

# CASE STUDY: TITANIC DATA ANALYSIS

- DataFrames
  - loading & basic analysis
  - sorting & subsetting
  - creating new columns

## 1. Data loading & Basic Analysis

### 1. Import Libraries

```
import pandas as pd
import matplotlib.pyplot as plt
```

To Analyze the data first we can read the all data and then we have some methods to analyze the data, Methods are given below:

### 1. Read Titanic Comma Separated file as `titanic_df`

```
titanic_df = pd.read_csv('titanic.csv')
titanic_df.head()
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

  

	SibSp	\	Name	Sex	Age
0			Braund, Mr. Owen Harris	male	22.0
1					
1	1		Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0
1					
2			Heikkinen, Miss. Laina	female	26.0
0					
3			Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0
1					
4			Allen, Mr. William Henry	male	35.0
0					

  

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S

### 1. Basic Analysis with `.info()`, `.describe()` methods.

```
titanic_df.info() # info method tells about, how many columns are null  
& non-null also tells about datatype.
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 891 entries, 0 to 890  
Data columns (total 12 columns):  
#   Column      Non-Null Count  Dtype  
---  -  
0   PassengerId  891 non-null    int64  
1   Survived     891 non-null    int64  
2   Pclass       891 non-null    int64  
3   Name         891 non-null    object  
4   Sex          891 non-null    object  
5   Age          714 non-null    float64  
6   SibSp        891 non-null    int64  
7   Parch        891 non-null    int64  
8   Ticket       891 non-null    object  
9   Fare         891 non-null    float64  
10  Cabin        204 non-null    object  
11  Embarked     889 non-null    object  
dtypes: float64(2), int64(5), object(5)  
memory usage: 66.2+ KB
```

```
titanic_df.describe() # Quick overview for numeric values
```

	PassengerId	Survived	Pclass	Age	SibSp	\
count	891.000000	891.000000	891.000000	714.000000	891.000000	
mean	446.000000	0.383838	2.308642	29.699118	0.523008	
std	257.353842	0.486592	0.836071	14.526497	1.102743	
min	1.000000	0.000000	1.000000	0.420000	0.000000	
25%	223.500000	0.000000	2.000000	20.125000	0.000000	
50%	446.000000	0.000000	3.000000	28.000000	0.000000	
75%	668.500000	1.000000	3.000000	38.000000	1.000000	
max	891.000000	1.000000	3.000000	80.000000	8.000000	

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

#### 1. Basic Analysis with `.shape`, `.columns`, `.values`, `.index`

```
titanic_df.shape # IT gives the numbers of rows and columns  
(891, 12)
```

```
titanic_df.columns # It returns the name of all columns
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age',
      'SibSp',
      'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')

titanic_df.values # It gives the all rows/observations of every person
array([[1, 0, 3, ..., 7.25, nan, 'S'],
       [2, 1, 1, ..., 71.2833, 'C85', 'C'],
       [3, 1, 3, ..., 7.925, nan, 'S'],
       ...,
       [889, 0, 3, ..., 23.45, nan, 'S'],
       [890, 1, 1, ..., 30.0, 'C148', 'C'],
       [891, 0, 3, ..., 7.75, nan, 'Q']], dtype=object)

titanic_df.index # It gives the starting point, end point and steps of
provided data
RangeIndex(start=0, stop=891, step=1)
```

## 2. Sorting & SubSetting

1. Sorting Data Set by Single columns `.sort_values()`
  - `by=""`
  - `ascending=""`
  - `na_position=""`

```
titanic_df.sort_values(by='PassengerId') #It follows the Ascending
order means the smallest or first or earliest in the order will appear
at the top of the list
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
..	...	...	...	
886	887	0	2	
887	888	1	1	
888	889	0	3	
889	890	1	1	
890	891	0	3	

