AMERICAN INTERNATIONAL UNIVERSITY-BANGLADESH (AIUB)

FACULTY OF SCIENCE & TECHNOLOGY

DEPARTMENT OF CSE

INTRODUCTION TO DATA SCIENCE

SUMMER 2022-2023

SECTION: C

Submitted by

Name: - AHMED IMTIAZ

ID: - 20-42933-1

Section: C

TABLE OF CONTENTS: -

PROJECT DESCRIPTION: - Here, we have been provided with the modified version of the Titanic dataset. In the given dataset, there is the age given who was on the ship and the gender, whether they were male, female or child. In the dataset, how many people had their siblings that data is also provided. The fare and the ticket price, the class of that people, and from where the passengers boarded the ship this data was also given in the dataset. Lastly, the persons who survived that tragic death that information has been provided in the dataset.

Data Science Project Mid Term.R × Titanic_Modified ×

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	0	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	1	38.00	1	0	71.2833	C	First	woman	FALL	1
3	1	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	1	35.00	1	0	53.1000	S	First	woman	FALL	1
5	0	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	0	NA	0	0	8.4583	Q	Third	man	TRUE	0
7	0	54.00	0	0	51.8625	S	First	man	TRUE	0
8	0	2.00	3	1	21.0750	S	Third	child	FALSE	0
9	1	27.00	0	2	11.1333	S	Third	woman	FALSE	1
10	1	14.00	1	0	30.0708	C	Second	child	FALSE	1
11	1	4.00	1	1	16.7000	S	Third	child	FALSE	1
12	1	58.00	0	0	26.5500	S	First	woman	TRUE	1
13	NA	20.00	0	0	8.0500	S	Third	man	TRUE	0
14	0	39.00	1	5	31.2750	S	Third	man	FALSE	0
15	1	14.00	0	0	7.8542	S	Third	child	TRUE	0
16	1	55.00	0	0	16.0000	S	Second	woman	TRUE	1
17	0	2.00	4	1	29.1250	Q	Third	child	FALSE	0
18	0	NA	0	0	13.0000	S	Second	man	TRUE	1
19	1	31.00	1	0	18.0000	S	Third	woman	FALSE	0
20	1	NA	0	0	7.2250	C	Third	woman	TRUE	1
21	0	35.00	0	0	26.0000	S	Second	man	TRUE	0
22	0	34.00	0	0	13.0000	S	Second	man	TRUE	1
23	1	15.00	0	0	8.0292	Q	Third	child	TRUE	1

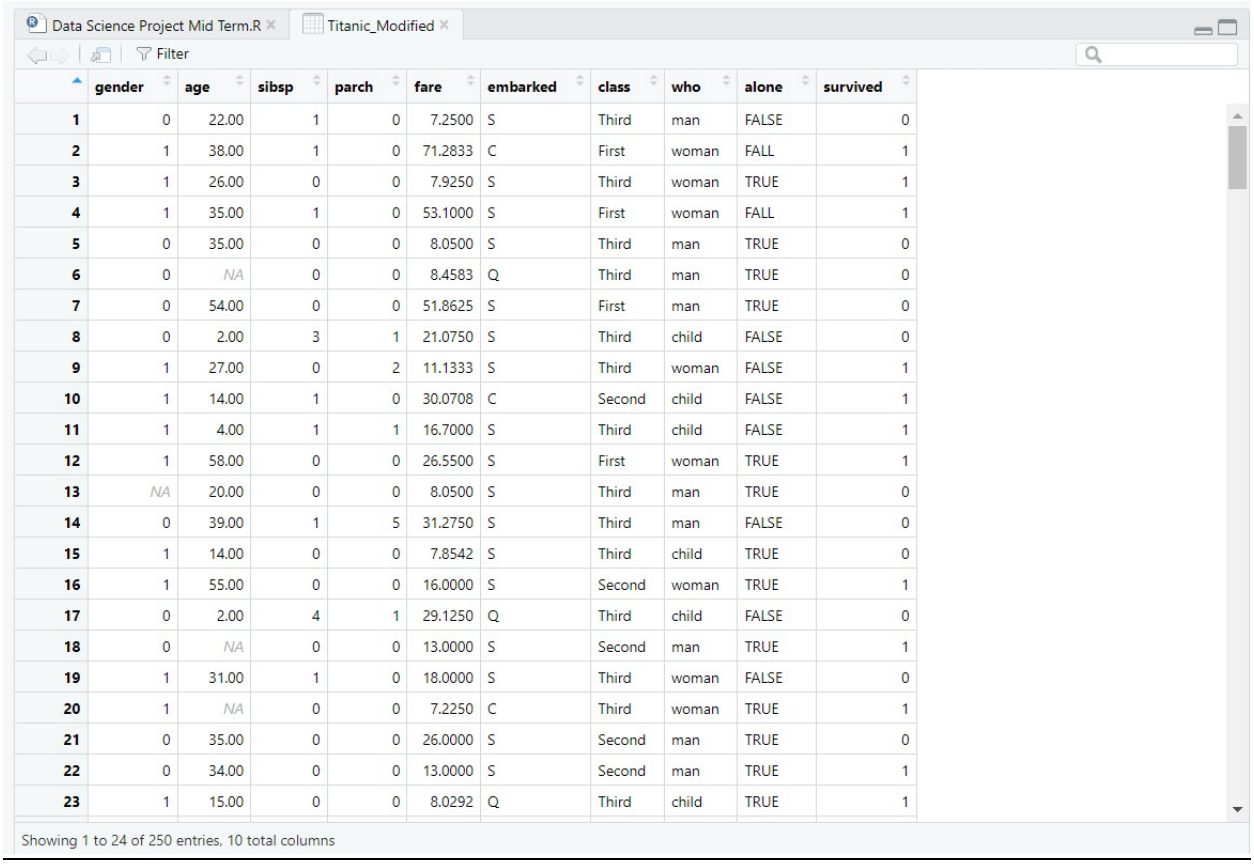
Showing 1 to 24 of 250 entries, 10 total columns

