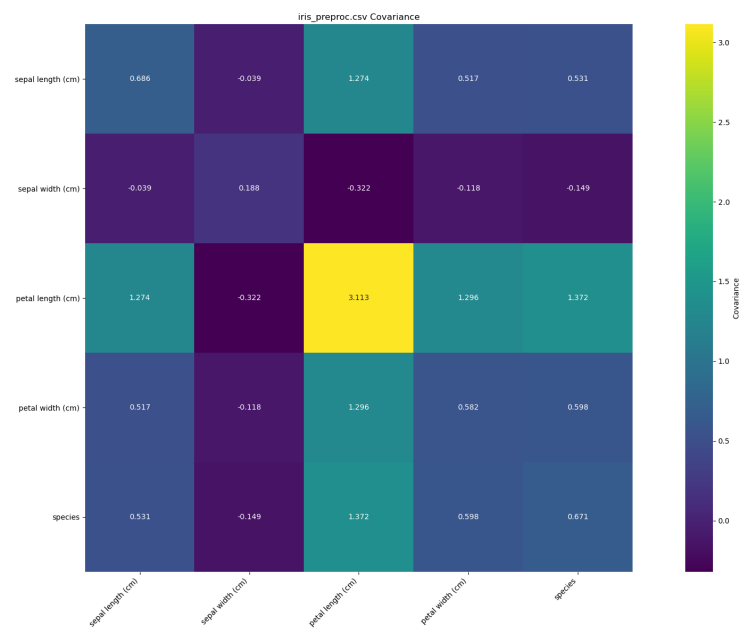


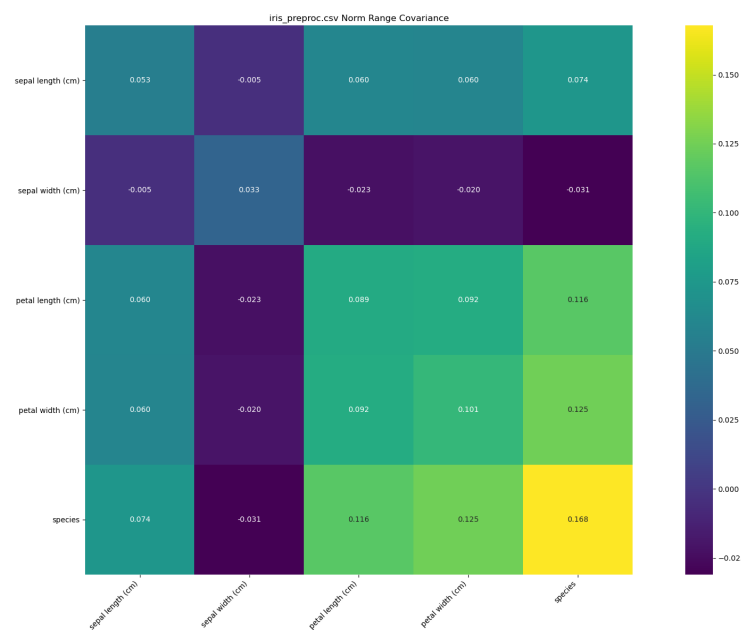
# Part 1: Iris Demo

## Covariance and Heatmaps

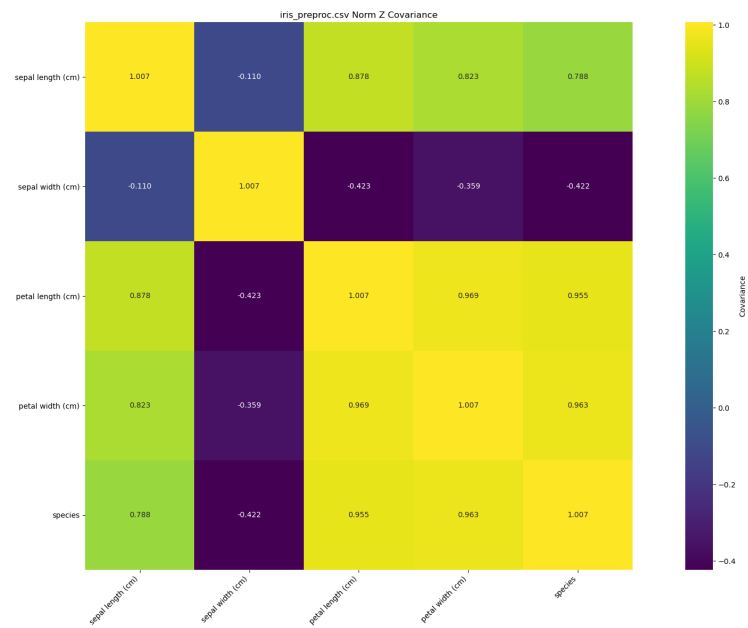
Raw Covariance:



Normalized by Range Covariance:

