Notes 8.1.2022

Issues with project so far:

* With categorical data, you need to make sure they are turned to factors and represented with bar charts and not histograms.
* Data cleaning
  + Need to get rid of the outlier
  + Change it to a NULL value (use na.omit)

Final Project:

* Send over a nice markdown document like you’re send a report for a business.
  + Label things Correctly
  + Don’t make things too complex

Regularization:

* If we did poorly on the previous assignment, as long as we run the model in the right way that’s fine.

Lecture

Two types of “learning”

* Supervised learning: you have the dependent variables
* Unsupervised learning: independent variables, look at characteristics to predict an outcome
  + recommendation engine
    - Ex. Algorithm of tiktok: % video watched, liked, commented, etc. -> what videos you’ll like etc.

What is distance?

* Euclidean distance = A­­2 + B2 = C2 or Sqrt (X2