* house.shape

(1460,81)

81 Columns Datatypes– Float(3), Int(35), Object(43)

* house.info()

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Id 1460 non-null int64

1 MSSubClass 1460 non-null int64

2 MSZoning 1460 non-null object

3 LotFrontage 1201 non-null float64

4 LotArea 1460 non-null int64

5 Street 1460 non-null object

6 Alley 91 non-null object

7 LotShape 1460 non-null object

8 LandContour 1460 non-null object

9 Utilities 1460 non-null object

10 LotConfig 1460 non-null object

11 LandSlope 1460 non-null object

12 Neighborhood 1460 non-null object

13 Condition1 1460 non-null object

14 Condition2 1460 non-null object

15 BldgType 1460 non-null object

16 HouseStyle 1460 non-null object

17 OverallQual 1460 non-null int64

18 OverallCond 1460 non-null int64

19 YearBuilt 1460 non-null int64

20 YearRemodAdd 1460 non-null int64

21 RoofStyle 1460 non-null object

22 RoofMatl 1460 non-null object

23 Exterior1st 1460 non-null object

24 Exterior2nd 1460 non-null object

25 MasVnrType 1452 non-null object

26 MasVnrArea 1452 non-null float64

27 ExterQual 1460 non-null object

28 ExterCond 1460 non-null object

29 Foundation 1460 non-null object

30 BsmtQual 1423 non-null object

31 BsmtCond 1423 non-null object

32 BsmtExposure 1422 non-null object

33 BsmtFinType1 1423 non-null object

34 BsmtFinSF1 1460 non-null int64

35 BsmtFinType2 1422 non-null object

36 BsmtFinSF2 1460 non-null int64

37 BsmtUnfSF 1460 non-null int64

38 TotalBsmtSF 1460 non-null int64

39 Heating 1460 non-null object

40 HeatingQC 1460 non-null object

41 CentralAir 1460 non-null object

42 Electrical 1459 non-null object

43 1stFlrSF 1460 non-null int64

44 2ndFlrSF 1460 non-null int64

45 LowQualFinSF 1460 non-null int64

46 GrLivArea 1460 non-null int64

47 BsmtFullBath 1460 non-null int64

48 BsmtHalfBath 1460 non-null int64

49 FullBath 1460 non-null int64

50 HalfBath 1460 non-null int64

51 BedroomAbvGr 1460 non-null int64

52 KitchenAbvGr 1460 non-null int64

53 KitchenQual 1460 non-null object

54 TotRmsAbvGrd 1460 non-null int64

55 Functional 1460 non-null object

56 Fireplaces 1460 non-null int64

57 FireplaceQu 770 non-null object

58 GarageType 1379 non-null object

59 GarageYrBlt 1379 non-null float64

60 GarageFinish 1379 non-null object

61 GarageCars 1460 non-null int64

62 GarageArea 1460 non-null int64

63 GarageQual 1379 non-null object

64 GarageCond 1379 non-null object

65 PavedDrive 1460 non-null object

66 WoodDeckSF 1460 non-null int64

67 OpenPorchSF 1460 non-null int64

68 EnclosedPorch 1460 non-null int64

69 3SsnPorch 1460 non-null int64

70 ScreenPorch 1460 non-null int64

71 PoolArea 1460 non-null int64

72 PoolQC 7 non-null object

73 Fence 281 non-null object

74 MiscFeature 54 non-null object

75 MiscVal 1460 non-null int64

76 MoSold 1460 non-null int64

77 YrSold 1460 non-null int64

78 SaleType 1460 non-null object

79 SaleCondition 1460 non-null object

80 SalePrice 1460 non-null int64

* house.describe(include='object').T

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **count** | **unique** | **top** | **freq** |
| **MSZoning** | 1460 | 5 | RL | 1151 |
| **Street** | 1460 | 2 | Pave | 1454 |
| **Alley** | 91 | 2 | Grvl | 50 |
| **LotShape** | 1460 | 4 | Reg | 925 |
| **LandContour** | 1460 | 4 | Lvl | 1311 |
| **Utilities** | 1460 | 2 | AllPub | 1459 |
| **LotConfig** | 1460 | 5 | Inside | 1052 |
| **LandSlope** | 1460 | 3 | Gtl | 1382 |
| **Neighborhood** | 1460 | 25 | NAmes | 225 |
| **Condition1** | 1460 | 9 | Norm | 1260 |
| **Condition2** | 1460 | 8 | Norm | 1445 |
| **BldgType** | 1460 | 5 | 1Fam | 1220 |
| **HouseStyle** | 1460 | 8 | 1Story | 726 |
| **RoofStyle** | 1460 | 6 | Gable | 1141 |
| **RoofMatl** | 1460 | 8 | CompShg | 1434 |
| **Exterior1st** | 1460 | 15 | VinylSd | 515 |
| **Exterior2nd** | 1460 | 16 | VinylSd | 504 |
| **MasVnrType** | 1452 | 4 | None | 864 |
| **ExterQual** | 1460 | 4 | TA | 906 |
| **ExterCond** | 1460 | 5 | TA | 1282 |
| **Foundation** | 1460 | 6 | PConc | 647 |
| **BsmtQual** | 1423 | 4 | TA | 649 |
| **BsmtCond** | 1423 | 4 | TA | 1311 |
| **BsmtExposure** | 1422 | 4 | No | 953 |
| **BsmtFinType1** | 1423 | 6 | Unf | 430 |
| **BsmtFinType2** | 1422 | 6 | Unf | 1256 |
| **Heating** | 1460 | 6 | GasA | 1428 |
| **HeatingQC** | 1460 | 5 | Ex | 741 |
| **CentralAir** | 1460 | 2 | Y | 1365 |
| **Electrical** | 1459 | 5 | SBrkr | 1334 |
| **KitchenQual** | 1460 | 4 | TA | 735 |
| **Functional** | 1460 | 7 | Typ | 1360 |
| **FireplaceQu** | 770 | 5 | Gd | 380 |
| **GarageType** | 1379 | 6 | Attchd | 870 |
| **GarageFinish** | 1379 | 3 | Unf | 605 |
| **GarageQual** | 1379 | 5 | TA | 1311 |
| **GarageCond** | 1379 | 5 | TA | 1326 |
| **PavedDrive** | 1460 | 3 | Y | 1340 |
| **PoolQC** | 7 | 3 | Gd | 3 |
| **Fence** | 281 | 4 | MnPrv | 157 |
| **MiscFeature** | 54 | 4 | Shed | 49 |
| **SaleType** | 1460 | 9 | WD | 1267 |
| **SaleCondition** | 1460 | 6 | Normal | 1198 |

