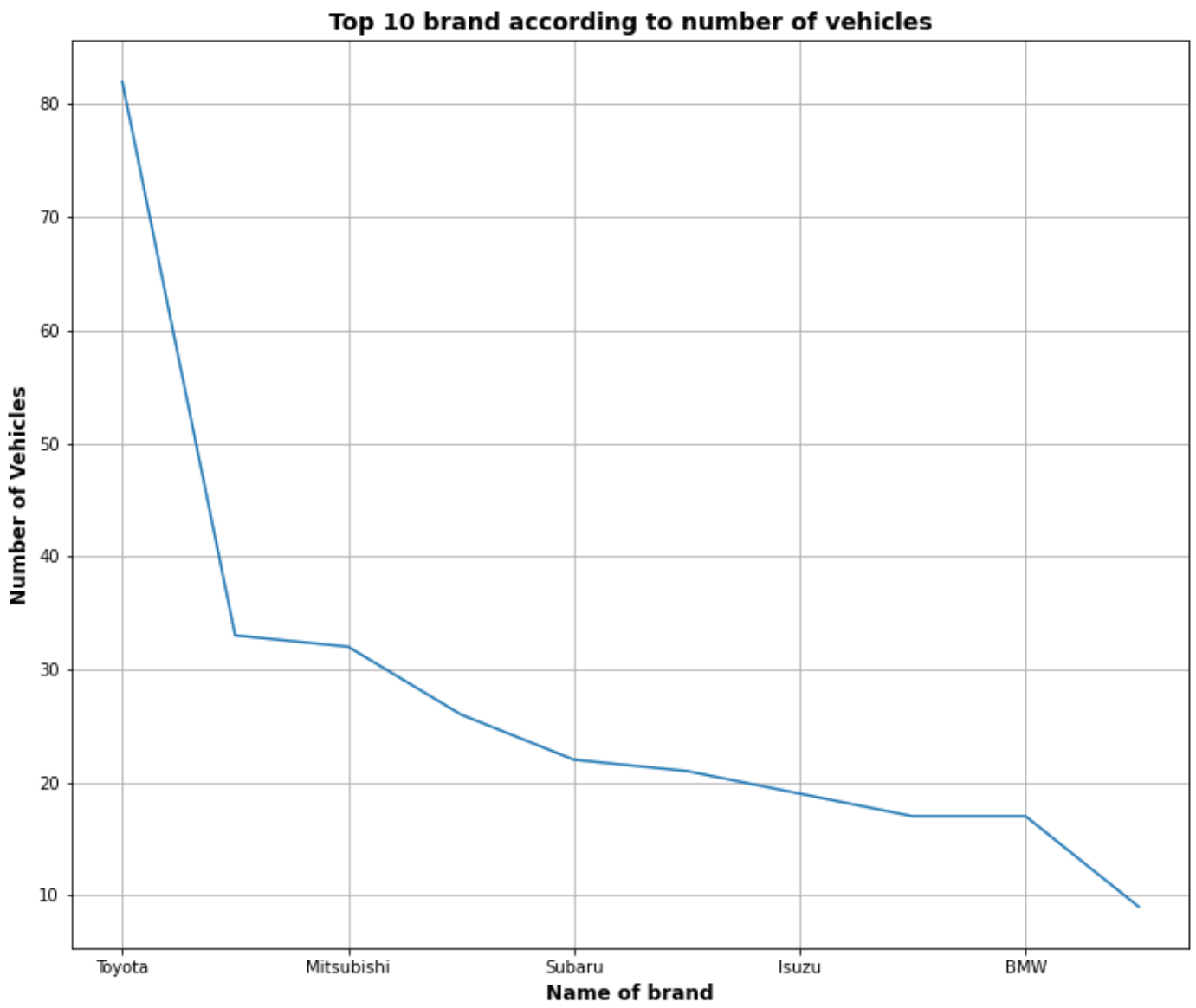


In [91]:

```
plt.figure(figsize=(12,10))
plt.title("Top 10 brand according to number of vehicles", fontsize=14, fontweight='bold'
top_10.plot.line()
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-regions.png')

plt.show()
```

