Real Estate House Price Prediction – Feature Engineering.

The Real Estate data contains the following features,

**Transaction date, House age, Distance to the nearest MRT station, Number of convenience stores, Latitude, Longitude.**

With target column **House price of unit area.**

**Drop Unnecessary Columns.**

Removed **Transaction date** column.

Feature columns: **5**

Target columns: **1**

**Real Estate data.**

Feature columns: **6**

Target columns: **1**

**Handling Missing values and Infinite values.**

* Filled missing values with **median.**
* Replaced infinite values with **NAN** then replaced with **median**.

**Derived Interaction Features.**

Created features,

* **Price per convenience store.**
* **Price per distance.**
* **Age impact**.

Feature columns: **8**

Target columns: **1**

**Handling Outliers.**

Removed outliers from

* **Distance to the nearest MRT station.**
* **Price per convenience store.**
* **Price per distance.**
* **Age impact.**

**Feature transformation.**

Applied **log transformation** to,

* **Distance to the nearest MRT station.**
* **Price per distance.**

**Feature Scaling.**

Used **RobustScaler** to scale both features and target variables.

**Feature Selection**.

Selected important features for modeling,

* **House age**
* **Distance to nearest MRT station**
* **Number of convenience store**
* **Latitude**
* **Longitude**
* **Price per convenience store**
* **Price per distance**
* **Age impact.**

Feature columns: **8**

Target columns: **1**