Importing Dataset (Titanic - Modified)

Code Segment

```
> library(readxl)
Warning message:
package 'readxl' was built under R version 4.3.1
> Titanic_Modified <- read_excel("D:/INTRODUCTION TO DATA SCIENCE/Project/Titanic - Modified.xlsx")
> View(Titanic_Modified)
>
```

Output



Data Science Project Mid Term.R | Titanic_Modified

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	0	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	1	38.00	1	0	71.2833	C	First	woman	FALL	1
3	1	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	1	35.00	1	0	53.1000	S	First	woman	FALL	1
5	0	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	0	NA	0	0	8.4583	Q	Third	man	TRUE	0
7	0	54.00	0	0	51.8625	S	First	man	TRUE	0
8	0	2.00	3	1	21.0750	S	Third	child	FALSE	0
9	1	27.00	0	2	11.1333	S	Third	woman	FALSE	1
10	1	14.00	1	0	30.0708	C	Second	child	FALSE	1
11	1	4.00	1	1	16.7000	S	Third	child	FALSE	1
12	1	58.00	0	0	26.5500	S	First	woman	TRUE	1
13	NA	20.00	0	0	8.0500	S	Third	man	TRUE	0
14	0	39.00	1	5	31.2750	S	Third	man	FALSE	0
15	1	14.00	0	0	7.8542	S	Third	child	TRUE	0
16	1	55.00	0	0	16.0000	S	Second	woman	TRUE	1
17	0	2.00	4	1	29.1250	Q	Third	child	FALSE	0
18	0	NA	0	0	13.0000	S	Second	man	TRUE	1
19	1	31.00	1	0	18.0000	S	Third	woman	FALSE	0
20	1	NA	0	0	7.2250	C	Third	woman	TRUE	1
21	0	35.00	0	0	26.0000	S	Second	man	TRUE	0
22	0	34.00	0	0	13.0000	S	Second	man	TRUE	1
23	1	15.00	0	0	8.0292	Q	Third	child	TRUE	1

Showing 1 to 24 of 250 entries, 10 total columns

Here we imported the dataset using the library function readxl in the data frame and then viewed the dataset.

Structure of the Dataset (Titanic - Modified)

Code Segment

```
4  str(Titanic_Modified)
```

Output

```
> str(Titanic_Modified)
tibble [250 × 10] (S3: tbl_df/tbl/data.frame)
 $ gender  : num [1:250] 0 1 1 1 0 0 0 0 1 1 ...
 $ age     : num [1:250] 22 38 26 35 35 NA 54 2 27 14 ...
 $ sibsp   : num [1:250] 1 1 0 1 0 0 0 3 0 1 ...
 $ parch   : num [1:250] 0 0 0 0 0 0 0 1 2 0 ...
 $ fare    : num [1:250] 7.25 71.28 7.92 53.1 8.05 ...
 $ embarked: chr [1:250] "S" "C" "S" "S" ...
 $ class   : chr [1:250] "Third" "First" "Third" "First" ...
 $ who     : chr [1:250] "man" "woman" "woman" "woman" ...
 $ alone   : chr [1:250] "FALSE" "FALL" "TRUE" "FALL" ...
 $ survived: num [1:250] 0 1 1 1 0 0 0 0 1 1 ...
>
```

We have shown the dataset structure by writing the code str, and then the whole structure appears.

Summary of the Dataset (Titanic - Modified)

Code Segment

```
6  summary(Titanic_Modified)
```

Output

```
> summary(Titanic_Modified)
     gender            age             sibsp           parch            fare           embarked
 Min.   :0.0000   Min.   :  0.83   Min.   :0.000   Min.   :0.000   Min.   :  0.000   Length:250
 1st Qu.:0.0000   1st Qu.: 19.00   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:  8.034   Class :character
 Median :0.0000   Median : 27.00   Median :0.000   Median :0.000   Median : 13.977   Mode  :character
 Mean   :0.3629   Mean   : 33.33   Mean   :0.656   Mean   :0.392   Mean   : 26.588
 3rd Qu.:1.0000   3rd Qu.: 37.00   3rd Qu.:1.000   3rd Qu.:0.000   3rd Qu.: 29.094
 Max.   :1.0000   Max.   :455.00   Max.   :8.000   Max.   :5.000   Max.   :263.000
 NA's   :13       NA's   :48
    class               who               alone             survived
 Length:250         Length:250         Length:250         Min.   :0.000
 Class :character   Class :character   Class :character   1st Qu.:0.000
 Mode  :character   Mode  :character   Mode  :character   Median :0.000
                                                          Mean   :0.344
                                                          3rd Qu.:1.000
                                                          Max.   :1.000

>
```

Here we showed the summary of the dataset by writing the code summary, and then all the summary of the dataset showed up like the median of the gender is 0.00, the mean is 0.3629, with 13 NA values.

