BAN130 - Programming for Analytics

NBB Section

Project Report

Titanic Survival Analysis

Group 7

Group Members:

1. Name - Udbhav Singh Chauhan

   Student ID - 146782206

   E-Mail ID - uschauhan@myseneca.ca

2. Name - Pruthviben Jamin Patel

   Email ID - pjpatel44@myseneca.ca

   Student ID – 139602205

1. Introduction

In this project we are going to analyze the chances of survival of Titanic passengers, based on their socio-economic status, gender, age, and port of embarkation. The dataset we're using is named titanic_data.csv and its specifications are available here.

Our base dataset will have the following columns:

Pclass - (ticket class)

Sex - (gender)

Age - (age)

embarked - (port of embarkation)

Precisely, we will investigate the following questions:

How does the ticket class impact the chances of survival?

How does gender impact the chances of survival?

How does age impact the chances of survival?

How does the port of embarkation impact the chances of survival?

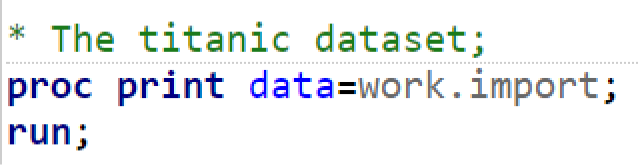
We are going to do the cleaning and pre-processing of the dataset first, then we are going to check the outliers, after that we are going to check how these attributes are affecting the survival and at last, we have also used the visualizations to make analysis clearer using bar graph, box plots, mosaic plot.

2. Dataset Description

The dataset that we are going to use is titanic dataset.

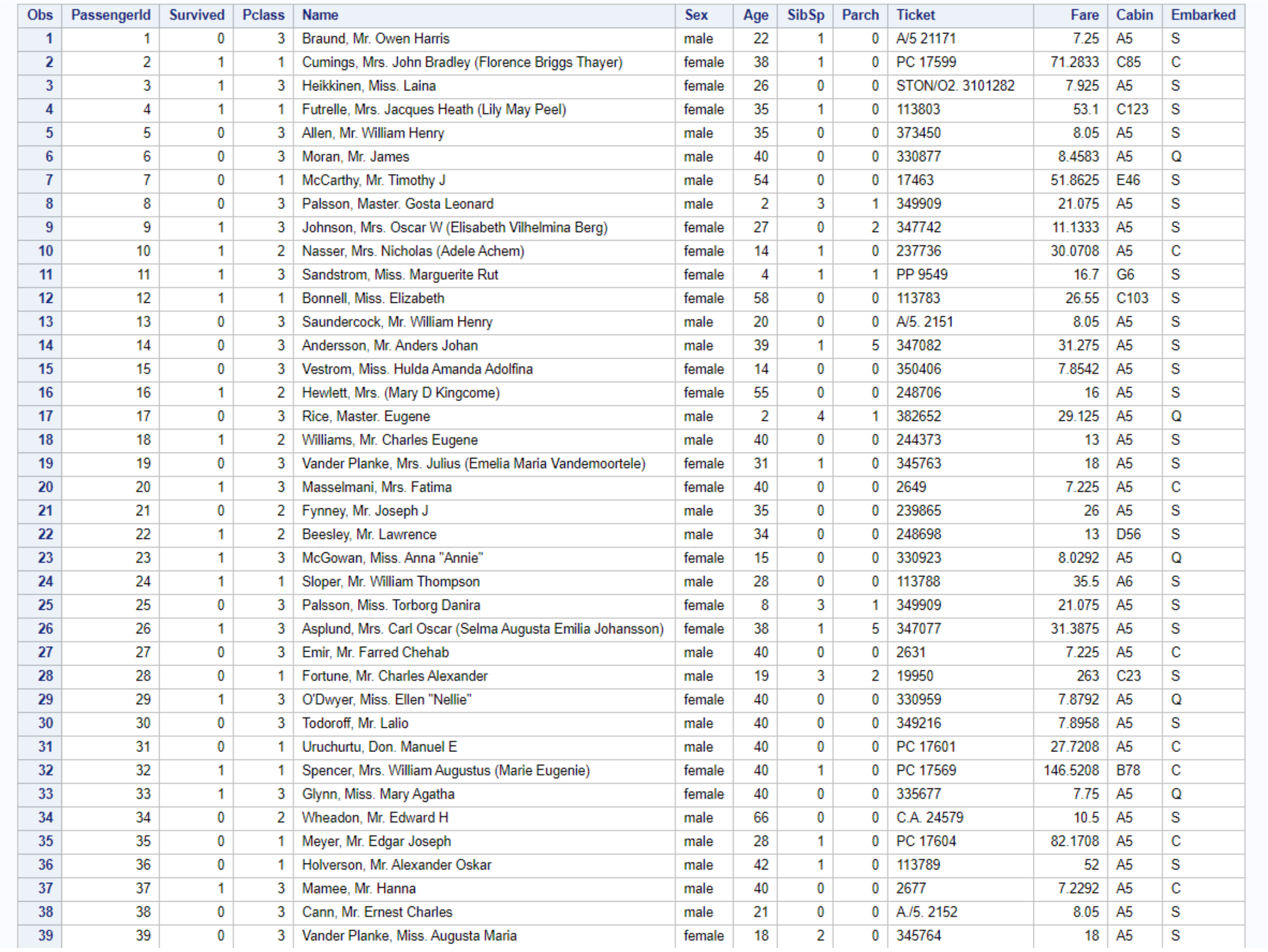
There are 891 data values and 12 attributes/features of the dataset.

SAS Code:



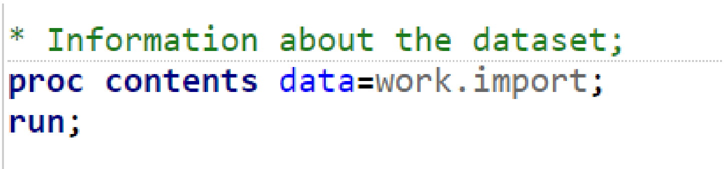
```
* The titanic dataset;
proc print data=work.import;
run;
```

Output:



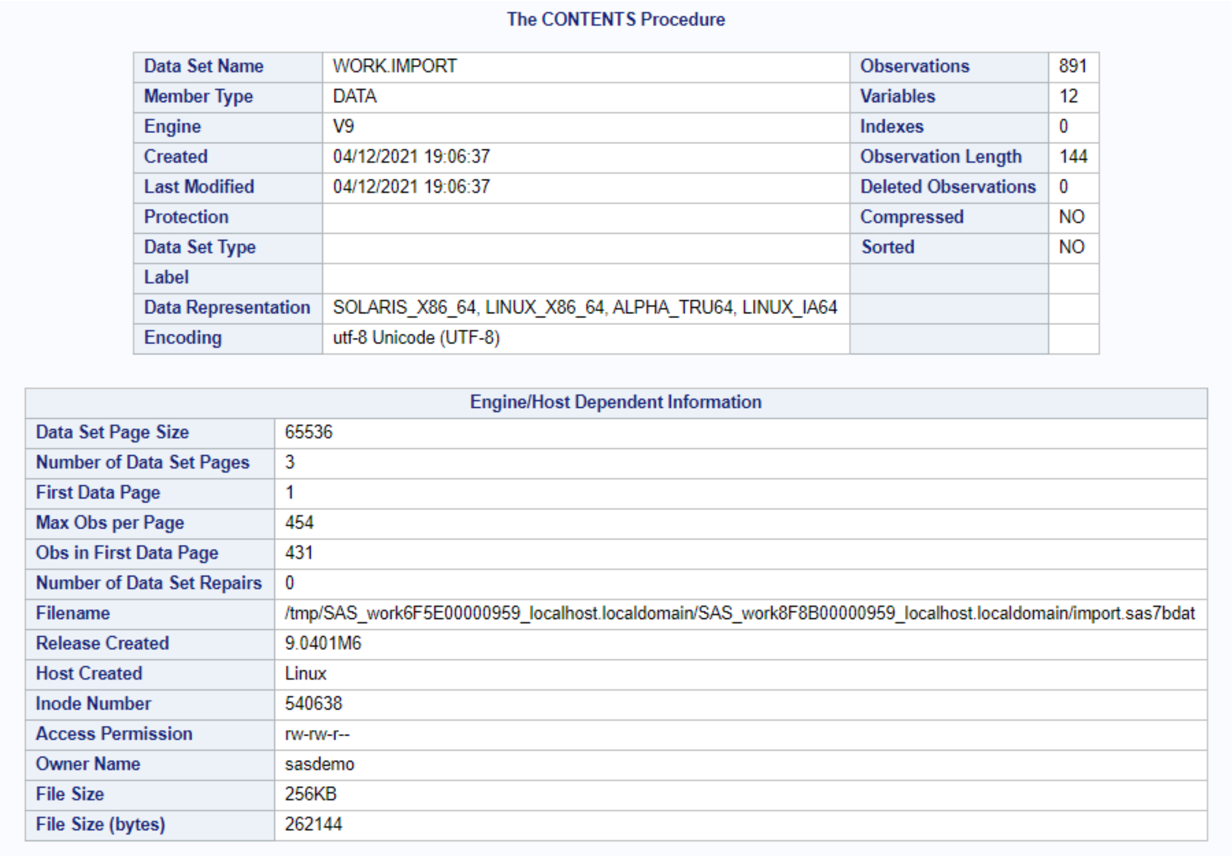
Obs	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25	A5	S
2	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925	A5	S
4	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S
5	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05	A5	S
6	6	0	3	Moran, Mr. James	male	40	0	0	330877	8.4583	A5	Q
7	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075	A5	S
9	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333	A5	S
10	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708	A5	C
11	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.7	G6	S
12	12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S
13	13	0	3	Saundercock, Mr. William Henry	male	20	0	0	A/5. 2151	8.05	A5	S
14	14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31.275	A5	S
15	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.8542	A5	S
16	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16	A5	S
17	17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.125	A5	Q
18	18	1	2	Williams, Mr. Charles Eugene	male	40	0	0	244373	13	A5	S
19	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	female	31	1	0	345763	18	A5	S
20	20	1	3	Masselmani, Mrs. Fatima	female	40	0	0	2649	7.225	A5	C
21	21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26	A5	S
22	22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698	13	D56	S
23	23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923	8.0292	A5	Q
24	24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788	35.5	A6	S
25	25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909	21.075	A5	S
26	26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)	female	38	1	5	347077	31.3875	A5	S
27	27	0	3	Emir, Mr. Farred Chehab	male	40	0	0	2631	7.225	A5	C
28	28	0	1	Fortune, Mr. Charles Alexander	male	19	3	2	19950	263	C23	S
29	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	40	0	0	330959	7.8792	A5	Q
30	30	0	3	Todoroff, Mr. Lalio	male	40	0	0	349216	7.8958	A5	S
31	31	0	1	Uruchurtu, Don. Manuel E	male	40	0	0	PC 17601	27.7208	A5	C
32	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female	40	1	0	PC 17569	146.5208	B78	C
33	33	1	3	Glynn, Miss. Mary Agatha	female	40	0	0	335677	7.75	A5	Q
34	34	0	2	Wheadon, Mr. Edward H	male	66	0	0	C.A. 24579	10.5	A5	S
35	35	0	1	Meyer, Mr. Edgar Joseph	male	28	1	0	PC 17604	82.1708	A5	C
36	36	0	1	Holverson, Mr. Alexander Oskar	male	42	1	0	113789	52	A5	S
37	37	1	3	Mamee, Mr. Hanna	male	40	0	0	2677	7.2292	A5	C
38	38	0	3	Cann, Mr. Ernest Charles	male	21	0	0	A./5. 2152	8.05	A5	S
39	39	0	3	Vander Planke, Miss. Augusta Maria	female	18	2	0	345764	18	A5	S

SAS Code:



```
* Information about the dataset;
proc contents data=work.import;
run;
```

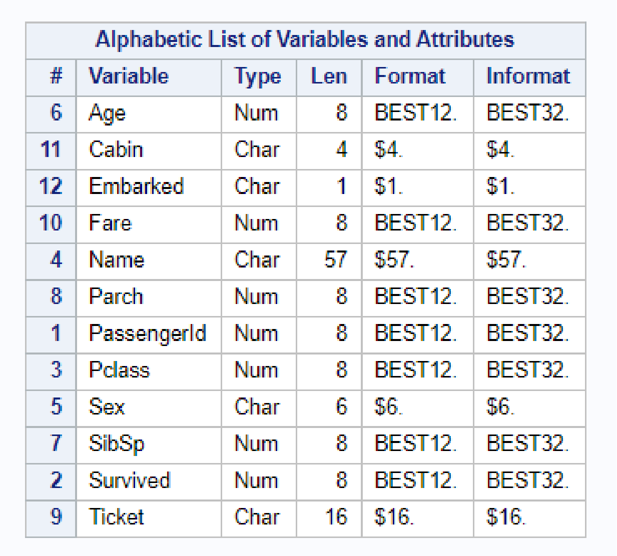
Output:



The CONTENTS Procedure

Data Set Name	WORK.IMPORT	Observations	891
Member Type	DATA	Variables	12
Engine	V9	Indexes	0
Created	04/12/2021 19:06:37	Observation Length	144
Last Modified	04/12/2021 19:06:37	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	3
First Data Page	1
Max Obs per Page	454
Obs in First Data Page	431
Number of Data Set Repairs	0
Filename	/tmp/SAS_work6F5E00000959_localhost.localdomain/SAS_work8F8B00000959_localhost.localdomain/import.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	540638
Access Permission	rw-rw-r--
Owner Name	sasdemo
File Size	256KB
File Size (bytes)	262144



Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
6	Age	Num	8	BEST12.	BEST32.
11	Cabin	Char	4	$4.	$4.
12	Embarked	Char	1	$1.	$1.
10	Fare	Num	8	BEST12.	BEST32.
4	Name	Char	57	$57.	$57.
8	Parch	Num	8	BEST12.	BEST32.
1	PassengerId	Num	8	BEST12.	BEST32.
3	Pclass	Num	8	BEST12.	BEST32.
5	Sex	Char	6	$6.	$6.
7	SibSp	Num	8	BEST12.	BEST32.
2	Survived	Num	8	BEST12.	BEST32.
9	Ticket	Char	16	$16.	$16.

Metadata/Data Dictionary

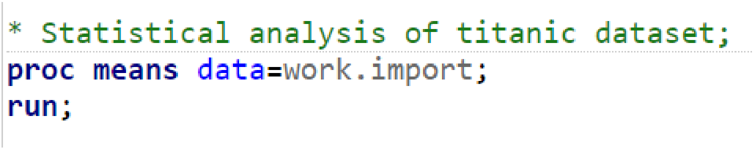
Variable Name	Description	Type
survival	Survival of the passenger (0 = No; 1 = Yes)	Number
class	Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)	Number
name	Name of the passenger	Character
sex	Sex of the passenger	Character
age	Age of the passenger	Number
sibsp	Number of Siblings/Spouses Aboard	Number
parch	Number of Parents/Children Aboard	Number
ticket	Ticket Number	Character
fare	Passenger Fare	Number
cabin	Cabin	Character
embarked	Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)	Character

3. Exploratory Analysis

Statistical Analysis

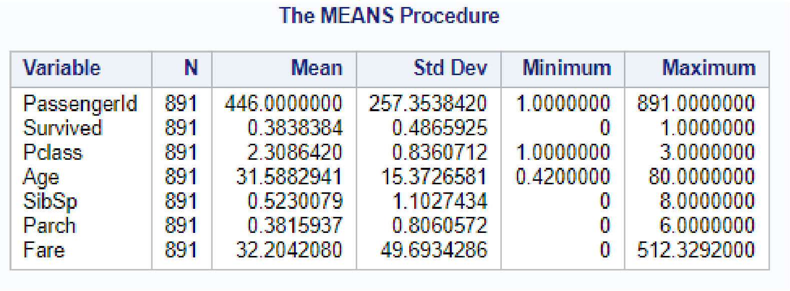
We can get the information about different statistical variables such as min, max, mean, standard deviation, variance.

