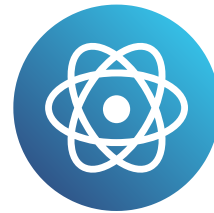


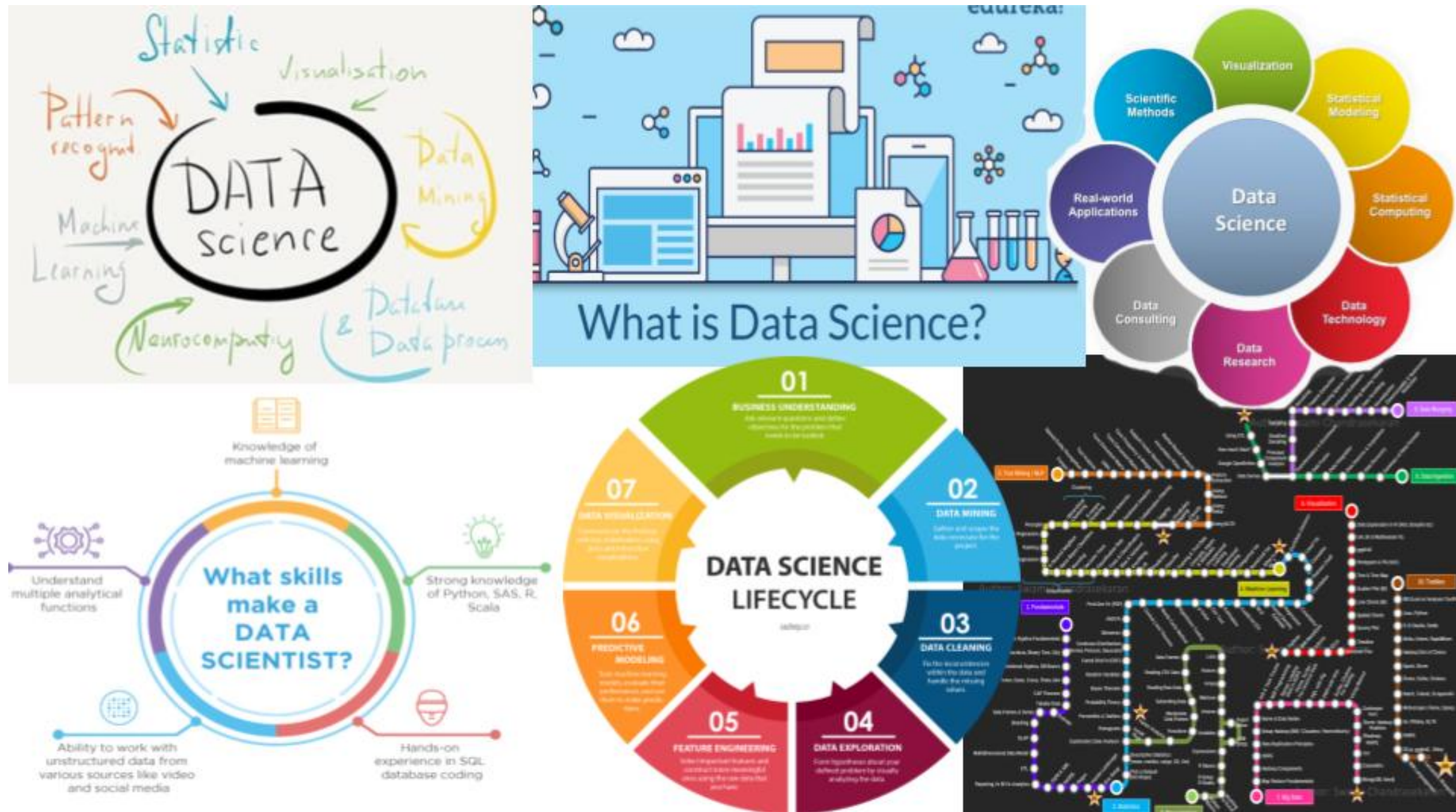
# What is Data Science?

