df\_train

Autogenerated data summary from dataMaid

2024-03-24 20:09:16.883747

# Data report overview

The dataset examined has the following dimensions:

| Feature | Result |
| --- | --- |
| Number of observations | 1460 |
| Number of variables | 81 |

### Checks performed

The following variable checks were performed, depending on the data type of each variable:

|  | character | factor | labelled | haven labelled | numeric | integer | logical | Date |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Identify miscoded missing values | × | × | × | × | × | × |  | × |
| Identify prefixed and suffixed whitespace | × | × | × | × |  |  |  |  |
| Identify levels with < 6 obs. | × | × | × | × |  |  |  |  |
| Identify case issues | × | × | × | × |  |  |  |  |
| Identify misclassified numeric or integer variables | × | × | × | × |  |  |  |  |
| Identify outliers |  |  |  |  | × | × |  | × |

Please note that all numerical values in the following have been rounded to 2 decimals.

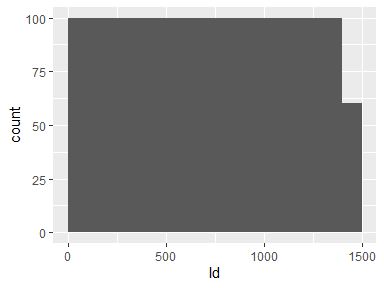
# Summary table

|  | Variable class | # unique values | Missing observations | Any problems? |
| --- | --- | --- | --- | --- |
| [Id](#id) | numeric | 1460 | 0.00 % |  |
| [MSSubClass](#mssubclass) | numeric | 15 | 0.00 % | × |
| [MSZoning](#mszoning) | character | 5 | 0.00 % |  |
| [LotFrontage](#lotfrontage) | numeric | 111 | 17.74 % | × |
| [LotArea](#lotarea) | numeric | 1073 | 0.00 % | × |
| [Street](#street) | character | 2 | 0.00 % |  |
| [Alley](#alley) | character | 3 | 93.77 % |  |
| [LotShape](#lotshape) | character | 4 | 0.00 % |  |
| [LandContour](#landcontour) | character | 4 | 0.00 % |  |
| [Utilities](#utilities) | character | 2 | 0.00 % | × |
| [LotConfig](#lotconfig) | character | 5 | 0.00 % | × |
| [LandSlope](#landslope) | character | 3 | 0.00 % |  |
| [Neighborhood](#neighborhood) | character | 25 | 0.00 % | × |
| [Condition1](#condition1) | character | 9 | 0.00 % | × |
| [Condition2](#condition2) | character | 8 | 0.00 % | × |
| [BldgType](#bldgtype) | character | 5 | 0.00 % |  |
| [HouseStyle](#housestyle) | character | 8 | 0.00 % |  |
| [OverallQual](#overallqual) | numeric | 10 | 0.00 % | × |
| [OverallCond](#overallcond) | numeric | 9 | 0.00 % | × |
| [YearBuilt](#yearbuilt) | numeric | 112 | 0.00 % | × |
| [YearRemodAdd](#yearremodadd) | numeric | 61 | 0.00 % |  |
| [RoofStyle](#roofstyle) | character | 6 | 0.00 % | × |
| [RoofMatl](#roofmatl) | character | 8 | 0.00 % | × |
| [Exterior1st](#exterior1st) | character | 15 | 0.00 % | × |
| [Exterior2nd](#exterior2nd) | character | 16 | 0.00 % | × |
| [MasVnrType](#masvnrtype) | character | 5 | 0.55 % |  |
| [MasVnrArea](#masvnrarea) | numeric | 328 | 0.55 % |  |
| [ExterQual](#exterqual) | character | 4 | 0.00 % |  |
| [ExterCond](#extercond) | character | 5 | 0.00 % | × |
| [Foundation](#foundation) | character | 6 | 0.00 % | × |
| [BsmtQual](#bsmtqual) | character | 5 | 2.53 % |  |
| [BsmtCond](#bsmtcond) | character | 5 | 2.53 % | × |
| [BsmtExposure](#bsmtexposure) | character | 5 | 2.60 % |  |
| [BsmtFinType1](#bsmtfintype1) | character | 7 | 2.53 % |  |
| [BsmtFinSF1](#bsmtfinsf1) | numeric | 637 | 0.00 % | × |
| [BsmtFinType2](#bsmtfintype2) | character | 7 | 2.60 % |  |
| [BsmtFinSF2](#bsmtfinsf2) | numeric | 144 | 0.00 % | × |
| [BsmtUnfSF](#bsmtunfsf) | numeric | 780 | 0.00 % |  |
| [TotalBsmtSF](#totalbsmtsf) | numeric | 721 | 0.00 % | × |
| [Heating](#heating) | character | 6 | 0.00 % | × |
| [HeatingQC](#heatingqc) | character | 5 | 0.00 % | × |
| [CentralAir](#centralair) | character | 2 | 0.00 % |  |
| [Electrical](#electrical) | character | 6 | 0.07 % | × |
| [1stFlrSF](#stflrsf) | numeric | 753 | 0.00 % | × |
| [2ndFlrSF](#ndflrsf) | numeric | 417 | 0.00 % |  |
| [LowQualFinSF](#lowqualfinsf) | numeric | 24 | 0.00 % | × |
| [GrLivArea](#grlivarea) | numeric | 861 | 0.00 % | × |
| [BsmtFullBath](#bsmtfullbath) | numeric | 4 | 0.00 % | × |
| [BsmtHalfBath](#bsmthalfbath) | numeric | 3 | 0.00 % | × |
| [FullBath](#fullbath) | numeric | 4 | 0.00 % |  |
| [HalfBath](#halfbath) | numeric | 3 | 0.00 % |  |
| [BedroomAbvGr](#bedroomabvgr) | numeric | 8 | 0.00 % | × |
| [KitchenAbvGr](#kitchenabvgr) | numeric | 4 | 0.00 % | × |
| [KitchenQual](#kitchenqual) | character | 4 | 0.00 % |  |
| [TotRmsAbvGrd](#totrmsabvgrd) | numeric | 12 | 0.00 % | × |
| [Functional](#functional) | character | 7 | 0.00 % | × |
| [Fireplaces](#fireplaces) | numeric | 4 | 0.00 % | × |
| [FireplaceQu](#fireplacequ) | character | 6 | 47.26 % |  |
| [GarageType](#garagetype) | character | 7 | 5.55 % |  |
| [GarageYrBlt](#garageyrblt) | numeric | 98 | 5.55 % | × |
| [GarageFinish](#garagefinish) | character | 4 | 5.55 % |  |
| [GarageCars](#garagecars) | numeric | 5 | 0.00 % | × |
| [GarageArea](#garagearea) | numeric | 441 | 0.00 % | × |
| [GarageQual](#garagequal) | character | 6 | 5.55 % | × |
| [GarageCond](#garagecond) | character | 6 | 5.55 % | × |
| [PavedDrive](#paveddrive) | character | 3 | 0.00 % |  |
| [WoodDeckSF](#wooddecksf) | numeric | 274 | 0.00 % |  |
| [OpenPorchSF](#openporchsf) | numeric | 202 | 0.00 % | × |
| [EnclosedPorch](#enclosedporch) | numeric | 120 | 0.00 % | × |
| [3SsnPorch](#ssnporch) | numeric | 20 | 0.00 % | × |
| [ScreenPorch](#screenporch) | numeric | 76 | 0.00 % | × |
| [PoolArea](#poolarea) | numeric | 8 | 0.00 % | × |
| [PoolQC](#poolqc) | character | 4 | 99.52 % | × |
| [Fence](#fence) | character | 5 | 80.75 % |  |
| [MiscFeature](#miscfeature) | character | 5 | 96.30 % | × |
| [MiscVal](#miscval) | numeric | 21 | 0.00 % | × |
| [MoSold](#mosold) | numeric | 12 | 0.00 % |  |
| [YrSold](#yrsold) | numeric | 5 | 0.00 % |  |
| [SaleType](#saletype) | character | 9 | 0.00 % | × |
| [SaleCondition](#salecondition) | character | 6 | 0.00 % | × |
| [SalePrice](#saleprice) | numeric | 663 | 0.00 % | × |

