**Summary:**

* The Smart question I propose considers the relationship between dwelling neighborhood and dwelling prices. In particular, I propose we attempt to determine whether different DC neighborhoods are associated with different home values. As part of our analysis, we will consider the extent to which heterogeneity across average home values by neighborhood are associated with different home characteristics, vs. location or “neighborhood effects.” To do so, we can first determine whether housing characteristics (see below for examples) are significantly different between neighborhoods. Subsequently, we can construct a linear regression model to demonstrate the relationship between neighborhood and housing prices while controlling for other confounding variables.

**Main Smart Question:**

* Are dwellings in the DC metropolitan area priced significantly differently from one another based on quadrant?

**Secondary Smart Questions:**

* Do dwellings in different DC quadrant share similar characteristics? (Chi-Square tests for categorical variables and ANOVA for continuous)
  + Average Dwelling Age
  + Date of last refurbishment (Year, treated as continuous)
  + Number of Bedrooms
  + Type of Dwelling
  + Number of Baths
  + Square footage
  + Air Conditioning (categorical)
  + Stories
  + Kitchens
  + Fireplaces
  + Condition (categorical)
  + Land Area of Property
  + Dwelling Type (cateogorical)
* How do each of the above characteristics relate to the overall sale price of a home in DC? (Linear Regression)
  + Dependent Variable – Dwelling Price
  + Independent Variables considered in above analysis
  + Distance from nearest Metro station?
  + Distance from Center City (US Capital Building)
  + Color line of nearest metro station

**Methodology:**

For our research, we can utilize chi-square, t-tests, and or ANOVA tests to determine whether the above housing characteristics differ across neighborhoods. After determining whether individual dwelling characteristics are significantly different across neighborhoods, we can fit a linear regression model to determine the association between individual neighborhoods and home values.

**Next Steps:**

* Determine the limit/scope of our analysis. Start to understand our data
  + Do we include residential and condominiums? What are the difference between the two? (complete)
  + What do all of the variables that are not readily intuitive mean? For example, USECODE? (complete)
* Clean Data
  + Clear duplicated or missing values (complete)
  + Create an appropriate methodology to identify and, if needed, remove outlier observations
* Analyze differences in dwelling characteristics between neighborhoods
  + How can Chi-Square Tests be used?
  + How can ANOVA be used?
  + T-Tests?
* Fit a regression model to the data, including neighborhood fixed effects and controls for house characteristics
  + Dependent Variable – House price
  + Independent Variable – house characteristics
  + Fixed Effects – Quadrant dummies
* Report Findings
  + R-markdown Report
  + GIS visualization? (In Progress)
  + PowerPoint presentation