



Rensselaer

why not change the world?®

Lab2: Linear Regression

Ahmed Eleish

February 7th, 2025

ITWS-4600/ITWS-6600/MATP-4450/CSCI-4960

Tetherless World Constellation
Rensselaer Polytechnic Institute



Lab 2

Files:

<https://rpi.box.com/s/oyl72mf3jw5z6t457rbyphjug083imih>



Exercise 1: linear models

- Your exercise: Examine the influence of various variables on property price
- Dataset: NY-House-Dataset.csv

Using the code in NYHousing_Linear_Regression.R do the following:

- Fit 3 linear models with Price as the response variable and with combinations of PropertySqFt, Beds, and Bath as predictors.
- Do any data cleaning (outliers, etc.) to get the best possible models. Remember that you may need to transform a variable for a better fit.
- For each model:
 - Print model summary stats.
 - Plot the most significant variable vs Price with the best fit line.
 - Plot a scatter plot of the residuals for the model.
- Compare the 3 models and decide which you believe is most useful. You don't need to tell me which one.

Please push to your github repository:

1. All your code in a *.R or *.MD file
2. Summary stats and select plots from 3 linear models in a pdf/word doc

Thanks!
Have a great weekend!