

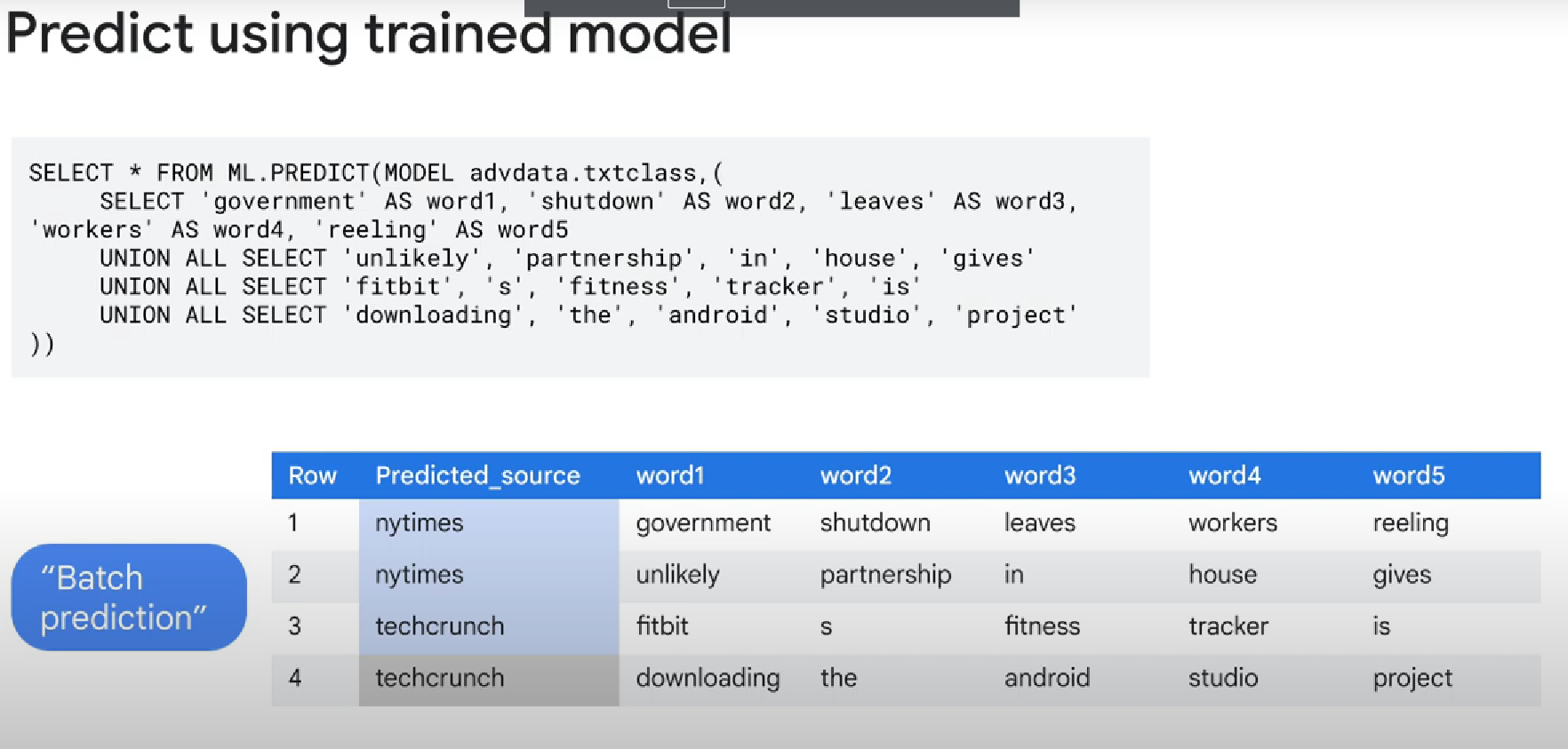
(??????)

High precision: low false positive rate.

Recall: correctly predicted positive observations/ All observations (how many you got right out of both true positive and false negative)