	Name	Sex	Age
SibSp \			
0	Braund, Mr. Owen Harris	male	22.0
1			
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0

```

1
2                               Heikkinen, Miss. Laina  female  26.0
0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0
1
4                               Allen, Mr. William Henry   male  35.0
0
..                               ...      ...      ...
...
886                               Montvila, Rev. Juozas    male  27.0
0
887                               Graham, Miss. Margaret Edith  female  19.0
0
888       Johnston, Miss. Catherine Helen "Carrie"  female   NaN
1
889                               Behr, Mr. Karl Howell     male  26.0
0
890                               Dooley, Mr. Patrick       male  32.0
0

```

```

      Parch      Ticket    Fare Cabin Embarked
0         0          A/5 21171    7.2500   NaN      S
1         0          PC 17599   71.2833   C85      C
2         0  STON/O2. 3101282    7.9250   NaN      S
3         0          113803   53.1000  C123      S
4         0          373450    8.0500   NaN      S
..      ...      ...      ...      ...      ...
886        0          211536   13.0000   NaN      S
887        0          112053   30.0000   B42      S
888        2          W./C. 6607   23.4500   NaN      S
889        0          111369   30.0000  C148      C
890        0          370376    7.7500   NaN      Q

```

[891 rows x 12 columns]

```

# If I want to set it in descending order so
titanic_df.sort_values(by='Age', ascending=False, na_position='last')
# na_position will keep the non fill-up values in the last

```

	PassengerId	Survived	Pclass	
Name \				
630	631	1	1	Barkworth, Mr. Algernon Henry
Wilson				
851	852	0	3	Svensson, Mr.
Johan				
493	494	0	1	Artagaveytia, Mr.
Ramon				
96	97	0	1	Goldschmidt, Mr.
George B				
116	117	0	3	Connors, Mr.

```

Patrick
..      ...      ...      ...
...
859      860      0      3      Razi, Mr.
Raihed
863      864      0      3      Sage, Miss. Dorothy Edith
"Dolly"
868      869      0      3      van Melkebeke, Mr.
Philemon
878      879      0      3      Laleff, Mr.
Kristo
888      889      0      3      Johnston, Miss. Catherine Helen
"Carrie"

```

	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
630	male	80.0	0	0	27042	30.0000	A23	S
851	male	74.0	0	0	347060	7.7750	NaN	S
493	male	71.0	0	0	PC 17609	49.5042	NaN	C
96	male	71.0	0	0	PC 17754	34.6542	A5	C
116	male	70.5	0	0	370369	7.7500	NaN	Q
..	...	...	...	...	...	...	...	...
859	male	NaN	0	0	2629	7.2292	NaN	C
863	female	NaN	8	2	CA. 2343	69.5500	NaN	S
868	male	NaN	0	0	345777	9.5000	NaN	S
878	male	NaN	0	0	349217	7.8958	NaN	S
888	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

[891 rows x 12 columns]

#### 1. Sorting DataSet by Multiple columns `.sort_values()`

```
titanic_df.sort_values(['Survived', 'Sex'])
```

	PassengerId	Survived	Pclass	\
14	15	0	3	
18	19	0	3	
24	25	0	3	
38	39	0	3	
40	41	0	3	
..	...	...	...	
838	839	1	3	
839	840	1	1	
857	858	1	1	
869	870	1	3	
889	890	1	1	

	Name	Sex	Age
SibSp \			
14	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0
0			

18	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0
1			
24	Palsson, Miss. Torborg Danira	female	8.0
3			
38	Vander Planke, Miss. Augusta Maria	female	18.0
2			
40	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	female	40.0
1			
..	...	...	...
...			
838	Chip, Mr. Chang	male	32.0
0			
839	Marechal, Mr. Pierre	male	NaN
0			
857	Daly, Mr. Peter Denis	male	51.0
0			
869	Johnson, Master. Harold Theodor	male	4.0
1			
889	Behr, Mr. Karl Howell	male	26.0
0			

