

# Introduction to Pandas

## Loading the dependencies

In [1]:

```
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')
```

## Loading the dataset

In [2]:

```
data = pd.read_csv('./titanic.csv')
data.head()
```

Out[2]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500

In [3]:

```
data.head(10)
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708



In [4]:

```
data.tail()
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	nan
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	nan
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	nan
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C 85
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	nan

In [5]:

```
data.sample(10)
```

Out[5]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	
198	199	1	3	Madigan, Miss. Margaret "Maggie"	female	NaN	0	0	370370	7.7500
603	604	0	3	Torber, Mr. Ernst William	male	44.0	0	0	364511	8.0500
814	815	0	3	Tomlin, Mr. Ernest Portage	male	30.5	0	0	364499	8.0500
242	243	0	2	Coleridge, Mr. Reginald Charles	male	29.0	0	0	W./C. 14263	10.5000
108	109	0	3	Rekic, Mr. Tido	male	38.0	0	0	349249	7.8958
351	352	0	1	Williams-Lambert, Mr. Fletcher Fellows	male	NaN	0	0	113510	35.0000
150	151	0	2	Bateman, Rev. Robert James	male	51.0	0	0	S.O.P. 1166	12.5250
230	231	1	1	Harris, Mrs. Henry Birkhardt (Irene Wallach)	female	35.0	1	0	36973	83.4750
382	383	0	3	Tikkanen, Mr. Juho	male	32.0	0	0	STON/O 2. 3101293	7.9250
83	84	0	1	Carrau, Mr. Francisco M	male	28.0	0	0	113059	47.1000



In [6]:

data.info()

```
<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: 83.7+ KB
```

In [7]:

data.describe()

Out[7]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
<b>count</b>	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
<b>mean</b>	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
<b>std</b>	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
<b>min</b>	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
<b>25%</b>	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
<b>50%</b>	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
<b>75%</b>	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
<b>max</b>	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [8]:

```
data.describe().T
```

Out[8]:

	count	mean	std	min	25%	50%	75%	max
<b>PassengerId</b>	891.0	446.000000	257.353842	1.00	223.5000	446.0000	668.5	891.0000
<b>Survived</b>	891.0	0.383838	0.486592	0.00	0.0000	0.0000	1.0	1.0000
<b>Pclass</b>	891.0	2.308642	0.836071	1.00	2.0000	3.0000	3.0	3.0000
<b>Age</b>	714.0	29.699118	14.526497	0.42	20.1250	28.0000	38.0	80.0000
<b>SibSp</b>	891.0	0.523008	1.102743	0.00	0.0000	0.0000	1.0	8.0000
<b>Parch</b>	891.0	0.381594	0.806057	0.00	0.0000	0.0000	0.0	6.0000
<b>Fare</b>	891.0	32.204208	49.693429	0.00	7.9104	14.4542	31.0	512.3292

In [9]:

```
data.size
```

Out[9]:

10692

In [10]:

```
data.shape
```

Out[10]:

(891, 12)

In [11]:

```
data.columns
```

Out[11]:

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

In [12]:

```
data.Pclass.dtype
```

Out[12]:

dtype('int64')

**Null count**

In [13]:

```
data.isnull().sum()
```

Out[13]:

```
PassengerId      0
Survived          0
Pclass           0
Name             0
Sex              0
Age             177
SibSp            0
Parch           0
Ticket           0
Fare             0
Cabin           687
Embarked         2
dtype: int64
```

### Null percentage

In [14]:

```
data.isnull().mean()*100
```

Out[14]:

```
PassengerId      0.000000
Survived          0.000000
Pclass           0.000000
Name             0.000000
Sex              0.000000
Age             19.865320
SibSp            0.000000
Parch           0.000000
Ticket           0.000000
Fare             0.000000
Cabin           77.104377
Embarked         0.224467
dtype: float64
```

In [15]:

```
data.duplicated().sum()
```

Out[15]:

0

Dropping the null values

In [16]:

```
data.dropna()
```

Out[16]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
1	2	1	1Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
3	4	1	3Futelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
6	7	0	6McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
10	11	1	10Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000
11	12	1	11Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500
...	...	...	...	...	...	...	...	...	...
871	872	1	871Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542
872	873	0	872Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000
879	880	1	879Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583
887	888	1	887Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
889	890	1	889Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000

