American International University-Bangladesh (AIUB)

Course Code: CSC4180

Course Title: INTRODUCTION TO DATA SCIENCE

Project Title: Data Preparation and Exploration On Titanic Dataset.

Course: INTRODUCTION TO DATA SCIENCE

Submitted By

Name: Mir Faris

ID: 20-42241-1

Section: [B]

FACULTY OF SCIENCE & TECHNOLOGY

Submitted To

TOHEDUL ISLAM

Project overview:

Titanic dataset, which is widely used in the field of data science and machine learning. This dataset consists of information about the passengers who were on board the titanic during its first voyage in 1912, which tragically ended in the ship sinking. The dataset includes different characteristics of the passengers, such as their age, gender, class, fare, and whether they survived or not. The purpose of this project is to analyze the titanic dataset, extract valuable insights from the data, and potentially create predictive models to understand the factors that influenced passenger survival during the titanic disaster.

Dataset description:

The titanic dataset consists of multiple variables (columns) that describe different aspects of the passengers. Here is a brief description of the key variables in the dataset:

Survived: indicates whether a passenger survived or not (0 = no, 1 = yes)

Pclass: passenger class (1 = first class, 2 = second class, 3 = third class)

Name: passenger's name

Sex: passenger's gender (male or female)

Age: passenger's age in years

Sibsp: number of siblings/spouses aboard the titanic

Parch: number of parents/children aboard the titanic

Ticket: ticket number

Fare: fare paid for the ticket

Cabin: cabin number

Embarked: port of embarkation (c = cherbourg, q = queenstown, s = southampton)

Project solution design:

loading and preprocessing the enormous dataset, doing data analysis and visualization, engineering pertinent features, training predictive models, assessing their performance, interpreting the findings, and documenting the findings are the initial steps. We examine the dataset, deal with the missing values, and modify the data types. Techniques for analysis and visualization are used, as well as feature engineering. The outcomes of the training and testing of machine learning models are analyzed and summarized in a project report. This methodical technique guarantees a thorough review of the data and offers useful insights.

Data pre-processing:

Importing the dataset: Initially, the read.csv function in R is used to read a dataset into the computer from the "dataset.csv" csv file. The generated data frame is then associated with the dataset variable. The dataset variable's contents are then shown in the console by using the print function. This code's output shows every row and column in the dataset as well as all of its data.

```
1  dataset<-read.csv("Dataset_midterm_Section(B).csv")
2  print(dataset)
```

The output provides a way to verify that the data has been read in correctly and allows for a quick visual inspection of the dataset. The output also provides information about the structure of the data, such as the number of columns and rows and the type of data in each column. Overall, this code and its output are a necessary first step in any data analysis project in R.

```
> dataset<-read.csv("Dataset_midterm_Section(B).csv")
> print(dataset)
    Gender  age sibsp parch     fare embarked  class     who alone survived
1        0 24.0     0     0   7.7958        S  Third   mannn  TRUE        0
2        0 17.0     0     0   8.6625        S  Third     man  TRUE        0
3        1 21.0     0     0   7.7500        Q  Third  womann  TRUE        0
4        1   NA     0     0   7.6292        Q  Third   woman  TRUE        0
5        1 37.0     0     0   9.5875        S  Third womannn  TRUE        0
6       NA 16.0     0     0  86.5000        S  First   woman  TRUE        1
7        0 18.0     1     0 108.9000        C  First   mannn FALSE        0
8        1 33.0     0     2  26.0000        S Second   woman FALSE        1
9        0   NA     0     0  26.5500        S  First     man  TRUE        1
10       0 28.0     0     0  22.5250        S  Third     man  TRUE        0
11       0 26.0     0     0  56.4958        S  Third     man  TRUE        1
12       0 29.0     0     0   7.7500        Q  Third     man  TRUE        1
13       0   NA     0     0   8.0500        S  Third     man  TRUE        0
14       0 36.0     0     0  26.2875        S            man  TRUE        1
15       1 54.0     1     0  59.4000        C  First   woman FALSE        1
16       0 24.0     0     0   7.4958        S  Third     man  TRUE        0
17       0 47.0     0     0  34.0208        S  First     man  TRUE        0
18       1 34.0     0     0  10.5000        S Second   woman  TRUE        1
19       0   NA     0     0  24.1500        Q  Third     man  TRUE        0
20       1 36.0     1     0  26.0000        S Second   woman FALSE        1
21       0 32.0     0     0   7.8958        S  Third     man  TRUE        0
22       1 30.0     0     0  93.5000        S  First   woman  TRUE        1
23      NA 22.0     0     0   7.8958        S  Third     man  TRUE        0
24       0   NA     0     0   7.2250        C  Third     man  TRUE        0
25       1 44.0     0     1  57.9792        C  First   woman FALSE        1
26       0   NA     0     0   7.2292        C  Third     man  TRUE        0
27       0 40.5     0     0   7.7500        Q  Third     man  TRUE        0
28       1 50.0     0     0  10.5000        S Second   woman  TRUE        1
29       0   NA     0     0 221.7792        S  First     man  TRUE        0
30       0 39.0     0     0   7.9250        S  Third     man  TRUE        0
31       0 23.0     2     1  11.5000        S Second     man FALSE        0
32      NA  2.0     1     1  26.0000        S Second   child FALSE        1
33       0   NA     0     0   7.2292        C  Third     man  TRUE        0
34       0 17.0     1     1   7.2292        C  Third     man FALSE        0
35       1   NA     0     2  22.3583        C            woman FALSE        1
36       1 30.0     0     0   8.6625        S  Third   woman  TRUE        0
37       1  7.0     0     2  26.2500        S Second   child FALSE        1
38       0 45.0     0     0  26.5500        S  First     man  TRUE        0
39       1 30.0     0     0 106.4250        C  First   woman  TRUE        1
40       0   NA     0     0  14.5000        S  Third     man  TRUE        0
```

```
41       1 22.0     0     2  49.5000        C  First   woman FALSE        1
42       1 36.0     0     2  71.0000        S  First   woman FALSE        1
43      NA  9.0     4     2  31.2750        S  Third   child FALSE        0
44       1 11.0     4     2  31.2750        S  Third   child FALSE        0
45       0 32.0     1     0  26.0000        S Second     man FALSE        1
46       0 50.0     1     0 106.4250        C  First     man FALSE        0
47       0 64.0     0     0  26.0000        S  First     man  TRUE        0
48       1 19.0     1     0  26.0000        S Second   woman FALSE        1
49       0    NA    0     0  13.8625        C Second     man  TRUE        1
50       0 33.0     1     1  20.5250        S  Third     man FALSE        0
51       0  8.0     1     1  36.7500        S Second   child FALSE        1
52       0 17.0     0     2 110.8833        C  First     man FALSE        1
53       0 27.0     0     0  26.0000        S Second     man  TRUE        0
54       0    NA    0     0   7.8292        Q  Third     man  TRUE        0
55       0 22.0     0     0   7.2250        C  Third     man  TRUE        1
56       1 22.0     0     0   7.7750        S  Third   woman  TRUE        1
57       0 62.0     0     0  26.5500        S  First     man  TRUE        0
58       1 48.0     1     0  39.6000        C  First   woman FALSE        1
59       0    NA    0     0 227.5250        C  First     man  TRUE        0
60       1 39.0     1     1  79.6500        S  First   woman FALSE        1
61       1 36.0     1     0  17.4000        S  Third   woman FALSE        1
62       0    NA    0     0   7.7500        Q  Third     man  TRUE        0
63       0 40.0     0     0   7.8958        S  Third     man  TRUE        0
64       0 28.0     0     0  13.5000        S            man  TRUE        0
65       0    NA    0     0   8.0500        S  Third     man  TRUE        0
66       1    NA    0     0   8.0500        S  Third   woman  TRUE        0
67       0 24.0     2     0  24.1500        S  Third     man FALSE        0
68       0 19.0     0     0   7.8958        S  Third     man  TRUE        0
69       1 29.0     0     4  21.0750        S  Third   woman FALSE        0
70       0    NA    0     0   7.2292        C  Third     man  TRUE        0
71       0 32.0     0     0   7.8542        S  Third     man  TRUE        1
72       0 62.0     0     0  10.5000        S Second     man  TRUE        1
73       1 53.0     2     0  51.4792        S  First   woman FALSE        1
74       0 36.0     0     0  26.3875        S  First     man  TRUE        1
75       1    NA    0     0   7.7500        Q  Third   woman  TRUE        1
76       0 16.0     0     0   8.0500        S  Third     man  TRUE        0
77       0 19.0     0     0  14.5000        S  Third     man  TRUE        0
78       1 34.0     0     0  13.0000        S Second   woman  TRUE        1
79       1 39.0     1     0  55.9000        S  First   woman FALSE        1
80       1    NA    1     0  14.4583        C  Third   woman FALSE        0
81       0 32.0     0     0   7.9250        S  Third     man  TRUE        1
82       1 25.0     1     1  30.0000        S Second   woman FALSE        1
83       1 39.0     1     1 110.8833        C  First   woman FALSE        1
84       0 54.0     0     0  26.0000        S Second     man  TRUE        0
85       0 36.0     0     0  40.1250        C  First     man  TRUE        0
86       0    NA    0     0   8.7125        C  Third     man  TRUE        0
87       1 18.0     0     2  79.6500        S  First   woman FALSE        1
88       0 47.0     0     0  15.0000        S Second     man  TRUE        0
89       0 60.0     1     1  79.2000        C  First     man FALSE        1
90       0 22.0     0     0   8.0500        S  Third     man  TRUE        0
91       0    NA    0     0   8.0500        S  Third     man  TRUE        0
92       0 35.0     0     0   7.1250        S  Third     man  TRUE        0
93       1 52.0     1     0  78.2667        C  First   woman FALSE        1
94       0 47.0     0     0   7.2500        S  Third     man  TRUE        0
95       1    NA    0     2   7.7500        Q  Third   woman FALSE        0
96       0 37.0     1     0  26.0000        S Second     man FALSE        0
97       0 36.0     1     1  24.1500        S  Third     man FALSE        0
98       1    NA    0     0  33.0000        S Second   woman  TRUE        1
99       0 49.0     0     0   0.0000        S            man  TRUE        0
100      0    NA    0     0   7.2250        C  Third     man  TRUE        0
 [ reached 'max' / getOption("max.print") -- omitted 201 rows ]
```