SAS Code:



```
* Statistical analysis of titanic dataset;
proc means data=work.import;
run;
```

Output:



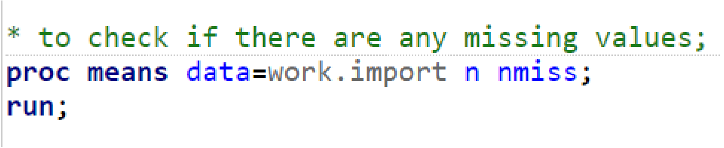
The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
PassengerId	891	446.0000000	257.3538420	1.0000000	891.0000000
Survived	891	0.3838384	0.4865925	0	1.0000000
Pclass	891	2.3086420	0.8360712	1.0000000	3.0000000
Age	891	31.5882941	15.3726581	0.4200000	80.0000000
SibSp	891	0.5230079	1.1027434	0	8.0000000
Parch	891	0.3815937	0.8060572	0	6.0000000
Fare	891	32.2042080	49.6934286	0	512.3292000

Checking Missing Values:

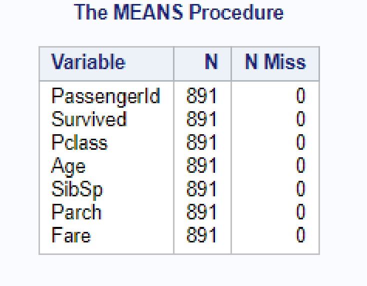
- To check if there are any missing values in the dataset or not, we use means procedure with nmiss. We can see from the output that there are no missing values.

SAS Code:



```
* to check if there are any missing values;
proc means data=work.import n nmiss;
run;
```

Output:

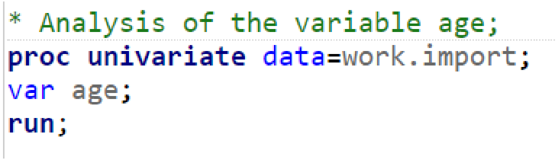


The MEANS Procedure

Variable	N	N Miss
PassengerId	891	0
Survived	891	0
Pclass	891	0
Age	891	0
SibSp	891	0
Parch	891	0
Fare	891	0

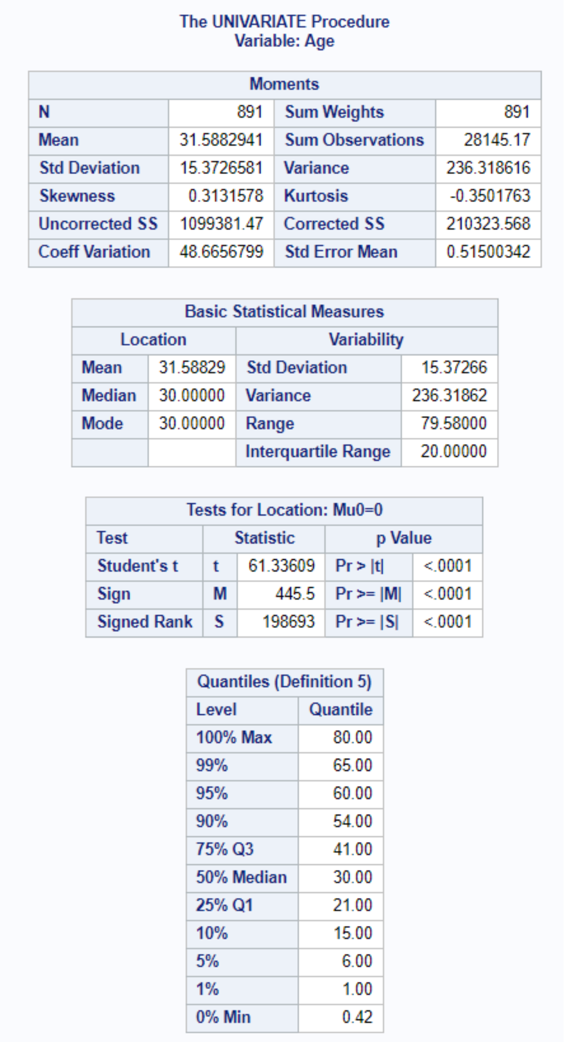
- Next, we use Univariate procedure with the variables age.

SAS Code:



```
* Analysis of the variable age;
proc univariate data=work.import;
var age;
run;
```

Output:



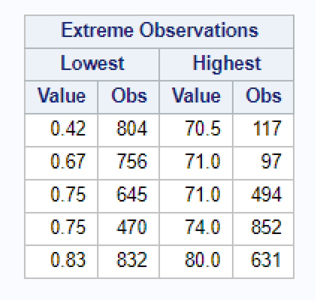
The UNIVARIATE Procedure
Variable: Age

Moments			
N	891	Sum Weights	891
Mean	31.5882941	Sum Observations	28145.17
Std Deviation	15.3726581	Variance	236.318616
Skewness	0.3131578	Kurtosis	-0.3501763
Uncorrected SS	1099381.47	Corrected SS	210323.568
Coeff Variation	48.6656799	Std Error Mean	0.51500342

Basic Statistical Measures			
Location		Variability	
Mean	31.58829	Std Deviation	15.37266
Median	30.00000	Variance	236.31862
Mode	30.00000	Range	79.58000
		Interquartile Range	20.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	61.33609	Pr > |t|	<.0001
Sign	M	445.5	Pr >= |M|	<.0001
Signed Rank	S	198693	Pr >= |S|	<.0001

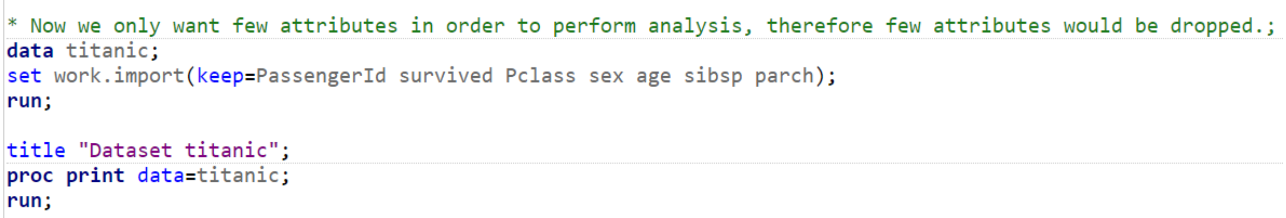
Quantiles (Definition 5)	
Level	Quantile
100% Max	80.00
99%	65.00
95%	60.00
90%	54.00
75% Q3	41.00
50% Median	30.00
25% Q1	21.00
10%	15.00
5%	6.00
1%	1.00
0% Min	0.42



Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.42	804	70.5	117
0.67	756	71.0	97
0.75	645	71.0	494
0.75	470	74.0	852
0.83	832	80.0	631

- Now we only want few attributes in order to perform analysis, therefore few attributes would be dropped.

