Undertanding the cartwheel data set

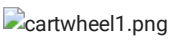
The notebook aims to undertand the content of the cartwheel data set.

Acknowledgments

- Data from https://www.coursera.org/ from the course "Understanding and Visualizing Data with Python" by University of Michigan

Cartwheel data set

1. A cartwheel



2. The dataset description
    - The dataset used here is an extension from the original cartwheel dataset from cursera
    - Total numer of observations: 28
    - Many observations/measurements/recordings of the characteristics/attributes/variables of cartwheel executions
    - Variables: Age, Gender, GenderGroup, Glasses, GlassesGroup, Height, Wingspan, CWDistance, ... (X variables)

Importing and inspecting the data

```
# Define where you are running the code: colab or local
RunInColab          = True     # (False: no  | True: yes)

# If running in colab:
if RunInColab:
    # Mount your google drive in google colab
    from google.colab import drive
    drive.mount('/content/drive')

    # Find location
    #!pwd
    #!ls
    #!ls "/content/drive/My Drive/Colab Notebooks/MachineLearningWithPython/"

    # Define path del proyecto
    Ruta            = "/content/drive/My Drive/Colab Notebooks/MachineLearningWithPython/"

else:
    # Define path del proyecto
    Ruta            = ""
```

```
 Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
# Import the packages that we will be using
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

# Dataset url
url = Ruta + "iris.csv"

# Load the dataset
df = pd.read_csv(url,names=['SepalLength', 'SepalWidth', 'PetalLength','PetalWidth','FlowerType'])



# Print the dataset
df
```

	SepalLength	SepalWidth	PetalLength	PetalWidth	FlowerType
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica

```
# Print the number of rows
fila = len(df.index)
print('Número de Filas:', fila)
```

```
Número de Filas: 150
```

```
# Print the number of columns
columnas = df.columns
Ncolumnas= len(columnas)

print('Número de Columnas:', Ncolumnas)
```

```
Número de Columnas: 5
```

Activity: work with the iris dataset

1. Load the iris.csv file in your computer and understand the dataset
2. How many observations (rows) are in total?
3. How many variables (columns) are in total? What do they represent?
4. How many observations are for each type of flower?
5. What is the type of data for each variable?
6. What are the units of each variable?

✓ 0 s se ejecutó 21:40