**KU Data Analytics Bootcamp -Trilogy/2U -Final Project**

**Presentation Date: 07/11/2020**

**Team: JuST Data Members: Jemi, Sanita, Tom**

**Alternate analysis methods to LinearRegression Modeling**

When we started looking into this project, we each looked at a couple different Machine Learning models. Although the LinearRegression model seemed best, we did explore others.

Initial examination of the SVM/SVR analysis Model looked like it would be a good fit as our pre-analysis showed some heavy outliers and SVR is supposed to be good for outlier understanding. However, after initial attempts, it looked like this model would require too much time to understand and it was abandoned.

Another method examined was the HubbleRegression model. This model employed CrossValidation and Kfold scoring and is known for its lower bias than other methods. This model appeared to have some good potential, and initially did provide additional information – it was hoped that we could compare this model’s output to that of the LinearRegression model. However, after further review with the TA’s and better understanding of the data, we decided to stay with the LinearRegression model and run multiple variable sets and compare the results.

Ultimately, after running the ML LinearRegression model and using 2, 4 & 6 independent variables we discovered that with the max X variable predictor set we maximized both our R2 and MSE to give us the best results.