

Sample output of the Data

Saturday, March 28, 2020 02:51:23 PM 1

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		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		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		S

The CONTENTS Procedure

Data Set Name	WORK.FULL	Observations	1309
Member Type	DATA	Variables	12
Engine	V9	Indexes	0
Created	03/28/2020 14:51:24	Observation Length	144
Last Modified	03/28/2020 14:51:24	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	131072
Number of Data Set Pages	2
First Data Page	1
Max Obs per Page	909
Obs in First Data Page	885
Number of Data Set Repairs	0
Filename	/saswork/SAS_workD0B800009B10_odaws02-usw2.oda.sas.com/SAS_work950000009B10_odaws02-usw2.oda.sas.com/full.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	880583
Access Permission	rw-r--r--
Owner Name	sp16670
File Size	384KB
File Size (bytes)	393216

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.

The CONTENTS Procedure

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
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.

Contingency Table of Male and Female and their Titles

The FREQ Procedure

Frequency

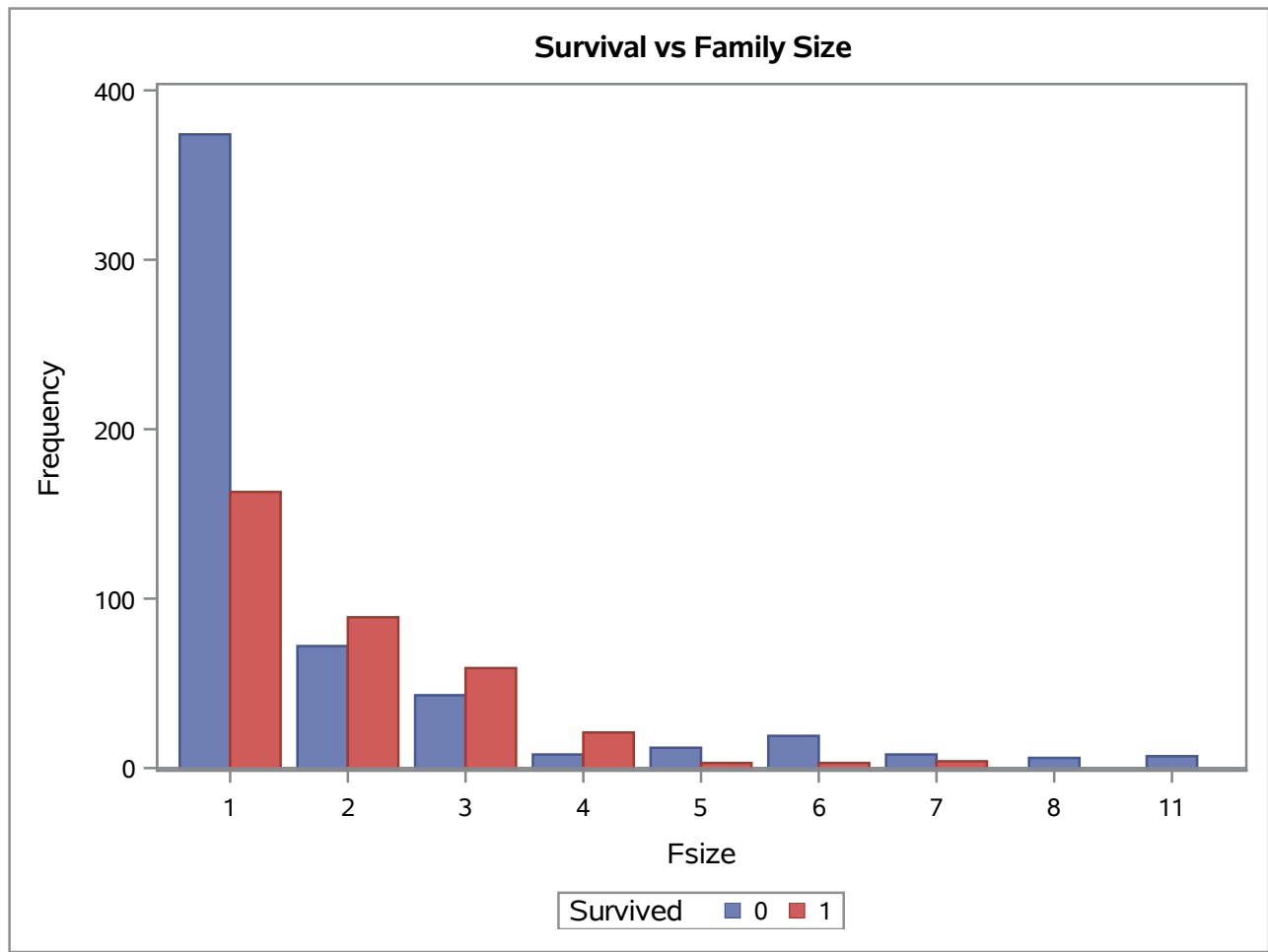
Table of Sex by Title_col											
Sex	Title_col										
	Capt.	Col.	Don.	Dona.	Dr.	Jonkheer.	Lady.	Major.	Master.	Miss.	Mlle.
female	0	0	0	1	1	0	1	0	0	260	2
male	1	4	1	0	7	1	0	2	61	0	0
Total	1	4	1	1	8	1	1	2	61	260	2

Table of Sex by Title_col								
Sex	Title_col							
	Mme.	Mr.	Mrs.	Ms.	Rev.	Sir.	the	Total
female	1	0	197	2	0	0	1	466
male	0	757	0	0	8	1	0	843
Total	1	757	197	2	8	1	1	1309