Then creating a new variable called xdataset which is a copy of the dataset variable created in the previous step

```
4  xdataset <- dataset
5  print(xdataset)
```

The creation of xdataset allows for a backup copy of the original dataset to be kept while any necessary changes are made to the data. This ensures that the original data is not lost or overwritten accidentally. The output of this code is useful for verifying that the copy was made

correctly and allows for a quick visual comparison of the two datasets to ensure they are identical.

Identifying missing values:

To identify missing value “is.na” function to check for missing values in the xdataset variable. The is.na function returns a logical value of true for each missing value in the data and false for all other values.

```
> is.na(xdataset)
      Gender   age sibsp parch  fare embarked class   who alone survived
 [1,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [2,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [3,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [4,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [5,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [6,]   TRUE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [7,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [8,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [9,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[10,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[11,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[12,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[13,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[14,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[15,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[16,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[17,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[18,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[19,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[20,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[21,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[22,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[23,]   TRUE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[24,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[25,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[26,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[27,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[28,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[29,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[30,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[31,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[32,]   TRUE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[33,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[34,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[35,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[36,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[37,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[38,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[39,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[40,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
```

```
 [40,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [41,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [42,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [43,]   TRUE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [44,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [45,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [46,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [47,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [48,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [49,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [50,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [51,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [52,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [53,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [54,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [55,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [56,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [57,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [58,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [59,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [60,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [61,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [62,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [63,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [64,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [65,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [66,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [67,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [68,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [69,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [70,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [70,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [71,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [72,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [73,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [74,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [75,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [76,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [77,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [78,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [79,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [80,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [81,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [82,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [83,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [84,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [85,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [86,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [87,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [88,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [89,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [90,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [91,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [92,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [93,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [94,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [95,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [96,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [97,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [98,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [99,]  FALSE FALSE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
[100,]  FALSE  TRUE FALSE FALSE FALSE    FALSE FALSE FALSE FALSE    FALSE
 [ reached getOption("max.print") -- omitted 201 rows ]
```

The output of this code is a matrix that has the same dimensions as the xdataset variable. The matrix contains true values in the cells where the corresponding value in xdataset is missing and false values in all other cells. This output is useful for identifying any missing values in the dataset and determining the extent to which missing values may affect subsequent analysis.

```
> sum(is.na(xdataset))
[1] 65
>
```

The sum() function is applied to the logical matrix from the previous step to count the total number of missing values in the dataset. Which is 65.

