```
#Install Packages
install.packages("titanic")
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' ...
 done
```

```
install.packages("dplyr")
```

```
install.packages("tidyverse")
```

```
#Load Library
library(titanic)
library(dplyr)
library(tidyverse)
```

```
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


Warning message in system("timedatectl", intern = TRUE):
“running command 'timedatectl' had status 1”
Warning message:
“Failed to locate timezone database”
── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1
```

```
glimpse(titanic_train)
```

```
Rows: 891
Columns: 12
$ PassengerId <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,…
$ Survived    <int> 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1…
$ Pclass      <int> 3, 1, 3, 1, 3, 3, 1, 3, 3, 2, 3, 1, 3, 3, 3, 2, 3, 2, 3, 3…
$ Name        <chr> "Braund, Mr. Owen Harris", "Cumings, Mrs. John Bradley (Fl…
$ Sex         <chr> "male", "female", "female", "female", "male", "male", "mal…
$ Age         <dbl> 22, 38, 26, 35, 35, NA, 54, 2, 27, 14, 4, 58, 20, 39, 14, …
$ SibSp       <int> 1, 1, 0, 1, 0, 0, 0, 3, 0, 1, 1, 0, 0, 1, 0, 0, 4, 0, 1, 0…
$ Parch       <int> 0, 0, 0, 0, 0, 0, 0, 1, 2, 0, 1, 0, 0, 5, 0, 0, 1, 0, 0, 0…
$ Ticket      <chr> "A/5 21171", "PC 17599", "STON/O2. 3101282", "113803", "37…
$ Fare        <dbl> 7.2500, 71.2833, 7.9250, 53.1000, 8.0500, 8.4583, 51.8625,…
$ Cabin       <chr> "", "C85", "", "C123", "", "", "E46", "", "", "", "G6", "C…
$ Embarked    <chr> "S", "C", "S", "S", "S", "Q", "S", "S", "S", "C", "S", "S"…
```

```
#Cleansing DATA
titanic_train_cl <- na.omit(titanic_train)
```

```
#Factor Sex Column
titanic_train_cl$Sex <- factor(titanic_train_cl$Sex,
                               levels = c("male","female"), labels = c(-1,1), ord
glimpse(titanic_train_cl)
```

```
Rows: 714
Columns: 12
$ PassengerId <int> 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19…
$ Survived    <int> 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1…
$ Pclass      <int> 3, 1, 3, 1, 3, 1, 3, 3, 2, 3, 1, 3, 3, 3, 2, 3, 3, 2, 2, 3…
$ Name        <chr> "Braund, Mr. Owen Harris", "Cumings, Mrs. John Bradley (Fl…
$ Sex         <ord> -1, 1, 1, 1, -1, -1, -1, 1, 1, 1, 1, -1, -1, 1, 1, -1, 1, …
$ Age         <dbl> 22, 38, 26, 35, 35, 54, 2, 27, 14, 4, 58, 20, 39, 14, 55, …
$ SibSp       <int> 1, 1, 0, 1, 0, 0, 3, 0, 1, 1, 0, 0, 1, 0, 0, 4, 1, 0, 0, 0…
$ Parch       <int> 0, 0, 0, 0, 0, 0, 1, 2, 0, 1, 0, 0, 5, 0, 0, 1, 0, 0, 0, 0…
$ Ticket      <chr> "A/5 21171", "PC 17599", "STON/O2. 3101282", "113803", "37…
$ Fare        <dbl> 7.2500, 71.2833, 7.9250, 53.1000, 8.0500, 51.8625, 21.0750…
$ Cabin       <chr> "", "C85", "", "C123", "", "E46", "", "", "", "G6", "C103"…
$ Embarked    <chr> "S", "C", "S", "S", "S", "S", "S", "S", "C", "S", "S", "S"…
```

```
## Split Data
set.seed(88)
n <- nrow(titanic_train_cl)
id <- sample(1:n, size=n*0.7) ## 70% train 30% test
train_data -> titanic_train_cl[id,]
glimpse(train_data)
test_data <- titanic_train_cl[-id,]
glimpse(test_data)
#> nrow(train_data)#[1] 499
#> nrow(test_data)#[1] 215
```

```
Rows: 499
Columns: 12
$ PassengerId <int> 346, 742, 551, 506, 104, 837, 431, 667, 196, 805, 501, 750