* house.describe().T

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |  |
| **Id** | 1460 | 730.5 | 421.61 | 1 | 365.75 | 730.5 | 1095.25 | 1460 |
| **MSSubClass** | 1460 | 56.89726 | 42.30057 | 20 | 20 | 50 | 70 | 190 |
| **LotFrontage** | 1201 | 70.04996 | 24.28475 | 21 | 59 | 69 | 80 | 313 |
| **LotArea** | 1460 | 10516.83 | 9981.265 | 1300 | 7553.5 | 9478.5 | 11601.5 | 215245 |
| **OverallQual** | 1460 | 6.099315 | 1.382997 | 1 | 5 | 6 | 7 | 10 |
| **OverallCond** | 1460 | 5.575342 | 1.112799 | 1 | 5 | 5 | 6 | 9 |
| **YearBuilt** | 1460 | 1971.268 | 30.2029 | 1872 | 1954 | 1973 | 2000 | 2010 |
| **YearRemodAdd** | 1460 | 1984.866 | 20.64541 | 1950 | 1967 | 1994 | 2004 | 2010 |
| **MasVnrArea** | 1452 | 103.6853 | 181.0662 | 0 | 0 | 0 | 166 | 1600 |
| **BsmtFinSF1** | 1460 | 443.6397 | 456.0981 | 0 | 0 | 383.5 | 712.25 | 5644 |
| **BsmtFinSF2** | 1460 | 46.54932 | 161.3193 | 0 | 0 | 0 | 0 | 1474 |
| **BsmtUnfSF** | 1460 | 567.2404 | 441.867 | 0 | 223 | 477.5 | 808 | 2336 |
| **TotalBsmtSF** | 1460 | 1057.429 | 438.7053 | 0 | 795.75 | 991.5 | 1298.25 | 6110 |
| **1stFlrSF** | 1460 | 1162.627 | 386.5877 | 334 | 882 | 1087 | 1391.25 | 4692 |
| **2ndFlrSF** | 1460 | 346.9925 | 436.5284 | 0 | 0 | 0 | 728 | 2065 |
| **LowQualFinSF** | 1460 | 5.844521 | 48.62308 | 0 | 0 | 0 | 0 | 572 |
| **GrLivArea** | 1460 | 1515.464 | 525.4804 | 334 | 1129.5 | 1464 | 1776.75 | 5642 |
| **BsmtFullBath** | 1460 | 0.425342 | 0.518911 | 0 | 0 | 0 | 1 | 3 |
| **BsmtHalfBath** | 1460 | 0.057534 | 0.238753 | 0 | 0 | 0 | 0 | 2 |
| **FullBath** | 1460 | 1.565068 | 0.550916 | 0 | 1 | 2 | 2 | 3 |
| **HalfBath** | 1460 | 0.382877 | 0.502885 | 0 | 0 | 0 | 1 | 2 |
| **BedroomAbvGr** | 1460 | 2.866438 | 0.815778 | 0 | 2 | 3 | 3 | 8 |
| **KitchenAbvGr** | 1460 | 1.046575 | 0.220338 | 0 | 1 | 1 | 1 | 3 |
| **TotRmsAbvGrd** | 1460 | 6.517808 | 1.625393 | 2 | 5 | 6 | 7 | 14 |
| **Fireplaces** | 1460 | 0.613014 | 0.644666 | 0 | 0 | 1 | 1 | 3 |
| **GarageYrBlt** | 1379 | 1978.506 | 24.68973 | 1900 | 1961 | 1980 | 2002 | 2010 |
| **GarageCars** | 1460 | 1.767123 | 0.747315 | 0 | 1 | 2 | 2 | 4 |
| **GarageArea** | 1460 | 472.9801 | 213.8048 | 0 | 334.5 | 480 | 576 | 1418 |
| **WoodDeckSF** | 1460 | 94.24452 | 125.3388 | 0 | 0 | 0 | 168 | 857 |
| **OpenPorchSF** | 1460 | 46.66027 | 66.25603 | 0 | 0 | 25 | 68 | 547 |
| **EnclosedPorch** | 1460 | 21.95411 | 61.11915 | 0 | 0 | 0 | 0 | 552 |
| **3SsnPorch** | 1460 | 3.409589 | 29.31733 | 0 | 0 | 0 | 0 | 508 |
| **ScreenPorch** | 1460 | 15.06096 | 55.75742 | 0 | 0 | 0 | 0 | 480 |
| **PoolArea** | 1460 | 2.758904 | 40.17731 | 0 | 0 | 0 | 0 | 738 |
| **MiscVal** | 1460 | 43.48904 | 496.123 | 0 | 0 | 0 | 0 | 15500 |
| **MoSold** | 1460 | 6.321918 | 2.703626 | 1 | 5 | 6 | 8 | 12 |
| **YrSold** | 1460 | 2007.816 | 1.328095 | 2006 | 2007 | 2008 | 2009 | 2010 |
| **SalePrice** | 1460 | 180921.2 | 79442.5 | 34900 | 129975 | 163000 | 214000 | 755000 |

* null\_columns=house.columns[house.isnull().any()]
* list(null\_columns)

