**Data Understanding**

Data Source

This data is downloaded from kaggle. All data is available at Open Data D.C. The residential and address point data is managed by the Office of the Chief Technology Officer

Data Description

This data is about Washington DC Residential. This data set have 158957 rows with 48 columns.

Overview

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset Statistics** | **Count** | **Variable Types** | **Count** |
| Number of variables | 49 | Numeric | 25 |
| Number of observations | 158957 | Categorical | 23 |
| Missing cells | 1251928 | Unsupported | 1 |
| Missing cells (%) | 16.1 % |  |  |
| Duplicate rows | 0 |  |  |
| Duplicate rows (%) | 0.0 % |  |  |
| Total size in memory | 59.5 MiB |  |  |
| Average record size in memory | 392.0 B |  |  |

Columns Explanation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No** | **Feature** | **Data Type** | **Zeros**  **(%)** | **Null (%)** | **Distinct (%)** | **Description** |
| 1 | #Unnamed: 0 | Int64 | < 0.1 % | 0 % | 100 % | Number of Index |
| 2 | BATHRM | int64 | < 0.1 % | 0 % | < 0.1 % | Number of bathrooms |
| 3 | HF\_BATHRM | int64 | 58.6 % | 0 % | < 0.1 % | Number of half bathroom  (No bathtub or shower) |
| 4 | HEAT | object | - | 0 % | < 0.1 % | Heating type |
| 5 | AC | Object | - | 0 % | < 0.1 % | Air Conditioner availability (Y/N) |
| 6 | NUM\_UNITS | float64 | 0.1 % | 32.9 % | < 0.1 % | Number of units |
| 7 | ROOMS | int64 | 0.1 % | 0 % | < 0.1 % | Number of rooms |
| 8 | BEDRM | int64 | 3.3 % | 0 % | < 0.1 % | Number of bedrooms |
| 9 | AYB | float64 | 0 % | 0.2 % | 0.1 % | The earliest time the main portion of the building was built |
| 10 | YR\_RMDL | float64 | 0 % | 49.1 % | 0.1 % | The year structure was remodeled |
| 11 | EYB | int64 | 0 % | 0 % | 0.1 % | The year an improvement was built more recent than actual year built |
| 12 | STORIES | float64 | < 0.1 % | 32.9 % | < 0.1 % | Number of stories in primary dwelling |
| 13 | SALEDATE | object | - | 16.8 % | 5.2 % | Date of most recent sale |
| 14 | PRICE | float64 | 0 % | 38.2 % | 13.7 % | Price of most recent sale |
| 15 | QUALIFIED | object | - | 0 % | < 0.1 % | Internally used indicator to reflect if a sale is representative of market value according to the office's internal criteria |
| 16 | SALE\_NUM | int64 | 0 % | 0 % | < 0.1 % | Number of times it's been sold since May 2014 |
| 17 | GBA | float64 | < 0.1 % | 32.9 % | 4.5 % | Gross building area in square feet |
| 18 | BLDG\_NUM | int64 | - | 0 % | <0.1 % | Building number on the property |
| 19 | STYLE | object | - | 32.9 % | < 0.1 % | Type of story |
| 20 | STRUCT | object | - | 32.9 % | < 0.1 % | Building structure |
| 21 | GRADE | object | - | 32.9 % | < 0.1 % | Property Grade |
| 22 | CNDTN | object | - | 32.9 % | < 0.1 % | Property Condition |
| 23 | EXTWALL | object | - | 32.9 % | < 0.1 % | Exterior wall type |
| 24 | ROOF | object | - | 32.9 % | < 0.1 % | Roof type |
| 25 | INTWALL | object | - | 32.9 % | < 0.1 % | Interior wall type |
| 26 | KITCHENS | float64 | 0.1 % | 32.9 % | < 0.1 % | Number of kitchens |
| 27 | FIREPLACES | int64 | 65.3 % | 0 % | < 0.1 % | Number of fireplaces |
| 28 | USECODE | int64 | 0 % | 0 % | < 0.1 % | Property use code type |
| 29 | LANDAREA | int64 | < 0.1 % | 0 % | 7.1 % | Land area of property in square feet |
| 30 | GIS\_LAST\_MOD\_DTTM | object | - | 0 % | < 0.1 % | Last modified data |
| 31 | SOURCE | object | - | 0 % | < 0.1 % | Raw data source |
| 32 | CMPLX\_NUM | float64 | 0 % | 67.1 % | 5.6 % | Complex number |
| 33 | LIVING\_GBA | float64 | < 0.1 % | 67.1 % | 4.2 % | Gross building area in square feet |
| 34 | FULLADDRESS | object | - | 33.3 % | 99.9 % | Full street address |
| 35 | CITY | object | - | 33.3 % | < 0.1 % | City |
| 36 | STATE | object | - | 33.3 % | < 0.1 % | State |
| 37 | ZIPCODE | float64 | 0 % | < 0.1 % | < 0.1 % | Zipcode |
| 38 | NATIONALGRID | object | - | 33.3 % | 99.9 % | Address location national grid coordinate spatial address |
| 39 | LATITUDE | float64 | 0 % | < 0.1 % | 66.4% | Latitude |
| 40 | LONGITUDE | float64 | 0 % | < 0.1 % | 66.6 % | Longitude |
| 41 | ASSESSMENT\_NBHD | object | - | < 0.1 % | < 0.1 % | Neighborhood ID |
| 42 | ASSESSMENT\_SUBNBHD | object | - | 20.5 % | 0.1 % | Subneighborhood ID |
| 43 | CENSUS\_TRACT | float64 | 0 % | < 0.1 % | 0.1 % | Census tract |
| 44 | CENSUS\_BLOCK | object | - | 33.3 % | 3.6 % | Census block |
| 45 | WARD | object | - | < 0.1 % | < 0.1 % | Ward (district is divided into eight wards, each with approximately 75,000 residents) |
| 46 | SQUARE | object | - | 0 % | - | Square (Part of Square Suffix Lot (SSL) an address identifier in DC) |
| 47 | X | float64 | 0 % | 0.1 % | 2.1 % | Longitude |
| 48 | Y | float64 | 0 % | 0.1 % | 2.1 % | Latitude |
| 49 | QUADRANT | object | - | 0.1 % | < 0.1 % | City quadrant (NE, SE, SW, NW) |