# Variable list

## Id

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 1460 |
| Median | 730.5 |
| 1st and 3rd quartiles | 365.75; 1095.25 |
| Min. and max. | 1; 1460 |



## MSSubClass

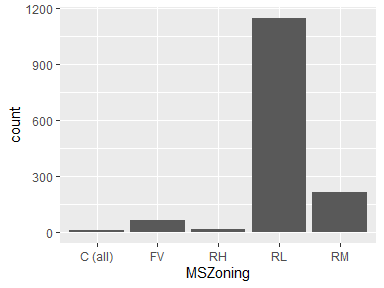
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 15 |
| Median | 50 |
| 1st and 3rd quartiles | 20; 70 |
| Min. and max. | 20; 190 |



* Note that the following possible outlier values were detected: "160", "180", "190".

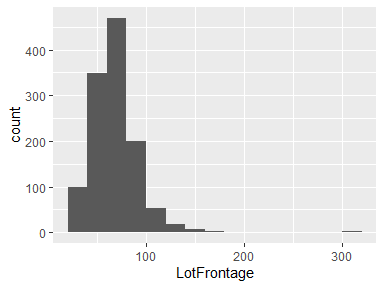
## MSZoning

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “RL” |



## LotFrontage

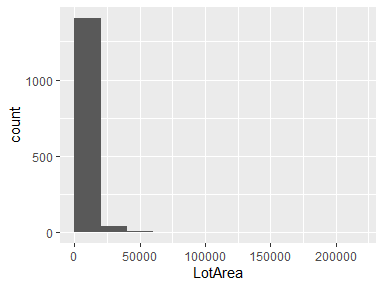
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 259 (17.74 %) |
| Number of unique values | 110 |
| Median | 69 |
| 1st and 3rd quartiles | 59; 80 |
| Min. and max. | 21; 313 |



* Note that the following possible outlier values were detected: "21", "24", "112", "114", "115", …, "160", "168", "174", "182", "313" (19 values omitted).

## LotArea

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 1073 |
| Median | 9478.5 |
| 1st and 3rd quartiles | 7553.5; 11601.5 |
| Min. and max. | 1300; 215245 |



* Note that the following possible outlier values were detected: "1300", "1477", "1491", "1526", "1533", …, "70761", "115149", "159000", "164660", "215245" (83 values omitted).

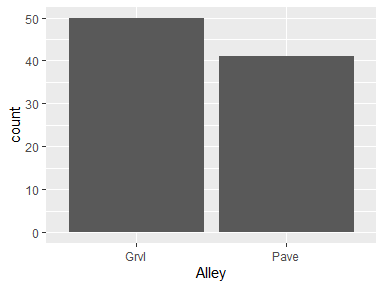
## Street

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 2 |
| Mode | “Pave” |



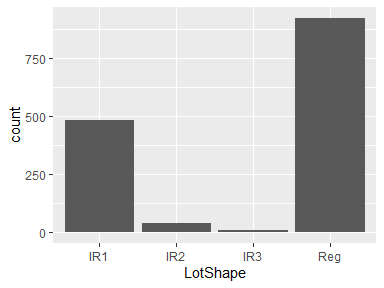
## Alley

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 1369 (93.77 %) |
| Number of unique values | 2 |
| Mode | “Grvl” |



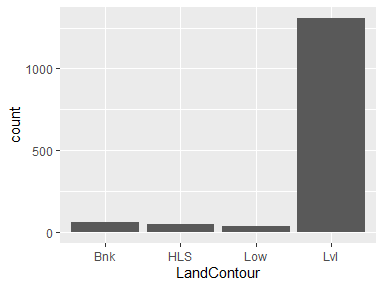
## LotShape

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 4 |
| Mode | “Reg” |



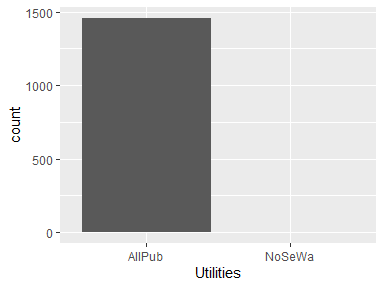
## LandContour

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 4 |
| Mode | “Lvl” |



## Utilities

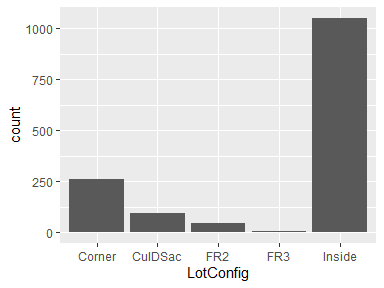
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 2 |
| Mode | “AllPub” |



* Note that the following levels have at most five observations: "NoSeWa".

