

### 3.4 Exercise do it yourself:

*Download Source Data Set from GitHub link* : <https://github.com/Smartbrain2024/Mastering_AI_2.git>

**Chapters/Chp\_03/3.4/DIY\_Exercise/Melbourne\_housing\_FULL.csv**

**Context:**

The original dataset contains price of different properties in Melbourne Housing Market and distinctive characteristics of those properties. Here we have considered the housing price data on a particular date, so that we can analyze how the different attributes of the properties relate to their price on a particular day. Attributes describing the location of the property, the type of property, the method by which it was sold, the price at which it was sold, etc. . are present.

**Problem Statement:**

The objective of this case study is to analyze the data which will help us draw insights and identify factors that affect the price of property.

#### Learning Summary:

* Visualize the dataset, take a feel of the different attributes present.
* Use the ‘Remove Duplicate Rows’ module to remove the duplicate rows.
* Use the ‘Edit metadata’ module to convert the categorical type variables.
* Perform univariate & Bivariate Analysis in the visualization window using different plots.
* Treat the missing values using the ‘Clean Missing Data’ module.
* Treat Outliers using the ‘Clip Values’ module.
* Use the ‘Threshold Filter’ & ‘Apply Filter’ modules for cleaning anomalous data.
* Apply the ‘Math Operation’ module for any feature engineering.
* Perform One Hot Encoding using the ‘Convert to Indicator Values’ module.

|  |  |
| --- | --- |
| Attribute Name | Description |
| Suburb | Suburb where property is located |
| Address | Address of the property |
| Rooms | Number of rooms |
| Type | h - house, cottage, villa, semi, terrace; u - unit, duplex;  t - townhouse; |
| SellerG | Real Estate Agent |
| Date | Date sold of property |
| Distance | Distance from CBD (Melbourne city center) in Kilometers. |
| ParkingArea | Parking Area Type |
| Attribute Name | Description |
| Method | S - property sold  SP - property sold prior PI - property passed in  PN - sold prior not disclosed SN - sold not disclosed.  VB - vendor bid.  W - withdrawn prior to auction SA - sold after auction  SS - sold after auction price not disclosed |
| Postcode | Post Code |
| Regionname | General Region (West, Northwest, North, Northeast …etc. .) |
| Propertycount | Number of properties that exist in the suburb. |
| Price | Price in Australian dollars |