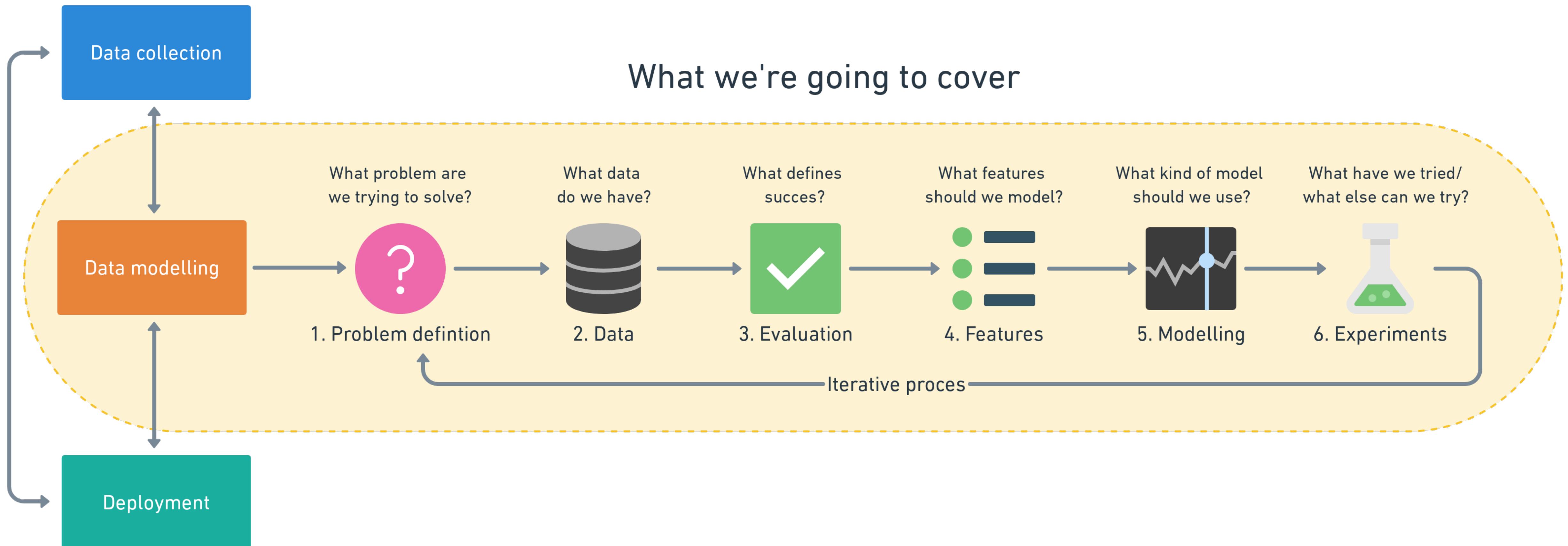


Steps in a full machine learning project



What we're going to cover

4. Features



“What do we already know about the data?”

Different features of data

ID	Feature variables				Target variable
	Weight	Sex	Heart Rate	Chest pain	
4326	110Kg	M	81	4	yes
5681	64Kg	F	61	1	no
7911	81Kg	m	57	0	no

Table 1.0: Patient records

Different features of data

ID	weight	Sex	Heart Rate	Chest pain	Heart disease?
4326	110Kg	M	81	4	Yes
5681	64Kg	F	61	1	No
7911	81Kg	m	57	0	No

Numerical features

Categorical features

Table 1.0: Patient records

Different features of data

ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?	Derived feature
4326	110Kg	M	81	4	Yes	visit in last year?
5681	64Kg	F	61	1	No	Yes
7911	81Kg	M	57	0	No	No