DATA SCIENCE FOR BUSINESS

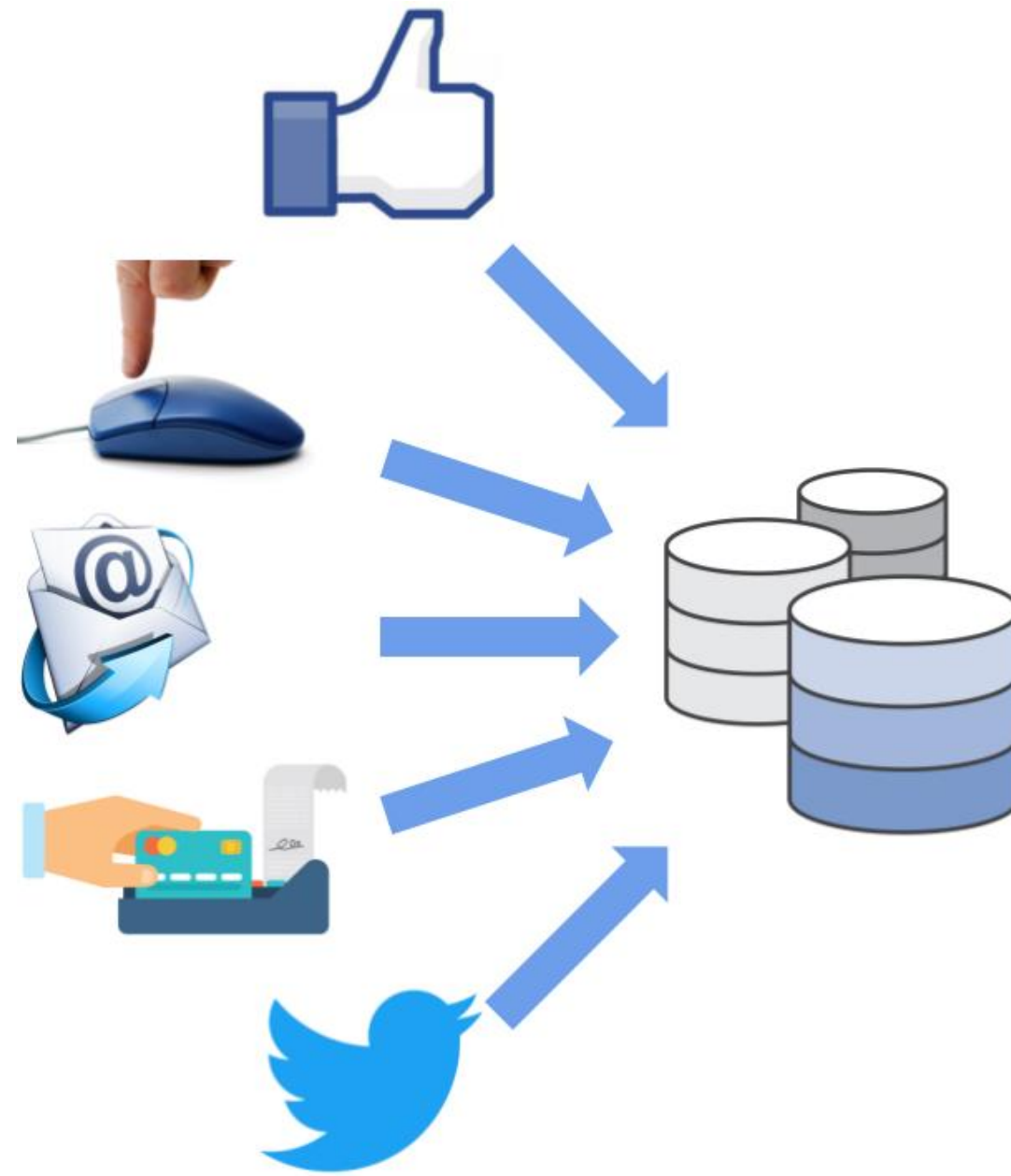


**Mari Nazary**  
VP of Content, DataCamp

# Let's ask Google!



# Making data work for you



# 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?

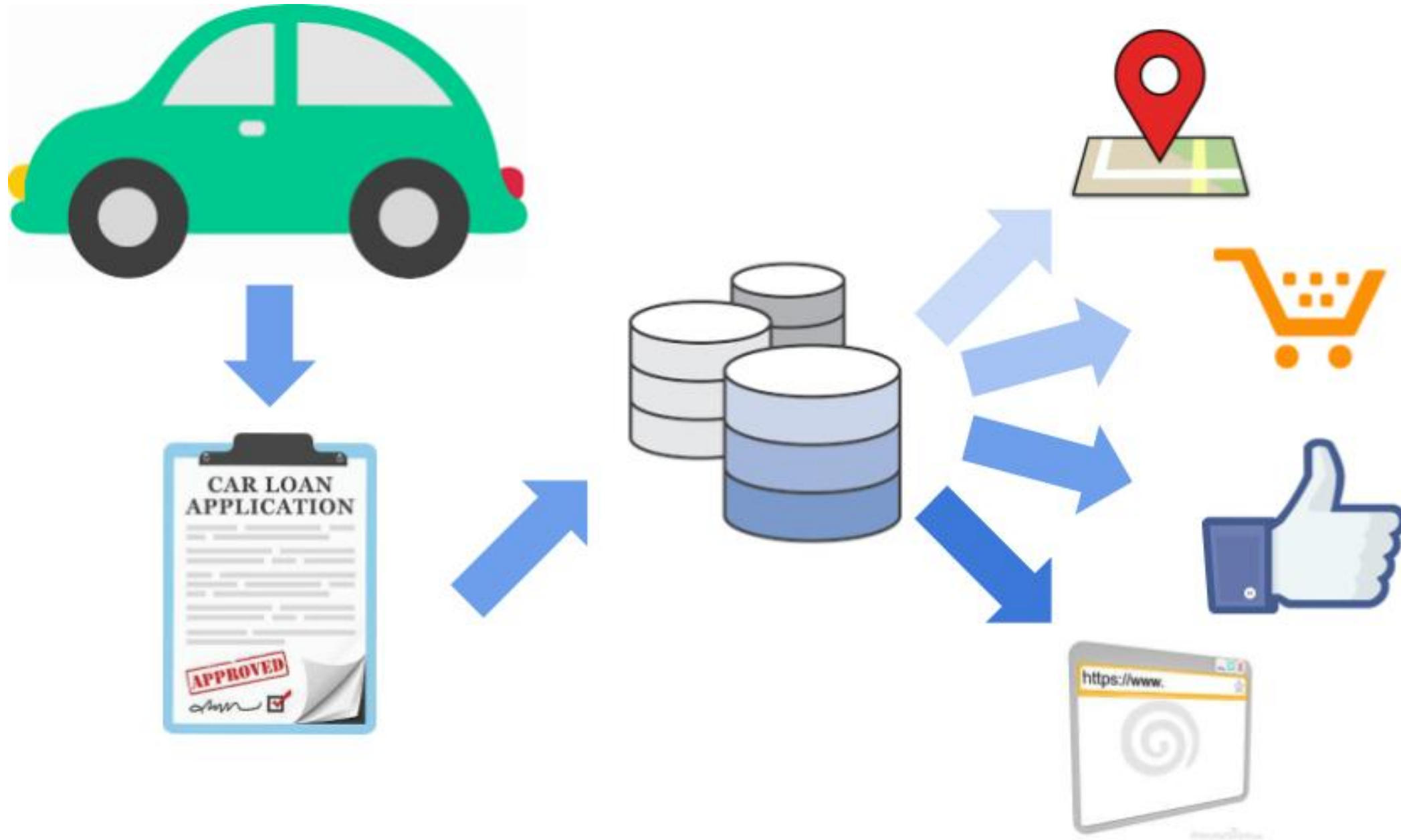


# Why now?





# Why now?

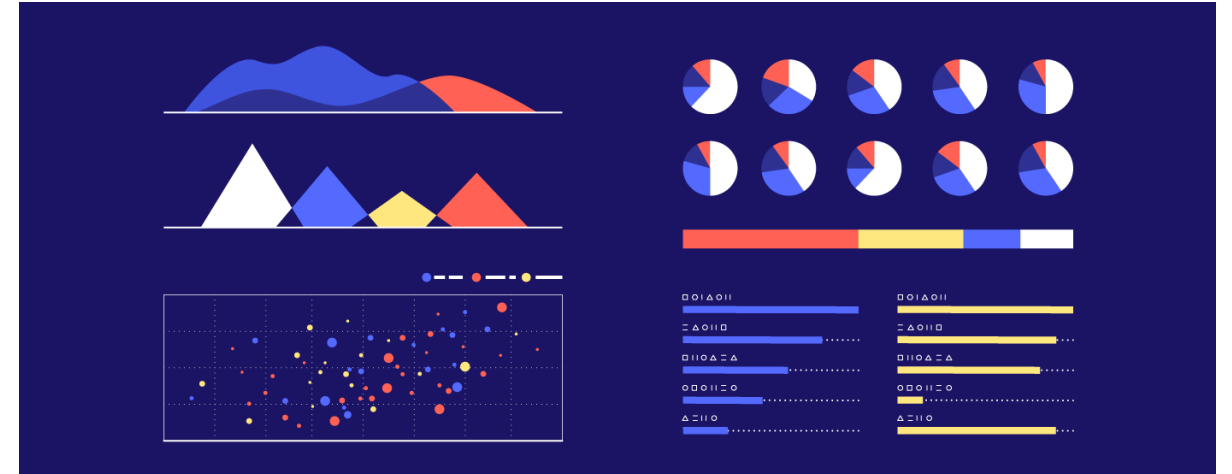


