For **each** of the scenarios below, answer the following questions. You do not have to explain your answers other than to explain where your targets would come from (Are they in the dataset already? Do you need to create them by hand?) and how you would make any non-numerical inputs numerical, if required by the algorithm you choose.

1. What type of machine learning problem (regression, classification, clustering) do you think this is?
2. If this is a supervised problem, what will you use as your targets (aka labels) and how will you get them? If this is an unsupervised problem, just write "none".
3. What processing do you need to do to your input data? (How will you handle non-numerical inputs? Do you plan to do any scaling? etc.)
4. What type(s) of model(s) would you try? Remember to start with the simplest thing that might work! The types of models we've talked about are linear regression, decision trees, random forest, logistic regression, naive bayes, K-means, DBSCAN, and fully connected neural networks.
5. What validation metric(s) would you use to decide how well you're doing?
6. What ethical challenges do the data collection, creation, and/or use of this model create? If you feel there aren’t any, just say “None”.

Scenarios:

1. You are playing fantasy football and want to predict how many points each player will score next season. You have their stats, including points scored, from last season, plus their height, weight, and position -- except for players new to the league, for whom you don't have last year's stats, only height, weight, and position.
   1. This is a regression problem.
   2. The labels are the points scored, which is provided.
   3. 0-1 normalization and ohe.
   4. Neural networks and linear regression (but would perform likely poorly).
   5. r^2.
   6. None.
2. You have customer reviews, each one of which has a rating from 1 (worst) to 10 (best) and some text. The reviews vary greatly in their length. You would like to use this to write a model that can predict if text is positive, negative, or neutral, along with a probability score (e.g., 68% likely to be positive, 30% neutral, 2% negative).
   1. This is a classification problem.
   2. The labels will be the ratings (converted to positive, negative, or neutral).
   3. Bag of words, TF-IDF, or word vectors must be used.
   4. Naive bayes.
   5. Accuracy and f-score.
   6. None.
3. You have data from a movie streaming service that consists of lists of movies that each user has watched as well as information about each movie: title, names of the stars, genre, and length. You would like to make a model that will help you decide what movies to recommend to users.
   1. This is a classification problem.
   2. None.
   3. OHE and bag of words, TF-IDF, or word vectors.
   4. NN.
   5. Watchtime of the recommendations.
   6. The company learns a lot of personal info about the user through the data on what they watch.
4. You want to predict whether a random stranger owns a cat, a dog, or neither, based on things that they like on Facebook. You decide to train your model on your friends, and write a program to collect all of their public Likes.
   1. This is a classification problem.
   2. Whether they own a cat, dog, or neither. They would have to ask their friends for this information.
   3. Bag of words, TF-IDF, or word vectors.
   4. NN.
   5. Accuracy and f-score.
   6. The program that colelcts all of their public likes can be viewed as invasive. Also, this model strives to determine private info from public info, which is problematic.
5. You want a model to predict the number of deer that will be born in a breeding season. You have a large amount of historical data, and each row consists of the following information for the breeding season of a particular area and species:
   1. number of fawns born
   2. the genus and species
   3. number of does sighted during the mating season
   4. vegetation quality during the mating season ("low", "average", or "high")
6. This is a regression problem.
7. The number of fawns born column would be the labels.
8. OHE.
9. Linear regression and NN.
10. r^2.
11. None.