**The London School of Economics & Political Science**

**ST447 Data Analysis and Statistical Methods :**

**Final Project**

**Predicting The Best Driving Centre for Sarah**



**Akeel Ahamed**

**31st January 2018**

**Table of Contents**

To do:

Basic Plots in Data Analysis.

Justify why im using linear regression.

Strengths and weaknesses the model doesn’t address

Identify potential tweaks to model to improve adjusted R^2: See what to do with multicollinearity.

Explain ESS TSS and other similar vague terms.

Read data in xlsx

Use for loops

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

Assumptions:

* The experience level of individuals(both male and female between the ages of 17-25 are almost negligible).
* XYZ needs to take the test immediately(within next few days) and thus cannot wait until she is atleast year older.
* Places marked “Bury” and “Bury (Manchester)” are the same. Note that the places named “Bury” have been renamed to the former for ease of entry of formulae in R.
* Places marked “Wood Green” and “Wood Green (London)” are the same. Note that the places named “Wood Green” have been renamed to the former for ease of entry of formulae in R.

References:

* Introduction to Statistical Learning

<http://www-bcf.usc.edu/~gareth/ISL/>

* Using Linear Regression to Predict Energy Output of a Power Plant

<https://www.r-bloggers.com/using-linear-regression-to-predict-energy-output-of-a-power-plant/>

* Image credits : https://www.google.co.uk/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwiLsYrAw-vYAhUMLsAKHXyGC34QjRwIBw&url=https%3A%2F%2Fsurety1.com%2Fsurety-bond-driving-schools%2F&psig=AOvVaw0n\_\_Vet9kMT2cXiekl0BiH&ust=1516708969648038

Drawbacks:

* Outliers affect mean.