Undertanding the cartwheel data set

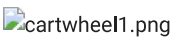
The notebook aims to undertand the content of the cartwheel data set.

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/"

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

```
Mounted at /content/drive
Hola Drive
```

Haz doble clic (o ingresa) para editar

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

# Dataset url
url = Ruta + "A01641179/datasets/cartwheel/cartwheel.csv"

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



# Print the dataset
dataset
```

12	13	25.0	F	1	Y	1	65.00	64.5	92	Y
13	14	23.0	F	1	N	0	61.50	57.5	66	Y
14	15	31.0	M	2	Y	1	73.00	74.0	72	Y
15	16	26.0	M	2	Y	1	71.00	72.0	115	Y
16	17	26.0	F	1	N	0	61.50	59.5	90	N
17	18	27.0	M	2	N	0	66.00	66.0	74	Y
18	19	23.0	M	2	Y	1	70.00	69.0	64	Y
19	20	24.0	F	1	Y	1	68.00	66.0	85	Y
20	21	23.0	M	2	Y	1	69.00	67.0	66	N
21	22	29.0	M	2	N	0	71.00	70.0	101	Y
22	23	25.0	M	2	N	0	70.00	68.0	82	Y
23	24	26.0	M	2	N	0	69.00	71.0	63	Y
24	25	23.0	F	1	Y	1	65.00	63.0	67	N
25	26	28.0	M	2	N	0	75.00	76.0	111	Y
26	27	24.0	M	2	N	0	78.40	71.0	92	Y
27	28	25.0	M	2	Y	1	76.00	73.0	107	Y
28	29	32.0	F	1	Y	1	63.00	60.0	75	Y
29	30	38.0	F	1	Y	1	61.50	61.0	78	Y
30	31	27.0	F	1	Y	1	62.00	60.0	72	Y
31	32	33.0	F	1	Y	1	65.30	64.0	91	Y
32	33	38.0	F	1	N	0	64.00	63.0	86	Y
33	34	27.0	M	2	N	0	77.00	75.0	100	Y
34	35	24.0	F	1	N	0	67.80	62.0	98	Y
35	36	27.0	M	2	N	0	68.00	66.0	74	Y
36	37	25.0	F	1	Y	1	65.00	64.5	92	Y
37	38	26.0	F	1	N	0	61.50	59.5	90	Y
38	39	31.0	M	2	Y	1	73.00	74.0	72	Y
39	40	30.0	M	2	Y	1	69.50	66.0	96	Y
40	41	23.0	F	1	N	0	70.40	71.0	66	Y
41	42	26.0	M	2	Y	1	73.50	72.0	115	Y
42	43	28.0	F	1	Y	1	72.50	72.0	81	Y
43	44	26.0	F	1	Y	1	72.00	72.0	92	Y
44	45	30.0	F	1	Y	1	66.00	64.0	85	Y
45	46	39.0	F	1	N	0	64.00	63.0	87	Y
46	47	27.0	M	2	N	0	78.00	75.0	72	N
47	48	24.0	M	2	N	0	79.50	75.0	82	N
48	49	28.0	M	2	N	0	77.80	76.0	99	Y
49	50	30.0	F	1	N	0	74.60	NaN	71	Y
50	51	NaN	M	2	N	0	71.00	70.0	101	Y
51	52	27.0	M	2	N	0	NaN	71.5	103	Y

```
# Print the number of rows
df = pd.read_csv(url ) #===> reads in all the rows, but skips the first one as it is a header..

total_rows=len(df.axes[0]) #===> Axes of 0 is for a row
print("Numero de filas: "+str(total_rows))
```

```
Numero de filas: 52
```

```
# Print the number of columns

total_cols=len(df.axes[1]) #===> Axes of 0 is for a column
print("Numero·de·columnas:·"+str(total_cols))
```

```
Numero de columnas: 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?

Load the iris.csv file in your computer and understand the dataset

```
# 1.Load the iris.csv file in your computer and understand the dataset
  # Import the packages that we will be using
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

# Dataset url
url = Ruta + "A01641179/datasets/iris/iris.csv"

dataset·=·pd.read_csv(url·)

df = pd.Cov = pd.read_csv(url , sep=',', names=["Sepal length (cm)", "Sepal width (cm)", "Petal length (cm)", "Petal width (cm)","Target"])


# Print the dataset
print(df)
```

```
     Sepal length (cm)  Sepal width (cm)  Petal length (cm)  Petal width (cm)  \
0                  5.1               3.5                1.4               0.2
1                  4.9               3.0                1.4               0.2
2                  4.7               3.2                1.3               0.2
3                  4.6               3.1                1.5               0.2
4                  5.0               3.6                1.4               0.2
..                 ...               ...                ...               ...
145                6.7               3.0                5.2               2.3
146                6.3               2.5                5.0               1.9
147                6.5               3.0                5.2               2.0
148                6.2               3.4                5.4               2.3
149                5.9               3.0                5.1               1.8

              Target
0        Iris-setosa
1        Iris-setosa
2        Iris-setosa
3        Iris-setosa
4        Iris-setosa
..               ...
145   Iris-virginica
```

```
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica

[150 rows x 5 columns]
```

```
# 2.How many observations (rows) are in total?
# Print the number of rows

total_rows=len(df.axes[0]) #===> Axes of 0 is for a row
print("Numero de filas: "+str(total_rows))
```

```
Numero de filas: 150
```

How many variables (columns) are in total? What do they represent?

```
# 3.How many variables (columns) are in total? What do they represent?
# Print the number of columns

total_cols=len(df.axes[1]) #===> Axes of 0 is for a column
print("Numero de columnas: "+str(total_cols))
```

```
Numero de columnas: 5
```

How many observations are for each type of flower?

```
# 4.How many observations are for each type of flower?
df = pd.read_csv(url , header = None)
df[4].value_counts()
```

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

5.What is the type of data for each variable? Int64

6.What are the units of each variable? Int