

## **Assessment report**

### **Quality issues**

#### **twitter\_archive**

- some columns not needed
- timestamp wrong type
- some names in the name variable are not accurate
- some value in rating\_denominator should be 10s
- some values in rating\_numerator should not be more than 10
- from the source variable need to extract the source
- tweet\_id is numeric make it str
- remove non tweets wich has valus in the retweet columns
- expanded\_urls has some missing values which mean they do not have pictures

#### **image\_predictions**

- change column names to something descreptive
- tweet\_id is numeric make it str

#### **tweet\_json**

- tweet\_id is numeric make it str

### **Tidiness issues**

- the columns (doggo, floofer, pupper and puppo) in twitter\_archive need to combined in one column
- creat master table to join all 3 tables on tweet\_id

### **Clean**

### **Define**

#### **twitter\_archive**

- remove non tweets wich has valus in the retweet columns by querying twitter\_archive with out retweet rows
- expanded\_urls has some missing values which mean they do not have pictures, removes these rows

- the columns (doggo, floofer, pupper and puppo) in twitter\_archive need to be combined in one column, joining

- some values in rating\_denominator should be 10s
- some values in rating\_numerator should not be more than 10
- extracting the source from source variable
- some names in the name variable are not accurate, will replace 'a' with None
- timestamp wrong type
- tweet\_id is numeric make it str
- drop the columns are not needed

#### **image\_predictions**

- change column names to something descriptive
- tweet\_id is numeric make it str

#### **tweet\_json**

- tweet\_id is numeric make it str

#### **Create master table**