# The data science workflow

## Data collection



## Exploration and visualization



## Experimentation and prediction



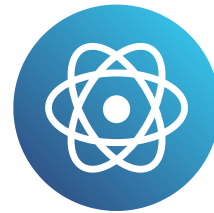


# Let's practice!

DATA SCIENCE FOR BUSINESS

# Applications of Data Science

DATA SCIENCE FOR BUSINESS



**Mari Nazary**  
VP of Content, DataCamp

# More case studies

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

# Case study: fraud detection



# Case study: fraud detection



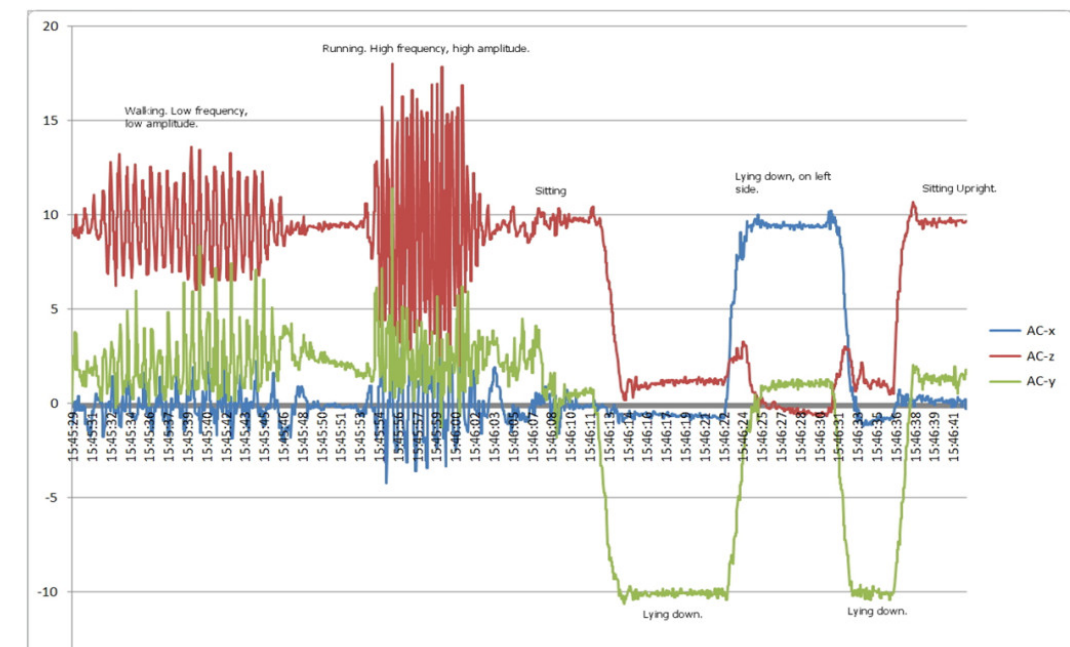
Amount	Date	Type	...
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

# 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

- 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



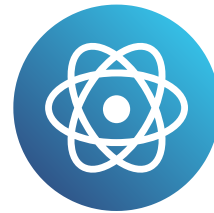
# Let's practice!

