Undertanding the cartwheel data set

VANESSA M. CUEVAS ARROYO A01634064

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

1. The dataset description
   - The dataset used here is an extension from the original cartwheel dataset from cursera
   - Total numer of observations: 52
   - 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

```python
# Define where you are running the code: colab or local
RunInColab        = False     # (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
```

```python
        Notebooks/MachineLearningWithPython/"

else:
    # Define path del proyectoC:\Users\vanec\Desktop\6to semestre\
Analitica\TC1002S\NotebooksStudents\A01634064\datasets\cartwheel\
carthweel.csv
    Ruta            = "datasets/cartwheel/cartwheel.csv"

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

# Dataset url



# Load the dataset
datos  = pd.read_csv(Ruta)

# Print the dataset
print(datos)
```

```
    ID   Age Gender  GenderGroup Glasses  GlassesGroup  Height
Wingspan  \
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
9   10  33.0      F            1       Y             1   63.00
60.0
10  11  30.0      M            2       Y             1   69.50
66.0
11  12  28.0      F            1       Y             1   62.75
58.0
12  13  25.0      F            1       Y             1   65.00
64.5
```

13	14	23.0	F	1	N	0	61.50	57.5
14	15	31.0	M	2	Y	1	73.00	74.0
15	16	26.0	M	2	Y	1	71.00	72.0
16	17	26.0	F	1	N	0	61.50	59.5
17	18	27.0	M	2	N	0	66.00	66.0
18	19	23.0	M	2	Y	1	70.00	69.0
19	20	24.0	F	1	Y	1	68.00	66.0
20	21	23.0	M	2	Y	1	69.00	67.0
21	22	29.0	M	2	N	0	71.00	70.0
22	23	25.0	M	2	N	0	70.00	68.0
23	24	26.0	M	2	N	0	69.00	71.0
24	25	23.0	F	1	Y	1	65.00	63.0
25	26	28.0	M	2	N	0	75.00	76.0
26	27	24.0	M	2	N	0	78.40	71.0
27	28	25.0	M	2	Y	1	76.00	73.0
28	29	32.0	F	1	Y	1	63.00	60.0
29	30	38.0	F	1	Y	1	61.50	61.0
30	31	27.0	F	1	Y	1	62.00	60.0
31	32	33.0	F	1	Y	1	65.30	64.0
32	33	38.0	F	1	N	0	64.00	63.0
33	34	27.0	M	2	N	0	77.00	75.0
34	35	24.0	F	1	N	0	67.80	62.0
35	36	27.0	M	2	N	0	68.00	66.0
36	37	25.0	F	1	Y	1	65.00	64.5
37	38	26.0	F	1	N	0	61.50	59.5

38	39	31.0	M	2	Y	1	73.00
		74.0					
39	40	30.0	M	2	Y	1	69.50
		66.0					
40	41	23.0	F	1	N	0	70.40
		71.0					
41	42	26.0	M	2	Y	1	73.50
		72.0					
42	43	28.0	F	1	Y	1	72.50
		72.0					
43	44	26.0	F	1	Y	1	72.00
		72.0					
44	45	30.0	F	1	Y	1	66.00
		64.0					
45	46	39.0	F	1	N	0	64.00
		63.0					
46	47	27.0	M	2	N	0	78.00
		75.0					
47	48	24.0	M	2	N	0	79.50
		75.0					
48	49	28.0	M	2	N	0	77.80
		76.0					
49	50	30.0	F	1	N	0	74.60
		NaN					
50	51	NaN	M	2	N	0	71.00
		70.0					
51	52	27.0	M	2	N	0	NaN
		71.5					

	CWDistance	Complete	CompleteGroup	Score
0	79	Y	1.0	7
1	70	Y	1.0	8
2	85	Y	1.0	7
3	87	Y	1.0	10
4	72	N	0.0	4
5	81	N	0.0	3
6	107	Y	1.0	10
7	98	Y	1.0	9
8	106	N	0.0	5
9	65	Y	1.0	8
10	96	Y	1.0	6
11	79	Y	1.0	10
12	92	Y	1.0	6
13	66	Y	1.0	4
14	72	Y	1.0	9
15	115	Y	1.0	6
16	90	N	0.0	10
17	74	Y	1.0	5
18	64	Y	1.0	3
19	85	Y	1.0	8

```
20        66        N        0.0     2
21       101        Y        1.0     8
22        82        Y        1.0     4
23        63        Y        1.0     5
24        67        N        0.0     3
25       111        Y        1.0    10
26        92        Y        1.0     7
27       107        Y        1.0     8
28        75        Y        1.0     8
29        78        Y        1.0     7
30        72        Y        1.0     8
31        91        Y        1.0     7
32        86        Y        1.0    10
33       100        Y        1.0     8
34        98        Y        1.0     9
35        74        Y        1.0     5
36        92        Y        1.0     6
37        90        Y        1.0     9
38        72        Y        1.0     9
39        96        Y        1.0     6
40        66        Y        1.0     4
41       115        Y        1.0     6
42        81        Y        1.0    10
43        92        Y        1.0     8
44        85        Y        1.0     7
45        87        Y        1.0    10
46        72        N        0.0     7
47        82        N        0.0     8
48        99        Y        1.0     9
49        71        Y        1.0     9
50       101        Y        NaN     8
51       103        Y        1.0    10
```

```python
# Print the number of rows

print("Number of rows: ", len(datos))
```

```
Number of rows:  52
```

```python
# Print the number of columns
print("number of columns", len(datos.columns))
```

```
number of 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?

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?

```python
#Load the dataset iris.csv

# Load the dataset
datosIris  = pd.read_csv('datasets/iris/iris.csv')

# Print the number of rows

print("Number of rows: ", len(datosIris))
```

Number of rows: 149

```python
# Print the number of columns
print("number of columns", len(datosIris.columns))

datosIris.dtypes
```

number of columns 5

```
5.1              float64
3.5              float64
1.4              float64
0.2              float64
Iris-setosa       object
dtype: object
```

```python
datosIris.columns = ["PetalWidth", "PetalLength", "SepalWidth",
"SepalLength", "Type"]

datosIris['Type'].value_counts()['Iris-setosa']
```

49

```python
datosIris['Type'].value_counts()['Iris-versicolor']
```

50

```python
datosIris['Type'].value_counts()['Iris-virginica']
```

50

3.What do they represent?

Las columnas representan las partes de las flores en Sepal y Nepal con su 'weight' and 'Height'

1.   How many observations are for each type of flower? 4

2. What is the type of data for each variable? float

3. What are the units of each variable? cm