Feature Identification:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Numerical Features** | | **Categorical Features** | |
| **Feature** | **Description** | **Feature** | **Description** |
| 1 | #Unnamed: 0 | Number of Index | HEAT | Heating type |
| 2 | BATHRM | Number of bathrooms | AC | Air Conditioner availability (Y/N) |
| 3 | HF\_BATHRM | Number of half bathroom  (No bathtub or shower) | SALEDATE | Date of most recent sale |
| 4 | NUM\_UNITS | Number of units | QUALIFIED | Internally used indicator to reflect if a sale is representative 5of market value according to the office's internal criteria |
| 5 | ROOMS | Number of rooms | STYLE | Type of story |
| 6 | BEDRM | Number of bedrooms | STRUCT | Building structure |
| 7 | AYB | The earliest time the main portion of the building was built | GRADE | Property Grade |
| 8 | YR\_RMDL | The year structure was remodeled | CNDTN | Property Condition |
| 9 | EYB | The year an improvement was built more recent than actual year built | EXTWALL | Exterior wall type |
| 10 | STORIES | Number of stories in primary dwelling | ROOF | Roof type |
| 11 | PRICE | Price of most recent sale | INTWALL | Interior wall type |
| 12 | SALE\_NUM | Number of times it's been sold since May 2014 | GIS\_LAST\_MOD\_DTTM | Last modified data |
| 13 | GBA | Gross building area in square feet | SOURCE | Raw data source |
| 14 | BLDG\_NUM | Building number on the property | FULLADDRESS | Full street address |
| 15 | KITCHENS | Number of kitchens | CITY | City |
| 16 | FIREPLACES | Number of fireplaces | STATE | State |
| 17 | USECODE | Property use code type | NATIONALGRID | Address location national grid coordinate spatial address |
| 18 | LANDAREA | Land area of property in square feet | ASSESSMENT\_NBHD | Neighborhood ID |
| 19 | CMPLX\_NUM | Complex number | ASSESSMENT\_SUBNBHD | Subneighborhood ID |
| 20 | LIVING\_GBA | Gross building area in square feet | CENSUS\_BLOCK | Census block |
| 21 | ZIPCODE | Zipcode | WARD | Ward (district is divided into eight wards, each with approximately 75,000 residents) |
| 22 | LATITUDE | Latitude | SQUARE | Square (Part of Square Suffix Lot (SSL) an address identifier in DC) |
| 23 | LONGITUDE | Longitude | QUADRANT | City quadrant (NE, SE, SW, NW) |
| 24 | CENSUS\_TRACT | Census tract |  |  |
| 25 | X | Longitude |  |  |
| 26 | Y | Latitude |  |  |

