



# From statistics to serverless: building intelligent applications on OpenShift 4

William Benton

Marius Bogoevici

Sophie Watson



# What you'll learn in this session

Processes and principles to solve problems with ML

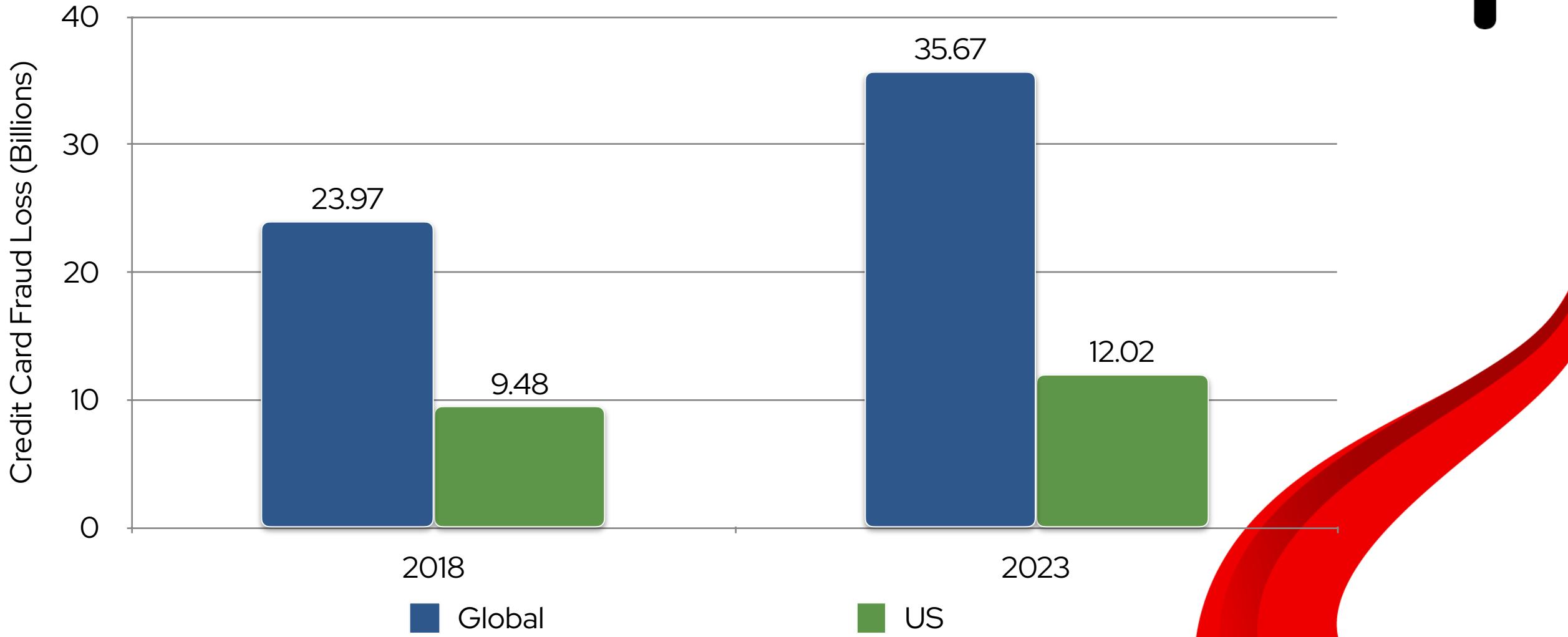
Some special considerations for time-series and payments data

How to train, evaluate, and deploy ML models

How OpenShift makes it possible to efficiently manage  
deployments with serverless infrastructure



## Card Fraud Losses in Numbers



# Fighting fraud with machine learning

- Drivers:
  - **Loss** and risk: financial and reputational
  - Increased **complexity** of money flows and fraud scenarios beyond human analysis
  - Minimize loss impact - act early
  - **Proactive and adaptive** fraud detection processes are required
- Outcomes:
  - Glean **insights** from large quantities of data and a variety of data sources
  - Increased **accuracy** of fraud detection process
  - **Adaptability** to changing conditions
- Challenges with machine learning
  - **Explainability**
  - **Performance**
  - **Operationalization**

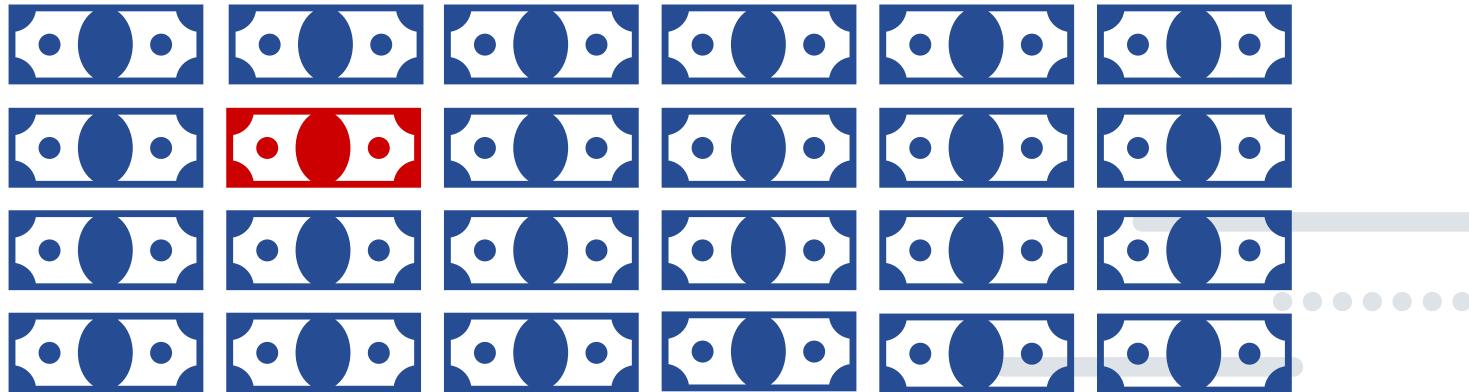




# codifying problem and metrics

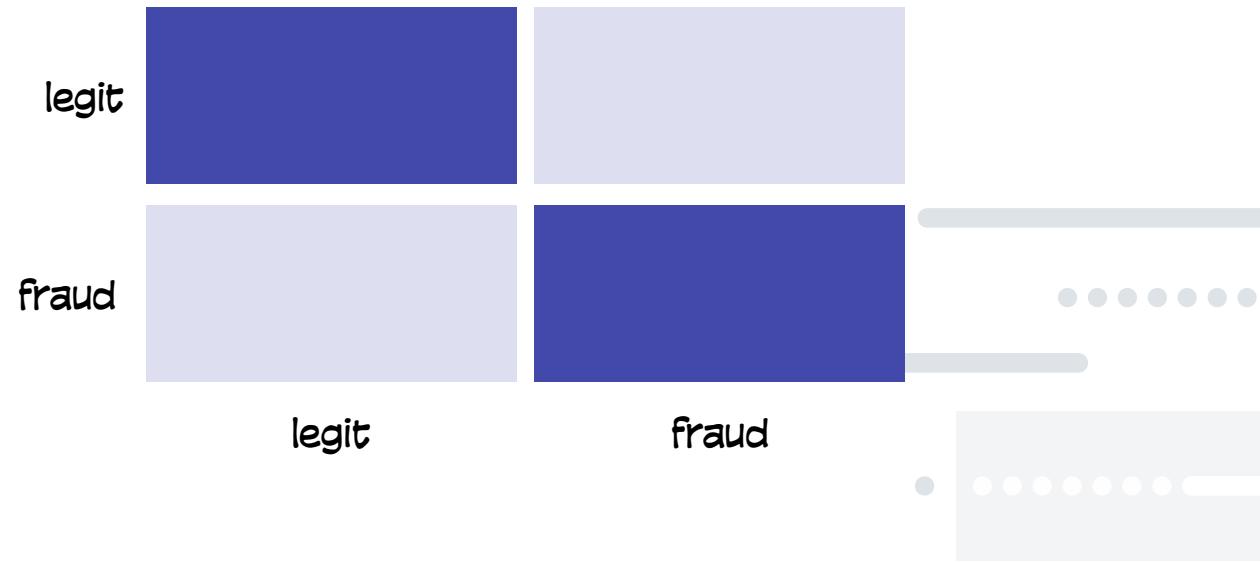


# codifying problem and metrics



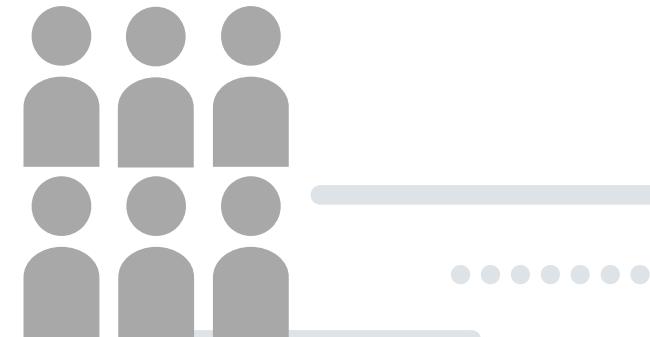
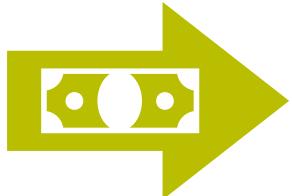
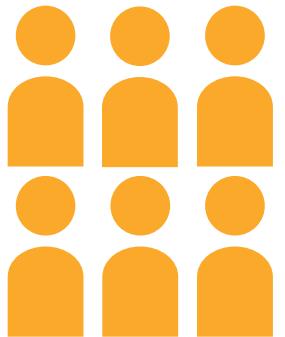


# codifying problem and metrics





# data collection and cleaning





# data collection and cleaning



- transaction time
- transaction amount
- transaction type (in-person, online, contactless, etc.)
- merchant and user IDs
- other details



# data collection and cleaning



- transaction time
- transaction amount
- transaction type (in-person, online, contactless, etc.)
- merchant and user IDs
- other details



# data collection and cleaning



- transaction time
- transaction amount
- transaction type (in-person, online, contactless, etc.)
- merchant and user IDs
- other details



# data collection and cleaning



- transaction time
  - transaction amount
  - transaction type (in-person, online, contactless, etc.)
  - merchant and user IDs
  - other details
- what was local time for this user  
when this transaction happened?



## data collection and cleaning



- transaction time
- transaction amount
- transaction type (in-person, online, contactless, etc.)
- merchant and user IDs
- other details

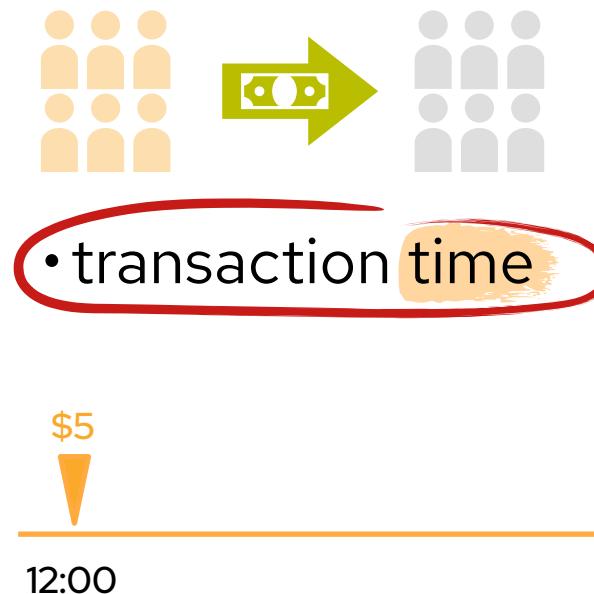
how long has it been since the last transaction this user initiated?

.....

what was local time for this user when this transaction happened?



## data collection and cleaning



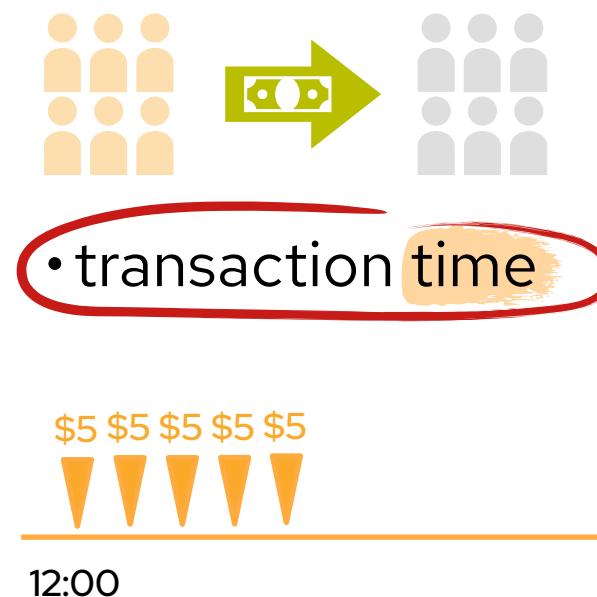
how long has it been since the last transaction this user initiated?

12:00

13:00



## data collection and cleaning



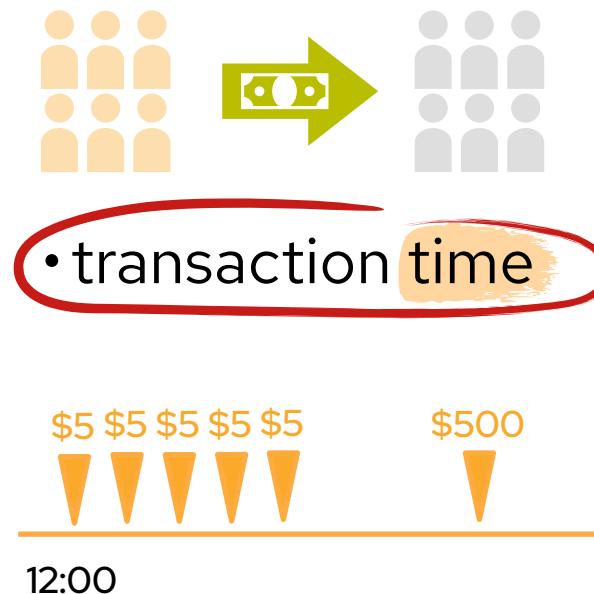
how long has it been since the last transaction this user initiated?

