

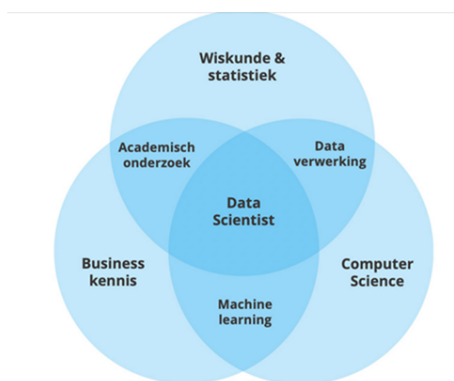


## Portfolio assignment 1

- What is Data Science?
  - You will be split up into breakrooms with 3-4 students each.
  - 10 min: Do online research to answer the question "What is Data Science?"\*
  - 10 min: Create a PowerPoint slide with a summary of your results\*
- Optional: Same assignment but the question is "What does a Data Scientist do?"\*

\*Do not use Avans school materials as a source for this assignment

Data science is a method of transforming data into valuable business strategies. It is a tool for making sound decisions to increase profits, improve the customer journey and spot opportunities. It is a methodology for extracting insights from (semi) structured data collected primarily from databases and unstructured data. Unstructured data includes data collected from non-specified formats. This can be anything like a video, images, social media or even emails and Word files. Organizations are using data science to analyze the status of business goals. This manifests itself in various forms to grow sales and reduce costs.



## Data Science

- PostNL, KLM en CBS
- Algoritmen & Wiskunde
- Grote kwantiteit data
- Communicatie
- Verbeteren bedrijfsproces
- Verbanden tussen data
- Verschillende bronnen
- Generen nieuwe data uit bestaande data

Think for example of product development, detecting bottlenecks in production and customer experience optimization. Data science is a good innovative way to make economic predictions. Data science is the generation of new data from existing data. The outcome is therefore not a service but an information product. Data science makes the creation of new data products possible.

The numerical analysis of business processes with the aim of improving these business processes: this is often called analytics in the management sciences. Analytics, or business analytics, or sometimes a little bit more advanced analytics: in essence, it all means the same thing. The improvement objective depends on the context. Think of increasing efficiency, improving quality, or reducing risks.

Of course, there are many ways to analyze this, but analytics is about quantitative, numerical analyses, often based on statistics and algorithms. Data science is a field that focuses on gaining insights from data. Data scientists use programming skills, statistical knowledge, and sector and organizational knowledge to make important connections.



## Portfolio assignment 2

**avans**  
hogeschool

- What are the most popular data science tools?
  - You will be split up into breakrooms with 3-4 students each.
  - 10 min: Do online research to answer the question "What are the most popular data science tools?"\*
  - 10 min: Create a PowerPoint slide with a summary of your results
- Optional: Same assignment but the question is "What are the characteristics that make these tools popular for data science tasks?"\*

\*Do not use Avans school materials as a source for this assignment

There are a lot of languages that can be used to create your software. Here are a few languages used for Big Data, along with the popularity.

### Favoriete data Science tools

Software	2019 % share	2018 % share	2017 % share
Python	65.8%	65.6%	59.0%
RapidMiner	51.2%	52.7%	31.9%
R Language	46.6%	48.5%	56.6%
Excel	34.8%	39.1%	31.5%
Anaconda	33.9%	33.4%	24.3%
SQL Language	32.8%	39.6%	39.2%
Tensorflow	31.7%	29.9%	22.7%
Keras	26.6%	22.2%	10.7%
scikit-learn	25.5%	24.4%	21.9%
Tableau	22.1%	26.4%	21.8%
Apache Spark	21.0%	21.5%	25.5%

Different languages have different purposes. Python is very high-level and has a wide variety of use-cases, so this language will mainly be used throughout our project.