

# USA Housing

Peace Peter

7/15/2021

## Importing the data

```
USA_Housing <- read.csv("C:/Users/PEACE/Desktop/yyy/USA_Housing.csv")
library(dplyr)
## Warning: package 'dplyr' was built under R version 3.6.3
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
library(caTools)
## Warning: package 'caTools' was built under R version 3.6.3
glimpse(USA_Housing)
## Rows: 5,000
## Columns: 7
## $ Avg..Area.Income      <dbl> 79545.46, 79248.64, 61287.07, 63345
##   .24, 5~
## $ Avg..Area.House.Age   <dbl> 5.682861, 6.002900, 5.865890, 7.188
##   236, 5~
## $ Avg..Area.Number.of.Rooms <dbl> 7.009188, 6.730821, 8.512727, 5.586
##   729, 7~
## $ Avg..Area.Number.of.Bedrooms <dbl> 4.09, 3.09, 5.13, 3.26, 4.23, 4.04,
##   3.41,~
## $ Area.Population       <dbl> 23086.80, 40173.07, 36882.16, 34310
##   .24, 2~
## $ Price                 <dbl> 1059033.6, 1505890.9, 1058988.0, 12
##   60616.~
## $ Address               <fct> "208 Michael Ferry Apt. 674\nLaurab
##   ury, N~
```

# Removing unwanted rows

```
Housing <- select(USA_Housing, -c(Address,Avg..Area.House.Age,Avg..Area.Number.of.Rooms))
```

```
head(Housing)
```

##	Avg..Area.Income	Avg..Area.Number.of.Bedrooms	Area.Population	Price
## 1	79545.46	4.09	23086.80	1059033.6
## 2	79248.64	3.09	40173.07	1505890.9
## 3	61287.07	5.13	36882.16	1058988.0
## 4	63345.24	3.26	34310.24	1260616.8
## 5	59982.20	4.23	26354.11	630943.5
## 6	80175.75	4.04	26748.43	1068138.1

```
glimpse(Housing)
```

```
## Rows: 5,000
## Columns: 4
## $ Avg..Area.Income      <dbl> 79545.46, 79248.64, 61287.07, 63345.24, 5~
## $ Avg..Area.Number.of.Bedrooms <dbl> 4.09, 3.09, 5.13, 3.26, 4.23, 4.04, 3.41, ~
## $ Area.Population       <dbl> 23086.80, 40173.07, 36882.16, 34310.24, 2~
## $ Price                 <dbl> 1059033.6, 1505890.9, 1058988.0, 1260616.~
```

```
summary(Housing)
```

##	Avg..Area.Income	Avg..Area.Number.of.Bedrooms	Area.Population
## Min.	: 17797	Min. :2.000	Min. : 172.6
## 1st Qu.:	61481	1st Qu.:3.140	1st Qu.:29403.9
## Median :	68804	Median :4.050	Median :36199.4
## Mean :	68583	Mean :3.981	Mean :36163.5
## 3rd Qu.:	75783	3rd Qu.:4.490	3rd Qu.:42861.3
## Max. :	107702	Max. :6.500	Max. :69621.7
##	Price		
## Min. :	15939		
## 1st Qu.:	997577		

```
## Median :1232669
## Mean   :1232073
## 3rd Qu.:1471210
## Max.    :2469066
```

## Splitting the data into training and testing data set

```
set.seed(2)
split <- sample.split(Housing, SplitRatio = 0.7)
split
## [1] TRUE FALSE FALSE TRUE
train <- subset(Housing, split="TRUE")
test <- subset(Housing, split="FALSE")
```

## Creating the model

```
Model <- lm(Price ~.,data=train)
summary(Model)
##
## Call:
## lm(formula = Price ~ ., data = train)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-804919	-149650	-7095	152555	862200

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-9.645e+05	2.536e+04	-38.03	<2e-16 ***
Avg..Area.Income	2.131e+01	2.917e-01	73.06	<2e-16 ***
Avg..Area.Number.of.Bedrooms	4.799e+04	2.520e+03	19.05	<2e-16 ***
Area.Population	1.504e+01	3.133e-01	48.01	<2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 219800 on 4996 degrees of freedom
## Multiple R-squared:  0.6129, Adjusted R-squared:  0.6127
## F-statistic: 2637 on 3 and 4996 DF,  p-value: < 2.2e-16

pred <- predict(Model, test)
```

