**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.

Columns Explanation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No** | **Feature** | **Data Type** | **Null Count** | **Null Percentage** | **Unique Data** | **Description** |
| 1 | #Unnamed: 0 | Int64 | 0 | 0 | 158957 | Number of Index |
| 2 | BATHRM | int64 | 0 | 0 | 15 | Number of bathrooms |
| 3 | HF\_BATHRM | int64 | 0 | 0 | 10 | Number of half bathroom  (No bathtub or shower) |
| 4 | HEAT | object | 0 | 0 | 14 | Heating type |
| 5 | AC | object | 0 | 0 | 3 | Air Conditioner availability (Y/N) |
| 6 | NUM\_UNITS | float64 | 52261 | 32.88 | 7 | Number of units |
| 7 | ROOMS | int64 | 0 | 0 | 40 | Number of rooms |
| 8 | BEDRM | int64 | 0 | 0 | 20 | Number of bedrooms |
| 9 | AYB | float64 | 271 | 0.17 | 220 | The earliest time the main portion of the building was built |
| 10 | YR\_RMDL | float64 | 78029 | 49.09 | 110 | The year structure was remodeled |
| 11 | EYB | int64 | 0 | 0 | 135 | The year an improvement was built more recent than actual year built |
| 12 | STORIES | float64 | 52305 | 32.91 | 40 | Number of stories in primary dwelling |
| 13 | SALEDATE | object | 26770 | 16.84 | 6937 | Date of most recent sale |
| 14 | PRICE | float64 | 60741 | 38.21 | 13496 | Price of most recent sale |
| 15 | QUALIFIED | object | 0 | 0 | 2 | 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 | 15 | Number of times it's been sold since May 2014 |
| 17 | GBA | float64 | 52261 | 32.88 | 4764 | Gross building area in square feet |
| 18 | BLDG\_NUM | int64 | 0 | 0 | 5 | Building number on the property |
| 19 | STYLE | object | 52261 | 32.88 | 18 | Type of story |
| 20 | STRUCT | object | 52261 | 32.88 | 9 | Building structure |
| 21 | GRADE | object | 52261 | 32.88 | 13 | Property Grade |
| 22 | CNDTN | object | 52261 | 32.88 | 7 | Property Condition |
| 23 | EXTWALL | object | 52261 | 32.88 | 25 | Exterior wall type |
| 24 | ROOF | object | 52261 | 32.88 | 16 | Roof type |
| 25 | INTWALL | object | 52261 | 32.88 | 12 | Interior wall type |
| 26 | KITCHENS | float64 | 52262 | 32.88 | 8 | Number of kitchens |
| 27 | FIREPLACES | int64 | 0 | 0 | 20 | Number of fireplaces |
| 28 | USECODE | int64 | 0 | 0 | 16 | Property use code type |
| 29 | LANDAREA | int64 | 0 | 0 | 11359 | Land area of property in square feet |
| 30 | GIS\_LAST\_MOD\_DTTM | object | 0 | 0 | 2 | Last modified data |
| 31 | SOURCE | object | 0 | 0 | 2 | Raw data source |
| 32 | CMPLX\_NUM | float64 | 106696 | 67.12 | 2913 | Complex number |
| 33 | LIVING\_GBA | float64 | 106696 | 67.12 | 2216 | Gross building area in square feet |
| 34 | FULLADDRESS | object | 52917 | 33.29 | 105978 | Full street address |
| 35 | CITY | object | 52906 | 33.28 | 1 | City |
| 36 | STATE | object | 52906 | 33.28 | 1 | State |
| 37 | ZIPCODE | float64 | 1 | 0 | 24 | Zipcode |
| 38 | NATIONALGRID | object | 52906 | 33.28 | 105949 | Address location national grid coordinate spatial address |
| 39 | LATITUDE | float64 | 1 | 0 | 105522 | Latitude |
| 40 | LONGITUDE | float64 | 1 | 0 | 105935 | Longitude |
| 41 | ASSESSMENT\_NBHD | object | 1 | 0 | 57 | Neighborhood ID |
| 42 | ASSESSMENT\_SUBNBHD | object | 32551 | 20.48 | 121 | Subneighborhood ID |
| 43 | CENSUS\_TRACT | float64 | 1 | 0 | 176 | Census tract |
| 44 | CENSUS\_BLOCK | object | 52906 | 33.28 | 3848 | Census block |
| 45 | WARD | object | 1 | 0 | 8 | Ward (district is divided into eight wards, each with approximately 75,000 residents) |
| 46 | SQUARE | object | 0 | 0 | 3512 | Square (Part of Square Suffix Lot (SSL) an address identifier in DC) |
| 47 | X | float64 | 237 | 0.15 | 3291 | Longitude |
| 48 | Y | float64 | 237 | 0.15 | 3291 | Latitude |
| 49 | QUADRANT | object | 237 | 0.15 | 4 | City quadrant (NE, SE, SW, NW) |