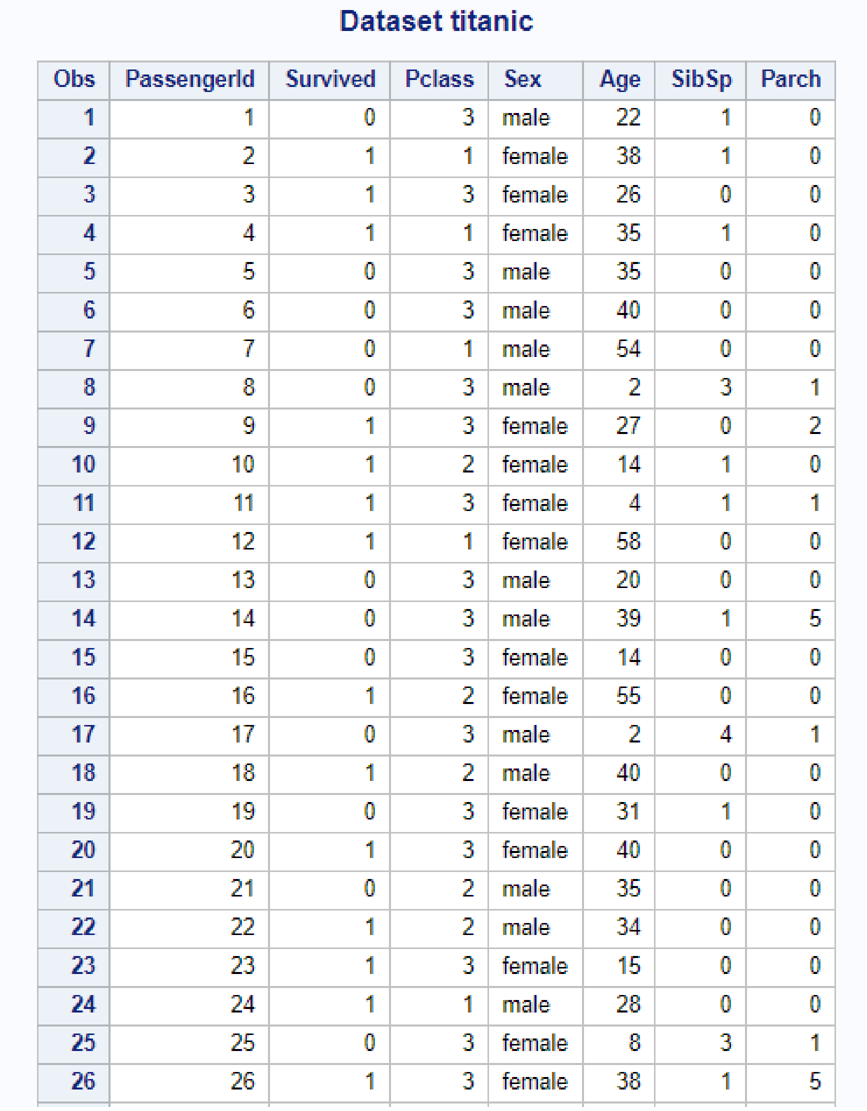
SAS Code:



```
* Now we only want few attributes in order to perform analysis, therefore few attributes would be dropped.;
data titanic;
set work.import(keep=PassengerId survived Pclass sex age sibsp parch);
run;

title "Dataset titanic";
proc print data=titanic;
run;
```

Output:

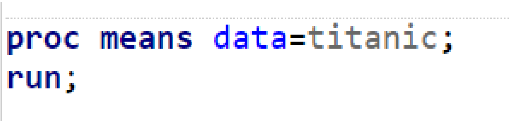


Dataset titanic

Obs	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch
1	1	0	3	male	22	1	0
2	2	1	1	female	38	1	0
3	3	1	3	female	26	0	0
4	4	1	1	female	35	1	0
5	5	0	3	male	35	0	0
6	6	0	3	male	40	0	0
7	7	0	1	male	54	0	0
8	8	0	3	male	2	3	1
9	9	1	3	female	27	0	2
10	10	1	2	female	14	1	0
11	11	1	3	female	4	1	1
12	12	1	1	female	58	0	0
13	13	0	3	male	20	0	0
14	14	0	3	male	39	1	5
15	15	0	3	female	14	0	0
16	16	1	2	female	55	0	0
17	17	0	3	male	2	4	1
18	18	1	2	male	40	0	0
19	19	0	3	female	31	1	0
20	20	1	3	female	40	0	0
21	21	0	2	male	35	0	0
22	22	1	2	male	34	0	0
23	23	1	3	female	15	0	0
24	24	1	1	male	28	0	0
25	25	0	3	female	8	3	1
26	26	1	3	female	38	1	5

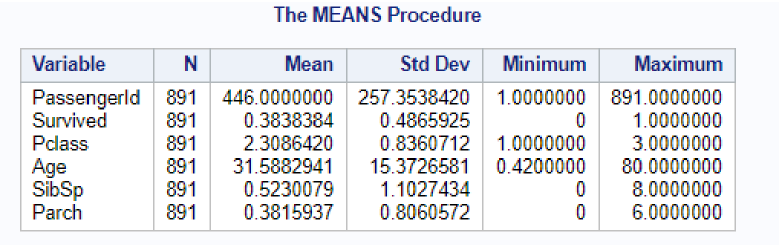
- Now, we can again check the statistical analysis of our dataset titanic

SAS Code:



```
proc means data=titanic;
run;
```

Output:



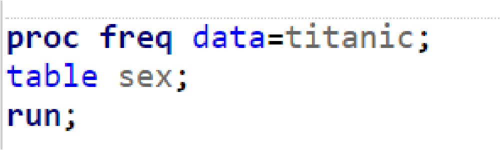
The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
PassengerId	891	446.0000000	257.3538420	1.0000000	891.0000000
Survived	891	0.3838384	0.4865925	0	1.0000000
Pclass	891	2.3086420	0.8360712	1.0000000	3.0000000
Age	891	31.5882941	15.3726581	0.4200000	80.0000000
SibSp	891	0.5230079	1.1027434	0	8.0000000
Parch	891	0.3815937	0.8060572	0	6.0000000

Variable – Sex:

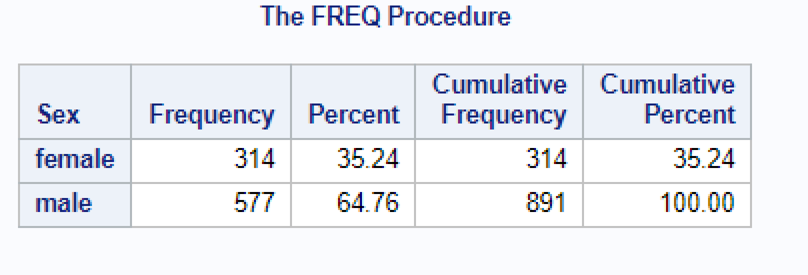
- Here we can see in the output that only six attributes are visible, that is because variable Sex is non-numerical data. Now we can apply FREQ procedure for categorical values.

SAS Code:



```
proc freq data=titanic;
table sex;
run;
```

Output:

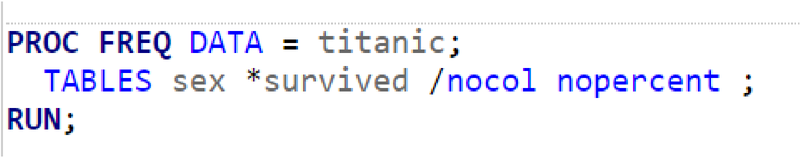


The FREQ Procedure

Sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
female	314	35.24	314	35.24
male	577	64.76	891	100.00

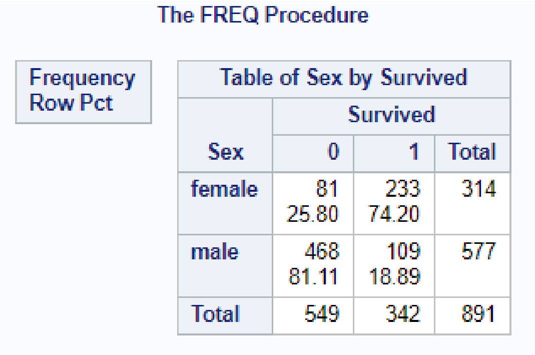
- Now we can use FREQ PROC to determine the survivorship by sex on the Titanic.

SAS Code:



```
PROC FREQ DATA = titanic;
  TABLES sex *survived /nocol nopercent ;
RUN;
```

Output:



The FREQ Procedure

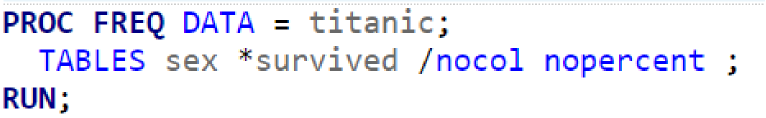
Frequency Row Pct

Table of Sex by Survived

	Survived		
Sex	0	1	Total
female	81 25.80	233 74.20	314
male	468 81.11	109 18.89	577
Total	549	342	891

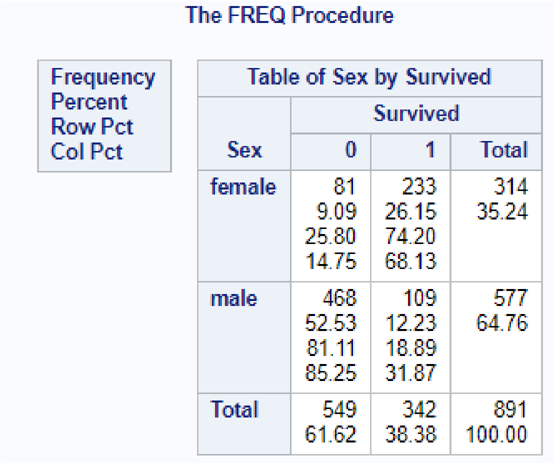
- We can see that 74.20% of women survived and 18.89% of men.