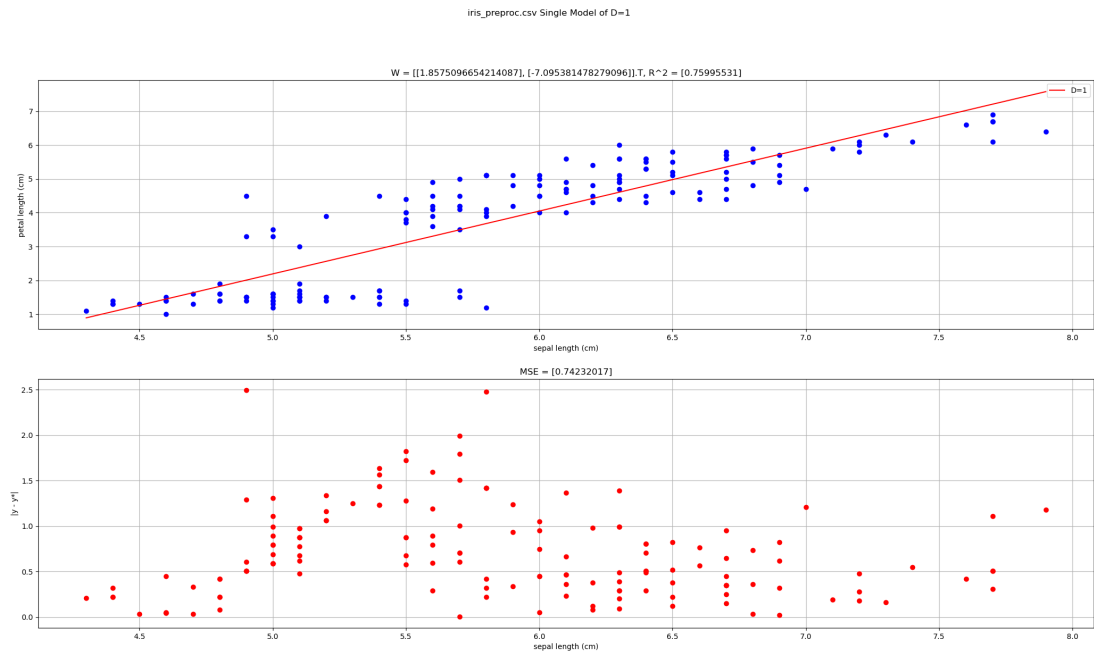


Normalized by Z-Score Covariance:

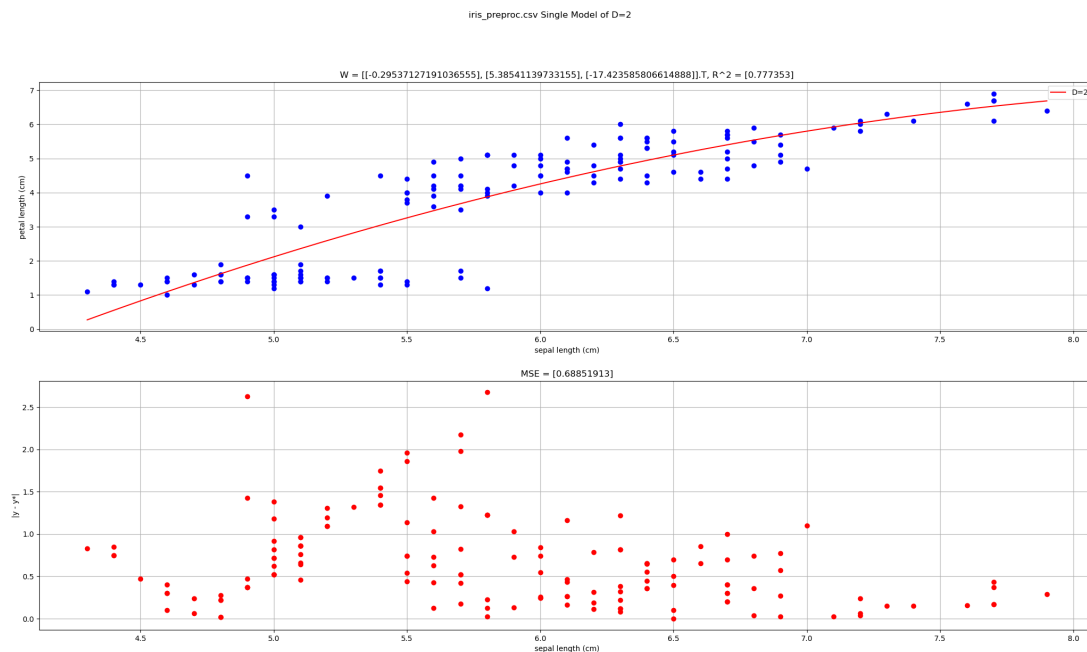


# Regression Models

Single Simple Model:



## Polynomial Single Model:



## Part II: Real-World Analysis

### 1. Describe your dataset

My dataset contains stats about pokemon. The following list of features is from the [kaggle page](#) I got the dataset from. They are not in order as they appear in the dataset.

