nhehk7ow0

December 13, 2024

1 DATA Cleaning , Missing Value Treatment

[2]:
```
#Name :  Devesh J Arbat
#Roll no. : 06
#Section : A
```

[3]:
```
#Aim: To perform Data Processing, Data Cleaning, Missing Value Treatment
```

[4]:
```
import pandas as pd
```

[5]:
```
import os
```

[6]:
```
os.getcwd()
```

[6]:
```
'C:\\Users\\91876'
```

[7]:
```
os.chdir("C:\\Users\\91876\\OneDrive\\Desktop\\Data Science")
```

[8]:
```
data=pd.read_csv("titanic.csv")
```

[9]:
```
data
```

[9]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	S	2
1	1.0	1.0	Allison, Master. Hudson Trevor									
2	1.0	0.0	Allison, Miss. Helen Loraine									
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton									
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)									
…	…	…	…									
1305	3.0	0.0	Zabour, Miss. Thamine									
1306	3.0	0.0	Zakarian, Mr. Mapriededer									
1307	3.0	0.0	Zakarian, Mr. Ortin									
1308	3.0	0.0	Zimmerman, Mr. Leo									
1309	NaN	NaN	NaN									

```
1       male   0.9167    1.0    2.0  113781  151.5500  C22 C26        S   11
2     female   2.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
3       male  30.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
4     female  25.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
…      …       …    …    …      …        …       …       …  …
1305  female     NaN    1.0    0.0    2665   14.4542      NaN        C  NaN
1306    male  26.5000    0.0    0.0    2656    7.2250      NaN        C  NaN
1307    male  27.0000    0.0    0.0    2670    7.2250      NaN        C  NaN
1308    male  29.0000    0.0    0.0  315082    7.8750      NaN        S  NaN
1309     NaN     NaN    NaN    NaN     NaN       NaN      NaN      NaN  NaN

       body                        home.dest
0       NaN                     St Louis, MO
1       NaN  Montreal, PQ / Chesterville, ON
2       NaN  Montreal, PQ / Chesterville, ON
3     135.0  Montreal, PQ / Chesterville, ON
4       NaN  Montreal, PQ / Chesterville, ON
…     …                                 …
1305    NaN                             NaN
1306  304.0                             NaN
1307    NaN                             NaN
1308    NaN                             NaN
1309    NaN                             NaN

[1310 rows x 14 columns]
```

[10]:
```
data.head(40)
```

[10]:
```
    pclass  survived                                             name  \
0      1.0       1.0                    Allen, Miss. Elisabeth Walton
1      1.0       1.0                   Allison, Master. Hudson Trevor
2      1.0       0.0                     Allison, Miss. Helen Loraine
3      1.0       0.0             Allison, Mr. Hudson Joshua Creighton
4      1.0       0.0  Allison, Mrs. Hudson J C (Bessie Waldo Daniels)
5      1.0       1.0                                  Anderson, Mr. Harry
6      1.0       1.0                Andrews, Miss. Kornelia Theodosia
7      1.0       0.0                            Andrews, Mr. Thomas Jr
8      1.0       1.0    Appleton, Mrs. Edward Dale (Charlotte Lamson)
9      1.0       0.0                          Artagaveytia, Mr. Ramon
10     1.0       0.0                            Astor, Col. John Jacob
11     1.0       1.0  Astor, Mrs. John Jacob (Madeleine Talmadge Force)
12     1.0       1.0                     Aubart, Mme. Leontine Pauline
13     1.0       1.0                      Barber, Miss. Ellen "Nellie"
14     1.0       1.0             Barkworth, Mr. Algernon Henry Wilson
15     1.0       0.0                                  Baumann, Mr. John D
16     1.0       0.0                           Baxter, Mr. Quigg Edmond
17     1.0       1.0   Baxter, Mrs. James (Helene DeLaudeniere Chaput)
```