12:00

13:00



## data collection and cleaning



how long has it been since the last transaction this user initiated?



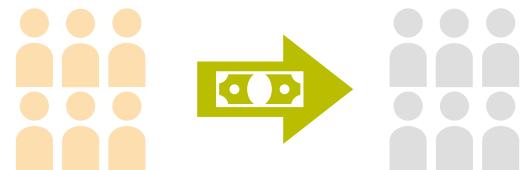
# data collection and cleaning



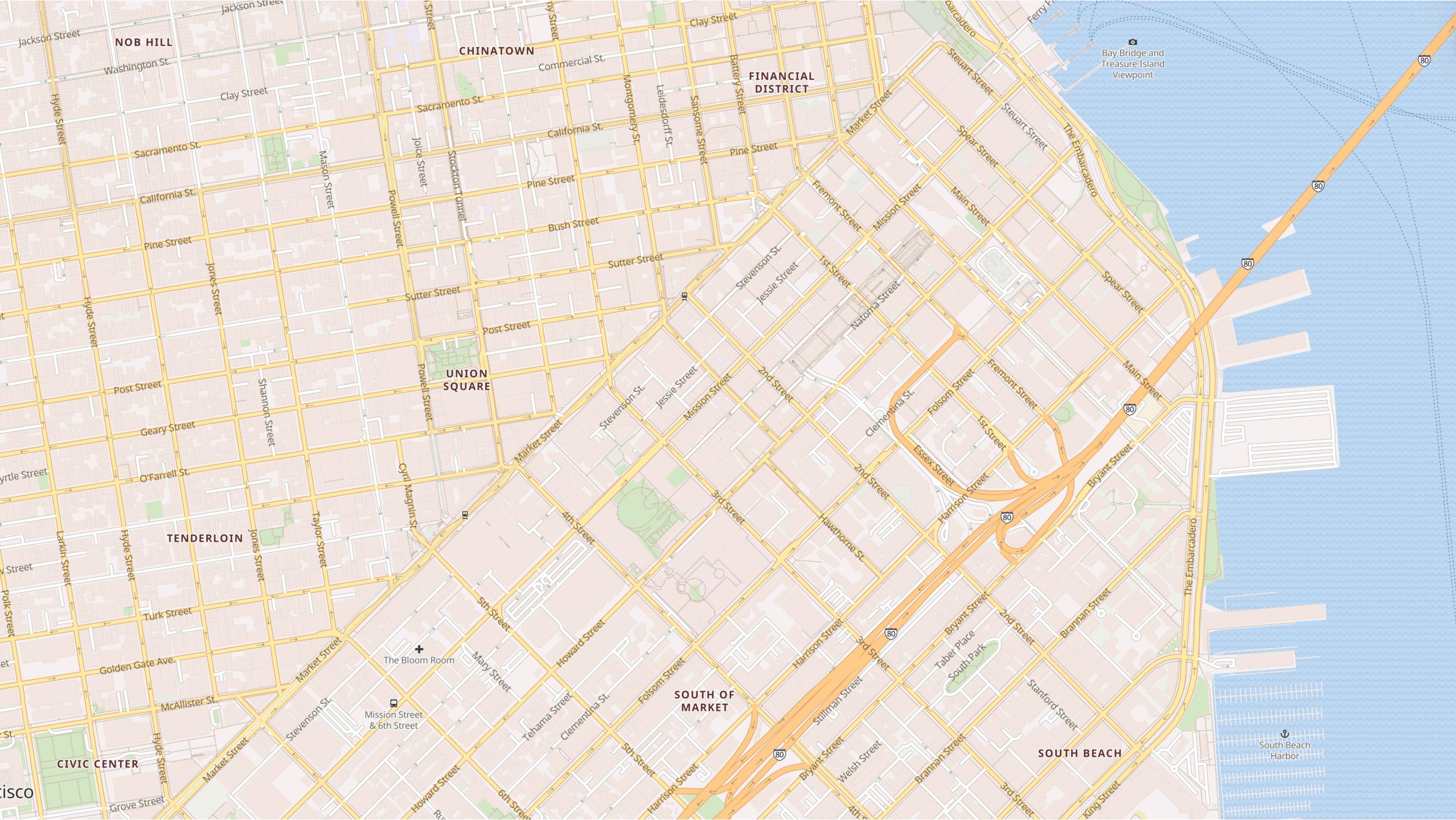
- transaction time
- transaction amount
- transaction type (in-person, online, contactless, etc.)
- merchant and user IDs
- other details