DATA SCIENCE FOR BUSINESS



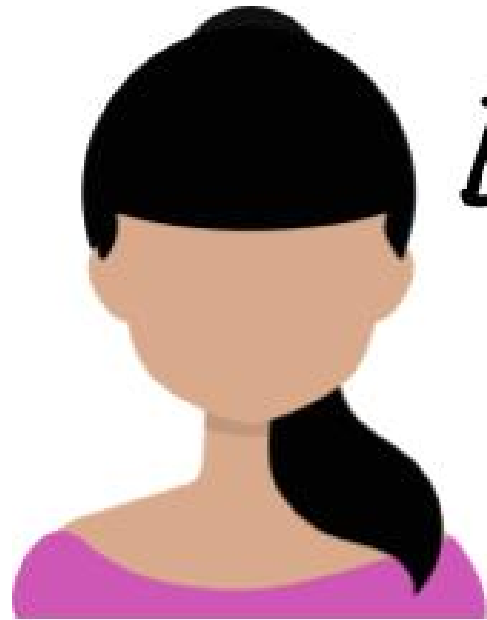
# Building a data science team

DATA SCIENCE FOR BUSINESS

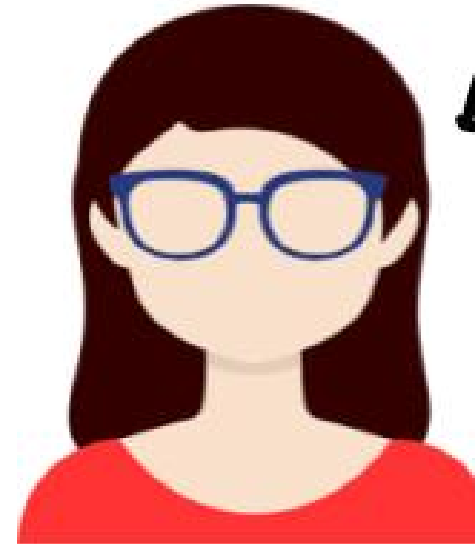


**Mari Nazary**  
VP of Content, DataCamp

# Members of your team



*Data Engineer*

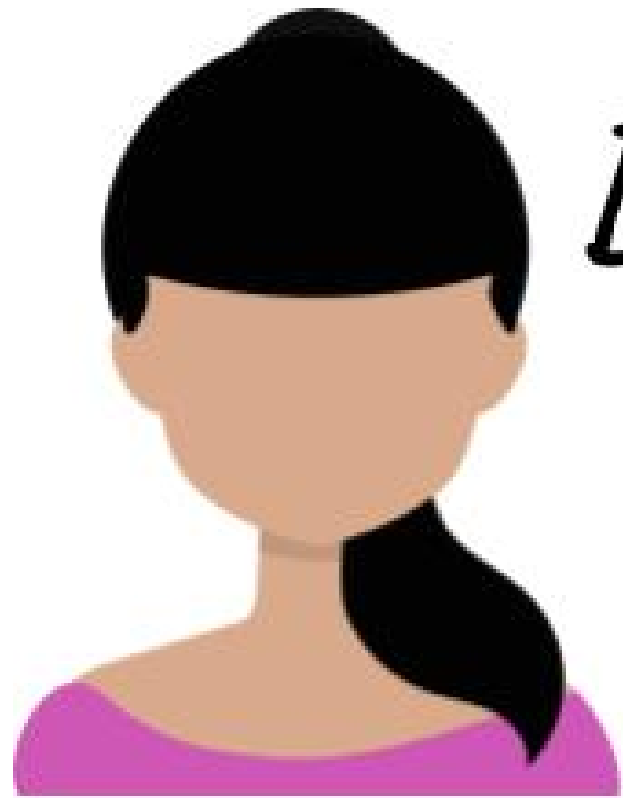


*Data Analyst*



*Machine Learning  
Scientist*

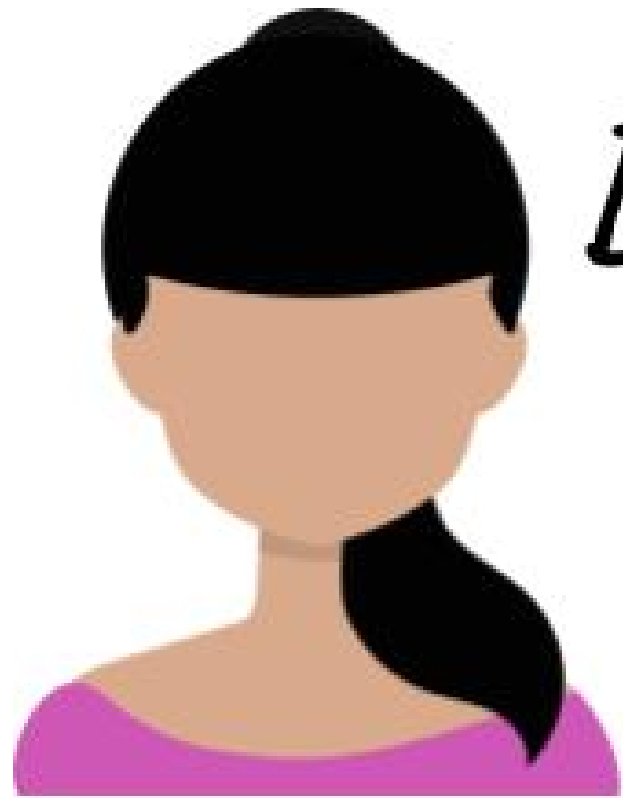
# Data engineering



*Data Engineer*

- Information architects
- Build storage solutions
- Maintain data access

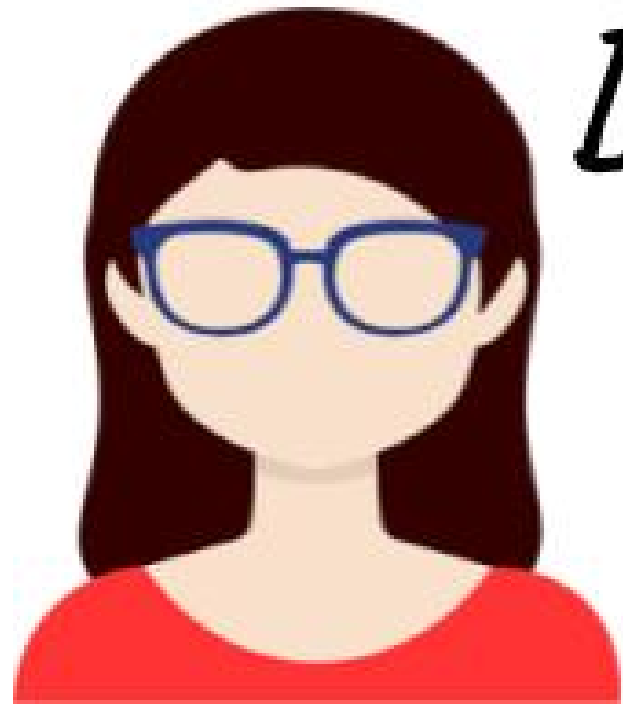
# Data engineering tools



*Data Engineer*

- SQL
  - Storing large quantities of data
- Java, Scala, or Python
  - Programming languages for processing data and automating tasks