	Parch	Ticket	Fare	Cabin	Embarked
14	0	350406	7.8542	NaN	S
18	0	345763	18.0000	NaN	S
24	1	349909	21.0750	NaN	S
38	0	345764	18.0000	NaN	S
40	0	7546	9.4750	NaN	S
..	...	...	...	...	...
838	0	1601	56.4958	NaN	S
839	0	11774	29.7000	C47	C
857	0	113055	26.5500	E17	S
869	1	347742	11.1333	NaN	S
889	0	111369	30.0000	C148	C

[891 rows x 12 columns]

*# Now If I want to set one in ascending order and second in descending order so*

```
titanic_df.sort_values(['Survived', 'Sex'], ascending=[True, False])
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
4	5	0	3	
5	6	0	3	
6	7	0	1	
7	8	0	3	
..	...	...	...	
874	875	1	2	
875	876	1	3	
879	880	1	1	

```

880      881      1      2
887      888      1      1

```

		Name	Sex	Age
SibSp \				
0		Braund, Mr. Owen Harris	male	22.0
1				
4		Allen, Mr. William Henry	male	35.0
0				
5		Moran, Mr. James	male	NaN
0				
6		McCarthy, Mr. Timothy J	male	54.0
0				
7		Palsson, Master. Gosta Leonard	male	2.0
3				
..		...	...	...
.				
874		Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0
1				
875		Najib, Miss. Adele Kiamie "Jane"	female	15.0
0				
879	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)		female	56.0
0				
880	Shelley, Mrs. William (Imanita Parrish Hall)		female	25.0
0				
887	Graham, Miss. Margaret Edith		female	19.0
0				

  

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
4	0	373450	8.0500	NaN	S
5	0	330877	8.4583	NaN	Q
6	0	17463	51.8625	E46	S
7	1	349909	21.0750	NaN	S
..	...	...	...	...	...
874	0	P/PP 3381	24.0000	NaN	C
875	0	2667	7.2250	NaN	C
879	1	11767	83.1583	C50	C
880	1	230433	26.0000	NaN	S
887	0	112053	30.0000	B42	S

[891 rows x 12 columns]

## 7. Subsetting Single column & Multiple columns

```

# Single column subsetting
titanic_df['Name']

```

```

0      Braund, Mr. Owen Harris
1  Cumings, Mrs. John Bradley (Florence Briggs Th...

```

```

2           Heikkinen, Miss. Laina
3   Futrelle, Mrs. Jacques Heath (Lily May Peel)
4           Allen, Mr. William Henry
...
886           Montvila, Rev. Juozas
887           Graham, Miss. Margaret Edith
888   Johnston, Miss. Catherine Helen "Carrie"
889           Behr, Mr. Karl Howell
890           Dooley, Mr. Patrick
Name: Name, Length: 891, dtype: object

```

```

# Multiple columns subsetting
titanic_df[['Survived', 'Name']]

```

	Survived	Name
0	0	Braund, Mr. Owen Harris
1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...
2	1	Heikkinen, Miss. Laina
3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)
4	0	Allen, Mr. William Henry
..	...	...
886	0	Montvila, Rev. Juozas
887	1	Graham, Miss. Margaret Edith
888	0	Johnston, Miss. Catherine Helen "Carrie"
889	1	Behr, Mr. Karl Howell
890	0	Dooley, Mr. Patrick

```
[891 rows x 2 columns]
```

```

# Second way of multiple columns subsetting
new_titanic_df = ['Name', 'Survived']
titanic_df[new_titanic_df]

```

	Name	Survived
0	Braund, Mr. Owen Harris	0
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	1
2	Heikkinen, Miss. Laina	1
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1
4	Allen, Mr. William Henry	0
..	...	...
886	Montvila, Rev. Juozas	0
887	Graham, Miss. Margaret Edith	1
888	Johnston, Miss. Catherine Helen "Carrie"	0
889	Behr, Mr. Karl Howell	1
890	Dooley, Mr. Patrick	0

```
[891 rows x 2 columns]
```

```

# If I want specific ages so
titanic_df['Age'] > 20 # It returns the value in boolean

```



```
0      True
1      True
2      True
3      True
4      True
```

```
...
886    True
887   False
888   False
889    True
890    True
```

Name: Age, Length: 891, dtype: bool

