

DATA ANALYSES AND VISUALIZATION FOR BEGINNERS



Mr. Ngazetungue Muheue
@muheuenga
muheuenga@gmail.com

PyCon NA 2019 , Windhoek
22 February 2019



P Y N A M
S O C I E T Y

Understanding Data Analysis and Visualization

- 1. Data analysis** - is a process of inspecting, cleansing, transforming, and modeling data to discovering useful information, informing conclusions, and supporting decision-making.

Help us to:

To make informed decisions
improve efficiency, increase profits and achieve organisational goals.

- 2. Data visualization** - plotting library for data visualization

Help us to:

Using charts or graphs to visualize large amounts of complex data
Identify areas that need attention or improvement.

Installation

Anaconda

Jupyter Notebook

Pandas - data manipulation and analysis.

Matplotlib - plotting library for data visualization

Workshop file

File for this workshop is on github.

<https://github.com/muheue/Presentation>

DATA ANALYSIS

Import pandas as pd

```
import pandas as pd
```

Reading Data in Panda

```
#Read Headers # Read each Column
print(learners.head(5)) print(learners[['Year', 'Region', 'NLearners']])

# Read Each Row
print(learners["Year"])

#print(learner.iloc[0:4])
for index, row in df.iterrows():
    print(index, row['Year'])

learners.loc[df['Type 1'] == "muheue"]
## Read a specific location (R,C)
print(learners.iloc[2,3])
```

DATA VISUALIZATION

Matplotlib

```
Import matplotlib.pyplot as pd
```

Our Plotting Area

```
plt.title("Grade 12 Performance")
plt.plot(x,y)
plt.ylabel("Number of Learners")
plt.xlabel("Year")
plt.show()
```

```
plt.show()
```

Plotting Data

```
# Plotting Data for each region
```

```
plt.plot(Kun.Year, Kun.Failed, Ero.Year, Ero.Failed)
plt.legend(["Kunene","Erongo"])
plt.show()
```

```
# Print the data of kunene region
```

```
Ero= learners[learners.Region == "Erongo"] print(Ero)
```

Our Regions and their Data

```
plt.title("Grade 12 Performance")
plt.plot(x,y)
plt.ylabel("Number of Learners")
plt.xlabel("Year")
plt.show()
```

```
plt.show()
```

Our Plotting Area

```
plt.title("Grade 12 Performance")
plt.plot(x,y)
plt.ylabel("Number of Learners")
plt.xlabel("Year")
plt.show()
```

```
plt.show()
```

Our Plotting Area

```
plt.title("Grade 12 Performance")
plt.plot(x,y)
plt.ylabel("Number of Learners")
plt.xlabel("Year")
plt.show()
```

```
plt.show()
```

Our Plotting Area

File for this workshop is on github.

<https://github.com/muheue/Presentation>

Links

<https://matplotlib.org/gallery/index.html>
<http://pandas.pydata.org/pandas-docs/stable/install.html#installing-pandas>
<https://www.anaconda.com/distribution/>
http://matplotlib.org/api/pyplot_summary.html
<http://pandas.pydata.org/pandas-docs/stable/>

Thank you!!!

Matplotlib
Pandas
Jupyter
Numpy
Anaconda



PyCon Namibia 2019

22 Windhoek
February 2019

PYNNAM
S O C I E T Y