

## 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)

### Nombre y matricula

Andres Benjamin Antelis Moreno A01637683

### ▼ 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/MyDrive/!Tec stuff/!Uni/Semestre 2/Semana Tec1/TC1002S/NotebooksStudents/A01637683"

else:
    # Define path del proyecto
    Ruta = ""

    Mounted at /content/drive

# Import the packages that we will be using
import pandas as pd

# Dataset url
Dataset="/content/drive/MyDrive/!Tec stuff/!Uni/Semestre 2/Semana Tec1/TC1002S/NotebooksProfessor/datasets/cartwheel/cartwheel.csv"

# Load the dataset
db=pd.read_csv(Dataset)

# Print the dataset
print(db)
```

```
# Print the number of rows
len(db.index)
```

52

```
# Print the number of columns
print(len(db.columns))
```

12

## ▼ 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?

149

3. How many variables (columns) are in total? What do they represent? 5, they represent the values of the characteristics assigned to each of the flowers
4. How many observations are for each type of flower? 4
5. What is the type of data for each variable? Numerical float numbers
6. What are the units of each variable? Tnee values of each of the variables on the iris cvs are float numbers

```
data2="/content/drive/MyDrive/!Tec stuff/!Uni/Semestre 2/Semana Tec1/TC1002S/NotebooksProfessor/datasets/iris/iris.csv"
```

```
dbI=pd.read_csv(data2)
print(dbI)
```

```
len(dbI.index)
```

149

```
print(len(dbI.columns))
```

5

```
print(type(dbI))
```

```
<class 'pandas.core.frame.DataFrame'>
```