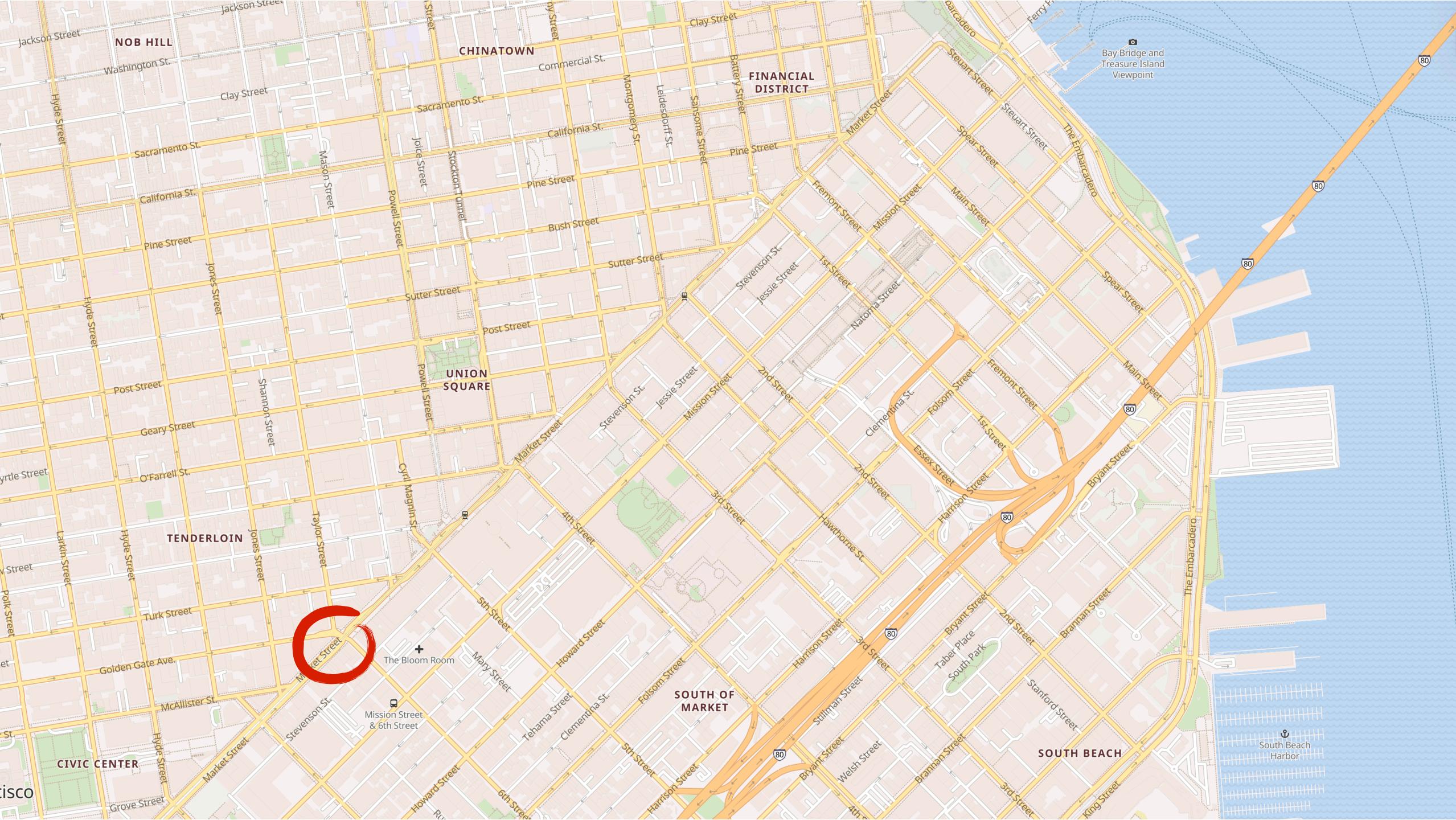


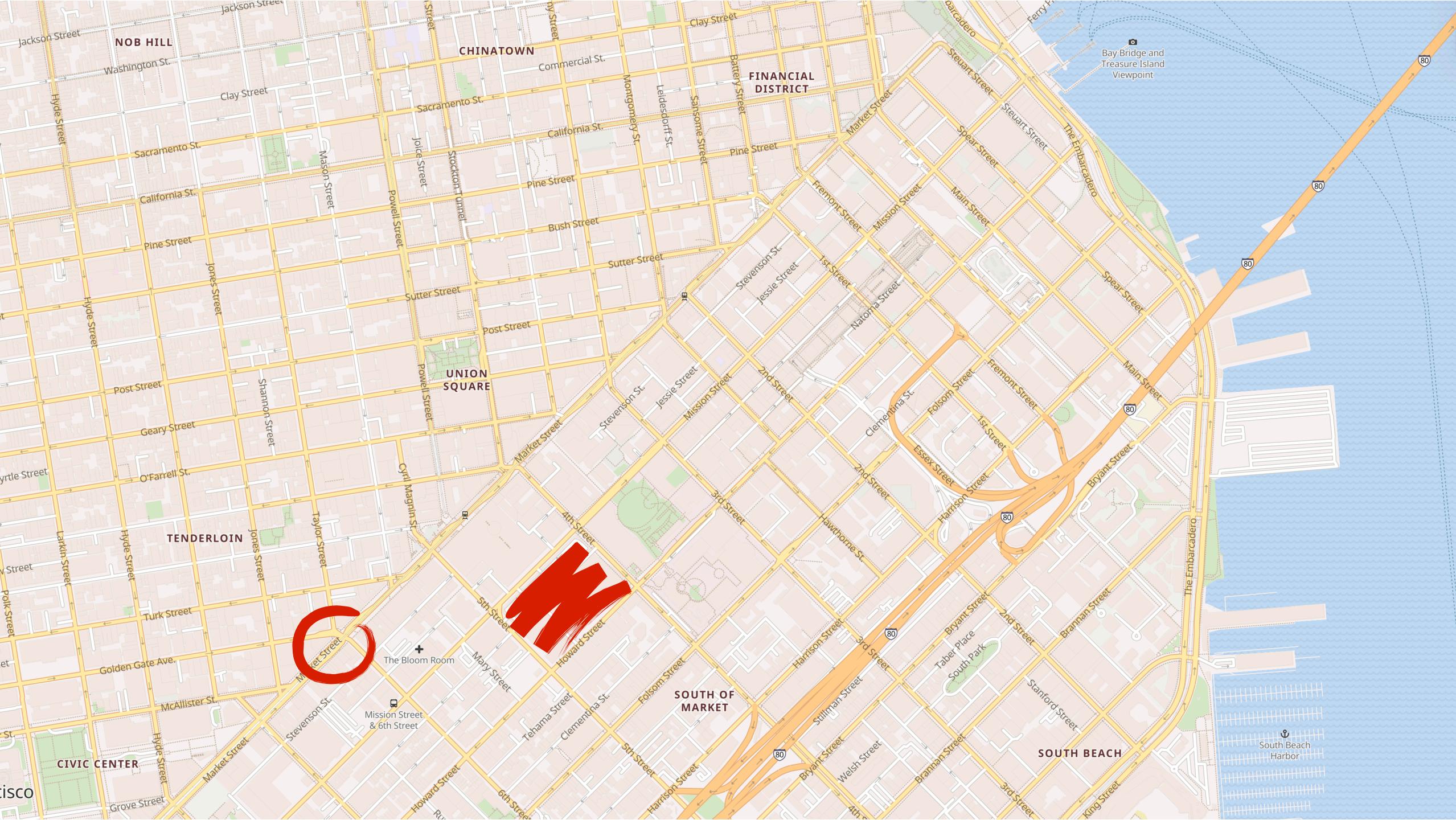
# data collection and cleaning

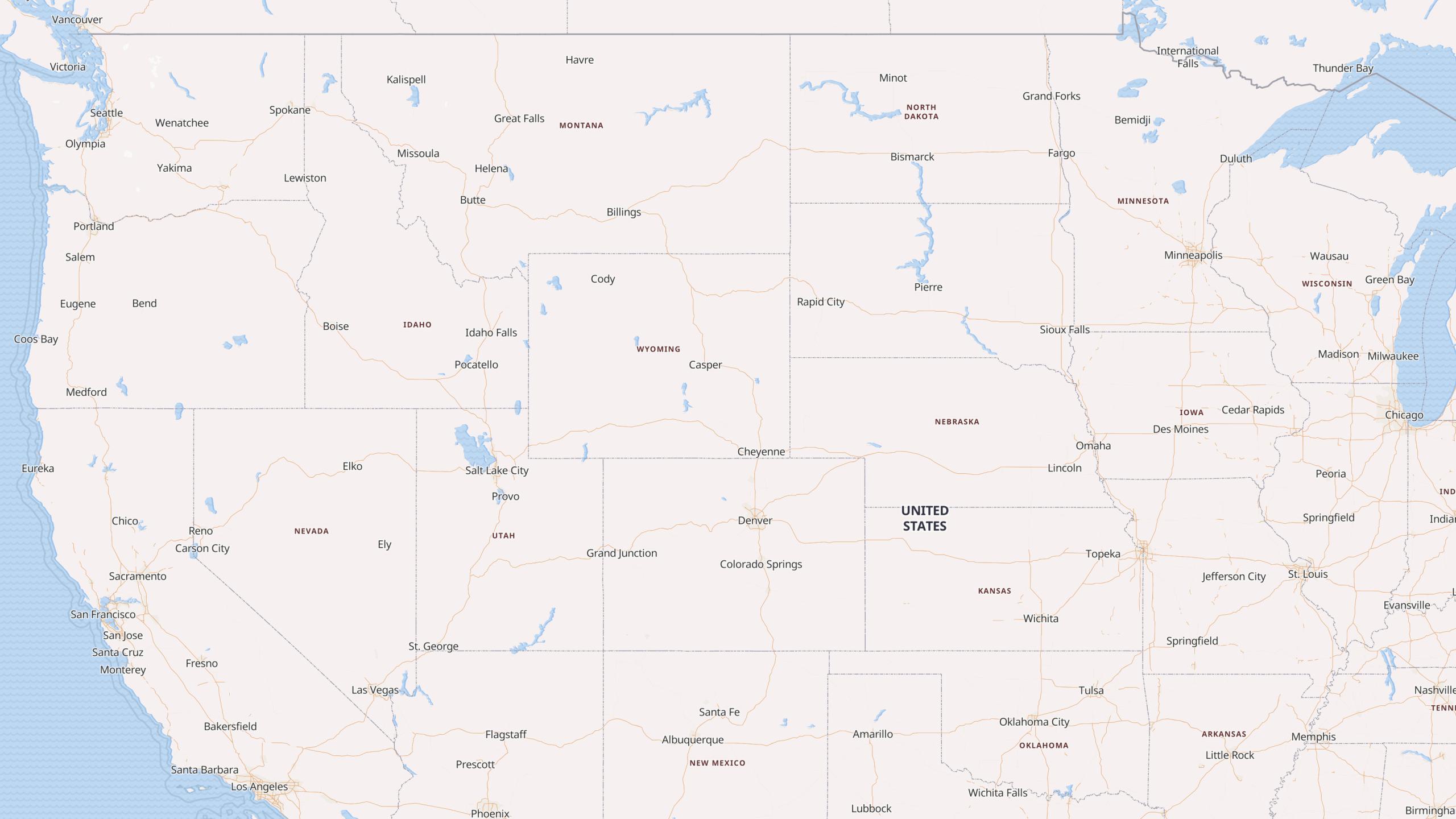


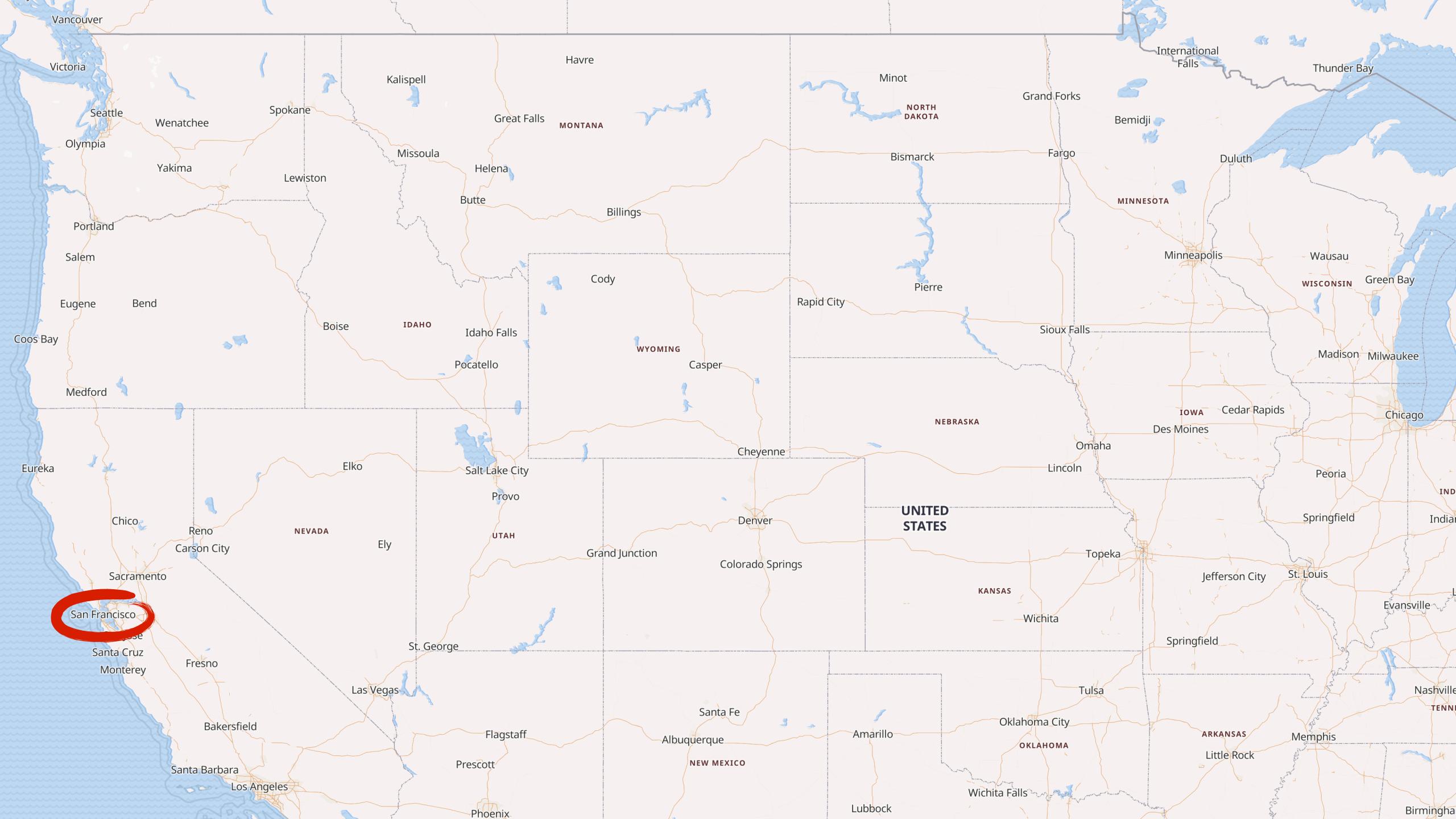
- transaction time
- transaction amount
- transaction type (in-person, online, contactless, etc.)
- merchant and user IDs
- other details

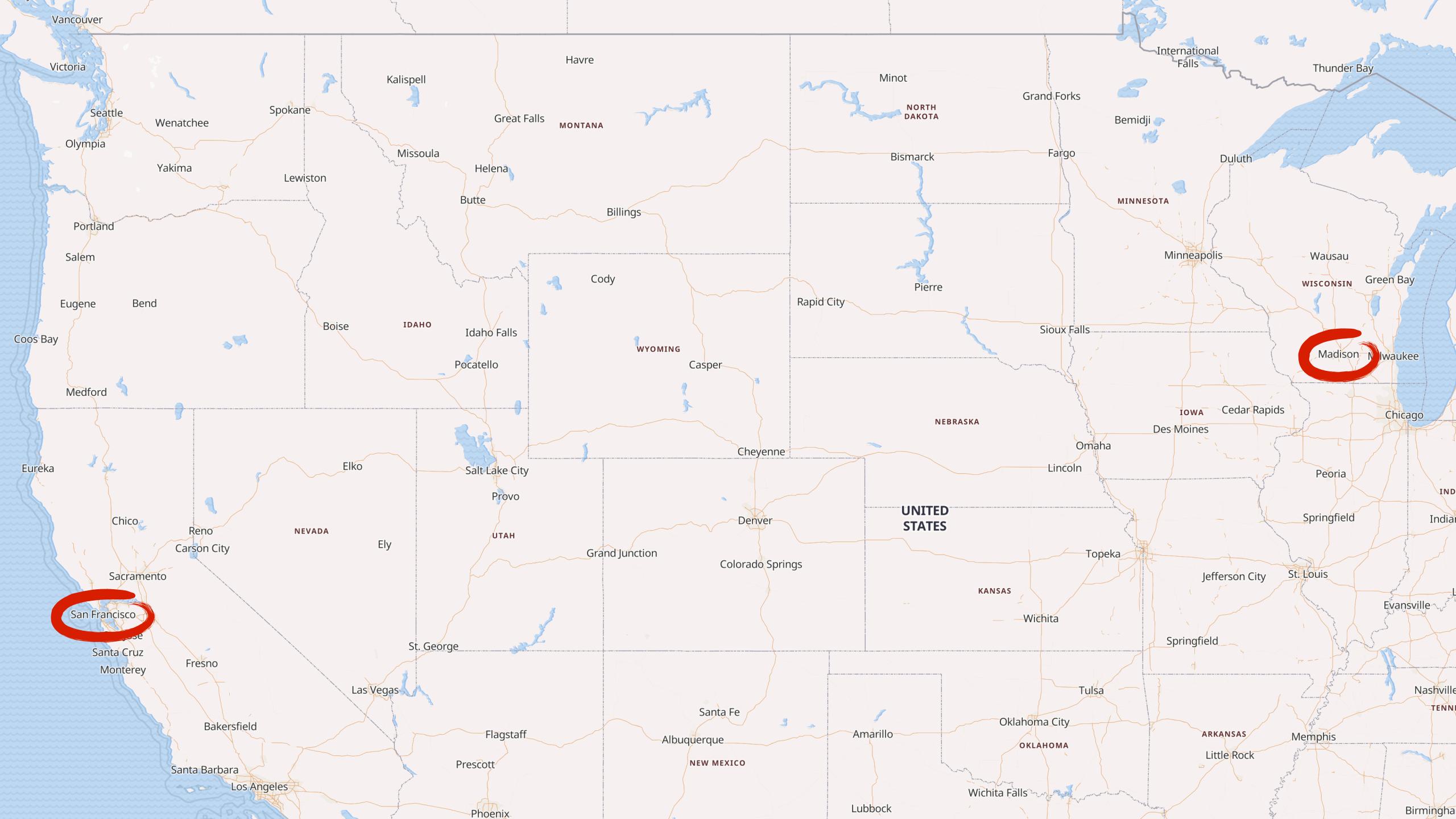






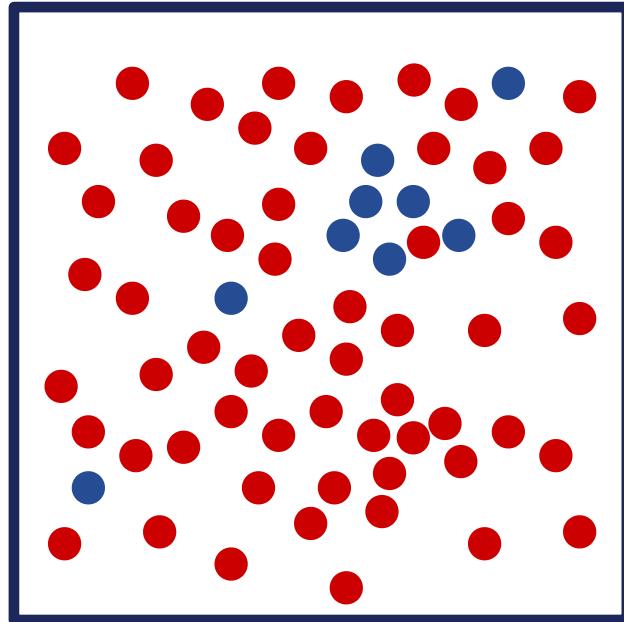






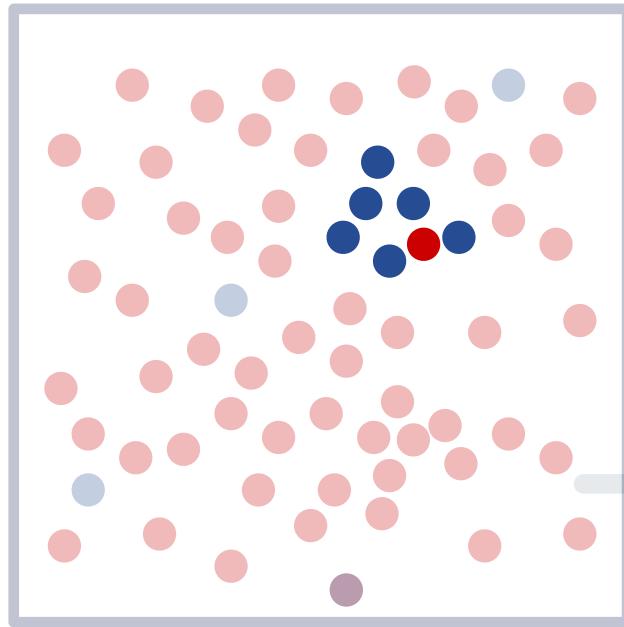


**feature  
engineering**



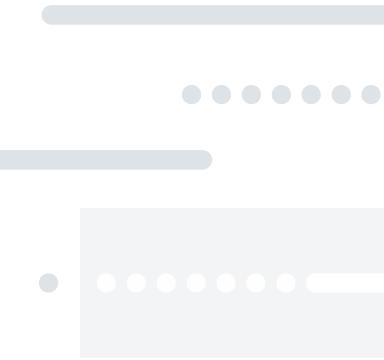
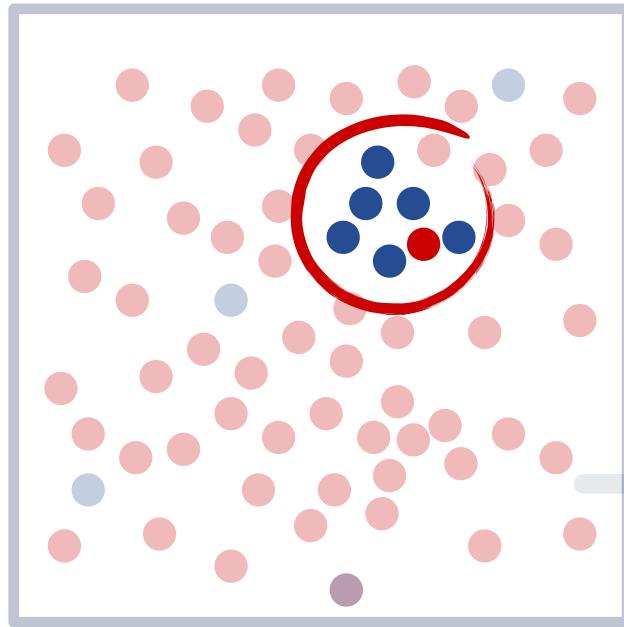