Table 1.0 : Patient records

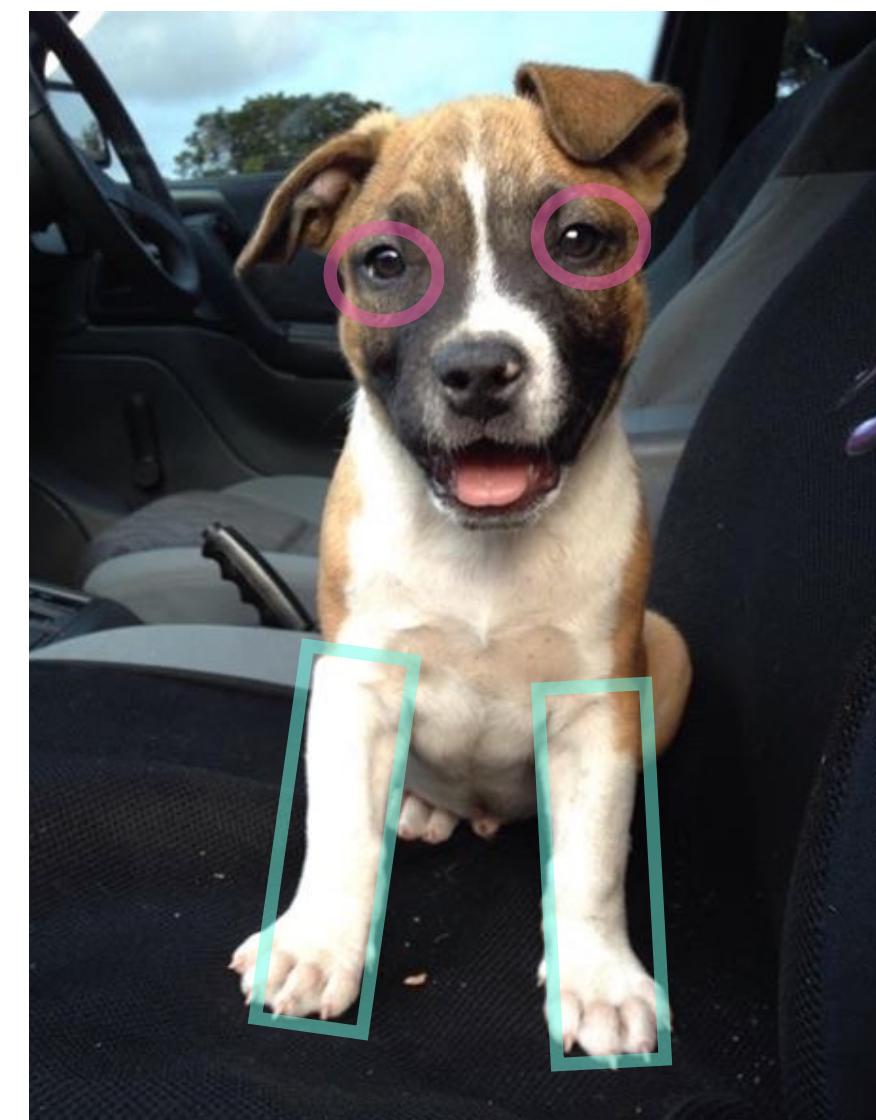
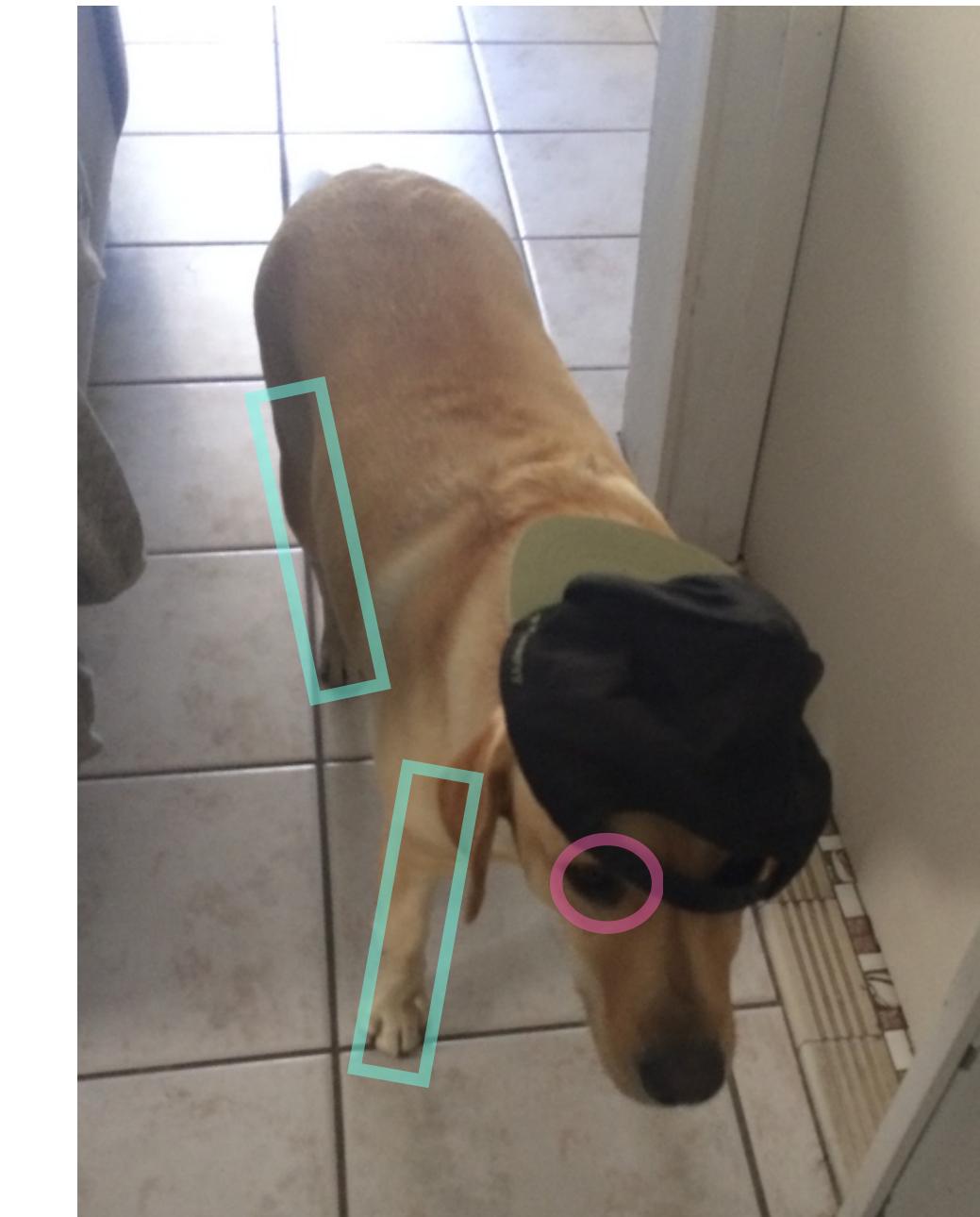
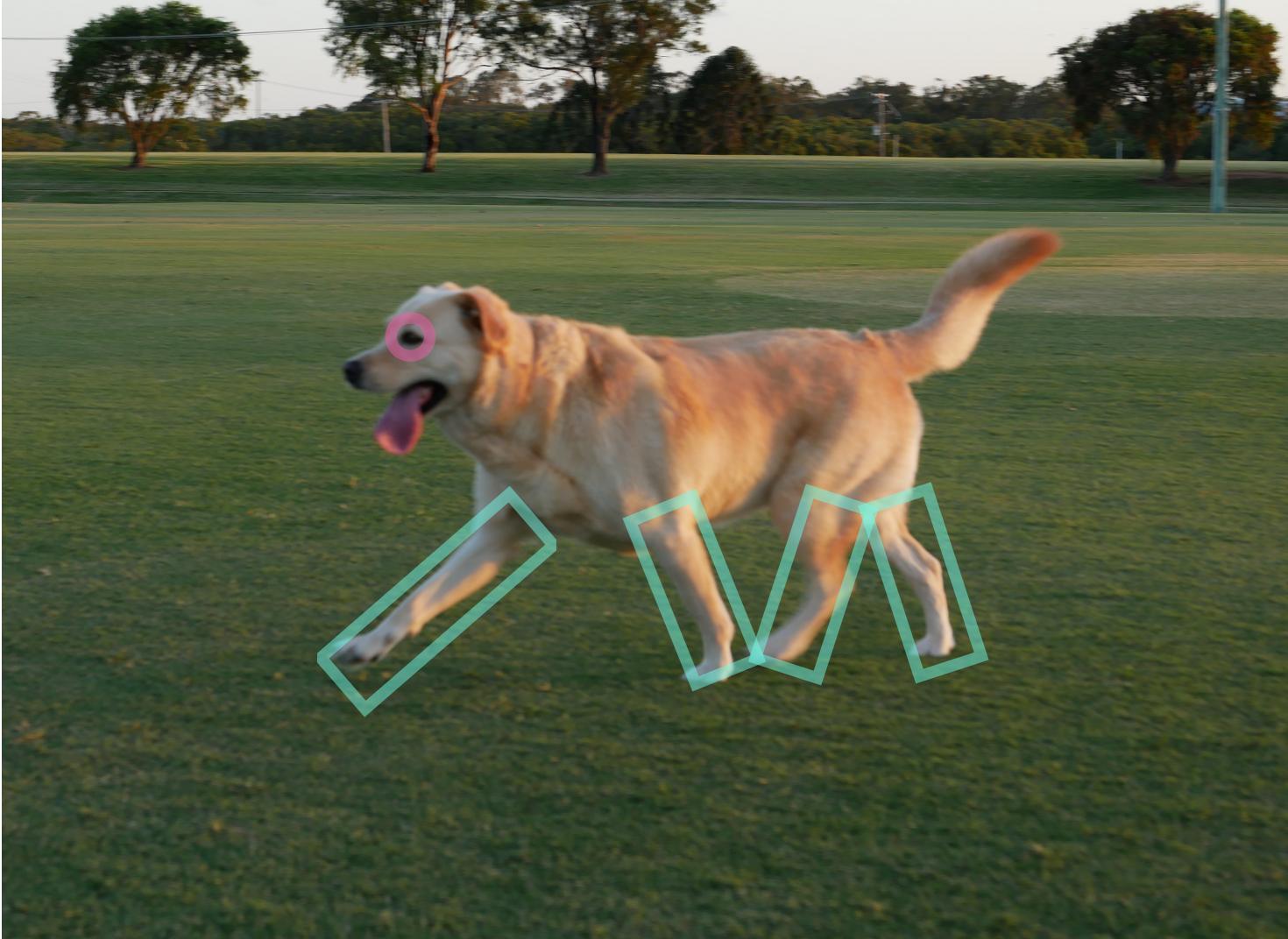
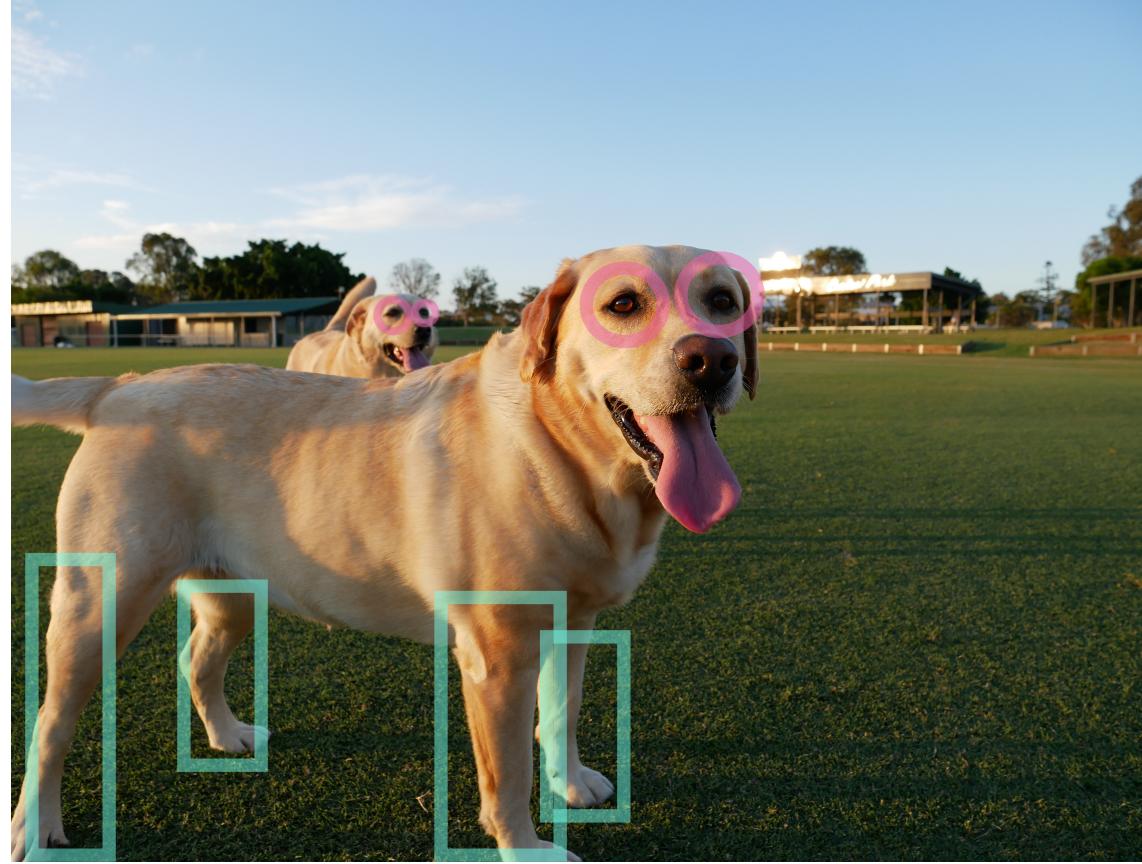
Numerical features

Categorical features

Feature engineering

Looking at different features of data and creating new ones/altering existing ones

Different features of data



What features should you use?

ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?	Most eaten food
4326	110Kg	M	81	4	yes	Fries
5681	64Kg	F	61	1	NO	?
7911	81Kg	m	57	0	NO	?

Table 1.0: Patient records

Want > 10% coverage

Feature coverage

How many samples have different features? Ideally, every sample has the same features.

**What are features of your
problems?**