**Taipei Real Estate Prediction (ML part)**

**Procedure:**

1. Scan through independent variables and try to have some ideas of which variables might be important to the dependent variable.
2. Check the distribution of dependent variable using density plot and its correlation with independent variables using heatmap.
   * If possible, try to construct a pairplot to check the relationship among variables
3. Check if missing values exist
   * Missing value exist:
     1. If more than a threshold (i.e 15%) of data is missing, we should delete the corresponding variable.
4. Feature engineering (most time consuming)
   * Create/delete variables
   * See if variables contain similar categorical values
5. Outliers removal
6. Check assumptions on dependent variable (stat big four – normality, QQplot, residual plot, outlier plot)
   * If assumption fails, try to apply different transformations. For example, log transformation.
7. Model fitting
   * Export trained model in pickle format
   * Export other needed pickle files, such as locations
8. Built website using flask
   * Html in templates folder
   * Flask app
   * Background image in static folder
9. Deploy on Heroku
   * Need to upload all the code and requirements on github