Attributes Names of the Dataset (Titanic - Modified)

Code Segment

```
8  names(Titanic_Modified)
```

Output

```
> names(Titanic_Modified)
 [1] "gender"   "age"      "sibsp"    "parch"    "fare"     "embarked" "class"    "who"      "alone"
[10] "survived"
>
```

Here we have shown all the names of the dataset's attributes using code names like gender, age, sibsp, etc.

Attributes Data Types of the Dataset (Titanic - Modified)

Code Segment

```
typeof(Titanic_Modified$gender)
typeof(Titanic_Modified$age)
typeof(Titanic_Modified$sibsp)
typeof(Titanic_Modified$parch)
typeof(Titanic_Modified$fare)
typeof(Titanic_Modified$embarked)
typeof(Titanic_Modified$class)
typeof(Titanic_Modified$who)
typeof(Titanic_Modified$alone)
typeof(Titanic_Modified$survived)
```

Output

```
> typeof(Titanic_Modified$gender)
[1] "double"
> typeof(Titanic_Modified$age)
[1] "double"
> typeof(Titanic_Modified$sibsp)
[1] "double"
> typeof(Titanic_Modified$parch)
[1] "double"
> typeof(Titanic_Modified$fare)
[1] "double"
> typeof(Titanic_Modified$embarked)
[1] "character"
> typeof(Titanic_Modified$class)
[1] "character"
> typeof(Titanic_Modified$who)
[1] "character"
> typeof(Titanic_Modified$alone)
[1] "character"
> typeof(Titanic_Modified$survived)
[1] "double"
```

Here we have shown the data types in the dataset, like fare, which is a double attribute, and class is a character attribute.

Removing NA values from the Dataset (Titanic - Modified)

Code Segment

```
20
21  Titanic_Modified <- na.omit(Titanic_Modified)
```

Output

Before applying the code

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	0	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	1	38.00	1	0	71.2833	C	First	woman	FALL	1
3	1	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	1	35.00	1	0	53.1000	S	First	woman	FALL	1
5	0	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	0	*NA*	0	0	8.4583	Q	Third	man	TRUE	0
7	0	54.00	0	0	51.8625	S	First	man	TRUE	0
8	0	2.00	3	1	21.0750	S	Third	child	FALSE	0
9	1	27.00	0	2	11.1333	S	Third	woman	FALSE	1
10	1	14.00	1	0	30.0708	C	Second	child	FALSE	1
11	1	4.00	1	1	16.7000	S	Third	child	FALSE	1
12	1	58.00	0	0	26.5500	S	First	woman	TRUE	1
13	*NA*	20.00	0	0	8.0500	S	Third	man	TRUE	0
14	0	39.00	1	5	31.2750	S	Third	man	FALSE	0

After applying the code

Data Science Project Mid Term.R* | Titanic_Modified

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	0	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	1	38.00	1	0	71.2833	C	First	woman	FALL	1
3	1	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	1	35.00	1	0	53.1000	S	First	woman	FALL	1
5	0	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	0	54.00	0	0	51.8625	S	First	man	TRUE	0
7	0	2.00	3	1	21.0750	S	Third	child	FALSE	0
8	1	27.00	0	2	11.1333	S	Third	woman	FALSE	1
9	1	14.00	1	0	30.0708	C	Second	child	FALSE	1
10	1	4.00	1	1	16.7000	S	Third	child	FALSE	1
11	1	58.00	0	0	26.5500	S	First	woman	TRUE	1
12	0	39.00	1	5	31.2750	S	Third	man	FALSE	0
13	1	14.00	0	0	7.8542	S	Third	child	TRUE	0
14	1	55.00	0	0	16.0000	S	Second	woman	TRUE	1

Here we have used the na.omit code to remove all the null values from the dataset.

Removing Noisy values from the Dataset (Titanic - Modified) (Gender)

Code Segment

```
22
23 Titanic_Modified$gender[Titanic_Modified$gender == 0] <- "male"
24 Titanic_Modified
25
26 Titanic_Modified$gender[Titanic_Modified$gender == 1] <- "female"
27 Titanic_Modified
```

Output

Before applying the code

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	0	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	1	38.00	1	0	71.2833	C	First	woman	FALL	1
3	1	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	1	35.00	1	0	53.1000	S	First	woman	FALL	1
5	0	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	0	NA	0	0	8.4583	Q	Third	man	TRUE	0
7	0	54.00	0	0	51.8625	S	First	man	TRUE	0
8	0	2.00	3	1	21.0750	S	Third	child	FALSE	0
9	1	27.00	0	2	11.1333	S	Third	woman	FALSE	1
10	1	14.00	1	0	30.0708	C	Second	child	FALSE	1
11	1	4.00	1	1	16.7000	S	Third	child	FALSE	1
12	1	58.00	0	0	26.5500	S	First	woman	TRUE	1
13	NA	20.00	0	0	8.0500	S	Third	man	TRUE	0
14	0	39.00	1	5	31.2750	S	Third	man	FALSE	0

After applying the code

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	male	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	female	38.00	1	0	71.2833	C	First	woman	FALL	1
3	female	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	female	35.00	1	0	53.1000	S	First	woman	FALL	1
5	male	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	male	54.00	0	0	51.8625	S	First	man	TRUE	0
7	male	2.00	3	1	21.0750	S	Third	child	FALSE	0
8	female	27.00	0	2	11.1333	S	Third	woman	FALSE	1
9	female	14.00	1	0	30.0708	C	Second	child	FALSE	1
10	female	4.00	1	1	16.7000	S	Third	child	FALSE	1
11	female	58.00	0	0	26.5500	S	First	woman	TRUE	1
12	male	39.00	1	5	31.2750	S	Third	man	FALSE	0
13	female	14.00	0	0	7.8542	S	Third	child	TRUE	0
14	female	55.00	0	0	16.0000	S	Second	woman	TRUE	1

We have removed noisy values from the dataset before applying the code. Gender had the value of 0 and 1. Gender should have the attribute of character, not the double.