Contingency Table of Male and Female and their Titles - 2 after alteration

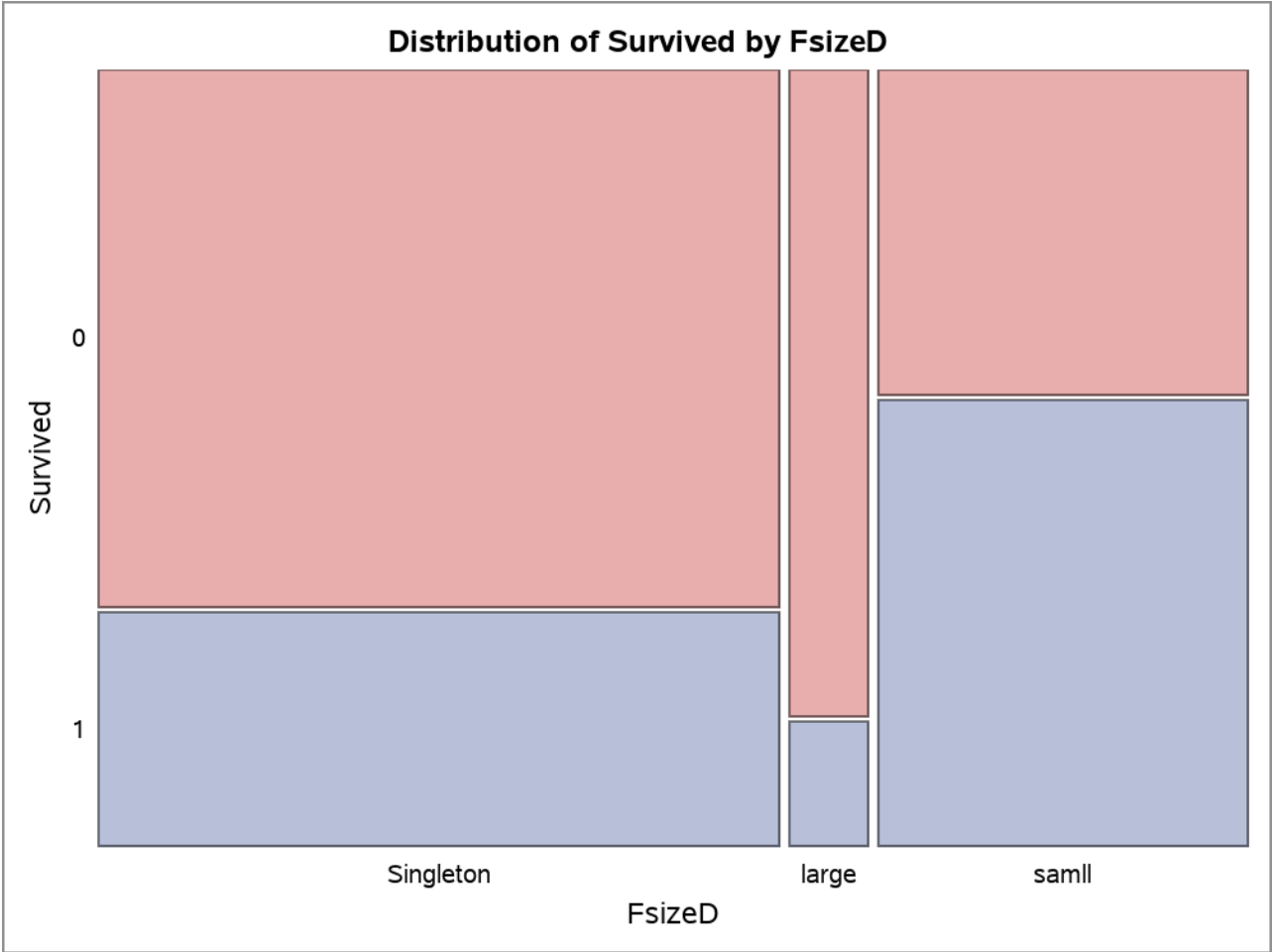
The FREQ Procedure

Frequency	Table of Sex by Title_col								
	Sex	Title_col							
		Master	Miss	Mr	Mrs	Rare	the	Total	
		female	0	264	0	198	3	1	466
		male	61	0	757	0	25	0	843
	Total	61	264	757	198	28	1	1309	