## LotConfig

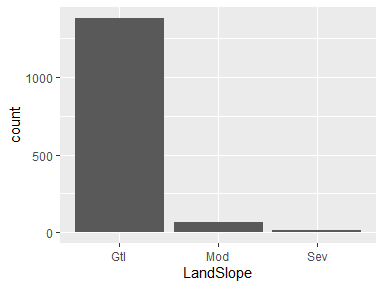
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “Inside” |



* Note that the following levels have at most five observations: "FR3".

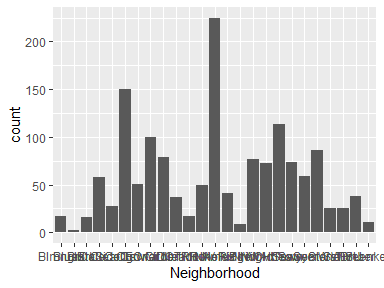
## LandSlope

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 3 |
| Mode | “Gtl” |



## Neighborhood

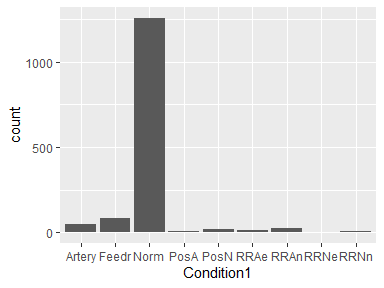
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 25 |
| Mode | “NAmes” |



* Note that the following levels have at most five observations: "Blueste".

## Condition1

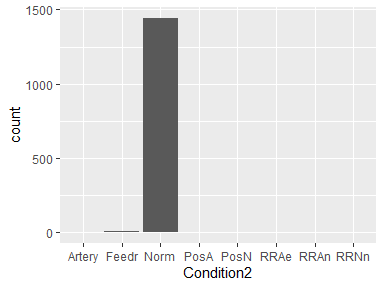
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 9 |
| Mode | “Norm” |



* Note that the following levels have at most five observations: "RRNe", "RRNn".

## Condition2

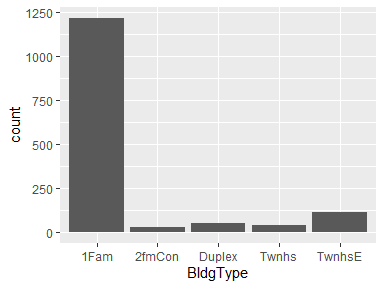
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 8 |
| Mode | “Norm” |



* Note that the following levels have at most five observations: "Artery", "PosA", "PosN", "RRAe", "RRAn", "RRNn".

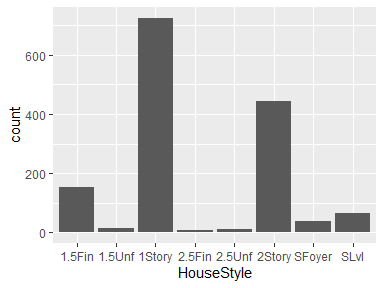
## BldgType

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “1Fam” |



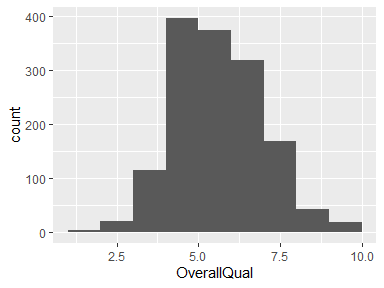
## HouseStyle

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 8 |
| Mode | “1Story” |



## OverallQual

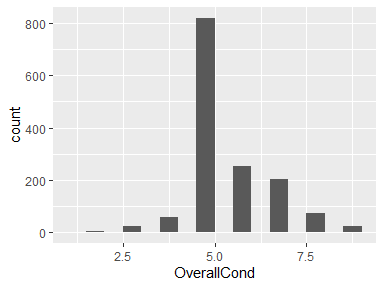
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 10 |
| Median | 6 |
| 1st and 3rd quartiles | 5; 7 |
| Min. and max. | 1; 10 |



* Note that the following possible outlier values were detected: "1".

## OverallCond

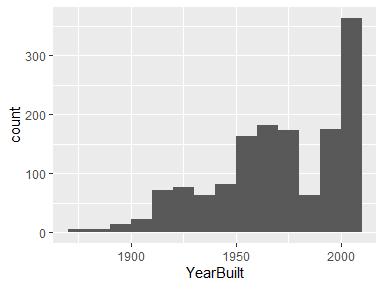
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 9 |
| Median | 5 |
| 1st and 3rd quartiles | 5; 6 |
| Min. and max. | 1; 9 |



* Note that the following possible outlier values were detected: "1", "2", "3", "4".

## YearBuilt

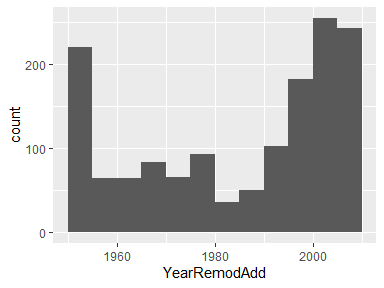
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 112 |
| Median | 1973 |
| 1st and 3rd quartiles | 1954; 2000 |
| Min. and max. | 1872; 2010 |



* Note that the following possible outlier values were detected: "1872", "1875", "1880", "1882", "1885".

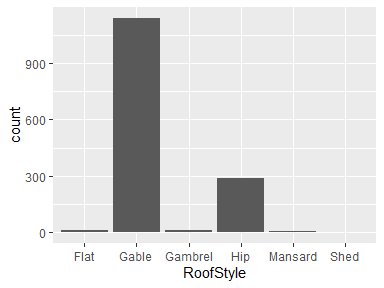
## YearRemodAdd

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 61 |
| Median | 1994 |
| 1st and 3rd quartiles | 1967; 2004 |
| Min. and max. | 1950; 2010 |



## RoofStyle

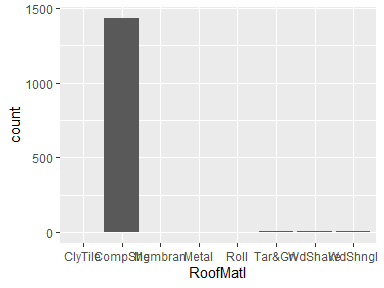
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 6 |
| Mode | “Gable” |



* Note that the following levels have at most five observations: "Shed".

## RoofMatl

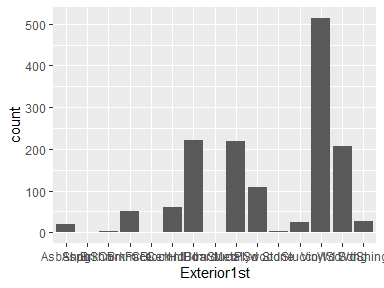
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 8 |
| Mode | “CompShg” |



* Note that the following levels have at most five observations: "ClyTile", "Membran", "Metal", "Roll", "WdShake".