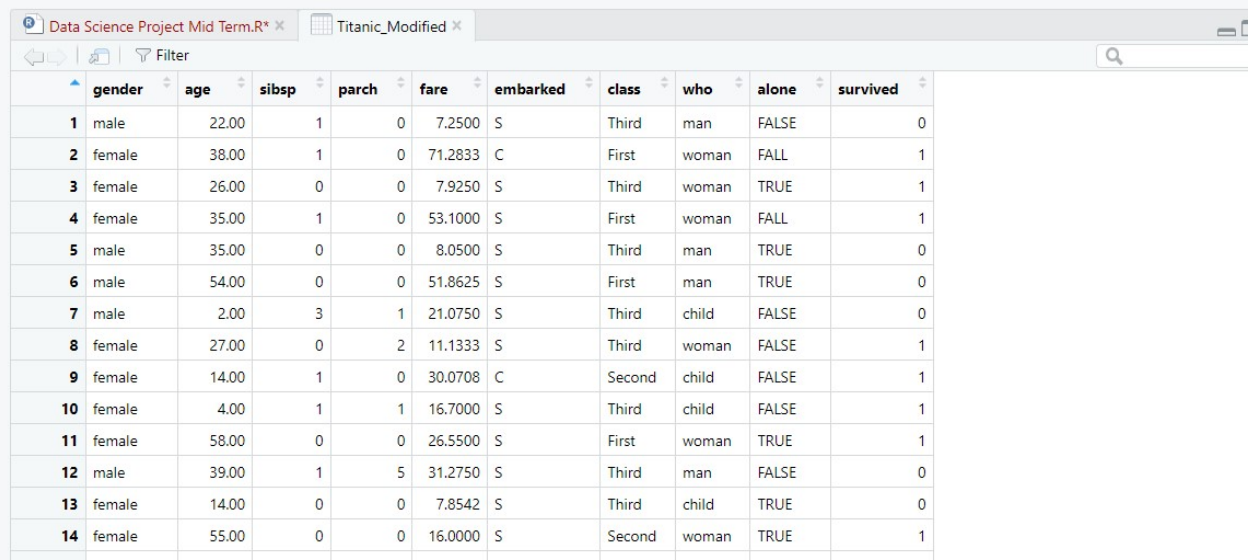
Removing Noisy values from the Dataset (Titanic - Modified) (FALL)

Code Segment

```
28
29 Titanic_Modified$alone[Titanic_Modified$alone == "FALL"] <- "FALSE"
30 Titanic_Modified
```

Output

Before applying the code

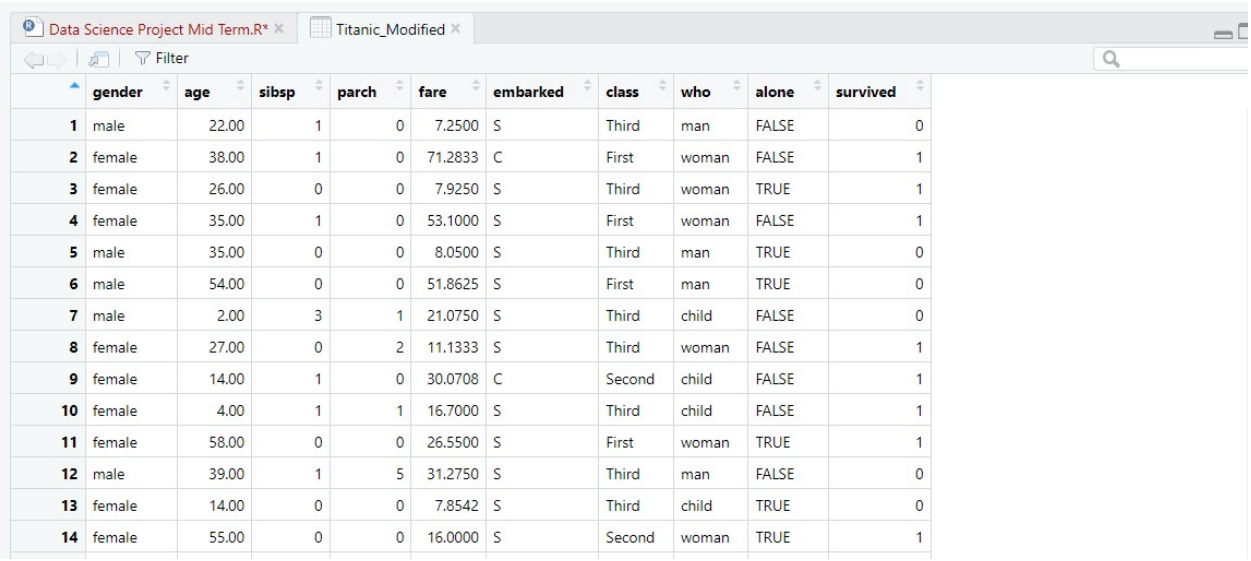


Data Science Project Mid Term.R* | Titanic_Modified

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	male	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	female	38.00	1	0	71.2833	C	First	woman	FALL	1
3	female	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	female	35.00	1	0	53.1000	S	First	woman	FALL	1
5	male	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	male	54.00	0	0	51.8625	S	First	man	TRUE	0
7	male	2.00	3	1	21.0750	S	Third	child	FALSE	0
8	female	27.00	0	2	11.1333	S	Third	woman	FALSE	1
9	female	14.00	1	0	30.0708	C	Second	child	FALSE	1
10	female	4.00	1	1	16.7000	S	Third	child	FALSE	1
11	female	58.00	0	0	26.5500	S	First	woman	TRUE	1
12	male	39.00	1	5	31.2750	S	Third	man	FALSE	0
13	female	14.00	0	0	7.8542	S	Third	child	TRUE	0
14	female	55.00	0	0	16.0000	S	Second	woman	TRUE	1

