**Term Project**

Matthew Litaker

Bellevue University

DSC 530-T301: Data Exploration and Analysis

Prof. Matthew Metzger

February 25, 2025

**Statistical/Hypothetical Question**

Which physical feature has the most significant effect on housing costs?

**Exploratory Data Analysis Outcome**

After observing the results from all models and graphs, I determined that “Area” (square footage of the house) had the most significant effect on housing costs. Though the permutation test comparing “Price” and “Area” showed a null value result produced by the simulation, I was able to draw the conclusion based on the actual correlation value between “Area” and “Price,” which had a greater value than all other variables tested. Additionally, the Cumulative Distribution Function supported this conclusion by showing that houses begin to rise in value as square footage becomes larger.

**What do you feel was missed during the analysis?**

I feel that running each model on each variable would have benefited the understanding of the analysis. I stayed in line with the criteria of the project; however, running additional models may have changed the outcome.

**Were there any variables you felt could have helped in the analysis?**

I believe having a variable that would have identified the location of the house would have benefited the analysis. I conducted the analysis assuming that all houses were in the same zip code; however, knowing exactly where each house was located could have affected the outcome or identification of outliers due to obvious price gaps based on the zoning of the house.

**Were there any assumptions made you felt were incorrect?**

Besides the location of the houses used in the dataset mentioned earlier, the dataset turned out to be straightforward, with well-identified variables. With this, I did not have to make another assumption and felt that my location assumption was accurate.

**What challenges did you face? What did you not fully understand?**

My greatest challenge was determining which model was best suited to test the hypothesis. After running a few different tests, I felt that the permutation test ended up fitting the question as well as the dataset the best. However, after conducting this project, I learned that understanding different models and how well they suit the analysis is an area in which I need to expand my knowledge.