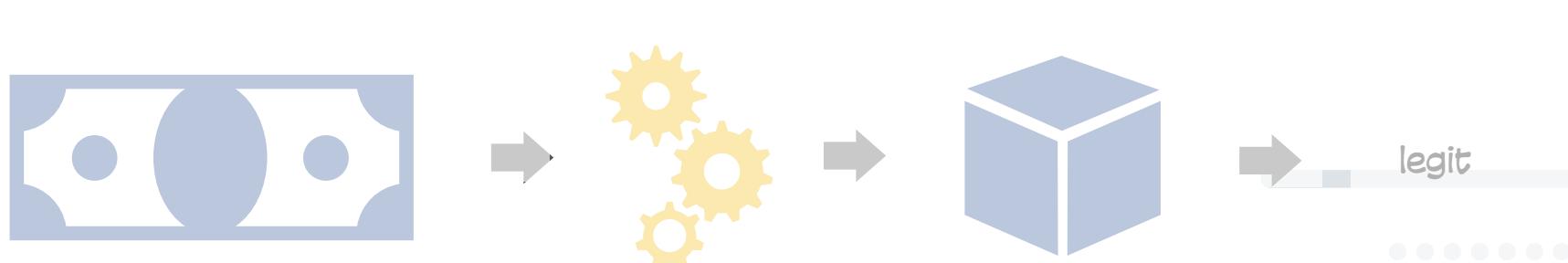
# feature engineering





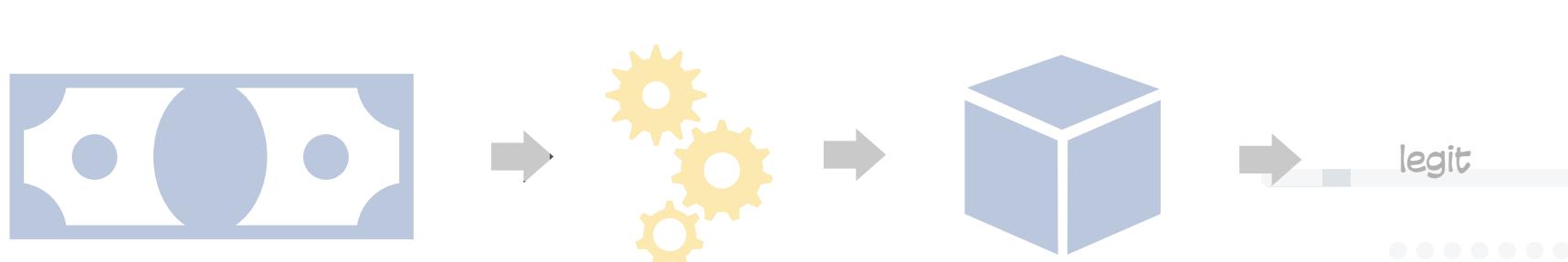
# feature engineering





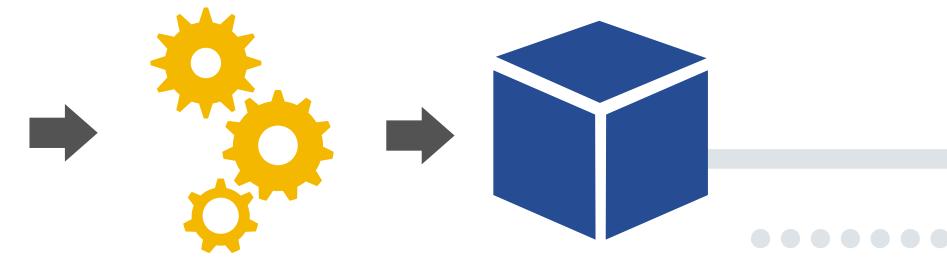
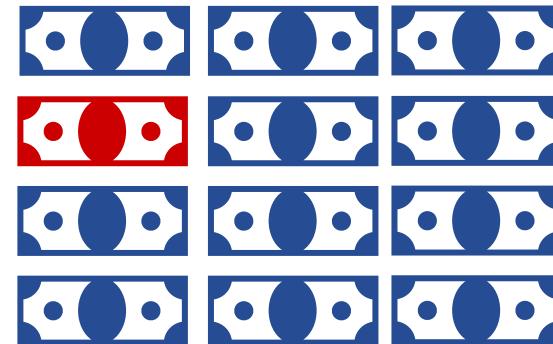


**model  
training  
and tuning**



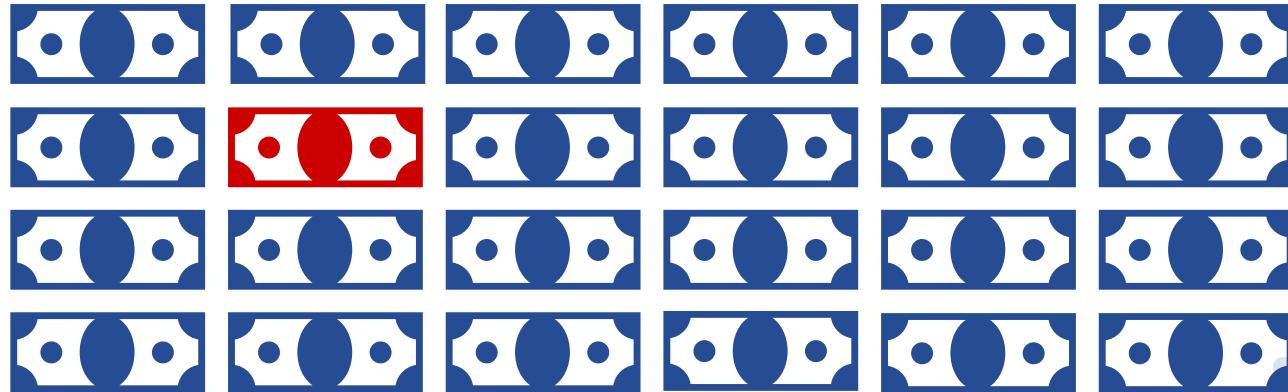


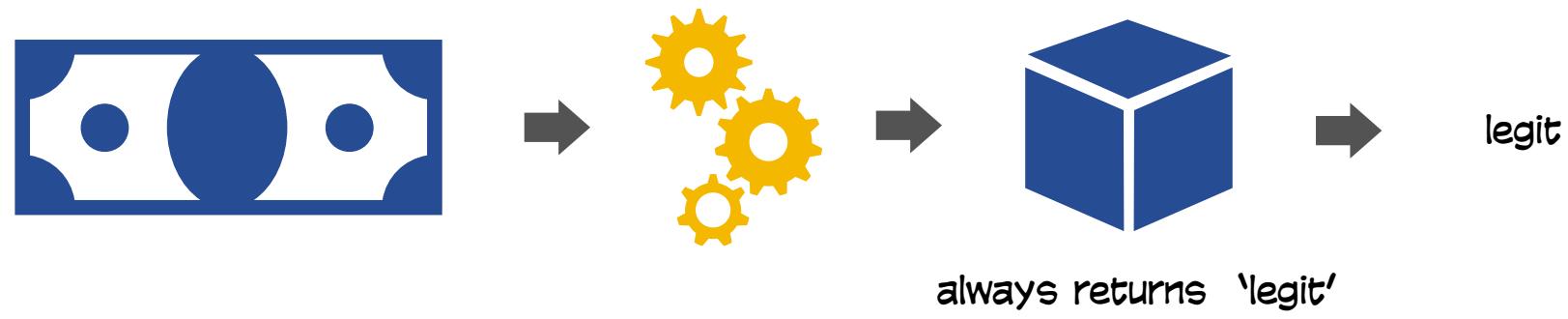
**model  
training  
and tuning**

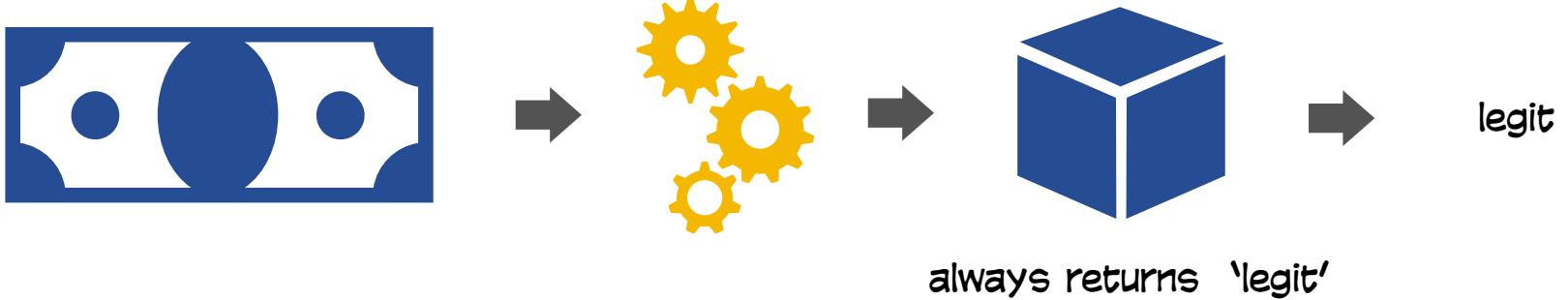




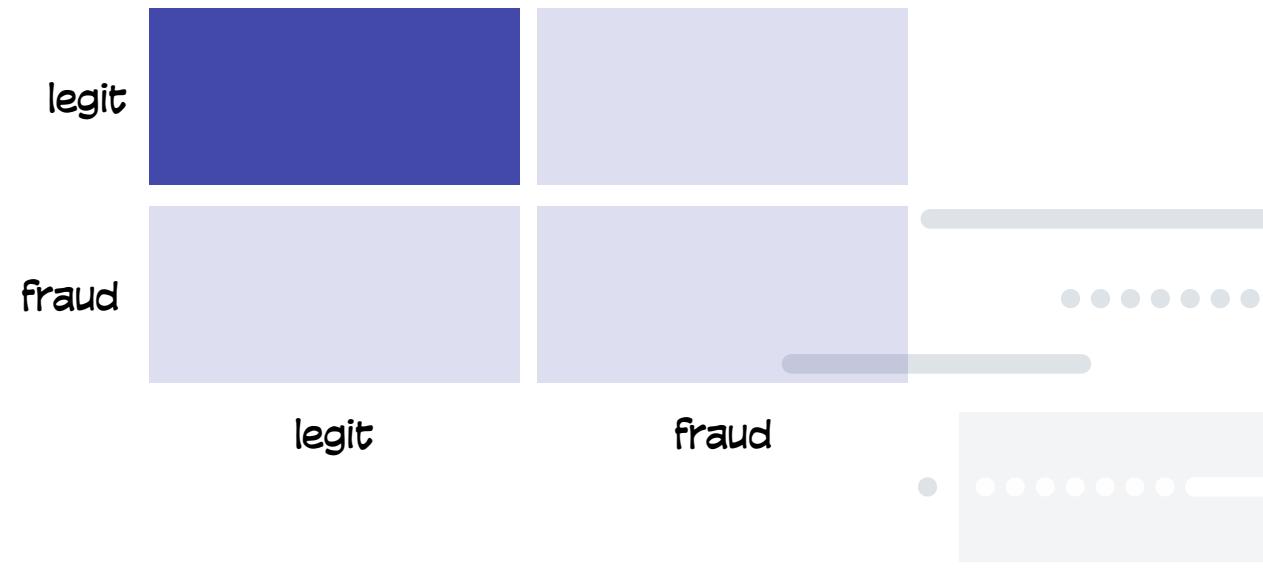
**model  
training  
and tuning**

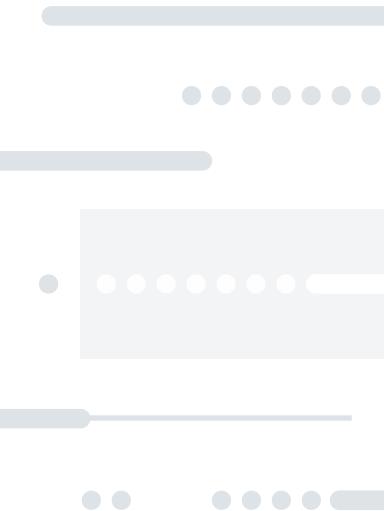
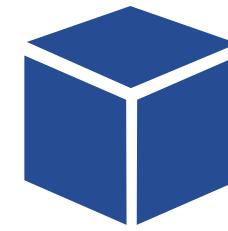
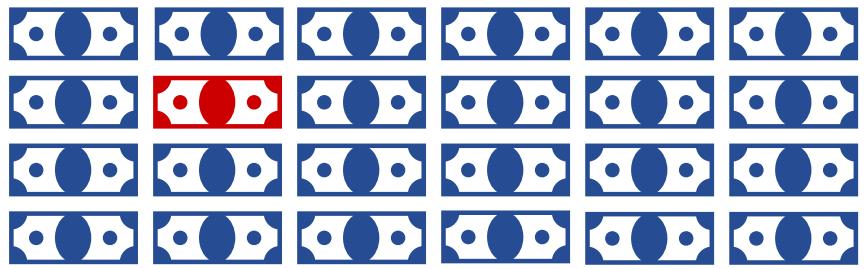


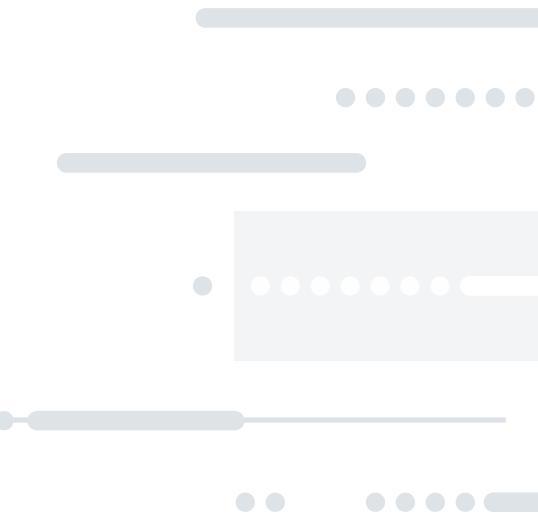
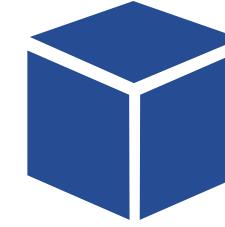
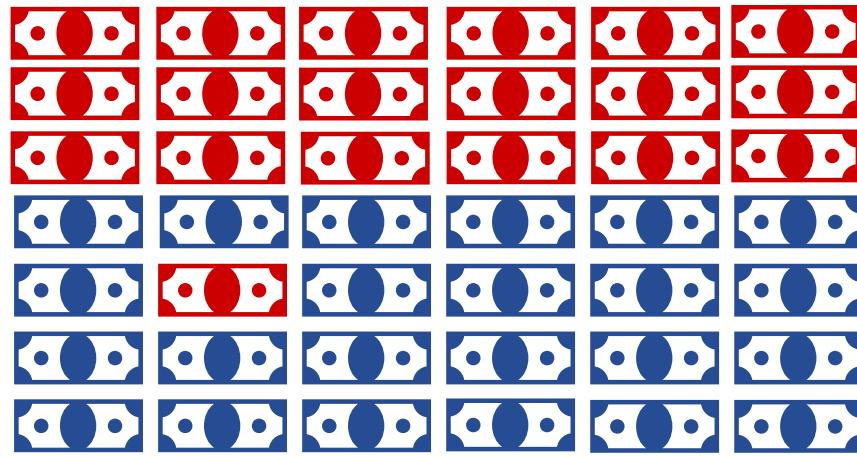




