



Session 2 / starter

Python refresh

No Python teaching here



DISCLAIMER : not my preferred language

FREE ebooks

A collection of free ebooks,
without stealing royalties

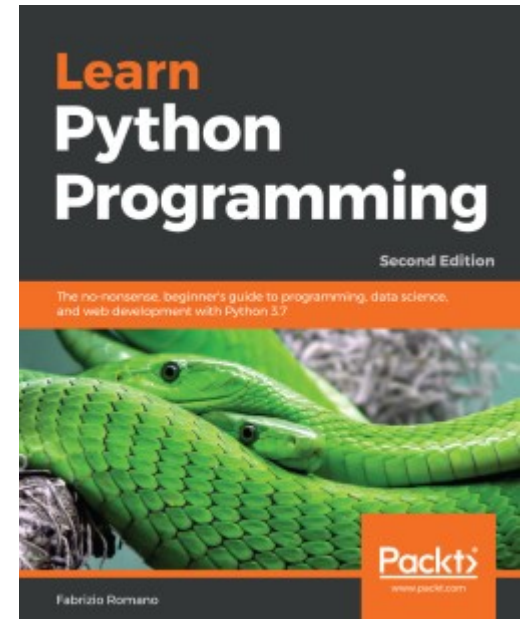
See

<https://www.packtpub.com/free-learning>

PLUS

Everyday a book is offered for the day

Disclaimer : its my publisher



The internet is full of resources about Python

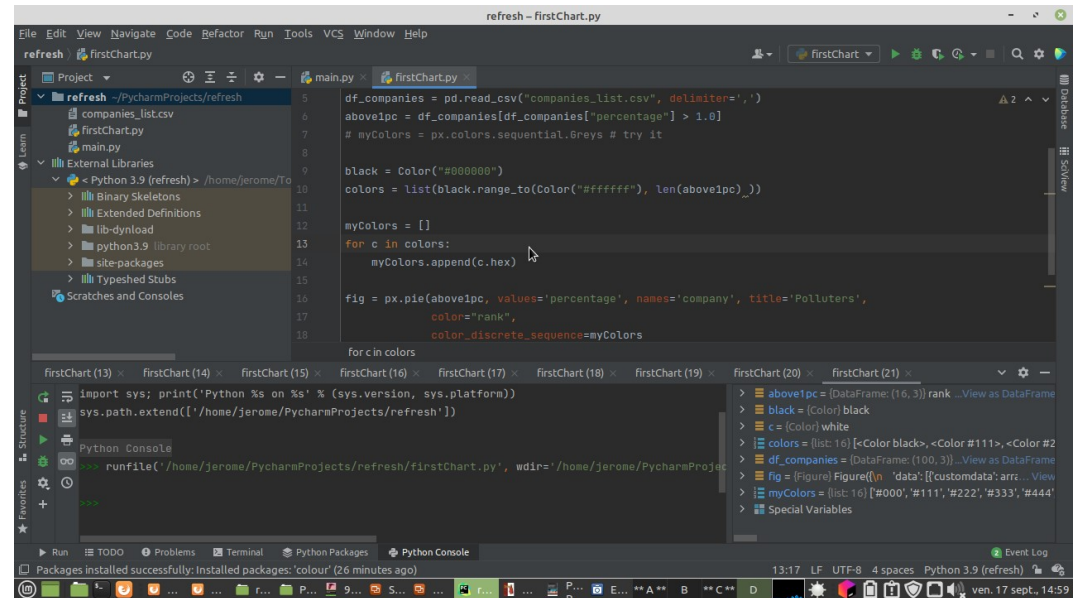
Many are free



Free IDE

JetBrains' PyCharm is licensed for free to students

**Use your conda
Interpreter with it**



Python files end with .py
Indentation is meaningful

« def » is for defining functions

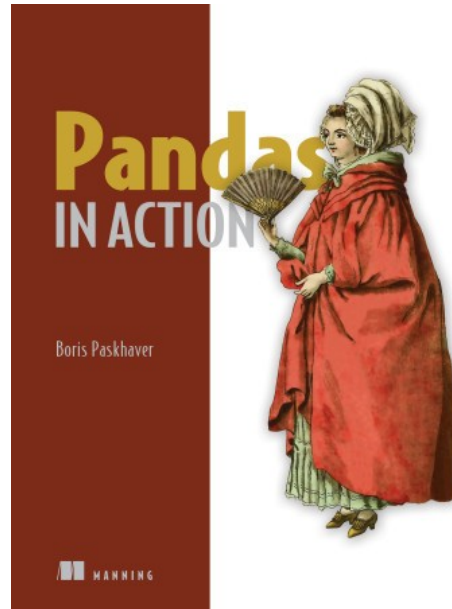
No ; at line ends

Unneeded spaces are marked in IDE

(\$€€) Recommended readings (non free)