Handling missing values:

performing a procedure to deal with missing values in a dataset. The missing values are found in specific columns, namely **gender, age, sibsp, parch, fare, alone, and survived**. To address these missing values, the code utilizes the mean of each respective column as a replacement. This is achieved through the mean function and logical indexing. The missing values are then replaced with the mean values using the assignment operator, effectively filling in the gaps in the dataset.

```
> xdataset$Gender[is.na(xdataset$Gender)]<-mean(xdataset$Gender,na.rm=TRUE)
> xdataset$age[is.na(xdataset$age)]<-mean(xdataset$age,na.rm=TRUE)
> xdataset$sibsp[is.na(xdataset$sibsp)]<-mean(xdataset$sibsp,na.rm=TRUE)
> xdataset$parch[is.na(xdataset$parch)]<-mean(xdataset$parch,na.rm=TRUE)
> xdataset$fare[is.na(xdataset$fare)]<-mean(xdataset$fare,na.rm=TRUE)
> xdataset$alone[is.na(xdataset$alone)]<-mean(xdataset$alone,na.rm=TRUE)
> xdataset$survived[is.na(xdataset$survived)]<-mean(xdataset$survived,na.rm=TRUE)
> print(xdataset)
       Gender      age sibsp parch     fare embarked  class     who alone survived
1   0.0000000 24.00000     0     0   7.7958        S  Third   mannn     1        0
2   0.0000000 17.00000     0     0   8.6625        S  Third     man     1        0
3   1.0000000 21.00000     0     0   7.7500        Q  Third  womann     1        0
4   1.0000000 34.03508     0     0   7.6292        Q  Third   woman     1        0
5   1.0000000 37.00000     0     0   9.5875        S  Third womannn     1        0
6   0.3198653 16.00000     0     0  86.5000        S  First   woman     1        1
7   0.0000000 18.00000     1     0 108.9000        C  First   mannn     0        0
8   1.0000000 33.00000     0     2  26.0000        S Second   woman     0        1
9   0.0000000 34.03508     0     0  26.5500        S  First     man     1        1
10  0.0000000 28.00000     0     0  22.5250        S  Third     man     1        0
11  0.0000000 26.00000     0     0  56.4958        S  Third     man     1        1
12  0.0000000 29.00000     0     0   7.7500        Q  Third     man     1        1
13  0.0000000 34.03508     0     0   8.0500        S  Third     man     1        0
14  0.0000000 36.00000     0     0  26.2875        S            man     1        1
15  1.0000000 54.00000     1     0  59.4000        C  First   woman     0        1
16  0.0000000 24.00000     0     0   7.4958        S  Third     man     1        0
17  0.0000000 47.00000     0     0  34.0208        S  First     man     1        0
18  1.0000000 34.00000     0     0  10.5000        S Second   woman     1        1
19  0.0000000 34.03508     0     0  24.1500        Q  Third     man     1        0
20  1.0000000 36.00000     1     0  26.0000        S Second   woman     0        1
21  0.0000000 32.00000     0     0   7.8958        S  Third     man     1        0
22  1.0000000 30.00000     0     0  93.5000        S  First   woman     1        1
23  0.3198653 22.00000     0     0   7.8958        S  Third     man     1        0
24  0.0000000 34.03508     0     0   7.2250        C  Third     man     1        0
25  1.0000000 44.00000     0     1  57.9792        C  First   woman     0        1
26  0.0000000 34.03508     0     0   7.2292        C  Third     man     1        0
27  0.0000000 40.50000     0     0   7.7500        Q  Third     man     1        0
28  1.0000000 50.00000     0     0  10.5000        S Second   woman     1        1
29  0.0000000 34.03508     0     0 221.7792        S  First     man     1        0
30  0.0000000 39.00000     0     0   7.9250        S  Third     man     1        0
```

```
70  0.0000000 34.03508     0     0   7.2292        C  Third     man     1        0
71  0.0000000 32.00000     0     0   7.8542        S  Third     man     1        1
72  0.0000000 62.00000     0     0  10.5000        S Second     man     1        1
73  1.0000000 53.00000     2     0  51.4792        S  First   woman     0        1
74  0.0000000 36.00000     0     0  26.3875        S  First     man     1        1
75  1.0000000 34.03508     0     0   7.7500        Q  Third   woman     1        1
76  0.0000000 16.00000     0     0   8.0500        S  Third     man     1        0
77  0.0000000 19.00000     0     0  14.5000        S  Third     man     1        0
78  1.0000000 34.00000     0     0  13.0000        S Second   woman     1        1
79  1.0000000 39.00000     1     0  55.9000        S  First   woman     0        1
80  1.0000000 34.03508     1     0  14.4583        C  Third   woman     0        0
81  0.0000000 32.00000     0     0   7.9250        S  Third     man     1        1
82  1.0000000 25.00000     1     1  30.0000        S Second   woman     0        1
83  1.0000000 39.00000     1     1 110.8833        C  First   woman     0        1
84  0.0000000 54.00000     0     0  26.0000        S Second     man     1        0
85  0.0000000 36.00000     0     0  40.1250        C  First     man     1        0
86  0.0000000 34.03508     0     0   8.7125        C  Third     man     1        0
87  1.0000000 18.00000     0     2  79.6500        S  First   woman     0        1
88  0.0000000 47.00000     0     0  15.0000        S Second     man     1        0
89  0.0000000 60.00000     1     1  79.2000        C  First     man     0        1
90  0.0000000 22.00000     0     0   8.0500        S  Third     man     1        0
91  0.0000000 34.03508     0     0   8.0500        S  Third     man     1        0
92  0.0000000 35.00000     0     0   7.1250        S  Third     man     1        0
93  1.0000000 52.00000     1     0  78.2667        C  First   woman     0        1
94  0.0000000 47.00000     0     0   7.2500        S  Third     man     1        0
95  1.0000000 34.03508     0     2   7.7500        Q  Third   woman     0        0
96  0.0000000 37.00000     1     0  26.0000        S Second     man     0        0
97  0.0000000 36.00000     1     1  24.1500        S  Third     man     0        0
98  1.0000000 34.03508     0     0  33.0000        S Second   woman     1        1
99  0.0000000 49.00000     0     0   0.0000        S            man     1        0
100 0.0000000 34.03508     0     0   7.2250        C  Third     man     1        0
 [ reached 'max' / getOption("max.print") -- omitted 201 rows ]
> 
31  0.0000000 23.00000     2     1  11.5000        S Second     man     0        0
32  0.3198653  2.00000     1     1  26.0000        S Second   child     0        1
33  0.0000000 34.03508     0     0   7.2292        C  Third     man     1        0
34  0.0000000 17.00000     1     1   7.2292        C  Third     man     0        0
35  1.0000000 34.03508     0     2  22.3583        C           woman     0        1
36  1.0000000 30.00000     0     0   8.6625        S  Third   woman     1        0
37  1.0000000  7.00000     0     2  26.2500        S Second   child     0        1
38  0.0000000 45.00000     0     0  26.5500        S  First     man     1        0
39  1.0000000 30.00000     0     0 106.4250        C  First   woman     1        1
40  0.0000000 34.03508     0     0  14.5000        S  Third     man     1        0
41  1.0000000 22.00000     0     2  49.5000        C  First   woman     0        1
42  1.0000000 36.00000     0     2  71.0000        S  First   woman     0        1
43  0.3198653  9.00000     4     2  31.2750        S  Third   child     0        0
44  1.0000000 11.00000     4     2  31.2750        S  Third   child     0        0
45  0.0000000 32.00000     1     0  26.0000        S Second     man     0        1
46  0.0000000 50.00000     1     0 106.4250        C  First     man     0        0
47  0.0000000 64.00000     0     0  26.0000        S  First     man     1        0
48  1.0000000 19.00000     1     0  26.0000        S Second   woman     0        1
49  0.0000000 34.03508     0     0  13.8625        C Second     man     1        1
50  0.0000000 33.00000     1     1  20.5250        S  Third     man     0        0
51  0.0000000  8.00000     1     1  36.7500        S Second   child     0        1
52  0.0000000 17.00000     0     2 110.8833        C  First     man     0        1
53  0.0000000 27.00000     0     0  26.0000        S Second     man     1        0
54  0.0000000 34.03508     0     0   7.8292        Q  Third     man     1        0
55  0.0000000 22.00000     0     0   7.2250        C  Third     man     1        1
56  1.0000000 22.00000     0     0   7.7750        S  Third   woman     1        1
57  0.0000000 62.00000     0     0  26.5500        S  First     man     1        0
58  1.0000000 48.00000     1     0  39.6000        C  First   woman     0        1
59  0.0000000 34.03508     0     0 227.5250        C  First     man     1        0
60  1.0000000 39.00000     1     1  79.6500        S  First   woman     0        1
61  1.0000000 36.00000     1     0  17.4000        S  Third   woman     0        1
62  0.0000000 34.03508     0     0   7.7500        Q  Third     man     1        0
63  0.0000000 40.00000     0     0   7.8958        S  Third     man     1        0
64  0.0000000 28.00000     0     0  13.5000        S            man     1        0
65  0.0000000 34.03508     0     0   8.0500        S  Third     man     1        0
66  1.0000000 34.03508     0     0   8.0500        S  Third   woman     1        0
67  0.0000000 24.00000     2     0  24.1500        S  Third     man     0        0
68  0.0000000 19.00000     0     0   7.8958        S  Third     man     1        0
69  1.0000000 29.00000     0     4  21.0750        S  Third   woman     0        0
70  0.0000000 34.03508     0     0   7.2292        C  Third     man     1        0
```

