



Introduction to Data

Ethical, Social, and Legal Issues

Dr Christopher Burr
@chrisdburr
18th September 2021

**The
Alan Turing
Institute**

Presentation

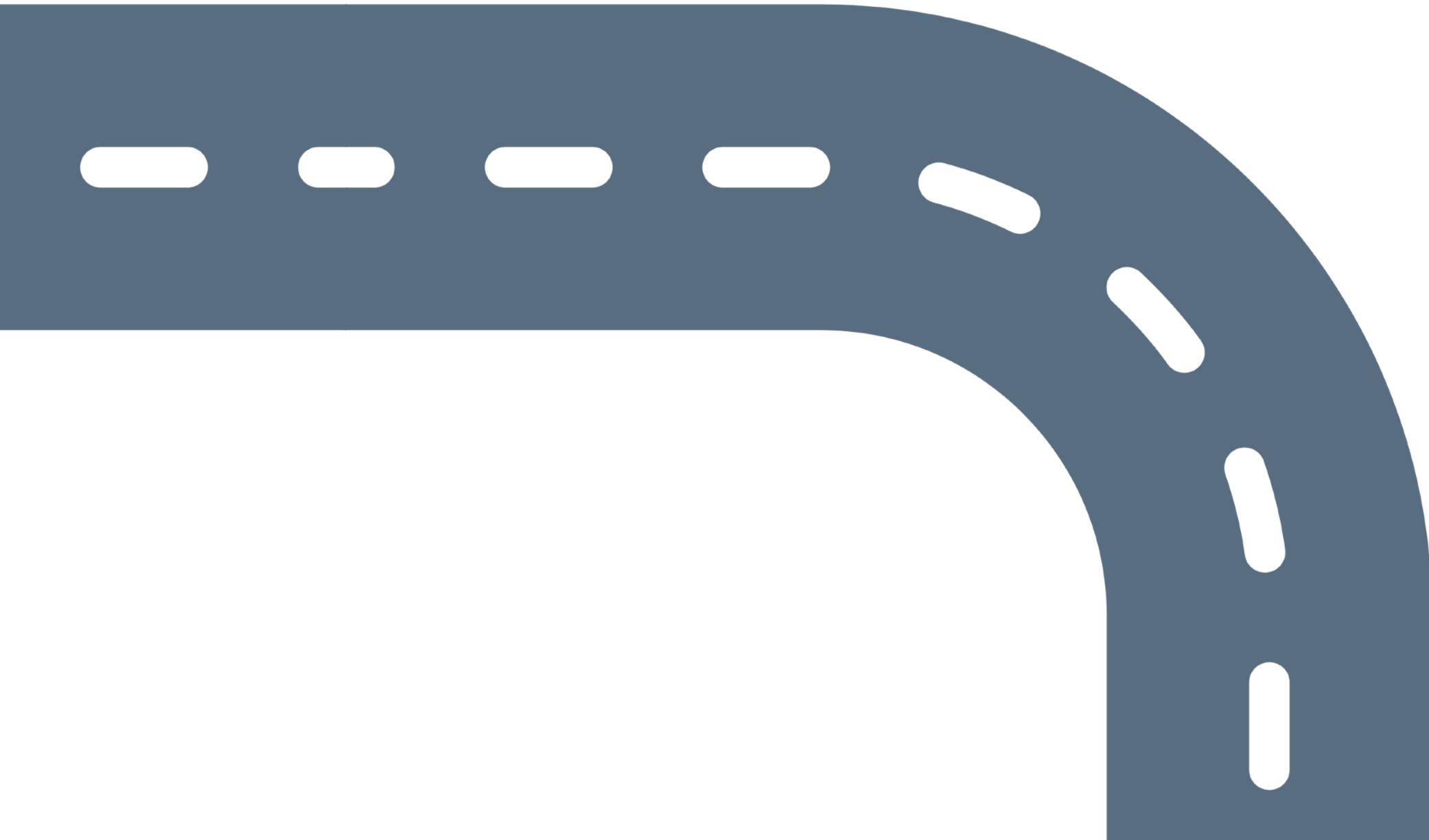
Outline

1. Data Journey (Part 1)
2. Activity: Exploring Data
3. Data Journey (Part 2)
4. Activity: Topics for Reflection



Part 1

Data Journey



Data Journey

Project Planning

A local police force want to use data-driven technology to reduce crime



Data Journey

Project Planning

They plan to develop a facial recognition system, which can automatically identify wanted suspects or known criminals in public spaces