Property Overview:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **House Features** | | **Construction Details** | |
| **Based on:** | **Consist of:** | **Based on:** | **Consist of:** |
| 1 | Interior Details | * Bedrooms * Bathrooms * Half Bathrooms | Property | * Story * Style |
| 2 | Heating | * Forced Air * Electric Based * etc | Wall Type | * Exterior * Interior |
| 3 | Cooling | * Air Conditioner | Roof Type | * Slate * Concrete |
| 4 | Interior Features | * Kitchens * Fireplaces | Condition | * Property Condition * Property Grade |
| 5 | Other Features | * Gross Building Area * Living Gross Building Area | Notable Dates | * Year Built * Year Remodel |

Property Condition

The assessment of the condition is varied from different individuals. Hence, there is no uniform rule to classify properties into differernt levels of condition. Here is a detailed explanation of condition from Marshall & Swift Condition Assessment (page E-6).

**Excellent Condition** - All items that can normally be repaired or refinished have recently been corrected, such as new roofing, paint, furance overhaul, state of the art components, etc. With no functional inadequacies of any consequence and all major short-lived components in like-new condition, the overall effective age has been substantially reduced upon complete revitilization of the structure regardless of the actual chronological age.

**Very Good Condition** - All items well maintained, many having been overhauled and repaired as they’ve showed signs of wear, increasing the life expectancy and lowering the effective age with little deterioration or obsolesence evident with a high degree of utility.

**Good Condition** - No obvious maintenance required but neither is everything new. Appearance and utility are above the standard and the overall effective age will be lower than the typical property.

**Average Condition** - Some evidence of deferred maintenance and normal obsolescence with age in that a few minor repairs are needed along with some refinishing. But with all major components still functional and contributing toward an extended life expectancy, effective age and utility is standard for like properties of its class and usage.

**Fair Condition** (Badly worn) - Much repair needed. Many items need refinishing or overhauling, deferred maintenance obvious, inadequate building utility and services all shortening the life expectancy and increasing the effective age.

**Poor Condition** (Worn Out) - Repair and overall needed on painted surfaces, roofing, plumbing, heating, numerous functional inadequacies, substandard utilities etc. (found only in extraordinary circumstances). Excessive deferred maintenance and abuse, limited value-in-use, approaching abandonment or major reconstruction, reuse or change in occupancy is imminent. Effective age is near the end of the scale regardless of the actual chronological age.

Property type use codes

Property use codes are utilized to categorize and group similar types of properties for easy identification. Data for properties are collected and the properties are grouped by use for various types of analysis and comparison (Property use code manual page 1).

