

Getting Started with Python

Vehicle Analysis Project

Author : Meshack Muli

Date: 27/3/2022

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

```
In [3]: ▶ os.listdir()
```

```
Out[3]: ['.ipynb_checkpoints',
'0770-digital-painting-with-krita.pdf',
'Baker_PlantingTreesCombatDrought_2021.md',
'bandicut-setup.exe',
'binance-setup.exe',
'BrowserSetup_FJ0TIZC.exe',
'BrowserSetup_XBQX4mM.exe',
'BrowserSetup_XORIP0m.exe',
'Business-Economics.docx',
'cleaned_stock1.csv',
'desktop.ini',
'EOSC 310-Assignment 3-Template-UPDATED (1) (1).docx',
'EOSC 310-Assignment 3-Template-UPDATED (1) (2).docx',
'EOSC 310-Assignment 3-Template-UPDATED (1).docx',
'exness4setup.exe',
'inkscape-1.1.2_2022-02-05_b8e25be833-x64.exe',
'install_virtualdj_2021_b6800_pc.msi',
'Internal Job Opportunities.pdf',
'JUDITH WAYUA CV.docx',
'krita-x64-5.0.2-setup.exe',
'Learn Python in One Day and Learn It Well_ Python for Beginners with Hand
s-on Project. The only book you need to start coding in Python immediately
( PDFDrive ).epub',
'new_daily_prices1.csv',
'NiceHashQuickMinerV1001.exe',
'project.py',
'Python Exercises ( PDFDrive ).pdf',
'python-3.10.2-amd64.exe',
'python_project_vehicle_dataset-Copy1.ipynb',
'SpotifySetup (1).exe',
'SpotifySetup.exe',
'student_copy_pandas_workbook.ipynb',
'student_workbook_stocks.ipynb',
'Telegram Desktop',
'TradingView (1).appinstaller',
'TradingView (2).appinstaller',
'TradingView.appinstaller',
'tsetup-x64.3.4.3.exe.4bm8ggg.partial',
'tsetup-x64.3.4.8.exe',
'Tyler.Perry.s.If.Loving.You.Is.Wrong.S01E01.(WEB-DL).(HDEpic.com).mp4',
'Tyler.Perry.s.If.Loving.You.Is.Wrong.S01E02.(WEB-DL).(HDEpic.com).mp4',
'Tyler.Perry.s.If.Loving.You.Is.Wrong.S01E03.(WEB-DL).(HDEpic.com).mp4',
'Untitled.ipynb',
'vehicle_data1.csv',
'VLC media player - Installer _hsGz (1).exe',
'VLC media player - Installer _hsGz.exe',
'WhatsAppSetup.exe',
'WindowsPCHealthCheckSetup.msi']
```

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

```
In [5]: ▶ df = pd.read_csv('vehicle_data1.csv')
df.head()
```

Out[5]:

	title	category	region	parent_region	condition	attrs	brand	color	model
0	Toyota Land Cruiser Prado 2016 Black	Cars	Mvita	Mombasa	Foreign Used	First registration, No faults	Toyota	Black	Land Cruiser Prado
1	Mazda Demio 2014 Brown	Cars	Langata	Nairobi	Foreign Used	First owner, No faults	Mazda	Brown	Demio
2	Clean NV300 Caravan 2014 Model Diesel 16 Seater	Buses & Microbuses	Kilimani	Nairobi	Foreign Used		Nissan	NaN	Caravan (Urvan)
3	Toyota Crown 2014 Pearl	Cars	Kilimani	Nairobi	Foreign Used	No faults	Toyota	Pearl	Crown
4	Honda Fit 2014 Black	Cars	Mvita	Mombasa	Foreign Used	No faults	Honda	Black	Fit

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

```
In [ ]: ▶
```

```
In [6]: ▶ mask = df['category'] == 'Trucks & Trailers'
```

```
In [7]: # all the rows in the dataframe that have parent_region Mombasa
Buses_df = df[mask].copy()
Buses_df.head()
```

Out[7]:

	title	category	region	parent_region	condition	attrs	brand	color	mod
14	Very Clean Isuzu FRR Truck 2015 Model	Trucks & Trailers	Thome	Nairobi	Used	Used	Isuzu	White	SERIE
20	Truck Lorry	Trucks & Trailers	Shanzu	Mombasa	Used	Used	Ashok Leyland	NaN	192
27	Volvo NL12 for Sale	Trucks & Trailers	Embakasi	Nairobi	Used	Used	Volvo	NaN	Na
30	Mitsubishi Fuso Canter	Trucks & Trailers	Mombasa CBD	Mombasa	Used	Used	Mitsubishi	NaN	Na
36	Isuzu Elf, Year 2014 manual Transmission	Trucks & Trailers	Mombasa CBD	Mombasa	Brand New	Brand New	Isuzu	NaN	Na

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

```
In [17]: # top 10 vehicles with highest price
Buses_df.nlargest(5, 'price')
```

