```
# EMILIO BERBER MALDONADO - A01640603
# ACT 1: Cartwheel and Iris
# SEMANA TEC
```

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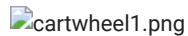
The notebook aims to undertand the content of the cartwheel data set.

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

cartwheel1.png

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           = ""

     Mounted at /content/drive

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


# Dataset url
url = "/content/drive/MyDrive/cartwheel.csv"

# Load the dataset
data = pd.read_csv(url)
```

```
# Print the dataset
data
```

	ID	Age	Gender	GenderGroup	Glasses	GlassesGroup	Height	Wingspan	CWDista
0	1	56.0	F	1	Y	1	62.00	61.0	
1	2	26.0	F	1	Y	1	62.00	60.0	
2	3	33.0	F	1	Y	1	66.00	64.0	
3	4	39.0	F	1	N	0	64.00	63.0	
4	5	27.0	M	2	N	0	73.00	75.0	
5	6	24.0	M	2	N	0	75.00	71.0	
6	7	28.0	M	2	N	0	75.00	76.0	
7	8	22.0	F	1	N	0	65.00	62.0	
8	9	29.0	M	2	Y	1	74.00	73.0	

```
# Print the number of rows
Nrows = data.shape[0]
Nrows
```

```
52
```

| 13 | 14 | 23.0 | F | 1 | N | 0 | 61.50 | 57.5 | |

```
# Print the number of columns
Ncol = data.shape[1]
Ncol
```

```
12
```

| 17 | 18 | 27.0 | M | 2 | N | 0 | 66.00 | 66.0 | |

```
data.shape
```

```
(52, 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?

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?

| 30 | 31 | 27.0 | F | 1 | Y | 1 | 62.00 | 60.0 | |

```
dataIris = pd.read_csv("/content/drive/MyDrive/iris.csv", header = None)
dataIris #1
```

	0	1	2	3	4
0	5.1	3.5	1.4	0.2	Iris-setosa
46	44.9 2.0	1.4	0.2	Iris-setosa	N	0	78.00	75.0

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Nrows= dataIris.shape[0]
Nrows #2
```

```
150
```

| 146 | 6.3 | 2.5 | 5.0 | 1.9 | Iris-virginica |

```
Ncol = dataIris.shape[1]
Ncol #3
```

```
5
```

```
dataIris.shape
```

```
(150, 5)
```

```
dataIris[4].value_counts() #4
```

```
Iris-setosa       50
Iris-versicolor   50
Iris-virginica    50
Name: 4, dtype: int64
```

```
dataIris.dtypes #5
```

```
0    float64
1    float64
2    float64
3    float64
4     object
dtype: object
```

```
# 6
# UNIDADES:
# 0    float64 --> cm [Sepalo_Largo]
# 1    float64 --> cm [Sepalo_Ancho]
# 2    float64 --> cm [Petalo_Largo]
# 3    float64 --> cm [Petalo_Ancho]
```

✓ 0 s   se ejecutó 07:55