After applying the code



Data Science Project Mid Term.R* | Titanic_Modified

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	male	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	female	38.00	1	0	71.2833	C	First	woman	FALSE	1
3	female	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	female	35.00	1	0	53.1000	S	First	woman	FALSE	1
5	male	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	male	54.00	0	0	51.8625	S	First	man	TRUE	0
7	male	2.00	3	1	21.0750	S	Third	child	FALSE	0
8	female	27.00	0	2	11.1333	S	Third	woman	FALSE	1
9	female	14.00	1	0	30.0708	C	Second	child	FALSE	1
10	female	4.00	1	1	16.7000	S	Third	child	FALSE	1
11	female	58.00	0	0	26.5500	S	First	woman	TRUE	1
12	male	39.00	1	5	31.2750	S	Third	man	FALSE	0
13	female	14.00	0	0	7.8542	S	Third	child	TRUE	0
14	female	55.00	0	0	16.0000	S	Second	woman	TRUE	1

We have removed noisy values from the dataset before applying the code. alone had the value of fall, which will be false. After applying the code, the fall value changed to false.

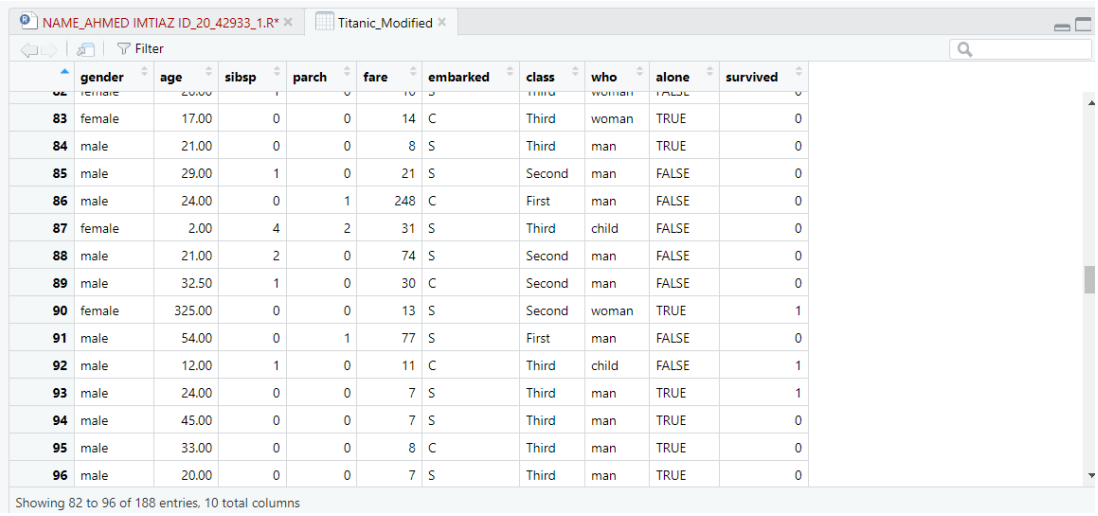
Removing Noisy values from the Dataset (Titanic - Modified) (AGE)

Code Segment

```
34 Titanic_Modified$age[Titanic_Modified$age == 325] <- 32.5
35 Titanic_Modified
36
37 Titanic_Modified$age[Titanic_Modified$age == 365] <- 36.5
38 Titanic_Modified
39
40 Titanic_Modified$age[Titanic_Modified$age == 455] <- 45.5
41 Titanic_Modified
42
```

Output

Before applying the code



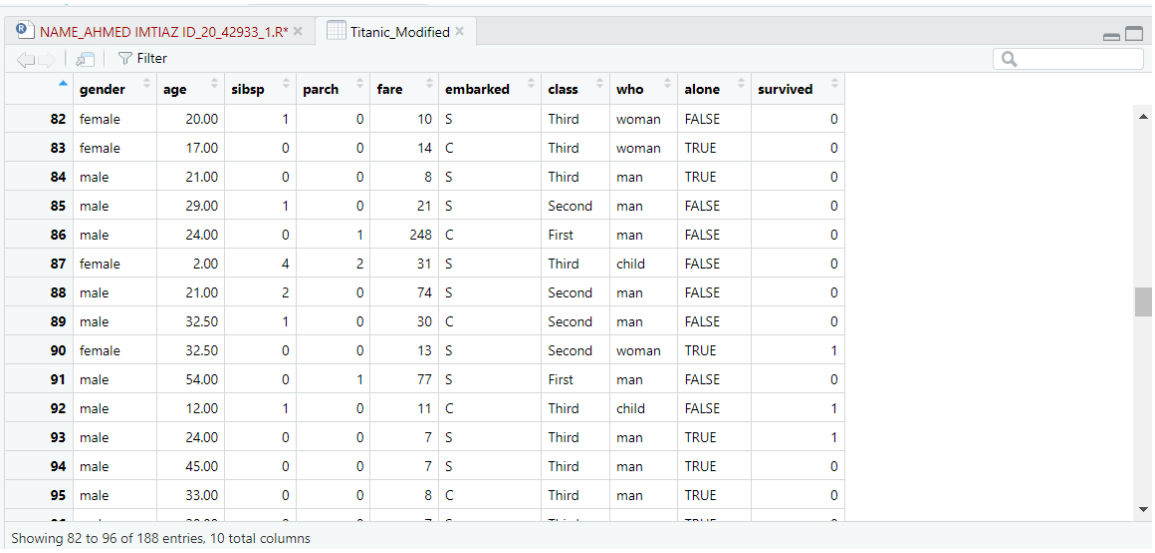
NAME_AHMED IMTIAZ ID_20_42933_1.R* | Titanic_Modified