## Exterior1st

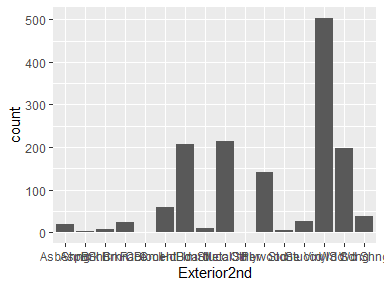
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 15 |
| Mode | “VinylSd” |



* Note that the following levels have at most five observations: "AsphShn", "BrkComm", "CBlock", "ImStucc", "Stone".

## Exterior2nd

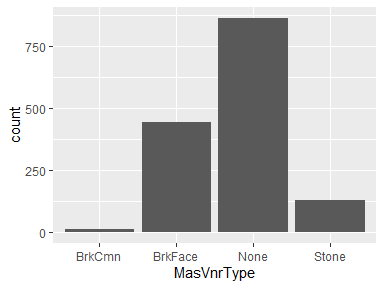
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 16 |
| Mode | “VinylSd” |



* Note that the following levels have at most five observations: "AsphShn", "CBlock", "Other", "Stone".

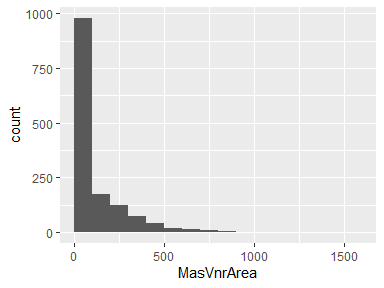
## MasVnrType

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 8 (0.55 %) |
| Number of unique values | 4 |
| Mode | “None” |



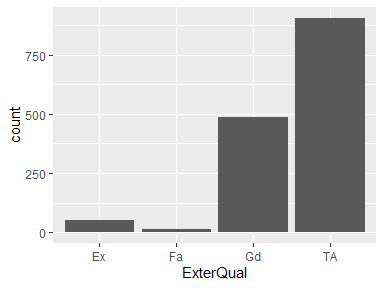
## MasVnrArea

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 8 (0.55 %) |
| Number of unique values | 327 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 166 |
| Min. and max. | 0; 1600 |



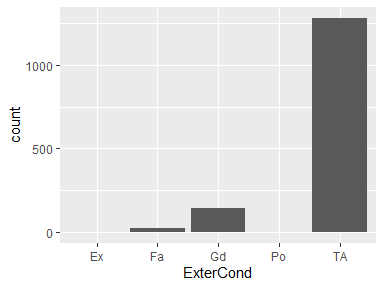
## ExterQual

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 4 |
| Mode | “TA” |



## ExterCond

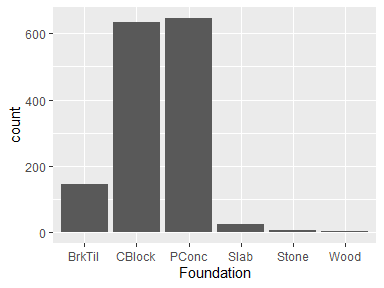
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “TA” |



* Note that the following levels have at most five observations: "Ex", "Po".

## Foundation

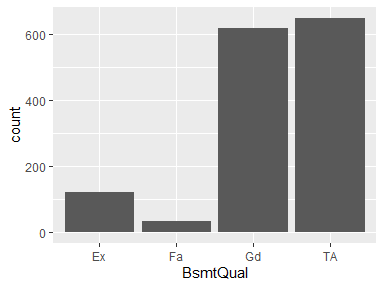
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 6 |
| Mode | “PConc” |



* Note that the following levels have at most five observations: "Wood".

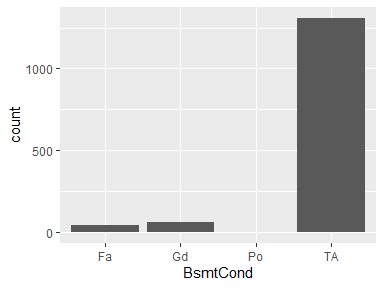
## BsmtQual

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 37 (2.53 %) |
| Number of unique values | 4 |
| Mode | “TA” |



## BsmtCond

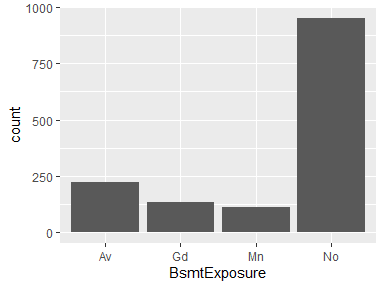
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 37 (2.53 %) |
| Number of unique values | 4 |
| Mode | “TA” |



* Note that the following levels have at most five observations: "Po".

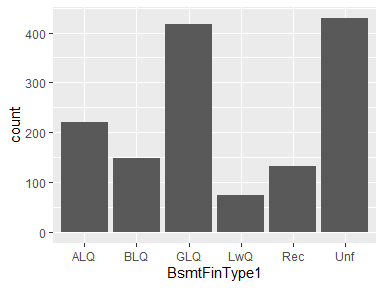
## BsmtExposure

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 38 (2.6 %) |
| Number of unique values | 4 |
| Mode | “No” |



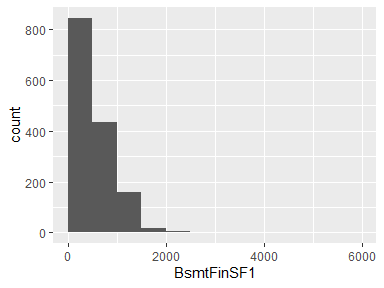
## BsmtFinType1

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 37 (2.53 %) |
| Number of unique values | 6 |
| Mode | “Unf” |



## BsmtFinSF1

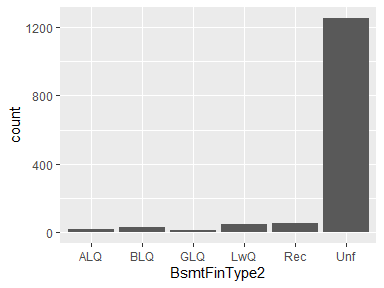
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 637 |
| Median | 383.5 |
| 1st and 3rd quartiles | 0; 712.25 |
| Min. and max. | 0; 5644 |



* Note that the following possible outlier values were detected: "2096", "2188", "2260", "5644".

