**Melbourne Housing Snapshot**

**Dataset source:** <https://www.kaggle.com/datasets/dansbecker/melbourne-housing-snapshot/data>

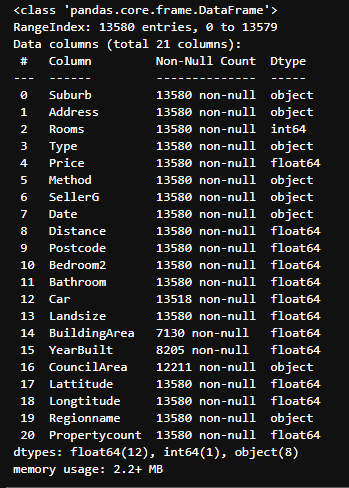
**Done By:** Ahmed Tarfa 27/9/2024

**About Data:**  
The dataset contains sales records of a real estate agency. It was scraped from publicly available results posted weekly on Domain.com.au.

**Dataset Structure:**  
The dataset includes the following columns:

* **Suburb:** The suburb where the property is located.
* **Address:** The street address of the property.
* **Rooms:** Number of rooms in the property.
* **Type:** Property type (e.g.,  br - bedroom(s); h - house,cottage,villa, semi,terrace; u - unit, duplex; t - townhouse; dev site - development site; o res - other residential.).
* **Price:** Sale price of the property by dollar.
* **Method:** Sale method (e.g., S - property sold; SP - property sold prior; PI - property passed in; PN - sold prior not disclosed; SN - sold not disclosed; NB - no bid; VB - vendor bid; W - withdrawn prior to auction; SA - sold after auction; SS - sold after auction price not disclosed. N/A - price or highest bid not available).
* **SellerG:** The seller's agency.
* **Date:** Sale date.
* **Distance:** Distance from the central business district (CBD).
* **Postcode:** Property's postcode.
* **Bedroom2:** Number of bedrooms in the property.
* **Bathroom:** Number of bathrooms in the property.
* **Car:** Number of car spaces available.
* **Landsize:** Land size of the property in square meters.
* **BuildingArea:** Building area of the property in square meters.
* **YearBuilt:** Year when the property was built.
* **CouncilArea:** Governing council for the area.
* **Lattitude:** Latitude coordinates of the property.
* **Longtitude:** Longitude coordinates of the property.
* **Regionname:** The region name, typically representing the broader area or city.
* **Propertycount:** The total number of properties in the suburb.

**Data Types:**



1. Date change to date time with format “y-m-d”
2. Car change to integer
3. Bedroom2 change to integer
4. Bathroom change to integer
5. Postcode change to integer
6. Year of building change to integer
7. Property count change to integer

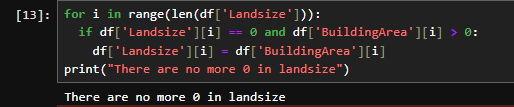
**Nulls:**



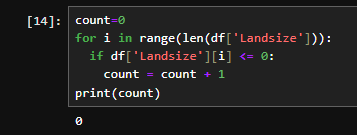
1. Car fill with 0
2. Building Area fill with mean
3. Year of build fill with mean
4. Council area fill with mode

**Other problems:**

1-Land size equals to zero but building size more than 0 , so I had make the land size equals to building size

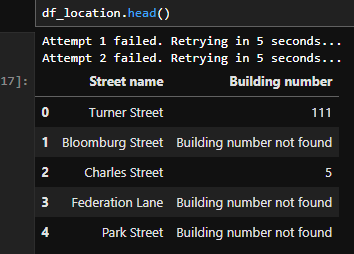
 

CHECK:



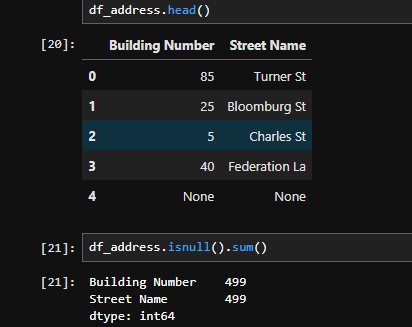
2-Address check using Lattitude Longtitude