18	1.0	1.0	Bazzani, Miss. Albina
19	1.0	0.0	Beattie, Mr. Thomson
20	1.0	1.0	Beckwith, Mr. Richard Leonard
21	1.0	1.0	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)
22	1.0	1.0	Behr, Mr. Karl Howell
23	1.0	1.0	Bidois, Miss. Rosalie
24	1.0	1.0	Bird, Miss. Ellen
25	1.0	0.0	Birnbaum, Mr. Jakob
26	1.0	1.0	Bishop, Mr. Dickinson H
27	1.0	1.0	Bishop, Mrs. Dickinson H (Helen Walton)
28	1.0	1.0	Bissette, Miss. Amelia
29	1.0	1.0	Bjornstrom-Steffansson, Mr. Mauritz Hakan
30	1.0	0.0	Blackwell, Mr. Stephen Weart
31	1.0	1.0	Blank, Mr. Henry
32	1.0	1.0	Bonnell, Miss. Caroline
33	1.0	1.0	Bonnell, Miss. Elizabeth
34	1.0	0.0	Borebank, Mr. John James
35	1.0	1.0	Bowen, Miss. Grace Scott
36	1.0	1.0	Bowerman, Miss. Elsie Edith
37	1.0	1.0	Bradley, Mr. George ("George Arthur Brayton")
38	1.0	0.0	Brady, Mr. John Bertram
39	1.0	0.0	Brandeis, Mr. Emil

	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	\
0	female	29.0000	0.0	0.0	24160	211.3375	B5	S	2	
1	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	S	11	
2	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	
3	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	
4	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	
5	male	48.0000	0.0	0.0	19952	26.5500	E12	S	3	
6	female	63.0000	1.0	0.0	13502	77.9583	D7	S	10	
7	male	39.0000	0.0	0.0	112050	0.0000	A36	S	NaN	
8	female	53.0000	2.0	0.0	11769	51.4792	C101	S	D	
9	male	71.0000	0.0	0.0	PC 17609	49.5042	NaN	C	NaN	
10	male	47.0000	1.0	0.0	PC 17757	227.5250	C62 C64	C	NaN	
11	female	18.0000	1.0	0.0	PC 17757	227.5250	C62 C64	C	4	
12	female	24.0000	0.0	0.0	PC 17477	69.3000	B35	C	9	
13	female	26.0000	0.0	0.0	19877	78.8500	NaN	S	6	
14	male	80.0000	0.0	0.0	27042	30.0000	A23	S	B	
15	male	NaN	0.0	0.0	PC 17318	25.9250	NaN	S	NaN	
16	male	24.0000	0.0	1.0	PC 17558	247.5208	B58 B60	C	NaN	
17	female	50.0000	0.0	1.0	PC 17558	247.5208	B58 B60	C	6	
18	female	32.0000	0.0	0.0	11813	76.2917	D15	C	8	
19	male	36.0000	0.0	0.0	13050	75.2417	C6	C	A	
20	male	37.0000	1.0	1.0	11751	52.5542	D35	S	5	
21	female	47.0000	1.0	1.0	11751	52.5542	D35	S	5	
22	male	26.0000	0.0	0.0	111369	30.0000	C148	C	5	

23	female	42.0000	0.0	0.0	PC 17757	227.5250	NaN	C	4
24	female	29.0000	0.0	0.0	PC 17483	221.7792	C97	S	8
25	male	25.0000	0.0	0.0	13905	26.0000	NaN	C	NaN
26	male	25.0000	1.0	0.0	11967	91.0792	B49	C	7
27	female	19.0000	1.0	0.0	11967	91.0792	B49	C	7
28	female	35.0000	0.0	0.0	PC 17760	135.6333	C99	S	8
29	male	28.0000	0.0	0.0	110564	26.5500	C52	S	D
30	male	45.0000	0.0	0.0	113784	35.5000	T	S	NaN
31	male	40.0000	0.0	0.0	112277	31.0000	A31	C	7
32	female	30.0000	0.0	0.0	36928	164.8667	C7	S	8
33	female	58.0000	0.0	0.0	113783	26.5500	C103	S	8
34	male	42.0000	0.0	0.0	110489	26.5500	D22	S	NaN
35	female	45.0000	0.0	0.0	PC 17608	262.3750	NaN	C	4
36	female	22.0000	0.0	1.0	113505	55.0000	E33	S	6
37	male	NaN	0.0	0.0	111427	26.5500	NaN	S	9
38	male	41.0000	0.0	0.0	113054	30.5000	A21	S	NaN
39	male	48.0000	0.0	0.0	PC 17591	50.4958	B10	C	NaN

