Machine Learning algorithms

During the machine learning part of the assignment, we decided it would be best to use logistic regression to predict if a country had recycled 25% or more of its total waste, we did this by taking the year and the total weight of trash collected from 38 different OECD participating countries. Something interesting we found in the data is how close every countries trash amount was, excluding a single country, the United States, whom had more than three times the amount of trash then the next nearest country. Unfortunately, the model didn’t work very well, and with nearly every combination of trash amount and year, it had predicted that the country recycled more than 25% of its total trash, which is a nice idea but also simply untrue. Next, we decided to see what would be best to predict a cities air quality using a couple of simple factors. To predict air quality we used, dew, humidity, pressure, temperature, and wind speed. When using a multiclass logistic regression we found that humidity was the best for predicting air quality, with dew and temperature having some effect, but not surprisingly pressure and wind speed didn’t have much of an impact, with pressure being nearly the same across the globe and the wind speed just not telling us much of anything. Last but not least, we finished of with a clustering model of different countries arable land as a percentage of total land, these countries Urban land in km^2 and each countries political rights scored by the [Freedom House Index](https://freedomhouse.org/explore-the-map?type=fiw&year=2025). The data had mediocre clustering. We used the k-means method, with k = 7. Most of the data clustered into 1 cluster, since many countries don’t have very large urban areas, even if they have varying levels of arable land. But outliers do seem to be a concern with this data, with China, the US, and India all being very far away from every other country within the data. This might have worked better with other forms of clustering such as DBSCAN, because we wouldn’t have been so influenced by these couple of outliers.