Titanic Disaster

```
library(readr)
library(here)
library(rsample)
library(recipes)
library(parsnip)
library(workflows)
library(yardstick)
```

Load Data

```
titanic <- read_csv(here("src", "data", "train.csv"))
```

```
head(titanic)
```

```
# A tibble: 6 x 12
  PassengerId Survived Pclass Name    Sex     Age SibSp Parch Ticket  Fare Cabin
        <dbl>    <dbl>  <dbl> <chr>   <chr> <dbl> <dbl> <dbl> <chr>  <dbl> <chr>
1           1        0      3 Braund~ male     22     1     0 A/5 2~  7.25 <NA>
2           2        1      1 Cuming~ fema~    38     1     0 PC 17~ 71.3  C85
3           3        1      3 Heikki~ fema~    26     0     0 STON/~  7.92 <NA>
4           4        1      1 Futrel~ fema~    35     1     0 113803 53.1  C123
5           5        0      3 Allen,~ male     35     0     0 373450  8.05 <NA>
6           6        0      3 Moran,~ male     NA     0     0 330877  8.46 <NA>
# i 1 more variable: Embarked <chr>
```

Train-Test-Split

```r
set.seed(42)

split <- initial_split(titanic)
train_data <- training(split)
test_data <- testing(split)
```

```r
nrow(train_data)
```

```
[1] 668
```

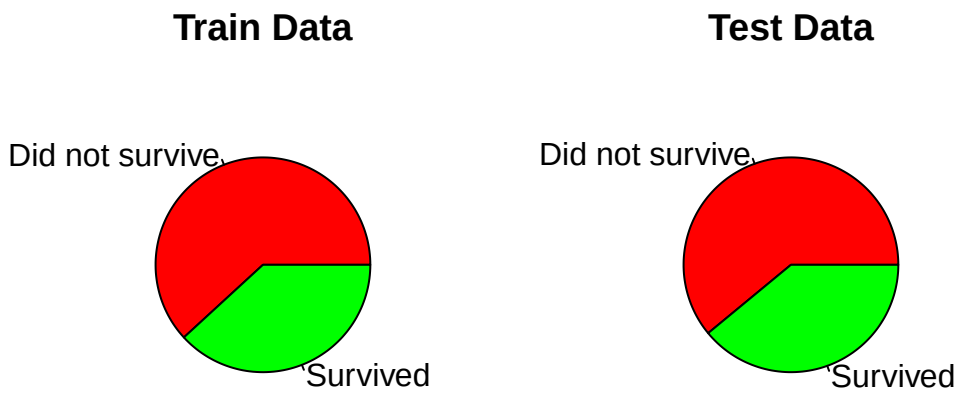
```r
nrow(test_data)
```

```
[1] 223
```

```r
par(mfrow = c(1, 2))

# Pie plot for train data
pie(
  table(train_data$Survived),
  labels = c("Did not survive", "Survived"),
  main = "Train Data",
  col = c("red", "green")
)

# Pie plot for test data
pie(
  table(test_data$Survived),
  labels = c("Did not survive", "Survived"),
  main = "Test Data",
  col = c("red", "green")
)
```



Preprocess Data

```
train_data$Survived <- as.factor(train_data$Survived)

titanic_recipe <- recipe(Survived ~ ., data = train_data) %>%
  step_rm(PassengerId, Name, Ticket, Cabin) %>%
  step_impute_mean(Age) %>%
  step_impute_mode(Embarked) %>%
  step_mutate(across(c(Pclass, Sex, Embarked), as.factor)) %>%
  step_dummy(c(Pclass, Sex, Embarked)) %>%
  step_normalize(all_of(c("Age", "Fare")))
```

```
prep_recipe <- prep(titanic_recipe, training = train_data)
```

```
train_processed <- bake(prep_recipe, new_data = train_data)
head(train_processed)
```

```
# A tibble: 6 x 10
     Age SibSp Parch    Fare Survived Pclass_X2 Pclass_X3 Sex_male Embarked_Q
   <dbl> <dbl> <dbl>   <dbl> <fct>        <dbl>     <dbl>    <dbl>      <dbl>
```

```
1  0         0     0 -0.473  0                 0         1         1          1
2 -0.592     0     0 -0.483  0                 0         1         1          0
3  1.99      0     0 -0.467  0                 0         1         1          0
4 -0.284     1     0 -0.343  0                 0         1         1          0
5 -0.707     0     0 -0.483  0                 0         1         1          0
6 -0.823     1     1  0.0883 0                 1         0         1          0
# i 1 more variable: Embarked_S <dbl>
```

Fit Model

```
log_reg_model <- logistic_reg() %>%
  set_engine("glm") %>%
  set_mode("classification")
```

```
titanic_workflow <- workflow() %>%
  add_model(log_reg_model) %>%
  add_recipe(titanic_recipe)
```

```
titanic_fit <- fit(titanic_workflow, data = train_data)
```

Predict

```
test_data$Survived <- as.factor(test_data$Survived)
predictions <- predict(titanic_fit, new_data = test_data) %>%
  bind_cols(test_data)
```

```
conf_mat(predictions, truth = Survived, estimate = .pred_class)
```

```
          Truth
Prediction   0   1
         0 115  30
         1  21  57
```

```
accuracy(predictions, truth = Survived, estimate = .pred_class)
```

```
# A tibble: 1 x 3
  .metric  .estimator .estimate
  <chr>    <chr>          <dbl>
1 accuracy binary         0.771
```

Simulate File Upload

```
train_data <- read_csv(here("src", "data", "train.csv"))
train_data$Survived = as.factor(train_data$Survived)
model <- fit(titanic_workflow, data = train_data)
```

```
test_data <- read_csv(here("src", "data", "test.csv"))
prediction <- predict(model, new_data = test_data)
result <- data.frame(
  PassengerId = test_data$PassengerId,
  Survived = prediction$.pred_class
)
```

```
head(result)
```

```
  PassengerId Survived
1         892        0
2         893        0
3         894        0
4         895        0
5         896        1
6         897        0
```