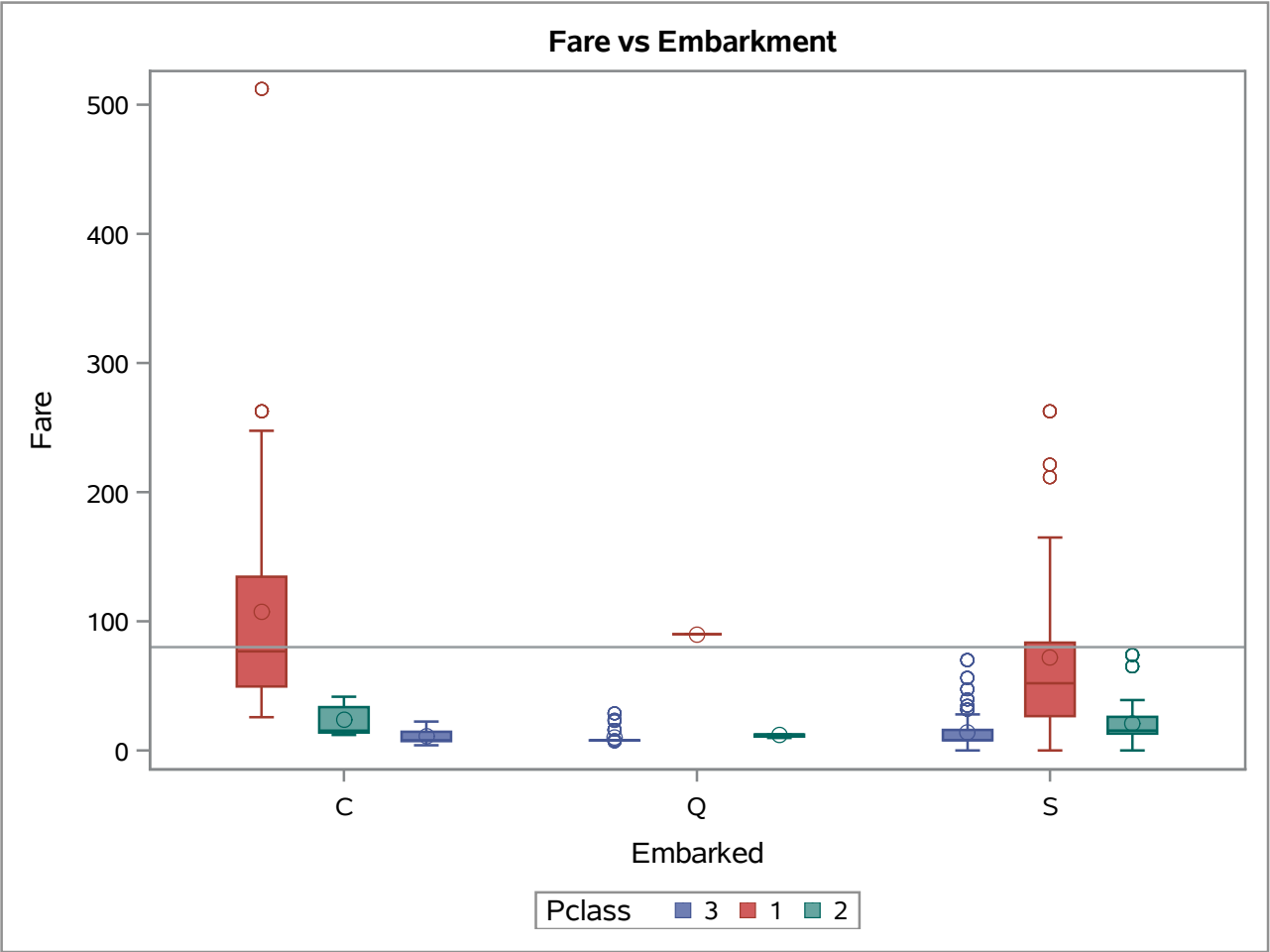


The FREQ Procedure

Percent Col Pct	Table of Survived by FsizeD				
	Survived	FsizeD			
		Singleton	large	samll	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	
Frequency Missing = 418					