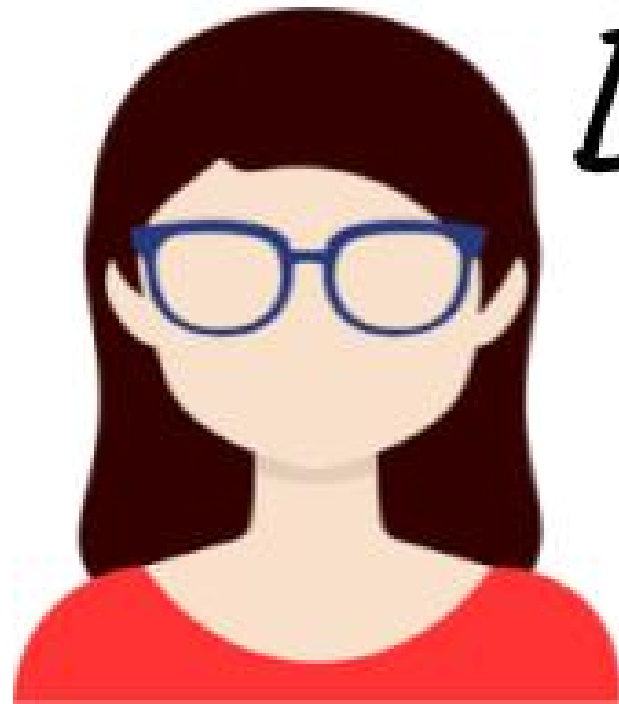
# Data analysis



*Data Analyst*

- Creating dashboards
- Hypothesis testing
- Data visualization

# Data analysis tools



*Data Analyst*

- Spreadsheets (Excel or Google Sheets)
  - Simple storage and analysis
- SQL
  - Large-scale analysis
- BI Tools (Tableau, Power BI, Looker)
  - Dashboarding and sharing information



# Machine learning



*Machine Learning  
Scientist*

- Predictions and extrapolations
- Classification
- Stock price prediction
- Image processing
- Automated text analysis