|  |  |  |
| --- | --- | --- |
| **Code** | **Type** | **Description** |
| 11 | Residential Row Single Family | Single-family dwelling with 2 walls built as common walls with another structure, 2 exposed walls; primarily used as place of abode. |
| 12 | Residential Detached Single Family | Free-standing dwelling with open space around it and in all exterior walls; primarily used as abode. |
| 13 | Residential Semi-Detached Single Family | Structure with 1 dwelling place, 1 wall built as common wall with another structure, 3 exposed walls; primarily used as abode. |
| 15 | Residential Mixed Use | Single-family property with commercial (usually office) space in part of house. If use is mostly single-family, lot may be eligible for a Homestead Deduction. Mixed-use eligible. |
| 16 | Residential Condo Horizontal | Enclosed space of 1 or more rooms, occupying all or part of 1 or more floors; entrance no higher than 3 floors; single-family use; may/may not have parking, laundry, patio, etc. |
| 17 | Residential Condo Vertical | Enclosed space of 1 or more rooms, occupying all/part of 1 or more floors; in structure with elevator; more than 3 floors. Original primaries use single-family. May have parking, laundry, patio, etc. |
| 19 | Residential Single Family Miscellaneous | All other residential-single family uses not otherwise coded. |
| 23 | Residential Flats Less than 5 | Structure with more than 1 single family unit, less than 5; usually self-contained, under 1 roof; few accessory uses; in some cases, owner occupies 1 unit; built for this |
| 24 | Residential Conversions less than 5 | Structure with more than 1 single-family unit, but less than 5; usually self-contained, under 1 roof; few accessory uses; 1 unit may be owner-occupied; original primary use not multi-family. |
| 29 | Residential Multifamily Miscellaneous | All other residential multi-family uses not otherwise noted. Mixed-use eligible. |
| 39 | Residential Transient Miscellaneous | All other residential transient not otherwise coded. |
| 41 | Store Small 1 Story | Structure used primarily for retail sales; row, attached, or detached; with/without accessory uses; with/without living quarters. |
| 81 | Religious | Structure devoted to public worship; housing for and/or education of clergy/officials connected to religious activity; religious communities. |
| 83 | Educational | Structure devoted to any level of public/private instruction. May include administrative, accessory functions; parking, retail sales, secondary use. |
| 116 | Condo Horizontal Combined | Unit in a structure with entrance no higher than 3 floors; designed primarily for single family residential use; accessory uses. Abuts primary unit; owner entitled to lower (Class 1) tax rate, but not Homestead Deduction. |
| 117 | Condo Vertical Combined | Unit in structure with entrance no higher than 3 floors, designed primarily for single family residential use: accessory uses. Abuts primary unit; owner entitled to lower (Class 1) tax rate, but not Homestead Deduction. |

The area in Washington DC is divided into several **Quadrant**. Each **Quadrant** is furhter divided into **Ward** and each Ward is divided into Assessment Neighborhood.

**Quadrant and Ward**

The Quadrant in Washington DC is Northwest, Northeast, Southeast, Southwest, and each Quadrant is further divided into Ward.

* **Ward 1** consist of: Adams Morgan, Columbia Heights, Howard University, etc
* **Ward 2** consist of: Burleith, Chinatown, Downtown, etc
* **Ward 3** consist of: Chevy Chase, Cleveland Park, Colony Hill, etc
* **Ward 4** consist of: Bright Wood Park, Crestwood, Petworth, etc
* **Ward 5** consist of: Arboretum, Brentwood, Brookland, etc
* **Ward 6** consist of: Capiton Hill, Kingman Park, Navy Yard, etc
* **Ward 7** consist of: Benning Heights, Benning Ridge, Burrville, etc
* **Ward 8** consist of: Anacosita, Barry Farm, Bellevue, etc

**Distribution Style**

Residential style sale in Washington DC is divided into 17 Categories, which is:

* **1 story:** house consisting of a ground storey only.
* **1.5 story:** house that has the master bedroom suite on the main floor and all other bedrooms on the second floor.
* **2 story:** houses have the master bedroom suite and additional bedrooms located on the second level of the home.
* **2.5 story:** house would mean its a house that is 2 storeys with a loft.
* **3 story:** houses would mean its a house that is 3 storeys, usually have a ground floor used for the living space and kitchen.
* **3.5 story:** house would mean its a house that is 3 storeys with a loft.
* **4 story:** houses would mean its a house that is 4 storeys.
* **4.5 story:** house would mean its a house that is 4 storeys with a loft.
* **Split Level:** (also called a tri-level home) is a style of house in which the floor levels are staggered. There are typically two short sets of stairs, one running upward to a bedroom level, and one going downward toward a basement area.
* **Split Foyer and Bi-Level:** same as split level but in another ward different called.
* **Vacant Property:** means residential, commercial, industrial, or mixed-use real property that has not been lawfully occupied and maintained, actively marketed for rental, or under active construction for a continuous period of forty-five (45) days or more.
* **Unfin & Fin:** meaning about the loft not yet completed.