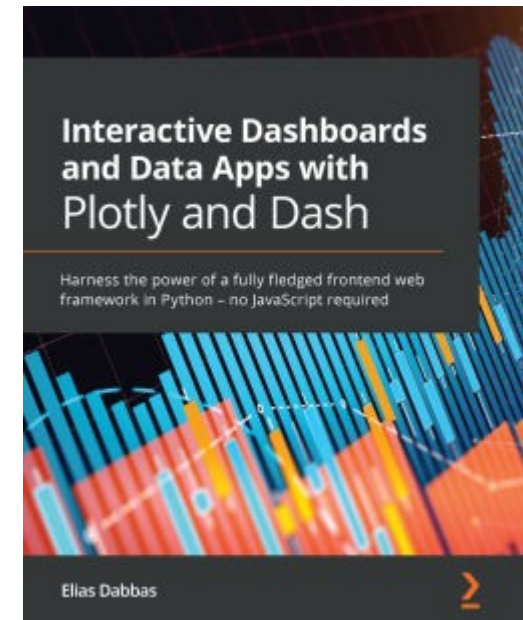
« **Pandas in Action** »
by **Boris Paskhaver**

August 2021
ISBN 9781617297434



« **Interactive Dashboards and Data Apps with Plotly and Dash** »
By **Elias Dabbas @ Packt publishing**

ISBN 9781800568914



Pandas

When working with **tabular data**, such as data stored in spreadsheets or databases, pandas is the right tool for you. pandas will help you to **explore**, **clean**, and **process** your data.

In pandas, a data table is called a **DataFrame**.

Panda documentation is great

https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html

Read a CSV file into a dataframe

```
import pandas as pd
```

```
df_companies = pd.read_csv("companies_list.csv",  
delimiter=',')  
print(df_companies.head())
```

Filter a dataframe

```
import pandas as pd
```

```
df_companies = pd.read_csv("companies_list.csv",  
delimiter=',')
```

```
above1pc = df_companies[df_companies["percentage"] >  
1.0]
```

```
print( above1pc )
```

Use a df with plotly

```
import pandas as pd
import plotly.express as px

df_companies = pd.read_csv("companies_list.csv", delimiter=',')
above1pc = df_companies[df_companies["percentage"] > 1.0]

fig = px.pie(above1pc, values='percentage', names='company',
title='Polluters')
fig.show()
```

This should open a web browser

Colors (built in scale) 1/2

```
import pandas as pd
import plotly.express as px
from colour import Color

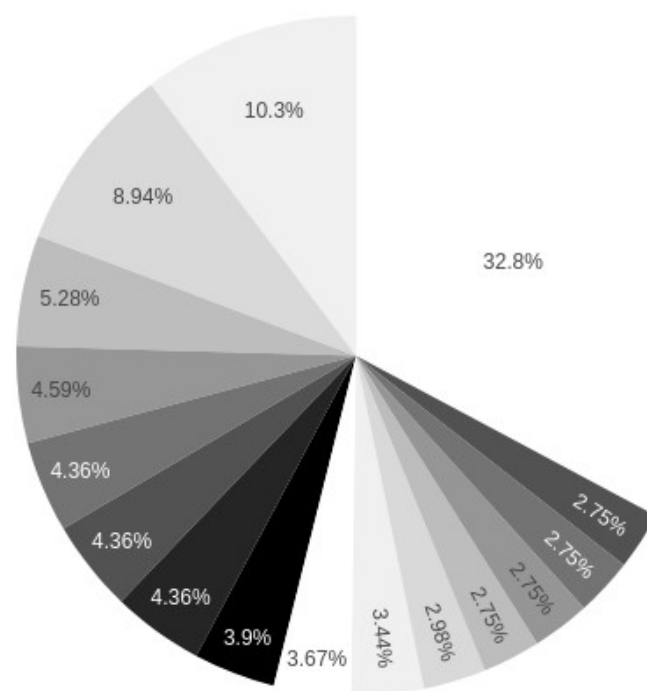
df_companies = pd.read_csv("companies_list.csv", delimiter=',')
above1pc = df_companies[df_companies["percentage"] > 1.0]

myColors = px.colors.sequential.Greys # try it

fig = px.pie(above1pc, values='percentage', names='company',
             title='Polluters',
             color="rank",
             color_discrete_sequence=myColors
             )
fig.show()
```