98% accuracy

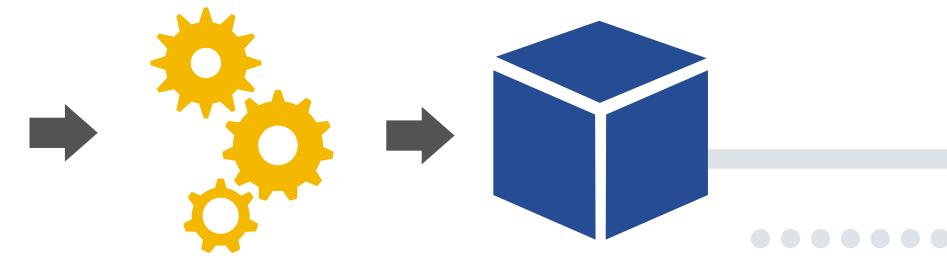
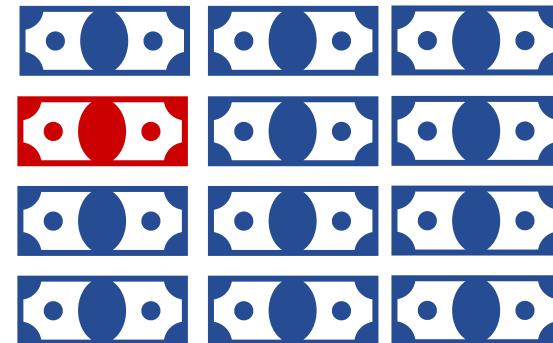


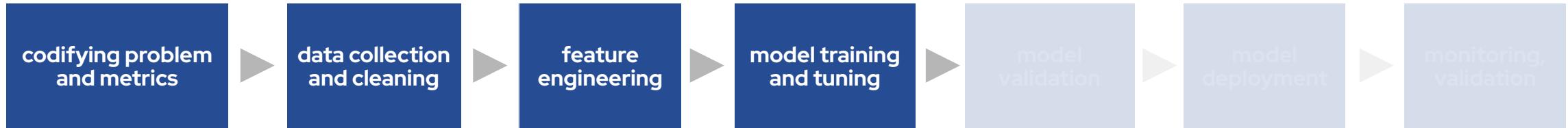




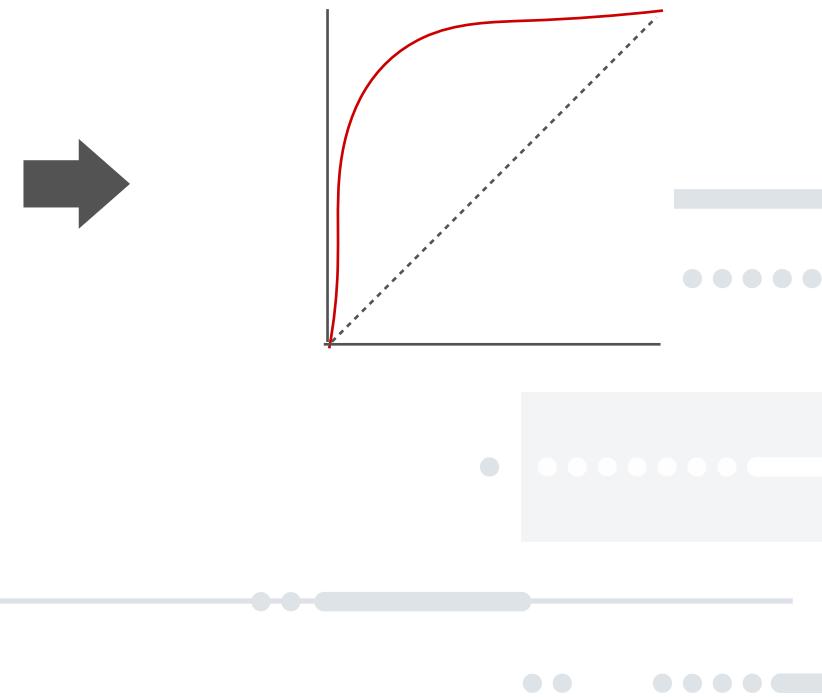
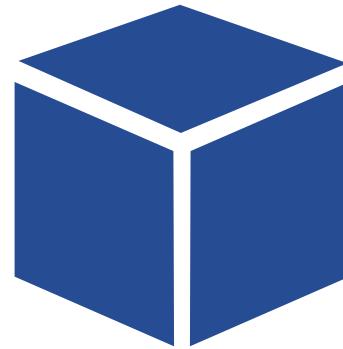


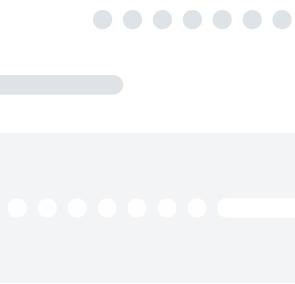
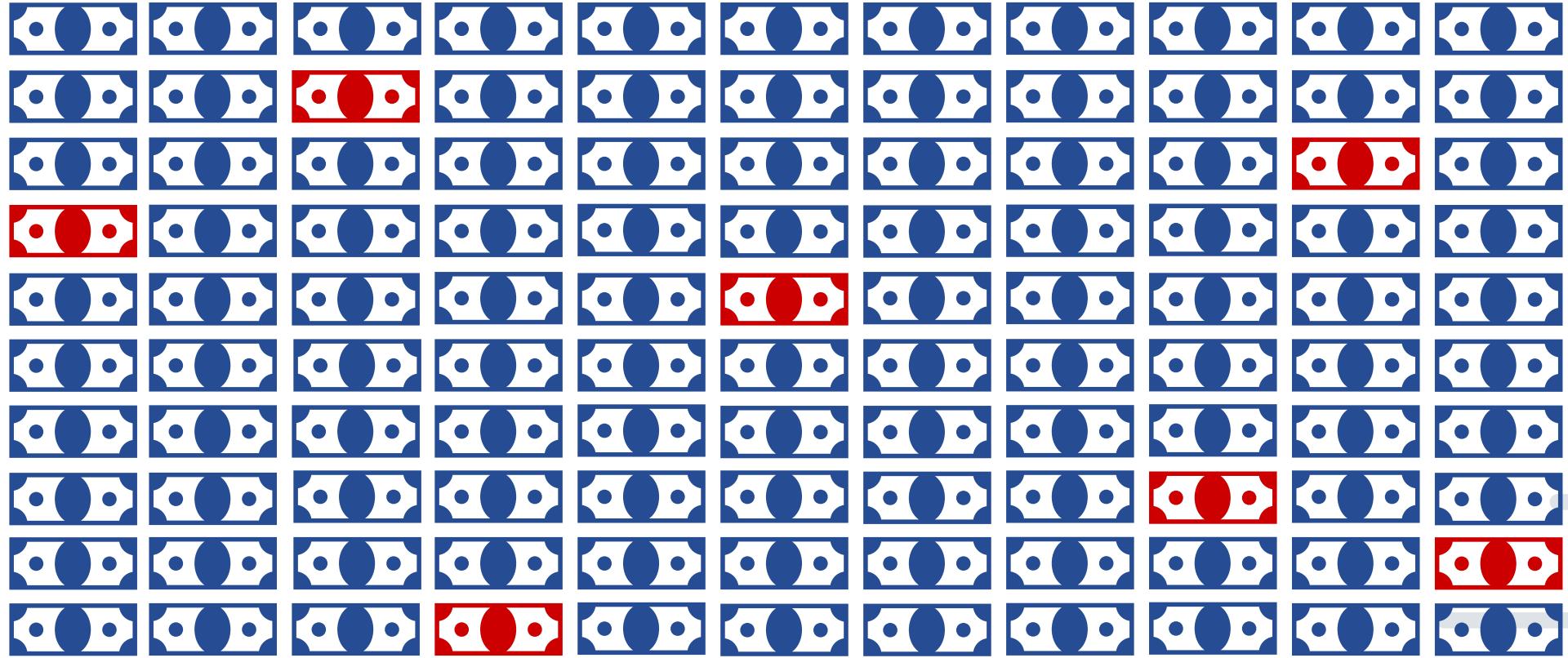
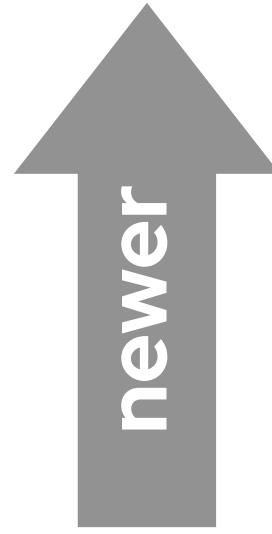
**model  
training  
and tuning**

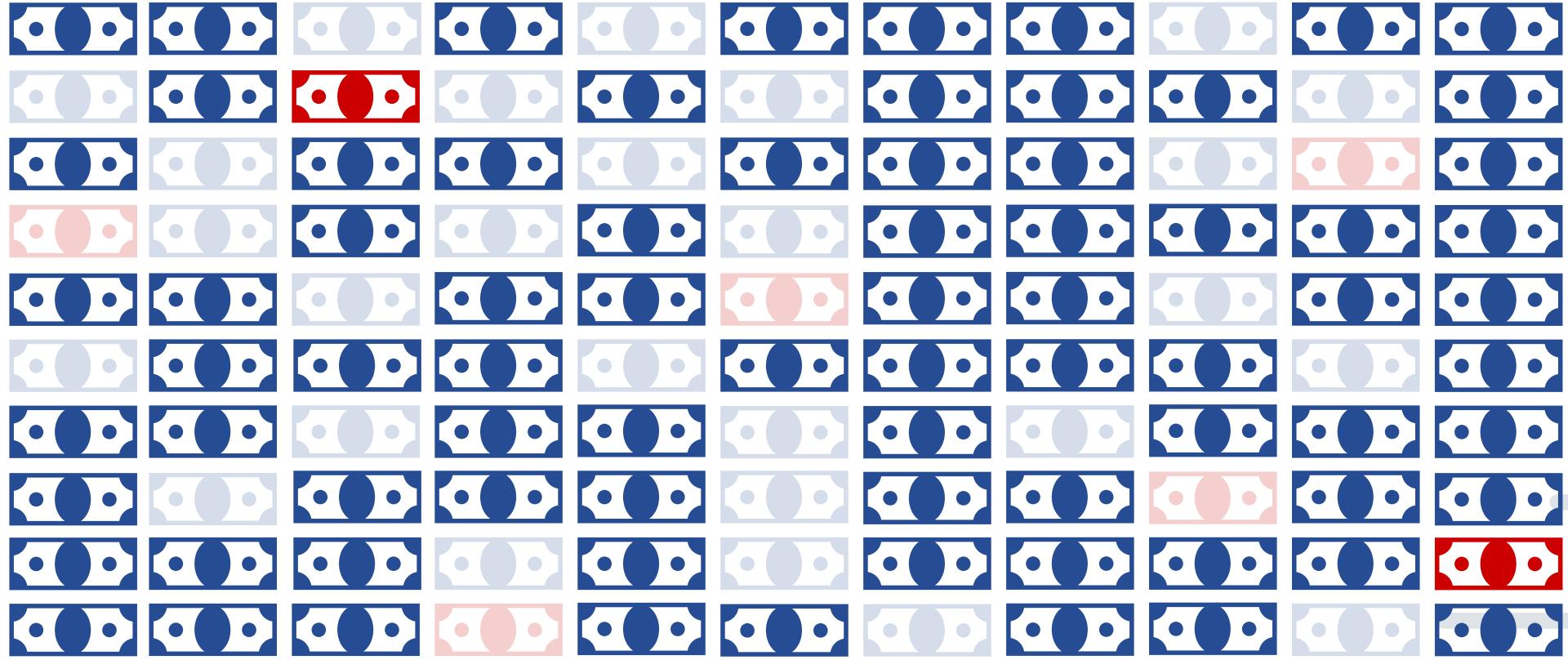
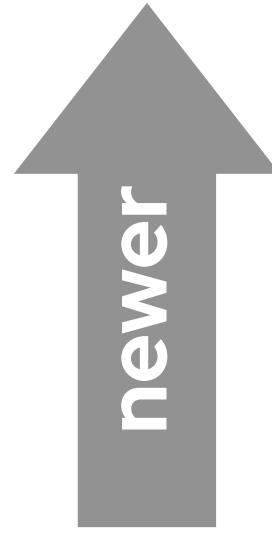