# Machine learning tools



*Machine Learning  
Scientist*

- Python and R
  - Programming languages for creating predictive models

# Review: members of your team

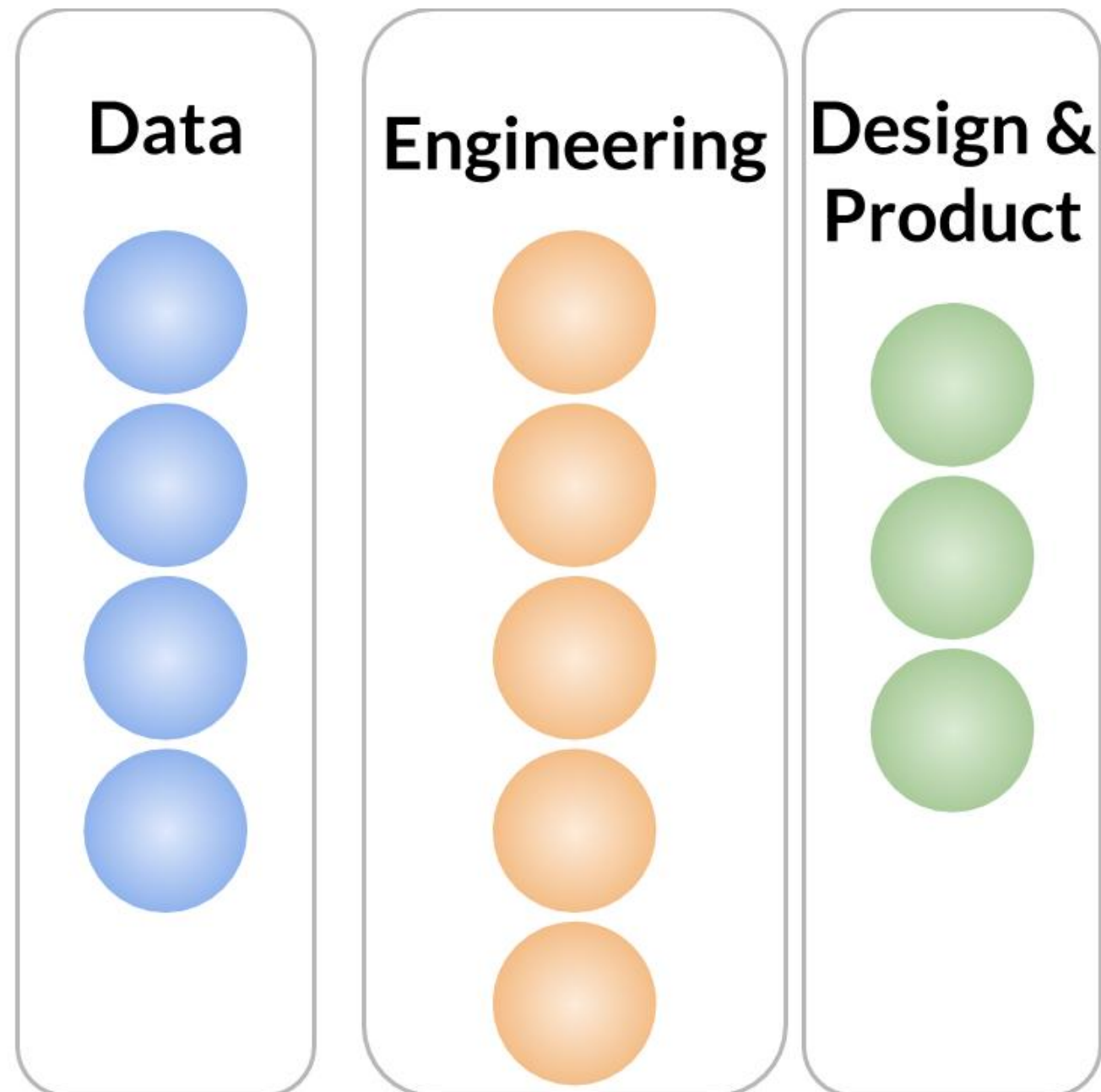


Data Engineer	Data Analyst	Machine Learning Scientist
Store and maintain data	Visualize and describe data	Model and predict with data
SQL + Java/Scala/Python	SQL + BI Tools + Spreadsheets	Python/R

