Predictive model : Residence sale price in District of Columbia region.

Author : Pavan Kumar Kulkarni

CAPSTONE MILESTONE REPORT.

1. Background and problem statement.

Computer Assisted Mass Appraisal (CAMA) database. The dataset contains attribution on housing characteristics for residential properties, and was created as part of the DC Geographic Information System (DC GIS) for the DC Office of the Chief Technology Officer (OCTO) and participating D.C. government agencies.

Is it possible to build machine learning model which can predict the residential sale price using the attributes made available to public by DC government via these datasets?

2. Potential clients.

Anyone who is interested in residential property sale can use the model built. This includes

- **Residential agents**. - To estimate the sale price to get maximum value for their clients.
- **Homeowners** - Who are looking to sell residential properties. Getting know estimated sale price will help get maximum price and timely sale.
- **Home buyers** - Who are in market looking for residences. Estimated sale price based house attributes will help in zeroing on reasonable budget and expected residence attributes.
- **Public offices/Government agencies** -
  who need residential properties price for planning and making public policies.

3. Datasets and approach.

The Dataset is made available to public by DC government.
http://opendata.dc.gov/datasets/computer-assisted-mass-appraisal-residential.

Approach:
Explore the data to understand distribution, outliers, redundancies. Data wrangling to clean the data and make it ready for machine learning. Apply various regression models, use hyper parameter tunings and choose best model/s.

Tools/technologies used.
R language
R Markdown
Various packages available in R.