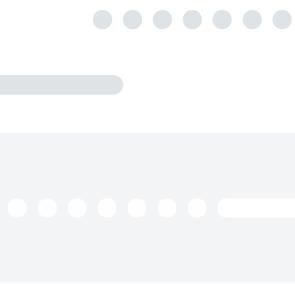
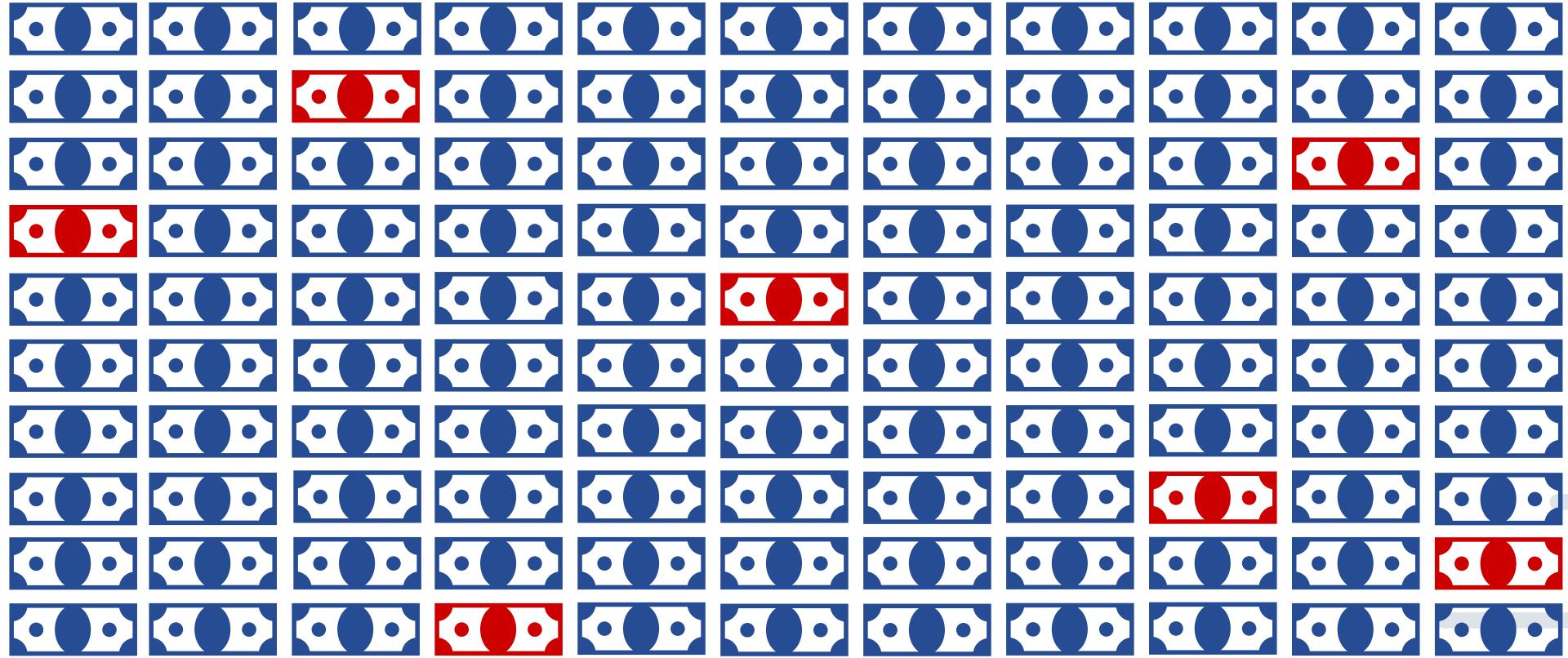
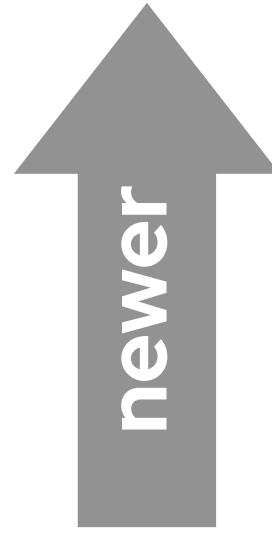


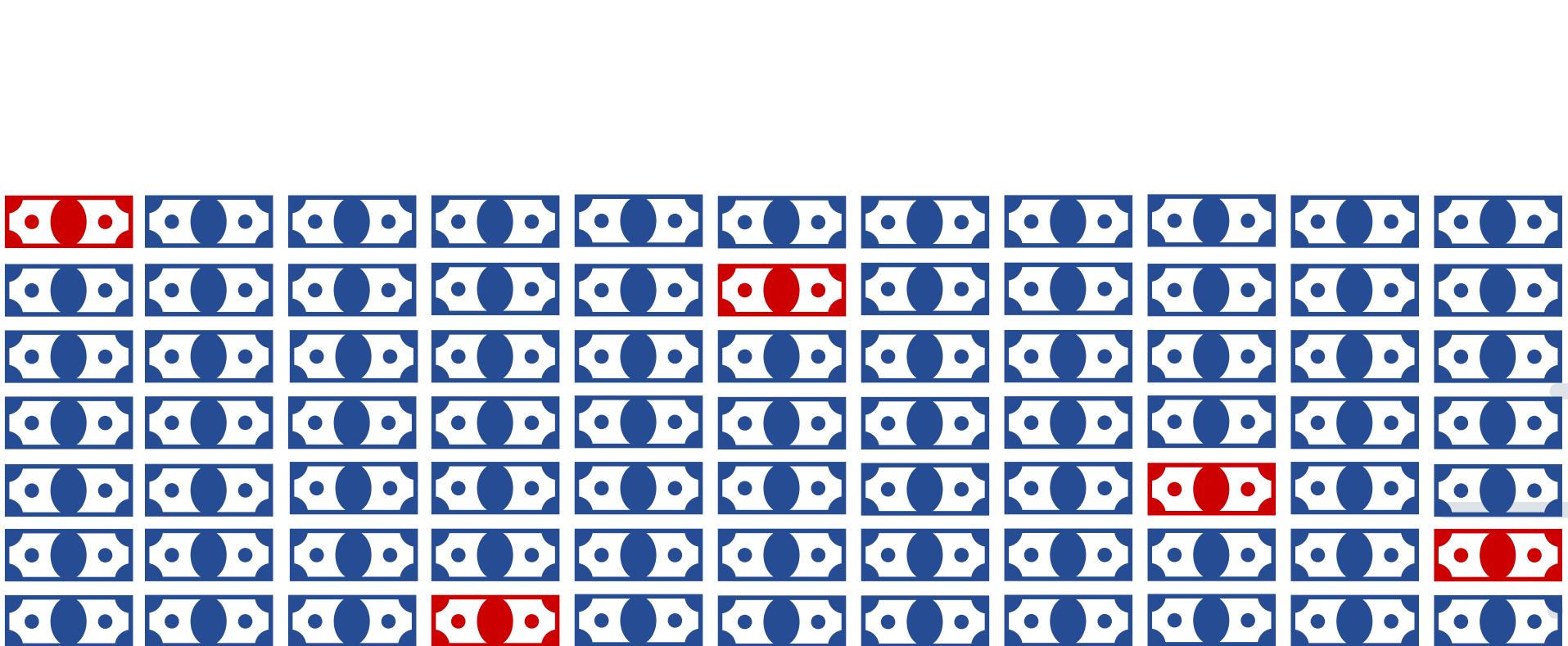
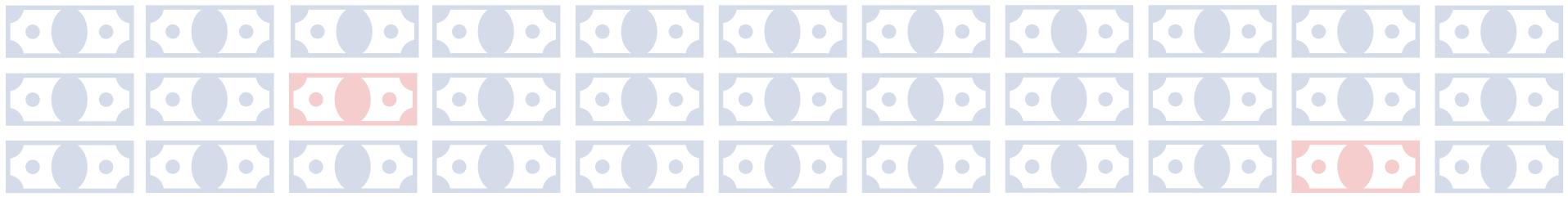
**model  
validation**

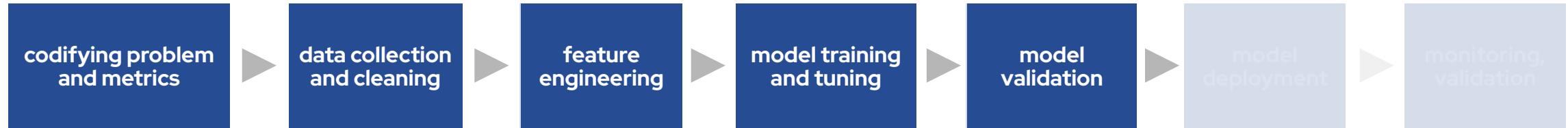






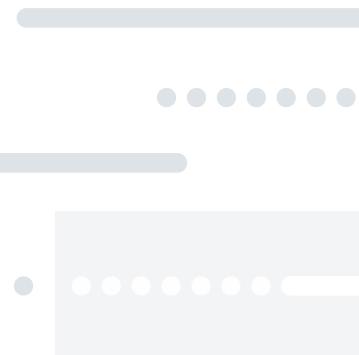
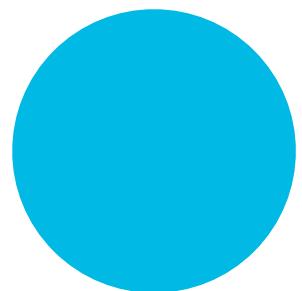
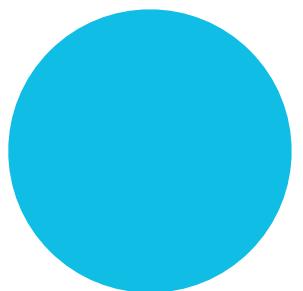