## Validation 1

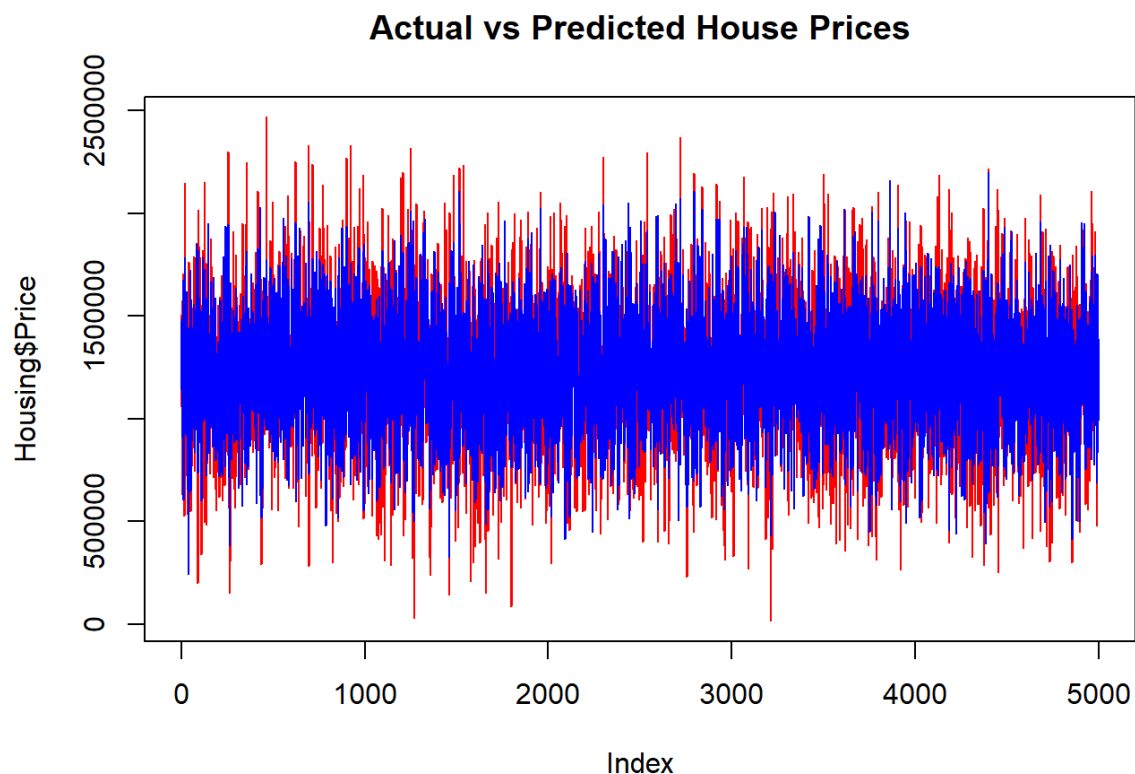
```
rmse <- sqrt(mean(pred-Housing$Price)^2)

rmse

## [1] 1.431299e-09
```

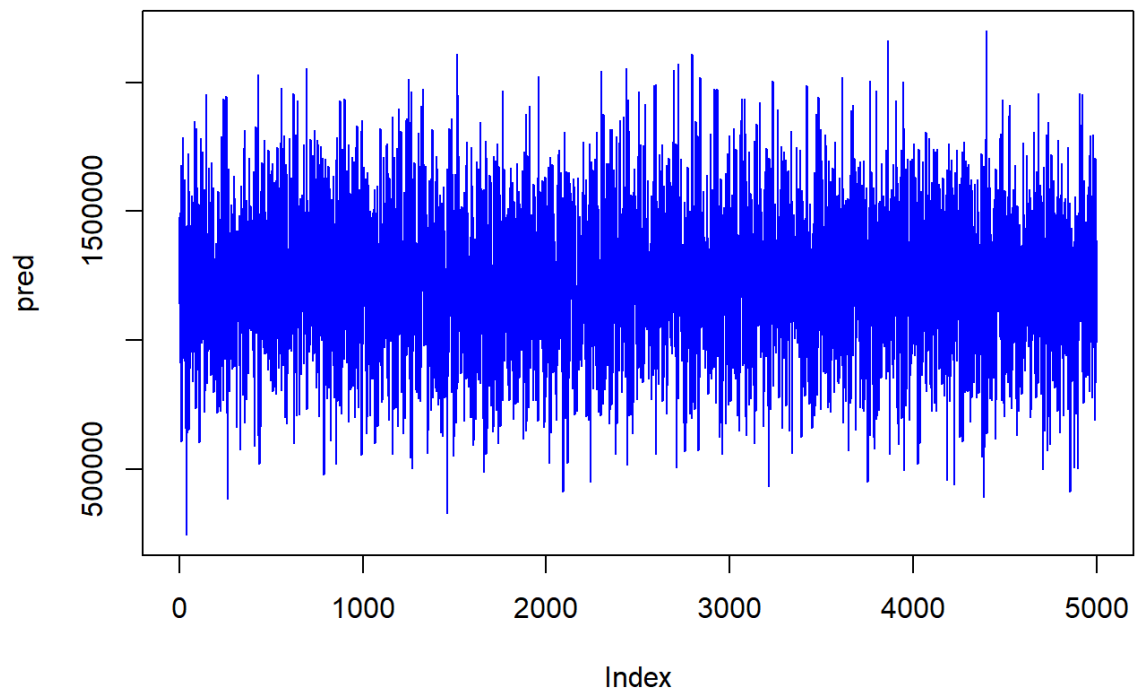
## Validation 2

```
plot(Housing$Price,type = "l",lty = 1.8,col = "red", main = "Actual vs Predicted House Prices")
lines(pred, type = "l", col = "blue")
```



```
plot(pred,type = "l",lty = 1.8,col = "blue", main = "Predicted House Prices")
```

## Predicted House Prices



```
plot(Housing$Price,type = "l",lty = 1.8,col = "red", main = "Actual House P  
rices")
```

**Actual House Prices**

