

```
In [3]: import numpy as np
import pandas as pd
import scipy
import sklearn
```

```
In [4]: # Reading the "modified" .CSV file except row 1

dataset = pd.read_csv('D:/Seagate_Sync/VOL/VOL00/VIVEK/Big Data Analytics Certification/BDA 104/Week_2/Titanic_modified.csv', dtype = 'unicode')
```

```
In [7]: # Drop the columns where all elements are missing values:  
dataset.dropna(how='all')
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877
6	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463
7	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549
11	12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
12	13	0	3	Saunderscock, Mr. William Henry	male	20	0	0	A/5. 2151
13	14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706
16	17	0	3	Rice, Master. Eugene	male	2	4	1	382652
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31	1	0	345763
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649
20	21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865
21	22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698
22	23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923
23	24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788
24	25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
25	26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	female	38	1	5	347077
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631
27	28	0	1	Fortune, Mr. Charles Alexander	male	19	3	2	19950
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959
29	30	0	3	Todoroff, Mr. Lalio	male	NaN	0	0	349216
...
861	862	0	2	Giles, Mr. Frederick Edward	male	21	1	0	28134
862	863	1	1	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48	0	0	17466
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343
864	865	0	2	Gill, Mr. John William	male	24	0	0	233866
865	866	1	2	Bystrom, Mrs. (Karolina)	female	42	0	0	236852
866	867	1	2	Duran y More, Miss. Asuncion	female	27	1	0	SC/PARIS 2149
867	868	0	1	Roebeling, Mr. Washington Augustus II	male	31	0	0	PC 17590

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777
869	870	1	3	Johnson, Master. Harold Theodor	male	4	1	1	347742
870	871	0	3	Balkic, Mr. Cerin	male	26	0	0	349248
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47	1	1	11751
872	873	0	1	Carlsson, Mr. Frans Olof	male	33	0	0	695
873	874	0	3	Vander Cruyssen, Mr. Victor	male	47	0	0	345765
874	875	1	2	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28	1	0	P/PP 3381
875	876	1	3	Najib, Miss. Adele Kiamie "Jane"	female	15	0	0	2667
876	877	0	3	Gustafsson, Mr. Alfred Ossian	male	20	0	0	7534
877	878	0	3	Petroff, Mr. Nedelio	male	19	0	0	349212
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56	0	1	11767

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25	0	1	230433
881	882	0	3	Markun, Mr. Johann	male	33	0	0	349257
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22	0	0	7552
883	884	0	2	Banfield, Mr. Frederick James	male	28	0	0	C.A./SOTO 34068
884	885	0	3	Sutehall, Mr. Henry Jr	male	25	0	0	SOTON/OC 392076
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39	0	5	382652
886	887	0	2	Montvila, Rev. Juozas	male	27	0	0	211536
887	888	1	1	Graham, Miss. Margaret Edith	female	19	0	0	112053
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607
889	890	1	1	Behr, Mr. Karl Howell	male	26	0	0	111369
890	891	0	3	Dooley, Mr. Patrick	male	32	0	0	370376

891 rows × 12 columns



