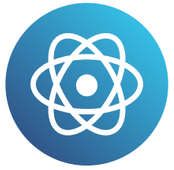
What is data science?

DATA SCIENCE FOR EVERYONE

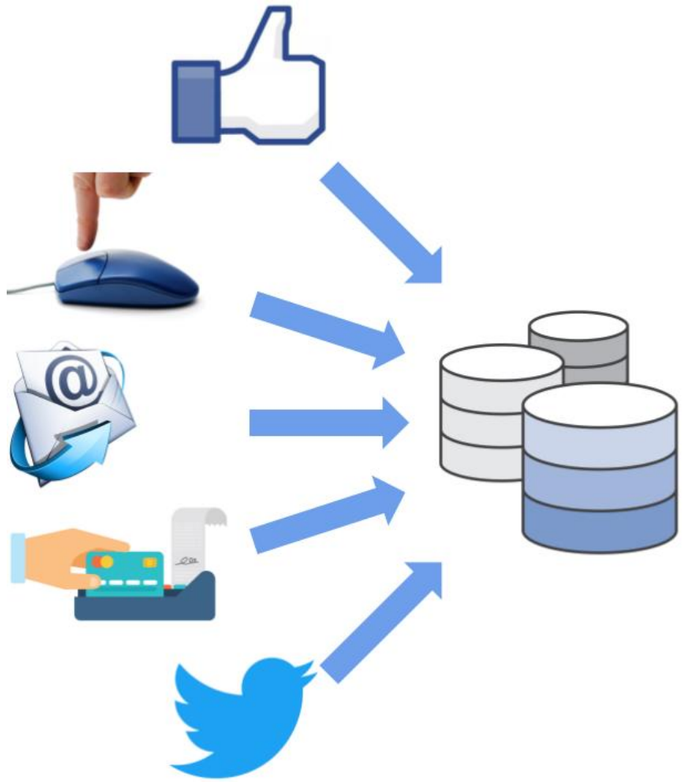


Lis Sulmont
Curriculum Manager, DataCamp

DataCamp

Let's ask Google!

Making data work for you



Use data to better describe the present or better predict the future

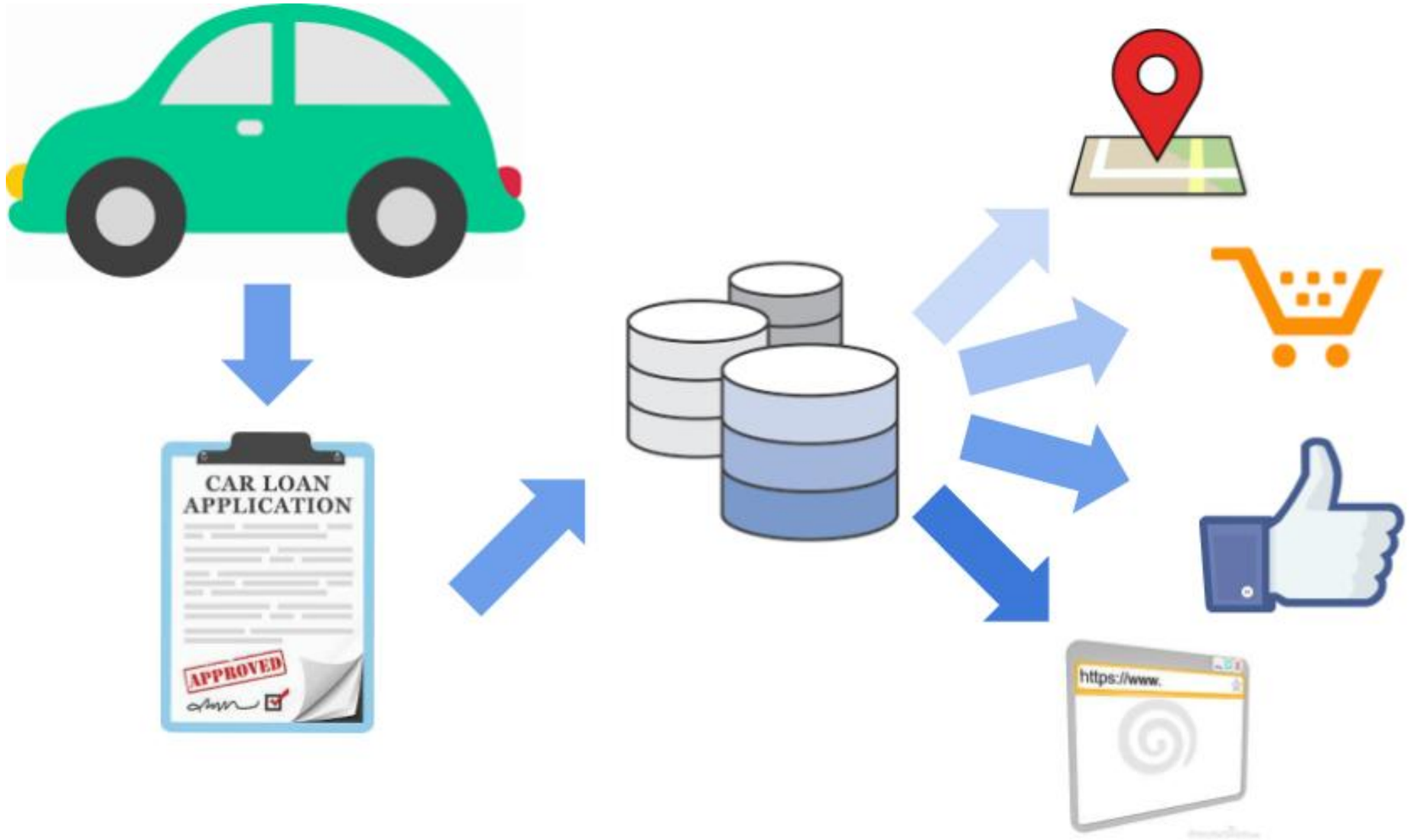
What can data do?

