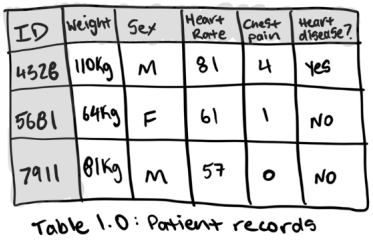
Features in Data

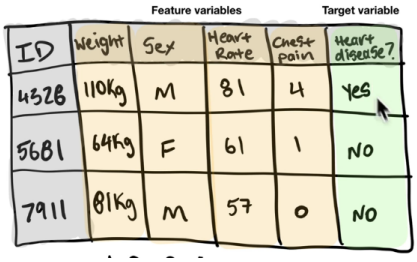
- **What do we know about the data?**
- **Different Features of data**



ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?
4328	110Kg	M	81	4	Yes
5681	64Kg	F	61	1	NO
7911	81Kg	M	57	0	NO

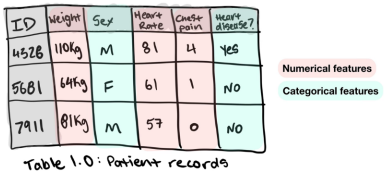
Table 1.0: Patient records

  - Weight
  - Sex
  - Heart Rate
  - Chest Pain
  - Use the feature to predict the target



ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?
4328	110Kg	M	81	4	Yes
5681	64Kg	F	61	1	NO
7911	81Kg	M	57	0	NO

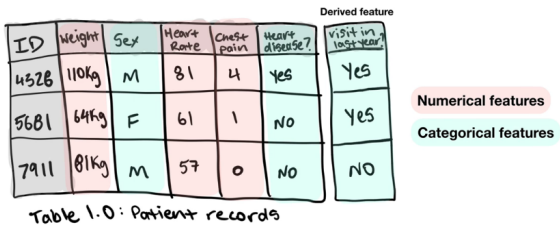
- **Types of Feature Variables**



ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?
4328	110Kg	M	81	4	Yes
5681	64Kg	F	61	1	NO
7911	81Kg	M	57	0	NO

Table 1.0: Patient records

  - Numerical features
  - Categorical features
  - Feature Eng - Looking at different features of data and creating new ones/altering existing ones

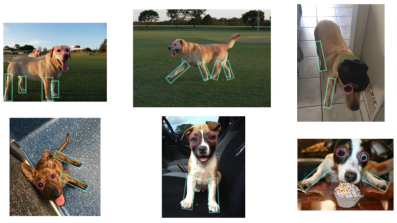


ID	Weight	Sex	Heart Rate	Chest pain	Heart disease?	Visit in last year?
4328	110Kg	M	81	4	Yes	Yes
5681	64Kg	F	61	1	NO	Yes
7911	81Kg	M	57	0	NO	NO

Table 1.0: Patient records

- **Features for unstructured data**

- Dogs Pictures



  - Legs
  - Eyes
  - Shapes

- **Features works in ML algo, when most of the samples have it**

  - Feature Coverage - How many samples have different features? Ideally every samples has the same features