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
83	female	17.00	0	0	14	C	Third	woman	TRUE	0
84	male	21.00	0	0	8	S	Third	man	TRUE	0
85	male	29.00	1	0	21	S	Second	man	FALSE	0
86	male	24.00	0	1	248	C	First	man	FALSE	0
87	female	2.00	4	2	31	S	Third	child	FALSE	0
88	male	21.00	2	0	74	S	Second	man	FALSE	0
89	male	32.50	1	0	30	C	Second	man	FALSE	0
90	female	325.00	0	0	13	S	Second	woman	TRUE	1
91	male	54.00	0	1	77	S	First	man	FALSE	0
92	male	12.00	1	0	11	C	Third	child	FALSE	1
93	male	24.00	0	0	7	S	Third	man	TRUE	1
94	male	45.00	0	0	7	S	Third	man	TRUE	0
95	male	33.00	0	0	8	C	Third	man	TRUE	0
96	male	20.00	0	0	7	S	Third	man	TRUE	0

Showing 82 to 96 of 188 entries, 10 total columns

After applying the code



NAME_AHMED IMTIAZ ID_20_42933_1.R* | Titanic_Modified

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
82	female	20.00	1	0	10	S	Third	woman	FALSE	0
83	female	17.00	0	0	14	C	Third	woman	TRUE	0
84	male	21.00	0	0	8	S	Third	man	TRUE	0
85	male	29.00	1	0	21	S	Second	man	FALSE	0
86	male	24.00	0	1	248	C	First	man	FALSE	0
87	female	2.00	4	2	31	S	Third	child	FALSE	0
88	male	21.00	2	0	74	S	Second	man	FALSE	0
89	male	32.50	1	0	30	C	Second	man	FALSE	0
90	female	32.50	0	0	13	S	Second	woman	TRUE	1
91	male	54.00	0	1	77	S	First	man	FALSE	0
92	male	12.00	1	0	11	C	Third	child	FALSE	1
93	male	24.00	0	0	7	S	Third	man	TRUE	1
94	male	45.00	0	0	7	S	Third	man	TRUE	0
95	male	33.00	0	0	8	C	Third	man	TRUE	0

Showing 82 to 96 of 188 entries, 10 total columns

Here we have converted the data from 325 to 32.5, a decimal value

Applying Data Formats in Dataset (Titanic - Modified) (Round-Fare)

Code Segment

```
31
32  Titanic_Modified$fare <- as.numeric(format(round(Titanic_Modified$fare, 0)))
33
```

Output

Before applying the code

Data Science Project Mid Term.R* | Titanic_Modified

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	male	22.00	1	0	7.2500	S	Third	man	FALSE	0
2	female	38.00	1	0	71.2833	C	First	woman	FALSE	1
3	female	26.00	0	0	7.9250	S	Third	woman	TRUE	1
4	female	35.00	1	0	53.1000	S	First	woman	FALSE	1
5	male	35.00	0	0	8.0500	S	Third	man	TRUE	0
6	male	54.00	0	0	51.8625	S	First	man	TRUE	0
7	male	2.00	3	1	21.0750	S	Third	child	FALSE	0
8	female	27.00	0	2	11.1333	S	Third	woman	FALSE	1
9	female	14.00	1	0	30.0708	C	Second	child	FALSE	1
10	female	4.00	1	1	16.7000	S	Third	child	FALSE	1
11	female	58.00	0	0	26.5500	S	First	woman	TRUE	1
12	male	39.00	1	5	31.2750	S	Third	man	FALSE	0
13	female	14.00	0	0	7.8542	S	Third	child	TRUE	0
14	female	55.00	0	0	16.0000	S	Second	woman	TRUE	1

After applying the code

Data Science Project Mid Term.R* | Titanic_Modified

Filter

	gender	age	sibsp	parch	fare	embarked	class	who	alone	survived
1	male	22.00	1	0	7	S	Third	man	FALSE	0
2	female	38.00	1	0	71	C	First	woman	FALSE	1
3	female	26.00	0	0	8	S	Third	woman	TRUE	1
4	female	35.00	1	0	53	S	First	woman	FALSE	1
5	male	35.00	0	0	8	S	Third	man	TRUE	0
6	male	54.00	0	0	52	S	First	man	TRUE	0
7	male	2.00	3	1	21	S	Third	child	FALSE	0
8	female	27.00	0	2	11	S	Third	woman	FALSE	1
9	female	14.00	1	0	30	C	Second	child	FALSE	1
10	female	4.00	1	1	17	S	Third	child	FALSE	1
11	female	58.00	0	0	27	S	First	woman	TRUE	1
12	male	39.00	1	5	31	S	Third	man	FALSE	0
13	female	14.00	0	0	8	S	Third	child	TRUE	0
14	female	55.00	0	0	16	S	Second	woman	TRUE	1