LotFrontage

Alley

MasVnrType

MasVnrArea

BsmtQual

BsmtCond

BsmtExposure

BsmtFinType1

BsmtFinType2

Electrical

FireplaceQu

GarageType

GarageYrBlt

GarageFinish

GarageQual

GarageCond

PoolQC

Fence

MiscFeature

* house[null\_columns].isnull().sum()\*100/len(house)

LotFrontage 17.739726

Alley 93.767123

MasVnrType 0.547945

MasVnrArea 0.547945

BsmtQual 2.534247

BsmtCond 2.534247

BsmtExposure 2.602740

BsmtFinType1 2.534247

BsmtFinType2 2.602740

Electrical 0.068493

FireplaceQu 47.260274

GarageType 5.547945

GarageYrBlt 5.547945

GarageFinish 5.547945

GarageQual 5.547945

GarageCond 5.547945

PoolQC 99.520548

Fence 80.753425

MiscFeature 96.301370

* for x in house.columns[house.isnull().any()]:

print('--------------------------------')

print('Column Name:',x)

print('--------------------------------')

print(house[x].value\_counts()\*100/len(house))

print(house[x].value\_counts())

--------------------------------

Column Name: LotFrontage

--------------------------------

60.0 9.794521

70.0 4.794521

80.0 4.726027

50.0 3.904110

75.0 3.630137

...

137.0 0.068493

141.0 0.068493

38.0 0.068493

140.0 0.068493

46.0 0.068493

Name: LotFrontage, Length: 110, dtype: float64

60.0 143

70.0 70

80.0 69

50.0 57

75.0 53

...

137.0 1

141.0 1

38.0 1

140.0 1

46.0 1

Name: LotFrontage, Length: 110, dtype: int64

--------------------------------

Column Name: Alley

--------------------------------

Grvl 3.424658

Pave 2.808219

Name: Alley, dtype: float64

Grvl 50

Pave 41

Name: Alley, dtype: int64

--------------------------------

Column Name: MasVnrType

--------------------------------

None 59.178082

BrkFace 30.479452

Stone 8.767123

BrkCmn 1.027397

Name: MasVnrType, dtype: float64

None 864

BrkFace 445

Stone 128

BrkCmn 15

Name: MasVnrType, dtype: int64

--------------------------------

Column Name: MasVnrArea

--------------------------------

0.0 58.972603

180.0 0.547945

72.0 0.547945

108.0 0.547945

120.0 0.479452

...

562.0 0.068493

89.0 0.068493

921.0 0.068493

762.0 0.068493

119.0 0.068493

Name: MasVnrArea, Length: 327, dtype: float64

0.0 861

180.0 8

72.0 8

108.0 8

120.0 7

...

