

INFOMDW: Data Wrangling

Introduction to the course

Hakim Qahtan

Department of Information and Computing Sciences

Utrecht University



Utrecht University

Today

- Who are we?
 - Introduction to the course content
 - Getting to know you a little
 - Break
 - Practicalities & course flow
 - Time for questions
-
- But first...



Who are we?



Utrecht University

The Teaching team

Hakim Qahtan
a.qahtan@uu.nl

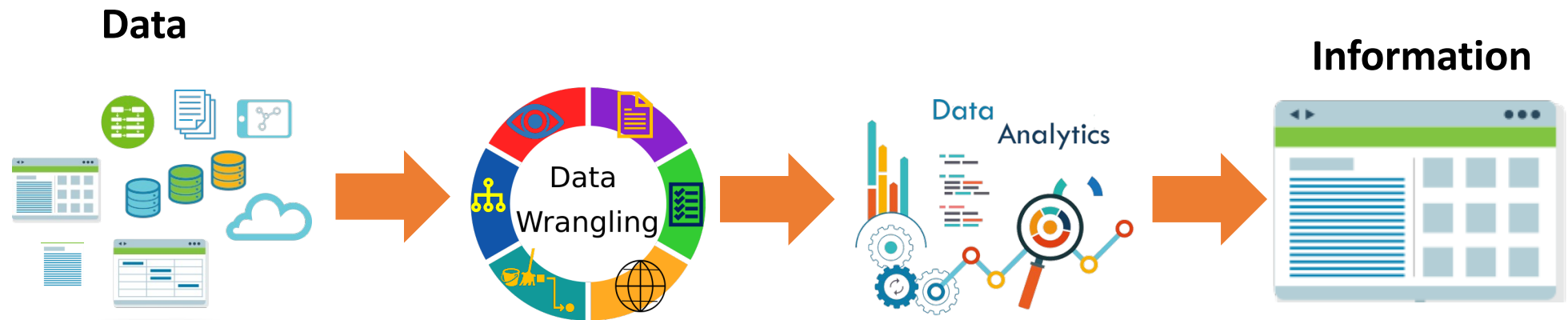


Konstantinos Nikoletos
k.nikoletos@uu.nl

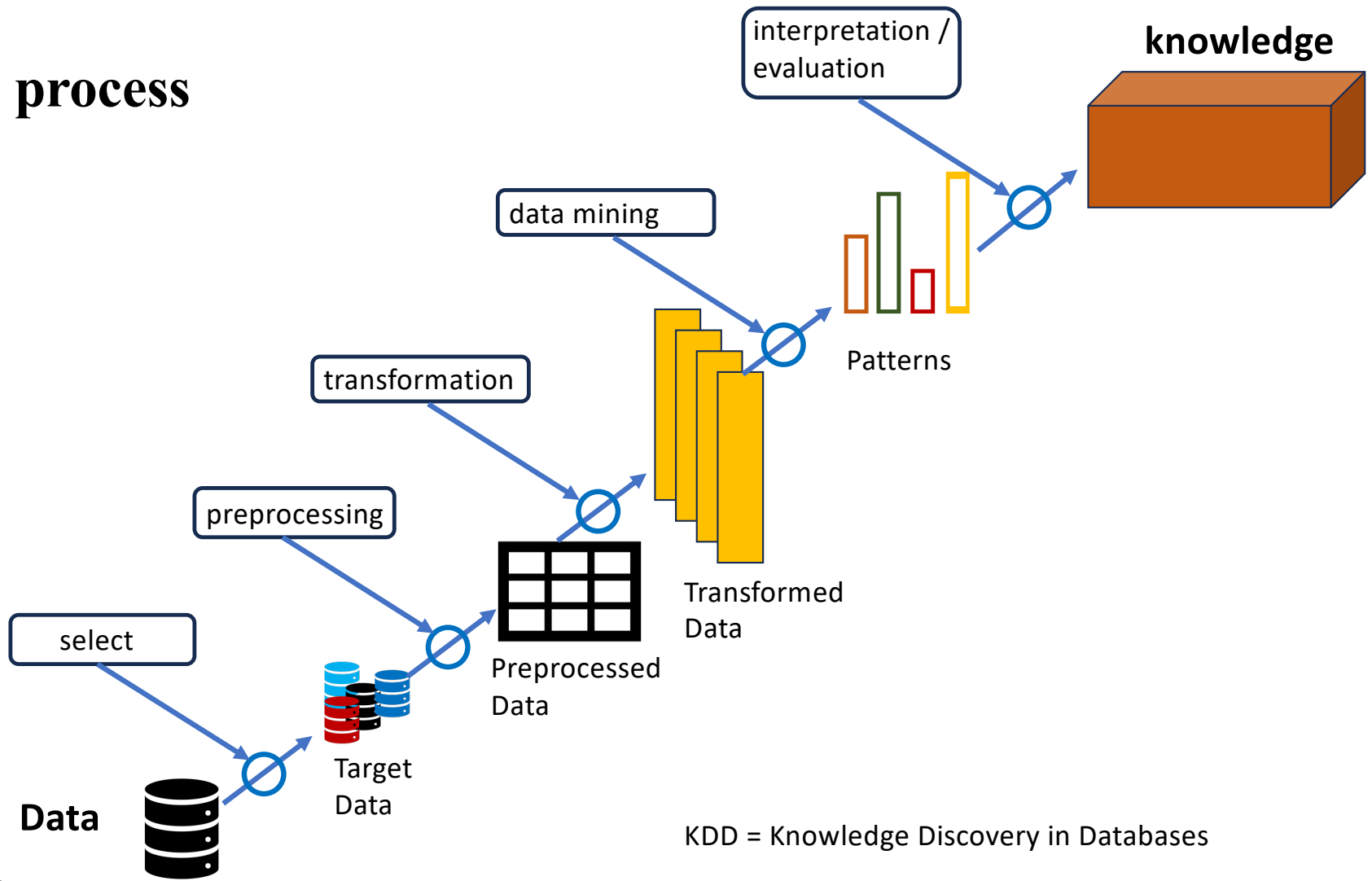


Course introduction

Extracting value from data

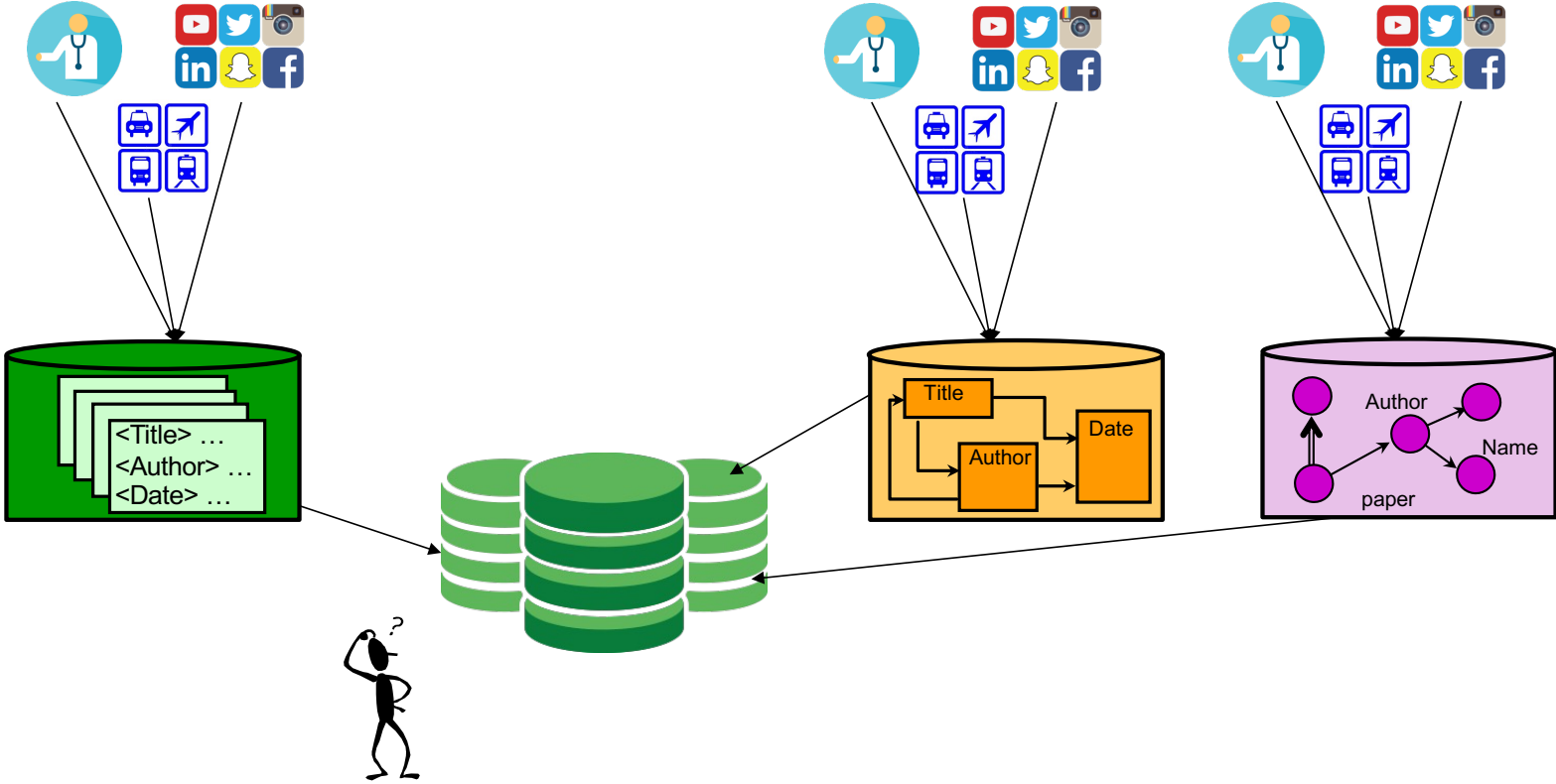