	body	home.dest
0	NaN	St Louis, MO
1	NaN	Montreal, PQ / Chesterville, ON
2	NaN	Montreal, PQ / Chesterville, ON
3	135.0	Montreal, PQ / Chesterville, ON
4	NaN	Montreal, PQ / Chesterville, ON
5	NaN	New York, NY
6	NaN	Hudson, NY
7	NaN	Belfast, NI
8	NaN	Bayside, Queens, NY
9	22.0	Montevideo, Uruguay
10	124.0	New York, NY
11	NaN	New York, NY
12	NaN	Paris, France
13	NaN	NaN
14	NaN	Hessle, Yorks
15	NaN	New York, NY
16	NaN	Montreal, PQ
17	NaN	Montreal, PQ
18	NaN	NaN
19	NaN	Winnipeg, MN
20	NaN	New York, NY
21	NaN	New York, NY
22	NaN	New York, NY
23	NaN	NaN
24	NaN	NaN
25	148.0	San Francisco, CA
26	NaN	Dowagiac, MI
27	NaN	Dowagiac, MI

```
28    NaN                                NaN
29    NaN  Stockholm, Sweden / Washington, DC
30    NaN                        Trenton, NJ
31    NaN                     Glen Ridge, NJ
32    NaN                     Youngstown, OH
33    NaN   Birkdale, England Cleveland, Ohio
34    NaN               London / Winnipeg, MB
35    NaN                    Cooperstown, NY
36    NaN    St Leonards-on-Sea, England Ohio
37    NaN                     Los Angeles, CA
38    NaN                        Pomeroy, WA
39  208.0                          Omaha, NE
```

[11]:
```
data.tail(10)
```

[11]:
```
      pclass  survived                                     name     sex   age  \
1300     3.0       1.0  Yasbeck, Mrs. Antoni (Selini Alexander)  female  15.0
1301     3.0       0.0                     Youseff, Mr. Gerious    male  45.5
1302     3.0       0.0                        Yousif, Mr. Wazli    male   NaN
1303     3.0       0.0                    Yousseff, Mr. Gerious    male   NaN
1304     3.0       0.0                     Zabour, Miss. Hileni  female  14.5
1305     3.0       0.0                    Zabour, Miss. Thamine  female   NaN
1306     3.0       0.0                Zakarian, Mr. Mapriededer    male  26.5
1307     3.0       0.0                      Zakarian, Mr. Ortin    male  27.0
1308     3.0       0.0                     Zimmerman, Mr. Leo    male  29.0
1309     NaN       NaN                                    NaN     NaN   NaN

      sibsp  parch  ticket     fare cabin embarked boat   body home.dest
1300    1.0    0.0    2659  14.4542   NaN        C  NaN    NaN       NaN
1301    0.0    0.0    2628   7.2250   NaN        C  NaN  312.0       NaN
1302    0.0    0.0    2647   7.2250   NaN        C  NaN    NaN       NaN
1303    0.0    0.0    2627  14.4583   NaN        C  NaN    NaN       NaN
1304    1.0    0.0    2665  14.4542   NaN        C  NaN  328.0       NaN
1305    1.0    0.0    2665  14.4542   NaN        C  NaN    NaN       NaN
1306    0.0    0.0    2656   7.2250   NaN        C  NaN  304.0       NaN
1307    0.0    0.0    2670   7.2250   NaN        C  NaN    NaN       NaN
1308    0.0    0.0  315082   7.8750   NaN        S  NaN    NaN       NaN
1309    NaN    NaN     NaN      NaN   NaN      NaN  NaN    NaN       NaN
```

[12]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1310 entries, 0 to 1309
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   pclass     1309 non-null   float64
```