183 rows × 12 columns





In [17]:

```
data.isnull().sum()
```

Out[17]:

```
PassengerId      0
Survived          0
Pclass           0
Name             0
Sex              0
Age             177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin           687
Embarked         2
dtype: int64
```

In [18]:

```
data.dropna(inplace=True)
```

In [19]:

```
data.isnull().sum()
```

Out[19]:

```
PassengerId      0
Survived          0
Pclass           0
Name             0
Sex              0
Age             0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin            0
Embarked         0
dtype: int64
```

## Dropping the duplicated values

In [20]:

```
data.drop_duplicates(inplace=True)
```

dropping the columns

In [21]:

```
data.drop(columns='PassengerId')
```

Out[21]:

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
1	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
6	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
10	1	3	Sandstrom, Miss. Marguerite	female	4.0	1	1	PP 9549	16.7000	G6	S

In [22]:

```
data.head()
```

Out[22]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500

In [23]:

```
data.drop(columns='PassengerId', inplace=True)
```

In [24]:

```
data.head()
```

Out[24]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embar
1	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
3	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
6	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	
10	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	
11	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	

## Slicing

In [ ]:

```
data.loc[50:60]
```

In [ ]:

```
data.iloc[10:30]
```

In [ ]:

```
data.iloc[:30]
```

In [ ]:

```
data.iloc[50:]
```

## Calling the columns

In [25]:

```
data.Survived
```

Out[25]:

```
1      1
3      1
6      0
10     1
11     1
```

..

```
871    1
872    0
879    1
887    1
889    1
```

Name: Survived, Length: 183, dtype: int64

In [26]:

```
data["Survived"]
```

Out[26]:

```
1      1
3      1
6      0
10     1
11     1
```

..

```
871    1
872    0
879    1
887    1
889    1
```

Name: Survived, Length: 183, dtype: int64

In [27]:

```
data.Pclass.value_counts()
```

Out[27]:

```
1    158
2     15
3     10
```

Name: Pclass, dtype: int64

In [28]:

```
data.Pclass.nunique()
```

Out[28]:

```
3
```

In [ ]:

```
data.Age.mean()
```

In [ ]:

```
data.Fare.sum()
```

Sorting

In [29]:

```
data.sort_values(by='Age')
```

Out[29]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
305	1	1	Allison, Master. Hudson Trevor	male	0.92	1	2	113781	151.5500	C22 C26	
183	1	2	Becker, Master. Richard F	male	1.00	2	1	230136	39.0000	F4	
205	0	3	Strom, Miss. Telma Matilda	female	2.00	0	1	347054	10.4625	G6	
297	0	1	Allison, Miss. Helen Lorraine	female	2.00	1	2	113781	151.5500	C22 C26	
340	1	2	Navratil, Master. Edmond Roger	male	2.00	1	1	230080	26.0000	F2	
...	...	...	...	...	...	...	...	...	...	...	
54	0	1	Ostby, Mr. Engelhart Cornelius	male	65.00	0	1	113509	61.9792	B30	
456	0	1	Millet, Mr. Francis Davis	male	65.00	0	0	13509	26.5500	E38	
745	0	1	Crosby, Capt. Edward Gifford	male	70.00	1	1	WE/P 5735	71.0000	B22	
96	0	1	Goldschmidt, Mr. George B	male	71.00	0	0	PC 17754	34.6542	A5	
630	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.00	0	0	27042	30.0000	A23	

183 rows × 11 columns



In [30]:

```
data.sort_values(by='Age', ascending=False)
```

Out[30]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
630	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.00	0	0	27042	30.0000	A23	
96	0	1	Goldschmidt, Mr. George B	male	71.00	0	0	PC 17754	34.6542	A5	
745	0	1	Crosby, Capt. Edward Gifford	male	70.00	1	1	WE/P 5735	71.0000	B22	
456	0	1	Millet, Mr. Francis Davis	male	65.00	0	0	13509	26.5500	E38	
54	0	1	Ostby, Mr. Engelhart Cornelius	male	65.00	0	1	113509	61.9792	B30	
...	...	...	...	...	...	...	...	...	...	...	...
297	0	1	Allison, Miss. Helen Loraine	female	2.00	1	2	113781	151.5500	C22 C26	
205	0	3	Strom, Miss. Telma Matilda	female	2.00	0	1	347054	10.4625	G6	
340	1	2	Navratil, Master. Edmond Roger	male	2.00	1	1	230080	26.0000	F2	
183	1	2	Becker, Master. Richard F	male	1.00	2	1	230136	39.0000	F4	
305	1	1	Allison, Master. Hudson Trevor	male	0.92	1	2	113781	151.5500	C22 C26	