KDD process

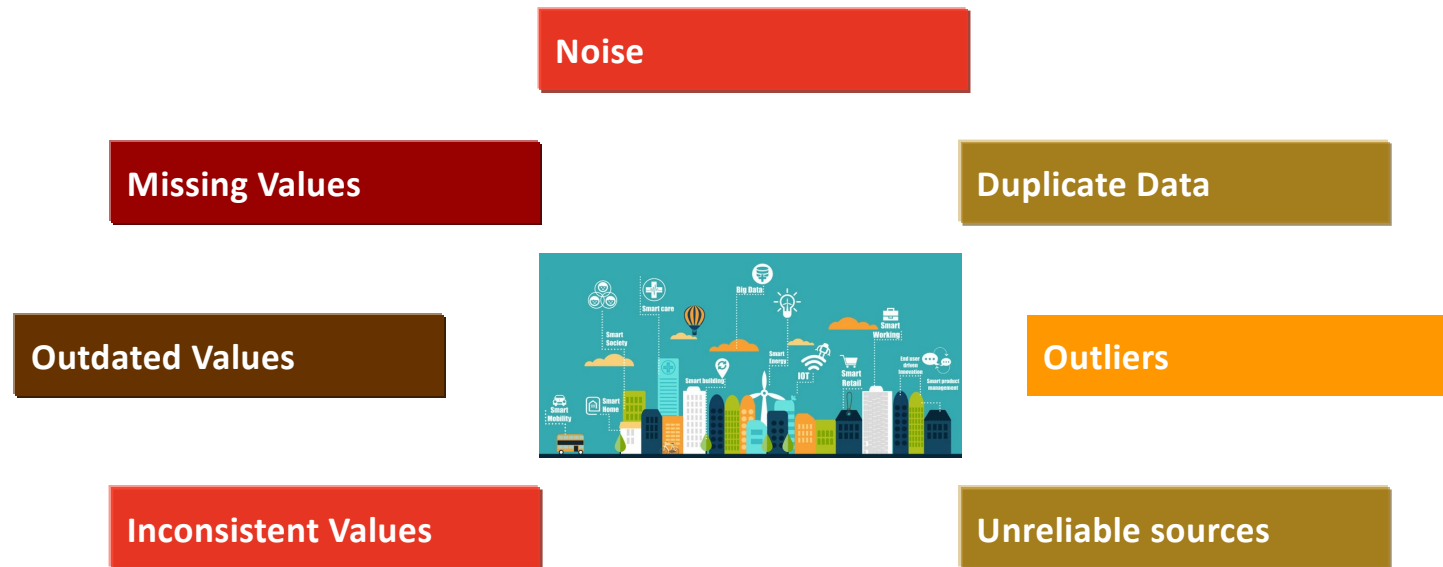


KDD = Knowledge Discovery in Databases

We collect a lot of data



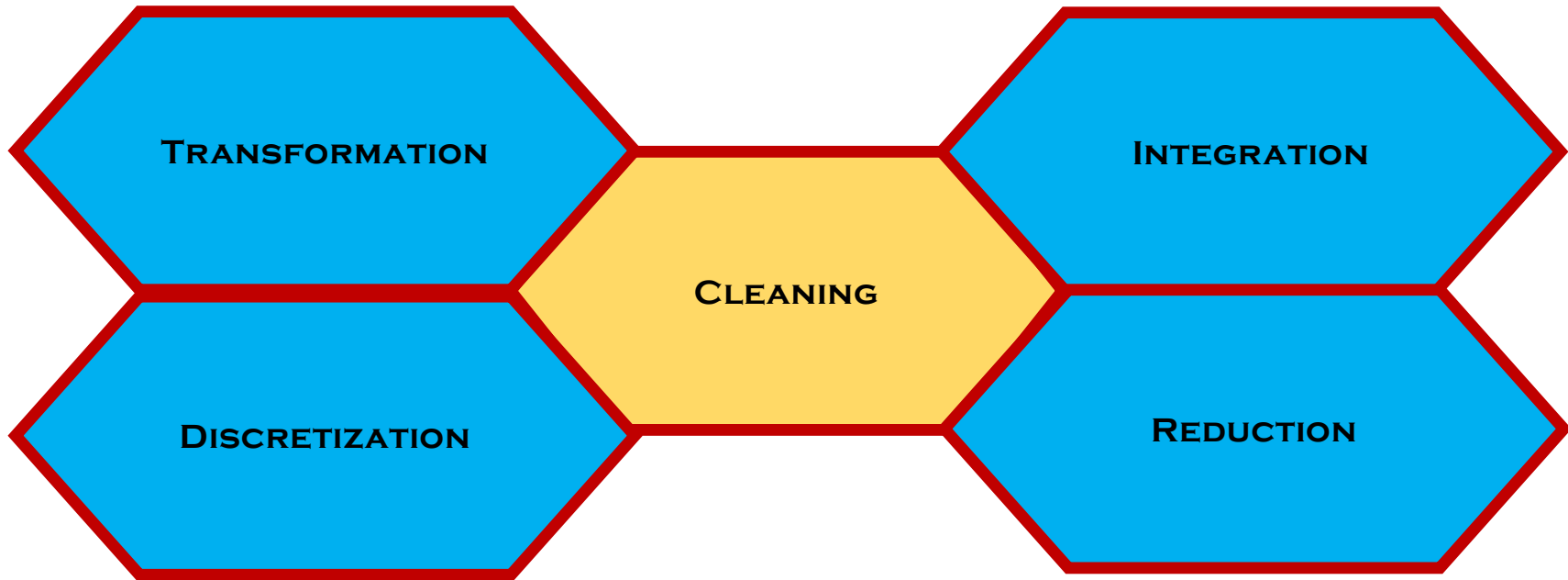
Data quality issues



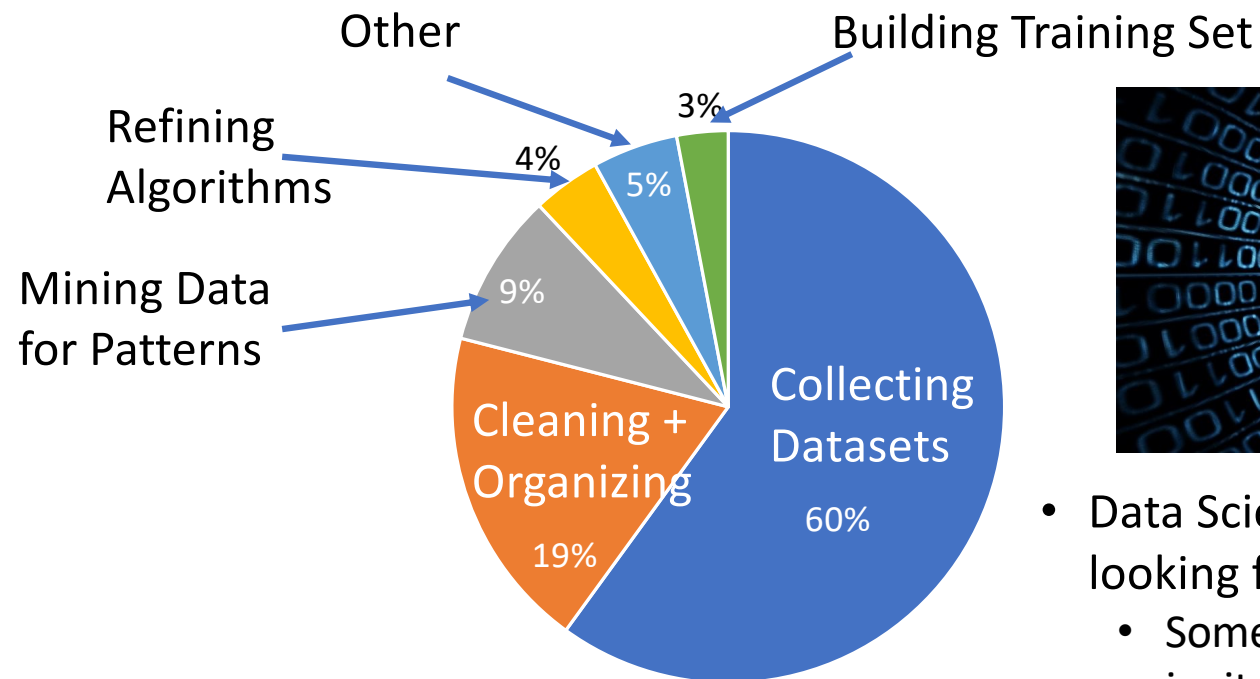
For any data retrieval or data analytiscs task,
it is critically important to know the quality of your data

For the US alone, poor data quality costs around US economy \$3 trillions a year (2016).

Data preparation

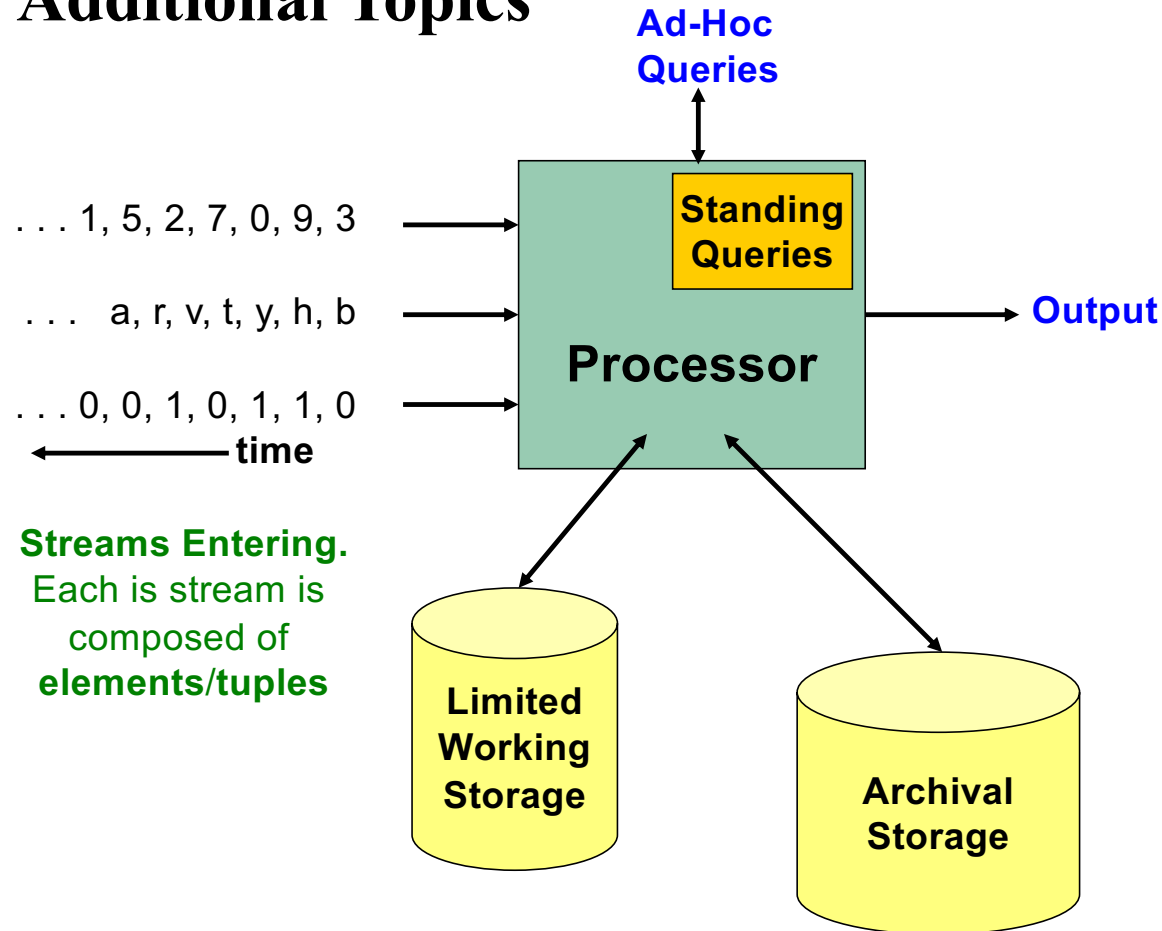


How are you going to spend your time?