```
In [9]: # Drop the columns where any of the elements are missing values  
dataset.dropna(how='any')
```


Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	7
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	5
6	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	5
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	1
11	12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	2
21	22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698	1
23	24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788	3
27	28	0	1	Fortune, Mr. Charles Alexander	male	19	3	2	19950	2
52	53	1	1	Harper, Mrs. Henry Sleeper (Myna Haxtun)	female	49	1	0	PC 17572	7
54	55	0	1	Ostby, Mr. Engelhart Cornelius	male	65	0	1	113509	6
62	63	0	1	Harris, Mr. Henry Birkhardt	male	45	1	0	36973	8
66	67	1	2	Nye, Mrs. (Elizabeth Ramell)	female	29	0	0	C.A. 29395	1

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
75	76	0	3	Moen, Mr. Sigurd Hansen	male	25	0	0	348123	7
88	89	1	1	Fortune, Miss. Mabel Helen	female	23	3	2	19950	2
92	93	0	1	Chaffee, Mr. Herbert Fuller	male	46	1	0	W.E.P. 5734	6
96	97	0	1	Goldschmidt, Mr. George B	male	71	0	0	PC 17754	3
97	98	1	1	Greenfield, Mr. William Bertram	male	23	0	1	PC 17759	6
102	103	0	1	White, Mr. Richard Frasar	male	21	0	1	35281	7
110	111	0	1	Porter, Mr. Walter Chamberlain	male	47	0	0	110465	5
118	119	0	1	Baxter, Mr. Quigg Edmond	male	24	0	1	PC 17558	2
124	125	0	1	White, Mr. Percival Wayland	male	54	0	1	35281	7
136	137	1	1	Newsom, Miss. Helen Monypeny	female	19	0	2	11752	2
137	138	0	1	Futrelle, Mr. Jacques Heath	male	37	1	0	113803	5
139	140	0	1	Giglio, Mr. Victor	male	24	0	0	PC 17593	7
151	152	1	1	Pears, Mrs. Thomas (Edith Wearne)	female	22	1	0	113776	6
170	171	0	1	Van der hoef, Mr. Wyckoff	male	61	0	0	111240	3

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
174	175	0	1	Smith, Mr. James Clinch	male	56	0	0	17764	3
177	178	0	1	Isham, Miss. Ann Elizabeth	female	50	0	0	PC 17595	2
183	184	1	2	Becker, Master. Richard F	male	1	2	1	230136	3
193	194	1	2	Navratil, Master. Michel M	male	3	1	1	230080	2
...
737	738	1	1	Lesurer, Mr. Gustave J	male	35	0	0	PC 17755	5
741	742	0	1	Cavendish, Mr. Tyrell William	male	36	1	0	19877	7
742	743	1	1	Ryerson, Miss. Susan Parker "Suzette"	female	21	2	2	PC 17608	2
745	746	0	1	Crosby, Capt. Edward Gifford	male	70	1	1	WE/P 5735	7
748	749	0	1	Marvin, Mr. Daniel Warner	male	19	1	0	113773	5
751	752	1	3	Moor, Master. Meier	male	6	0	1	392096	1
759	760	1	1	Roths, the Countess. of (Lucy Noel Martha Dye...	female	33	0	0	110152	8
763	764	1	1	Carter, Mrs. William Ernest (Lucile Polk)	female	36	1	2	113760	1
765	766	1	1	Hogeboom, Mrs. John C (Anna Andrews)	female	51	1	0	13502	7

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
772	773	0	2	Mack, Mrs. (Mary)	female	57	0	0	S.O./P.P. 3	1
779	780	1	1	Robert, Mrs. Edward Scott (Elisabeth Walton Mc...	female	43	0	1	24160	2
781	782	1	1	Dick, Mrs. Albert Adrian (Vera Gillespie)	female	17	1	0	17474	5
782	783	0	1	Long, Mr. Milton Clyde	male	29	0	0	113501	3
789	790	0	1	Guggenheim, Mr. Benjamin	male	46	0	0	PC 17593	7
796	797	1	1	Leader, Dr. Alice (Farnham)	female	49	0	0	17465	2
802	803	1	1	Carter, Master. William Thornton II	male	11	1	2	113760	1
806	807	0	1	Andrews, Mr. Thomas Jr	male	39	0	0	112050	0
809	810	1	1	Chambers, Mrs. Norman Campbell (Bertha Griggs)	female	33	1	0	113806	5
820	821	1	1	Hays, Mrs. Charles Melville (Clara Jennings Gr...	female	52	1	1	12749	9
823	824	1	3	Moor, Mrs. (Beila)	female	27	0	1	392096	1
835	836	1	1	Compton, Miss. Sara Rebecca	female	39	1	1	PC 17756	8

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
853	854	1	1	Lines, Miss. Mary Conover	female	16	0	1	PC 17592	3
857	858	1	1	Daly, Mr. Peter Denis	male	51	0	0	113055	2
862	863	1	1	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48	0	0	17466	2
867	868	0	1	Roebbling, Mr. Washington Augustus II	male	31	0	0	PC 17590	5
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47	1	1	11751	5
872	873	0	1	Carlsson, Mr. Frans Olof	male	33	0	0	695	5
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56	0	1	11767	8
887	888	1	1	Graham, Miss. Margaret Edith	female	19	0	0	112053	3
889	890	1	1	Behr, Mr. Karl Howell	male	26	0	0	111369	3

179 rows × 12 columns