$ Survived    <int> 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0
$ Pclass      <int> 2, 1, 1, 1, 3, 3, 1, 2, 1, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 1
$ Name        <chr> "Brown, Miss. Amelia \"Mildred\"", "Cavendish, Mr. Tyrell
$ Sex         <ord> 1, -1, -1, -1, -1, -1, -1, -1, 1, -1, -1, -1, -1, -1, -1,
$ Age         <dbl> 24, 36, 17, 18, 33, 21, 28, 25, 58, 27, 17, 31, 51, 28, 33
$ SibSp       <int> 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
$ Parch       <int> 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
$ Ticket      <chr> "248733", "19877", "17421", "PC 17758", "7540", "315097",
$ Fare        <dbl> 13.0000, 78.8500, 110.8833, 108.9000, 8.6542, 8.6625, 26.5
$ Cabin       <chr> "F33", "C46", "C70", "C65", "", "", "C52", "", "B80", "",
$ Embarked    <chr> "S", "S", "C", "C", "S", "S", "S", "S", "C", "S", "S", "Q"
Rows: 215
Columns: 12
$ PassengerId <int> 3, 5, 7, 8, 11, 15, 23, 24, 25, 31, 35, 38, 41, 50, 51, 54
$ Survived    <int> 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1
$ Pclass      <int> 3, 3, 1, 3, 3, 3, 3, 1, 3, 1, 1, 3, 3, 3, 3, 2, 2, 3, 3, 1
$ Name        <chr> "Heikkinen, Miss. Laina", "Allen, Mr. William Henry", "McC
```

```
## Train Model
model <- glm(Survived ~ Pclass + Age + Sex, data = train_data, family = "binomial
model
summary(model)
```

```
Call:  glm(formula = Survived ~ Pclass + Age + Sex, family = "binomial",
    data = train_data)

Coefficients:
(Intercept)       Pclass          Age        Sex.L
    3.75387     -1.25231     -0.03704      1.65601

Degrees of Freedom: 498 Total (i.e. Null);  495 Residual
Null Deviance:      672.8
Residual Deviance: 472  AIC: 480
```

```
Call:
glm(formula = Survived ~ Pclass + Age + Sex, family = "binomial",
    data = train_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6928  -0.7210  -0.4175   0.6756   2.4073

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.753872   0.561521   6.685 2.31e-11 ***
Pclass      -1.252311   0.162874  -7.689 1.48e-14 ***
Age         -0.037041   0.009088  -4.076 4.59e-05 ***
Sex.L        1.656014   0.171756   9.642  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

```
## Test Model
train_data$prob_survived <- predict(model,type = "response")
train_data$pred_survived <- ifelse(train_data$prob_survived >= 0.5,1,0)
glimpse(train_data$pred_survived)
```

```
 num [1:499] 1 0 1 1 0 0 1 0 1 0 ...
```

```
## Confusion matrix of Train Model
confM_titanic <- table(train_data$pred_survived,train_data$Survived,
                       dnn = c("Predicted","Actual"))
confM_titanic
```

```
         Actual
Predicted   0   1
        0 251  60
        1  47 141
```

```
# Evaluate Model
accuracy <- (confM_titanic[1,1] + confM_titanic[2,2])/sum(confM_titanic)
precision <- confM_titanic[2,2] / (confM_titanic[2,1]+confM_titanic[2,2])
recall <- confM_titanic[2,2] / (confM_titanic[1,2]+confM_titanic[2,2])
f1_score <- 2*((precision*recall)/(precision+recall))

#Print Results
cat("Accuracy:", accuracy, "\nPrecision:", precision, "\nRecall:", recall,"\nF1:"
```

```
Accuracy: 0.7855711
Precision: 0.75
Recall: 0.7014925
F1: 0.7249357
```