- Describe the current state of an organization or process
- Detect anomalous events
- Diagnose the causes of events and behaviors
- Predict future events

Why now?

Why now?

Why now?



The data science workflow

Data Collection & Storage

Data Preparation

Exploration & Visualization

Experimentation & Prediction

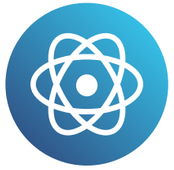
Let's practice!

DATA SCIENCE FOR EVERYONE

DataCamp

Applications of data science

DATA SCIENCE FOR EVERYONE



Lis Sulmont
Curriculum Manager, DataCamp

DataCamp

More case studies

- Traditional machine learning
- Internet of Things (IoT)
- Deep Learning

Case study: fraud detection

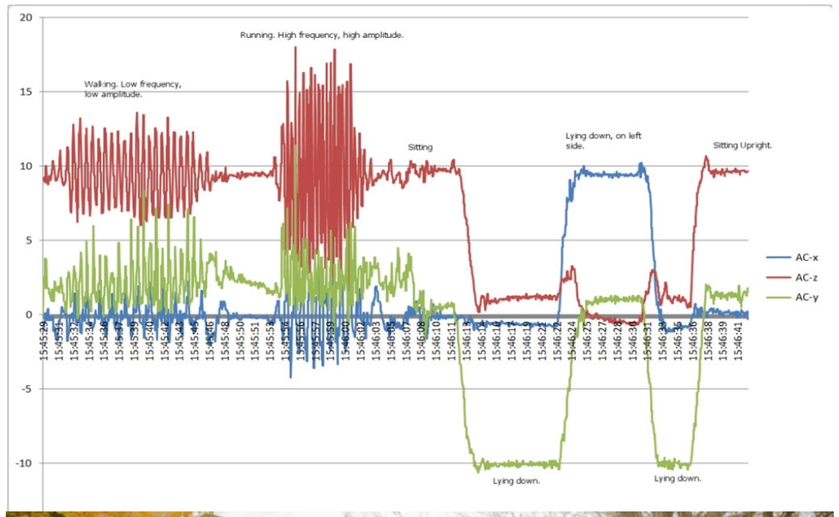
Case study: fraud detection

Amount	Date	Type	...
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

What do we need for machine learning?