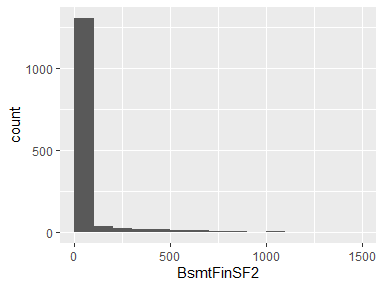
## BsmtFinType2

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 38 (2.6 %) |
| Number of unique values | 6 |
| Mode | “Unf” |



## BsmtFinSF2

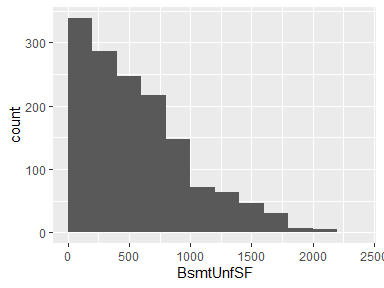
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 144 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 0 |
| Min. and max. | 0; 1474 |



* Note that the following possible outlier values were detected: "28", "32", "35", "40", "41", …, "1080", "1085", "1120", "1127", "1474" (133 values omitted).

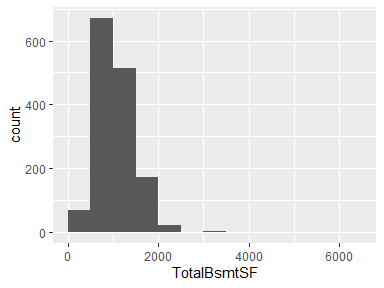
## BsmtUnfSF

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 780 |
| Median | 477.5 |
| 1st and 3rd quartiles | 223; 808 |
| Min. and max. | 0; 2336 |



## TotalBsmtSF

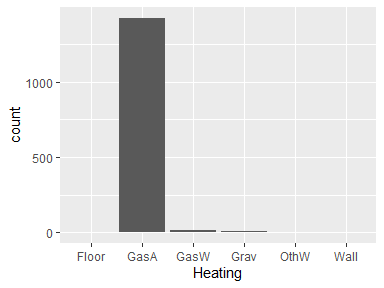
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 721 |
| Median | 991.5 |
| 1st and 3rd quartiles | 795.75; 1298.25 |
| Min. and max. | 0; 6110 |



* Note that the following possible outlier values were detected: "0", "105", "190", "264", "270", …, "3094", "3138", "3200", "3206", "6110" (10 values omitted).

## Heating

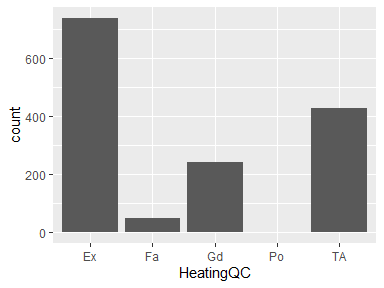
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 6 |
| Mode | “GasA” |



* Note that the following levels have at most five observations: "Floor", "OthW", "Wall".

## HeatingQC

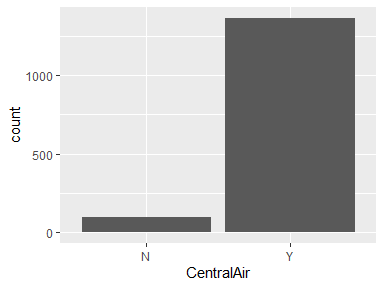
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “Ex” |



* Note that the following levels have at most five observations: "Po".

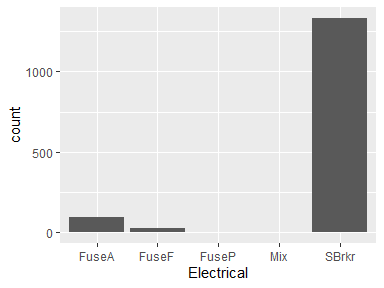
## CentralAir

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 2 |
| Mode | “Y” |



## Electrical

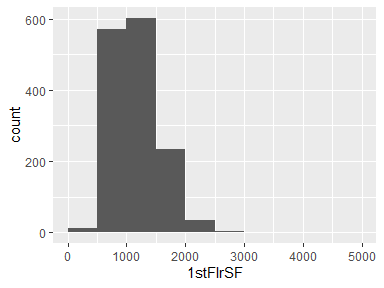
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 1 (0.07 %) |
| Number of unique values | 5 |
| Mode | “SBrkr” |



* Note that the following levels have at most five observations: "FuseP", "Mix".

## 1stFlrSF

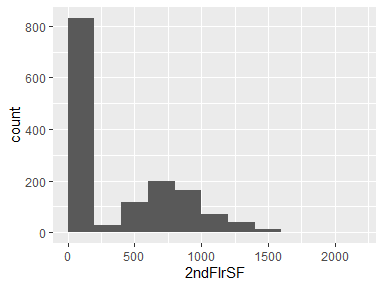
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 753 |
| Median | 1087 |
| 1st and 3rd quartiles | 882; 1391.25 |
| Min. and max. | 334; 4692 |



* Note that the following possible outlier values were detected: "334", "372", "438", "480", "483", …, "551", "2898", "3138", "3228", "4692" (6 values omitted).

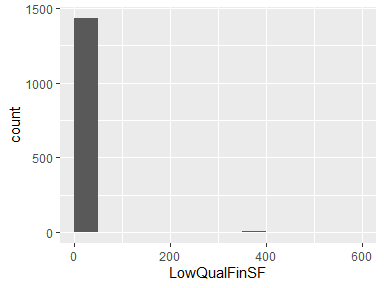
## 2ndFlrSF

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 417 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 728 |
| Min. and max. | 0; 2065 |



## LowQualFinSF

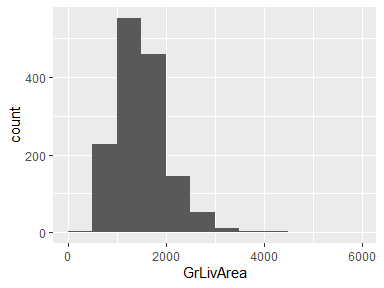
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 24 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 0 |
| Min. and max. | 0; 572 |



* Note that the following possible outlier values were detected: "53", "80", "120", "144", "156", …, "513", "514", "515", "528", "572" (13 values omitted).

## GrLivArea

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 861 |
| Median | 1464 |
| 1st and 3rd quartiles | 1129.5; 1776.75 |
| Min. and max. | 334; 5642 |

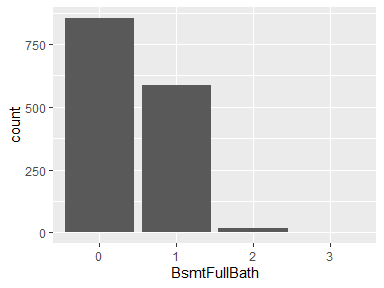


* Note that the following possible outlier values were detected: "2872", "2898", "2945", "2978", "3082", …, "3627", "4316", "4476", "4676", "5642" (12 values omitted).