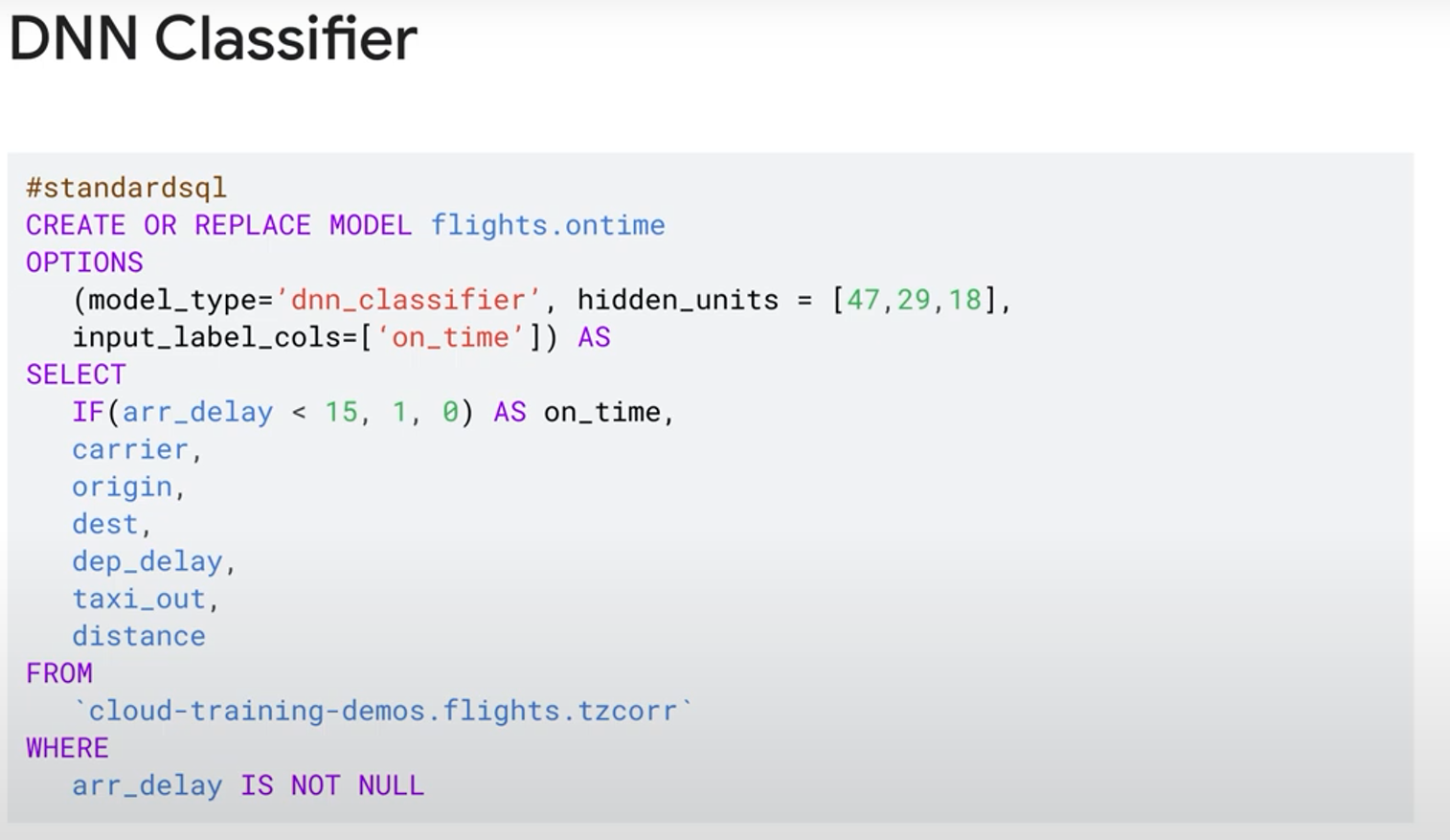
Accuracy: true positive + true negative / all set of observations

(??????)





Linear classifier/logistic regression: binary output, 0 or 1



DNN are used to modelize to linear relationships

Other models:



If the output isn’t binary:

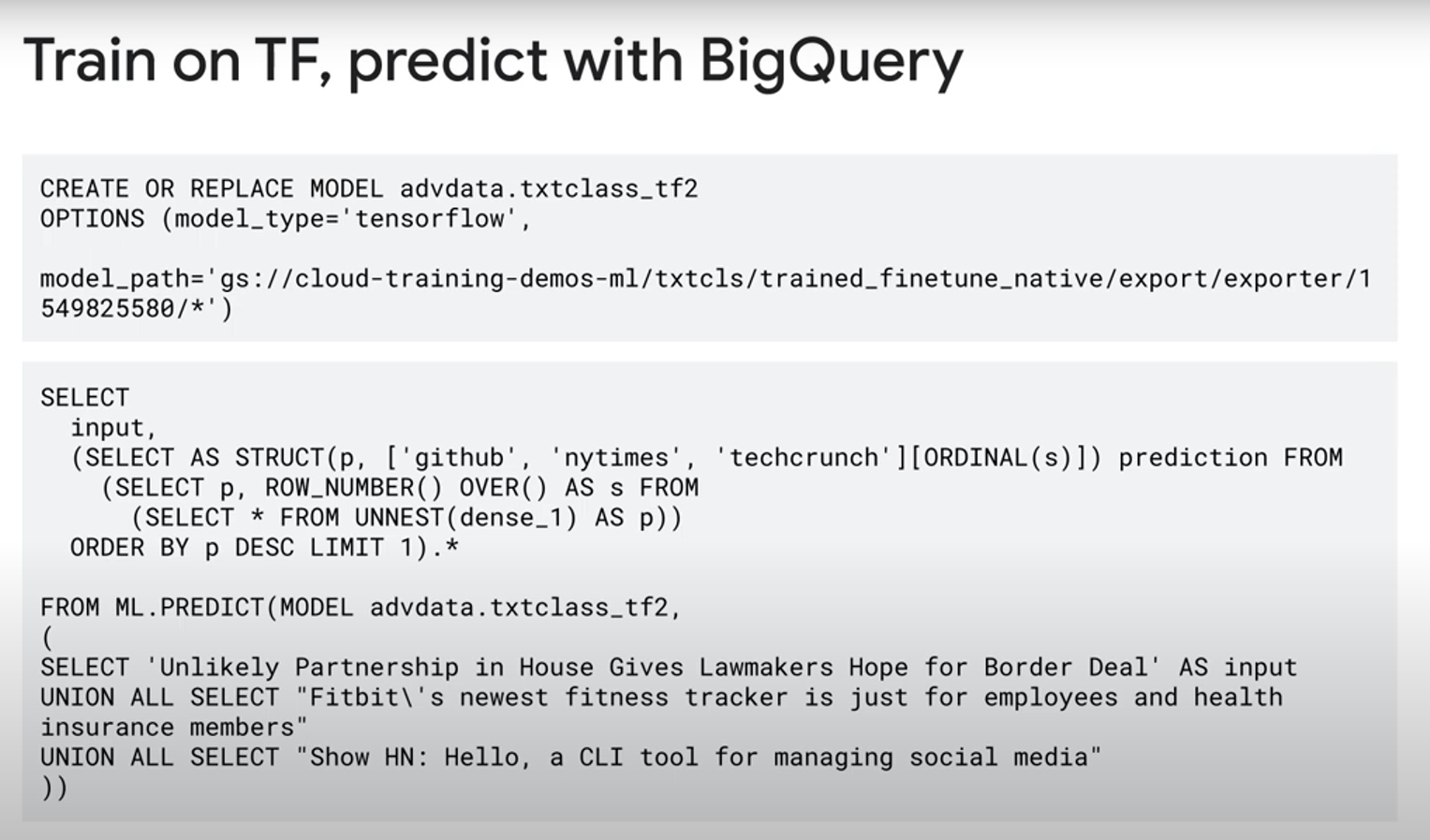




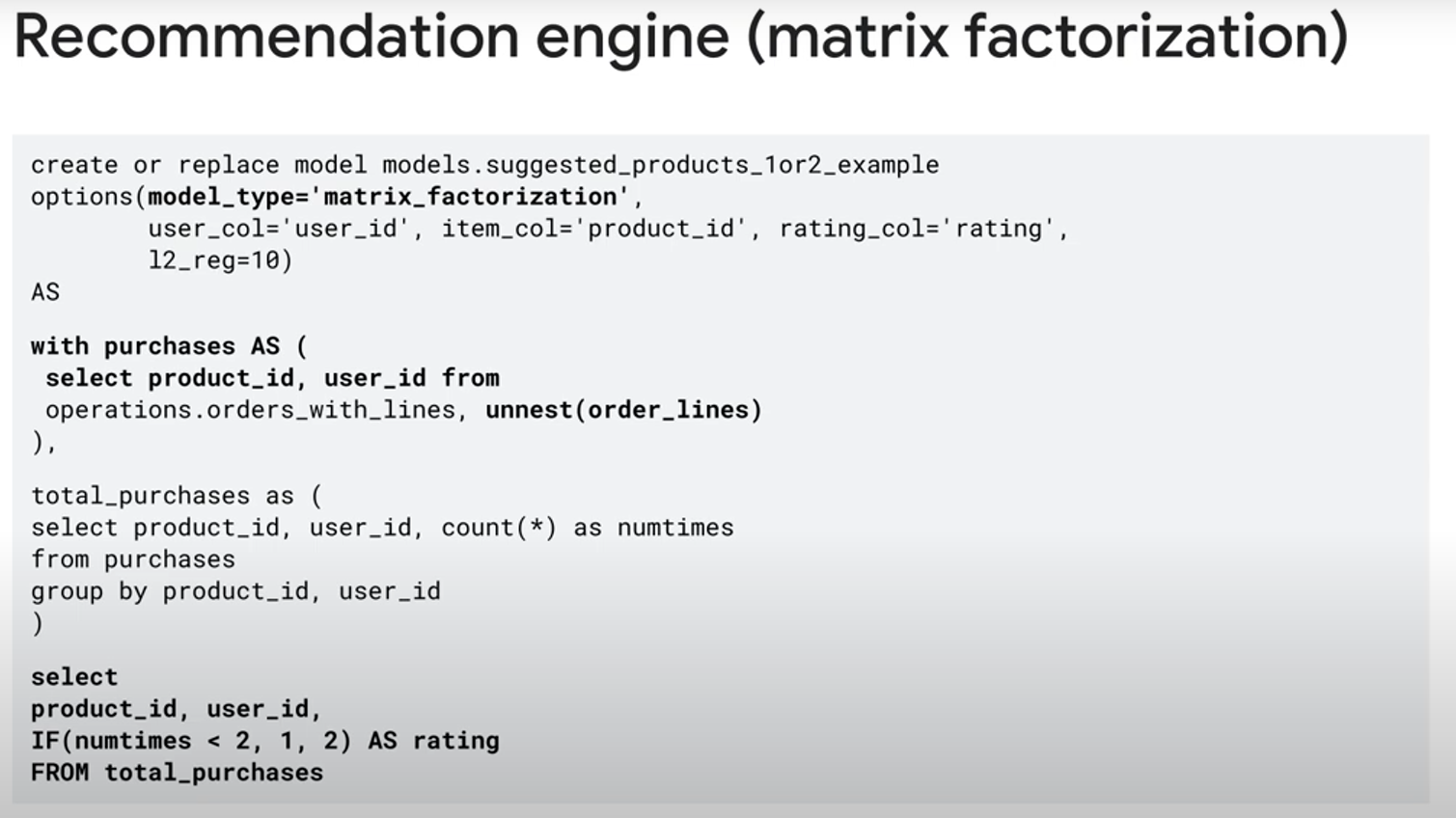