- A well-defined question
  - *"What is the probability that this transaction is fraudulent?"*
- A set of example data
  - *Old transactions labeled as "fraudulent" or "valid"*
- A new set of data to use our algorithm on
  - *New credit card transactions*

Case study: smart watch



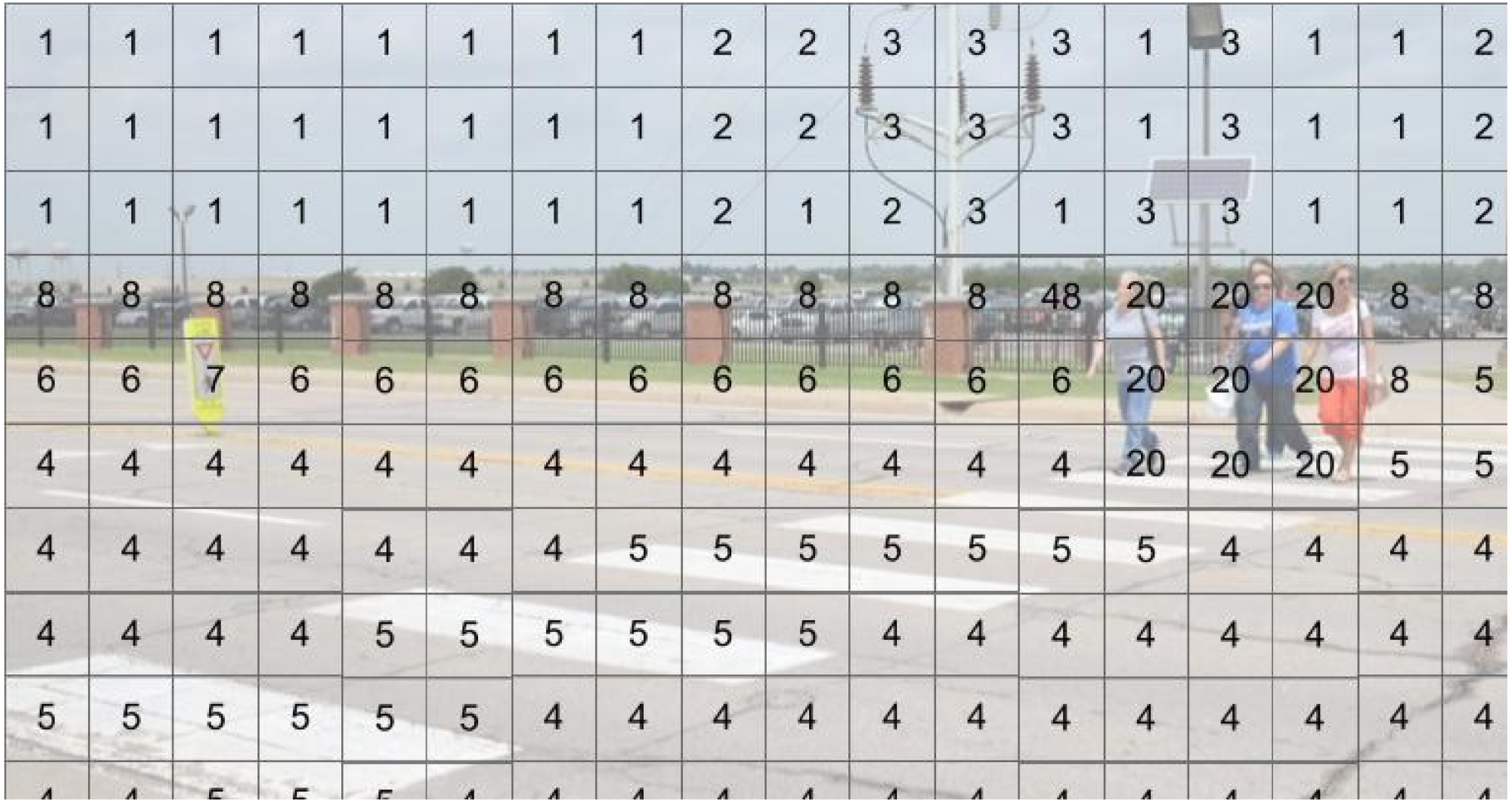
Internet of Things

Refers to gadgets that aren't standard computers

- Smart watches
- Internet-connected home security systems
- Electronic toll collection systems
- Building energy management systems
- Much, much more!

