

Understanding the cartwheel data set

The notebook aims to understand 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 coursera
- Total number 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/NotebooksProfessor"

else:
    # Define path del proyecto
    Ruta          = ""

 Mounted at /content/drive

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

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

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

# Print the dataset
df
```



20	21	23.0	M	2	Y	1	69.00	67.0	66
21	22	29.0	M	2	N	0	71.00	70.0	101
22	23	25.0	M	2	N	0	70.00	68.0	82
23	24	26.0	M	2	N	0	69.00	71.0	63
24	25	23.0	F	1	Y	1	65.00	63.0	67
25	26	28.0	M	2	N	0	75.00	76.0	111
26	27	24.0	M	2	N	0	78.40	71.0	92
27	28	25.0	M	2	Y	1	76.00	73.0	107
28	29	32.0	F	1	Y	1	63.00	60.0	75
29	30	38.0	F	1	Y	1	61.50	61.0	78

30	31	27.0	F	1	Y	1	62.00	60.0	72
31	32	33.0	F	1	Y	1	65.30	64.0	91
32	33	38.0	F	1	N	0	64.00	63.0	86
33	34	27.0	M	2	N	0	77.00	75.0	100
34	35	24.0	F	1	N	0	67.80	62.0	98
35	36	27.0	M	2	N	0	68.00	66.0	74
36	37	25.0	F	1	Y	1	65.00	64.5	92
37	38	26.0	F	1	N	0	61.50	59.5	90
38	39	31.0	M	2	Y	1	73.00	74.0	72
39	40	30.0	M	2	Y	1	69.50	66.0	96
40	41	23.0	F	1	N	0	70.40	71.0	66
41	42	26.0	M	2	Y	1	73.50	72.0	115
42	43	28.0	F	1	Y	1	72.50	72.0	81
43	44	26.0	F	1	Y	1	72.00	72.0	92
44	45	30.0	F	1	Y	1	66.00	64.0	85
45	46	39.0	F	1	N	0	64.00	63.0	87
46	47	27.0	M	2	N	0	78.00	75.0	72
47	48	24.0	M	2	N	0	79.50	75.0	82
48	49	28.0	M	2	N	0	77.80	76.0	99
49	50	30.0	F	1	N	0	74.60	NaN	71
50	51	NaN	M	2	N	0	71.00	70.0	101
51	52	27.0	M	2	N	0	NaN	71.5	103

Pasos
siguientes:

Generar código
con df

☒ Ver gráficos
recomendados

New interactive
sheet

```
df.head(5)
```



	ID	Age	Gender	GenderGroup	Glasses	GlassesGroup	Height	Wingspan	CWDistance	Co
0	1	56.0	F	1	Y	1	62.0	61.0	79	
1	2	26.0	F	1	Y	1	62.0	60.0	70	
2	3	33.0	F	1	Y	1	66.0	64.0	85	
3	4	39.0	F	1	N	0	64.0	63.0	87	
4	5	27.0	M	2	N	0	73.0	75.0	72	

Pasos
siguientes:

[Generar código con](#) df



[Ver gráficos recomendados](#)

[New interactive sheet](#)

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



52

```
print(pd.options.display.max_columns)
```



20

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



12

```
df.shape
```



(52, 12)

✓ Data types

```
df.dtypes
```



	0
ID	int64
Age	float64
Gender	object
GenderGroup	int64
Glasses	object
GlassesGroup	int64
Height	float64
Wingspan	float64
CWDistance	int64
Complete	object
CompleteGroup	float64
Score	int64

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?

```
# Import necessary libraries
import seaborn as sns

# Load the Iris dataset from seaborn
df = sns.load_dataset('iris')

# Check the first few rows of the dataset
df
```



	sepal_length	sepal_width	petal_length	petal_width	species	
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

Pasos siguientes:

Generar código con df

☒ Ver gráficos recomendados

New interactive sheet

Rows

```
df.shape[0]
```

 150

Columns

```
df.shape[0]
```

 150

Observations