CSc 496, Homework #4: Batted ball analysis using Statcast.

Due date: October 17th, 2024. **No late assignments will be accepted.**

In this assignment you will learn a little bit about Statcast data. This data is incredibly rich and allows you to do really interesting analyses.

You will use 2022 and 2024 batted ball data that I have already pulled from Baseball Savant and posted to Piazza. There are several things you are to do with this data. The first is to develop several regression models, each of which uses the file `all2024.csv`. Among other things, for each batted ball in the file (one per line), there are fields for the result (e.g., "field out", "home run"), launch speed, and launch angle.

- Perform two regressions to generate a prediction of whether a batted ball is a home run based on launch angle and launch speed. The first regression should be linear, and the second should be polynomial. These should be logistic regressions.

- Perform three regressions to generate a prediction of the distance of a batted ball. The first regression should be linear using *only* launch speed, the second regression should be linear using *only* launch angle, and the third should be polynomial using *only* launch angle.

- Perform two regressions to generate a prediction of the distance of a batted ball, using *both* launch speed and launch angle. The first regression should be linear, and the second should be polynomial.

The program should take as input (on the command line) a CSV file. (It will be `all2024.csv` when you are testing your program, but I reserve the right to test with a different CSV file.) Please make sure that each regresson outputs the mean squared error, the R^2, and the coefficients (one number per line). Nothing else should be output.

Next, you will investigate the effect of pulling the ball on OPS. The definition of a batted ball that is pulled is (1) for a left-handed hitter, the ball goes towards right field, and (2) for a right-handed hitter, the ball goes towards left field. For a ball hit to the opposite field, these are reversed. For a ball hit straightaway, for both right- and left-handed hitters, the ball goes towards center field. Note that Statcast data indicates direction using the integers that represent baseball positions; the relevant ones for you are 7 for left field, 8 for center field, and 9 for right field.

First, you will do the following.

- Compute the OPS for all hitters (for batted balls), for left-handed hiters, and for right-handed hitters.

- Do the same for balls that are pulled, hit straightaway, and are hit to the opposite field. The files `all2024.csv`, `pull2024.csv`, `straight2024.csv`, and `oppo2024.csv` have this data. Note that this is only the NL central data; do not worry about this. Note also that the union of the data in `pull2024.csv`, `straight2024.csv`, and `oppo2024.csv` is exactly `all2024.csv`.

- Repeat for the 2022 entire season; the naming convention is the same except that "2022" is used.

This program should take in as input (from the command line) a CSV file and then output three OPS values (all hitters, lefties, and then righties). Please print one value per line.

Finally, you will investigate the effect of pulling the ball on fly ball distance. Here, I only want your program to output (each on a separate line) the total number of fly balls, the number of fly balls pulled, the

average distance of pulled fly balls, the number of fly balls hit to the opposite field, and the average distance of fly balls hit to the opposite field. Your program must take two command line arguments: the first is the file of pulled balls and the second is the file of balls hit to the opposite field. Please note that these files contain all batted balls and you must filter out all batted balls that are not fly balls.

You should submit your python files on lectura using the `turnin` command; for this program, use the assignment name `csc496-f24-hw4`. **Here is a list of files to submit.**

- `launchLogitLinear.py`

- `launchLogitPoly.py`

- `launchSpeedOnlyLinear.py`

- `launchAngleOnlyLinear.py`

- `launchAngleOnlyPoly.py`

- `launchDistLinear.py`

- `launchDistPoly.py`

- `pullResultLogit.py`

- `hitDistanceAnalysis.py`