**DATA WRANGLING**

**Housing Data:**

The housing data is formally describe as “The Ames Housing Dataset” and was originally compiled by Dean De Cock. The objective was to build a data frame table consisting of major homes’ characteristics and their sale price.

The steps taken for these consisted of:

1. Downloading 2 different files:
   1. The major aspects and characteristics of each house sold between the years 2006 and 2010.
   2. The price for which each house was sold for.
2. Building each file separately into a panda’s data frame.
3. Merging these two data frames into a new table and filter columns to major homes’ characteristics and sale price, these included:
   1. “Bedrooms”, “Full Bathrooms”, “Lot Area Square Feet”, “Living Area Square Feet”, “Month Sold”, “Year Sold”, and “Sales Price”.
4. Searching for bad or negatives values:
5. There was not any “Not a number value(NaN)”
6. Searching for potential outliers:
   1. Built graphs and scatterplots to check for potential outliers:
      1. There were 3 properties with 0 values for bedrooms and bathrooms, the values were dropped, seemed to be typos and made not much sense to relation to sale price.
   2. Column “(GrLivArea): Living Area Square Feet” was filter to values under 5,000 Square Feet, after a data point was found to be highly isolated.
   3. For the most part the data showed normality.

**WEATHER DATA:**

The weather data was obtain from The National Center for Environmental Information(NOAA), the objective was to pull a data frame containing Average Precipitation, Average Maximum Temperature and Average Minimum Temperature on a Year- month basis from city Ames, Iowa.

The steps taken for this included:

1. Downloading data and opened to a data frame,
2. Subtracting needed rows and columns.
3. Building dates to monthly frequencies
4. Constructing negative or bad values such as (-9999) to Not a Number (NaN) category.
   1. These values represented missing data or data that was not yet received.
5. Building into a new data frame which included monthly averages for Precipitation, Maximum and Minimum Temperature.

Finally, after wrangling and cleaning the housing and weather data, these two were joined to one single data frame table.

**SUMMARY**

On a last brief summary, each dataset was separately downloaded, built into a data frame and wrangled as needed, at the end the two cleaned data sets were jointed to a single data frame table to be user for analysis.