Undertanding the cartwheel data set

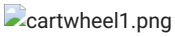
The notebook aims to undertand the content of the cartwheel data set.

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/semana:tec/datasets_iris/Iris.csv/"

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

```
Mounted at /content/drive
```

```
# Import the packages that we will be using
import pandas as pd
from sklearn.datasets import load_iris

# Dataset url
iris = load_iris()
# Read the .csv file and store it as a pandas Data Frame
df = pd.DataFrame(iris.data, columns=iris.feature_names)

# Add the species as a column for easier grouping
df['Class'] = iris.target_names[iris.target]


# Print the dataset
df.head(6)
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	Class
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa

Próximos pasos: Generar código con df | Ver gráficos recomendados | New interactive sheet

```
df.shape
```

```
(52, 12)
```

```
# Print the number of rows
print("Number of rows: ", df.shape[0])
```

```
El dataset tiene 52 filas
```

```
# Print the number of columns
print("Number of columns: ", df.shape[1])
```

```
Number of columns:  5
```

Data types

```
# What is the type of data for each variable?
print(df.dtypes)
```

```
sepal length (cm)    float64
sepal width (cm)     float64
petal length (cm)    float64
petal width (cm)     float64
Class                 object
dtype: object
```

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?

```
# Activity: work with the iris dataset
# How many observations (rows) are in total?
print(df.shape[0])

# How many variables (columns) are in total? What do they represent?
print(df.shape[1])

# How many observations are for each type of flower?
print(df.value_counts("Class"))

# What is the type of data for each variable?
print(df.dtypes)
```

```
150
5
Class
setosa        50
versicolor    50
virginica     50
Name: count, dtype: int64
sepal length (cm)    float64
sepal width (cm)     float64
petal length (cm)    float64
petal width (cm)     float64
Class                 object
dtype: object
```