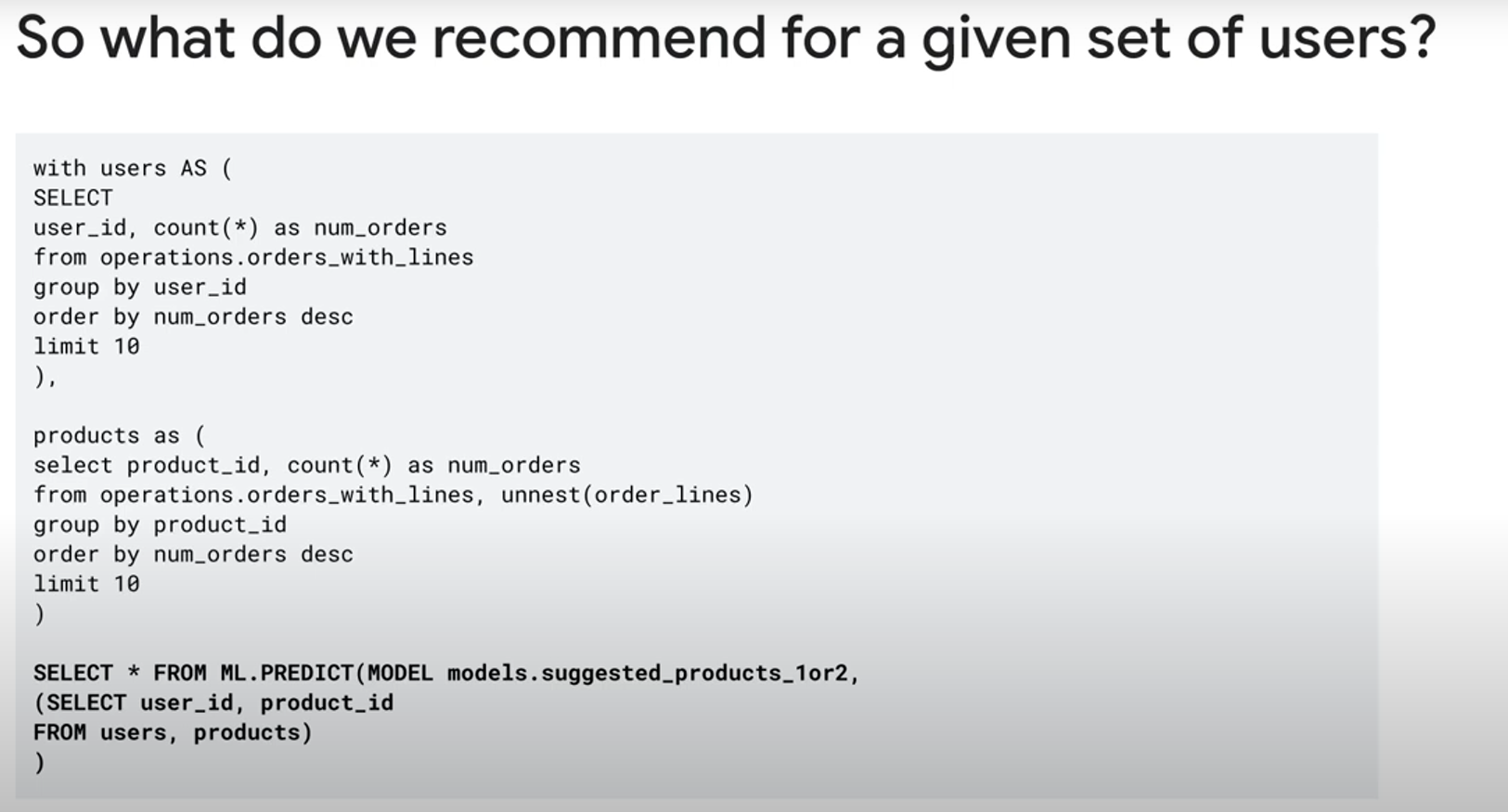
Classifier: Binary output

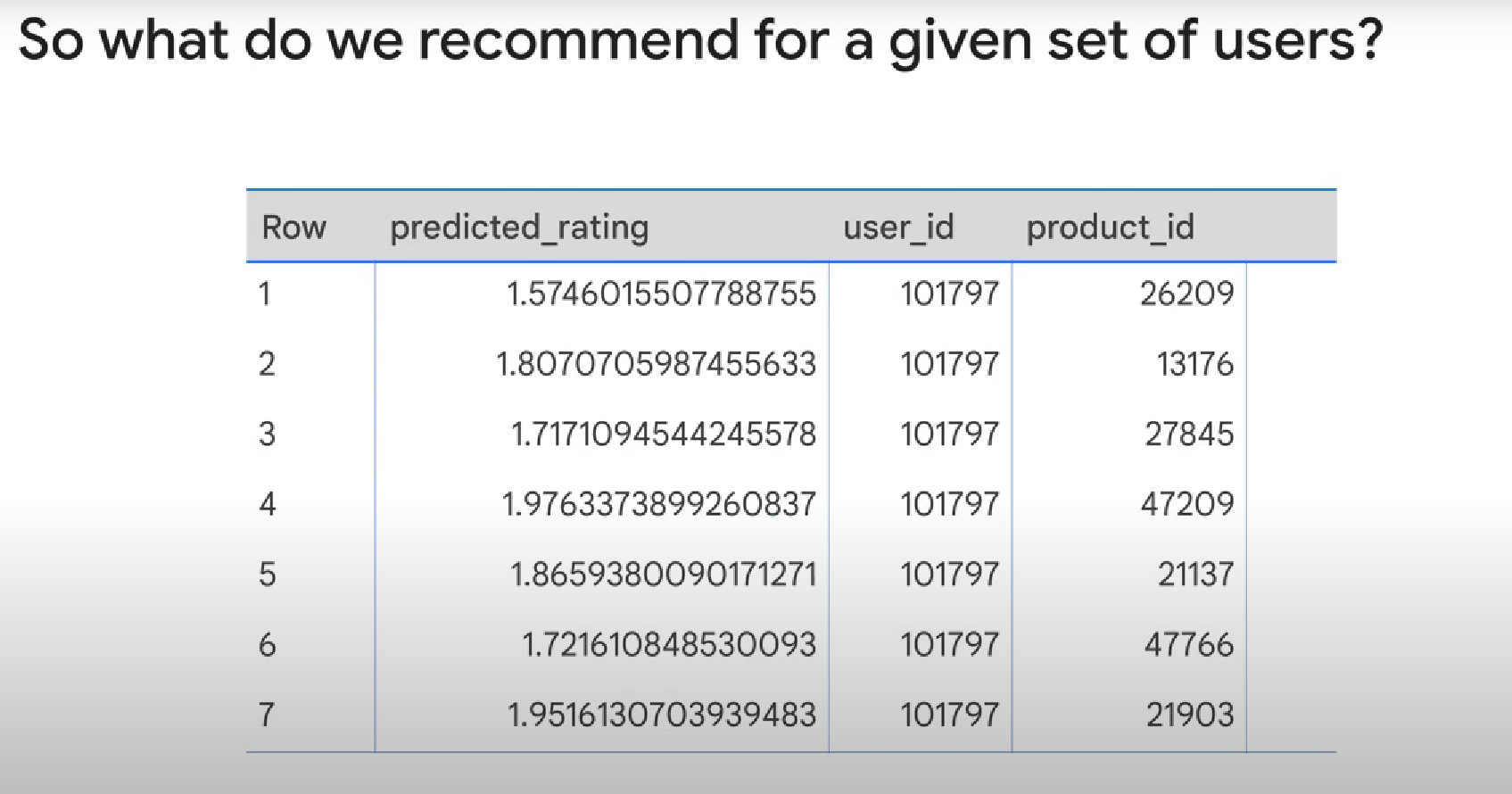
Regression: float output



To recommend a certain product to users:

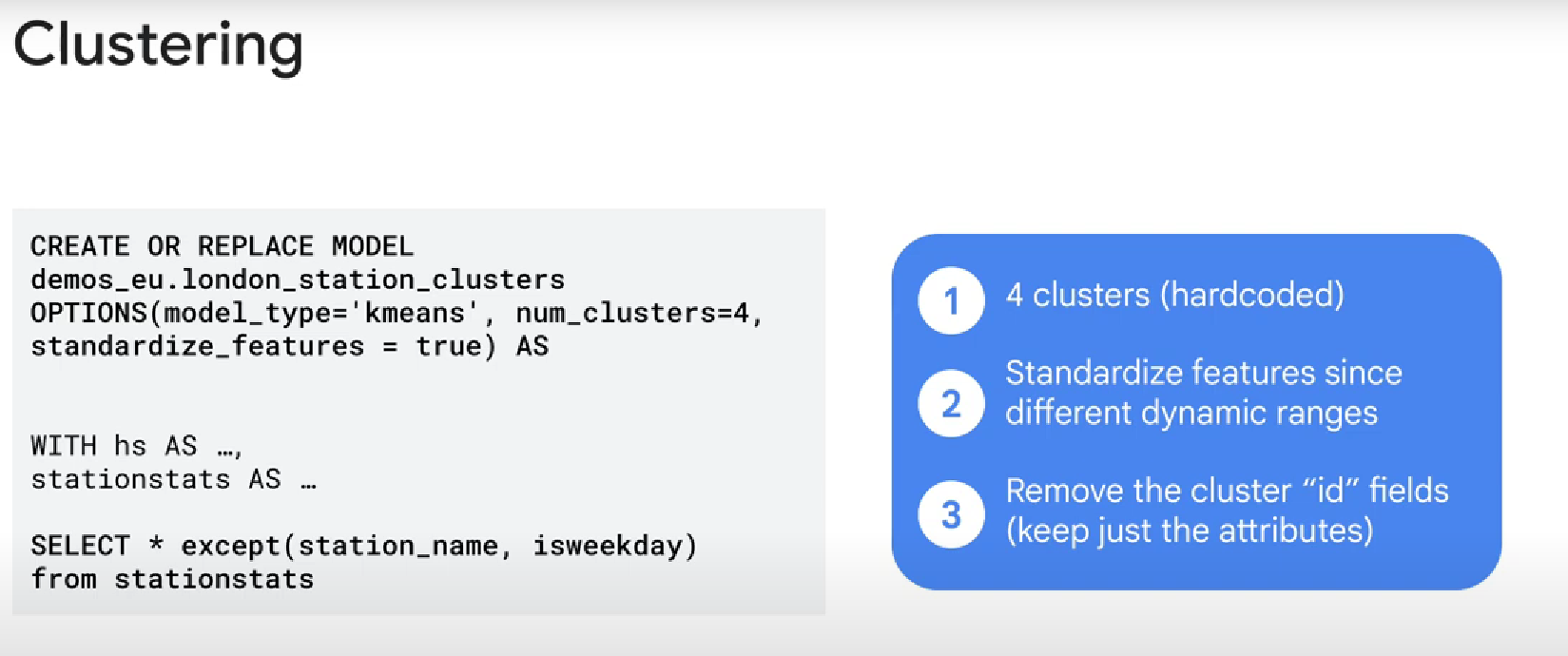


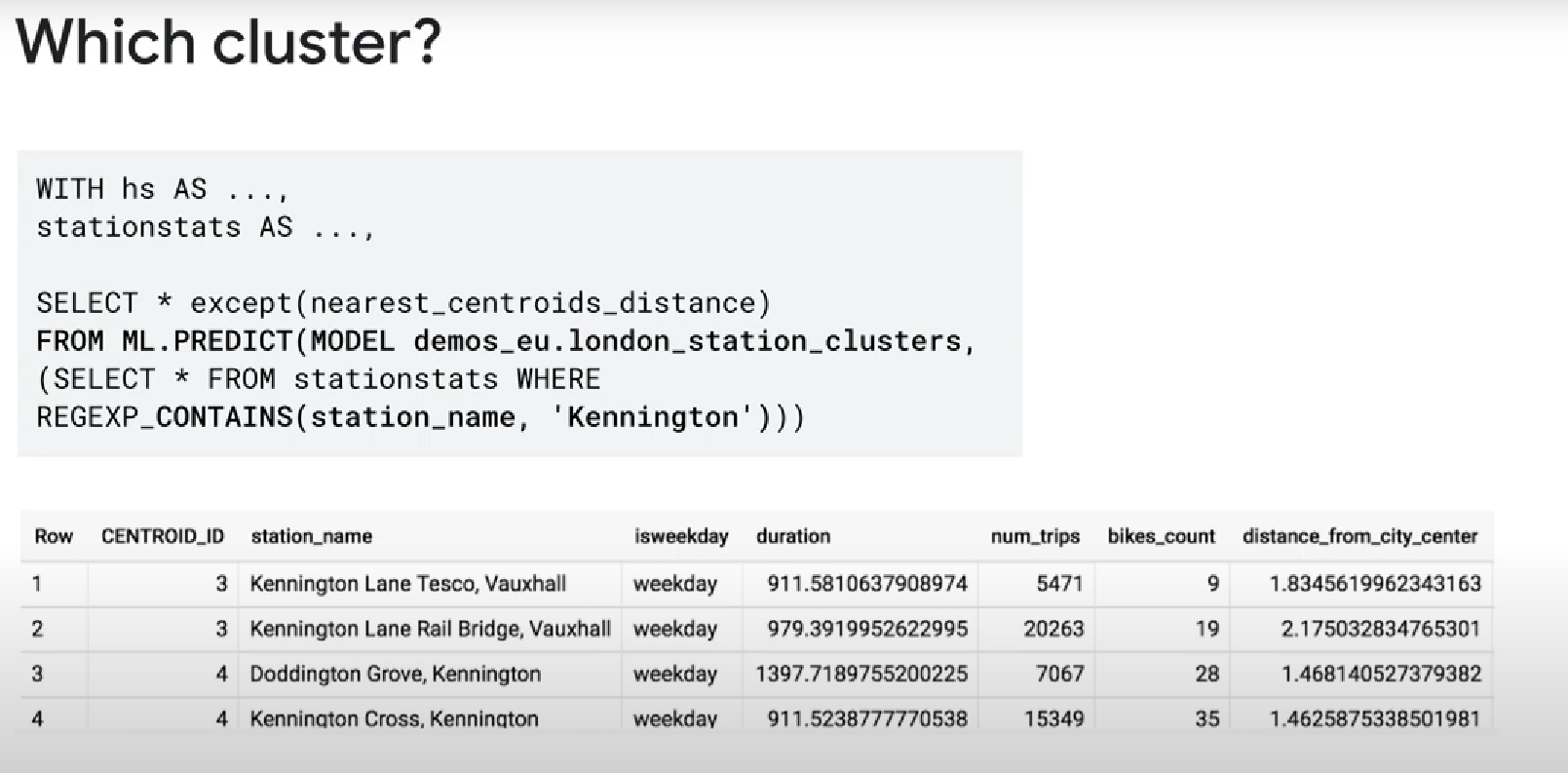




