Best set of parameters for part1: learning rate = .15, max-iterations = 4000

We believe this is the best solution as the gradient descent is completely applied. We have tested increasing the parameters of learning rate, but that simply leads to a bottleneck at a higher MSE value and wastes efficiency. The same goes for having a large iteration count, as it leads to redundant trials.

Part 2:

We are not fully satisfied that the regression model fits our dataset. We tested both the linear and gradient descent regression models, and our R2 values never increased past .04. However, we are confident that the model had fitted the X and Y training data appropriately and had used the same model to predict Y\_pred. It is possible that this dataset is not suited for linear regression.

Included is a plot of Temperature vs Biker Count, aka our independent vs dependent variable.

A close up of a map

Description automatically generated