Housing Price Dataset

|  |  |
| --- | --- |
| **Size** | 24 MB |
| **Dataset Characteristics:** | Multivariate |
| **Attribute Characteristics:** | Numerical (continuous), Categorical (discrete), Ordinal |
| **Associated Tasks:** | Regression |
| **Number of Instances:** | 339651 |
| **Number of Attributes:** | 11 |
| **Missing Values?** | Yes |
| **Irrelevant Features?** | Yes |
| **Area:** | Business |
| **Additional Details:** | N/A |

# Dataset Information:

This dataset contains the housing price of 26 cities in the US, detailed info such as the location and room type are given. You have to predict the housing price (last column). You cannot ignore any of the rows that contain missing values. We want to see how realistically you handle these missing values. Also there maybe irrelevant features available which may or may not be detrimental to the prediction task. You should eliminate such features (if any).

Attribute Information:

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Type** |
| date | Timestamp of the record | str |
| location | The location of house | str |
| type | The type of house | str |
| block | The block of house | str |
| street | The street of house | str |
| storey\_range | Story of the house | str |
| area\_sqm | Size of the house | str |
| price | Price of house | float |
| flat\_mode | The extend of new/old, denote by an uppercase letter. | str |
| commence\_date | It denotes when the house is licensed. | int |
| Quanlity\_Index | Air quality index by environmentalists | int |