

Name : Chaitanya More

Roll_no :- 01 "B"

Batch :- TB1-B2

Practical No 1

Perform the following operations using Python on any open source dataset (e.g., data.csv) the data types (i.e., character, numeric, integer, factor, and logical) of the variables in the.

1. Import all the required Python Libraries.
2. Locate an open source data from the web (e.g., <https://www.kaggle.com>). Provide a clear description of the data and its source (i.e., URL of the web site).
3. Load the Dataset into pandas dataframe.
4. Data Preprocessing: check for missing values in the data using pandas isnull(), describe() function to get some initial statistics. Provide variable descriptions. Types of variables etc. Check the dimensions of the data frame.
5. Data Formatting and Data Normalization: Summarize the types of variables by checking data set. If variables are not in the correct data type, apply proper type conversions.
6. Turn categorical variables into quantitative variables in Python.

```
In [2]: import pandas as pd
import numpy as np
```

```
In [3]: df=pd.read_csv("Titanic.csv")
```

```
In [5]: df
```

```
Out[5]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

```
In [7]: df.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 [8]: df.describe()
```

Out[8]:

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

In [9]:

df.isnull().sum()

Out[9]:

PassengerId0

Survived0

Pclass0

Name0

Sex0

Age177

SibSp0

Parch0

Ticket0

Fare0

Cabin687

Embarked2

dtype: int64

In [10]:

df.head()

Out[10]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [11]:

df.tail()

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

In [13]:

df['Fare']=df['Fare'].astype('int64')

In [14]:

df.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 int64
10 Cabin 204 non-null object
11 Embarked 889 non-null object
dtypes: float64(1), int64(6), object(5)
memory usage: 83.7+ KB

In [19]:

df.isnull().sum()

Out[19]: PassengerId 0
Survived 0
Pclass 0
Name 0
Age 177
SibSp 0
Parch 0
Ticket 0
Fare 0
Cabin 687
Embarked 2
Sex_female 0
dtype: int64

In [24]: df['Age'].replace(np.nan, df['Age'].mean(), inplace=True)

In [25]: df.isnull().sum()

Out[25]: PassengerId 0
Survived 0
Pclass 0
Name 0
Age 0
SibSp 0
Parch 0
Ticket 0
Fare 0
Cabin 687
Embarked 2
Sex_female 0
dtype: int64

In [26]: df["Cabin"].fillna("A32", inplace = True)

In [27]: df.isnull().sum()

Out[27]: PassengerId 0
Survived 0
Pclass 0
Name 0
Age 0
SibSp 0
Parch 0
Ticket 0
Fare 0
Cabin 0
Embarked 2
Sex_female 0
dtype: int64

In [28]: df

Out[28]:

	PassengerId	Survived	Pclass	Name	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Sex_female
0	1	0	3	Braund, Mr. Owen Harris	22.000000	1	0	A/5 21171	7	A32	S	0
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	38.000000	1	0	PC 17599	71	C85	C	1
2	3	1	3	Heikkinen, Miss. Laina	26.000000	0	0	STON/O2. 3101282	7	A32	S	1
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35.000000	1	0	113803	53	C123	S	1
4	5	0	3	Allen, Mr. William Henry	35.000000	0	0	373450	8	A32	S	0
...
886	887	0	2	Montvila, Rev. Juozas	27.000000	0	0	211536	13	A32	S	0
887	888	1	1	Graham, Miss. Margaret Edith	19.000000	0	0	112053	30	B42	S	1
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	29.699118	1	2	W./C. 6607	23	A32	S	1
889	890	1	1	Behr, Mr. Karl Howell	26.000000	0	0	111369	30	C148	C	0
890	891	0	3	Dooley, Mr. Patrick	32.000000	0	0	370376	7	A32	Q	0

891 rows × 12 columns

In []: