Project 4 Proposal
Group Members: Arlene Wimbley, Zac Blankenship, Denise Long, Caitlyn Epley
April 22, 2024

Overview:
The aim of our project is to explore real estate data to predict home prices. We will examine aspects of homes, such as square footage, number of bedrooms, etc. to predict the value of other homes.

The main question we will ask of the data is what a home is worth based on its features. We will also analyze the data to determine which features are most important in determining home price in order to better predict the home's price.

The machine learning model we will attempt to use is a random forest regressor. If time permits, we will use other models, such as multivariate linear regression, support vector machine, etc. to compare to the neural network model. We will use techniques, such as a correlation matrix to pick out the most important features to use for the model.

Dataset(s):
- USA Real Estate Dataset:
  https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset/data

Timeline:
- Monday, April 22: decide on topic, create proposal, find dataset(s)
- Tuesday, April 23: data cleanup, create data profile
- Thursday, April 25: finish data cleanup, start ML
- Monday, April 29: ML, visualization, other requirements
- Tuesday, April 30: visualization, other requirements
- Thursday, May 2: finish up presentation, finish up Github repo, finalize code
- Monday, May 6: Presentation

Responsibilities:
- Arlene: data cleanup, SQL database, narrative/README
- Denise: analysis, research
- Zac: data cleanup, visualizations, slides, GitHub cleanup
- Caitlyn: machine learning models, correlation matrix, narrative/README