Rough Idea:

Original: Take player attributes from FM20 (football manager 20) and see if we can predict player positions from their attributes. (eg pacy, high flair, good crossing and shooting -> winger; good tackling, marking, jumping reach, heading -> centreback). However, this data does not exist as a public dataset, (the game will auto-generate player attributes from a range for each individual save) and compiling on one's own would be too complicated.

Modified: FIFA, which is a game with slightly less detail than FM20, is released on PC and thus player attributes and overalls can exist as a public dataset. One was found on Kaggle: https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset?select=players_20.csv

And we will try to use this. While FIFA may be less extensive than FM when it comes to player roles (FM will specify, for instance, Complete Forward v Pressing Forward v Deep-Lying Forward v Target Man (all different roles for a central striker) and FIFA only really has Centre Forward and Striker, this should still be an extensive dataset to work with. On the whole, the dataset contains 18278 players, which should be enough to draw good conclusions.

2: I expect to be isolating player attributes and their preferred positions. Details like date of birth, player name, club and country aren't important, but Fifa contains attributes for every player. These include general attributes (pace, passing, shooting, dribbling, defending, physical), as well as certain subcategories, like: attacking, skill, movement, power, mentality and defending. Additionally, players are listed with their favourite positions. Players do have more than one (for instance, Mo Salah is both a striker and right wing) and I am not quite sure as of yet whether I will be working for only the first position listed or all (leaning towards all at the moment, although the first position listed is the player's strongest position). I may exclude goaltenders (this is possible, as the database allows one to select for position when picking what you want to see) due to their limited abilities outside of the net (as well, there are several "goalkeeper attributes", although all outfield players will have low goalkeeping stats and vice-versa, to a lesser extent).

I haven't looked too much into which algorithm I will use at the moment, although as this is a classification problem (finding best suited position for a player from given attributes), and from the algorithms I have seen in class, I presume that I will use something similar to k-nearest neighbours. Each outfield (ie non-gk) player could be classified as a 29-35 vector of his attributes, and my hypothesis is that different positions will see different groupings of these attributes. However, I will look for a better algorithm to navigate this problem in the future, as I haven't really thought of other algorithms. One potential problem is that as not every player is of the same overall ability, that some players may be "farther along" than other players who share the same position. Potentially I may have to look into normalizing player ability before I implement an algorithm like k-nearest neighbours, but I will see.

As this is a classification problem, I guess I will rely on the confusion matrix and accuracy/precision-recall/logistic loss to determine the success of my algorithm, but again I am not quite sure.

As for the webapp demo, I guess I could work on a webapp where either people could input values for attributes and see what position that best links to, or perhaps users could input a position and the attributes most closely associated with that position.