RFinalHYu

Haozhe (Jerry) Yu

2023-04-26

Data Import

Read in Raw Data

Read in the raw data directly form the url.

```
rawhouse <- read.csv("https://www4.stat.ncsu.edu/~online/ST308/Data/hyu23_house.csv")
```

Data Subsetting

Create a tibble from the read in data table with the following modifications:

1. Remove any observations where
    - the `SaleType` variable takes the value "Other" or
    - the `BedroomAbvGr` variable takes on a value less than or equal to 2
2. Create a new variable with a name of your choosing that is the `SalePrice` variable divided by 100000.
3. The `GarageArea` and `MSZoning` variables are removed

```
House <- rawhouse %>%
         filter(SaleType != "Other") %>%
         filter(BedroomAbvGr > 2) %>%
         mutate(SalePrice100k = SalePrice/100000) %>%
         select(-GarageArea, -MSZoning)
```

Now print out the first 10 observations and first 6 variables of House.

```
knitr::kable(
  head(House, 10),
  colnames( c( "SalePrice",
               "BsmtUnfSF",
               "OverallQual",
               "OpenPorchSF",
               "BedroomAbvGr",
               "YrSold"))
)
```

SalePrice	BsmtUnfSF	OverallQual	OpenPorchSF	BedroomAbvGr	YrSold	BsmtCond	BsmtFinType2	LotShape	GarageCond	SaleType	SalePrice100k
208500	150	7	61	3	2008	TA	Unf	Reg	TA	WD	2.085
181500	284	6	0	3	2007	TA	Unf	Reg	TA	WD	1.815
223500	434	7	42	3	2008	TA	Unf	IR1	TA	WD	2.235
140000	540	7	35	3	2006	Gd	Unf	IR1	TA	WD	1.400
250000	490	8	84	4	2008	TA	Unf	IR1	TA	WD	2.500
307000	317	8	57	3	2007	TA	Unf	Reg	TA	WD	3.070
200000	216	7	204	3	2009	TA	Other	IR1	TA	WD	2.000
279500	1494	7	33	3	2007	TA	Unf	IR1	TA	New	2.795
159000	468	5	102	3	2008	TA	Unf	Reg	TA	WD	1.590
139000	525	5	0	3	2009	TA	Unf	Reg	TA	COD	1.390