

# GERMAN CREDIT RISK ANALYSIS AND PREDICTIVE MODELLING

Surya Sashank Gundepudi

9/26/2021

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```

## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
library(dplyr)
```

```
##  
## 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(plotly)
```

```
## Loading required package: ggplot2  
  
##  
## Attaching package: 'plotly'  
  
## The following object is masked from 'package:ggplot2':  
##  
##   last_plot
```

```
## The following object is masked from 'package:stats':  
##  
## filter
```

```
## The following object is masked from 'package:graphics':  
##  
## layout
```

```
library(ggplot2)  
library(reshape2)  
library(tidyr)
```

```
##  
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:reshape2':  
##  
## smiths
```

```
library(descr)
```

```
# reading the clean data set which was created in mutating-data-for-eda.R  
data <- read.csv("data/eda-german-credit.csv")
```

```
data_good = data.frame(data[data["Outcome"] == 'Good', ])  
data_bad = data.frame(data[data["Outcome"] == 'Bad', ])
```

```
fig1 <- plot_ly(  
  data = data,  
  x = c("Good", "Bad"),  
  y = c(length(data$Outcome[data$Outcome == "Good"]),  
        length(data$Outcome[data$Outcome == "Bad"])),  
  type = "bar",  
  marker = list(color = c("Green", "Red"))  
)
```

```
fig1 <- fig1 %>% layout(  
  title = 'Target Variable Distribution',  
  xaxis = list(title = 'Outcome of Risk Taken'),  
  font=t,  
  plot_bgcolor = "#e5ecf6",  
  yaxis = list(title = 'Count'),  
  legend = list(title=list(text='Legend Title')))
```

```
# Un-comment this line to get the figure  
#fig1
```