- name: The English name of the Pokemon
- japanese\_name: The Original Japanese name of the Pokemon
- pokedex\_number: The entry number of the Pokemon in the National Pokedex
- percentage\_male: The percentage of the species that are male. Blank if the Pokemon is genderless.
- type1: The Primary Type of the Pokemon
- type2: The Secondary Type of the Pokemon
- classification: The Classification of the Pokemon as described by the Sun and Moon Pokedex
- height\_m: Height of the Pokemon in metres
- weight\_kg: The Weight of the Pokemon in kilograms
- capture\_rate: Capture Rate of the Pokemon
- base\_egg\_steps: The number of steps required to hatch an egg of the Pokemon
- abilities: A stringified list of abilities that the Pokemon is capable of having
- experience\_growth: The Experience Growth of the Pokemon
- base\_happiness: Base Happiness of the Pokemon
- against\_?: Eighteen features that denote the amount of damage taken against an attack of a particular type
- hp: The Base HP of the Pokemon
- attack: The Base Attack of the Pokemon
- defense: The Base Defense of the Pokemon
- sp\_attack: The Base Special Attack of the Pokemon
- sp\_defense: The Base Special Defense of the Pokemon

- speed: The Base Speed of the Pokemon
- generation: The numbered generation which the pokemon was first introduced
- is\_legendary: Denotes if the pokemon is legendary.

My target feature is capture rate.

## 2. Ask a question

My question is can you determine how hard a pokemon is to capture? I believe that regression analysis can help me answer this question because it is a question of whether there are associations between two variables. I want to be able to look at a certain feature of pokemon and be able to determine the capture\_rate of that pokemon and this is what regression analysis is for.

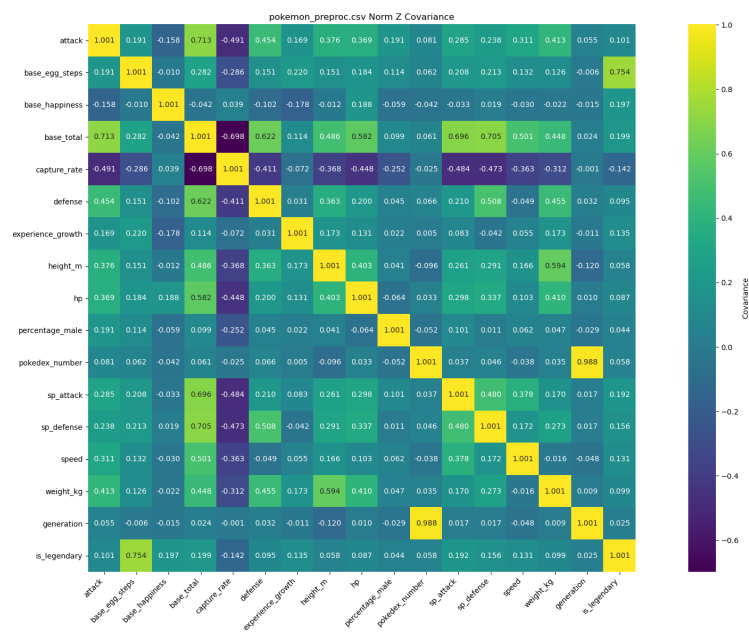
## 3. Preprocess

I did a bit of preprocessing before using this dataset. First I removed the japanese\_name feature because I was having some unicode issues. Since I'm using pandas the unicode characters aren't an issue with this project but just in case I got rid of them. I also removed the against\_? features because I was not interested in them and that's 18 less features to count the index of. Lastly I removed a single row because the capture\_rate value had non numeric characters in it which broke things. In the future I can probably find a way to handle that in the program in case there are multiple rows like that.