SAS Code:



```
PROC FREQ DATA = titanic;
  TABLES sex *survived /nocol nopercent ;
RUN;
```

Output:



The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of Sex by Survived

	Survived		
Sex	0	1	Total
female	81 9.09 25.80 14.75	233 26.15 74.20 68.13	314 35.24
male	468 52.53 81.11 85.25	109 12.23 18.89 31.87	577 64.76
Total	549 61.62	342 38.38	891 100.00

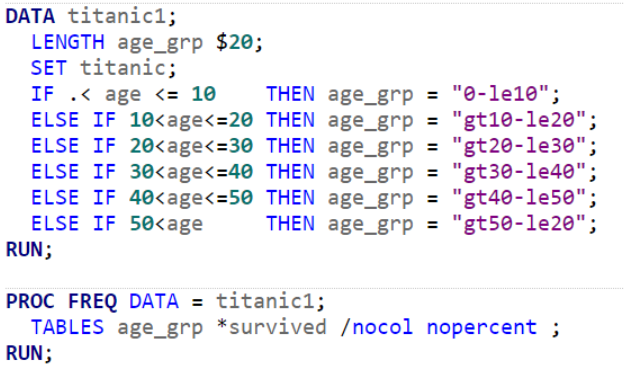
- This two-way frequency table can be read following way:

  1. 81 females did not survive (0) while 233 did (1).
  2. 52.3% of the total passengers were males and did not survive. 26.15% of the total passengers were female and did survive.
  3. Of the male passengers, 81.11% died vs 74.40% of the female passengers survived.
  4. For those that died 85.25% were male. For those that survived 68.13% were female.

Variable – Class:

- We can also include the variable "class" into the analysis proc freq analysis. An interesting finding is the high proportion of females who survived in first and second class (both over 90%) and the only 50% survival rate of females in 3rd class.

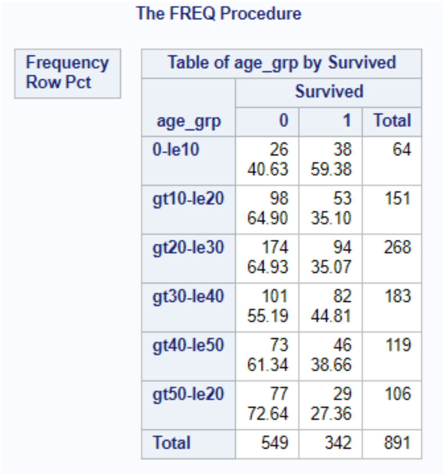
SAS Code:



```
DATA titanic1;
  LENGTH age_grp $20;
  SET titanic;
  IF .< age <= 10     THEN age_grp = "0-le10";
  ELSE IF 10<age<=20 THEN age_grp = "gt10-le20";
  ELSE IF 20<age<=30 THEN age_grp = "gt20-le30";
  ELSE IF 30<age<=40 THEN age_grp = "gt30-le40";
  ELSE IF 40<age<=50 THEN age_grp = "gt40-le50";
  ELSE IF 50<age      THEN age_grp = "gt50-le20";
RUN;

PROC FREQ DATA = titanic1;
  TABLES age_grp *survived /nocol nopercent ;
RUN;
```

Output:



The FREQ Procedure

Frequency
Row Pct

Table of age_grp by Survived

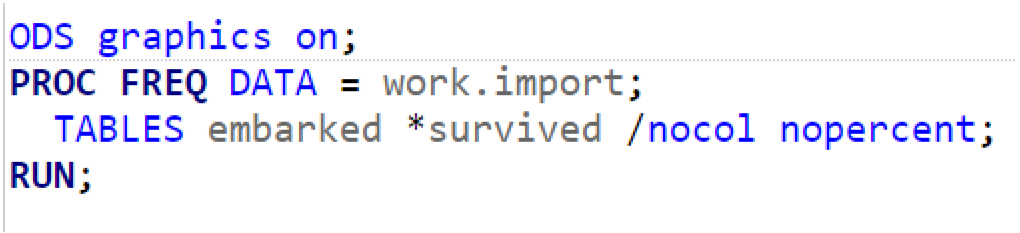
	Survived		
age_grp	0	1	Total
0-le10	26 40.63	38 59.38	64
gt10-le20	98 64.90	53 35.10	151
gt20-le30	174 64.93	94 35.07	268
gt30-le40	101 55.19	82 44.81	183
gt40-le50	73 61.34	46 38.66	119
gt50-le20	77 72.64	29 27.36	106
Total	549	342	891

- For example, this table shows that for children under the age of 10, 59.38% survived.

Variable- Embarked:

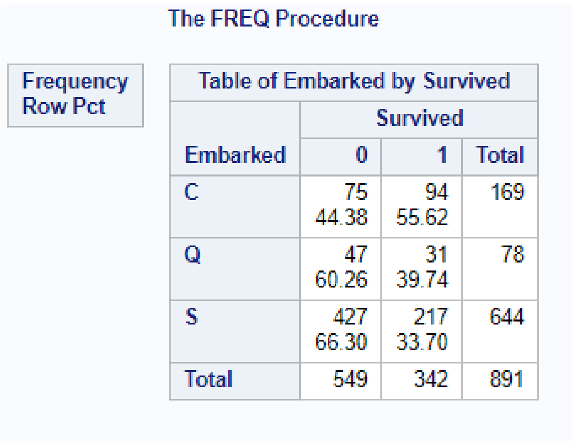
- Finally, we can see how the port where the passengers embarked made a difference in survivorship. We can also use the proc freq to graphically visualize this.

SAS Code:



```
ODS graphics on;
PROC FREQ DATA = work.import;
  TABLES embarked *survived /nocol nopercent;
RUN;
```

Output:



The FREQ Procedure

Frequency
Row Pct

Table of Embarked by Survived

	Survived		
Embarked	0	1	Total
C	75 44.38	94 55.62	169
Q	47 60.26	31 39.74	78
S	427 66.30	217 33.70	644
Total	549	342	891

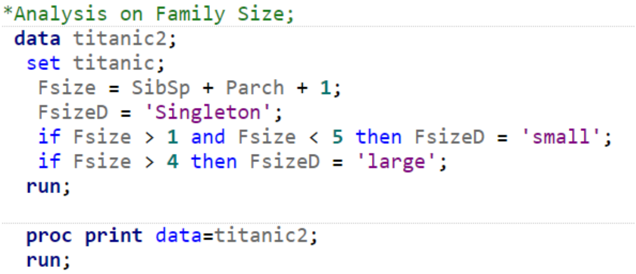
- For example, those that embarked from Cherbourne (C), 55.36% survived. A random finding with no basis for prediction use but interesting none the less.

Variable- Parch and Sibp

- Now we are calculating family size by adding the variables “Parch” and “Sibp”.

- Within the data step we are reading the sas table “full” and rewriting the table(using “set”) with edited variables “Fsize” and “FsizeD”. “FsizeD” is a discretized version of the “Fsize” variable which can be achieved using simple “if then” statement within SAS.