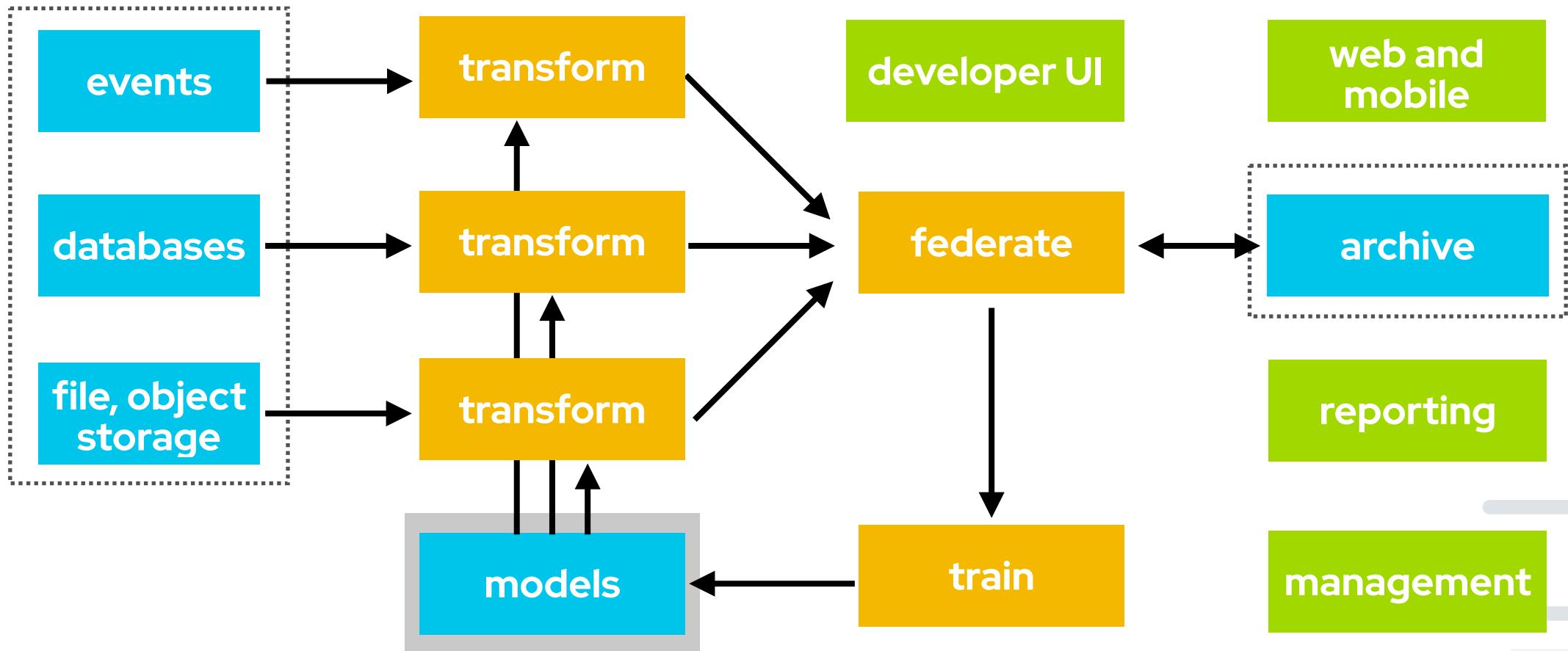




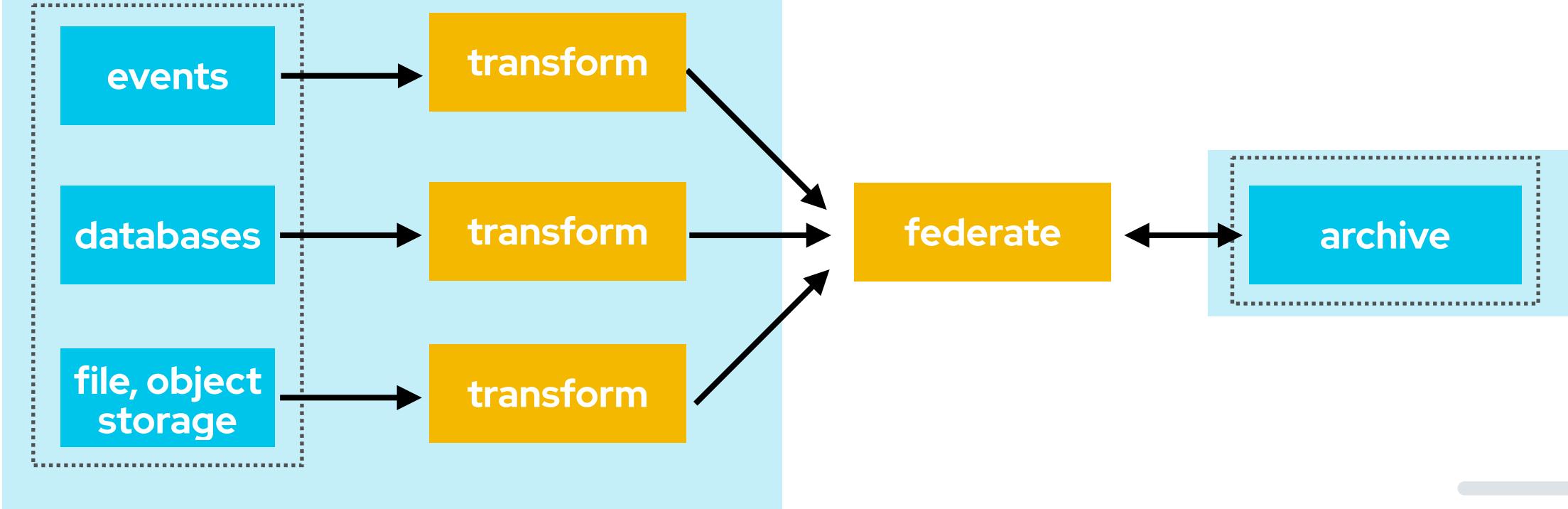
**model  
deployment**

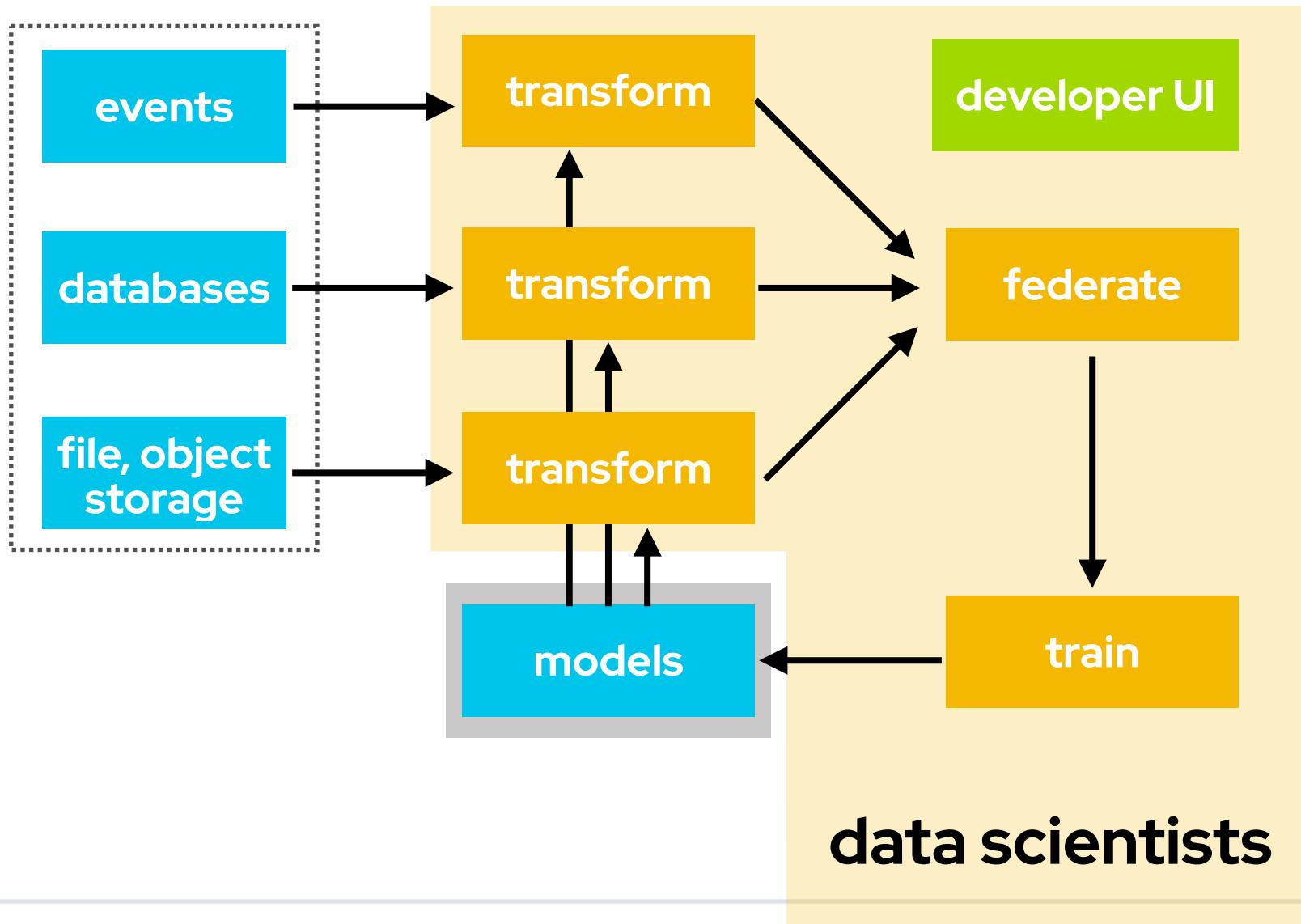
**monitoring,  
validation**



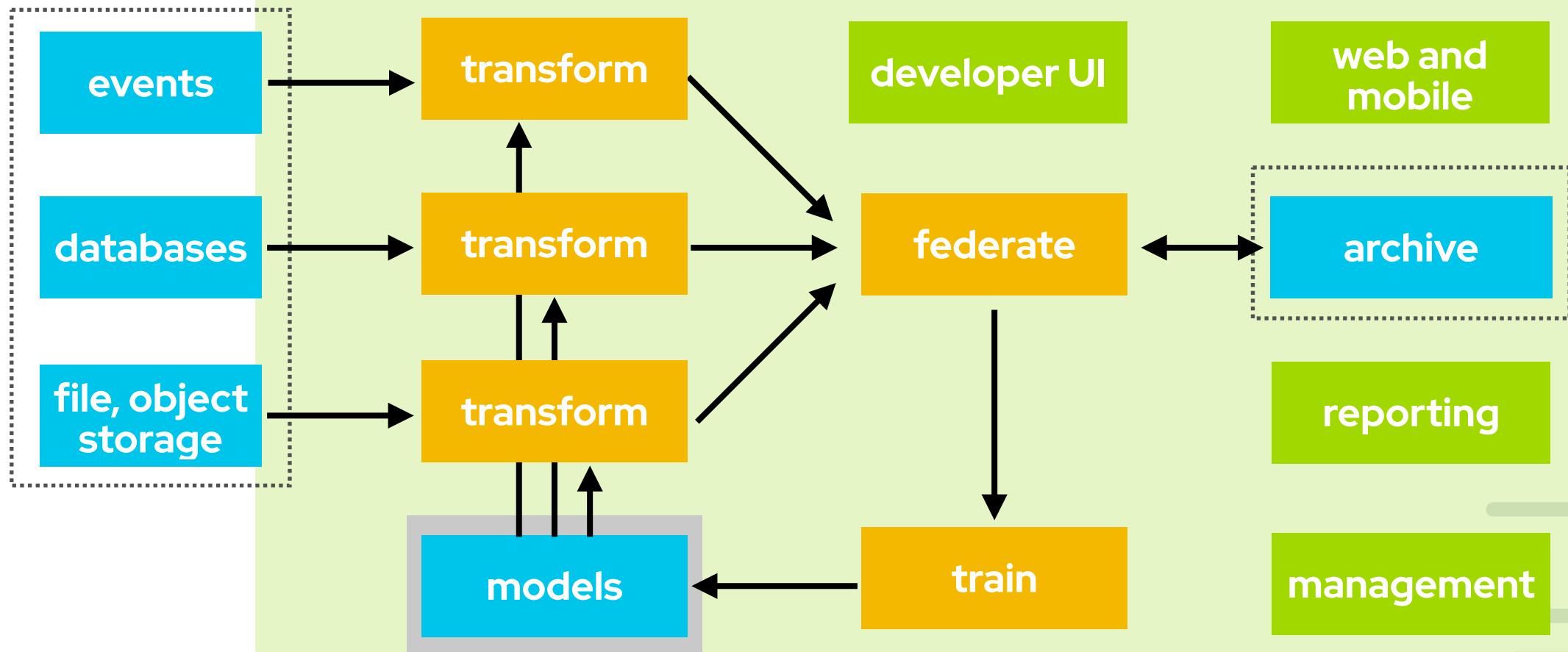


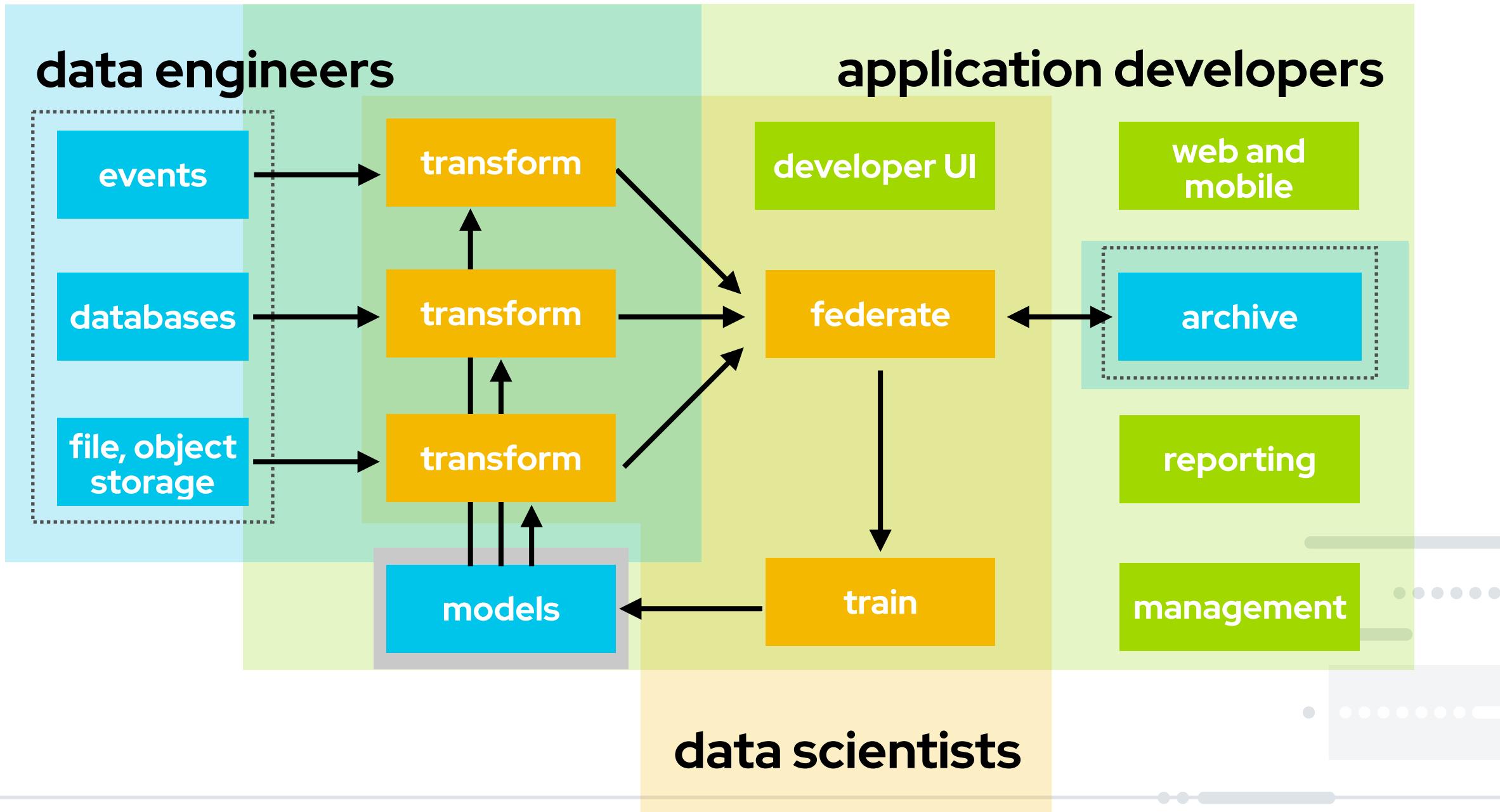
# data engineers



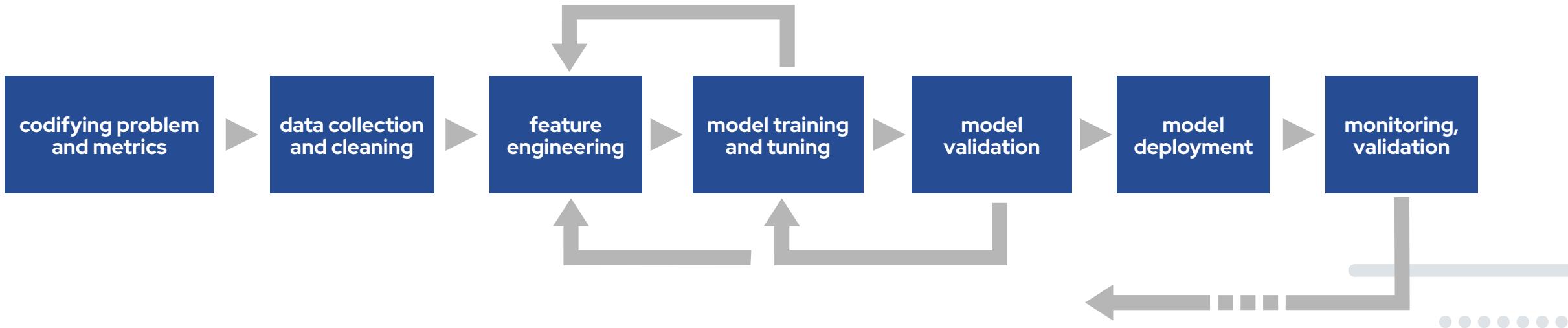


# application developers

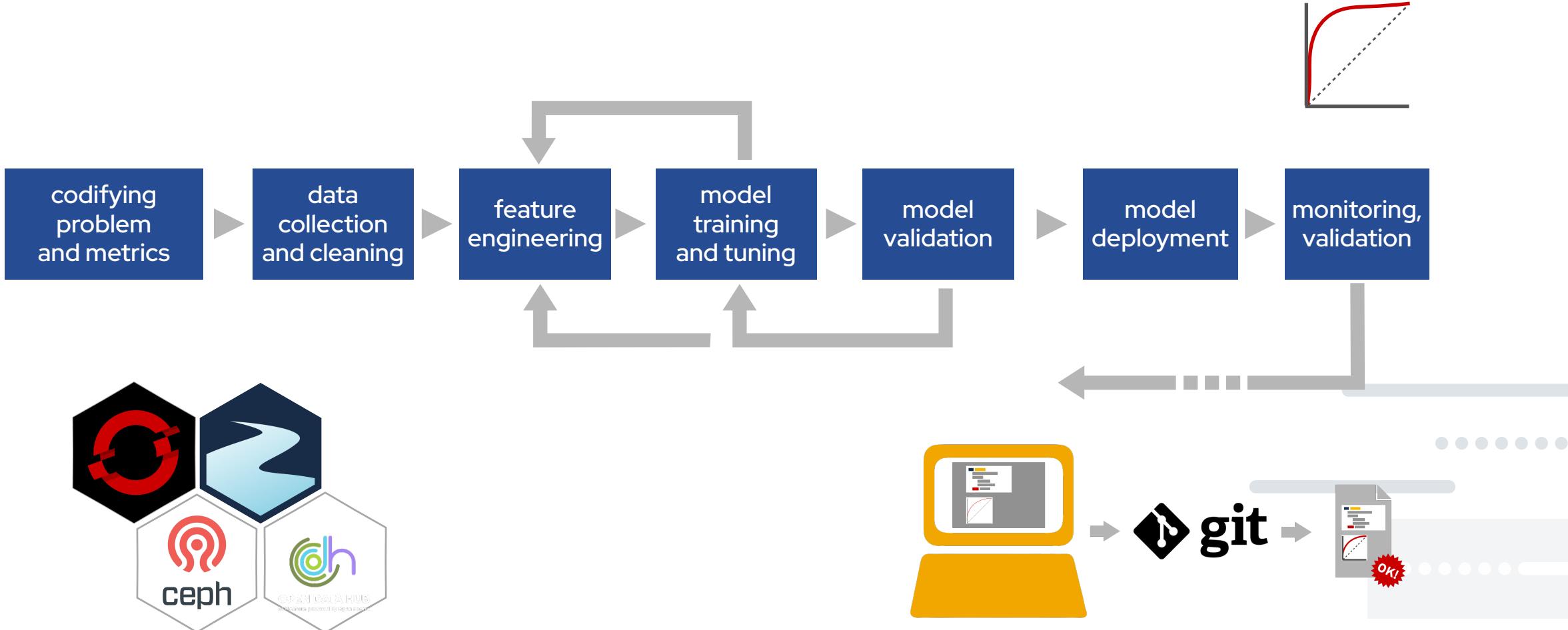


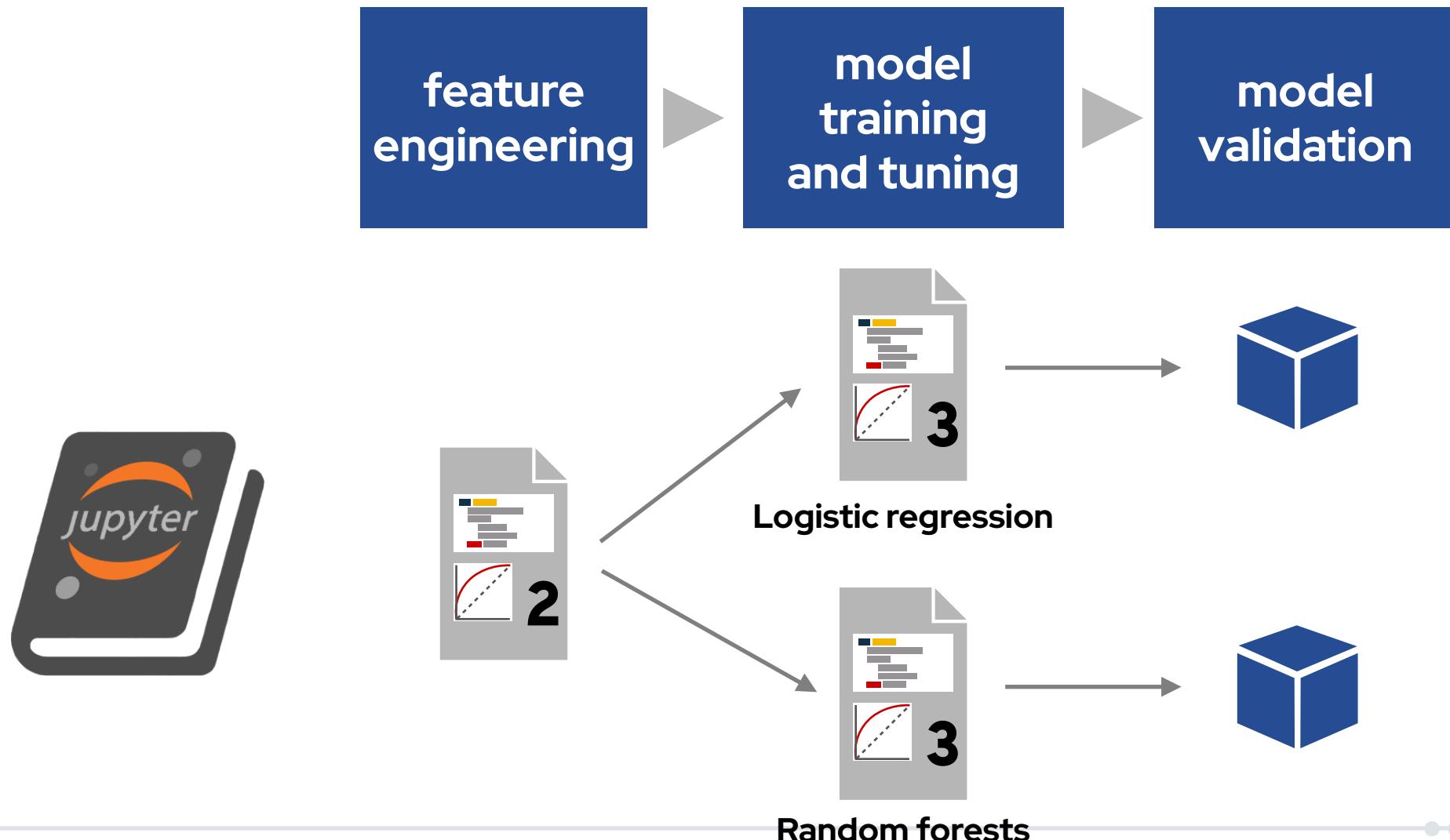




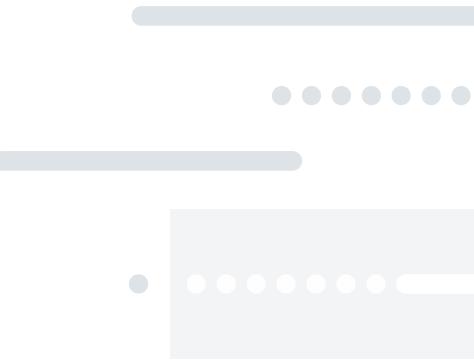
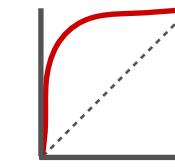