Out[17]:

	title	category	region	parent_region	condition	attrs	brand	color	mo
195	Mercedes-Benz Actros	Trucks & Trailers	Thome	Nairobi	Used	Used	Mercedes-Benz	NaN	N
222	Tata Signa LPK-1618 Tipper 10 Ton	Trucks & Trailers	Nairobi Central	Nairobi	Brand New	Brand New	Tata	NaN	M&H Ri Tru
103	Shacman F2000 Tipper	Trucks & Trailers	Municipality	Meru	Used	Used	Shacman	NaN	N
176	Isuzu Forward 7 Tonne Freezer	Trucks & Trailers	Tudor	Mombasa	Used	Used	Isuzu	NaN	SERI
62	Isuzu Elf, Year 2015 Manual	Trucks & Trailers	Mombasa CBD	Mombasa	Brand New	Brand New	Isuzu	NaN	N

To get only the titles

```
In [19]: Buses_df.nlargest(5, 'price')[['title', 'category', 'price']]
```

Out[19]:

	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

```
In [ ]:
```

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 [20]: ▶ # get number of rows with same region  
df['brand'].value_counts()
```

```
Out[20]: 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 [21]: ▶ # grab the top 5  
df['brand'].value_counts()[:10]
```

```
Out[21]: 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
```

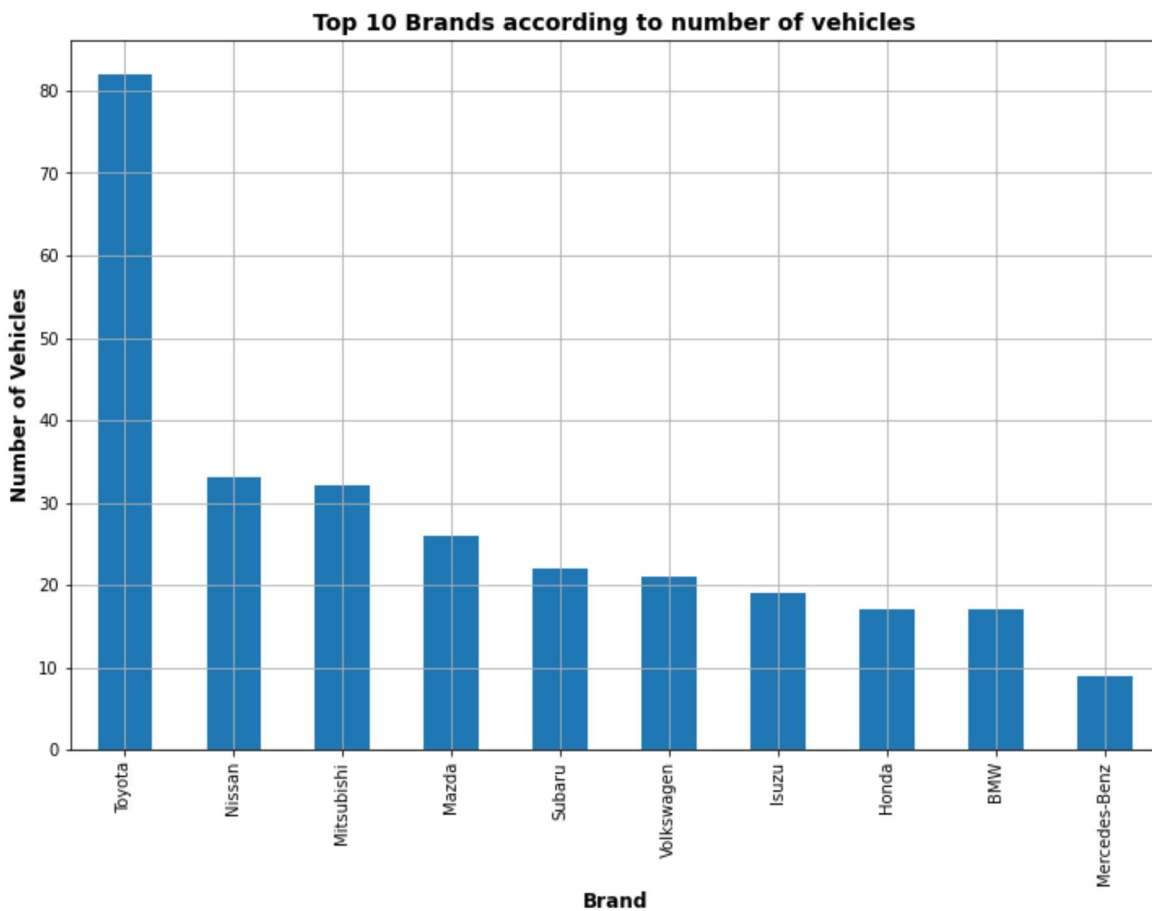
```
In [22]: ▶ # make it a variable  
top_5 = df['brand'].value_counts()[:10]
```

Now to create a bar plot of the top 5 regions

```
In [23]: ▶ plt.figure(figsize=(12,8))
plt.title("Top 10 Brands according to number of vehicles", fontsize=14, fontw
top_5.plot.bar()
plt.xlabel('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()
```



```
In [ ]: ▶
```