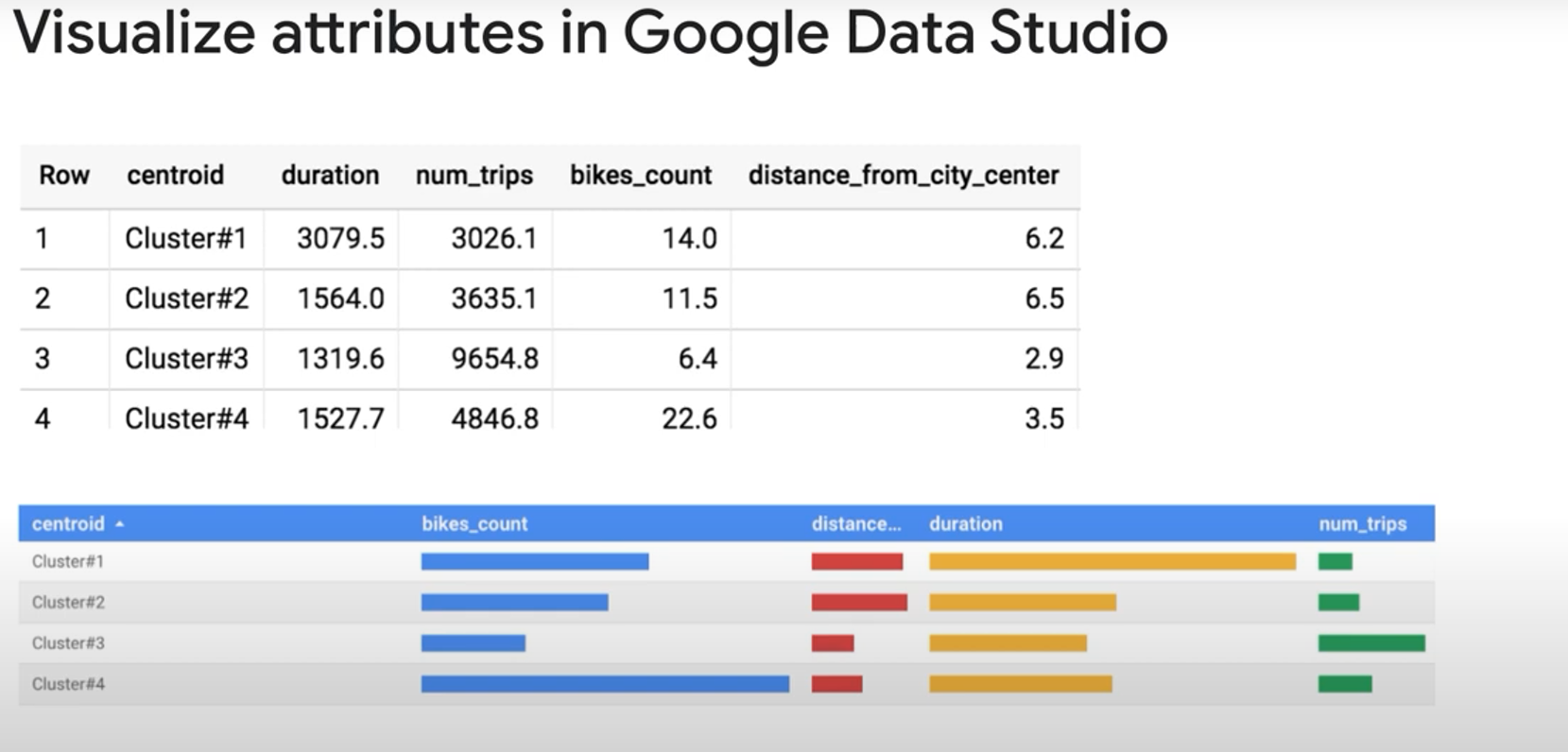
uNSUPERVISEd learning algorithm using Kmeans clustering