```
 1   survived   1309 non-null   float64
 2   name       1309 non-null   object
 3   sex        1309 non-null   object
 4   age        1046 non-null   float64
 5   sibsp      1309 non-null   float64
 6   parch      1309 non-null   float64
 7   ticket     1309 non-null   object
 8   fare       1308 non-null   float64
 9   cabin      295 non-null    object
 10  embarked   1307 non-null   object
 11  boat       486 non-null    object
 12  body       121 non-null    float64
 13  home.dest  745 non-null    object
dtypes: float64(7), object(7)
memory usage: 143.4+ KB
```

[13]:
```
data.describe()
```

[13]:

	pclass	survived	age	sibsp	parch
count	1309.000000	1309.000000	1046.000000	1309.000000	1309.000000
mean	2.294882	0.381971	29.881135	0.498854	0.385027
std	0.837836	0.486055	14.413500	1.041658	0.865560
min	1.000000	0.000000	0.166700	0.000000	0.000000
25%	2.000000	0.000000	21.000000	0.000000	0.000000
50%	3.000000	0.000000	28.000000	0.000000	0.000000
75%	3.000000	1.000000	39.000000	1.000000	0.000000
max	3.000000	1.000000	80.000000	8.000000	9.000000

	fare	body
count	1308.000000	121.000000
mean	33.295479	160.809917
std	51.758668	97.696922
min	0.000000	1.000000
25%	7.895800	72.000000
50%	14.454200	155.000000
75%	31.275000	256.000000
max	512.329200	328.000000

[14]:
```
data.shape
```

[14]: (1310, 14)

[15]:
```
data.size
```

[15]: 18340

[16]:
```
data.ndim
```

[16]: 2

[17]:
```
data.isna()
```

[17]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	\
0	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	
…	…	…	…	…	…	…	…	…	…	
1305	False	False	False	False	True	False	False	False	False	
1306	False	False	False	False	False	False	False	False	False	
1307	False	False	False	False	False	False	False	False	False	
1308	False	False	False	False	False	False	False	False	False	
1309	True	True	True	True	True	True	True	True	True	

	cabin	embarked	boat	body	home.dest
0	False	False	False	True	False
1	False	False	False	True	False
2	False	False	True	True	False
3	False	False	True	False	False
4	False	False	True	True	False
…	…	…	…	…	…
1305	True	False	True	True	True
1306	True	False	True	False	True
1307	True	False	True	True	True
1308	True	False	True	True	True
1309	True	True	True	True	True

[1310 rows x 14 columns]

[18]:
```
data.isna().any()
```

[18]:
```
pclass       True
survived     True
name         True
sex          True
age          True
sibsp        True
parch        True
ticket       True
fare         True
cabin        True
embarked     True
boat         True
body         True
```

```
home.dest    True
dtype: bool
```

[19]:
```
data.isna().sum()
```

[19]:
```
pclass          1
survived        1
name            1
sex             1
age           264
sibsp           1
parch           1
ticket          1
fare            2
cabin        1015
embarked        3
boat          824
body         1189
home.dest     565
dtype: int64
```

[20]:
```
data["age"].fillna(29.699118)
```

[20]:
```
0       29.000000
1        0.916700
2        2.000000
3       30.000000
4       25.000000
           …
1305    29.699118
1306    26.500000
1307    27.000000
1308    29.000000
1309    29.699118
Name: age, Length: 1310, dtype: float64
```

[21]:
```
data.isna().sum()
```

[21]:
```
pclass          1
survived        1
name            1
sex             1
age           264
sibsp           1
parch           1
ticket          1
fare            2
```

```
cabin        1015
embarked        3
boat          824
body         1189
home.dest     565
dtype: int64
```

[22]:
```
data.any()
```

[22]:
```
pclass       True
survived     True
name         True
sex          True
age          True
sibsp        True
parch        True
ticket       True
fare         True
cabin        True
embarked     True
boat         True
body         True
home.dest    True
dtype: bool
```

[23]:
```
data=data.dropna()
```

[24]:
```
data.any()
```

[24]:
```
pclass       False
survived     False
name         False
sex          False
age          False
sibsp        False
parch        False
ticket       False
fare         False
cabin        False
embarked     False
boat         False
body         False
home.dest    False
dtype: bool
```

[25]:
```
data.isna().sum()
```

```
[25]: pclass       0.0
      survived     0.0
      name         0.0
      sex          0.0
      age          0.0
      sibsp        0.0
      parch        0.0
      ticket       0.0
      fare         0.0
      cabin        0.0
      embarked     0.0
      boat         0.0
      body         0.0
      home.dest    0.0
      dtype: float64
```