

## **Machine Learning Class Task Two:**

**Category:** *Feature Selection*

**Sub-Category:** *Embedded L1 Numerical FS with One Hot Encoding*

**Alternative Models:** *Filters, Wrappers, with Binary Encoding or Backward Difference Encoding*

### ***Problem Definition:***

You are provided with the **Nairobi Office Price Simulated Ex** dataset.

- i. Prepare a dataset description file
- ii. Prepare the dataset for training
- iii. Identified the features to be used for training using L1 Embedding

### **Some Important References:**

- a. Pandas Data Frames Manipulations:

<http://pytlearn.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](https://scikit-learn.org/stable/modules/linear_model.html)