Case study: image recognition

Case study: image recognition



1	1	1	1	1	1	1	1	2	2	3	3	3	1	3	1	1	2
1	1	1	1	1	1	1	1	2	2	3	3	3	1	3	1	1	2
1	1	1	1	1	1	1	1	2	1	2	3	1	3	3	1	1	2
8	8	8	8	8	8	8	8	8	8	8	8	48	20	20	20	8	8
6	6	7	6	6	6	6	6	6	6	6	6	6	20	20	20	8	5
4	4	4	4	4	4	4	4	4	4	4	4	4	20	20	20	5	5
4	4	4	4	4	4	4	5	5	5	5	5	5	5	4	4	4	4
4	4	4	4	5	5	5	5	5	5	4	4	4	4	4	4	4	4
5	5	5	5	5	5	4	4	4	4	4	4	4	4	4	4	4	4
4	4	5	5	5	4	4	4	4	4	4	4	4	4	4	4	4	4

Deep learning

- Many neurons work together
- Requires much more training data
- Used in complex problems
  - Image classification
  - Language learning/understanding

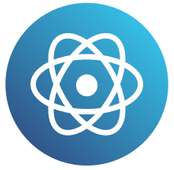
Let's practice!

DATA SCIENCE FOR EVERYONE

DataCamp