Data Journey

Data Collection & Processing



The police force have to collect two types of data for their system to work

Data Journey

Data Collection & Processing

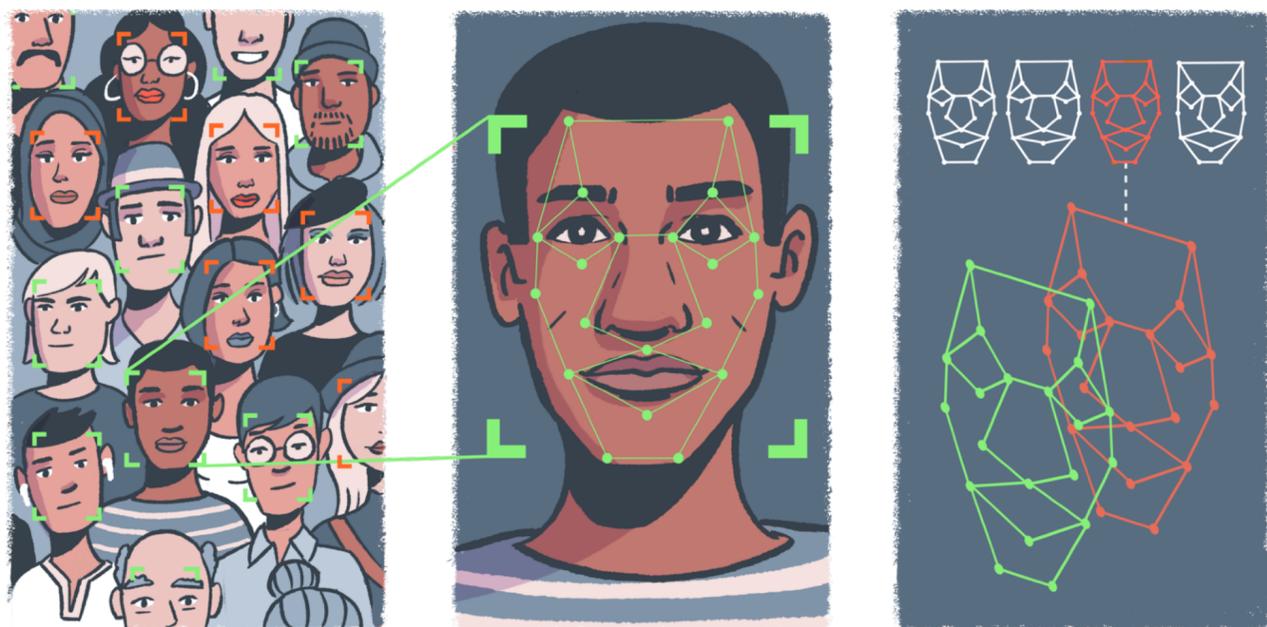


First, they need to collect images of the wanted suspects and known criminals.

The police already have access to this data.

Data Journey

Data Collection & Processing



Second, they need to collect images of members of the public that will be matched against their original dataset.

They need to collect this data

Data Journey

Data Exploration & Analysis

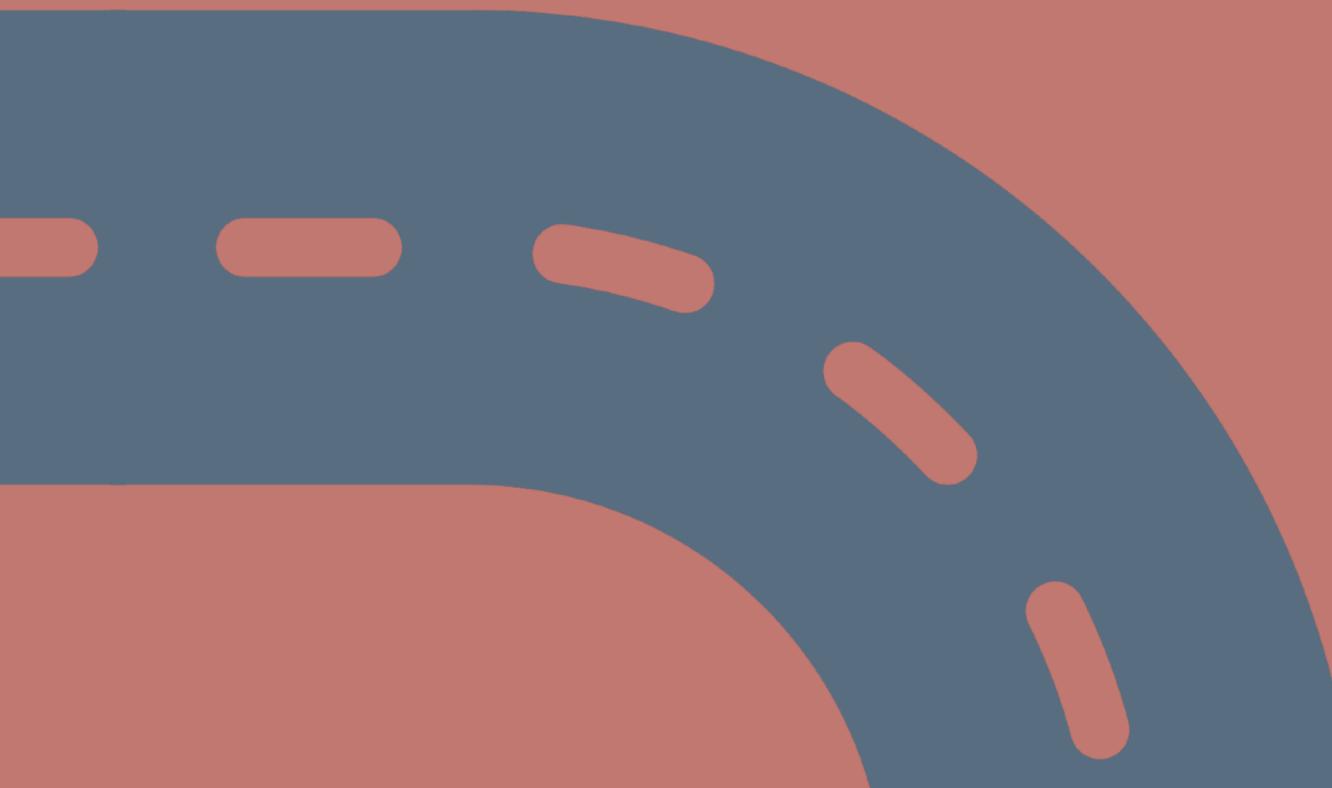
Once our police force have gathered all the data they can then explore whether it is good enough for their automated system to work properly