The result of the code would be a revised dataset in which the means of the missing columns (gender, age, sibsp, parch, fare, alone, and survived) would be used to fill in the blanks. This indicates that the average value of the non-missing values in that specific column would be used to fill in any missing values in those columns. The resulting dataset would have all the missing values replaced, ensuring that each column is complete with no missing data.

Data type conversion:

Age, gender, and fare are the three mentioned columns. The "as.integer" function is used in the code to change the data type of these columns from their initial format to integers. When the variables in these columns are meant to represent discrete numerical values, like categories or whole integers, this conversion is especially helpful. The code guarantees that the variables in these columns are correctly represented as whole numbers rather than as strings or floating-point values by changing the data type to integers. This conversion aids in consistency maintenance and simplifies further analysis or calculations that could call for integer numbers.

```
> xdataset$Gender <- as.integer(xdataset$Gender)
> xdataset$age <- as.integer(xdataset$age)
> xdataset$fare <- as.integer(xdataset$fare)
>
> print(xdataset)
    Gender age sibsp parch fare embarked  class     who alone survived
1        0  24     0     0    7        S  Third   mannn     1        0
2        0  17     0     0    8        S  Third     man     1        0
3        1  21     0     0    7        Q  Third  womann     1        0
4        1  34     0     0    7        Q  Third   woman     1        0
5        1  37     0     0    9        S  Third womannn     1        0
6        0  16     0     0   86        S  First   woman     1        1
7        0  18     1     0  108        C  First   mannn     0        0
8        1  33     0     2   26        S Second   woman     0        1
9        0  34     0     0   26        S  First     man     1        1
10       0  28     0     0   22        S  Third     man     1        0
11       0  26     0     0   56        S  Third     man     1        1
12       0  29     0     0    7        Q  Third     man     1        1
13       0  34     0     0    8        S  Third     man     1        0
14       0  36     0     0   26        S            man     1        1
15       1  54     1     0   59        C  First   woman     0        1
16       0  24     0     0    7        S  Third     man     1        0
17       0  47     0     0   34        S  First     man     1        0
18       1  34     0     0   10        S Second   woman     1        1
19       0  34     0     0   24        Q  Third     man     1        0
20       1  36     1     0   26        S Second   woman     0        1
21       0  32     0     0    7        S  Third     man     1        0
22       1  30     0     0   93        S  First   woman     1        1
23       0  22     0     0    7        S  Third     man     1        0
24       0  34     0     0    7        C  Third     man     1        0
25       1  44     0     1   57        C  First   woman     0        1
26       0  34     0     0    7        C  Third     man     1        0
27       0  40     0     0    7        Q  Third     man     1        0
28       1  50     0     0   10        S Second   woman     1        1
29       0  34     0     0  221        S  First     man     1        0
30       0  39     0     0    7        S  Third     man     1        0
31       0  23     2     1   11        S Second     man     0        0
32       0   2     1     1   26        S Second   child     0        1
33       0  34     0     0    7        C  Third     man     1        0
34       0  17     1     1    7        C  Third     man     0        0
35       1  34     0     2   22        C           woman     0        1
```

```
35       1  34     0     2   22        C          woman     0        1
36       1  30     0     0    8        S  Third   woman     1        0
37       1   7     0     2   26        S Second   child     0        1
38       0  45     0     0   26        S  First     man     1        0
39       1  30     0     0  106        C  First   woman     1        1
40       0  34     0     0   14        S  Third     man     1        0
41       1  22     0     2   49        C  First   woman     0        1
42       1  36     0     2   71        S  First   woman     0        1
43       0   9     4     2   31        S  Third   child     0        0
44       1  11     4     2   31        S  Third   child     0        0
45       0  32     1     0   26        S Second     man     0        1
46       0  50     1     0  106        C  First     man     0        0
47       0  64     0     0   26        S  First     man     1        0
48       1  19     1     0   26        S Second   woman     0        1
49       0  34     0     0   13        C Second     man     1        1
50       0  33     1     1   20        S  Third     man     0        0
51       0   8     1     1   36        S Second   child     0        1
52       0  17     0     2  110        C  First     man     0        1
53       0  27     0     0   26        S Second     man     1        0
54       0  34     0     0    7        Q  Third     man     1        0
55       0  22     0     0    7        C  Third     man     1        1
56       1  22     0     0    7        S  Third   woman     1        1
57       0  62     0     0   26        S  First     man     1        0
58       1  48     1     0   39        C  First   woman     0        1
59       0  34     0     0  227        C  First     man     1        0
60       1  39     1     1   79        S  First   woman     0        1
61       1  36     1     0   17        S  Third   woman     0        1
62       0  34     0     0    7        Q  Third     man     1        0
63       0  40     0     0    7        S  Third     man     1        0
64       0  28     0     0   13        S            man     1        0
65       0  34     0     0    8        S  Third     man     1        0
66       1  34     0     0    8        S  Third   woman     1        0
67       0  24     2     0   24        S  Third     man     0        0
68       0  19     0     0    7        S  Third     man     1        0
69       1  29     0     4   21        S  Third   woman     0        0
70       0  34     0     0    7        C  Third     man     1        0
71       0  32     0     0    7        S  Third     man     1        1
72       0  62     0     0   10        S Second     man     1        1
73       1  53     2     0   51        S  First   woman     0        1
74       0  36     0     0   26        S  First     man     1        1
75       1  34     0     0    7        Q  Third   woman     1        1
75       1  34     0     0    7        Q  Third   woman     1        1
76       0  16     0     0    8        S  Third     man     1        0
77       0  19     0     0   14        S  Third     man     1        0
78       1  34     0     0   13        S Second   woman     1        1
79       1  39     1     0   55        S  First   woman     0        1
80       1  34     1     0   14        C  Third   woman     0        0
81       0  32     0     0    7        S  Third     man     1        1
82       1  25     1     1   30        S Second   woman     0        1
83       1  39     1     1  110        C  First   woman     0        1
84       0  54     0     0   26        S Second     man     1        0
85       0  36     0     0   40        C  First     man     1        0
86       0  34     0     0    8        C  Third     man     1        0
87       1  18     0     2   79        S  First   woman     0        1
88       0  47     0     0   15        S Second     man     1        0
89       0  60     1     1   79        C  First     man     0        1
90       0  22     0     0    8        S  Third     man     1        0
91       0  34     0     0    8        S  Third     man     1        0
92       0  35     0     0    7        S  Third     man     1        0
93       1  52     1     0   78        C  First   woman     0        1
94       0  47     0     0    7        S  Third     man     1        0
95       1  34     0     2    7        Q  Third   woman     0        0
96       0  37     1     0   26        S Second     man     0        0
97       0  36     1     1   24        S  Third     man     0        0
98       1  34     0     0   33        S Second   woman     1        1
99       0  49     0     0    0        S            man     1        0
100      0  34     0     0    7        C  Third     man     1        0
 [ reached 'max' / getOption("max.print") -- omitted 201 rows ]
> 
```