Data science roles and tools

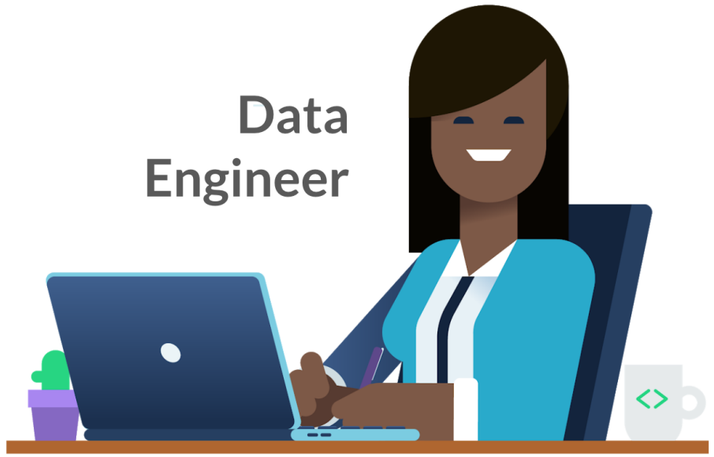
DATA SCIENCE FOR EVERYONE



Lis Sulmont
Curriculum Manager, DataCamp

DataCamp

Data Engineer



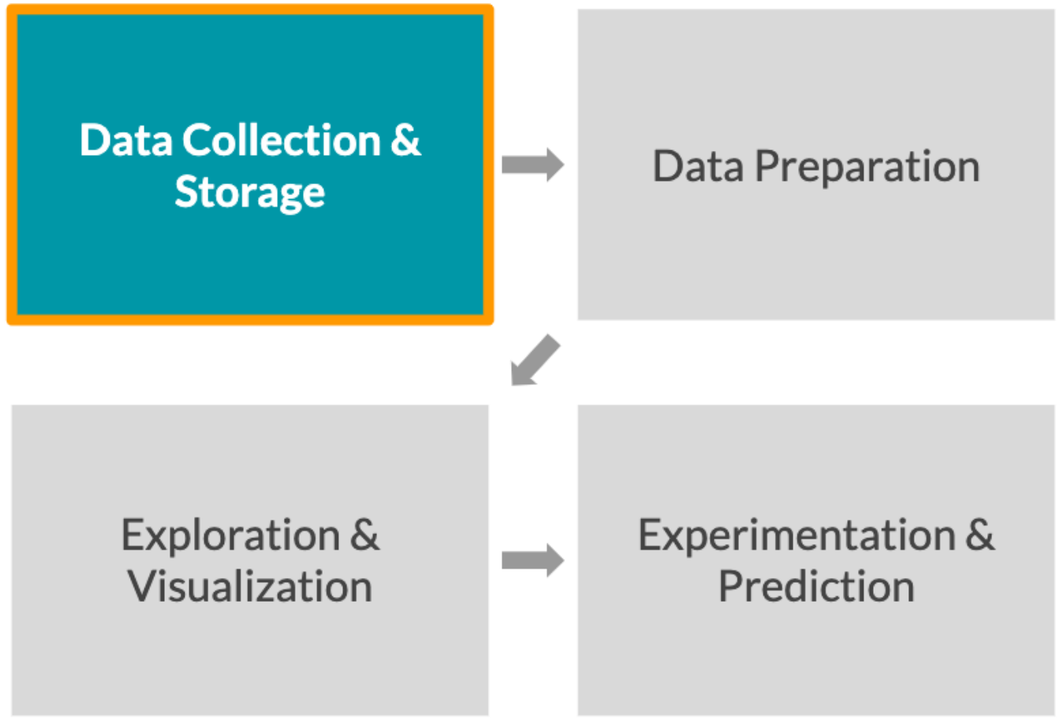
Data Analyst

Data Scientist

Machine Learning Scientist

Data engineer

- Information architects
- Build data pipelines and storage solutions
- Maintain data access

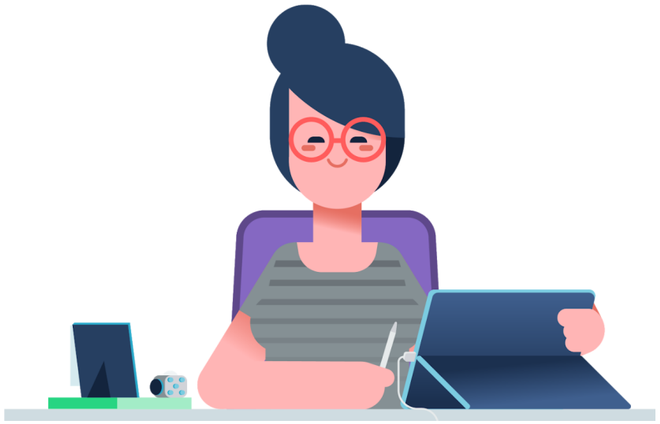


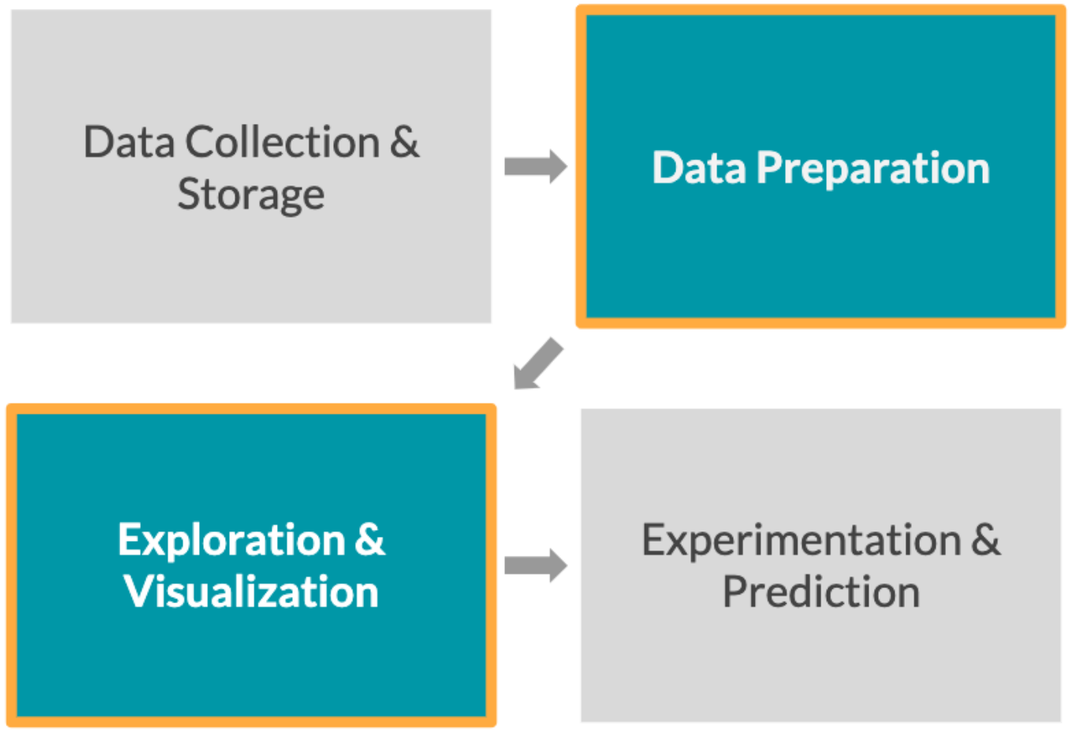
Data engineering tools

- **SQL**
  - To store and organize data
- **Java**, **Scala**, or **Python**
  - Programming languages to process data
- **Shell**
  - Command line to automate and run tasks
- **Cloud computing**
  - AWS, Azure, Google Cloud Platform