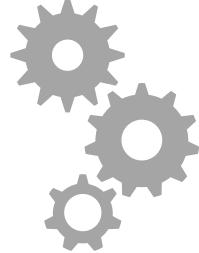
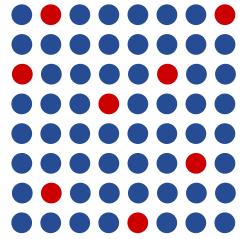


# jupyter

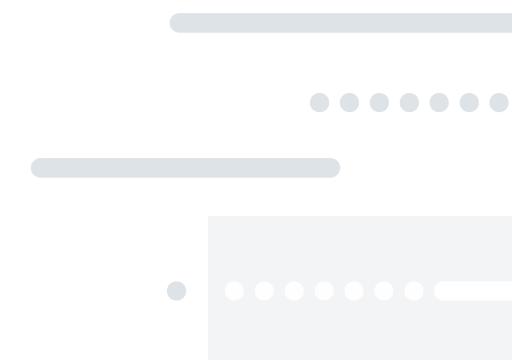
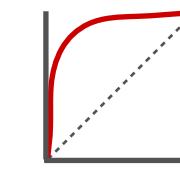
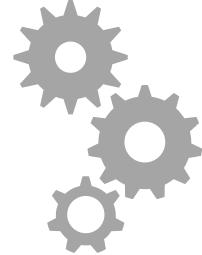
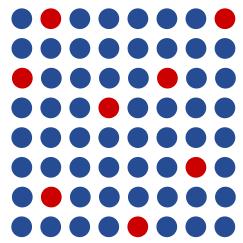
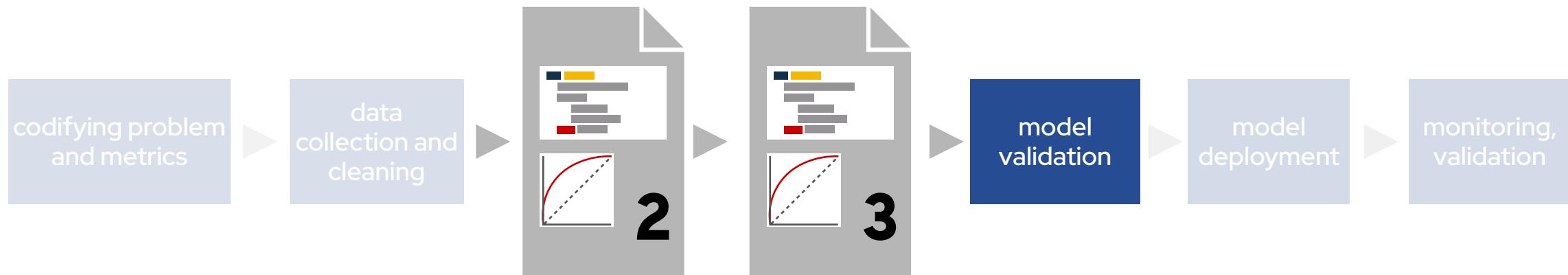


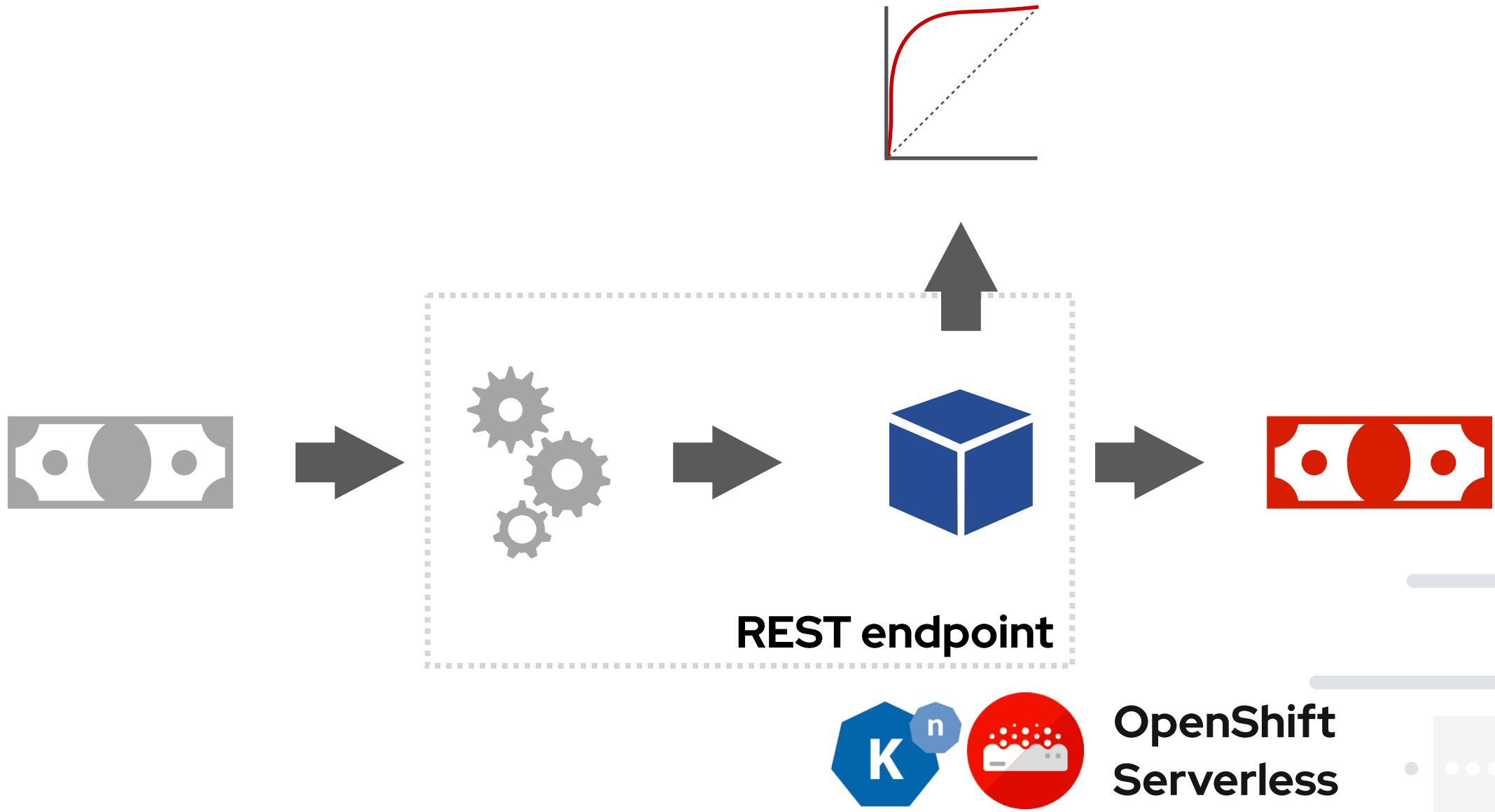


# OpenShift Pipelines



# OpenShift Pipelines





# Some useful links

- <https://www.openshift.com/learn/topics/ai-ml>
- <https://opendatahub.io/>
- <https://tekton.dev>