List of **Numerical** Features:

|  |  |  |
| --- | --- | --- |
| **No** | **Feature** | **Description** |
| 1 | #Unnamed: 0 | Number of Index |
| 2 | BATHRM | Number of bathrooms |
| 3 | HF\_BATHRM | Number of half bathroom  (No bathtub or shower) |
| 4 | NUM\_UNITS | Number of units |
| 5 | ROOMS | Number of rooms |
| 6 | BEDRM | Number of bedrooms |
| 7 | AYB | The earliest time the main portion of the building was built |
| 8 | YR\_RMDL | The year structure was remodeled |
| 9 | EYB | The year an improvement was built more recent than actual year built |
| 10 | STORIES | Number of stories in primary dwelling |
| 11 | PRICE | Price of most recent sale |
| 12 | SALE\_NUM | Number of times it's been sold since May 2014 |
| 13 | GBA | Gross building area in square feet |
| 14 | BLDG\_NUM | Building number on the property |
| 15 | KITCHENS | Number of kitchens |
| 16 | FIREPLACES | Number of fireplaces |
| 17 | USECODE | Property use code type |
| 18 | LANDAREA | Land area of property in square feet |
| 19 | CMPLX\_NUM | Complex number |
| 20 | LIVING\_GBA | Gross building area in square feet |
| 21 | ZIPCODE | Zipcode |
| 22 | LATITUDE | Latitude |
| 23 | LONGITUDE | Longitude |
| 24 | CENSUS\_TRACT | Census tract |
| 25 | X | Longitude |
| 26 | Y | Latitude |

List of **Categorical** Features:

|  |  |  |
| --- | --- | --- |
| **No** | **Feature** | **Description** |
| 1 | HEAT | Heating type |
| 2 | AC | Air Conditioner availability (Y/N) |
| 3 | SALEDATE | Date of most recent sale |
| 4 | QUALIFIED | Internally used indicator to reflect if a sale is representative 5of market value according to the office's internal criteria |
| 5 | STYLE | Type of story |
| 6 | STRUCT | Building structure |
| 7 | GRADE | Property Grade |
| 8 | CNDTN | Property Condition |
| 9 | EXTWALL | Exterior wall type |
| 10 | ROOF | Roof type |
| 11 | INTWALL | Interior wall type |
| 12 | GIS\_LAST\_MOD\_DTTM | Last modified data |
| 13 | SOURCE | Raw data source |
| 14 | FULLADDRESS | Full street address |
| 15 | CITY | City |
| 16 | STATE | State |
| 17 | NATIONALGRID | Address location national grid coordinate spatial address |
| 18 | ASSESSMENT\_NBHD | Neighborhood ID |
| 19 | ASSESSMENT\_SUBNBHD | Subneighborhood ID |
| 20 | CENSUS\_BLOCK | Census block |
| 21 | WARD | Ward (district is divided into eight wards, each with approximately 75,000 residents) |
| 22 | SQUARE | Square (Part of Square Suffix Lot (SSL) an address identifier in DC) |
| 23 | QUADRANT | City quadrant (NE, SE, SW, NW) |

House Features:

1. Interior details

* Bedrooms
* Bathrooms
* Half Bathrooms

1. Heating

* Forced Air
* Electric based
* Wall Furnace
* Etc

1. Cooling

* Air Conditioner

1. Interior features:

* Kitchens
* Fireplaces

1. Other features:

* Gross building area
* Living gross building area

Construction details:

1. Property:

* Story
* Style

1. Wall type:

* Exterior
* Interior

1. Roof type
2. Condition:

* Property Condition
* Property Grade
* Year Built

1. Notable Dates:

* Remodel year

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 storey 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.

The top 3 style of residential in Washington DC are:

* 2 story: 42097
* 3 story: 4347
* 2.5 story: 3013