

Logistic Regression on GermanCredit dataset in “caret” package using R.

```
knitr::opts_chunk$set(echo = FALSE, warning=FALSE, message = FALSE, tidy = TRUE)
```

```
## [1] "Exploratory data analysis ----->"
```

```
## [1] "The dimension of the data is"
```

```
## [1] 1000 60
```

```
## [1] "The dependent variable is CLASS"
```

```
## [1] 1 0
```

```
## [1] "Some of the features are"
```

```
## [1] "Duration" "Amount"
```

```
## [3] "InstallmentRatePercentage" "ResidenceDuration"
```

```
## [5] "Age" "NumberExistingCredits"
```

```
## [1] "The total number of missing values is 0"
```

```
## [1] "The number of unique levels in each column"
```

```
##           [,1]
```

```
## Duration           33
```

```
## Amount           921
```

```
## InstallmentRatePercentage 4
```

```
## ResidenceDuration 4
```

```
## Age           53
```

```
## NumberExistingCredits 4
```

```
## NumberPeopleMaintenance 2
```

```
## Telephone 2
```

```
## ForeignWorker 2
```

```
## CheckingAccountStatus.lt.0 2
```

```
## [1] "53 columns have only two levels"
```

```
## [1] "Data transformation using WOE ----->"
```

```
## CLASS ResidenceDuration
```

```
## 1 1 4
```

```
## 2 0 2
```

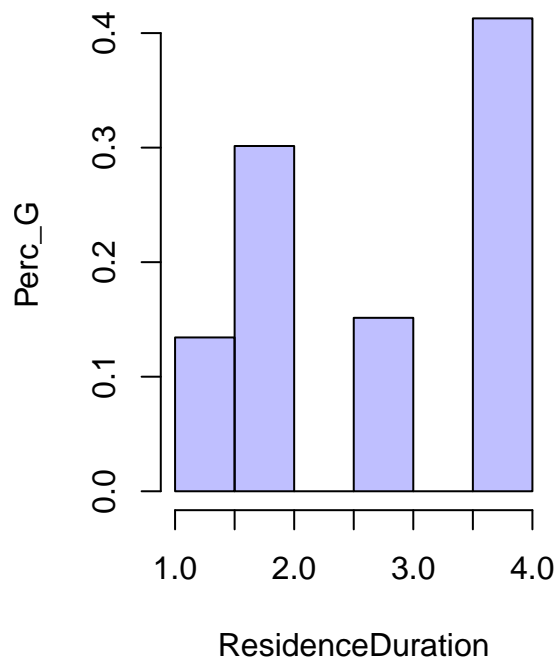
```
## 3 1 3
```

```
## 4 1 4
```

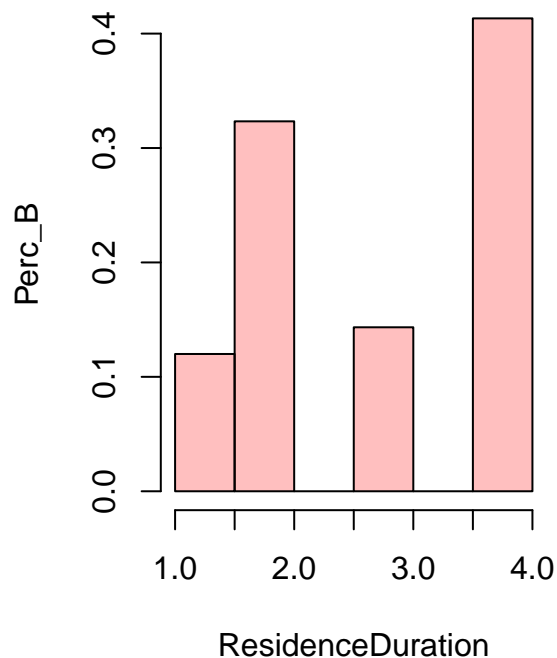
```
## 5 0 4
```

```
## 6 1 4
```

CLASS=1 (total number = 700)