Data analyst

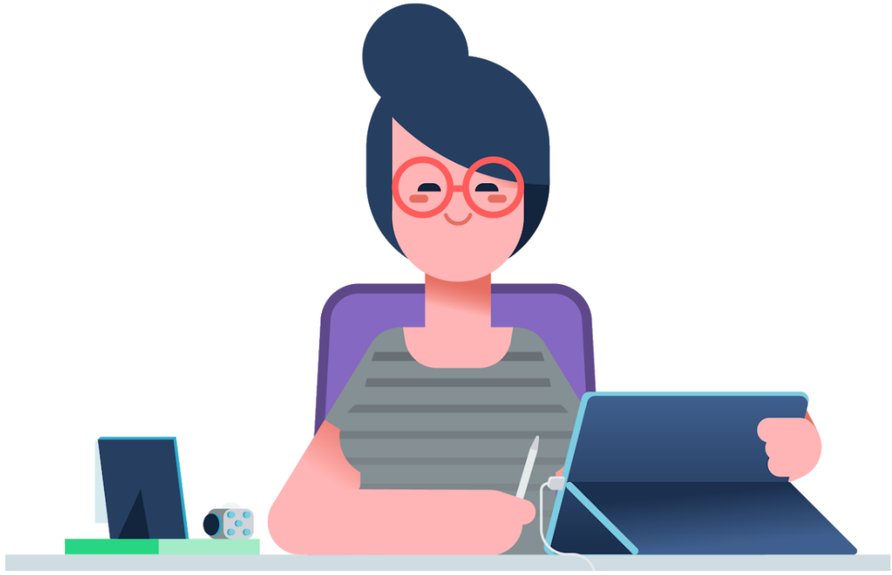
- Perform simpler analyses that describe data
- Reports and dashboards to summarize data
- Clean data for analysis