Fare	Pclass
80	1
80	1



The FREQ Procedure

Name	Frequency
Not Missing	1309

Sex	Frequency
Not Missing	1309

Ticket	Frequency
Not Missing	1309

Cabin	Frequency
Missing	1014
Not Missing	295

Embarked	Frequency
Not Missing	1309

Title_col	Frequency
Not Missing	1309

FsizeD	Frequency
Not Missing	1309

Deck	Frequency
Missing	1014
Not Missing	295

PassengerId	Frequency
Not Missing	1309

Survived	Frequency
Missing	418
Not Missing	891

Pclass	Frequency
Not Missing	1309

The FREQ Procedure

Age	Frequency
Missing	263
Not Missing	1046

SibSp	Frequency
Not Missing	1309

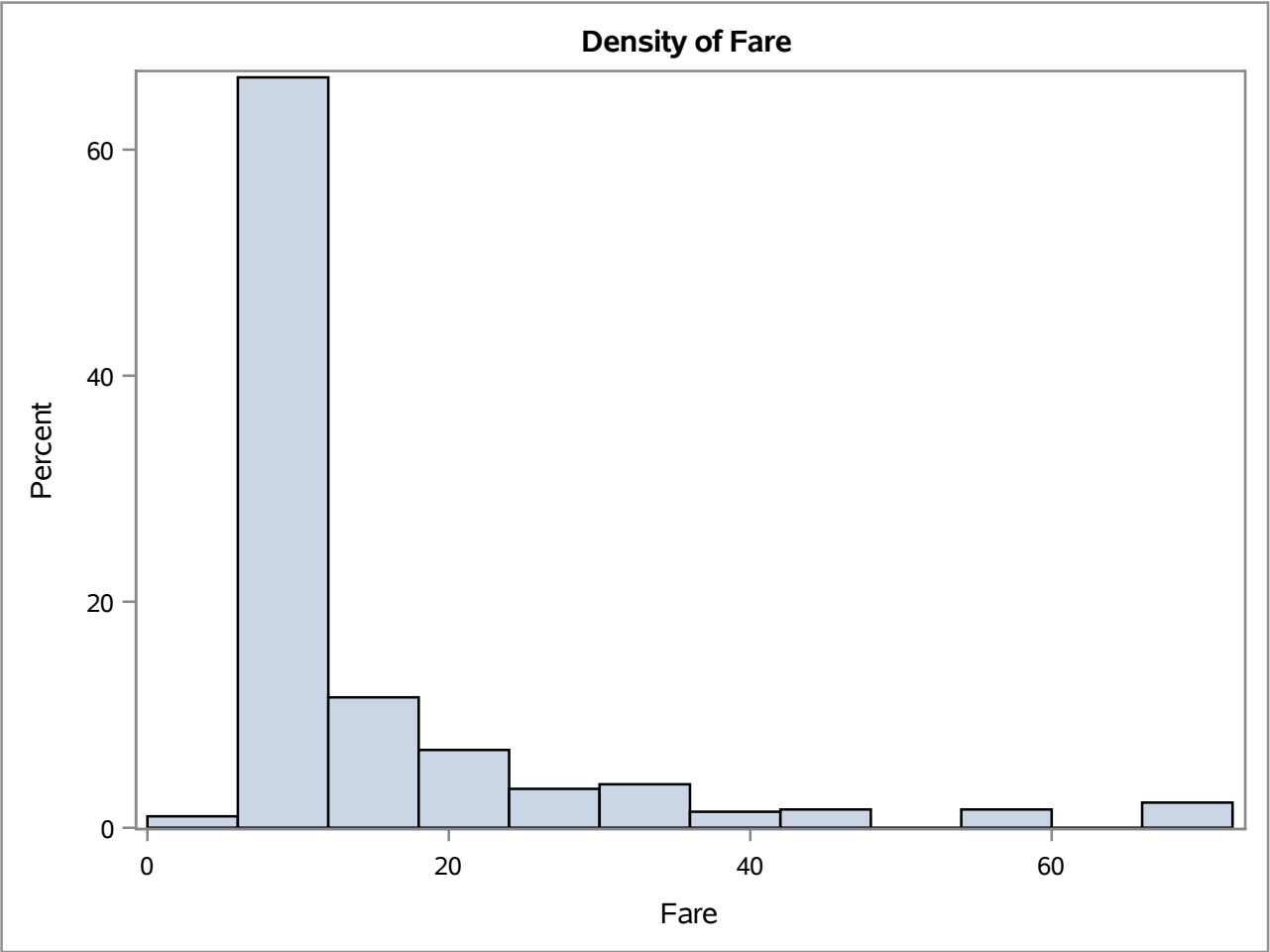
Parch	Frequency
Not Missing	1309

Fare	Frequency
Missing	1
Not Missing	1308

Fsize	Frequency
Not Missing	1309

PassengerId	Pclass	Name	Sex	SibSp	Parch	Ticket	Fare	Embarked	Title_col	Fsize	FsizeD	numMissing
1044	3	Storey, Mr. Thomas	male	0	0	3701	.	S	Mr	1	Singleton	1

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Title_col	Fsize	FsizeD	Deck
1044	.	3	Storey, Mr. Thomas	male	60.5	0	0	3701	.		S	Mr	1	Singleton	



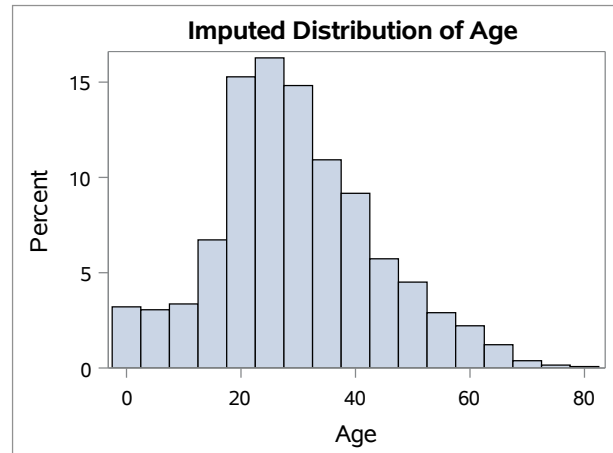
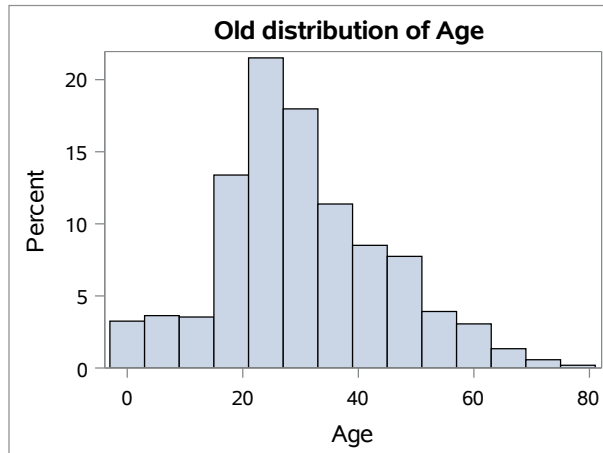
PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Title_col	Fsize	FsizeD	Deck
1044	.	3	Storey, Mr. Thomas	male	60.5	0	0	3701	8.05		S	Mr	1	Singleton	

