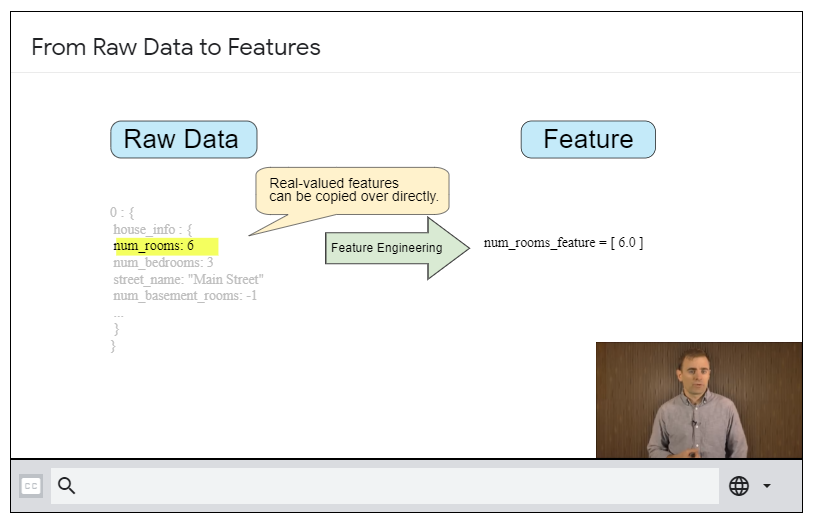
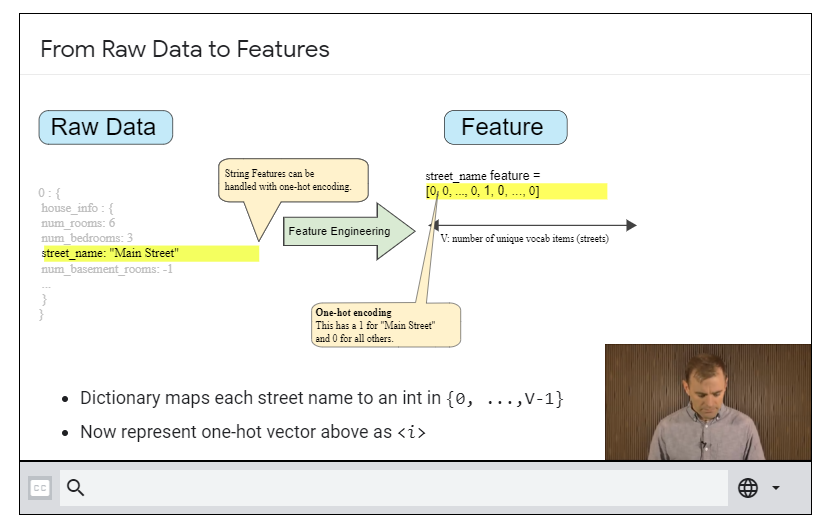
Representation

A machine learning model can't directly see, hear, or sense input examples. Instead, you must create a **representation** of the data to provide the model with a useful vantage point into the data's key qualities. That is, in order to train a model, you must choose the set of features that best represent the data.



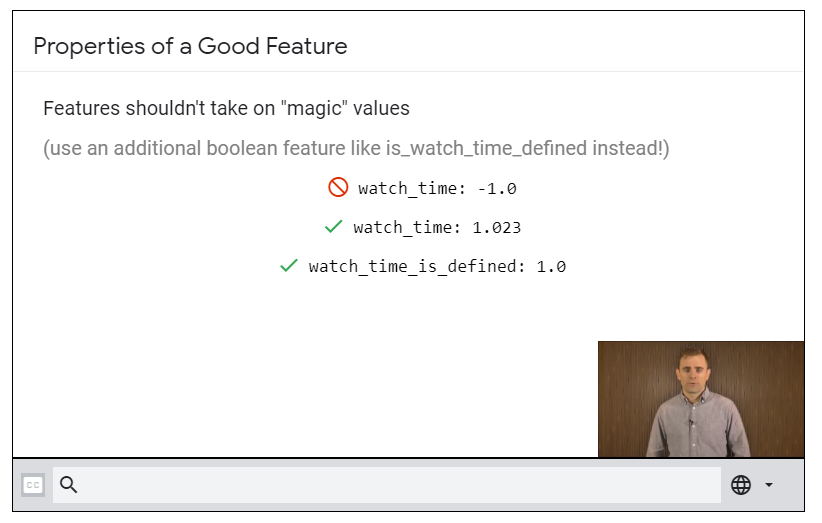


For data like num\_of\_rooms, we can directly incorporate into our model as a feature.

But for stuff where an attribute is a ‘string’, then we could make use of one-hot vectors.

For eg. Street name = “puthiyatheru” among other possible street names, we can encode it as: [0, 1, 0]

## **Properties of a Good Feature**

* Feature values should appear with non-zero value more than a small handful of times in the dataset. If the feature only has value very rarely, then it probably isn’t a good one & can be filtered out.
* Features should have meaningful values. For eg: House age is more meaningful when represented in units of years, rather than UNIX time or something.
* Features shouldn’t take on magic values. For eg. 
* 
* 