```
In [10]: # Keep only the rows which contain 2 missing values maximum  
dataset.dropna(thresh=2)
```

Out[10]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877
6	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463
7	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549
11	12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783

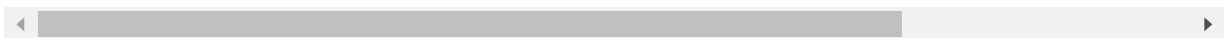
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
12	13	0	3	Saunderscock, Mr. William Henry	male	20	0	0	A/5. 2151
13	14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706
16	17	0	3	Rice, Master. Eugene	male	2	4	1	382652
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31	1	0	345763
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649
20	21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865
21	22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698
22	23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923
23	24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788
24	25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
25	26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	female	38	1	5	347077
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631
27	28	0	1	Fortune, Mr. Charles Alexander	male	19	3	2	19950
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959
29	30	0	3	Todoroff, Mr. Lalio	male	NaN	0	0	349216
...
861	862	0	2	Giles, Mr. Frederick Edward	male	21	1	0	28134
862	863	1	1	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48	0	0	17466
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343
864	865	0	2	Gill, Mr. John William	male	24	0	0	233866
865	866	1	2	Bystrom, Mrs. (Karolina)	female	42	0	0	236852
866	867	1	2	Duran y More, Miss. Asuncion	female	27	1	0	SC/PARIS 2149
867	868	0	1	Roebeling, Mr. Washington Augustus II	male	31	0	0	PC 17590

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777
869	870	1	3	Johnson, Master. Harold Theodor	male	4	1	1	347742
870	871	0	3	Balkic, Mr. Cerin	male	26	0	0	349248
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47	1	1	11751
872	873	0	1	Carlsson, Mr. Frans Olof	male	33	0	0	695
873	874	0	3	Vander Cruyssen, Mr. Victor	male	47	0	0	345765
874	875	1	2	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28	1	0	P/PP 3381
875	876	1	3	Najib, Miss. Adele Kiamie "Jane"	female	15	0	0	2667
876	877	0	3	Gustafsson, Mr. Alfred Ossian	male	20	0	0	7534
877	878	0	3	Petroff, Mr. Nedelio	male	19	0	0	349212
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56	0	1	11767

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25	0	1	230433
881	882	0	3	Markun, Mr. Johann	male	33	0	0	349257
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22	0	0	7552
883	884	0	2	Banfield, Mr. Frederick James	male	28	0	0	C.A./SOTO 34068
884	885	0	3	Sutehall, Mr. Henry Jr	male	25	0	0	SOTON/OC 392076
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39	0	5	382652
886	887	0	2	Montvila, Rev. Juozas	male	27	0	0	211536
887	888	1	1	Graham, Miss. Margaret Edith	female	19	0	0	112053
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607
889	890	1	1	Behr, Mr. Karl Howell	male	26	0	0	111369
890	891	0	3	Dooley, Mr. Patrick	male	32	0	0	370376

891 rows × 12 columns