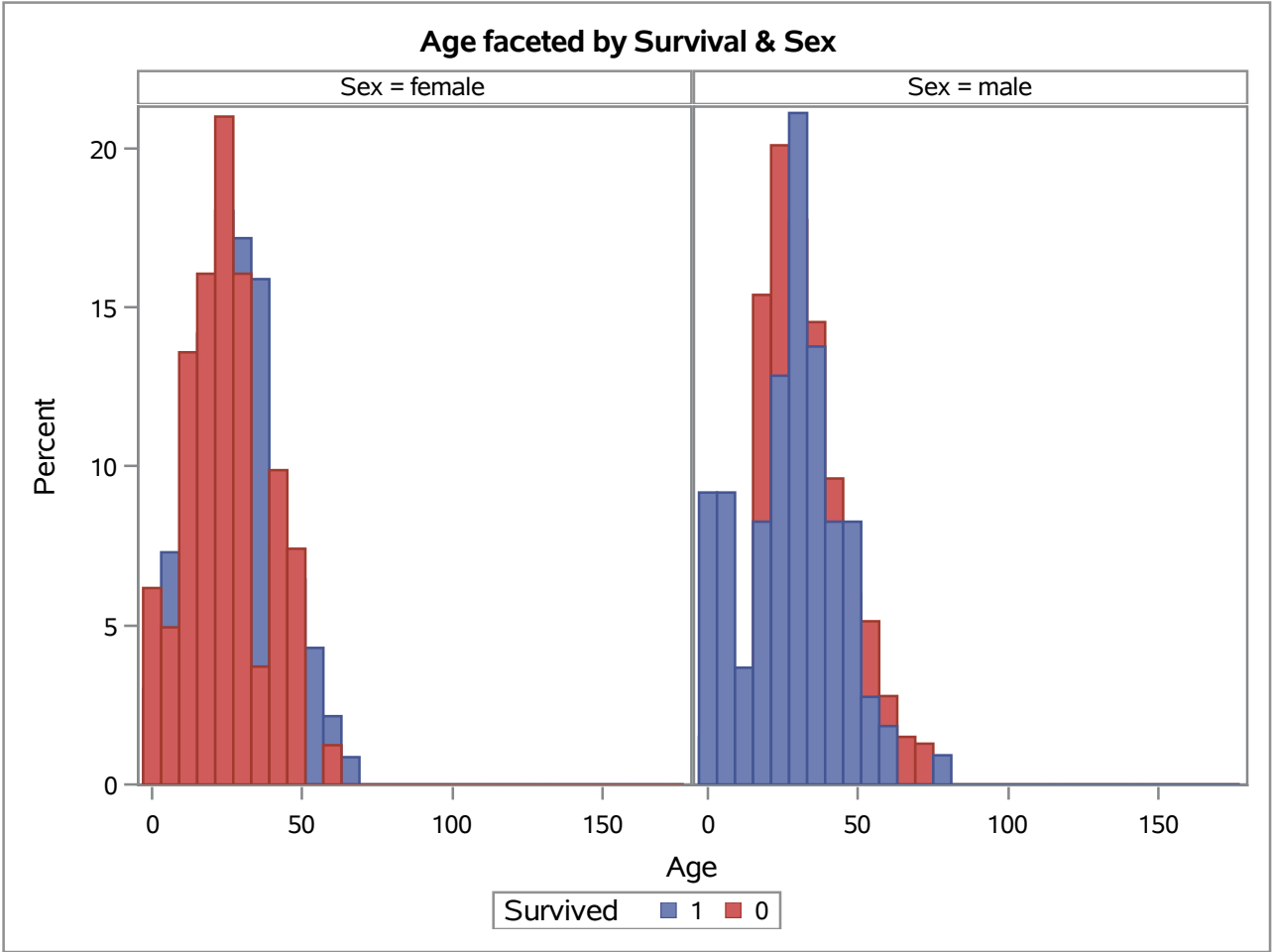
The MI Procedure

Model Information	
Data Set	WORK.FULL
Method	Monotone
Number of Imputations	1
Seed for random number generator	54321

Monotone Model Specification	
Method	Imputed Variables
Regression	Fsize Parch Age
Discriminant Function	Embarked FsizeD Title_col Sex

Missing Data Patterns														
Group	Pclass	Fsize	Parch	Embarked	FsizeD	Title_col	Sex	Age	Freq	Percent	Group Means			
											Pclass	Fsize	Parch	Age
1	X	X	X	X	X	X	X	X	1046	79.91	2.207457	1.923518	0.420650	29.881138
2	X	X	X	X	X	X	X	.	263	20.09	2.642586	1.726236	0.243346	.





The FREQ Procedure

Table of Child by Survived			
Child	Survived		
	0	1	Total
Adult	470	268	738
Child	79	74	153
Total	549	342	891

The FREQ Procedure

Frequency	Table of Mother by Survived			
	Mother	Survived		
		0	1	Total
		Mother	16	39
	Not Mother	533	303	836
	Total	549	342	891

The FREQ Procedure

Name	Frequency
Not Missing	1309

Sex	Frequency
Not Missing	1309

Ticket	Frequency
Not Missing	1309

Cabin	Frequency
Missing	1014
Not Missing	295

Embarked	Frequency
Not Missing	1309

Title_col	Frequency
Not Missing	1309

FsizeD	Frequency
Not Missing	1309

Deck	Frequency
Missing	1014
Not Missing	295

Child	Frequency
Not Missing	1309

Mother	Frequency
Not Missing	1309

PassengerId	Frequency
Not Missing	1309

The FREQ Procedure

Survived	Frequency
Missing	418
Not Missing	891

Pclass	Frequency
Not Missing	1309

Age	Frequency
Not Missing	1309

SibSp	Frequency
Not Missing	1309

Parch	Frequency
Not Missing	1309

Fare	Frequency
Not Missing	1309

Fsize	Frequency
Not Missing	1309

The HPFOREST Procedure

Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	Engine	Role	Path
WORK.TRAIN	V9	Input	On Client

Model Information		
Parameter	Value	
Variables to Try	3	(Default)
Maximum Trees	50	
Actual Trees	50	
Inbag Fraction	0.85	
Prune Fraction	0	(Default)
Prune Threshold	0.1	(Default)
Leaf Fraction	0.00001	(Default)
Leaf Size Setting	1	(Default)
Leaf Size Used	1	
Category Bins	30	(Default)
Interval Bins	100	
Minimum Category Size	5	(Default)
Node Size	100000	(Default)
Maximum Depth	20	(Default)
Alpha	1	(Default)
Exhaustive	5000	(Default)
Rows of Sequence to Skip	5	(Default)
Split Criterion	.	Gini
Preselection Method	.	Loh
Missing Value Handling	.	Valid value

