Capstone Proposal:

* PROBLEM:
  + How is it possible to predict the real estate value in DC?
* PURPOSE:
  + This knowledge can be used to advise real estate developers and investors.
  + This can also be used by the government and NGOs to problem-solve issues arising around high cost-of-living and segregation in the city.
* DATA DESCRIPTION:
  + I found the data from a Kaggle competition, D. C. Residential Properties. It is provided by Open Data DC.
  + https://www.kaggle.com/christophercorrea/dc-residential-properties
* STAGES:
  + Clean the data
    - Evaluate missing data. Remove or replace with meaningful values.
    - Reorganize, delete, or rename features as needed
  + Exploratory data analysis
    - Outlier evaluation and removal if needed
    - Ranges, measures of center
    - Distributions
  + Machine learning applications
    - Explore various predictive models to find the best solution
      * Linear regression
      * Logisitic regression
      * Clustering
      * Random Forest
  + Presentation
* DELIVERABLES:
  + Jupyter Notebook