```
# But if I want specific ages in numbers so
titanic_df[titanic_df['Age'] > 25]
```

	PassengerId	Survived	Pclass	\
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
6	7	0	1	
..	...	...	...	
883	884	0	2	
885	886	0	3	
886	887	0	2	
889	890	1	1	
890	891	0	3	

	SibSp	\	Name	Sex	Age
1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	
2	0	Heikkinen, Miss. Laina	female	26.0	
3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	
4	0	Allen, Mr. William Henry	male	35.0	
6	0	McCarthy, Mr. Timothy J	male	54.0	
..	...	...	...	...	
883	0	Banfield, Mr. Frederick James	male	28.0	
885	0	Rice, Mrs. William (Margaret Norton)	female	39.0	
886	0	Montvila, Rev. Juozas	male	27.0	
889		Behr, Mr. Karl Howell	male	26.0	

```

0
890                                Dooley, Mr. Patrick      male  32.0
0

   Parch      Ticket    Fare Cabin Embarked
1      0         PC 17599   71.2833   C85        C
2      0  STON/O2. 3101282    7.9250   NaN        S
3      0         113803   53.1000  C123        S
4      0         373450    8.0500   NaN        S
6      0         17463   51.8625   E46        S
..     ...         ...     ...     ...     ...
883     0  C.A./SOTON 34068   10.5000   NaN        S
885     5         382652   29.1250   NaN        Q
886     0         211536   13.0000   NaN        S
889     0         111369   30.0000  C148        C
890     0         370376    7.7500   NaN        Q

[413 rows x 12 columns]

```

## 8. Subsetting based on Specific Requirements

- Passengers of age greater than 30
- Passengers whose gender is 'Male' only
- Male Passengers who Survived in the Incident
- Passengers who PClass isin([1,2])
- Give Discounts of 10 percent

```

# Passengers of age greater than 30
titanic_df[titanic_df['Age'] > 30]

```

```

   PassengerId  Survived  Pclass \
1             2         1       1
3             4         1       1
4             5         0       3
6             7         0       1
11            12         1       1
..           ...         ...     ...
873           874         0       3
879           880         1       1
881           882         0       3
885           886         0       3
890           891         0       3

   Name                               Sex  Age
SibSp \
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0
1
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)      female  35.0
1
4  Allen, Mr. William Henry                          male    35.0

```

```

0
6          McCarthy, Mr. Timothy J    male  54.0
0
11         Bonnell, Miss. Elizabeth  female  58.0
0
..          ...          ...          ...
...
873         Vander Cruyssen, Mr. Victor    male  47.0
0
879    Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)  female  56.0
0
881         Markun, Mr. Johann    male  33.0
0
885         Rice, Mrs. William (Margaret Norton)  female  39.0
0
890         Dooley, Mr. Patrick    male  32.0
0

```

```

      Parch  Ticket   Fare Cabin Embarked
1         0   PC 17599  71.2833   C85         C
3         0  113803  53.1000  C123         S
4         0  373450   8.0500   NaN         S
6         0   17463  51.8625   E46         S
11        0  113783  26.5500  C103         S
..      ...      ...      ...      ...      ...
873        0  345765   9.0000   NaN         S
879        1   11767  83.1583   C50         C
881        0  349257   7.8958   NaN         S
885        5  382652  29.1250   NaN         Q
890        0  370376   7.7500   NaN         Q

```

[305 rows x 12 columns]

```

# I can also do subsetting on objects means strings
# Passengers whose gender is 'Male' only
titanic_df[titanic_df['Sex'] == 'male']

```

```

      PassengerId  Survived  Pclass          Name
Sex \
0              1         0        3  Braund, Mr. Owen Harris
male
4              5         0        3  Allen, Mr. William Henry
male
5              6         0        3    Moran, Mr. James
male
6              7         0        1  McCarthy, Mr. Timothy J
male
7              8         0        3  Palsson, Master. Gosta Leonard
male
..          ...          ...          ...          .

```