Data Journey

Data Exploration & Analysis

This is known as 'data analysis', and is a common process of data science

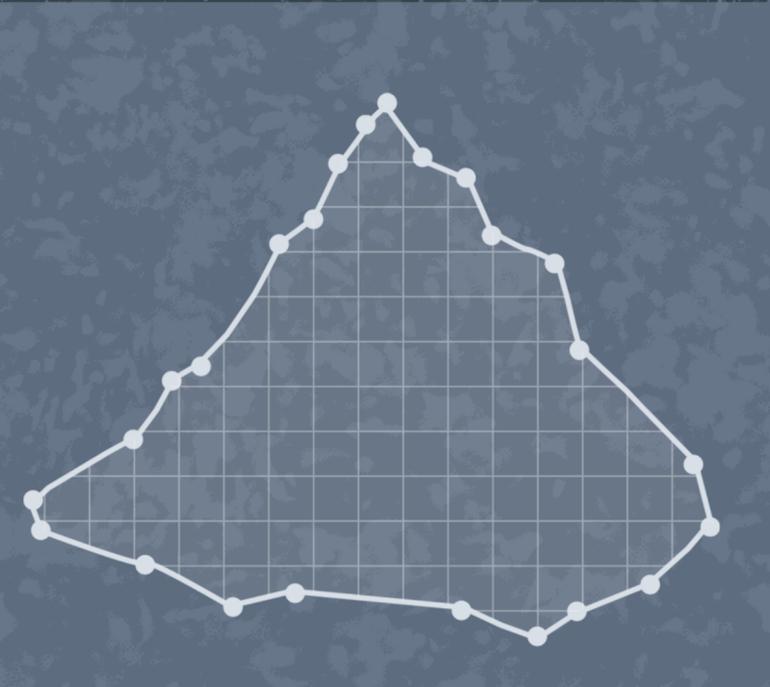
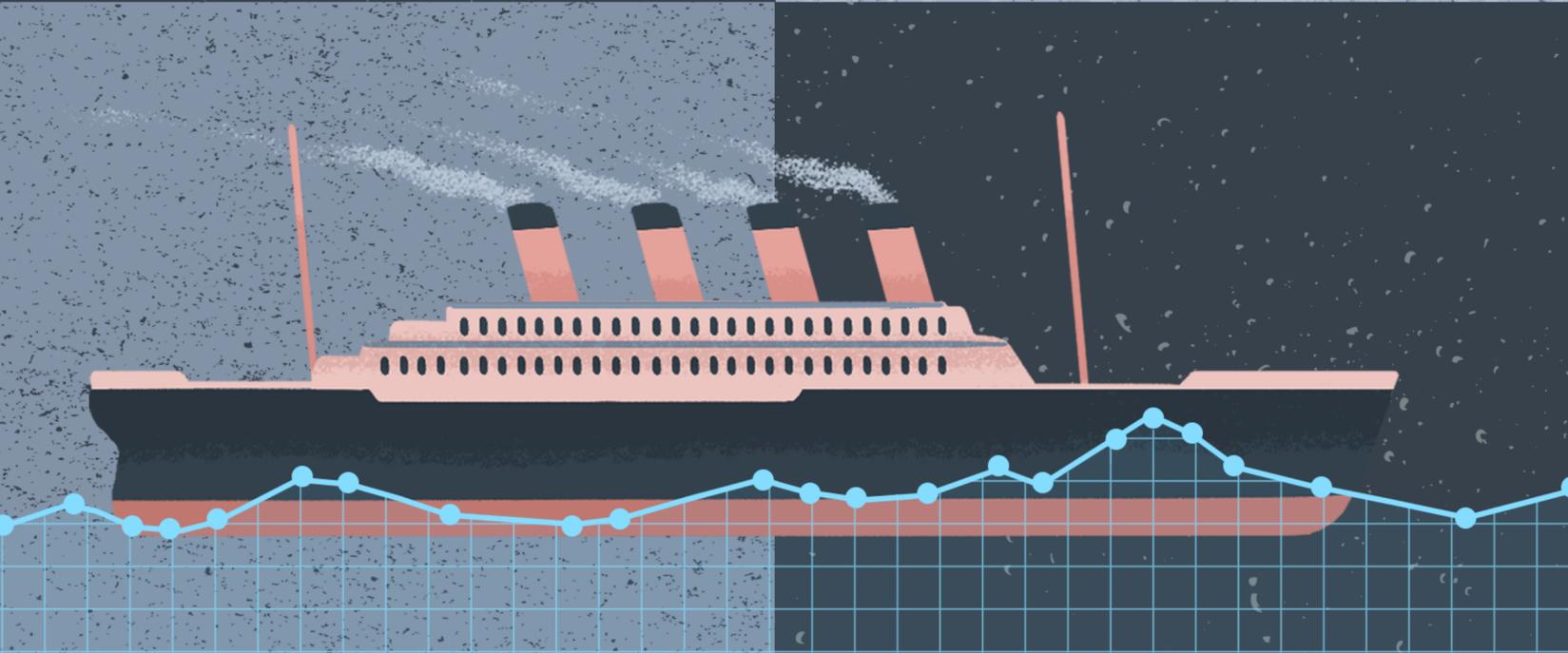
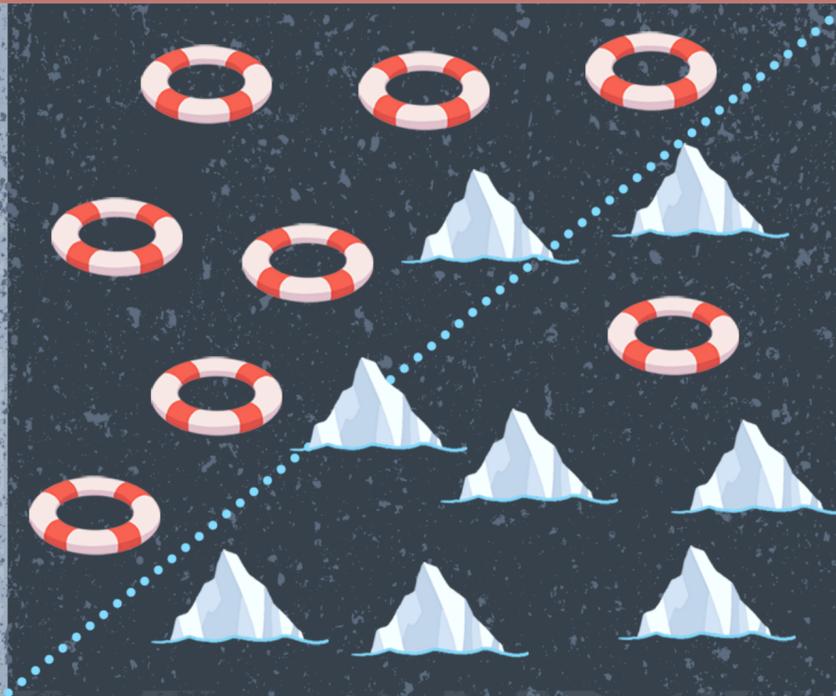