Showing 1 to 15 of 188 entries, 10 total columns

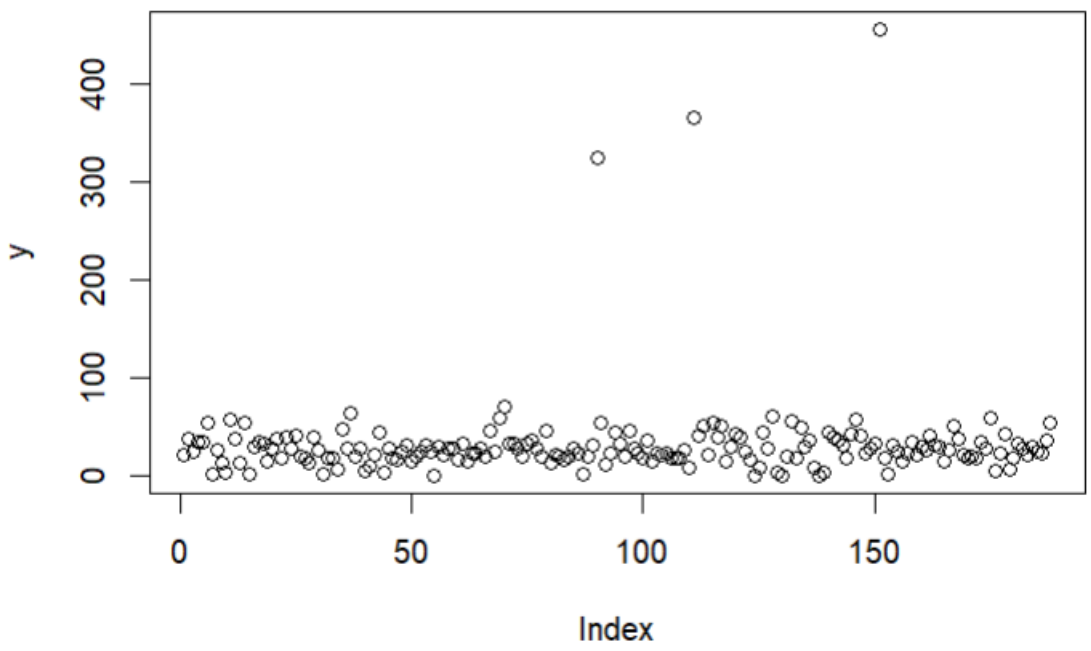
Here we have rounded up the fare values like before applying the code. The value of the fare was 7.2500. After applying the code, it got rounded up to 7 and became a whole number from an integer

Outliers in Dataset (Titanic - Modified) (AGE)

Code Segment

```
33
34 y <- Titanic_Modified$age
35 plot(y)
36
37
```

Output



Here we have used the code to plot the scatter diagram to show the outlier in the given dataset.

Finding Range in Dataset (Titanic - Modified) (AGE, Fare)

Code Segment

```
rdata = Titanic_Modified$age
print(rdata)
print(max(rdata, na.rm=TRUE))
print(min(rdata, na.rm=TRUE))
print(max(rdata, na.rm=TRUE))-print(min(rdata, na.rm=TRUE))


fdata = Titanic_Modified$fare
print(fdata)
print(max(fdata, na.rm=TRUE))
print(min(fdata, na.rm=TRUE))
print(max(fdata, na.rm=TRUE))-print(min(fdata, na.rm=TRUE))

```

Output

```
> rdata = Titanic_Modified$age
> print(rdata)
  [1] 22.00 38.00 26.00 35.00 35.00 54.00  2.00 27.00 14.00  4.00 58.00 39.00 14.00 55.00  2.00 31.00 35.00
 [18] 34.00 15.00 28.00 38.00 19.00 40.00 28.00 42.00 21.00 18.00 14.00 40.00 27.00  3.00 19.00 18.00  7.00
 [35] 49.00 29.00 65.00 21.00 28.50  5.00 11.00 22.00 45.00  4.00 29.00 19.00 17.00 26.00 32.00 16.00 21.00
 [52] 26.00 32.00 25.00  0.83 30.00 22.00 29.00 28.00 17.00 33.00 16.00 23.00 24.00 29.00 20.00 46.00 26.00
 [69] 59.00 71.00 34.00 34.00 28.00 21.00 33.00 37.00 28.00 21.00 47.00 14.50 22.00 20.00 17.00 21.00 29.00
 [86] 24.00  2.00 21.00 32.50 32.50 54.00 12.00 24.00 45.00 33.00 20.00 47.00 29.00 23.00 19.00 37.00 16.00
[103] 24.00 22.00 24.00 19.00 18.00 19.00 27.00  9.00 36.50 42.00 51.00 22.00 55.50 40.50 51.00 16.00 30.00
[120] 44.00 40.00 26.00 17.00  1.00  9.00 45.00 28.00 61.00  4.00  1.00 21.00 56.00 18.00 50.00 30.00 36.00
[137]  9.00  1.00  4.00 45.00 40.00 36.00 32.00 19.00 44.00 58.00 42.00 24.00 28.00 34.00 45.50 18.00  2.00
[154] 32.00 26.00 16.00 24.00 35.00 22.00 31.00 27.00 42.00 32.00 30.00 16.00 27.00 51.00 38.00 22.00 19.00
[171] 20.50 18.00 35.00 29.00 59.00  5.00 24.00 44.00  8.00 19.00 33.00 29.00 22.00 30.00 25.00 24.00 37.00
[188] 54.00
> print(max(rdata, na.rm=TRUE))
[1] 71
> print(min(rdata, na.rm=TRUE))
[1] 0.83
> print(max(rdata, na.rm=TRUE))-print(min(rdata, na.rm=TRUE))
[1] 71
[1] 0.83
[1] 70.17
>
>
> fdata = Titanic_Modified$fare
> print(fdata)
  [1]   7  71   8  53   8  52  21  11  30  17  27  31   8  16  29  18  26  13   8  36  31 263  28  82  52
 [26]   8  18  11   9  21  42   8  18  40  77  26  62  10   7  28  47   7  83  28  10   8   8   9  10  47
 [51]  74  14  56   8  29  12   9  10  47  10  16  34 263   8   8   8  61  21   7  35  23  26   8  77   9
 [76]   8   8   8  52  14   8  10  14   8  21 248  31  74  30  13  77  11   7   7   8   7  14  26  15  26
[101]  53   9  79   8  16   7  12  37   8  34  26  13  13  67   8  14  61   8   8  16  16   8   9  40  21
[126]  28  56  34  29  11   8  31   8  29  13   0  31  39  22  27  16   8  13  13  28 147   8  13  10   6
[151]   7   8  10  16  19   8   7  21   7 113   8  27  76  10   8  13   8  90   9  10   7  13  83   8  14
[176]  31  10  26  26  10  12  10   7   7   8  14  53  26
> print(max(fdata, na.rm=TRUE))
[1] 263
> print(min(fdata, na.rm=TRUE))
```