The output of the code would be an updated dataset where the gender, age, and fare columns have been converted to integers. This means that the original values in these columns, which may have been in a different data type (such as strings or floating-point numbers), will be converted to whole numbers. The resulting dataset will have these specific columns represented as integers, ensuring that they are suitable for analysis or computations that require discrete numerical values.

Transforming passenger class:

A code snippet that performs transformations on the "class" column of a dataset. The original column consists of string labels representing different classes ("first", "second", "third"). The code uses nested ifelse statements to convert these string labels into numerical values (1, 2, 3). The transformed numerical values are then stored back into the "class" column, replacing the original string labels. This conversion allows for easier analysis and computations that require numerical representations of the classes. Additionally, the code handles missing values in the "class" column. If any missing values are present, they are replaced with the mean value of the non-missing values in

```
> xdataset$class <- ifelse(xdataset$class == "Third",3,
+                          ifelse(xdataset$class == "Second",2,
+                                 ifelse(xdataset$class == "First",1,NA)))
>
> xdataset$class[is.na(xdataset$class)]<-mean(xdataset$class,na.rm=TRUE)
> xdataset$class <- as.integer(xdataset$class)
> print(xdataset)
    Gender age sibsp parch fare embarked class     who alone survived
1        0  24     0     0    7        S     3   mannn     1        0
2        0  17     0     0    8        S     3     man     1        0
3        1  21     0     0    7        Q     3  womann     1        0
4        1  34     0     0    7        Q     3   woman     1        0
5        1  37     0     0    9        S     3 womannn     1        0
6        0  16     0     0   86        S     1   woman     1        1
7        0  18     1     0  108        C     1   mannn     0        0
8        1  33     0     2   26        S     2   woman     0        1
9        0  34     0     0   26        S     1     man     1        1
10       0  28     0     0   22        S     3     man     1        0
11       0  26     0     0   56        S     3     man     1        1
12       0  29     0     0    7        Q     3     man     1        1
13       0  34     0     0    8        S     3     man     1        0
14       0  36     0     0   26        S     2     man     1        1
15       1  54     1     0   59        C     1   woman     0        1
16       0  24     0     0    7        S     3     man     1        0
17       0  47     0     0   34        S     1     man     1        0
18       1  34     0     0   10        S     2   woman     1        1
19       0  34     0     0   24        Q     3     man     1        0
20       1  36     1     0   26        S     2   woman     0        1
21       0  32     0     0    7        S     3     man     1        0
22       1  30     0     0   93        S     1   woman     1        1
23       0  22     0     0    7        S     3     man     1        0
24       0  34     0     0    7        C     3     man     1        0
25       1  44     0     1   57        C     1   woman     0        1
26       0  34     0     0    7        C     3     man     1        0
27       0  40     0     0    7        Q     3     man     1        0
28       1  50     0     0   10        S     2   woman     1        1
29       0  34     0     0  221        S     1     man     1        0
30       0  39     0     0    7        S     3     man     1        0
31       0  23     2     1   11        S     2     man     0        0
32       0   2     1     1   26        S     2   child     0        1
33       0  34     0     0    7        C     3     man     1        0
34       0  17     1     1    7        C     3     man     0        0
35       1  34     0     2   22        C     2   woman     0        1
```