Activity 1

Exploring Data



3RD CLASS
TICKET
MRS J. B CUMMINGS
38 YRS



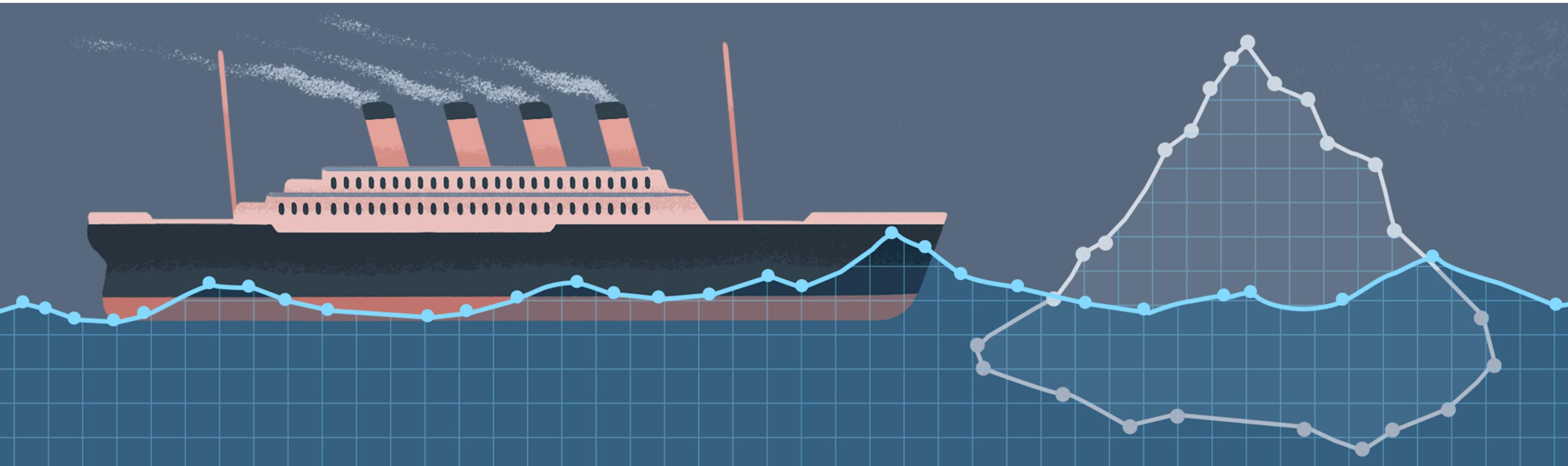
1ST CLASS
TICKET
MR T. J MCCARTHY
54 YRS

Activity 1

Exploring Data

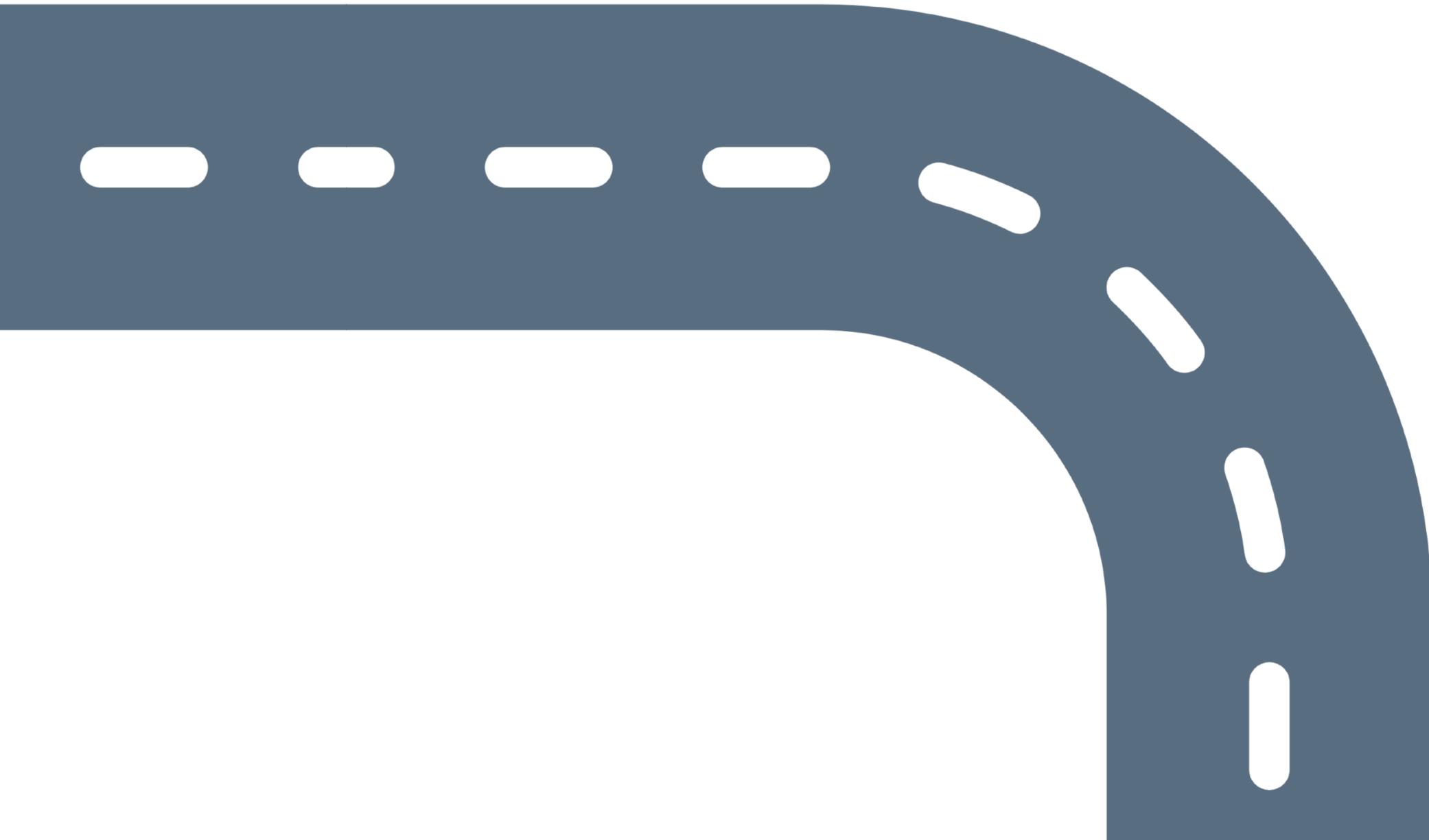
Do you think this person survived?

