

# **Data Science 1**

## **Herhaling en examen tips**

---

# Agenda

1. Minder kennen maar meer kunnen
2. Pycharm klaarzetten
3. Cheatpages
4. Maak zoveel mogelijk oefeningen
5. Bij denkvragen maak een dataset en experimenteer ermee

---

**Minder kennen maar  
meer kunnen**

An abstract graphic design featuring a white background. On the right side, there are several thick black geometric elements: a diagonal line segment pointing towards the center, a vertical line segment extending from the top to the bottom, and a horizontal line segment at the bottom right corner.

---

# Algemeen

- Wat zijn de verschillende meetniveau's

---

# Python algemeen... kunnen

Dictionary maken en gebruiken

Lijsten maken en gebruiken

Numpy array maken

Variabelen omzetten in andere vorm

For lus

if then else...

---

# Python algemeen... kunnen

Concatineren van strings

Printen van strings en getallen (f-format)

Rekenregels  $8+4-a*b/2^3$

Numpy array rekenregels (broadcasting)

---

# Data management ....kunnen

- Inlezen van dataframe

- Separator, skip rows, delimiter, header,...

- Dataframe maken van scratch

- Dataframe manipulaties uitvoeren:

- Kolom bijvoegen, wegdoen, veranderen
  - Rij toevoegen, wegdoen, veranderen
  - Alle rijen selecteren waar van de kolomwaarde aan voorwaarde voldoet
    - (voor kolom 'x' groter is dan 2 bijvoorbeeld)
  - Iloc, loc en [[]] kunnen gebruiken
  - Categorieën gebruiken
  - Nan Values droppen, vervangen,...

---

# Frequenties .. te kunnen

27/11/15

Berekenen van frequenties (absolute, relatieve, cumulatieve en relatief cumulatieve)

- Verschillende voorstellingen hangen af van
- ...meetniveau (betekenis)
- Voorbeeld histogram kan niet voor ....
- Matplotlib importeren en gebruiken
  - Boxplot, histogram, taartdiagram, staafdiagram, ..
  - (niet te kennen spiderdiagram)



---

# Centrummaten en spreidingsmaten: 27/11/15 kunnen berekenen van:

- Centrummaten:

- Mediaan, gemiddelde (verschillende soorten) , modus

- Spreidingsmaten:

- Standaardafwijking , IQR (dus ook percentielen), bereik,

- Uitschieters:

- Gewone en extreme kunnen bepalen (tip gebruik de functie)
- Ze kunnen verwijderen uit de dataframe

- Z scores:

- Waarom ? Hoe te bepalen?

**Project**

- pythonProjectDS1 C:\Users\baeke
- > .ipynb\_checkpoints
- > venvG library root
- > week2Python
- > week3DM
- > week4Freq
- > week5CSM

**Bookmarks**

- > External Libraries
- > Scratches and Consoles

**Structure**

maten - Slides.ipynb x 1\_Opgaveboek - centrummaten - oplossingen.ipynb x

[+][↑][↓][Run][Debug][Stop][Refresh][Undo][Redo] Markdown [Trash]

### Setup

⚠ 5 ⚠ 5 ✅ 20

```
In 2 1 # keep
      2 import ...
      7
      8 InteractiveShell.ast_node_interactivity = "all"
      9
     10
     11
     12
```

▼ C:\Users\baeke\l\AppData\Local\Temp\ipykernel\_11184\3061443004.py:5: DeprecationWarning: Importing display from IPython.core.display is deprecated since IPython 7.14, please import from IPython display  
from IPython.core.display import display

## Centrummaten

---

# Gebruik cheatpages of maak ze zelf, 27/11/15 maar bestudeer ze voor examen

- [Pandas cheat page](#)
- [Numpy cheat page](#)
- [Datacamp pandas cheat page](#)
- Zoek er zelf naar via Google

# **Maak zoveel mogelijk oefeningen**

- Probeer alle notebooks en kijk pas naar oplossing als het niet lukt, ook als het lukt kijk oplossing na.
- Bestudeer de oplossingen en probeer zonder te kijken opnieuw.
- En opnieuw tot je het kunt

