Module 1 - Housing Project

To keep the workflow clear this project was split into two notebooks:

1. Preprocessing
2. Analysis

Preprocessing

The purpose of the preprocessing stage is to ensure that a fully augmented and cleaned dataset is prepared before the start of the regression analysis stage. The aim is to ensure that all data cleaning takes place upstream.

Analysis Stage

The Analysis stage takes the cleaned and augmented data set as the only input to the process.
In the analysis phase data visualisations, correlation matrices, maps, univariate regressions and multivariate regressions are used to get a feel for the dataset.

There are two aims:

1) Describe the behaviour of the Seattle Housing market over the 2014-2015 period and the relationship between property features and prices.
2) Predict the behaviour of the market going forward

The notebooks are structured as follows:

(Preprocessing)

1. Import the libraries required in the analysis / active workbooks (Pandas, NumPy, Matplotlib, Seaborn, maths, statsmodel).
2. Import Raw Data file KC Seattle Housing Data Set containing descriptive variables of houses sold in 2014-2015 and sales prices.
3. Creating New Columns for new or transformed.
4. Data Familiarisation, Cleansing and Treatment.
5. Removal of Duplicates.
6. Standardisation.
7. Save down a cleaned data CSV to be used in Regression Analysis sheet.

(Analysis)

1. Import python libraries and the cleaned data .csv file.
2. Plot a correlation matrix.

3. Linear regression.
4. Multi-variate regression.
5. Data visualisation.
6. Outlier treatment.
7. Zipcode analysis.