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: running locally | True: running in Colab)

# If running in Colab:
if RunInColab:
    # Mount your Google Drive in Google Colab
    from google.colab import drive
    drive.mount('/content/drive')

    # Define the path of the project
    Ruta = "/content/drive/MyDrive/ITC/5toSem/semanaTecAn/"

else:
    # Define the path of the project for local use
    Ruta = "your/local/path/here"  # Replace with your local path if running locally

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

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

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

# Dataset url
url = "datasets/iris.csv"

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

```

```python
# Print the dataset
df
```

	sepal.length	sepal.width	petal.length	petal.width	variety
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Virginica
146	6.3	2.5	5.0	1.9	Virginica
147	6.5	3.0	5.2	2.0	Virginica
148	6.2	3.4	5.4	2.3	Virginica
149	5.9	3.0	5.1	1.8	Virginica

150 rows × 5 columns

Next steps: Generate code with df | View recommended plots | New interactive sheet

```python
df.shape
```

```
(150, 5)
```

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

print("El dataset tiene " + str(Nrows) + " filas")

```

```
El dataset tiene 150 filas
```

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

print("El dataset tiene " + str(Ncols) + " variables")

```

```
El dataset tiene 5 variables
```

Data types

```python
df.dtypes
```

	0
sepal.length	float64
sepal.width	float64
petal.length	float64
petal.width	float64
variety	object

dtype: object

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?

```
# Get the total number of rows (observations)
total_rows = len(df)
print(f"Total number of observations (rows): {total_rows}")
```

```
Total number of observations (rows): 150
```

```
# Get the number of columns (variables) and their names
total_columns = len(df.columns)
print(f"Total number of variables (columns): {total_columns}")
print("Columns and their representation:")
print(df.columns)
```

```
Total number of variables (columns): 5
Columns and their representation:
Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',
       'variety'],
      dtype='object')
```

```
# Count the number of observations for each type of flower
flower_counts = df['variety'].value_counts()
print("Number of observations for each type of flower:")
print(flower_counts)
```

```
Number of observations for each type of flower:
variety
Setosa        50
Versicolor    50
Virginica     50
Name: count, dtype: int64
```

```
# Check the data type of each variable
print("Data types of each variable:")
print(df.dtypes)
```

```
Data types of each variable:
sepal.length    float64
sepal.width     float64
petal.length    float64
petal.width     float64
variety          object
dtype: object
```

```
print("The units of each variable in the iris dataset are typically:")

print("Sepal Length and Width: Measured in centimeters (cm).")
print("Petal Length and Width: Measured in centimeters (cm).")
print("Variety: Categorical data representing the type of flower.")
```

```
The units of each variable in the iris dataset are typically:
Sepal Length and Width: Measured in centimeters (cm).
Petal Length and Width: Measured in centimeters (cm).
Variety: Categorical data representing the type of flower.
```