Name	Sex	Age	Class	Survived
Miss. Constance Mirium West	Female	5	2	?



Part 2

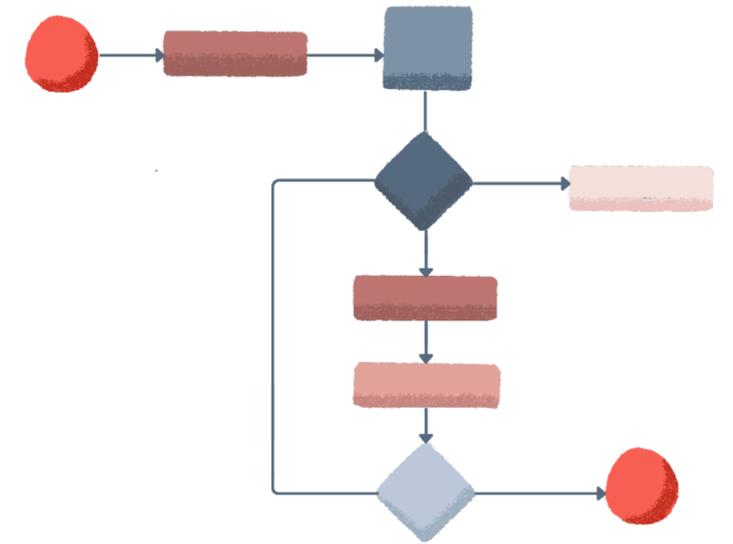
Data Journey



Data Journey

Data Use

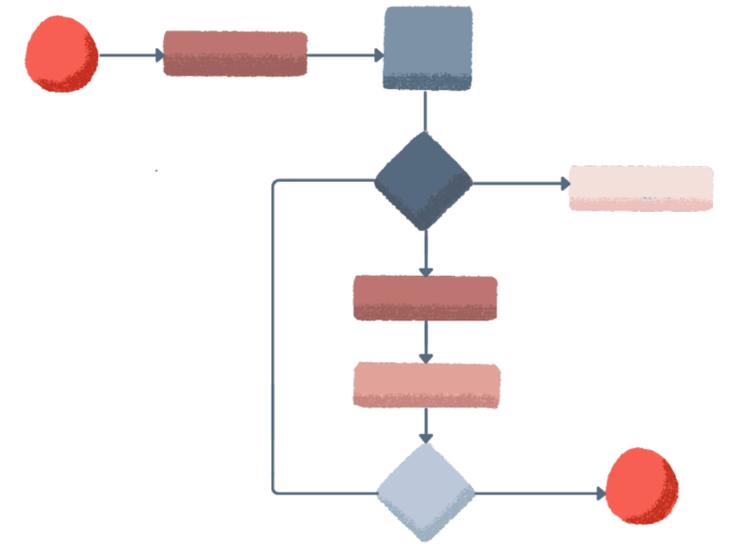
To develop a facial recognition system that matches images automatically the police use a technique known as 'machine learning'



Data Journey

Data Use

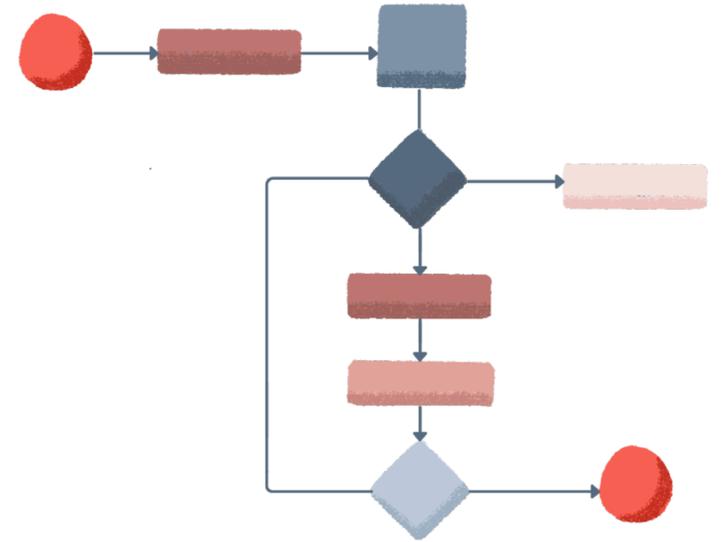
A computer model is trained to (a) identify individual faces in a crowd, taken from CCTV images, and then (b) try to match these faces with ones in an existing dataset