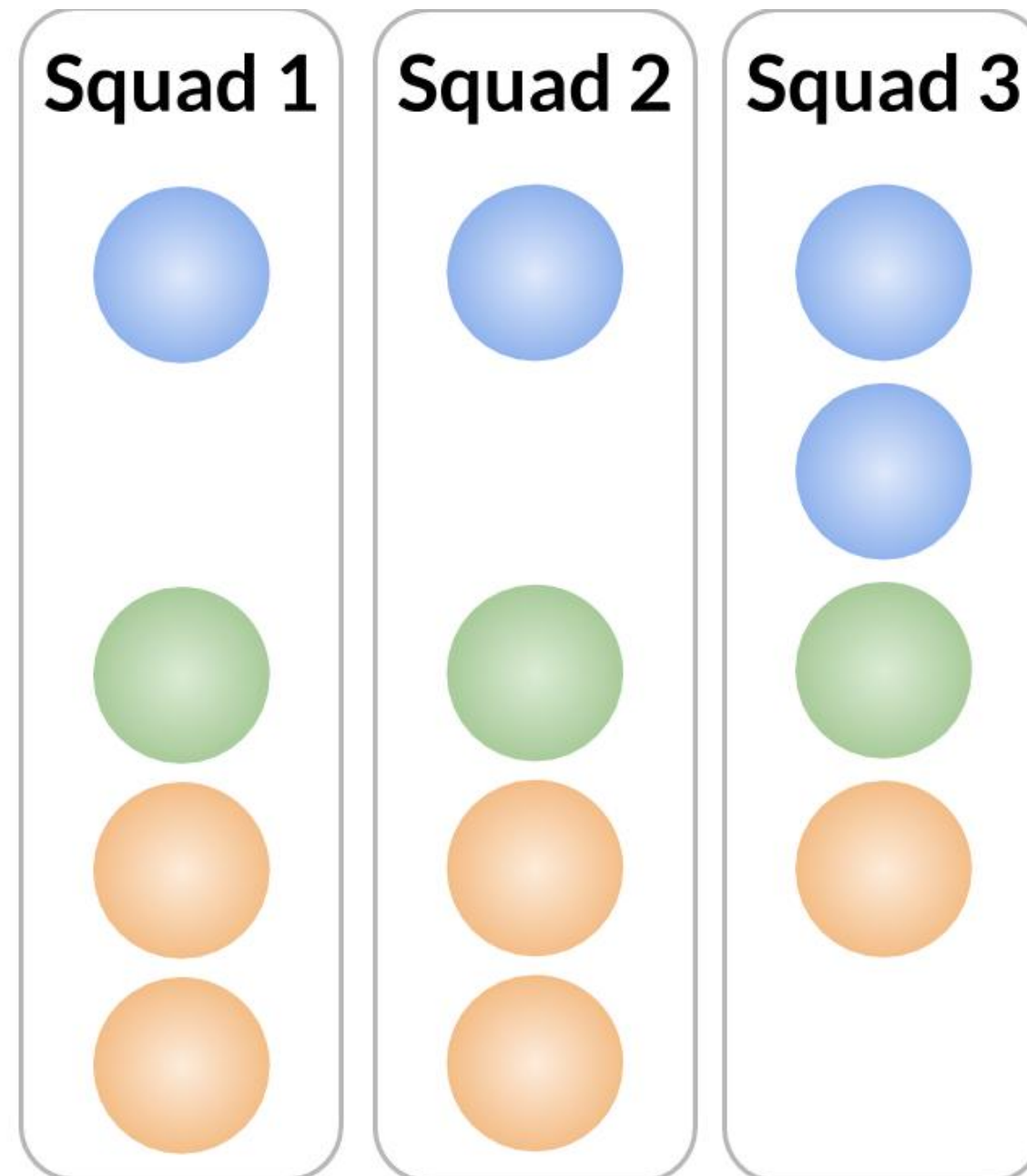
# Data science team structure

- Isolated
- Embedded
- Hybrid

# Team structure: isolated

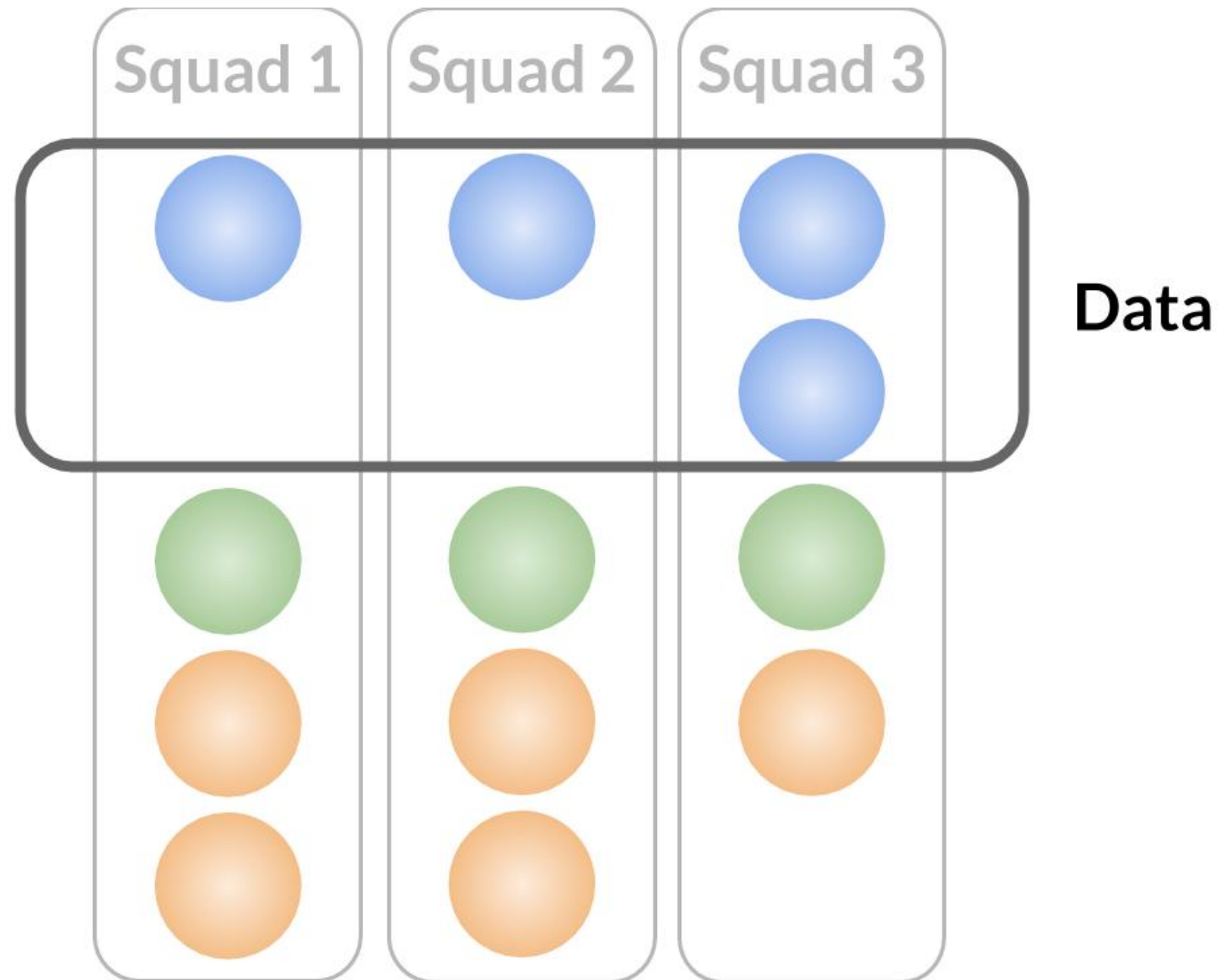


# Team structure: embedded





# Team structure: hybrid



# Let's practice!

DATA SCIENCE FOR BUSINESS