Df\_location was the address of Lattitude Longtitude only



Take the address column from df and split it in new df\_address as two columns

Building number , Street and check nulls

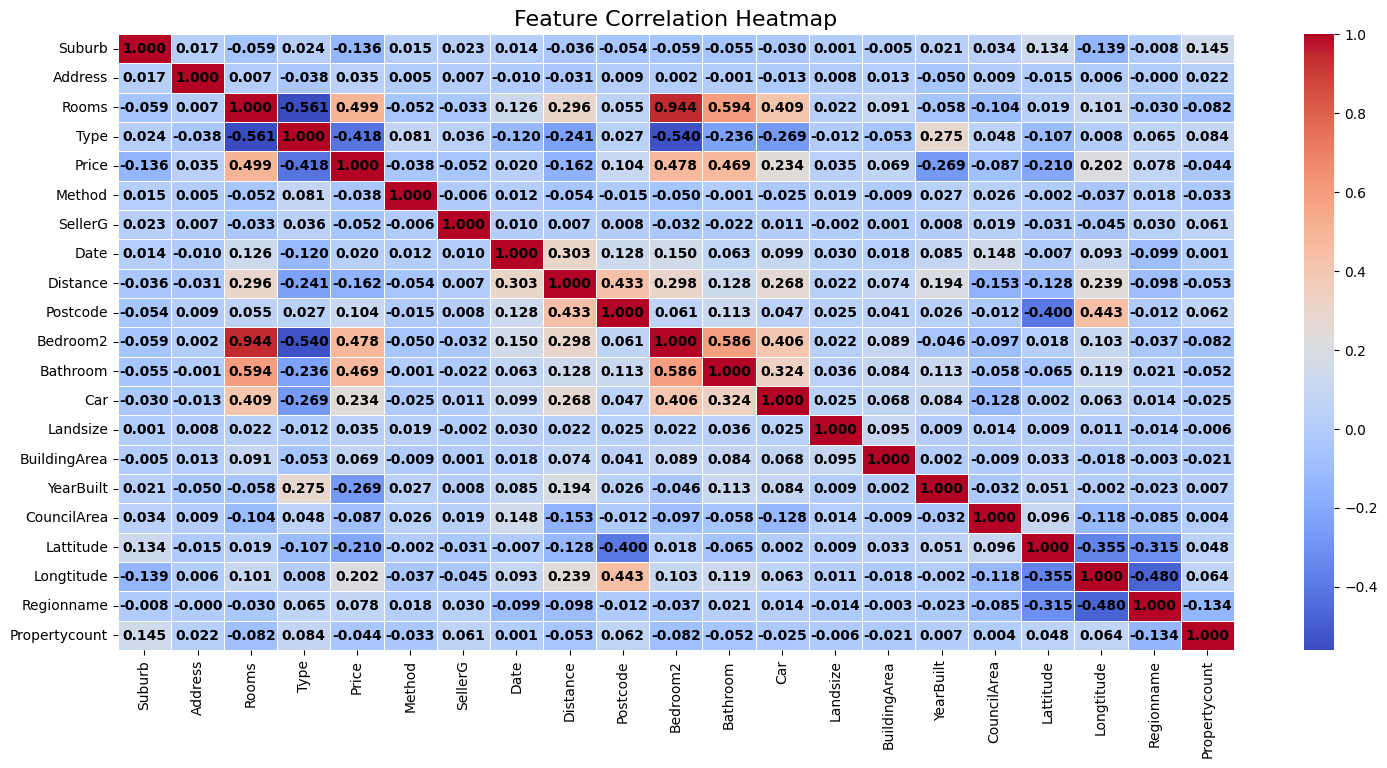


We had fill the nulls in df\_address and the row was nulls and missing in df\_location we had drop them and then take the df\_address and return it back to df[‘Address’]

