Team 12

Project name: "Location, location, location"

Objective: To quantify the impact of crime and school quality on home values in Minneapolis communities

Dependent variable: Average home price per square foot (Price_sqft) in each Minneapolis neighborhood (there are 11 geographically defined neighborhoods).
Null hypothesis: There is NO statistical difference in Price_sqft between neighborhoods.
Test hypothesis: Use ANOVA to test if any neighborhood has a statistically significant deviation from the mean. Price_sqft is continuous, neighborhood is categorical.
Outcome: If null hypothesis rejected, proceed to identify the causes (independent variables) of the differences in Price_sqft.

Independent variables to investigate:

1. **Home size (continuous)** – Does Price_sqft differ based on the size of the home? I.e. large homes command a higher Price_sqft than small homes. Do some neighborhoods have higher Price_sqft because the homes in that neighborhood are bigger?
2. **Crime rate (continuous)** – Does Price_sqft differ based on the crime rate of the neighborhood? Does high crime depress Price_sqft?
3. **Education (continuous)** – Does Price_sqft differ based on the rankings of the schools in the neighborhood. Does a good school increase Price_sqft?

Test 1: Test that the independent variables are not related to each other using pairwise linear regression on each combination and looking at R2. R2 should be low to include this variable as an independent variable.

Test 2: Test that the dependent variable is related to each independent variable using pairwise linear regression and looking at R2. R2 should be high to include a variable as an independent variable.

Test 3: Use multivariate linear regression to identify the impact of each independent variable on the dependent variable (coefficient). Depending on the outcomes of Test 1 and Test 2 this will be a multivariate linear regression with 1 dependent variable and 1-3 independent variables.