# Practical 1

##### Feature Selection (Duration: 40mins)

A property agent who specializes in the sale of resale flats wants to determine the significant factors contributing to the price of the resale flats in the last 10 years. Assuming all other external factors such as macro-economic factors are constant across the years, the property agent would like to identify factors that will affect how resale prices could change in the future based on the information within the dataset provided.

The **RESALE** data set contains 18 variables and over 18,000 observations. The variables in the data set are shown below with the appropriate roles and levels:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Measurement Level** | **Description** |
| Full Address | Nominal | Full address of resale flat sold |
| Story Range | Nominal | Resale flat’s floor range |
| Flat Model | Nominal | Resale flat model |
| Flat Type | Nominal | Resale flat type |
| Town | Nominal | Town |
| Town (group) | Nominal | Town (group) |
| Floor Area Sqm | Numeric | Floor Area Sqm |
| Latitude | Numeric | Latitude |
| Longitude | Numeric | Longitude |
| Lease Commence Date | Numeric | Lease Commence Date |
| Year-Month Sold | Numeric | Year-Month Sold |
| Nearest Hawker | Nominal | Nearest Hawker |
| Nearest MRT | Nominal | Nearest MRT |
| Nearest Primary School | Nominal | Nearest Primary School |
| Hawker Dist | Numeric | Distance between flat and nearest hawker |
| Mrt Dist | Numeric | Distance between flat and nearest MRT |
| School Dist | Numeric | Distance between flat and nearest school |
| Resale Price | Numeric | Resale Price (Target) |

Suppose your target variable is Resale Price, which variables will you select as features for the model?

1. Based on logic, which variables will you select?
2. Based on correlation analysis, which variables will you include or exclude?

\*\*Note for part b), only numerical variables can be for the correlation analysis.

Discuss you answer for part a in groups of 3-4, then share your conclusion.

For part b, go to the following URL: <https://sp2016-dev-ad.sit.nyp.edu.sg/SASLogon/login>

1. Download the “Resale.csv” file from BrightSpace.
2. Login using your lab login ID and password.
3. Follow the Step-by-Step Video in BrightSpace on how to create a correlation matrix in SAS Viya.
4. Discuss with your team on the variables to select.
5. Are there input variables with multicollinearity issue?

~~ End ~~