

# Project Creditworthiness

Dung Nguyen

## Step 1: Business and Data Understanding

What decisions needs to be made? Determining whether customers are creditworthy to give a loan to.

What data is needed to inform those decisions? Outcome: Credit-Application-Result.

What kind of model (Continuous, Binary, Non-Binary, Time-Series) do we need to use to help make these decisions? Binary model.

## Step 2: Building the Training Set

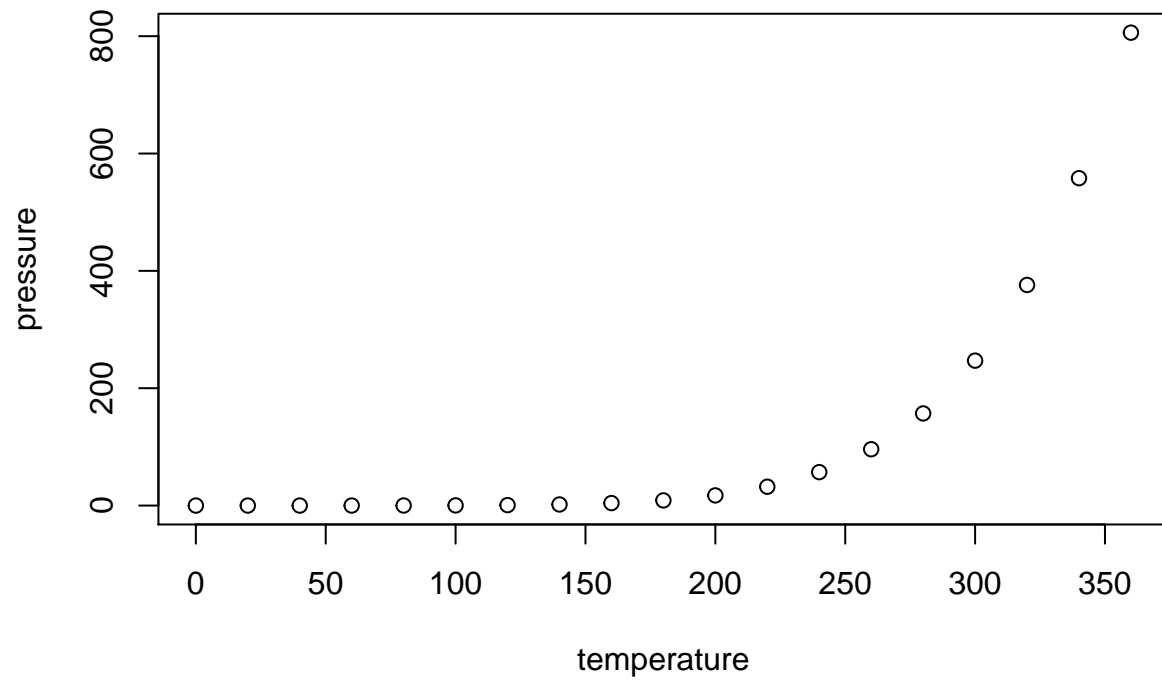
After importing, the data set were formatted as suggested.

```
library(tidyverse)
library(caret)
library(randomForest)
library(bst)
colnames(raw) <- str_replace_all(colnames(raw), '\\\\.', '_')
glimpse(raw)
```

```
## Rows: 500
## Columns: 20
## $ Credit_Application_Result      <chr> "Creditworthy", "Creditworthy", "...
## $ Account_Balance               <chr> "Some Balance", "Some Balance", "...
## $ Duration_of_Credit_Month       <int> 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 6, ...
## $ Payment_Status_of_Previous_Credit <chr> "Paid Up", "Paid Up", "No Problem...
## $ Purpose                       <chr> "Other", "Home Related", "Home Re...
## $ Credit_Amount                 <int> 1494, 1494, 1544, 3380, 343, 362,...
## $ Value_Savings_Stocks          <chr> "£100-£1000", "£100-£1000", "None...
## $ Length_of_current_employment   <chr> "< 1yr", "< 1yr", "1-4 yrs", "1-4...
## $ Instalment_per_cent            <int> 1, 1, 2, 1, 4, 4, 4, 3, 3, 2, 3, ...
## $ Guarantors                    <chr> "None", "None", "None", "None", "...
## $ Duration_in_Current_address     <int> 2, 2, 1, 1, 1, NA, NA, NA, 3, 4, ...
## $ Most_valuable_available_asset   <int> 1, 1, 1, 1, 1, 3, 2, 2, 1, 1, 1, ...
## $ Age_years                     <int> NA, 29, 42, 37, 27, 52, 24, 22, 2...
## $ Concurrent_Credits             <chr> "Other Banks/Depts", "Other Banks...
## $ Type_of_apartment              <int> 2, 2, 2, 2, 2, 2, 1, 2, 2, 1, 2, ...
## $ No_of_Credits_at_this_Bank      <chr> "1", "1", "More than 1", "1", "1"...
## $ Occupation                    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ No_of_dependents               <int> 2, 2, 2, 2, 1, 1, 2, 1, 1, 1, 1, ...
## $ Telephone                     <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, ...
## $ Foreign_Worker                 <int> 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
```

## 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.