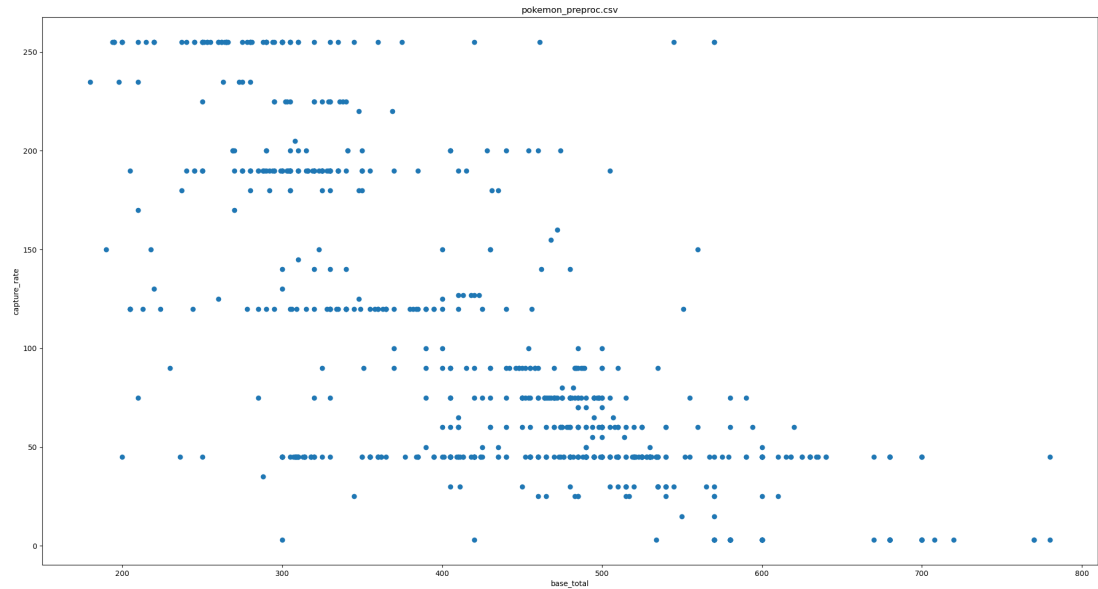
## 4. Identify the input feature

I believe that base\_total is the input feature that will help me most accurately predict capture\_rate. I identified this feature based on the covariance between it and capture\_rate and a scatterplot of the two features. base\_total had the highest covariance of all the preprocessed features. The covariance indicated a decent negative association and the scatterplot visually confirmed what the covariance suggested.

Pokemon Covariance:



Pokemon Scatterplot:

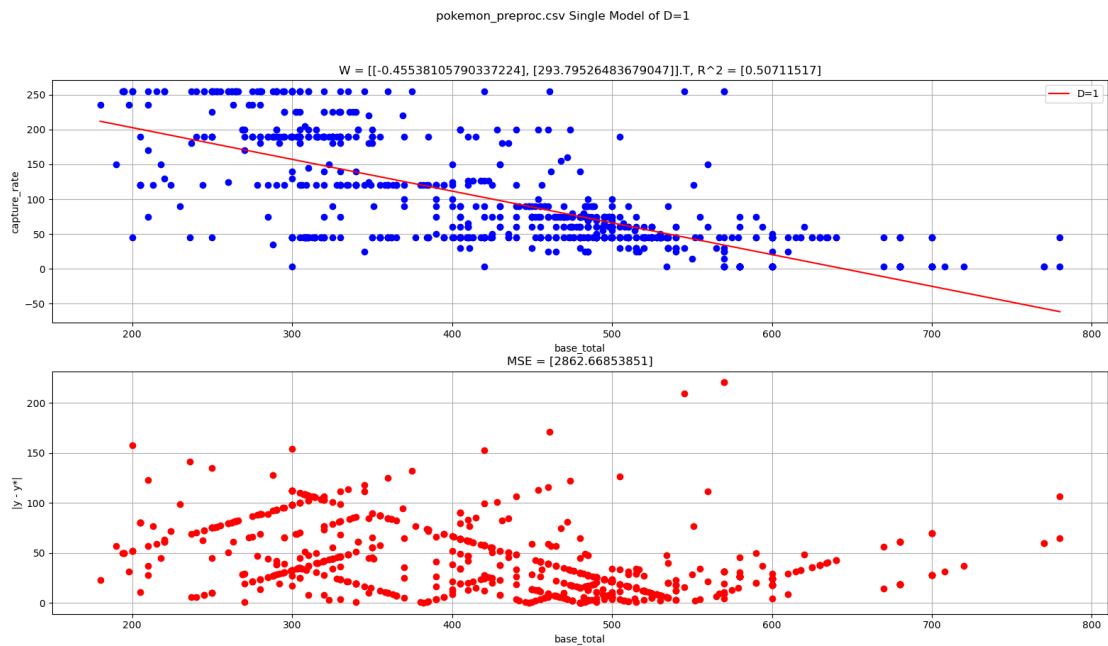


5. Simple single regression model

With the simple regression model there is an  $R^2$  of 0.507 and an MSE of 2863.

The first element of the weight matrix describes the slope of the trend of the data. Ours is -0.455 which indicates a gentle negative slope. This means that as our input variable increases the output variable tends to decrease. The second element of the weight matrix describes how the output variable behaves when the input variable is 0. This is not useful in our case because there are no input data values near 0.