Data Journey

Data Use

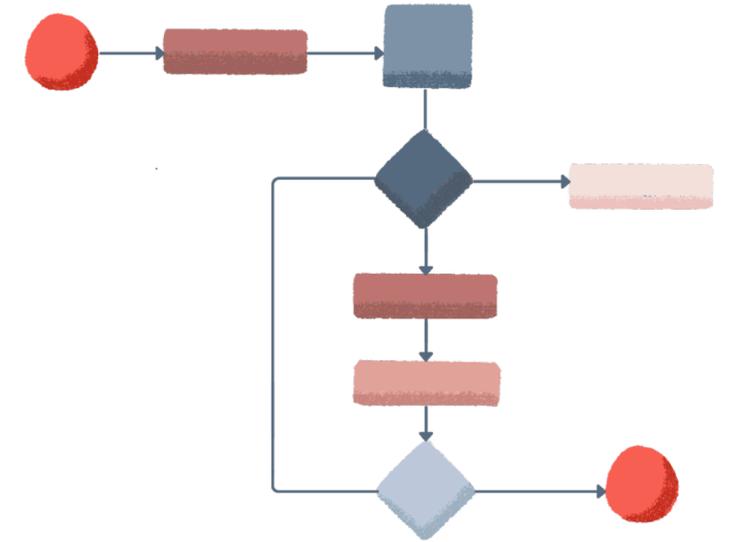
It learns by being told when it gets a match right or wrong, and uses this feedback to improve at the task



Data Journey

Data Use

The end result is an automated (or, algorithmic) system that can match new images that it has not previously seen



Data Journey

Project Documentation

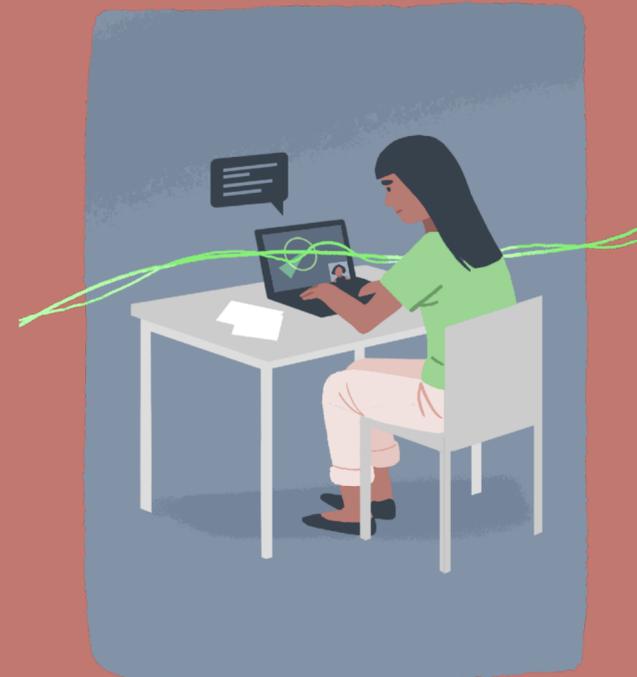
When the police test
their system in the real
world it does not
perform as expected



Data Journey

Project Documentation

The accuracy of their system is a lot worse than it was during the training phase



Data Journey

Project Documentation

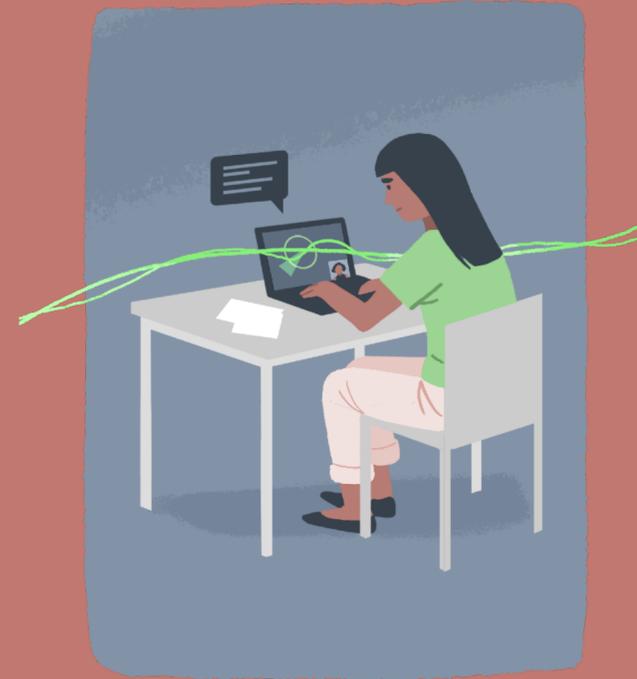
This is because the data
the police used to train
the model was not
sufficiently representative



Data Journey

Project Documentation

They return to the original documentation to identify the gaps, address the issue, and then update their documentation



Data Journey

Data Reuse

Once they have improved the accuracy of their system, the police decide to reuse the data



Data Journey

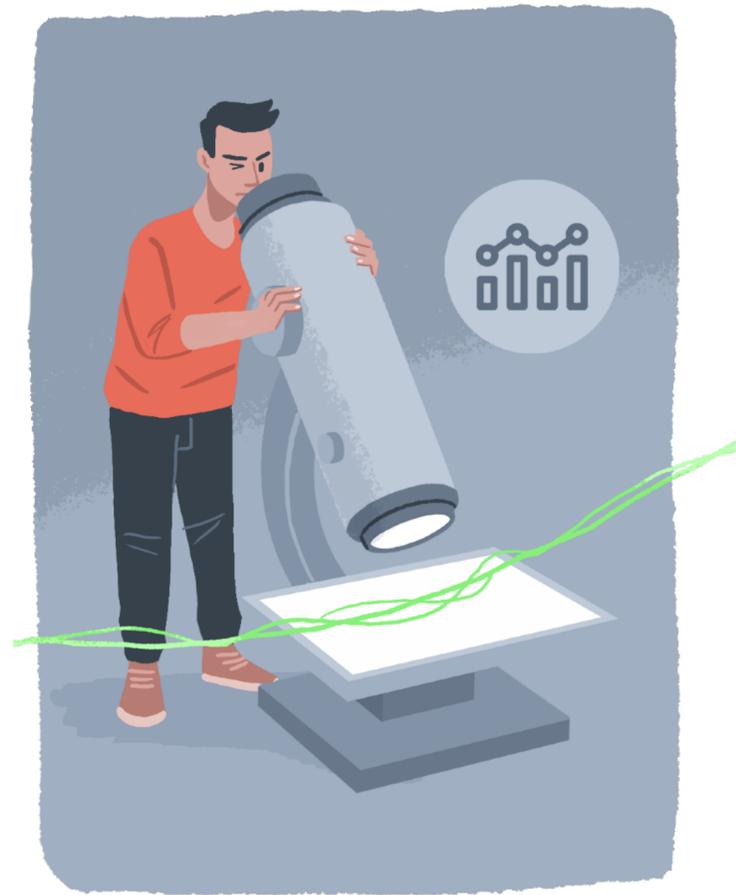
Data Reuse

They consider using their system in new locations and at different times of the day