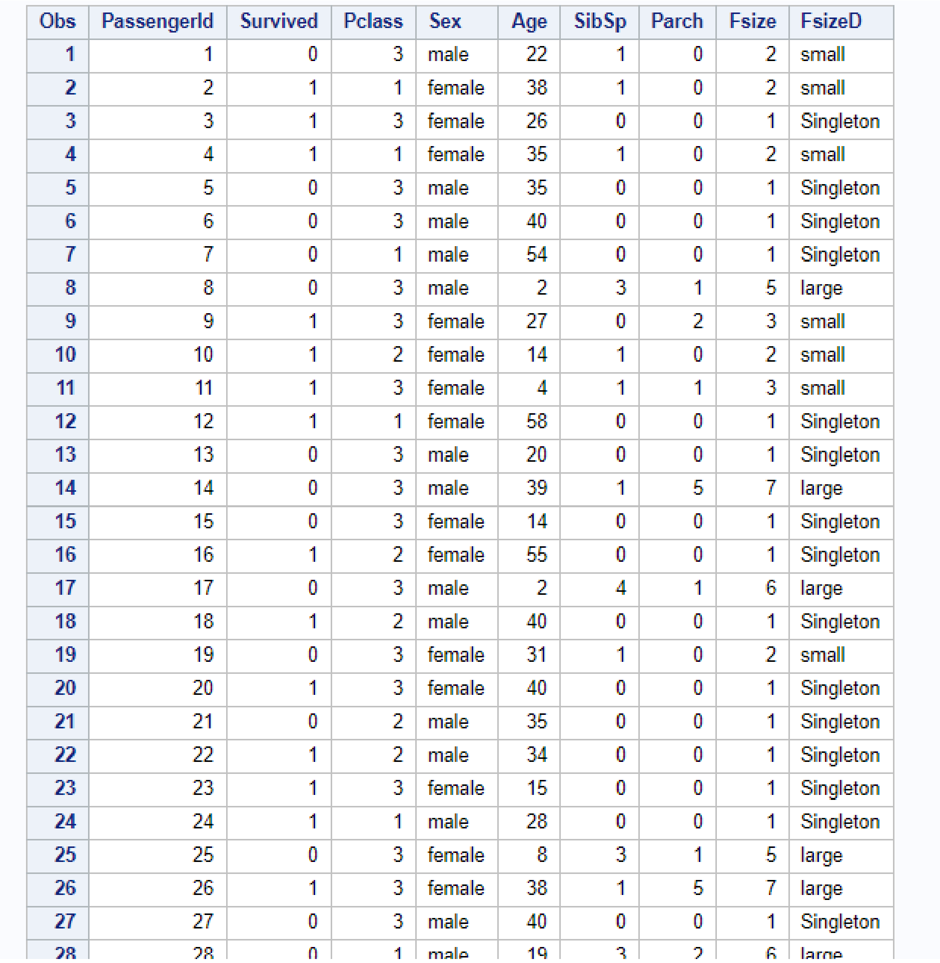
SAS Code:



```
*Analysis on Family Size;
 data titanic2;
  set titanic;
   Fsize = SibSp + Parch + 1;
   FsizeD = 'Singleton';
   if Fsize > 1 and Fsize < 5 then FsizeD = 'small';
   if Fsize > 4 then FsizeD = 'large';
  run;

  proc print data=titanic2;
  run;
```

Output:



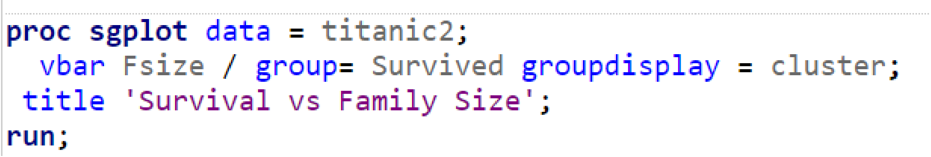
Obs	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fsize	FsizeD
1	1	0	3	male	22	1	0	2	small
2	2	1	1	female	38	1	0	2	small
3	3	1	3	female	26	0	0	1	Singleton
4	4	1	1	female	35	1	0	2	small
5	5	0	3	male	35	0	0	1	Singleton
6	6	0	3	male	40	0	0	1	Singleton
7	7	0	1	male	54	0	0	1	Singleton
8	8	0	3	male	2	3	1	5	large
9	9	1	3	female	27	0	2	3	small
10	10	1	2	female	14	1	0	2	small
11	11	1	3	female	4	1	1	3	small
12	12	1	1	female	58	0	0	1	Singleton
13	13	0	3	male	20	0	0	1	Singleton
14	14	0	3	male	39	1	5	7	large
15	15	0	3	female	14	0	0	1	Singleton
16	16	1	2	female	55	0	0	1	Singleton
17	17	0	3	male	2	4	1	6	large
18	18	1	2	male	40	0	0	1	Singleton
19	19	0	3	female	31	1	0	2	small
20	20	1	3	female	40	0	0	1	Singleton
21	21	0	2	male	35	0	0	1	Singleton
22	22	1	2	male	34	0	0	1	Singleton
23	23	1	3	female	15	0	0	1	Singleton
24	24	1	1	male	28	0	0	1	Singleton
25	25	0	3	female	8	3	1	5	large
26	26	1	3	female	38	1	5	7	large
27	27	0	3	male	40	0	0	1	Singleton
28	28	0	1	male	19	3	2	6	large

4. Visualization

- We can use different visualization for various attributes, in order to gain insights in our dataset.
- The next step is to visualize survival prospects for the difference observed genders with respect to various family sizes.
- To do this we use the “SGPLOT” function, which requires an input dataset. Then we visualize the frequency for each different family sizes, and feed in the other variables to group the visualization.

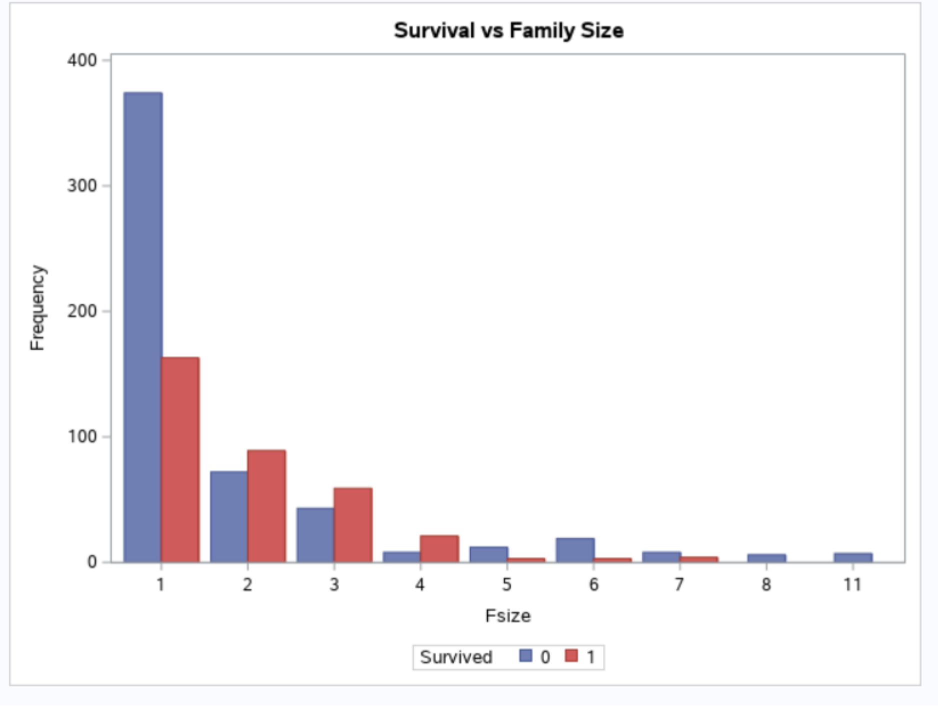
Variables: Family size:

SAS Code:



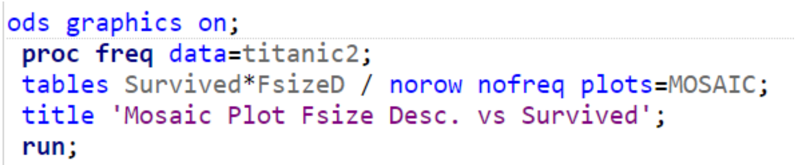
```
proc sgplot data = titanic2;
  vbar Fsize / group= Survived groupdisplay = cluster;
 title 'Survival vs Family Size';
run;
```

Output:



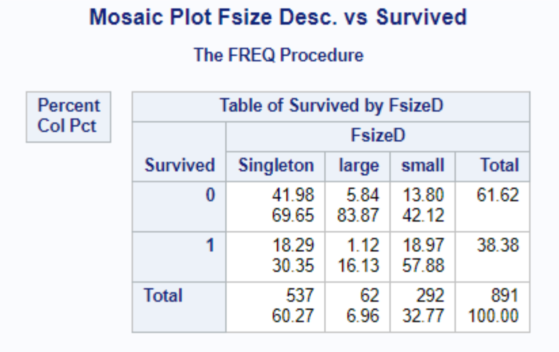
- The next step is the creation of a mosaic plot to visualize the same information as above, with an additional step of using discretized family size instead of the family sizes shown above.
- This process can be achieved using PROC FREQ, and passing an additional instruction to construct a mosaic plot.