Number of Observations	
Type	N
Number of Observations Read	891
Number of Observations Used	891

The HPFOREST Procedure

Baseline Fit Statistics	
Statistic	Value
Average Square Error	0.237
Misclassification Rate	0.384
Log Loss	0.666

Fit Statistics							
Number of Trees	Number of Leaves	Average Square Error (Train)	Average Square Error (OOB)	Misclassification Rate (Train)	Misclassification Rate (OOB)	Log Loss (Train)	Log Loss (OOB)
1	77	0.0906	0.164	0.1167	0.201	0.514	1.997
2	172	0.0763	0.146	0.0988	0.184	0.253	1.600
3	267	0.0720	0.170	0.0988	0.208	0.245	2.023
4	348	0.0714	0.164	0.0875	0.204	0.244	1.770
5	417	0.0727	0.161	0.0898	0.210	0.250	1.625
6	505	0.0710	0.163	0.0887	0.212	0.243	1.693
7	577	0.0721	0.162	0.0875	0.211	0.250	1.601
8	654	0.0716	0.156	0.0842	0.206	0.250	1.353
9	727	0.0724	0.154	0.0819	0.211	0.253	1.237
10	816	0.0715	0.151	0.0853	0.209	0.249	1.137
11	887	0.0720	0.154	0.0853	0.213	0.250	1.173
12	971	0.0720	0.148	0.0786	0.203	0.251	1.056
13	1045	0.0724	0.145	0.0831	0.201	0.251	1.037
14	1122	0.0731	0.144	0.0808	0.198	0.255	1.000
15	1206	0.0729	0.143	0.0774	0.194	0.254	0.966
16	1295	0.0732	0.143	0.0786	0.193	0.256	0.963
17	1377	0.0728	0.141	0.0786	0.190	0.255	0.924
18	1466	0.0726	0.142	0.0786	0.191	0.254	0.925
19	1561	0.0720	0.142	0.0786	0.192	0.252	0.947
20	1652	0.0719	0.141	0.0797	0.191	0.252	0.891
21	1737	0.0718	0.141	0.0786	0.194	0.252	0.887
22	1827	0.0714	0.141	0.0763	0.197	0.251	0.861
23	1889	0.0719	0.139	0.0786	0.195	0.252	0.832
24	1968	0.0721	0.139	0.0774	0.198	0.252	0.807
25	2057	0.0720	0.137	0.0774	0.196	0.252	0.775
26	2144	0.0720	0.136	0.0752	0.192	0.252	0.746
27	2215	0.0722	0.135	0.0786	0.192	0.253	0.678
28	2301	0.0720	0.135	0.0786	0.193	0.252	0.682