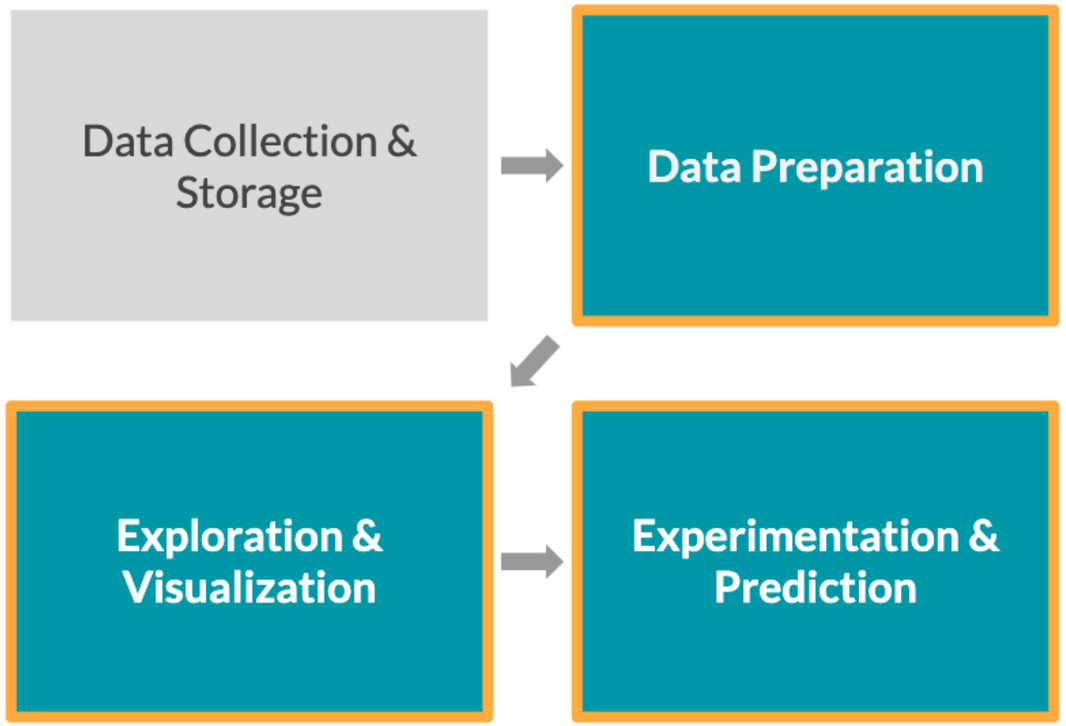
Data analyst tools

- **SQL**
  - Retrieve and aggregate data
- **Spreadsheets (Excel or Google Sheets)**
  - Simple analysis
- **BI Tools (Tableau, Power BI, Looker)**
  - Dashboards and visualizations
- *May have:* R and Python
  - Clean and analyze data



Data scientist

- Versed in statistical methods
- Run experiments and analyses for insights
- Traditional machine learning

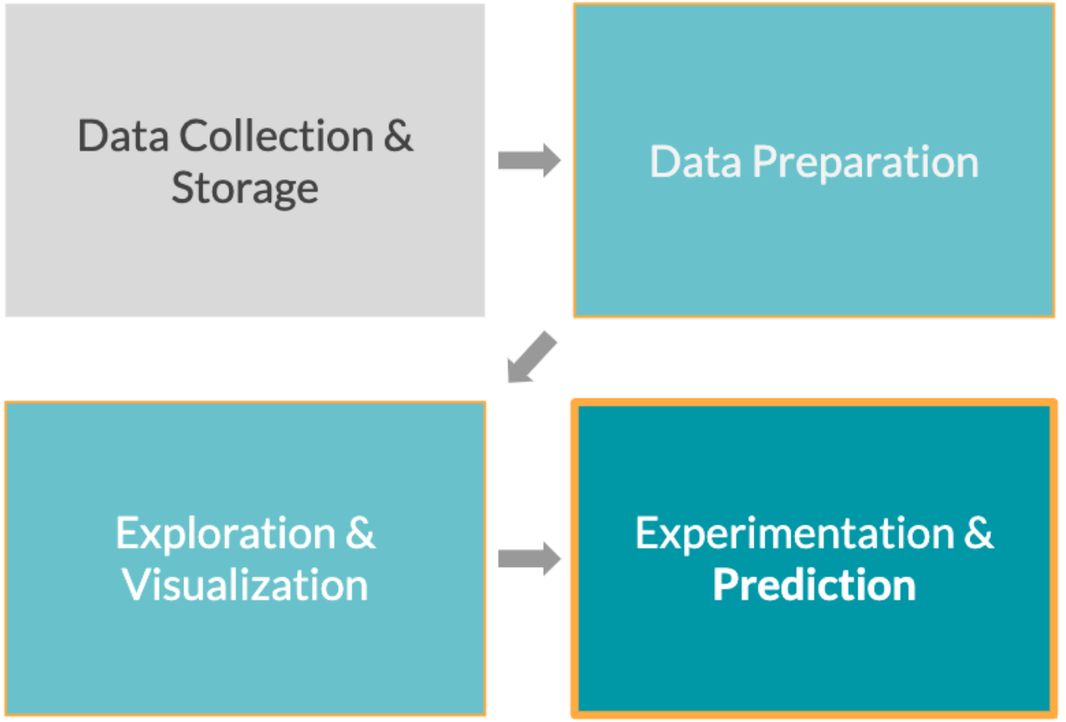


Data scientist tools

- **SQL**
  - Retrieve and aggregate data
- **Python and/or R**
  - With associated data science libraries, e.g., `pandas` (Python) and `tidyverse` (R)

Machine learning scientist

- Predictions and extrapolations
- Classification
- Deep Learning
  - Image processing
  - Natural language processing



Machine learning tools

- **Python and/or R**
  - With associated machine learning libraries, e.g., `TensorFlow` or `Spark`

Data Engineer	Data Analyst	Data Scientist	Machine Learning Scientist
Store and maintain data	Visualize and describe data	Gain insights from data	Predict with data
SQL + Java/Scala/Python	SQL + BI Tools + Spreadsheets	Python/R	Python/R

Let's practice!

DATA SCIENCE FOR EVERYONE

DataCamp