## BsmtFullBath

* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 4 |
| Mode | “0” |
| Reference category | 0 |

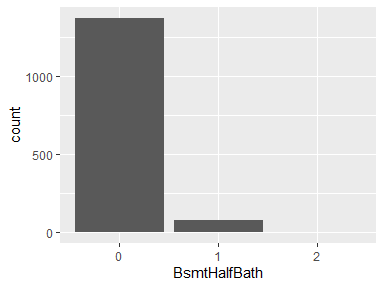


* Note that the following levels have at most five observations: "3".

## BsmtHalfBath

* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 3 |
| Mode | “0” |
| Reference category | 0 |

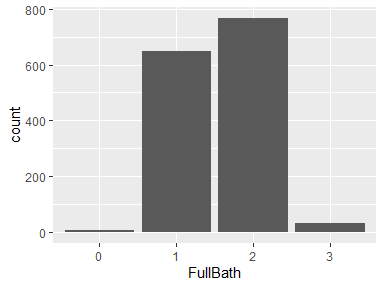


* Note that the following levels have at most five observations: "2".

## FullBath

* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

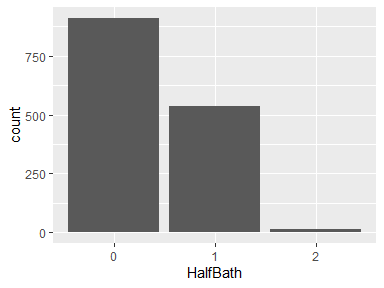
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 4 |
| Mode | “2” |
| Reference category | 0 |



## HalfBath

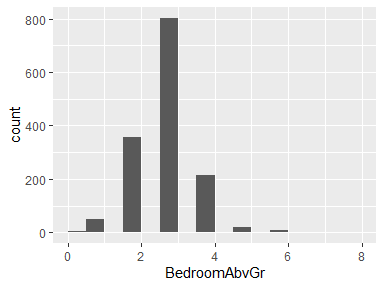
* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 3 |
| Mode | “0” |
| Reference category | 0 |



## BedroomAbvGr

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 8 |
| Median | 3 |
| 1st and 3rd quartiles | 2; 3 |
| Min. and max. | 0; 8 |

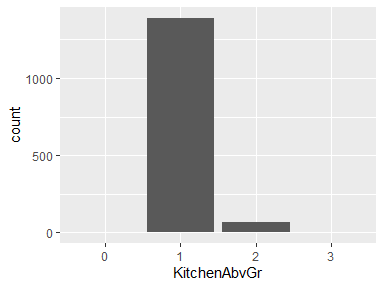


* The following suspected missing value codes enter as regular values: "8".
* Note that the following possible outlier values were detected: "4", "5", "6", "8".

## KitchenAbvGr

* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

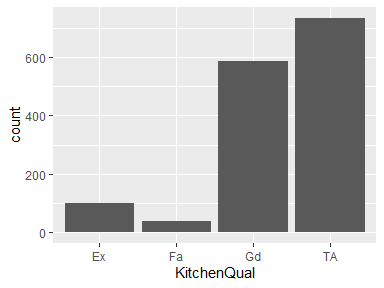
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 4 |
| Mode | “1” |
| Reference category | 0 |



* Note that the following levels have at most five observations: "0", "3".

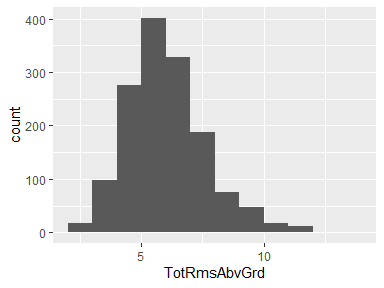
## KitchenQual

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 4 |
| Mode | “TA” |



## TotRmsAbvGrd

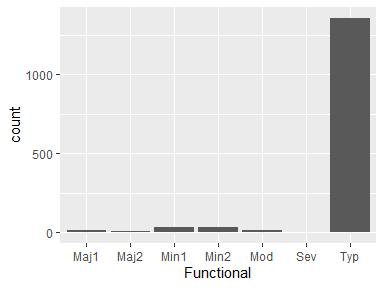
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 12 |
| Median | 6 |
| 1st and 3rd quartiles | 5; 7 |
| Min. and max. | 2; 14 |



* Note that the following possible outlier values were detected: "2", "3", "4".

## Functional

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 7 |
| Mode | “Typ” |

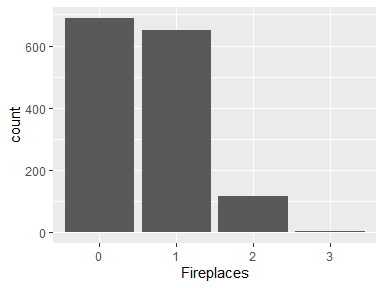


* Note that the following levels have at most five observations: "Maj2", "Sev".

## Fireplaces

* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

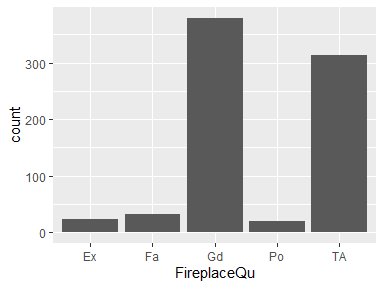
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 4 |
| Mode | “0” |
| Reference category | 0 |



* Note that the following levels have at most five observations: "3".