SAS Code:



```
ods graphics on;
 proc freq data=titanic2;
 tables Survived*FsizeD / norow nofreq plots=MOSAIC;
 title 'Mosaic Plot Fsize Desc. vs Survived';
 run;
```

Output:



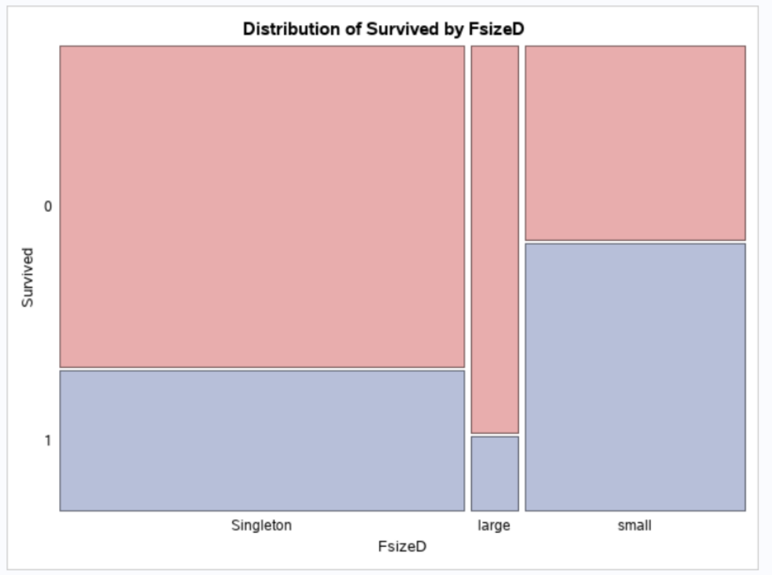
Mosaic Plot Fsize Desc. vs Survived

The FREQ Procedure

Percent
Col Pct

Table of Survived by FsizeD

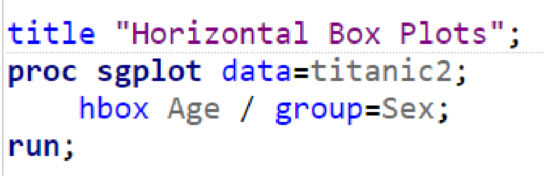
	FsizeD			
Survived	Singleton	large	small	Total
0	41.98 69.65	5.84 83.87	13.80 42.12	61.62
1	18.29 30.35	1.12 16.13	18.97 57.88	38.38
Total	537 60.27	62 6.96	292 32.77	891 100.00



Variables: Age and Sex

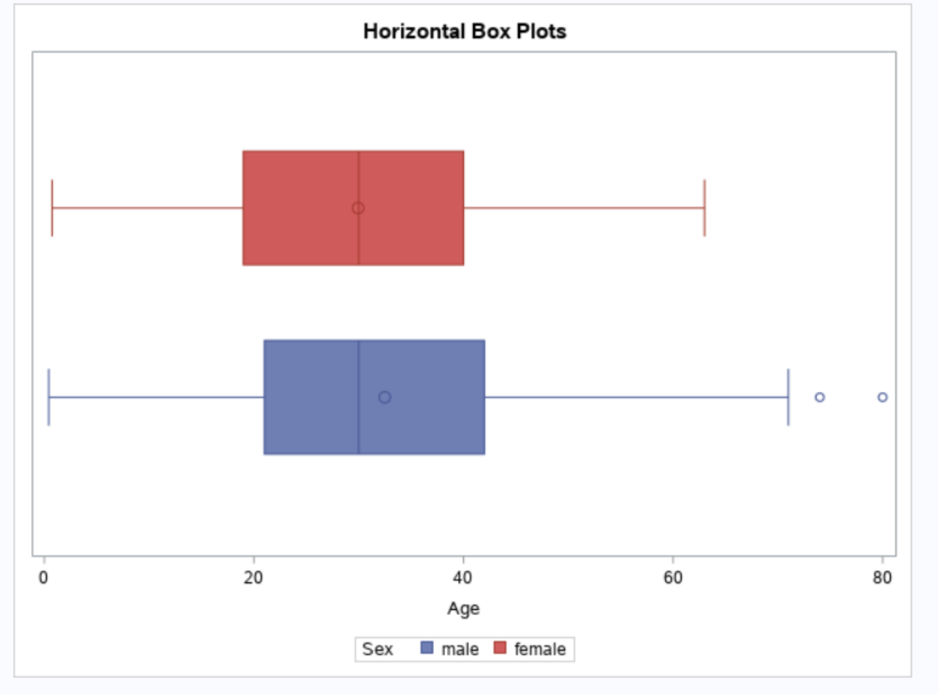
- Next, we can visualize the Age and Sex variables using box plots.

SAS Code:

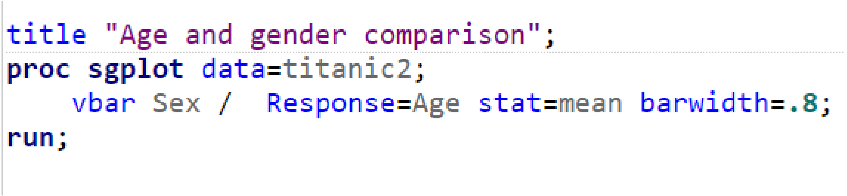


```
title "Horizontal Box Plots";
proc sgplot data=titanic2;
    hbox Age / group=Sex;
run;
```

Output:

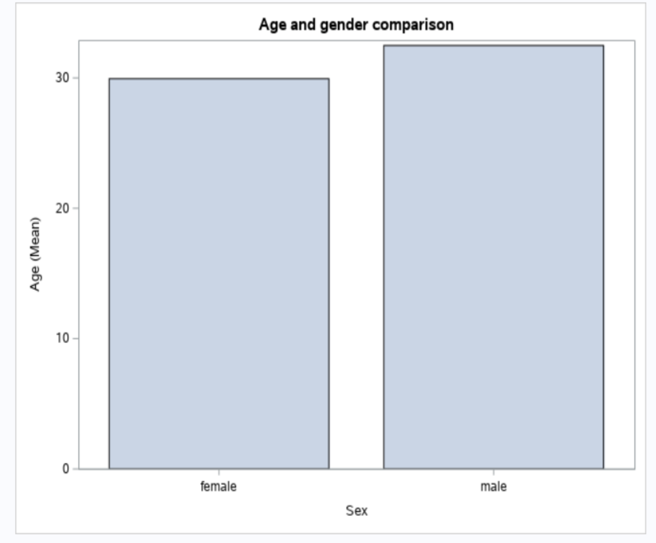


SAS Code:



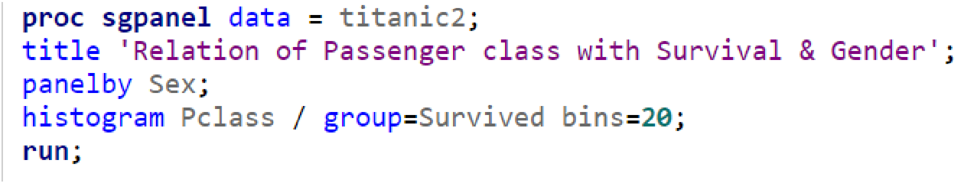
```
title "Age and gender comparison";
proc sgplot data=titanic2;
    vbar Sex /  Response=Age stat=mean barwidth=.8;
run;
```

