Double-click (or enter) to edit

Lucas Wong Mang A01639032

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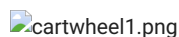
The notebook aims to undertand the content of the cartwheel data set.

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/Herramientas Computacionales/"

else:
    # Define path del proyecto
    Ruta          = ""

    Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount
```

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

# Dataset url
url = "cartwheel.csv"

# Load the dataset
df = pd.read_csv(Ruta+url)


# Print the dataset
df.head()
```

	ID	Age	Gender	GenderGroup	Glasses	GlassesGroup	Height	Wingspan	CWDistance	Complete	CompleteGroup	Score
0	1	56.0	F	1	Y	1	62.0	61.0	79	Y	1.0	7

```
# Print the number of rows
df.shape[0]
```

```
    52
```

```
# Print the number of columns
df.shape[1]
```

```
    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?

```
# 1 -- Load iris.csv
Ruta_General = "/content/drive/My Drive/Herramientas Computacionales/"
url = "iris.csv"
url_iris=Ruta_General+url

# Dataset
header·=·['s_length',·'s_width',·'p_length',·'p_width',·'class']
ds_iris·=·pd.read_csv(url_iris,·names=header)

# Print
ds_iris
```

	s_length	s_width	p_length	p_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
4	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

```
# 2 & 3 -- # of rows and columns
print(f"{ds_iris.shape[0]} rows and {ds_iris.shape[1]} columns")
```

```
    150 rows and 5 columns
```

```
# 4 -- # of observations per flower
ds_iris['class'].value_counts()
```

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

```
# 5 -- Types of data

ds_iris.dtypes
```

```
s_length    float64
s_width     float64
p_length    float64
p_width     float64
class        object
dtype: object
```

```
# 6 -- Units of each var
ds_iris.describe()
```

	s_length	s_width	p_length	p_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

```
print("Each in cm")
```

```
Each in cm
```

Colab paid products - Cancel contracts here