Machine Learning Class Task Two:

Category: *Data Cleaning*

Sub-Category: *Feature Engineering, Dimensionality Reduction*

Problem Definition:

You are provided with the **Housing Pricing** dataset.

  i.   Prepare a dataset description file
 ii.   Prepare the dataset for training
iii.   Fill in Missing Values
 iv.   Encode the dataset
  v.   Standardize the dataset
 vi.   Perfom feature extraction using PCA
vii.   Perform feature selection using L1
viii.   Train a regression model using PCA with 2 compnents

Some Important References:

  a. Pandas Data Frames Manipulations:
      http://pytolearn.csd.auth.gr/b4-pandas/40/moddfcols.html

  b. OHE using pandas:

      http://queirozf.com/entries/one-hot-encoding-a-feature-on-a-pandas-dataframe-an-example

  c. Scikit learn Library: data preprocessing –

      https://scikit-learn.org/stable/modules/preprocessing.html#encoding-categorical-features

  d. Regression using sklearn:

      https://scikit-learn.org/stable/modules/linear_model.html