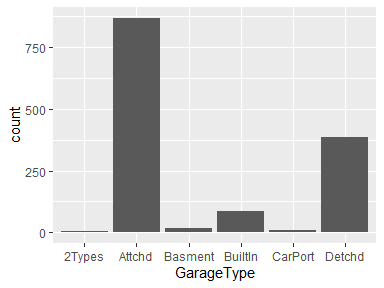
## FireplaceQu

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 690 (47.26 %) |
| Number of unique values | 5 |
| Mode | “Gd” |



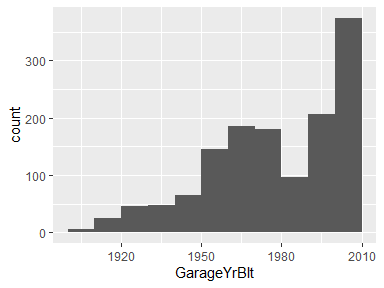
## GarageType

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 81 (5.55 %) |
| Number of unique values | 6 |
| Mode | “Attchd” |



## GarageYrBlt

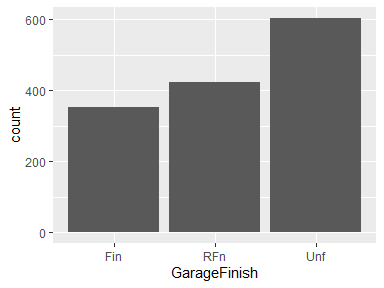
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 81 (5.55 %) |
| Number of unique values | 97 |
| Median | 1980 |
| 1st and 3rd quartiles | 1961; 2002 |
| Min. and max. | 1900; 2010 |



* Note that the following possible outlier values were detected: "1900".

## GarageFinish

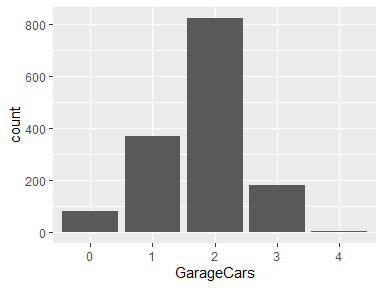
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 81 (5.55 %) |
| Number of unique values | 3 |
| Mode | “Unf” |



## GarageCars

* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

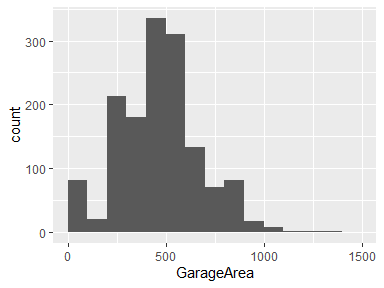
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “2” |
| Reference category | 0 |



* Note that the following levels have at most five observations: "4".

## GarageArea

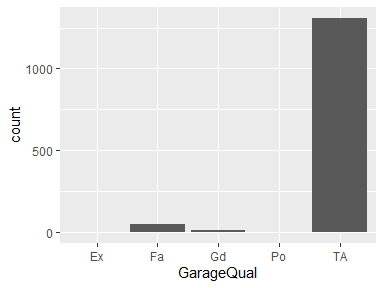
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 441 |
| Median | 480 |
| 1st and 3rd quartiles | 334.5; 576 |
| Min. and max. | 0; 1418 |



* Note that the following possible outlier values were detected: "852", "853", "856", "857", "858", …, "1220", "1248", "1356", "1390", "1418" (42 values omitted).

## GarageQual

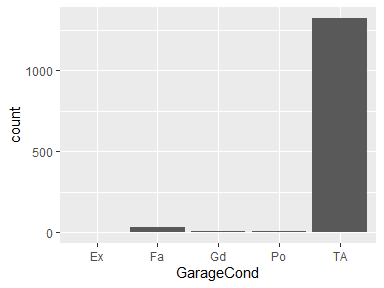
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 81 (5.55 %) |
| Number of unique values | 5 |
| Mode | “TA” |



* Note that the following levels have at most five observations: "Ex", "Po".

## GarageCond

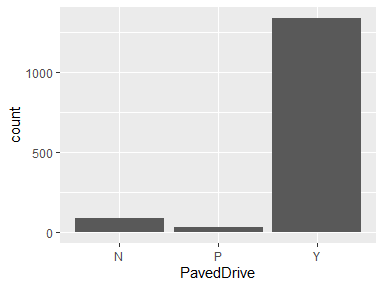
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 81 (5.55 %) |
| Number of unique values | 5 |
| Mode | “TA” |



* Note that the following levels have at most five observations: "Ex".

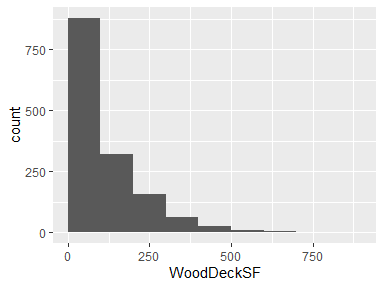
## PavedDrive

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 3 |
| Mode | “Y” |



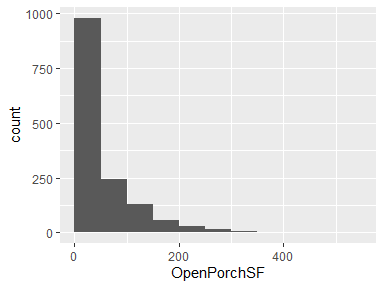
## WoodDeckSF

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 274 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 168 |
| Min. and max. | 0; 857 |



## OpenPorchSF

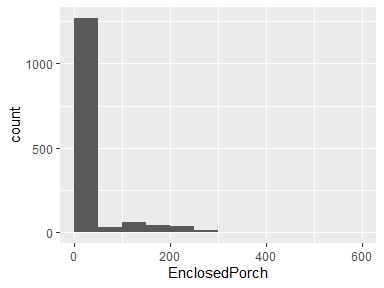
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 202 |
| Median | 25 |
| 1st and 3rd quartiles | 0; 68 |
| Min. and max. | 0; 547 |



* Note that the following possible outlier values were detected: "341", "364", "406", "418", "502", "523", "547".

## EnclosedPorch

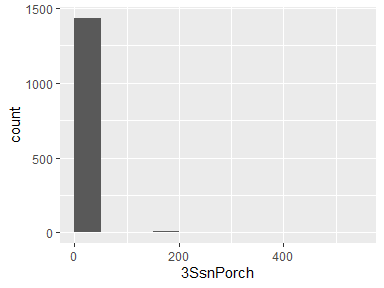
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 120 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 0 |
| Min. and max. | 0; 552 |



* Note that the following possible outlier values were detected: "19", "20", "24", "30", "32", …, "301", "318", "330", "386", "552" (109 values omitted).

## 3SsnPorch

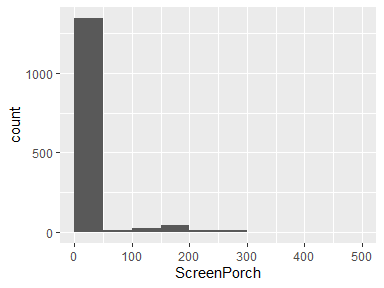
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 20 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 0 |
| Min. and max. | 0; 508 |



* Note that the following possible outlier values were detected: "23", "96", "130", "140", "144", …, "290", "304", "320", "407", "508" (9 values omitted).