Here we have used the code rdata to find out the range of the age and fare. The highest value, lowest value, and the middle value.

Finding Univariate values in Dataset (Titanic - Modified) (Age, Fare)

Code Segment

```
59 mean(Titanic_Modified$age)
60 median(Titanic_Modified$age)
61 names(sort(-table(Titanic_Modified$age)))[1]
62
63 mean(Titanic_Modified$fare)
64 median(Titanic_Modified$fare)
65 names(sort(-table(Titanic_Modified$fare)))[1]
```

Output

```
> mean(Titanic_Modified$age)
[1] 27.9592
> median(Titanic_Modified$age)
[1] 27
> names(sort(-table(Titanic_Modified$age)))[1]
[1] "19"
>
> mean(Titanic_Modified$fare)
[1] 28.00532
> median(Titanic_Modified$fare)
[1] 14
> names(sort(-table(Titanic_Modified$fare)))[1]
[1] "8"
>
```

Here we have found out the mean, mode, and median, which are the univariate values of the dataset of the age and fare.

Finding Standard Deviation values in Dataset (Titanic - Modified) (Age, Fare)

Code Segment

```
67 x <- Titanic_Modified$age
68 sd(x)
69 y <- Titanic_Modified$fare
70 sd(y)
71
```

Output

```
> x <- Titanic_Modified$age
> sd(x)
[1] 14.26584
> y <- Titanic_Modified$fare
> sd(y)
[1] 37.25307
>
```

Here we have found the standard deviation of the age and fare of the dataset.

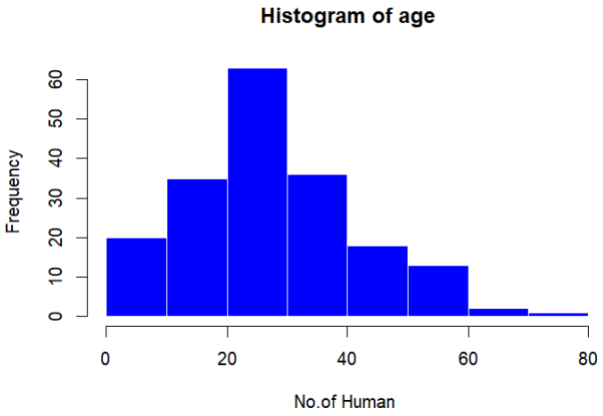
Histogram values Dataset (Titanic - Modified) (Age, Fare)

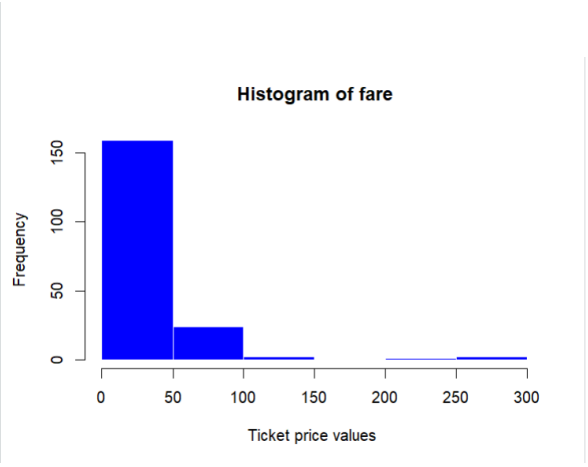
Code Segment

```
71
72 age <- Titanic_Modified$age
73 hist(age, xlab = "No.of Human ",
74      col = "blue", border = "white")
75
```

```
75
76 fare <- Titanic_Modified$fare
77 hist(fare, xlab = "Ticket price values ",
78      col = "blue", border = "white")
79
```

Output





Here we have shown the histogram of the age and fare of the dataset.