Data Journey

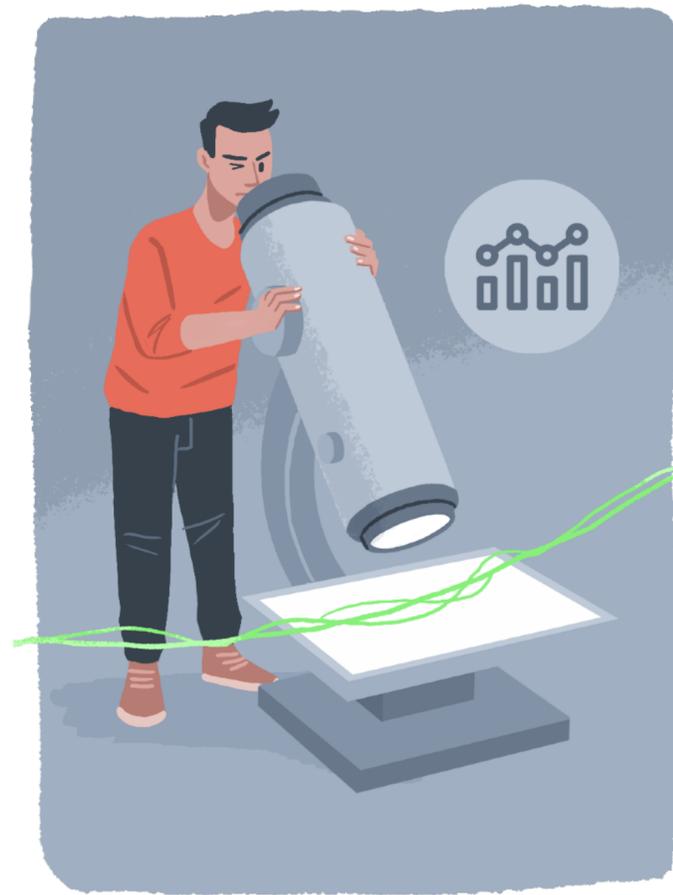
Data Monitoring & Deletion



The police force state that no images of members of the public gathered from CCTV are stored once they have determined that there is no match

Data Journey

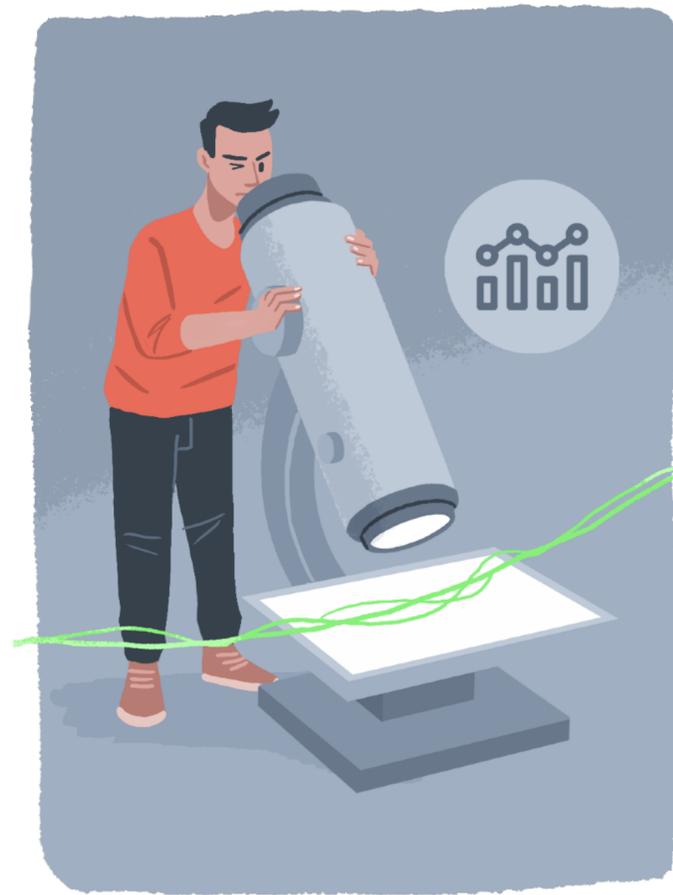
Data Monitoring & Deletion



Because of the previous issue with the low accuracy, they carefully monitor the performance of the new systems