the column. This imputation method ensures that the missing values are filled with a representative value, enabling a complete dataset for further analysis.

```
35       1  34     0     2   22        C     2   woman     0        1
36       1  30     0     0    8        S     3   woman     1        0
37       1   7     0     2   26        S     2   child     0        1
38       0  45     0     0   26        S     1     man     1        0
39       1  30     0     0  106        C     1   woman     1        1
40       0  34     0     0   14        S     3     man     1        0
41       1  22     0     2   49        C     1   woman     0        1
42       1  36     0     2   71        S     1   woman     0        1
43       0   9     4     2   31        S     3   child     0        0
44       1  11     4     2   31        S     3   child     0        0
45       0  32     1     0   26        S     2     man     0        1
46       0  50     1     0  106        C     1     man     0        0
47       0  64     0     0   26        S     1     man     1        0
48       1  19     1     0   26        S     2   woman     0        1
49       0  34     0     0   13        C     2     man     1        1
50       0  33     1     1   20        S     3     man     0        0
51       0   8     1     1   36        S     2   child     0        1
52       0  17     0     2  110        C     1     man     0        1
53       0  27     0     0   26        S     2     man     1        0
54       0  34     0     0    7        Q     3     man     1        0
55       0  22     0     0    7        C     3     man     1        1
56       1  22     0     0    7        S     3   woman     1        1
57       0  62     0     0   26        S     1     man     1        0
58       1  48     1     0   39        C     1   woman     0        1
59       0  34     0     0  227        C     1     man     1        0
60       1  39     1     1   79        S     1   woman     0        1
61       1  36     1     0   17        S     3   woman     0        1
62       0  34     0     0    7        Q     3     man     1        0
63       0  40     0     0    7        S     3     man     1        0
64       0  28     0     0   13        S     2     man     1        0
65       0  34     0     0    8        S     3     man     1        0
66       1  34     0     0    8        S     3   woman     1        0
67       0  24     2     0   24        S     3     man     0        0
68       0  19     0     0    7        S     3     man     1        0
69       1  29     0     4   21        S     3   woman     0        0
70       0  34     0     0    7        C     3     man     1        0
71       0  32     0     0    7        S     3     man     1        1
72       0  62     0     0   10        S     2     man     1        1
73       1  53     2     0   51        S     1   woman     0        1
74       0  36     0     0   26        S     1     man     1        1
75       1  34     0     0    7        Q     3   woman     1        1
76       0  16     0     0    8        S     3     man     1        0
77       0  19     0     0   14        S     3     man     1        0
78       1  34     0     0   13        S     2   woman     1        1
79       1  39     1     0   55        S     1   woman     0        1
80       1  34     1     0   14        C     3   woman     0        0
81       0  32     0     0    7        S     3     man     1        1
82       1  25     1     1   30        S     2   woman     0        1
83       1  39     1     1  110        C     1   woman     0        1
84       0  54     0     0   26        S     2     man     1        0
85       0  36     0     0   40        C     1     man     1        0
86       0  34     0     0    8        C     3     man     1        0
87       1  18     0     2   79        S     1   woman     0        1
88       0  47     0     0   15        S     2     man     1        0
89       0  60     1     1   79        C     1     man     0        1
90       0  22     0     0    8        S     3     man     1        0
91       0  34     0     0    8        S     3     man     1        0
92       0  35     0     0    7        S     3     man     1        0
93       1  52     1     0   78        C     1   woman     0        1
94       0  47     0     0    7        S     3     man     1        0
95       1  34     0     2    7        Q     3   woman     0        0
96       0  37     1     0   26        S     2     man     0        0
97       0  36     1     1   24        S     3     man     0        0
98       1  34     0     0   33        S     2   woman     1        1
99       0  49     0     0    0        S     2     man     1        0
100      0  34     0     0    7        C     3     man     1        0
 [ reached 'max' / getOption("max.print") -- omitted 201 rows ]
>
```

Reverting passenger class to labels:

The numerical transformations applied to the "class" column back to its original string labels ("first", "second", "third"). To achieve this, the code uses nested ifelse statements to conditionally assign the corresponding string label based on the numerical value. The original numerical values in the "class" column are replaced with their respective string labels. Once the transformation is complete, the resulting "class" column, now containing the original string labels, is printed to the console. This allows for the verification and inspection of the transformed column.

```
> xdataset$class <- ifelse(xdataset$class == 3,"Third",
+                          ifelse(xdataset$class == 2,"Second",
+                                 ifelse(xdataset$class == 1,"First",NA)))
> print(xdataset)
    Gender age sibsp parch fare embarked  class     who alone survived
1        0  24     0     0    7        S  Third   mannn     1        0
2        0  17     0     0    8        S  Third     man     1        0
3        1  21     0     0    7        Q  Third  womann     1        0
4        1  34     0     0    7        Q  Third   woman     1        0
5        1  37     0     0    9        S  Third womannn     1        0
6        0  16     0     0   86        S  First   woman     1        1
7        0  18     1     0  108        C  First   mannn     0        0
8        1  33     0     2   26        S Second   woman     0        1
9        0  34     0     0   26        S  First     man     1        1
10       0  28     0     0   22        S  Third     man     1        0
11       0  26     0     0   56        S  Third     man     1        1
12       0  29     0     0    7        Q  Third     man     1        1
13       0  34     0     0    8        S  Third     man     1        0
14       0  36     0     0   26        S Second     man     1        1
15       1  54     1     0   59        C  First   woman     0        1
16       0  24     0     0    7        S  Third     man     1        0
17       0  47     0     0   34        S  First     man     1        0
18       1  34     0     0   10        S Second   woman     1        1
19       0  34     0     0   24        Q  Third     man     1        0
20       1  36     1     0   26        S Second   woman     0        1
21       0  32     0     0    7        S  Third     man     1        0
22       1  30     0     0   93        S  First   woman     1        1
23       0  22     0     0    7        S  Third     man     1        0
24       0  34     0     0    7        C  Third     man     1        0
25       1  44     0     1   57        C  First   woman     0        1
26       0  34     0     0    7        C  Third     man     1        0
27       0  40     0     0    7        Q  Third     man     1        0
28       1  50     0     0   10        S Second   woman     1        1
29       0  34     0     0  221        S  First     man     1        0
30       0  39     0     0    7        S  Third     man     1        0
```

```
70       0  34     0     0    7        C  Third     man     1        0
71       0  32     0     0    7        S  Third     man     1        1
72       0  62     0     0   10        S Second     man     1        1
73       1  53     2     0   51        S  First   woman     0        1
74       0  36     0     0   26        S  First     man     1        1
75       1  34     0     0    7        Q  Third   woman     1        1
76       0  16     0     0    8        S  Third     man     1        0
77       0  19     0     0   14        S  Third     man     1        0
78       1  34     0     0   13        S Second   woman     1        1
79       1  39     1     0   55        S  First   woman     0        1
80       1  34     1     0   14        C  Third   woman     0        0
81       0  32     0     0    7        S  Third     man     1        1
82       1  25     1     1   30        S Second   woman     0        1
83       1  39     1     1  110        C  First   woman     0        1
84       0  54     0     0   26        S Second     man     1        0
85       0  36     0     0   40        C  First     man     1        0
86       0  34     0     0    8        C  Third     man     1        0
87       1  18     0     2   79        S  First   woman     0        1
88       0  47     0     0   15        S Second     man     1        0
89       0  60     1     1   79        C  First     man     0        1
90       0  22     0     0    8        S  Third     man     1        0
91       0  34     0     0    8        S  Third     man     1        0
92       0  35     0     0    7        S  Third     man     1        0
93       1  52     1     0   78        C  First   woman     0        1
94       0  47     0     0    7        S  Third     man     1        0
95       1  34     0     2    7        Q  Third   woman     0        0
96       0  37     1     0   26        S Second     man     0        0
97       0  36     1     1   24        S  Third     man     0        0
98       1  34     0     0   33        S Second   woman     1        1
99       0  49     0     0    0        S Second     man     1        0
100      0  34     0     0    7        C  Third     man     1        0
 [ reached 'max' / getOption("max.print") -- omitted 201 rows ]
30       0  39     0     0    7        S  Third     man     1        0
31       0  23     2     1   11        S Second     man     0        0
32       0   2     1     1   26        S Second   child     0        1
33       0  34     0     0    7        C  Third     man     1        0
34       0  17     1     1    7        C  Third     man     0        0
35       1  34     0     2   22        C Second   woman     0        1
36       1  30     0     0    8        S  Third   woman     1        0
37       1   7     0     2   26        S Second   child     0        1
38       0  45     0     0   26        S  First     man     1        0
39       1  30     0     0  106        C  First   woman     1        1
40       0  34     0     0   14        S  Third     man     1        0
41       1  22     0     2   49        C  First   woman     0        1
42       1  36     0     2   71        S  First   woman     0        1
43       0   9     4     2   31        S  Third   child     0        0
44       1  11     4     2   31        S  Third   child     0        0
45       0  32     1     0   26        S Second     man     0        1
46       0  50     1     0  106        C  First     man     0        0
47       0  64     0     0   26        S  First     man     1        0
48       1  19     1     0   26        S Second   woman     0        1
49       0  34     0     0   13        C Second     man     1        1
50       0  33     1     1   20        S  Third     man     0        0
51       0   8     1     1   36        S Second   child     0        1
52       0  17     0     2  110        C  First     man     0        1
53       0  27     0     0   26        S Second     man     1        0
54       0  34     0     0    7        Q  Third     man     1        0
55       0  22     0     0    7        C  Third     man     1        1
56       1  22     0     0    7        S  Third   woman     1        1
57       0  62     0     0   26        S  First     man     1        0
58       1  48     1     0   39        C  First   woman     0        1
59       0  34     0     0  227        C  First     man     1        0
60       1  39     1     1   79        S  First   woman     0        1
61       1  36     1     0   17        S  Third   woman     0        1
62       0  34     0     0    7        Q  Third     man     1        0
63       0  40     0     0    7        S  Third     man     1        0
64       0  28     0     0   13        S Second     man     1        0
65       0  34     0     0    8        S  Third     man     1        0
66       1  34     0     0    8        S  Third   woman     1        0
67       0  24     2     0   24        S  Third     man     0        0
68       0  19     0     0    7        S  Third     man     1        0
69       1  29     0     4   21        S  Third   woman     0        0
70       0  34     0     0    7        C  Third     man     1        0
```

