

- **Context**

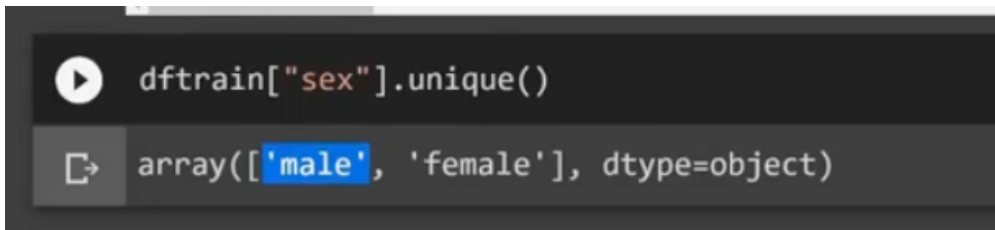
- he's loaded two datasets -> the training and testing dataset for the Titanic passengers
  - `deval.shape` <- the name of the data set. `shape` returns the number of rows and columns it has

- **Categorical vs numeric data**

- -> categorical data is discrete not continuous
- -> e.g the age isn't categorical
- -> categorical data (e.g the name of a place) is coded using an integer
  - or e.g male and female -> everything is coded into 1's and 0's
- -> doing maths on strings
- -> tensorflow can do this -> convert the names of places into numbers to be operated on

- **Feature columns for linear regression**

- -> what is feed into the linear model to make predictions
- -> the steps are
  - import the data sets
  - load the data sets
  - explore the data set
  - create the categoric and numeric columns (e.g sex / age etc)
- -> then he creates them into feature columns
  - -> these are the ones which e.g store the ages of the passengers etc
- -> he's printed out the unique values in the feature name
  - -> the values which weren't repeated
  - -> this (below) prints out the unique values



```
dftrain["sex"].unique()
array(['male', 'female'], dtype=object)
```

- -> he's done another example but for the name of the destinations for the passengers
- -> this creates a column which has the feature name and the different vocabulary associated with it
  - -> creating the columns associated with those different names / entrants

- **Numeric columns for linear regression**

- -> give the feature name and data type
- -> in this example there is a key and data type listed
- -> what he's doing is replacing the parameters for example 'male' and 'female', but with numbers instead
  - -> he hasn't actually made the model yet