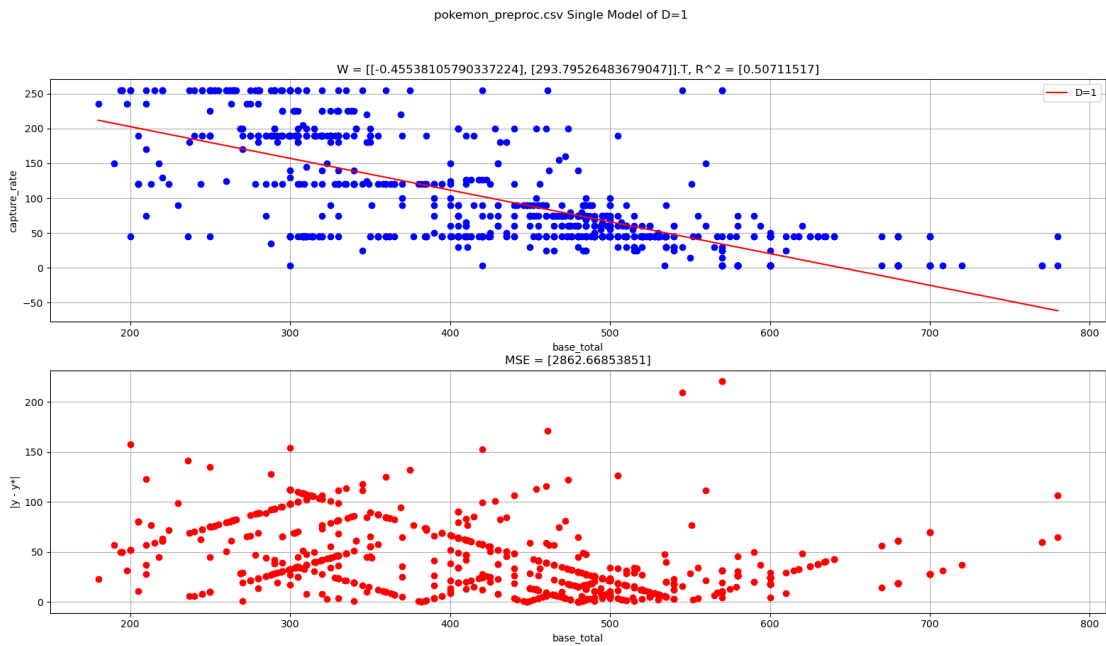
Pokemon Simple Single Regression:



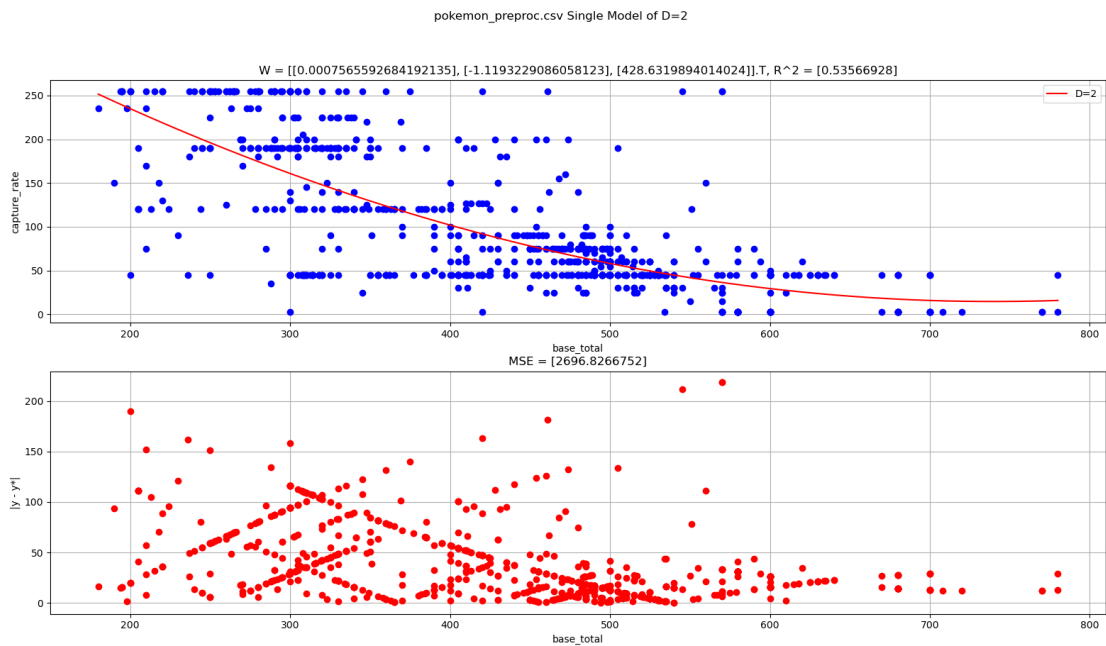
6. Polynomial single regression model

Pokemon Polynomial Single Regression (D=1 -> D=10):

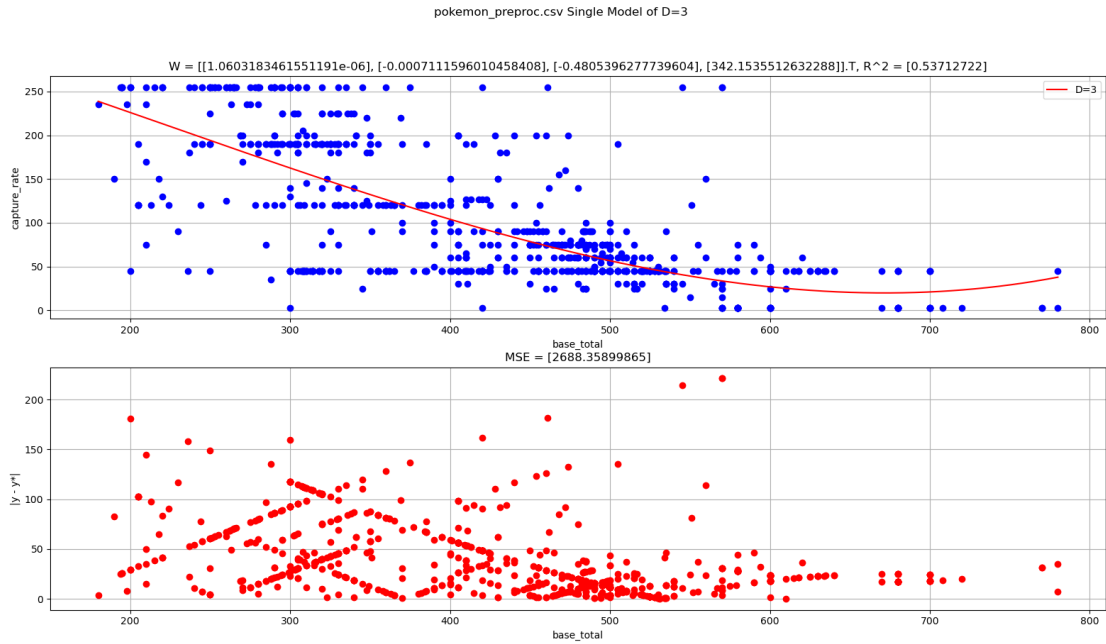
D: 1, RSq: [0.50711517], MSE: [2862.66853851]