```
In [11]: # Fill all missing values with the mean of the particular column  
dataset.fillna(dataset.mean())
```

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877
6	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463
7	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549
11	12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783

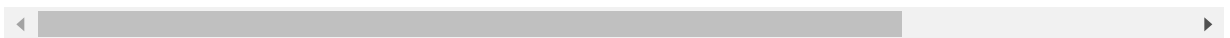
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
12	13	0	3	Saunderscock, Mr. William Henry	male	20	0	0	A/5. 2151
13	14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706
16	17	0	3	Rice, Master. Eugene	male	2	4	1	382652
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31	1	0	345763
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649
20	21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865
21	22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698
22	23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923
23	24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788
24	25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
25	26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	female	38	1	5	347077
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631
27	28	0	1	Fortune, Mr. Charles Alexander	male	19	3	2	19950
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959
29	30	0	3	Todoroff, Mr. Lalio	male	NaN	0	0	349216
...
861	862	0	2	Giles, Mr. Frederick Edward	male	21	1	0	28134
862	863	1	1	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48	0	0	17466
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343
864	865	0	2	Gill, Mr. John William	male	24	0	0	233866
865	866	1	2	Bystrom, Mrs. (Karolina)	female	42	0	0	236852
866	867	1	2	Duran y More, Miss. Asuncion	female	27	1	0	SC/PARIS 2149
867	868	0	1	Roebeling, Mr. Washington Augustus II	male	31	0	0	PC 17590

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777
869	870	1	3	Johnson, Master. Harold Theodor	male	4	1	1	347742
870	871	0	3	Balkic, Mr. Cerin	male	26	0	0	349248
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47	1	1	11751
872	873	0	1	Carlsson, Mr. Frans Olof	male	33	0	0	695
873	874	0	3	Vander Cruyssen, Mr. Victor	male	47	0	0	345765
874	875	1	2	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28	1	0	P/PP 3381
875	876	1	3	Najib, Miss. Adele Kiamie "Jane"	female	15	0	0	2667
876	877	0	3	Gustafsson, Mr. Alfred Ossian	male	20	0	0	7534
877	878	0	3	Petroff, Mr. Nedelio	male	19	0	0	349212
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56	0	1	11767

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25	0	1	230433
881	882	0	3	Markun, Mr. Johann	male	33	0	0	349257
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22	0	0	7552
883	884	0	2	Banfield, Mr. Frederick James	male	28	0	0	C.A./SOTO 34068
884	885	0	3	Sutehall, Mr. Henry Jr	male	25	0	0	SOTON/OC 392076
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39	0	5	382652
886	887	0	2	Montvila, Rev. Juozas	male	27	0	0	211536
887	888	1	1	Graham, Miss. Margaret Edith	female	19	0	0	112053
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607
889	890	1	1	Behr, Mr. Karl Howell	male	26	0	0	111369
890	891	0	3	Dooley, Mr. Patrick	male	32	0	0	370376

891 rows × 12 columns



```
In [13]: # Fill any missing value in column 'A' with the column median  
dataset['Fare'].fillna(dataset['Fare'].median())
```

```
Out[13]: 0      7.25
          1     71.2833
          2      7.925
          3      53.1
          4      8.05
          5      8.4583
          6     51.8625
          7     21.075
          8     11.1333
          9     30.0708
         10      16.7
         11      26.55
         12      8.05
         13     31.275
         14     7.8542
         15      16
         16     29.125
         17      13
         18      18
         19      7.225
         20      26
         21      13
         22     8.0292
         23     35.5
         24     21.075
         25     31.3875
         26      7.225
         27      263
         28      7.8792
         29      7.8958
          ...
        861     11.5
        862    25.9292
        863     69.55
        864      13
        865      13
        866    13.8583
        867    50.4958
        868      9.5
        869    11.1333
        870      7.8958
        871    52.5542
        872       5
        873       9
        874      24
        875      7.225
        876     9.8458
        877      7.8958
        878      7.8958
        879    83.1583
        880      26
        881      7.8958
        882    10.5167
        883     10.5
        884      7.05
        885     29.125
        886      13
```