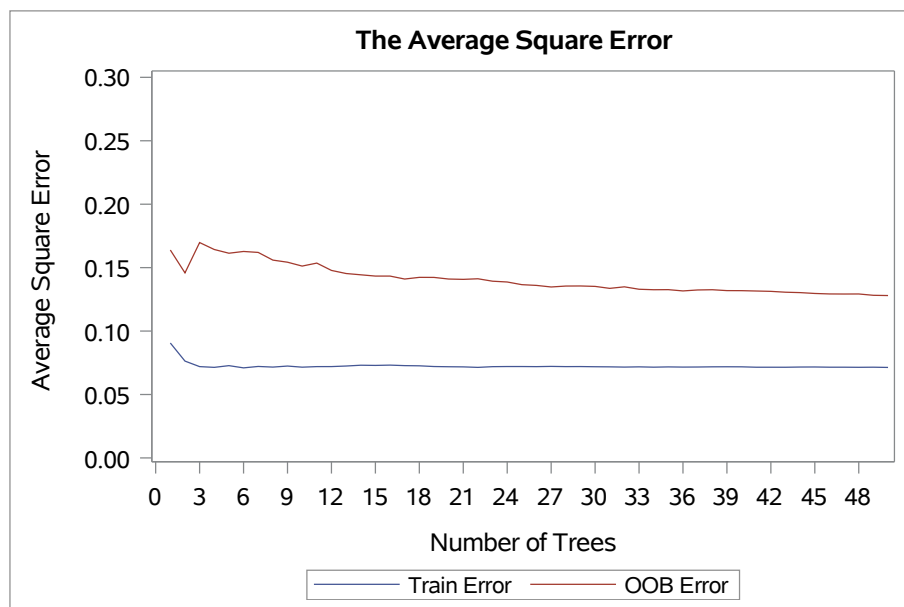
The HPFOREST Procedure

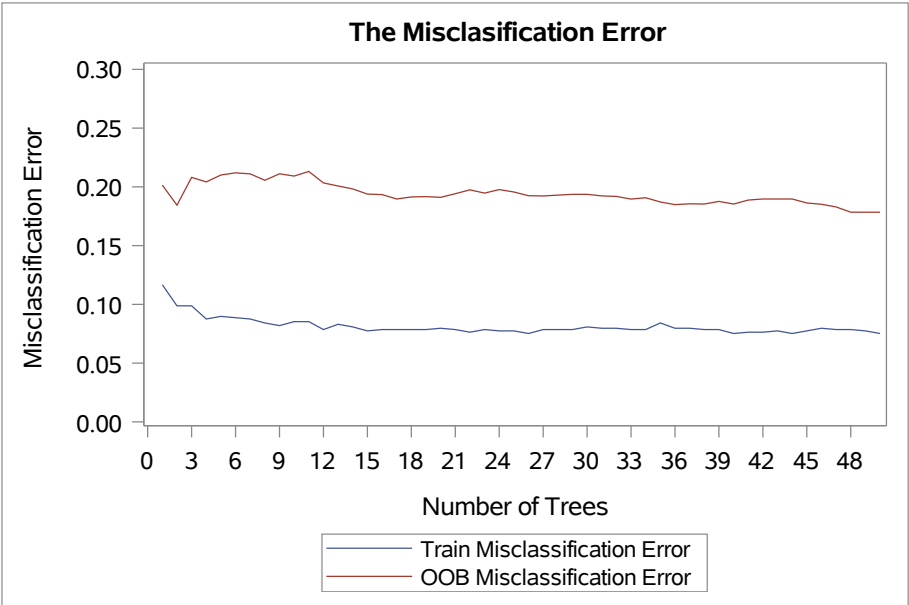
Fit Statistics							
Number of Trees	Number of Leaves	Average Square Error (Train)	Average Square Error (OOB)	Misclassification Rate (Train)	Misclassification Rate (OOB)	Log Loss (Train)	Log Loss (OOB)
29	2386	0.0720	0.136	0.0786	0.194	0.253	0.683
30	2473	0.0719	0.135	0.0808	0.194	0.252	0.681
31	2565	0.0718	0.134	0.0797	0.192	0.252	0.631
32	2658	0.0716	0.135	0.0797	0.192	0.251	0.658
33	2737	0.0718	0.133	0.0786	0.190	0.252	0.607
34	2824	0.0715	0.133	0.0786	0.191	0.251	0.607
35	2885	0.0718	0.133	0.0842	0.187	0.251	0.607
36	2969	0.0716	0.132	0.0797	0.185	0.251	0.582
37	3054	0.0717	0.132	0.0797	0.186	0.251	0.607
38	3138	0.0718	0.133	0.0786	0.185	0.252	0.608
39	3215	0.0719	0.132	0.0786	0.188	0.252	0.606
40	3296	0.0718	0.132	0.0752	0.185	0.252	0.584
41	3380	0.0715	0.132	0.0763	0.189	0.251	0.561
42	3455	0.0715	0.131	0.0763	0.190	0.251	0.560
43	3535	0.0715	0.131	0.0774	0.190	0.250	0.535
44	3604	0.0716	0.130	0.0752	0.190	0.251	0.535
45	3659	0.0717	0.130	0.0774	0.186	0.251	0.512
46	3755	0.0715	0.129	0.0797	0.185	0.250	0.512
47	3830	0.0715	0.129	0.0786	0.183	0.250	0.490
48	3917	0.0714	0.129	0.0786	0.178	0.250	0.491
49	4005	0.0715	0.128	0.0774	0.178	0.250	0.464
50	4093	0.0713	0.128	0.0752	0.178	0.250	0.462

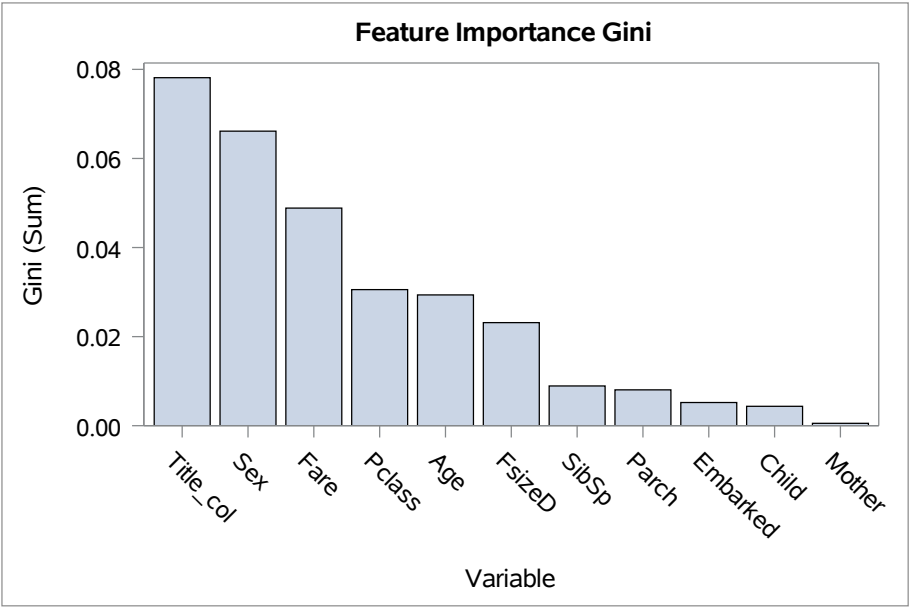
Loss Reduction Variable Importance					
Variable	Number of Rules	Gini	OOB Gini	Margin	OOB Margin
Title_col	204	0.078131	0.07140	0.156261	0.15047
Sex	75	0.066131	0.06638	0.132261	0.13230
Pclass	327	0.030554	0.03107	0.061108	0.06054
FsizeD	200	0.023142	0.02048	0.046284	0.04253
Child	146	0.004356	0.00080	0.008712	0.00426
Embarked	242	0.005208	-0.00047	0.010416	0.00397
Mother	71	0.000534	-0.00208	0.001068	-0.00138
Parch	392	0.008051	-0.00228	0.016102	0.00693