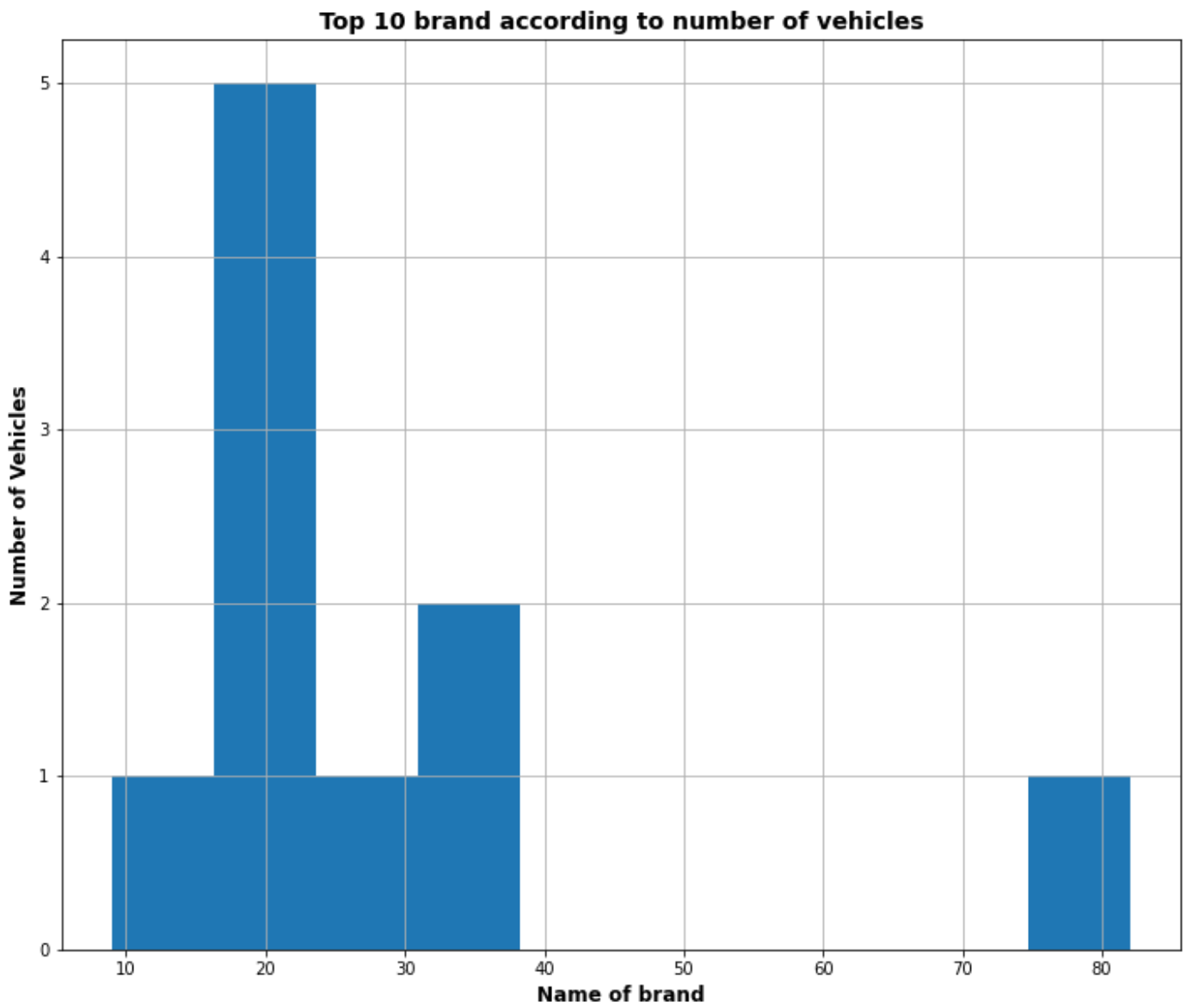


In [92]:

```
plt.figure(figsize=(12,10))
plt.title("Top 10 brand according to number of vehicles", fontsize=14, fontweight='bold'
top_10.plot.hist()
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-regions.png')

plt.show()
```