562.0 1

89.0 1

921.0 1

762.0 1

119.0 1

Name: MasVnrArea, Length: 327, dtype: int64

--------------------------------

Column Name: BsmtQual

--------------------------------

TA 44.452055

Gd 42.328767

Ex 8.287671

Fa 2.397260

Name: BsmtQual, dtype: float64

TA 649

Gd 618

Ex 121

Fa 35

Name: BsmtQual, dtype: int64

--------------------------------

Column Name: BsmtCond

--------------------------------

TA 89.794521

Gd 4.452055

Fa 3.082192

Po 0.136986

Name: BsmtCond, dtype: float64

TA 1311

Gd 65

Fa 45

Po 2

Name: BsmtCond, dtype: int64

--------------------------------

Column Name: BsmtExposure

--------------------------------

No 65.273973

Av 15.136986

Gd 9.178082

Mn 7.808219

Name: BsmtExposure, dtype: float64

No 953

Av 221

Gd 134

Mn 114

Name: BsmtExposure, dtype: int64

--------------------------------

Column Name: BsmtFinType1

--------------------------------

Unf 29.452055

GLQ 28.630137

ALQ 15.068493

BLQ 10.136986

Rec 9.109589

LwQ 5.068493

Name: BsmtFinType1, dtype: float64

Unf 430

GLQ 418

ALQ 220

BLQ 148

Rec 133

LwQ 74

Name: BsmtFinType1, dtype: int64

--------------------------------

Column Name: BsmtFinType2

--------------------------------

Unf 86.027397

Rec 3.698630

LwQ 3.150685

BLQ 2.260274

ALQ 1.301370

GLQ 0.958904

Name: BsmtFinType2, dtype: float64

Unf 1256

Rec 54

LwQ 46

BLQ 33

ALQ 19

GLQ 14

Name: BsmtFinType2, dtype: int64

--------------------------------

Column Name: Electrical

--------------------------------

SBrkr 91.369863

FuseA 6.438356

FuseF 1.849315

FuseP 0.205479

Mix 0.068493

Name: Electrical, dtype: float64

SBrkr 1334

FuseA 94

FuseF 27

FuseP 3

Mix 1

Name: Electrical, dtype: int64

--------------------------------

Column Name: FireplaceQu

--------------------------------

Gd 26.027397

TA 21.438356

Fa 2.260274

Ex 1.643836

Po 1.369863

Name: FireplaceQu, dtype: float64

Gd 380

TA 313

Fa 33

Ex 24

Po 20

Name: FireplaceQu, dtype: int64

--------------------------------

Column Name: GarageType

--------------------------------

Attchd 59.589041

Detchd 26.506849

BuiltIn 6.027397

Basment 1.301370

CarPort 0.616438

2Types 0.410959

Name: GarageType, dtype: float64

Attchd 870

Detchd 387

BuiltIn 88

Basment 19

CarPort 9

2Types 6

Name: GarageType, dtype: int64

--------------------------------

Column Name: GarageYrBlt

--------------------------------

2005.0 4.452055

2006.0 4.041096

2004.0 3.630137

2003.0 3.424658

2007.0 3.356164

...

1927.0 0.068493

1900.0 0.068493

1906.0 0.068493

1908.0 0.068493

1933.0 0.068493

Name: GarageYrBlt, Length: 97, dtype: float64

2005.0 65

2006.0 59

2004.0 53

2003.0 50

2007.0 49

..

1927.0 1

1900.0 1

1906.0 1

1908.0 1

1933.0 1

Name: GarageYrBlt, Length: 97, dtype: int64

--------------------------------

Column Name: GarageFinish

--------------------------------

Unf 41.438356

RFn 28.904110

Fin 24.109589

Name: GarageFinish, dtype: float64

Unf 605

RFn 422

Fin 352

Name: GarageFinish, dtype: int64

--------------------------------

Column Name: GarageQual

--------------------------------

TA 89.794521

Fa 3.287671

Gd 0.958904

Ex 0.205479

Po 0.205479

Name: GarageQual, dtype: float64

TA 1311

Fa 48

Gd 14

Ex 3

Po 3

Name: GarageQual, dtype: int64

--------------------------------

Column Name: GarageCond

--------------------------------

TA 90.821918

Fa 2.397260

Gd 0.616438

Po 0.479452

Ex 0.136986

Name: GarageCond, dtype: float64

TA 1326

Fa 35

Gd 9

Po 7

Ex 2

Name: GarageCond, dtype: int64

--------------------------------

Column Name: PoolQC

--------------------------------

Gd 0.205479

Ex 0.136986

Fa 0.136986

Name: PoolQC, dtype: float64

Gd 3

Ex 2

Fa 2

Name: PoolQC, dtype: int64

--------------------------------

Column Name: Fence

--------------------------------

MnPrv 10.753425

GdPrv 4.041096

GdWo 3.698630

MnWw 0.753425

Name: Fence, dtype: float64

MnPrv 157

GdPrv 59

GdWo 54

MnWw 11

Name: Fence, dtype: int64

--------------------------------

Column Name: MiscFeature

--------------------------------

Shed 3.356164

Gar2 0.136986

Othr 0.136986

TenC 0.068493

Name: MiscFeature, dtype: float64

Shed 49

Gar2 2

Othr 2

TenC 1

Name: MiscFeature, dtype: int64