Output:



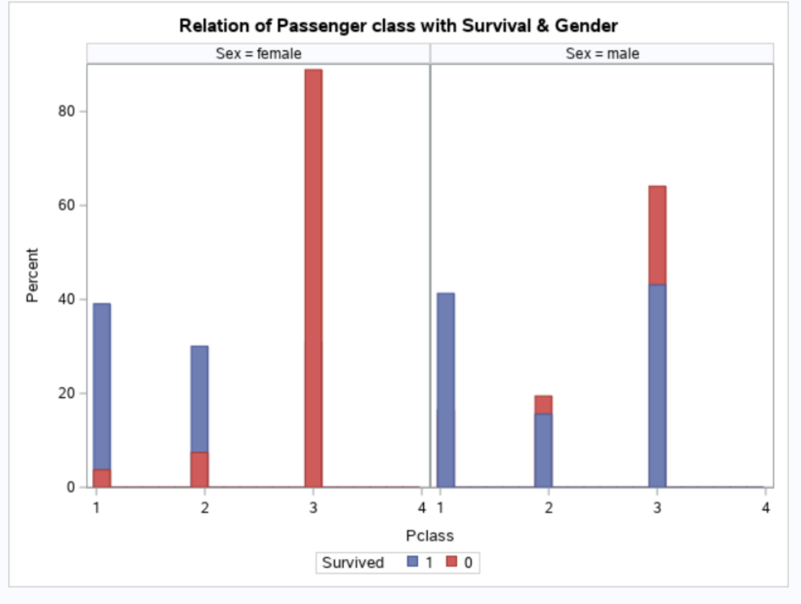
- We use histograms to see how well Passenger Class & Sex have an impact on survivability outcomes.

SAS Code:



```
proc sgpanel data = titanic2;
title 'Relation of Passenger class with Survival & Gender';
panelby Sex;
histogram Pclass / group=Survived bins=20;
run;
```

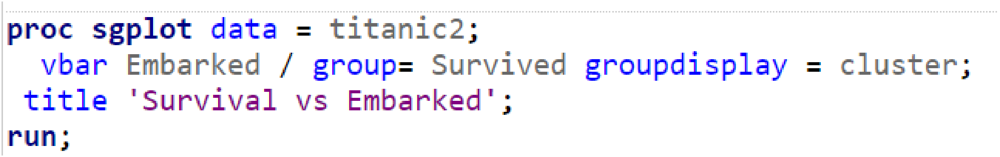
Output:



Variable: Embarked

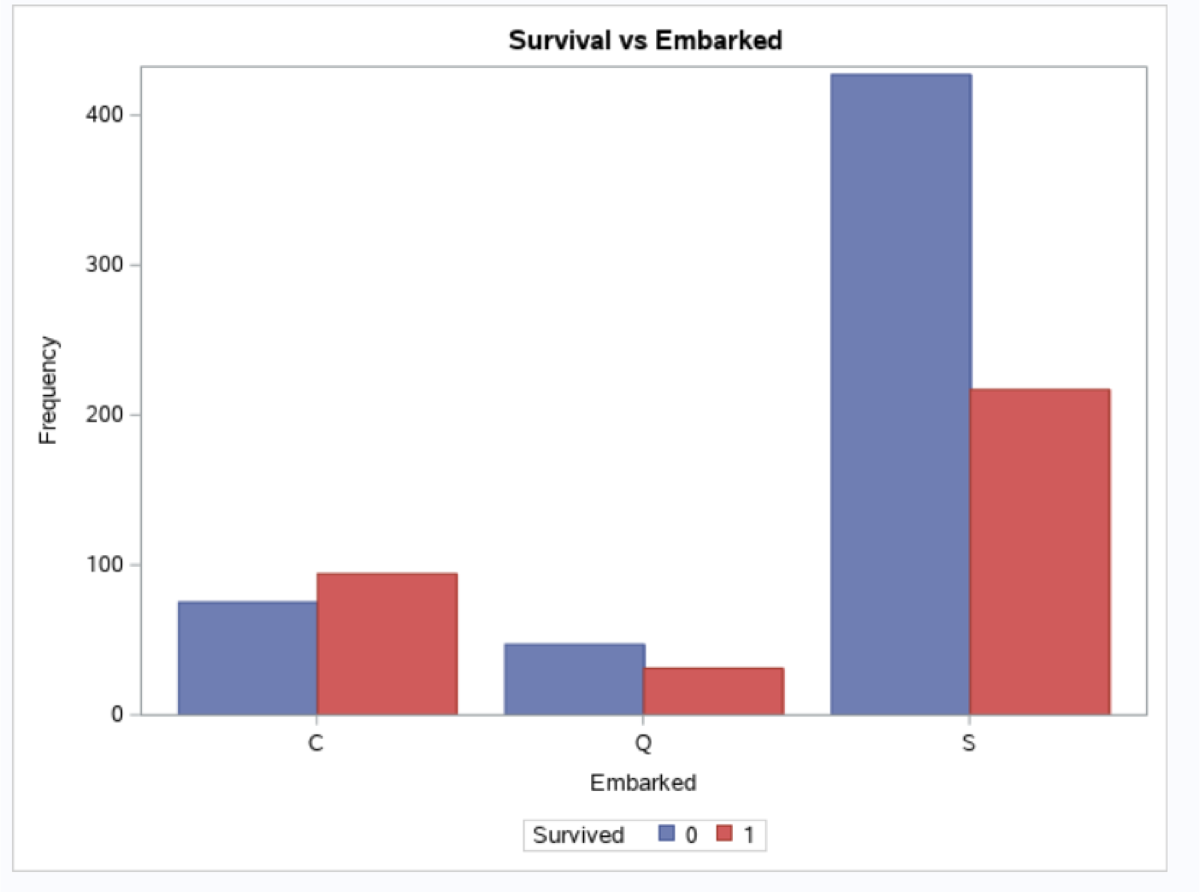
- We can analyze that how embarked has influence on the survival rate.

SAS Code:



```
proc sgplot data = titanic2;
  vbar Embarked / group= Survived groupdisplay = cluster;
 title 'Survival vs Embarked';
run;
```

Output:



5. Conclusion

- We can apply different machine learning models such as random forest, decision tree or logistic regression for predicting the survival rate.
- In this project we had done exploratory analysis of the titanic dataset.
- We have analyzed the correlation between different variables with the survival, in order to see that which variables has impact on survived variable.
- We have taken Age, Sex, Embarked and class variables and checked the relationship between those with survived.
- After the exploratory data analysis, following are the observations:
    - Upper class women did indeed have the highest probability of surviving, followed by middle class women and then lower-class women.
    - Among men, upper class men had greatest probability of survival, which was not much below that of lower-class women.
    - Additionally, the coefficient of age was negative, but was a very small value.
    - However, a person was lucky to survive and had the best chance of survival if they were a young, upper class woman.

Work distribution:

Udbhav Singh Chauhan:

- Data collection,
- Data Cleaning and Preprocessing,
- Statistical Analysis

Pruthviben Jamin Patel:

- Individual variable analysis,
- Correlation analysis,
- Data Visualization

References:

1. Titanic - Machine Learning from Disaster https://www.kaggle.com/c/titanic/data Retrieved on: 20th March 2021.
2. Cody, Ron. 2018. Learning SAS® by Example: A Programmer's Guide, Second Edition. Cary, NC: SAS Institute Inc.
3. SAS Help Center: How to use Data and PROC in SAS: https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/basess/n053a58fwk57v7n14h8x7y7u34y4.htm
4. Analyzing Titanic disaster using machine learning algorithms-Computing, Communication and Automation (ICCCA), 2017 International Conference on 21 December 2017, IEEE.
5. MICHAEL AARON WHITLEY, Using statistical learning to predict survival of passengers on the RMS Titanic by Michael Aaron Whitley, 2015.
6. Kunal Vyas, Zeshi Zheng, Lin Li, Titanic- Machine Learning From Disaster- 2015.