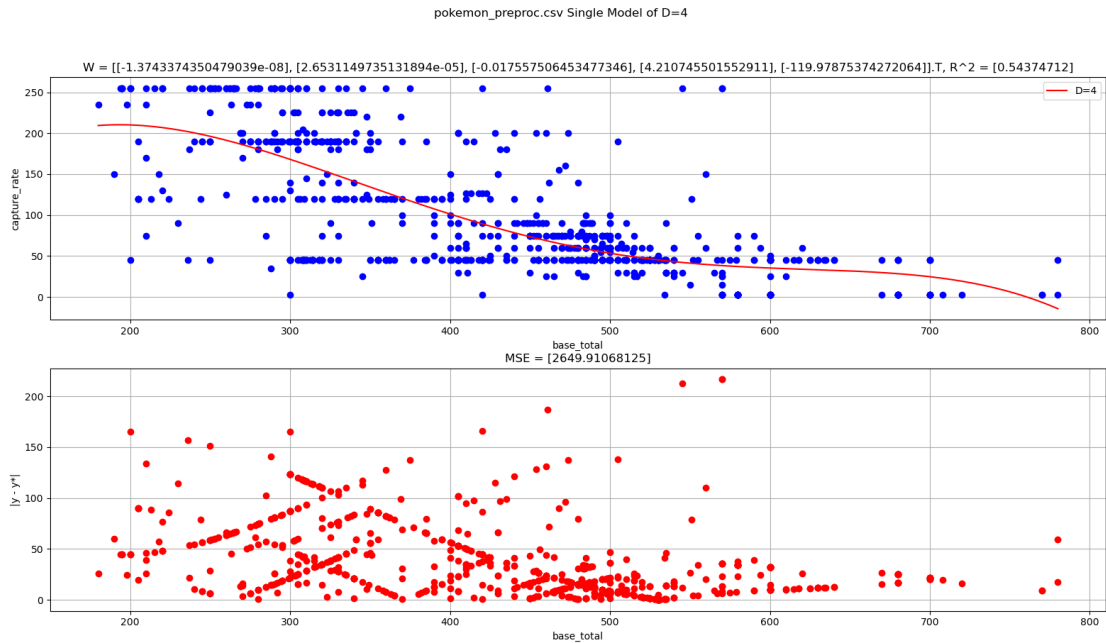
D: 2, RSq: [0.53566928], MSE: [2696.8266752]



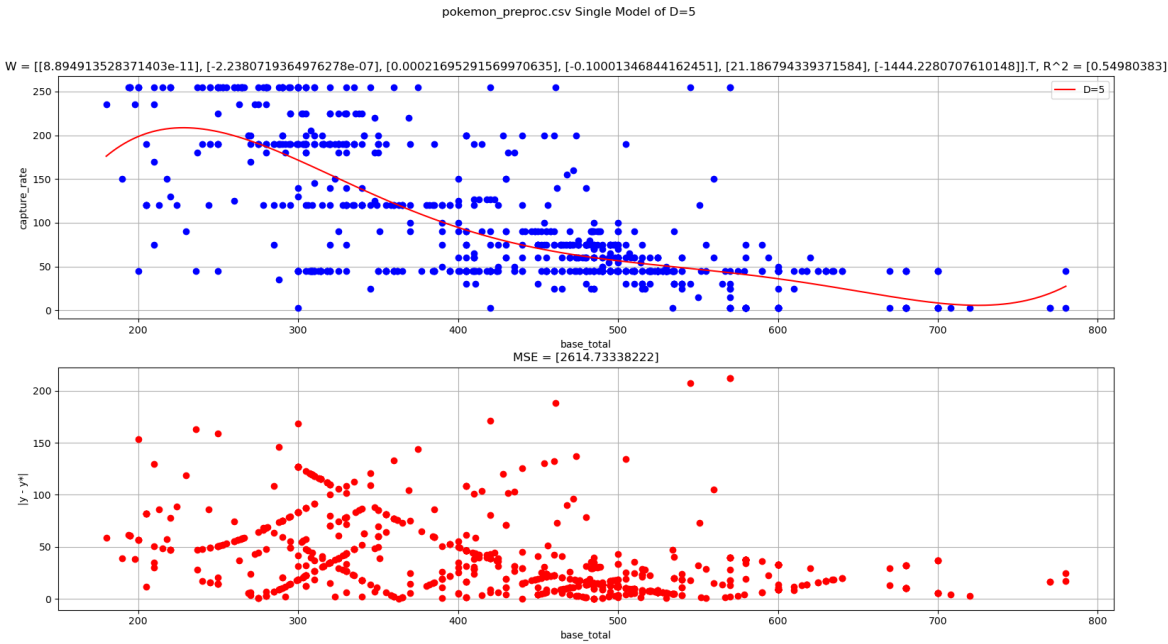
D: 3, RSq: [0.53712722], MSE: [2688.35899865]



