Stat 6021: Project 2 Proposal

Overview

The main goal of the proposal is for your group to come up with two questions that you would like to answer using the assigned data set. One question will be answered using linear regression, the other question will be answered using logistic regression. The proposal will enable me to assess whether what you are proposing is suitable for this class.

Deliverable

For this part of the project, you will submit a proposal as a PDF file. Failure to turn this in will result in a score of 0 for the entirety of Project 2.

Proposal

Your group's proposal should have the following sections:

1. Questions of Interest: Your group is to write **two** questions of interest that you would answer using your data set.
   (a) Write down your two questions of interest. These questions must be written without any technical jargon.
      - The first question would be answered using linear regression, so your response variable has to be quantitative.
      - The second question would be answered using logistic, so your response variable has to be binary.
   (b) In a separate paragraph, clearly specify the response variable for each question.
   (c) For each question, write a short paragraph to explain why the question is worth exploring. The more interesting, the better.
2. Data Visualizations:
   (a) Provide at least three data visualizations that provide some insight into each question of interest (so total six).
   (b) Provide some commentary for each data visualization and how they provide insight into each question of interest.

   Note: before creating the visualizations, be sure to split your data into training and test sets, using `set.seed(6021)`. The visualizations should be done only on the training set. The code below will perform this split.

```
Data<-read.csv("kc_house_data.csv", sep=",", header=TRUE)
set.seed(6021)
sample.data<-sample.int(nrow(Data), floor(.50*nrow(Data)), replace = F)
train<-Data[sample.data, ]
test<-Data[-sample.data, ]
```

Submission & Grading

- **One member of the group** will submit the Proposal as a **pdf document** via Canvas.
- If your does not submit a Proposal, your group will receive a score of 0 for the entirety of the project.

Feedback on the proposal may be provided. Groups will have the option of submitting one revised proposal (but this should be avoided, due to the tight time frame that we are working with).

Additional Guidelines

- The proposal should be no more than 8 pages.
- Also note that you have about 3 to 4 weeks to work on this project, so be realistic in your proposal.
- For the categorical response variable in logistic regression, be sure it is binary. We will not cover enough material to tackle categorical response variables with more than 2 classes.