```

..
883      884      0      2  Banfield, Mr. Frederick James
male
884      885      0      3      Sutehall, Mr. Henry Jr
male
886      887      0      2      Montvila, Rev. Juozas
male
889      890      1      1      Behr, Mr. Karl Howell
male
890      891      0      3      Dooley, Mr. Patrick
male

```

	Age	SibSp	Parch		Ticket	Fare	Cabin	Embarked
0	22.0	1	0		A/5 21171	7.2500	NaN	S
4	35.0	0	0		373450	8.0500	NaN	S
5	NaN	0	0		330877	8.4583	NaN	Q
6	54.0	0	0		17463	51.8625	E46	S
7	2.0	3	1		349909	21.0750	NaN	S
..	...	...	...		...	...	...	...
883	28.0	0	0	C.A./SOTON	34068	10.5000	NaN	S
884	25.0	0	0	SOTON/OQ	392076	7.0500	NaN	S
886	27.0	0	0		211536	13.0000	NaN	S
889	26.0	0	0		111369	30.0000	C148	C
890	32.0	0	0		370376	7.7500	NaN	Q

[577 rows x 12 columns]

*# Male Passengers who Survived in the Incident*

```
male_passengers = titanic_df['Sex'] == 'male'
```

```
servived_male = titanic_df['Survived'] == 1
```

```
titanic_df[male_passengers & servived_male]
```

	PassengerId	Survived	Pclass	Name
Sex \				
17	18	1	2	Williams, Mr. Charles Eugene
male				
21	22	1	2	Beesley, Mr. Lawrence
male				
23	24	1	1	Sloper, Mr. William Thompson
male				
36	37	1	3	Mamee, Mr. Hanna
male				
55	56	1	1	Woolner, Mr. Hugh
male				
..	...	...	...	...
...				
838	839	1	3	Chip, Mr. Chang
male				
839	840	1	1	Marechal, Mr. Pierre

male							
857	858	1	1			Daly, Mr. Peter Denis	
male							
869	870	1	3			Johnson, Master. Harold Theodor	
male							
889	890	1	1			Behr, Mr. Karl Howell	
male							

	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
17	NaN	0	0	244373	13.0000	NaN	S
21	34.0	0	0	248698	13.0000	D56	S
23	28.0	0	0	113788	35.5000	A6	S
36	NaN	0	0	2677	7.2292	NaN	C
55	NaN	0	0	19947	35.5000	C52	S
..	...	...	...	...	...	...	...
838	32.0	0	0	1601	56.4958	NaN	S
839	NaN	0	0	11774	29.7000	C47	C
857	51.0	0	0	113055	26.5500	E17	S
869	4.0	1	1	347742	11.1333	NaN	S
889	26.0	0	0	111369	30.0000	C148	C

[109 rows x 12 columns]

```
# Advance way to find Male Passengers who Survived in the Incident
titanic_df[(titanic_df['Sex'] == 'male') & (titanic_df['Survived'] == 1)]
```

	PassengerId	Survived	Pclass	Name
Sex \				
17	18	1	2	Williams, Mr. Charles Eugene
male				
21	22	1	2	Beesley, Mr. Lawrence
male				
23	24	1	1	Sloper, Mr. William Thompson
male				
36	37	1	3	Mamee, Mr. Hanna
male				
55	56	1	1	Woolner, Mr. Hugh
male				
..	...	...	...	...
...				
838	839	1	3	Chip, Mr. Chang
male				
839	840	1	1	Marechal, Mr. Pierre
male				
857	858	1	1	Daly, Mr. Peter Denis
male				
869	870	1	3	Johnson, Master. Harold Theodor
male				
889	890	1	1	Behr, Mr. Karl Howell

male

	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
17	NaN	0	0	244373	13.0000	NaN	S
21	34.0	0	0	248698	13.0000	D56	S
23	28.0	0	0	113788	35.5000	A6	S
36	NaN	0	0	2677	7.2292	NaN	C
55	NaN	0	0	19947	35.5000	C52	S
..	...	...	...	...	...	...	...
838	32.0	0	0	1601	56.4958	NaN	S
839	NaN	0	0	11774	29.7000	C47	C
857	51.0	0	0	113055	26.5500	E17	S
869	4.0	1	1	347742	11.1333	NaN	S
889	26.0	0	0	111369	30.0000	C148	C

[109 rows x 12 columns]

