STAT420: Statistical Modeling in R
Summer 2023
Data Analysis Project Proposal

Team: Random85

1. The names of the students who will be contributing to the group project.

Ian Vetter - ivette2
Thomas Richter - thomasr8
Vivek Patel - vivekdp2

2. A tentative title for the project.

An examination of trends in Ames, Iowa housing prices

3. Description of the data file (what they contain including number of variables and number of records). You do not necessarily have to list all the variables, but at least mention those of greatest importance.

We'll be analyzing a dataset of residential home prices and attributes in Ames, Iowa. This dataset features a healthy combination of 79 numeric and categorical variables, with around 3000 unique training examples. For the response variable, we'll aim to model the numeric SalePrice feature as a function of some of the most notable predictors, including lot area, year built, utilities, # of full and half baths, etc.

4. Background information on the data sets, including specific citation of their source (so that I can also access it).

The Ames housing dataset was assembled by Dean De Cock, and featured in a recent Kaggle data analysis competition:

https://www.kaggle.com/datasets/prevek18/ames-housing-dataset

5. A brief statement of the business, science, research, or personal interest you have in the data set which you hope to explore.

One of our members, Tom, is looking to buy a home in the near future. Finding key indicators of housing price sounds fascinating and potentially very helpful!

6. Evidence that the data can be loaded into `R`. Load the data, and print the first few values of the response variable as evidence.

https://github.com/ianvet31/AmesHousing_Analysis/blob/main/Data%20Analysis%20-%20MD.pdf data_analysis_main.Rmd

Knit on Save | Knit

Source | Visual

```
---
title: "Data Analysis - Ames House Prices"
author: "Random85"
date: "2023-07-20"
output: html_document
---

# The Dataset


```{r}
housing_data = read.csv("./Dataset/AmesHousing.csv")
View(housing_data)
str(housing_data)
```
```

```
'data.frame':	2930 obs. of  82 variables:
 $ Order          : int  1 2 3 4 5 6 7 8 9 10 ...
 $ PID            : int  526301100 526350040 526351010 526353030 527105010 527105030 527127150 527145080 527146030 527162130 ...
 $ MS.SubClass    : int  20 20 20 20 60 60 120 120 120 60 ...
 $ MS.Zoning      : chr  "RL" "RH" "RL" "RL" ...
 $ Lot.Frontage   : int  141 80 81 93 74 78 41 43 39 60 ...
 $ Lot.Area       : int  31770 11622 14267 11160 13830 9978 4920 5005 5389 7500 ...
 $ Street         : chr  "Pave" "Pave" "Pave" "Pave" ...
 $ Alley          : chr  NA NA NA NA ...
 $ Lot.Shape      : chr  "IR1" "Reg" "IR1" "Reg" ...
 $ Land.Contour   : chr  "Lvl" "Lvl" "Lvl" "Lvl" ...
 $ Utilities      : chr  "AllPub" "AllPub" "AllPub" "AllPub" ...
 $ Lot.Config     : chr  "Corner" "Inside" "Corner" "Corner" ...
 $ Land.Slope     : chr  "Gtl" "Gtl" "Gtl" "Gtl" ...
 $ Neighborhood   : chr  "NAmes" "NAmes" "NAmes" "NAmes" ...
 $ Condition.1    : chr  "Norm" "Feedr" "Norm" "Norm" ...
 $ Condition.2    : chr  "Norm" "Norm" "Norm" "Norm" ...
 $ Bldg.Type      : chr  "1Fam" "1Fam" "1Fam" "1Fam" ...
 $ House.Style    : chr  "1Story" "1Story" "1Story" "1Story" ...
 $ Overall.Qual   : int  6 5 6 7 5 6 8 8 8 7 ...
 $ Overall.Cond   : int  5 6 6 5 5 6 5 5 5 5 ...
 $ Year.Built     : int  1960 1961 1958 1968 1997 1998 2001 1992 1995 1999 ...
 $ Year.Remod.Add : int  1960 1961 1958 1968 1998 1998 2001 1992 1996 1999 ...
 $ Roof.Style     : chr  "Hip" "Gable" "Hip" "Hip" ...
 $ Roof.Matl      : chr  "CompShg" "CompShg" "CompShg" "CompShg" ...
 $ Exterior.1st   : chr  "BrkFace" "VinylSd" "Wd Sdng" "BrkFace" ...
 $ Exterior.2nd   : chr  "Plywood" "VinylSd" "Wd Sdng" "BrkFace" ...
```