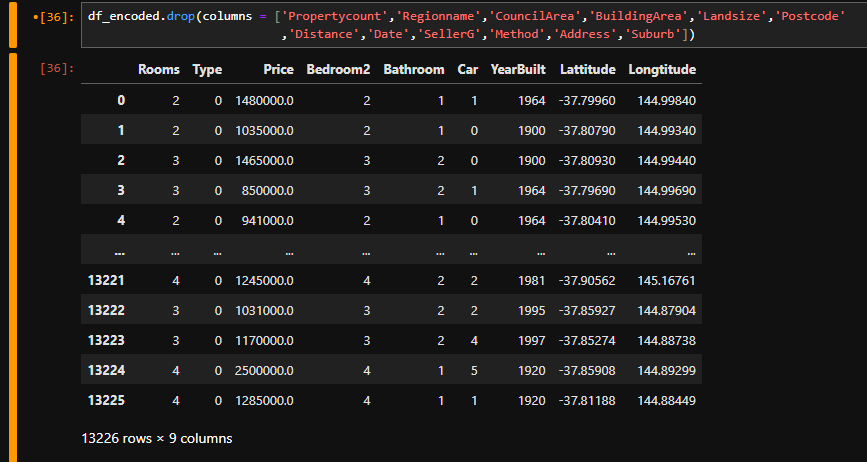
**Prepare the Model:**

1-Encode the data

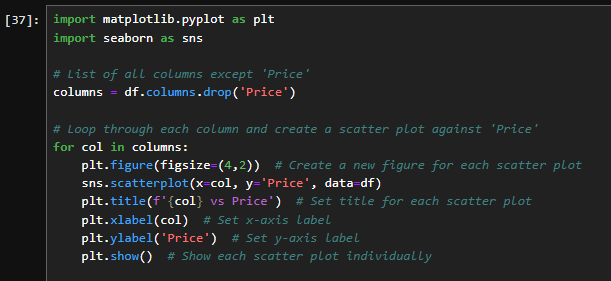
2-heatmap

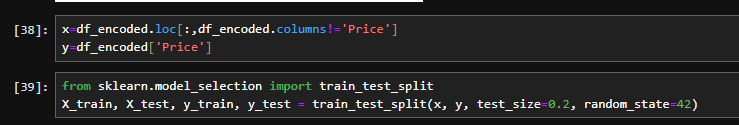
The more near to 1 is highly related to the other column

3-drop the low related to Price column



4- make scatter plot between price and all other columns to see the relation and choose the best algorithm

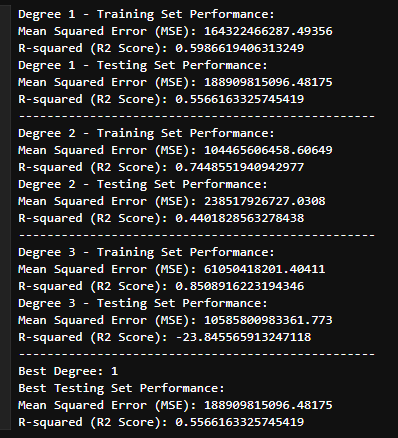


5-split data , y is the price and x is the other columns and split x and y to train and test , 20% of data for test and the other for train the model 

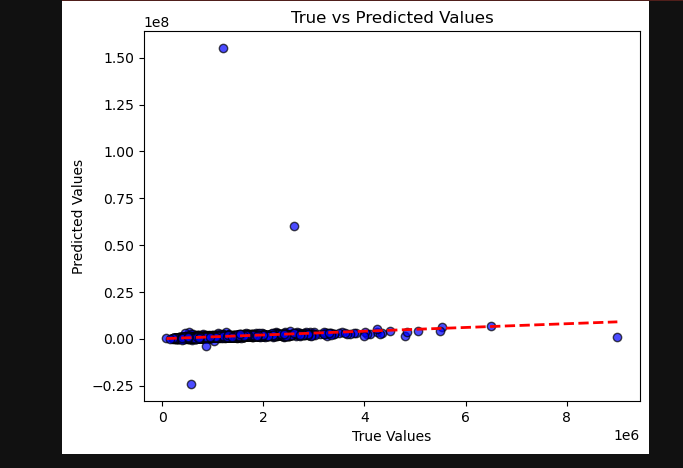
**Model:**

Use Linear Regression Model

Use polynomial and choose the best degree from (1,2,3) based on MSE



Visualize the true vs predicted values



**Power Bi visualization:**