```
#Passengers who PClass isin([1,2])  
# isin is used to identify that the provided data is in the column or  
# not and if the data is in the column so it will return that column  
titanic_df[titanic_df['Pclass'].isin([1, 2])]
```

	PassengerId	Survived	Pclass	\
1	2	1	1	
3	4	1	1	
6	7	0	1	
9	10	1	2	
11	12	1	1	
..	...	...	...	
880	881	1	2	
883	884	0	2	
886	887	0	2	
887	888	1	1	
889	890	1	1	

	Name	Sex	Age
SibSp \			
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0
1			
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0
1			
6	McCarthy, Mr. Timothy J	male	54.0
0			
9	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0
1			
11	Bonnell, Miss. Elizabeth	female	58.0
0			
..	...	...	...
...			
880	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0

```

0
883          Banfield, Mr. Frederick James    male  28.0
0
886          Montvila, Rev. Juozas          male  27.0
0
887          Graham, Miss. Margaret Edith    female 19.0
0
889          Behr, Mr. Karl Howell           male  26.0
0

```

```

      Parch      Ticket    Fare Cabin Embarked
1         0          PC 17599   71.2833    C85        C
3         0          113803   53.1000   C123        S
6         0          17463   51.8625    E46        S
9         0          237736   30.0708    NaN        C
11        0          113783   26.5500   C103        S
..        ...          ...      ...      ...      ...
880        1          230433   26.0000    NaN        S
883        0  C.A./SOTON 34068   10.5000    NaN        S
886        0          211536   13.0000    NaN        S
887        0          112053   30.0000    B42        S
889        0          111369   30.0000   C148        C

```

[400 rows x 12 columns]

```

# Give Discounts of 10 percent
titanic_df['Discounts'] = titanic_df.Fare / 10

titanic_df

```

```

      PassengerId  Survived  Pclass \
0               1         0        3
1               2         1        1
2               3         1        3
3               4         1        1
4               5         0        3
..            ...      ...      ...
886            887         0        2
887            888         1        1
888            889         0        3
889            890         1        1
890            891         0        3

```

```

      Name      Sex  Age
SibSp \
0      Braund, Mr. Owen Harris    male  22.0
1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0
1
2      Heikkinen, Miss. Laina    female  26.0

```

```

0
3      Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0
1
4      Allen, Mr. William Henry  male  35.0
0
..      ...      ...      ...
...
886      Montvila, Rev. Juozas  male  27.0
0
887      Graham, Miss. Margaret Edith  female  19.0
0
888      Johnston, Miss. Catherine Helen "Carrie"  female  NaN
1
889      Behr, Mr. Karl Howell  male  26.0
0
890      Dooley, Mr. Patrick  male  32.0
0

```

	Parch	Ticket	Fare	Cabin	Embarked	Discounts
0	0	A/5 21171	7.2500	NaN	S	0.72500
1	0	PC 17599	71.2833	C85	C	7.12833
2	0	STON/O2. 3101282	7.9250	NaN	S	0.79250
3	0	113803	53.1000	C123	S	5.31000
4	0	373450	8.0500	NaN	S	0.80500
..	...	...	...	...	...	...
886	0	211536	13.0000	NaN	S	1.30000
887	0	112053	30.0000	B42	S	3.00000
888	2	W./C. 6607	23.4500	NaN	S	2.34500
889	0	111369	30.0000	C148	C	3.00000
890	0	370376	7.7500	NaN	Q	0.77500

[891 rows x 13 columns]