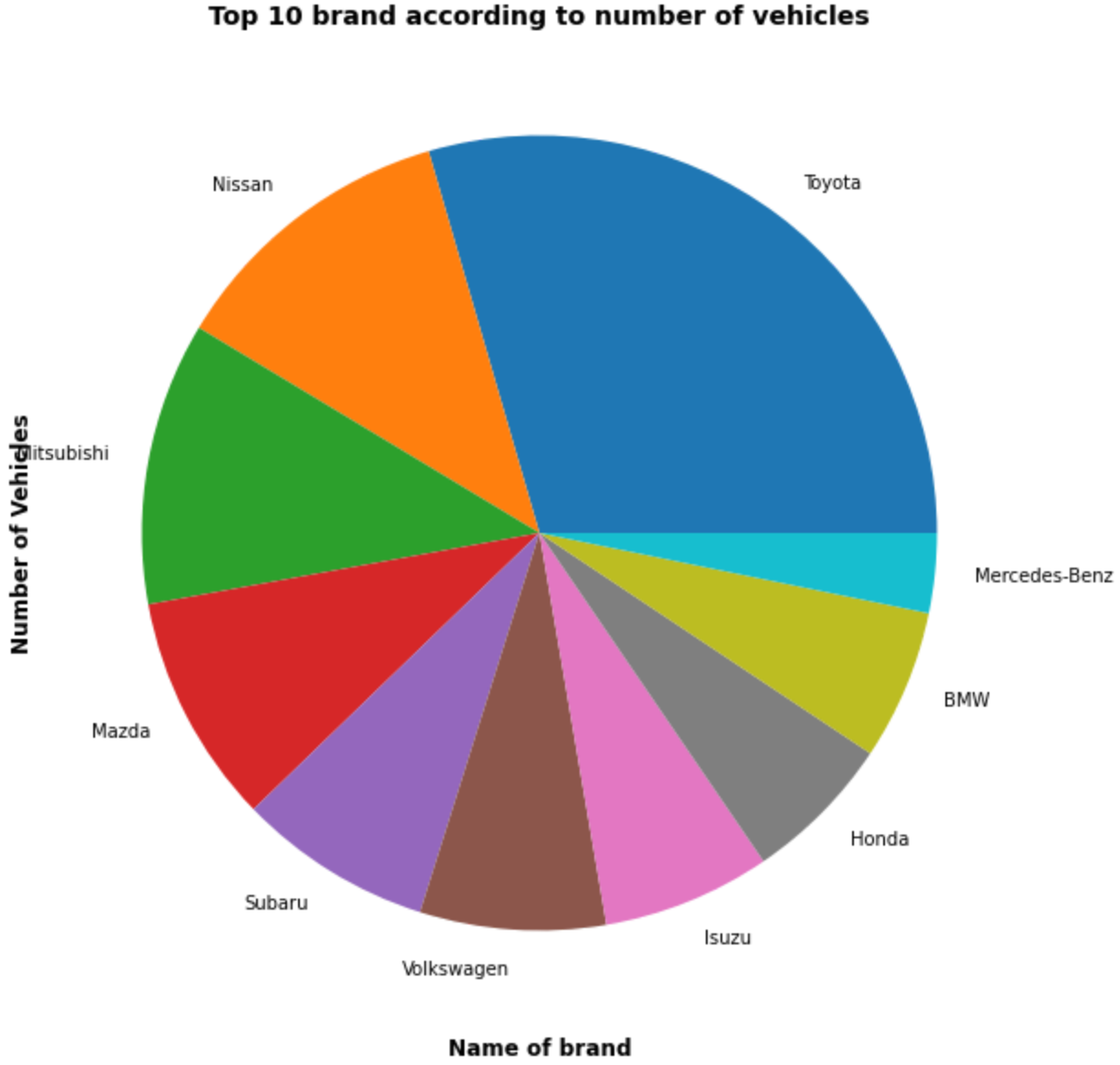


In [93]:

```
plt.figure(figsize=(12,10))
plt.title("Top 10 brand according to number of vehicles", fontsize=14, fontweight='bold'
top_10.plot.pie()
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-regions.png')

plt.show()
```


Top 10 brand according to number of vehicles
Toyota
Nissan
Mitsubishi
Mercedes-Benz
BMW
Mazda
Honda
Subaru
Isuzu
Volkswagen
Number of Vehicles
Name of brand