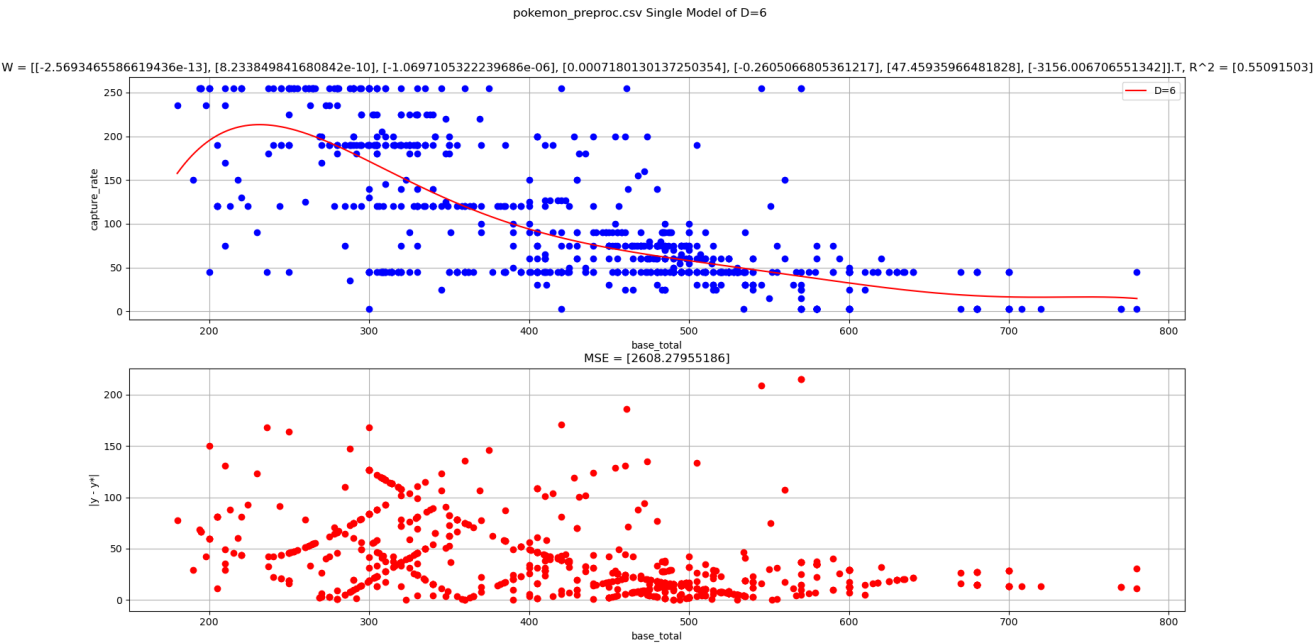
D: 4, RSq: [0.54374712], MSE: [2649.91068125]



D: 5, RSq: [0.54980383], MSE: [2614.73338222]