Converting "alone" column:

Transformation on the "alone" column of a dataset. The original column contains numerical values (0 and 1), which likely represent binary indicators (e.g., 0 for "not alone" and 1 for "alone"). The code utilizes the ifelse statement to conditionally assign string labels ("false" and "true") based on the numerical values in the "alone" column. This transformation converts the numerical values into their corresponding string labels. After the transformation is applied, the resulting "alone" column, now consisting of the string labels ("false" and "true"), is printed to the console. This allows for visual inspection and verification of the transformed column.

```
> xdataset$alone <- ifelse(xdataset$alone == 1,"TRUE",
+                          ifelse(xdataset$alone == 0,"FALSE",NA))
>
> print(xdataset)
    Gender age sibsp parch fare embarked  class     who alone survived
1        0  24     0     0    7        S  Third   mannn  TRUE        0
2        0  17     0     0    8        S  Third     man  TRUE        0
3        1  21     0     0    7        Q  Third  womann  TRUE        0
4        1  34     0     0    7        Q  Third   woman  TRUE        0
5        1  37     0     0    9        S  Third womannn  TRUE        0
6        0  16     0     0   86        S  First   woman  TRUE        1
7        0  18     1     0  108        C  First   mannn FALSE        0
8        1  33     0     2   26        S Second   woman FALSE        1
9        0  34     0     0   26        S  First     man  TRUE        1
10       0  28     0     0   22        S  Third     man  TRUE        0
11       0  26     0     0   56        S  Third     man  TRUE        1
12       0  29     0     0    7        Q  Third     man  TRUE        1
13       0  34     0     0    8        S  Third     man  TRUE        0
14       0  36     0     0   26        S Second     man  TRUE        1
15       1  54     1     0   59        C  First   woman FALSE        1
16       0  24     0     0    7        S  Third     man  TRUE        0
17       0  47     0     0   34        S  First     man  TRUE        0
18       1  34     0     0   10        S Second   woman  TRUE        1
19       0  34     0     0   24        Q  Third     man  TRUE        0
20       1  36     1     0   26        S Second   woman FALSE        1
21       0  32     0     0    7        S  Third     man  TRUE        0
22       1  30     0     0   93        S  First   woman  TRUE        1
23       0  22     0     0    7        S  Third     man  TRUE        0
24       0  34     0     0    7        C  Third     man  TRUE        0
25       1  44     0     1   57        C  First   woman FALSE        1
26       0  34     0     0    7        C  Third     man  TRUE        0
27       0  40     0     0    7        Q  Third     man  TRUE        0
28       1  50     0     0   10        S Second   woman  TRUE        1
29       0  34     0     0  221        S  First     man  TRUE        0
30       0  39     0     0    7        S  Third     man  TRUE        0
```

31	0	23	2	1	11	S	Second	man	FALSE	0
32	0	2	1	1	26	S	Second	child	FALSE	1
33	0	34	0	0	7	C	Third	man	TRUE	0
34	0	17	1	1	7	C	Third	man	FALSE	0
35	1	34	0	2	22	C	Second	woman	FALSE	1
36	1	30	0	0	8	S	Third	woman	TRUE	0
37	1	7	0	2	26	S	Second	child	FALSE	1
38	0	45	0	0	26	S	First	man	TRUE	0
39	1	30	0	0	106	C	First	woman	TRUE	1
40	0	34	0	0	14	S	Third	man	TRUE	0
41	1	22	0	2	49	C	First	woman	FALSE	1
42	1	36	0	2	71	S	First	woman	FALSE	1
43	0	9	4	2	31	S	Third	child	FALSE	0
44	1	11	4	2	31	S	Third	child	FALSE	0
45	0	32	1	0	26	S	Second	man	FALSE	1
46	0	50	1	0	106	C	First	man	FALSE	0
47	0	64	0	0	26	S	First	man	TRUE	0
48	1	19	1	0	26	S	Second	woman	FALSE	1
49	0	34	0	0	13	C	Second	man	TRUE	1
50	0	33	1	1	20	S	Third	man	FALSE	0
51	0	8	1	1	36	S	Second	child	FALSE	1
52	0	17	0	2	110	C	First	man	FALSE	1
53	0	27	0	0	26	S	Second	man	TRUE	0
54	0	34	0	0	7	Q	Third	man	TRUE	0
55	0	22	0	0	7	C	Third	man	TRUE	1
56	1	22	0	0	7	S	Third	woman	TRUE	1
57	0	62	0	0	26	S	First	man	TRUE	0
58	1	48	1	0	39	C	First	woman	FALSE	1
59	0	34	0	0	227	C	First	man	TRUE	0
60	1	39	1	1	79	S	First	woman	FALSE	1
61	1	36	1	0	17	S	Third	woman	FALSE	1
62	0	34	0	0	7	Q	Third	man	TRUE	0
63	0	40	0	0	7	S	Third	man	TRUE	0
64	0	28	0	0	13	S	Second	man	TRUE	0
65	0	34	0	0	8	S	Third	man	TRUE	0
66	1	34	0	0	8	S	Third	woman	TRUE	0
67	0	24	2	0	24	S	Third	man	FALSE	0
68	0	19	0	0	7	S	Third	man	TRUE	0
69	1	29	0	4	21	S	Third	woman	FALSE	0
70	0	34	0	0	7	C	Third	man	TRUE	0

