Carlos David Amezcua Canales - A01641742

Undertanding the cartwheel data set

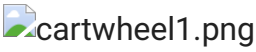
The notebook aims to undertand the content of the cartwheel data set.

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/Colab Notebooks/MachineLearningWithPyth

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

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

```
# Import the packages that we will be using

import matplotlib.pyplot as plt
import pandas as pd

# Dataset url

url_1 = Ruta + "datasets/cartwheel/cartwheel.csv"

# Load the dataset

dataset_1 = pd.read_csv(url_1)
```

```
# Print the dataset

display(dataset_1)
```

	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	
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	

```
# Print the number of rows

num_rows_1, num_cols_1 = dataset_1.shape

print(num_rows_1)
```

```
52
```

35	36	27.0	M	2	N	0	68.00	66.0

```
# Print the number of columns

print(num_cols_1)
```

```
12
```

39	40	30.0	M	2	Y	1	69.50	66.0

Activity: work with the iris dataset

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

```
# Dataset url

url_2 = Ruta + "datasets/iris/iris.csv"

# Load the dataset

dataset_2 = pd.read_csv(url_2, header = None,
                        names=['sepal_length', 'sepal_width', 'petal_length', 'petal_w

# Print the dataset

display()
```

	sepal_length	sepal_width	petal_length	petal_width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa

2. How many observations (rows) are in total?

...	...	...	...	...	...

```
# Print the number of rows

num_rows_2, num_cols_2 = dataset_2.shape

print(num_rows_2)
```

```
150
```

149	5.9	3.0	5.1	1.8	Iris-virginica

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

```
# Print the number of columns

print(num_cols_2)
```

```
5
```

They represent:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
    - Iris Setosa
    - Iris Versicolour
    - Iris Virginica

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

```
flower_counts = dataset_2["class"].value_counts()
```

```
print(flower_counts)
```

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

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

```
print(dataset_2.dtypes)
```

```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
class            object
dtype: object
```

6. What are the units of each variable?

   The units of measurement for the first four variables are centimeters.