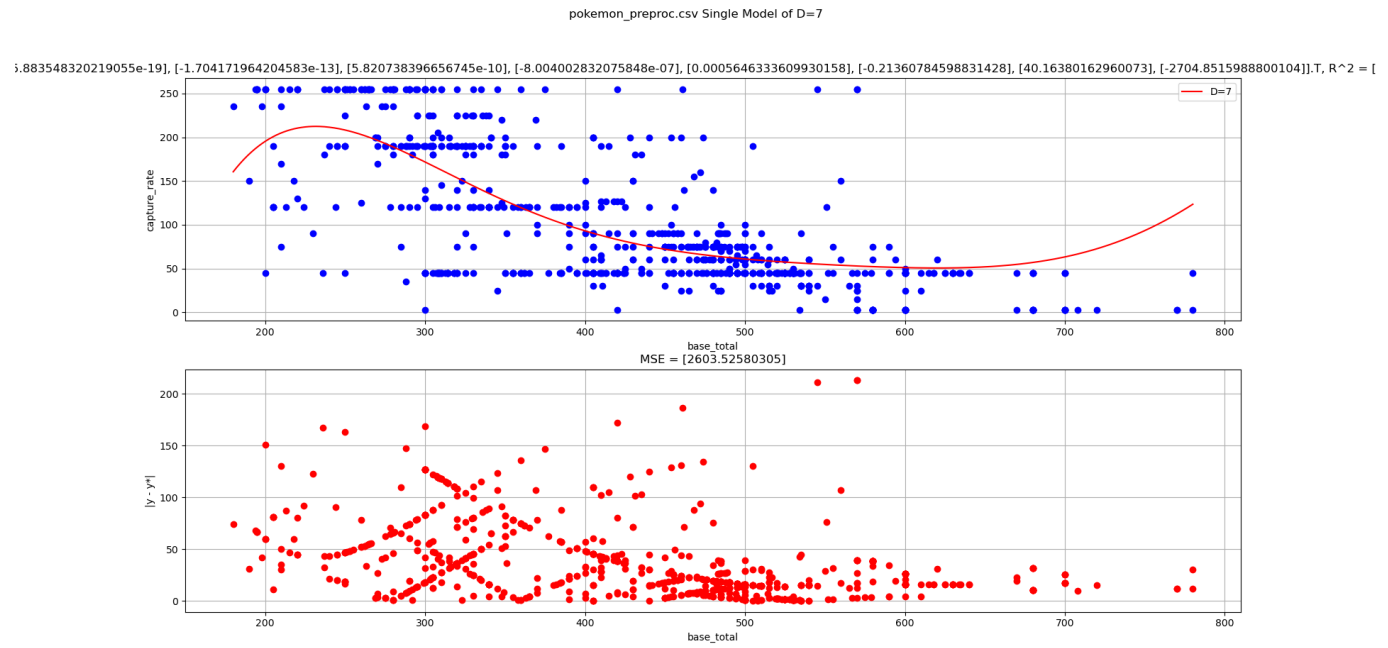


D: 6, RSq: [0.55091503], MSE: [2608.27955186]

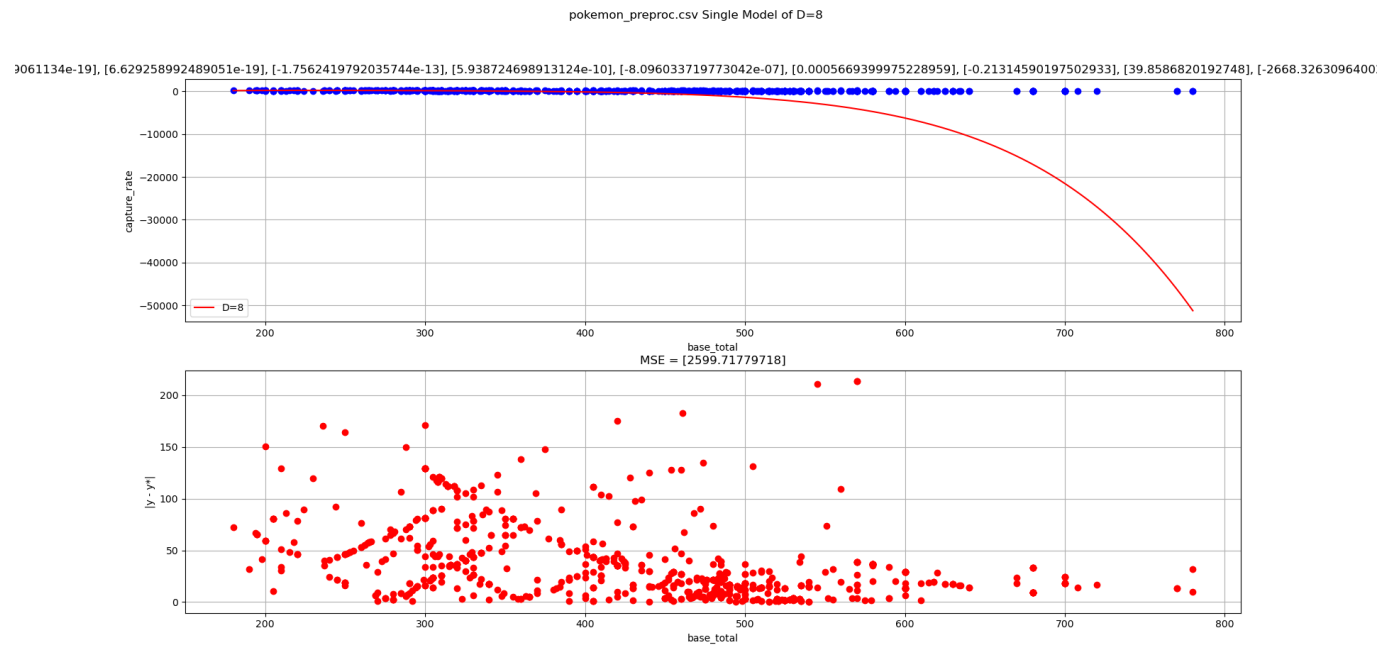




D: 7, RSq: [0.55173352], MSE: [2603.52580305]



D: 8, RSq: [0.55238917], MSE: [2599.71779718]



pokemon\_preproc.csv Single Model of D=9