Colors (built in scale) 1/2

Polluters



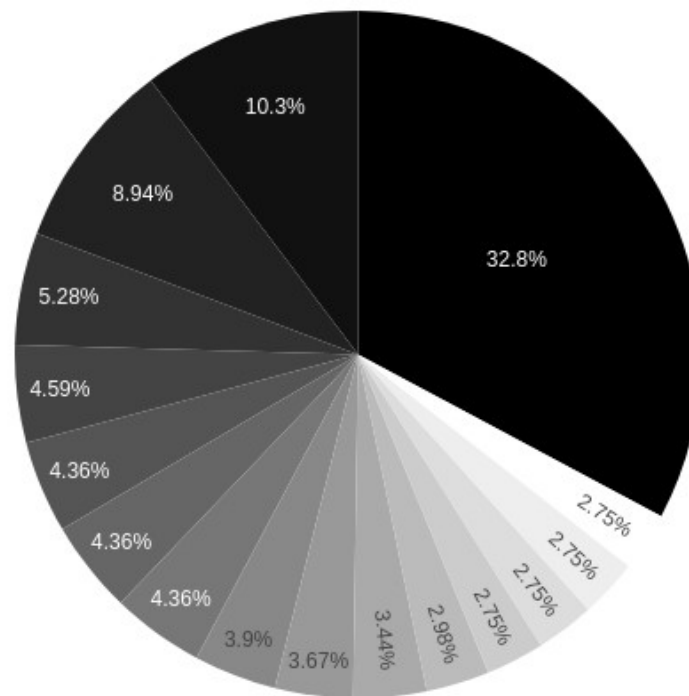
China (Coal)
Saudi Arabian Oil Company (Aramco)
Gazprom OAO
National Iranian Oil Co
ExxonMobil Corp
Coal India
Petroleos Mexicanos (Pemex)
Russia (Coal)
Royal Dutch Shell PLC
China National Petroleum Corp (CNPC)
BP PLC
Chevron Corp
Petroleos de Venezuela SA (PDVSA)
Abu Dhabi National Oil Co
Poland Coal
Peabody Energy Corp

Problem !

Solution

Build our own color scale

Polluters



- China (Coal)
- Saudi Arabian Oil Company (Aramco)
- Gazprom OAO
- National Iranian Oil Co
- ExxonMobil Corp
- Coal India
- Petroleos Mexicanos (Pemex)
- Russia (Coal)
- Royal Dutch Shell PLC
- China National Petroleum Corp (CNPC)
- BP PLC
- Chevron Corp
- Petroleos de Venezuela SA (PDVSA)
- Abu Dhabi National Oil Co
- Poland Coal
- Peabody Energy Corp

```
import pandas as pd
import plotly.express as px
from colour import Color

df_companies = pd.read_csv("companies_list.csv", delimiter=',')
above1pc = df_companies[df_companies["percentage"] > 1.0]

black = Color("#000000")
colors = list(black.range_to(Color("#ffffff"), len(above1pc))) # no loop in scale

myColors = []
for c in colors:
    myColors.append(c.hex)

fig = px.pie(above1pc, values='percentage', names='company', title='Polluters',
             color="rank",
             color_discrete_sequence=myColors
             )
fig.show()
```

A New library for your toolbelt

<https://pypi.org/project/colour/>

Sort ...

Need to sort a dataframe ?

df.sort_values(by='colName', ascending = True)

**df.sort_values(by=['colName','colName2'],
ascending = True)**

Methods to try

length(df)

df.size

df.shape

df.dtypes

df.iloc[123]

df.loc[«something»]

En route ! / Let's go

You got the machette
Now explore



<https://www.lookandlearn.com/history-images/A147955/Jungle-explorer-hacking-through-the-jungle?img=1&search=hacking&bool=phrase>