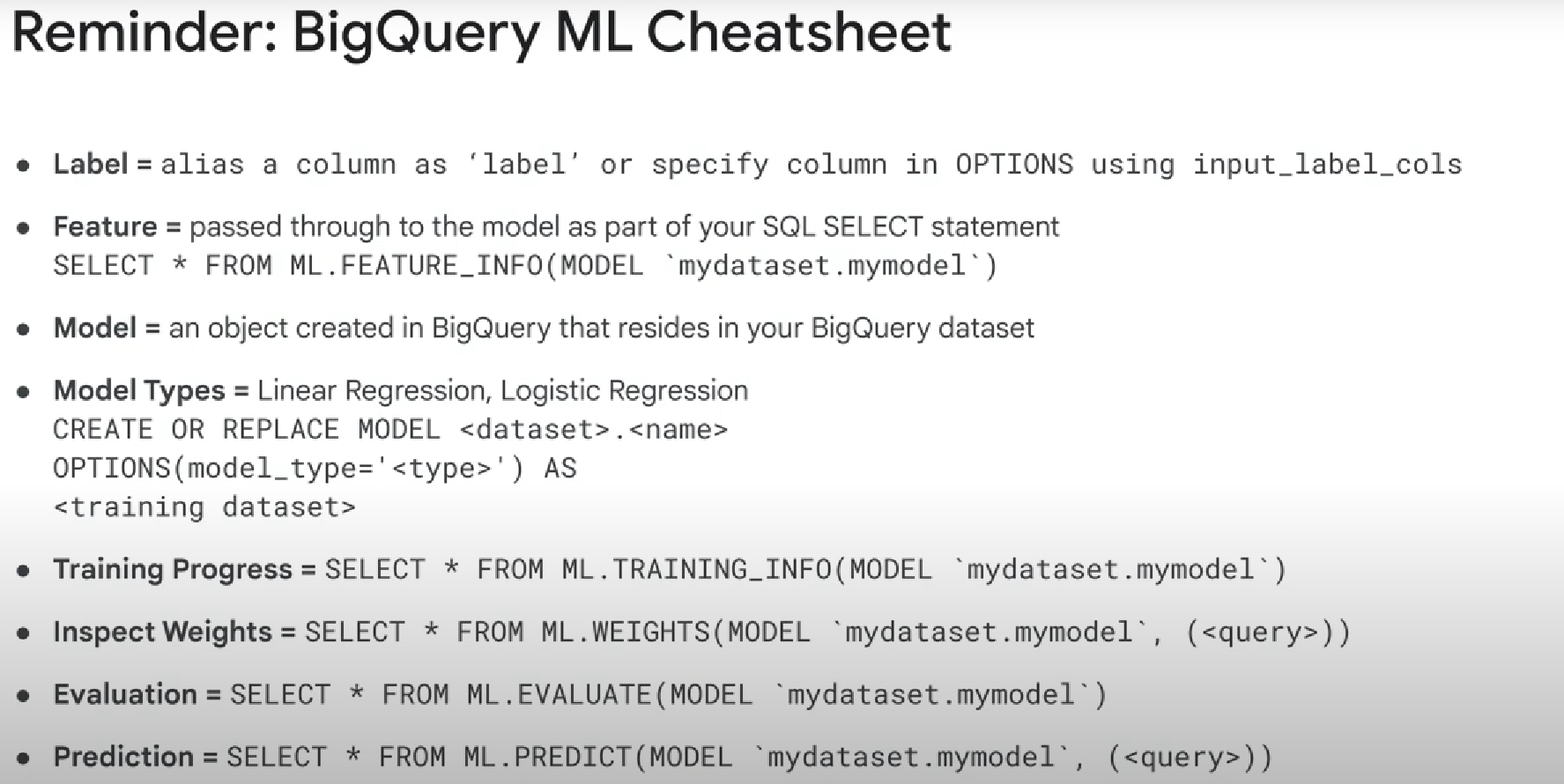
Unsupervized = you don’t have a label that you’re looking for. You try to group the elements of a dataset to K clusters, using their features.



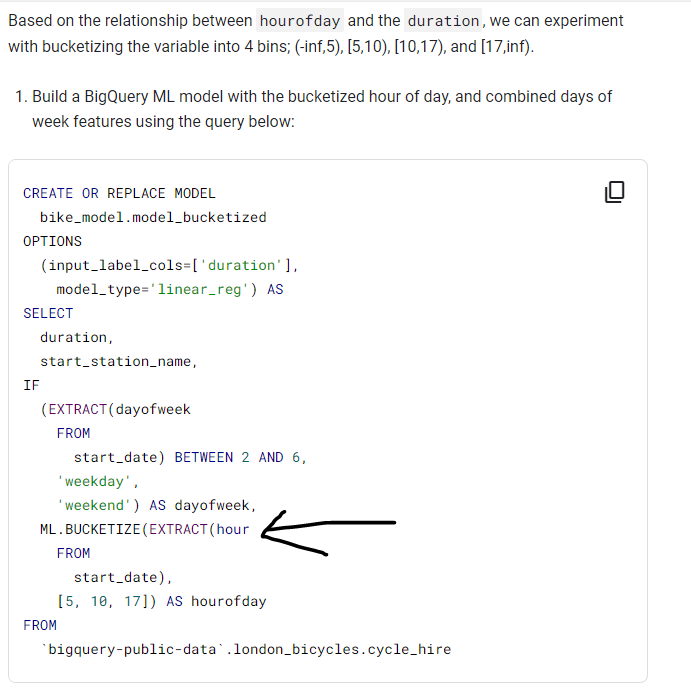


What to do with the clusters? ANALYZE THEM MANUALLY TO UNDERSTAND THE CLUSTERS!





Bucketize method (put a group of value in a single «bucket»/»value»)



You can add a data transformation in the model. For instance, if you have a field with a complete data+time, and you need to separate the time and the data as two different features, you can do it directly in the model. When using your model to predict things, you won’t need to transform your input, you can keep the field as is.

