Project 2 Info

Project Overview

This project will use a dataset of housing prices from King County, Washington, which includes the Seattle area. The dataset is a random sample of houses that sold in 2014 and 2015.

```
KingCo <- read_csv("https://raw.githubusercontent.com/STAT505/Project_2/main/KingCo.csv")
```

The dataset contains 6 measurements associated with 869 houses.

- bedrooms: number of bedrooms
- bathrooms: total number of bathrooms. Note: a full bathroom has toilet, sink, and shower or bathtub; a 1/2 bathroom generally has a toilet and sink; and a 1/4 bathroom has only a toilet.
- sqft_living: square footage of living space in the house
- sqft_living: square footage of the lot
- waterfront: binary indicator of whether a house has a waterfront lot

Your goal will be to build the best binary regression model to explain characteristics that lead to a million dollar house.

For this report, there is a maximum page limit of 5 pages (using the default RMD / PDF output settings). This will require judicious use of figures and concise writing. Note the page limit does not apply to the appendix with code.

The report should contain sections such as introduction, data overview + visualization, statistical procedures, results, and discussion. The following is a rubric that will be used to assess the report. All items are graded with the following scheme, where some have multipliers:

1. **No Credit**: Criterion was not addressed or was written in a way that was not understandable.
2. **Beginning**: Ideas are not clear and supporting ideas are not presented.
3. **Developing**: Ideas are identified but not well supported and developed or are minimally supported and developed.
4. **Advanced**: Ideas are clearly identified and are adequately supported and developed.

Report generalities	**Points**
Spelling, grammar, writing clarity, paragraphs & section labels	/12
Appendix with complete code / Citations and Acknowledgments	/4

Introduction	**Points**
Project motivation	/4
Research question	/4

Data Overview + Data Visualization	**Points**
Titles, Labels, and Captions	/4
Figure Clarity	/4

Data Overview + Data Visualization	**Points**
Figure Quality	/4

Statistical Procedures	**Points**
Define model to fit (with complete notation)	/8
Defense of model choice (using AIC or LOOIC/ELPD)	/4

Results	**Points**
State and assess model assumptions	/8
Summarize estimates from final model (with uncertainty intervals)	/4

Discussion	**Points**
Discuss Results in the context of the research question	/4
Scope of Inference: how can the results be generalized?	/4