## ScreenPorch

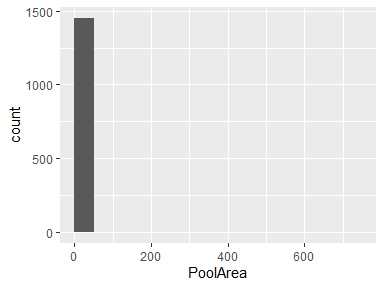
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 76 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 0 |
| Min. and max. | 0; 480 |



* Note that the following possible outlier values were detected: "40", "53", "60", "63", "80", …, "385", "396", "410", "440", "480" (65 values omitted).

## PoolArea

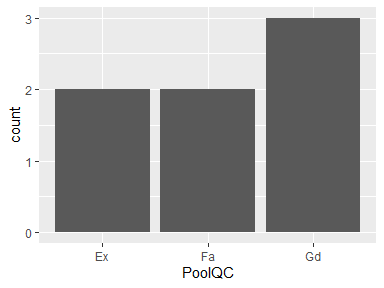
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 8 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 0 |
| Min. and max. | 0; 738 |



* Note that the following possible outlier values were detected: "480", "512", "519", "555", "576", "648", "738".

## PoolQC

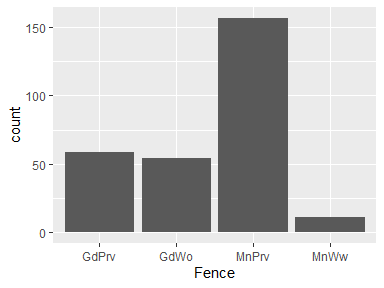
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 1453 (99.52 %) |
| Number of unique values | 3 |
| Mode | “Gd” |



* Note that the following levels have at most five observations: "Ex", "Fa", "Gd".

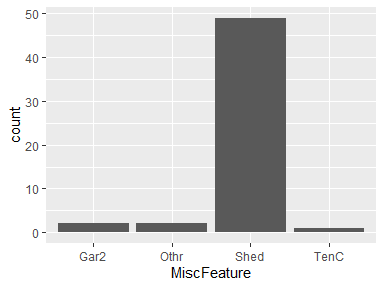
## Fence

| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 1179 (80.75 %) |
| Number of unique values | 4 |
| Mode | “MnPrv” |



## MiscFeature

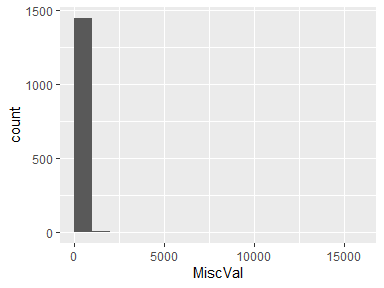
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 1406 (96.3 %) |
| Number of unique values | 4 |
| Mode | “Shed” |



* Note that the following levels have at most five observations: "Gar2", "Othr", "TenC".

## MiscVal

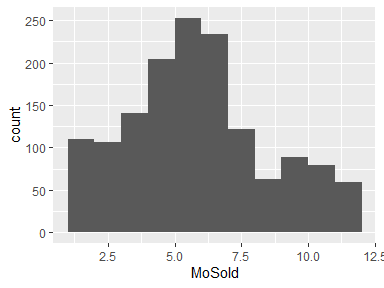
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 21 |
| Median | 0 |
| 1st and 3rd quartiles | 0; 0 |
| Min. and max. | 0; 15500 |



* Note that the following possible outlier values were detected: "54", "350", "400", "450", "480", …, "2000", "2500", "3500", "8300", "15500" (10 values omitted).

## MoSold

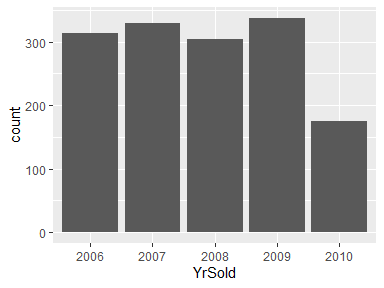
| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 12 |
| Median | 6 |
| 1st and 3rd quartiles | 5; 8 |
| Min. and max. | 1; 12 |



## YrSold

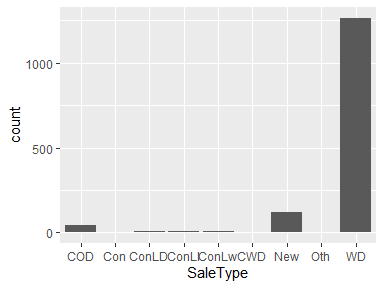
* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “2009” |
| Reference category | 2006 |



## SaleType

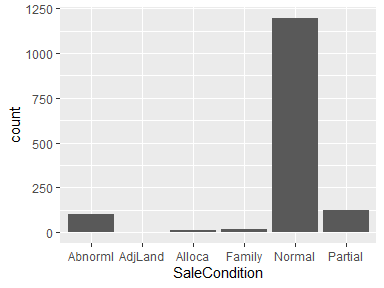
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 9 |
| Mode | “WD” |



* Note that the following levels have at most five observations: "Con", "ConLI", "ConLw", "CWD", "Oth".

## SaleCondition

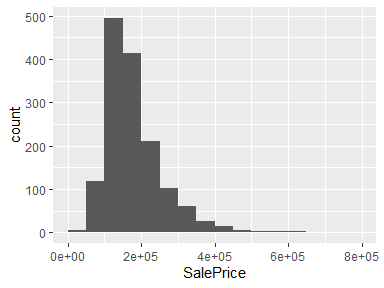
| Feature | Result |
| --- | --- |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 6 |
| Mode | “Normal” |



* Note that the following levels have at most five observations: "AdjLand".

## SalePrice

| Feature | Result |
| --- | --- |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 663 |
| Median | 163000 |
| 1st and 3rd quartiles | 129975; 214000 |
| Min. and max. | 34900; 755000 |



* Note that the following possible outlier values were detected: "34900", "35311", "37900", "39300", "40000", …, "582933", "611657", "625000", "745000", "755000" (28 values omitted).

Report generation information:

* Created by: Andre Rizzo (username: andre.rizzo).
* Report creation time: dom mar 24 2024 20:09:20
* Report was run from directory: C:/Users/andre.rizzo/OneDrive/Personal Projects/House prices/scripts/r
* dataMaid v1.4.1 [Pkg: 2021-10-08 from CRAN (R 4.3.1)]
* R version 4.3.3 (2024-02-29 ucrt).
* Platform: x86\_64-w64-mingw32/x64 (64-bit)(America/Sao\_Paulo).
* Function call: makeDataReport(data = df\_train)