Data Journey

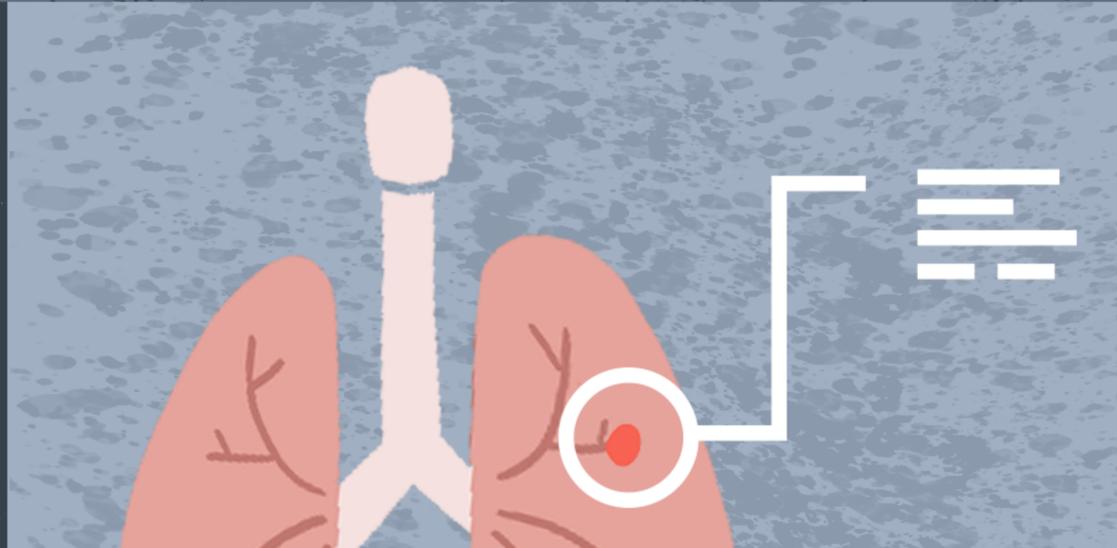
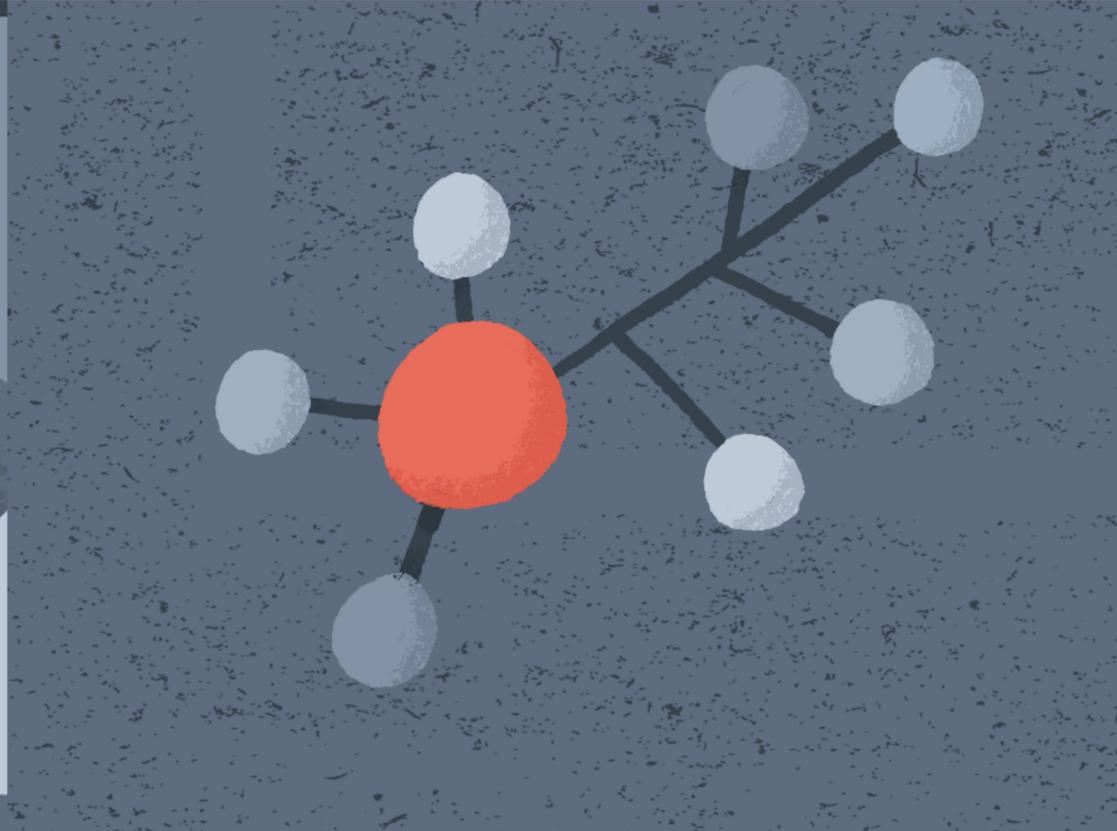
Data Monitoring & Deletion

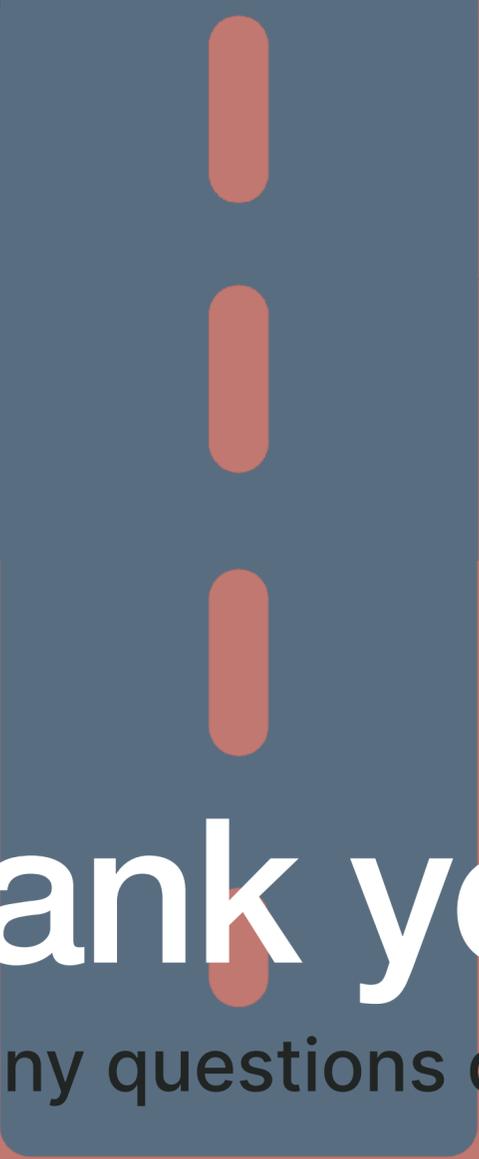


If the accuracy drops too low the police may have to stop the project altogether and delete all the data they have collected

Activity 2

Topics for Reflection





Thank you!

Please ask me any questions during the panel