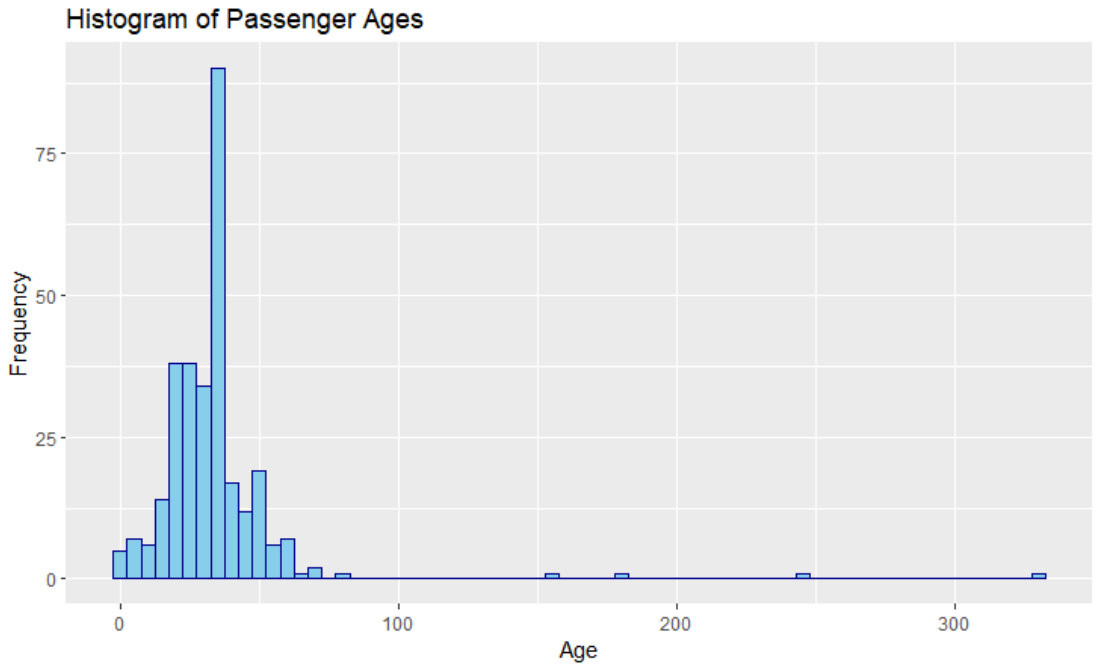
```
71       0  32     0     0    7        S  Third     man  TRUE        1
72       0  62     0     0   10        S Second     man  TRUE        1
73       1  53     2     0   51        S  First   woman FALSE        1
74       0  36     0     0   26        S  First     man  TRUE        1
75       1  34     0     0    7        Q  Third   woman  TRUE        1
76       0  16     0     0    8        S  Third     man  TRUE        0
77       0  19     0     0   14        S  Third     man  TRUE        0
78       1  34     0     0   13        S Second   woman  TRUE        1
79       1  39     1     0   55        S  First   woman FALSE        1
80       1  34     1     0   14        C  Third   woman FALSE        0
81       0  32     0     0    7        S  Third     man  TRUE        1
82       1  25     1     1   30        S Second   woman FALSE        1
83       1  39     1     1  110        C  First   woman FALSE        1
84       0  54     0     0   26        S Second     man  TRUE        0
85       0  36     0     0   40        C  First     man  TRUE        0
86       0  34     0     0    8        C  Third     man  TRUE        0
87       1  18     0     2   79        S  First   woman FALSE        1
88       0  47     0     0   15        S Second     man  TRUE        0
89       0  60     1     1   79        C  First     man FALSE        1
90       0  22     0     0    8        S  Third     man  TRUE        0
91       0  34     0     0    8        S  Third     man  TRUE        0
92       0  35     0     0    7        S  Third     man  TRUE        0
93       1  52     1     0   78        C  First   woman FALSE        1
94       0  47     0     0    7        S  Third     man  TRUE        0
95       1  34     0     2    7        Q  Third   woman FALSE        0
96       0  37     1     0   26        S Second     man FALSE        0
97       0  36     1     1   24        S  Third     man FALSE        0
98       1  34     0     0   33        S Second   woman  TRUE        1
99       0  49     0     0    0        S Second     man  TRUE        0
100      0  34     0     0    7        C  Third     man  TRUE        0
 [ reached 'max' / getOption("max.print") -- omitted 201 rows ]
> 
```

Data visualization:

the code checks if the "ggplot2" package is installed. If not, it installs the package using the install.packages function. Then, it loads the "ggplot2" library using the library function.

```
44 if (!require("ggplot2")) {
45   install.packages("ggplot2")
46 }
47 
48 library(ggplot2)
49 
50 ggplot(data = xdataset, aes(x = age)) +
51   geom_histogram(binwidth = 5, fill = "skyblue", color = "darkblue") +
52   labs(title = "Histogram of Passenger Ages", x = "Age", y = "Frequency")
```

A histogram is created using the ggplot function from the "ggplot2" package. The histogram represents the distribution of passenger ages from the xdataset. The geom_histogram function is used to create the histogram, specifying the binwidth, fill color, and line color. The labs function is used to set the title and labels for the x and y axes. The resulting plot is displayed.



Discussion & conclusion:

The analysis of the titanic passenger dataset involved data preprocessing steps and data visualization. The missing values were handled using mean imputation, ensuring that no missing data hindered subsequent analyses. Additionally, variable transformations were performed to enable more accurate analysis and interpretation. The resulting dataset, along with the generated histogram, provides valuable insights and sets the stage for further exploration and modeling tasks. It is important to note that any future analyses or modeling based on this dataset should consider the preprocessing steps performed here to ensure reliable results.