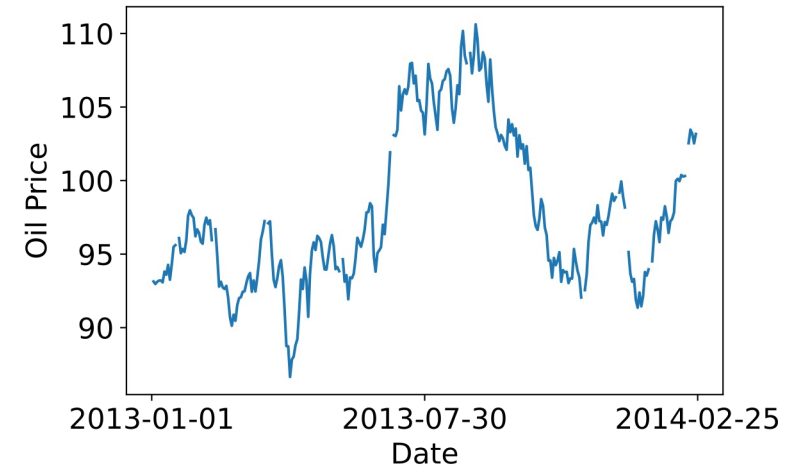


- Data Scientists: 80% of their time looking for and cleaning the data
 - Sometimes, they are called data janitors

Additional Topics

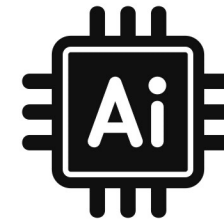


Algorithmic Fairness



This course is not

- Database course
- Systems course
- Artificial intelligence course
- Machine learning course



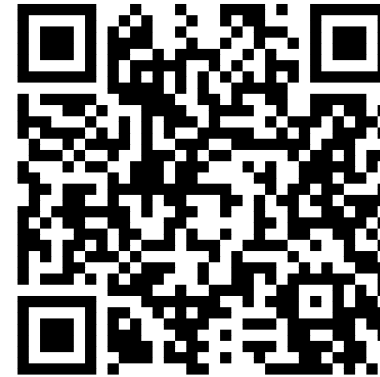


Go to
wooclap.com

Enter the
event code
in the top
banner

Event code

DW2627



<https://app.wooclap.com/DW2627>



Utrecht University

Break



Utrecht University

Practicalities



Utrecht University

<https://hqahtan.github.io/teaching.html>



Practicalities

- Everything is on the course website <https://hqahtan.github.io/teaching.html>, except:
 - Room locations (on mytimetable.uu.nl)
 - Announcements (in your email and on uu.brightspace.com)
 - Assignment hand-in (on uu.brightspace.com)
 - Exams (on the university digital assessment tool Remindo)
 - Lectures / tutorials will be in person



materials on the website are constantly under construction



Course flow

- This is a mandatory course.
- Every* week you will:



Read the required readings



Attend the lectures: Tuesday, Thursday



Attend and work on the hands-on tutorials: Thursday



Work on the group and the individual assignments



Review materials diligently for the midterm and final exams

* some weeks contain other stuff like exams and study time



Schedule & readings

- The course website contains a full schedule with required readings
- Reading materials can be found online for free! 🕵️ (you are resourceful students)



Lectures: Prepare for next lecture!

1. Look up reading materials in course website
2. Look up time and location for the lecture



Tutorials

- Hands-on tutorials with practical exercises
- Put into practice what you learnt in the lectures
- We incorporate real-world data and use-cases: APPLIED data science!
- One session per week with practical exercises from the theoretical lectures
- **The teaching assistant is your main point of contact for questions!**
- Tutorials provide skills needed to do the assignments



Assignments

- Every two weeks
 - Assignments are already available on the course website
 - deadlines on: weeks 3 and 7 for Group assignments and weeks 5 and 9 for individual assignments
- Hand in on Brightspace
- Group work!
 - You can assign yourself in a group starting from Friday (tomorrow) until Wednesday next week.
- **Plan ahead**



Assignments: prepare!

1. When is the deadline of the first assignment?
 1. Check the course website
2. Quickly check the content of the first assignment
3. Log in to Brightspace and find the INFOMDWR course
4. If we were organized coordinators:
 1. Find the group sign-up page there to see how you can sign up
 2. Find the assignment hand-in page to see how you can submit
5. If we are disorganized coordinators, don't worry, it'll be there soon 😊



Announcements

Further questions?



Utrecht University

See you in the tutorial session 🖐️
First session is optional!