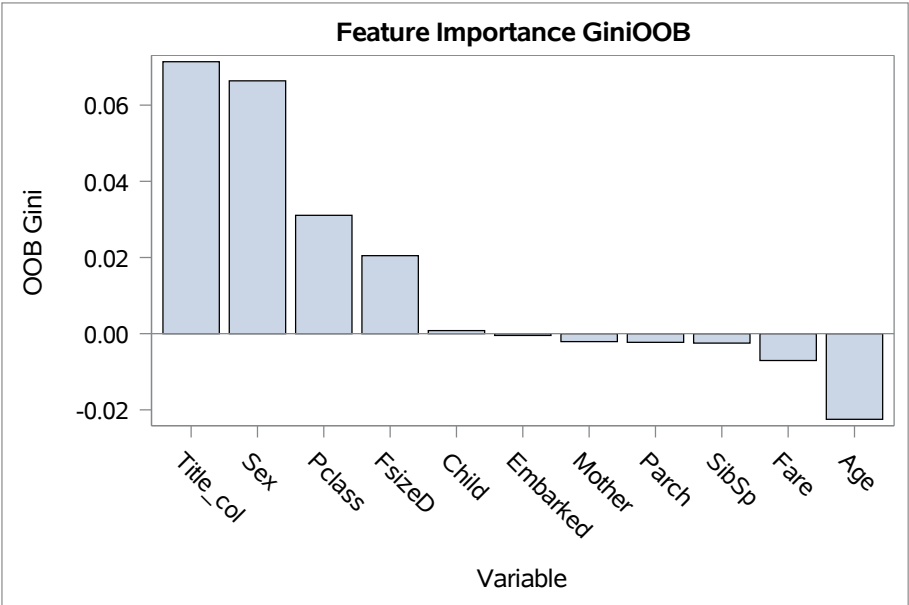
The HPFOREST Procedure

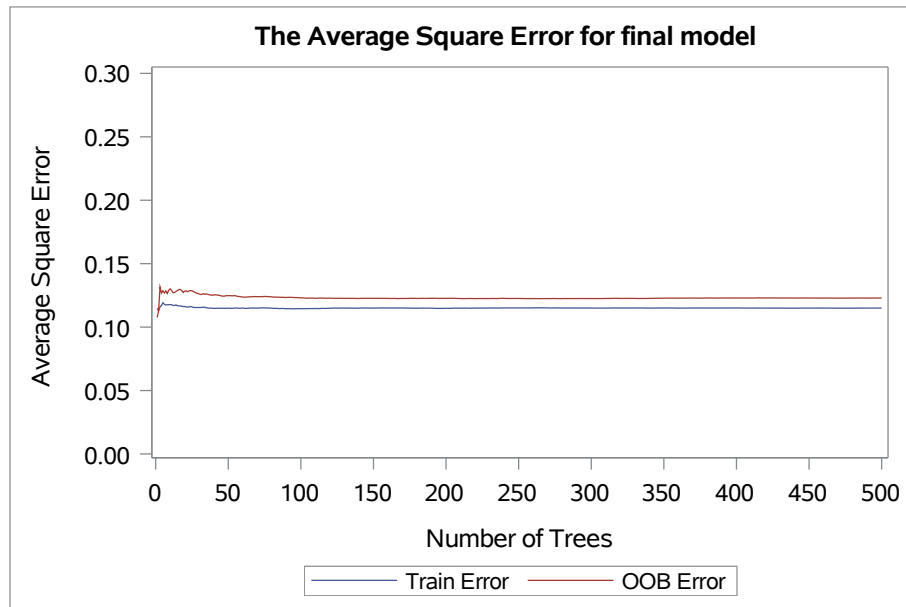
Loss Reduction Variable Importance					
Variable	Number of Rules	Gini	OOB Gini	Margin	OOB Margin
SibSp	363	0.008924	-0.00248	0.017848	0.00569
Fare	1096	0.048859	-0.00705	0.097718	0.03916
Age	927	0.029365	-0.02248	0.058729	0.00740

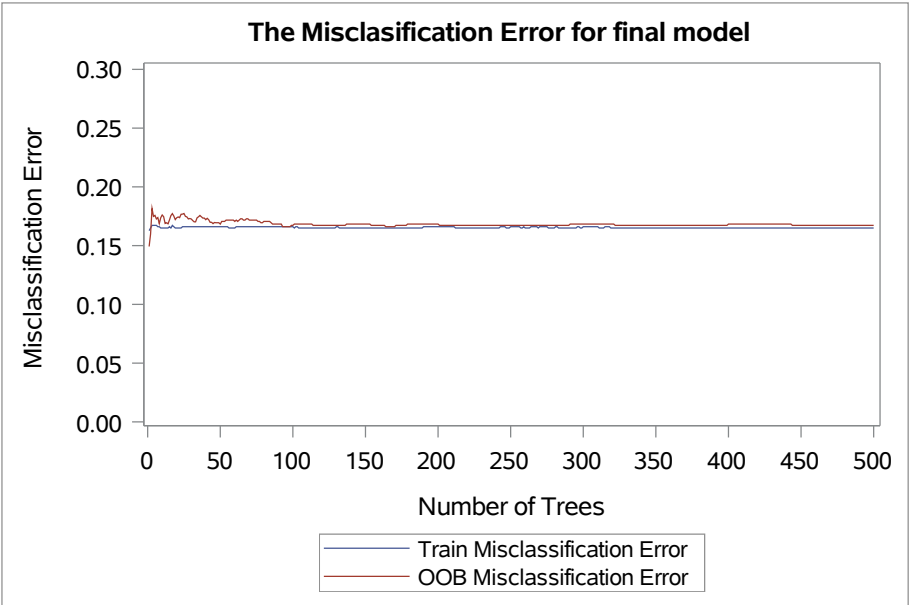












The Misclassification Error for final model

The HP4SCORE Procedure

Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	Engine	Role	Path
WORK.TEST	V9	Input	On Client
WORK.PREDICTIONS	V9	Output	On Client

Number of Observations	
Type	N
Number of Observations Read	418
Number of Observations Used	418
Sum of Frequencies Used	418