Algorithms HLT

Algorithm	Supervised, unsupervised, or reinforcement?	Function	Useful situation	Example of use
Linear regression	Supervised	It is used to train a model to predict how your data will behave depending on the variables that are input. There are two types of linear regression: simple, where there is only one independent variable, and multiple, where there is more than one independent variable. Simple: $y = b_0 + b_1x$ Multiple: $y = b_o + b_1x_1 + b_2x_2…+b_nx_n$ Whichever one is used, the model is used to find the best fit linear line, where the difference predicted values and the actual value are minimised.	Best used when you would need to know trends based on what variables you want to manipulate and whether those variables even make a significant difference to the trend. This could save time and costs and let stakeholders effectively strategise moving forward.	
Logistic regression	Supervised	Estimates the probability of an event occurring, based on the data that has been input. It uses previous observations to binarily classify outputs. Unlike linear regression, where a straight line is used to show a positive or negative correlation, a sigmoid function is used to, where real values are mapped into a value between 0 and 1. This is done using the formula: $f(x) = \frac{1}{1+ e^{-z}}$	In any situation when predictions are discrete/can be categorised. The simplest example would be determining whether an email is spam or not. Emails would first have to be identified as whether they are spam or not spam, either by researchers, or even users who report them. Then, typical characteristics of the email are extracted, such as unusual sender domains, typos, buzzwords etc, and they can be used to train the algorithm to classify emails as spam or not.	