183 rows × 11 columns



In [31]:

```
data.sort_values(by='Name')
```

Out[31]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
730	1	1	Allen, Miss. Elisabeth Walton	female	29.00	0	0	24160	211.3375	B5	
305	1	1	Allison, Master. Hudson Trevor	male	0.92	1	2	113781	151.5500	C22 C26	
297	0	1	Allison, Miss. Helen Loraine	female	2.00	1	2	113781	151.5500	C22 C26	
498	0	1	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.00	1	2	113781	151.5500	C22 C26	
460	1	1	Anderson, Mr. Harry	male	48.00	0	0	19952	26.5500	E12	
...	...	...	...	...	...	...	...	...	...	...	
124	0	1	White, Mr. Percival Wayland	male	54.00	0	1	35281	77.2875	D26	
102	0	1	White, Mr. Richard Frasar	male	21.00	0	1	35281	77.2875	D26	
318	1	1	Wick, Miss. Mary Natalie	female	31.00	0	2	36928	164.8667	C7	
377	0	1	Widener, Mr. Harry Elkins	male	27.00	0	2	113503	211.5000	C82	
325	1	1	Young, Miss. Marie Grice	female	36.00	0	0	PC 17760	135.6333	C32	

183 rows × 11 columns



In [32]:

```
data.sort_values(by='Name', ascending=False)
```

Out[32]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
325	1	1	Young, Miss. Marie Grice	female	36.00	0	0	PC 17760	135.6333	C32	
377	0	1	Widener, Mr. Harry Elkins	male	27.00	0	2	113503	211.5000	C82	
318	1	1	Wick, Miss. Mary Natalie	female	31.00	0	2	36928	164.8667	C7	
102	0	1	White, Mr. Richard Frasar	male	21.00	0	1	35281	77.2875	D26	
124	0	1	White, Mr. Percival Wayland	male	54.00	0	1	35281	77.2875	D26	
...	...	...	...	...	...	...	...	...	...	...	
460	1	1	Anderson, Mr. Harry	male	48.00	0	0	19952	26.5500	E12	
498	0	1	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.00	1	2	113781	151.5500	C22 C26	
297	0	1	Allison, Miss. Helen Loraine	female	2.00	1	2	113781	151.5500	C22 C26	
305	1	1	Allison, Master. Hudson Trevor	male	0.92	1	2	113781	151.5500	C22 C26	
730	1	1	Allen, Miss. Elisabeth Walton	female	29.00	0	0	24160	211.3375	B5	

183 rows × 11 columns





Filtering

In [33]:

```
data[data["Age"]>30]
```

Out[33]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
3	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
6	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	
11	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	
21	1	2	Beesley, Mr. Lawrence	male	34.0	0	0	248698	13.0000	D56	
...	...	...	...	...	...	...	...	...	...	...	
862	1	1	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48.0	0	0	17466	25.9292	D17	
867	0	1	Roebeling, Mr. Washington Augustus II	male	31.0	0	0	PC 17590	50.4958	A24	
871	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	
872	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B51 B53 B55	
879	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C50	

114 rows × 11 columns



In [34]:

```
data[data["Age"]==42]
```

Out[34]:

	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
621	1	1	Kimball, Mr. Edwin Nelson Jr	male	42.0	1	0	11753	52.5542	D19	
699	0	3	Humblen, Mr. Adolf Mathias Nicolai Olsen	male	42.0	0	0	348121	7.6500	F G63	
707	1	1	Calderhead, Mr. Edward Pennington	male	42.0	0	0	PC 17476	26.2875	E24	