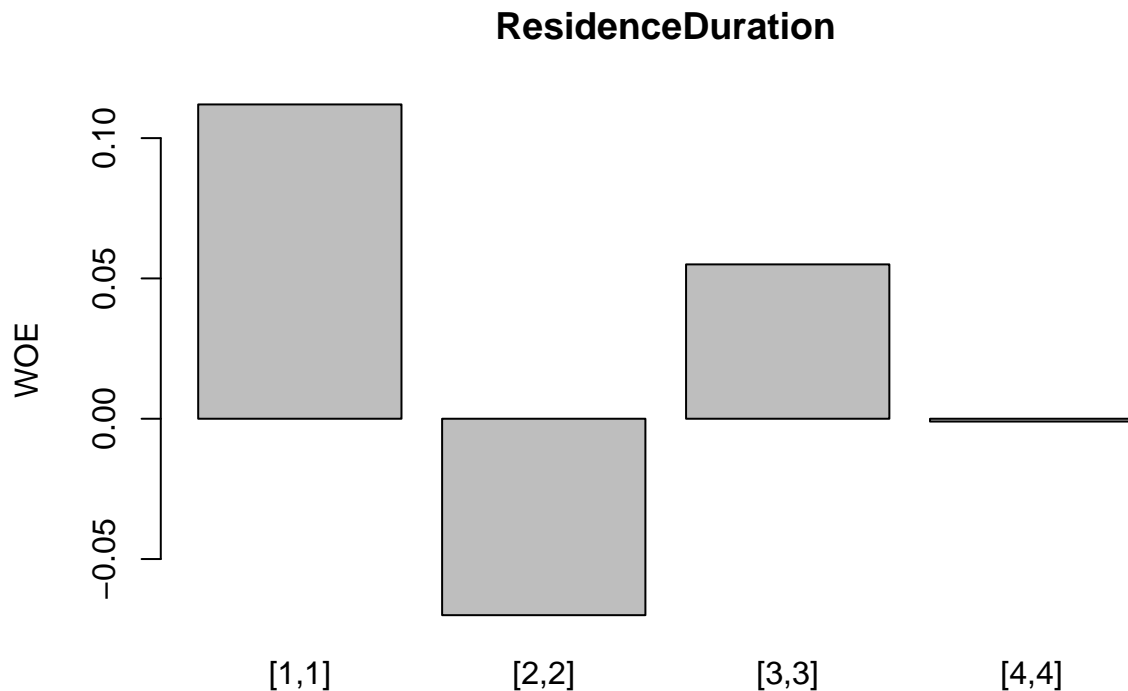


CLASS=0, (total number = 300)



##	ResidenceDuration	N	Percent	WOE	IV	Perc_G	Perc_B	N_G	N_B
## 1	[1,1]	130	0.130	0.112	0.002	0.134	0.120	94	36
## 2	[2,2]	308	0.308	-0.070	0.003	0.301	0.323	211	97
## 3	[3,3]	149	0.149	0.055	0.004	0.151	0.143	106	43
## 4	[4,4]	413	0.413	-0.001	0.004	0.413	0.413	289	124

##	CLASS	ResidenceDuration
## 1	1	-0.001
## 2	0	-0.070
## 3	1	0.055
## 4	1	-0.001
## 5	0	-0.001
## 6	1	-0.001



```
## [1] "Logistic regression result ----->"
```

```
## [1] "The confusion matrix using original data"
```

```
##           Actual
## Prediction  0   1
##           0  34  17
##           1  23 126
```

```
## [1] "The confusion matrix using transformed data"
```

```
##           Actual
## Prediction  0   1
##           0  36  19
##           1  21 124
```

```
## [1] "Type_I error = fp/(fp+tn)"
```

```
## [1] "Precision = tp/(tp+fp)"
```

```
##           Model Type_I Precision Accuracy    MSE
## 1   Original data 0.4035   0.8456     0.8 0.1562
## 2 Transformed data 0.3684   0.8552     0.8 0.1501
```

```
## [1] "Fitting result using transformed data ----->"
```

```
##           Estimate Std. Error  z value    Pr(>|z|)
## (Intercept)   0.9635406  3.3597682  0.2867878 7.742748e-01
## Duration      0.9152830  0.2176865  4.2045921 2.615534e-05
```

## Amount	0.9846595	0.3295985	2.9874512	2.813142e-03
## InstallmentRatePercentage	1.4320171	0.6466223	2.2146115	2.678674e-02
## ResidenceDuration	3.0532152	1.6750140	1.8227999	6.833371e-02
## Age	0.7649176	0.3100506	2.4670733	1.362225e-02