



**GHENT
UNIVERSITY**



imec

GEDISTRIBUEERDE GEGEVENSVVERWERKING

(E761040)

LAB SESSION 01 - EDA

24/02/2025

INTRODUCTION



- EDA = **E**xploratory **D**ata **A**alysis

- Why?

 - You have a new job at Netflix and on your first day you receive a dataset

 - Next week, you need to develop new features based on this dataset, how will you do this?

 - = Simple, **EDA** to the rescue

 - = Explore the data in the dataset with graphics, to reveal underlying structure and important information

 - *“The reason for the heavy reliance on graphics is that by its very nature the main role of EDA is to open-mindedly explore, and graphics gives the analysts unparalleled power to do so, enticing the data to reveal its structural secrets, and being always ready to gain some new, often unsuspected, insight into the data.”*

HOW TO START WITH EDA?

- 1) Clean-up the data
→ Data-cleanup with pandas
- 2) Combine all the data together
→ Merge datasets together with pandas
- 3) Extract information
→ Make contents of data “logical”
- 4) Gain insight
→ Draw graphs



LAB SESSION STRUCTURE

- 1) Introduction to pandas
- 2) Load and merge datasets together
- 3) Extract data and gain insights with graphs

→ We will use notebooks with python code
(Use ctrl+Enter to execute a cell)

<https://github.ugent.be/GDV/lab-eda>

- 4) Work on project



PROJECT NOTES

- Assignment: <https://github.ugent.be/GDV/project-eda>
- Group work, join a group on Ufora if you haven't yet!
- Questions, problems
 - ◆ Create a Github issue, tag @togoetha
- Submit your solution on Github by 02/03/2025 at 23:59