```

887      30
888    23.45
889      30
890     7.75
Name: Fare, Length: 891, dtype: object

```

```
In [14]: print( dataset.describe() )
```

```

      PassengerId  Survived  Pclass
\
count           891         891    891
unique           891          2      3
top             403          0      3  Humblen, Mr. Adolf Mathias Nicolai Olsen
freq              1        549    491

```

```

      Sex  Age  SibSp  Parch  Ticket   Fare  Cabin  Embarked
count    891  689    891    891     887    891    204      889
unique     2   70      7      7     680    248    147        3
top    male   24      0      0  347082    8.05     G6        S
freq     577   30   608   678         7     43      4     644

```

```
In [16]: categorical = dataset.dtypes[dataset.dtypes == "object"].index
print(categorical)
dataset[categorical].describe()
```

```

Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
      'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')

```

```
Out[16]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
count	891	891	891	891	891	689	891	891	887	891
unique	891	2	3	891	2	70	7	7	680	248
top	403	0	3	Humblen, Mr. Adolf Mathias Nicolai Olsen	male	24	0	0	347082	8.05
freq	1	549	491	1	577	30	608	678	7	43

```
In [ ]: # Identify categorical variables and explain why they cant be used in the raw form in the analysis
```

Categorical variables **in** their raw form may **not** be always numeric, For e.g. **in** Titanic dataset, categorical variable "Sex" has "Male" **and** "Female" **as** categories. So we need to convert them into numeric variables **in** order to use them **in** analysis. Moreover, some machine learning libraries do **not** take categorical variables **as** input.

```
In [18]: # Transform one of the categorical variables into numerical using label encoder method
```

```
from sklearn.preprocessing import LabelEncoder  
numeric = LabelEncoder()  
dataset['Sex']= numeric.fit_transform(dataset['Sex'].astype('str'))
```

```
In [19]: print(dataset['Sex'])
```

0	1
1	0
2	0
3	0
4	1
5	1
6	1
7	1
8	0
9	0
10	0
11	0
12	1
13	1
14	0
15	0
16	1
17	1
18	0
19	0
20	1
21	1
22	0
23	1
24	0
25	0
26	1
27	1
28	0
29	1
	..
861	1
862	0
863	0
864	1
865	0
866	0
867	1
868	1
869	1
870	1
871	0
872	1
873	1
874	0
875	0
876	1
877	1
878	1
879	0
880	0
881	1
882	0
883	1
884	1
885	0
886	1

887 0

888 0

889 1

890 1

Name: Sex, Length: 891, dtype: int64

```
In [26]: # Transform gender into numerical variable using dummy coding

dataset1 = pd.read_csv('D:/Seagate_Sync/VOL/VOL00/VIVEK/Big Data Analytics Cer
tification/BDA 104/Week_2/Titanic_original.csv', dtype = 'unicode')
dataset1=dataset1=pd.get_dummies(dataset1['Sex'])
```


In [27]: `print(dataset1)`

	female	male
0	0	1
1	1	0
2	1	0
3	1	0
4	0	1
5	0	1
6	0	1
7	0	1
8	1	0
9	1	0
10	1	0
11	1	0
12	0	1
13	0	1
14	1	0
15	1	0
16	0	1
17	0	1
18	1	0
19	1	0
20	0	1
21	0	1
22	1	0
23	0	1
24	1	0
25	1	0
26	0	1
27	0	1
28	1	0
29	0	1
..
861	0	1
862	1	0
863	1	0
864	0	1
865	1	0
866	1	0
867	0	1
868	0	1
869	0	1
870	0	1
871	1	0
872	0	1
873	0	1
874	1	0
875	1	0
876	0	1
877	0	1
878	0	1
879	1	0
880	1	0
881	0	1
882	1	0
883	0	1
884	0	1
885	1	0

886	0	1
887	1	0
888	1	0
889	0	1
890	0	1

[891 rows x 2 columns]