**DATS-6103-11**

**Spring 2023**

**Topic Proposal**

**Team 2:** Muhannad Alwhebie, Brian Gulko, Mengfei Hung, Kashyap Nimmagadda

**Topic:** DC Residential Property Sales

Our data comes from Open Data DC and describes the sale history for active properties listed among the District of Columbia’s real property tax assessment roll. The dataset contains about 108,000 rows and 39 columns describing property attributes such as area and number of bedrooms as well as sale information such as sale price and date. Our goal is to use analysis and models to better understand the relationship between these attributes and the effects they have on sale price.

**Attributes:**

SSL, BATHRM, HF\_BATHRM, HEAT, HEAT\_D, AC, NUM\_UNITS, ROOMS, BEDRM, AYB, YR\_RMDL, EYB, STORIES, SALEDATE, PRICE, QUALIFIED, SALE\_NUM, GBA, BLDG\_NUM, STYLE, STYLE\_D, STRUCT, STRUCT\_D, GRADE, GRADE\_D, CNDTN, CNDTN\_D, EXTWALL, EXTWALL\_D, ROOF, ROOF\_D, INTWALL, INTWALL\_D, KITCHENS, FIREPLACES, USECODE, LANDAREA, GIS\_LAST\_MOD\_DTTM, OBJECTID

**SMART Questions:**

1. What are the characteristics of an average residential property?

* Which heating type is the most common in residential properties in this dataset, and what is the percentage of properties with this heating type?
* What is the average number of rooms/bathrooms and half-bathrooms/kitchens in residential properties in this dataset?
* What is the average land area / gross building area of residential properties in this dataset, and how does this vary by number of bedrooms? How has it changed over time?
* What is the average GRADE of residential properties in this dataset

1. Which variables have an impact on sale price, and how strong is that impact?

We select the variables that have a high correlation with the sale price of a residential property to build the linear model of residential housing sale prices.

* Is there a correlation between the heating type and the sale price?
* Is there a correlation between the number of rooms/bedrooms/kitchens and the sale price of a residential property?
* Is there a correlation between land area / gross building area and sale price?
* Is there a correlation between GRADE of residential properties and sale price?

1. Did COVID-19 have an impact on residential sale prices? If so how big was that impact?

We believe these questions can provide valuable insights into the characteristics and trends of residential properties in our dataset.

**Models:** Linear regression, Multiple linear regression, K-Means clustering, KNN

**Dataset Source:** This data comes from Open Data DC’s Computer Assisted Mass Appraisal - Residential dataset which can be found at this link:

<https://opendata.dc.gov/datasets/DCGIS::computer-assisted-mass-appraisal-residential/explore>

Attribute information and descriptions can be found at this link:

<https://www.arcgis.com/sharing/rest/content/items/c5fb3fbe4c694a59a6eef7bf5f8bc49a/info/metadata/metadata.xml?format=default&output=html>

**GitHub Repo:** <https://github.com/kashyapnimmagadda/DATS-6103-TEAM2.git>