Handling Missing values

October 18, 2022

[1]:
```
import pandas as pd
import numpy as np

data2=pd.read_csv("data2.csv")
data2
```

[1]:

	Name	Python	Machine Learning	Age	AI	NLP	Cv
0	jay	90	?	23	90	?	23
1	raj	missing	89	NaN	missing	89	NaN
2	kumar	78	78	?	78	78	?
3	suraj	NaN	NaN	missing	NaN	NaN	missing
4	pawan	missing	NaN	21	missing	NaN	21
5	viraj	65	56	22	65	56	22
6	vijay	45	67	?	45	67	?
7	sumit	80	45	24	80	45	24
8	akash	70	67	missing	70	67	missing
9	sujit	NaN	NaN	22	NaN	NaN	22

[3]:
```
data2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   Name              10 non-null     object
 1   Python            8 non-null      object
 2   Machine Learning  7 non-null      object
 3   Age               9 non-null      object
 4   AI                8 non-null      object
 5   NLP               7 non-null      object
 6   Cv                9 non-null      object
dtypes: object(7)
memory usage: 688.0+ bytes
```

[4]:
```
data2["Python"].unique()
```

[4]:
```
array(['90', 'missing', '78', nan, '65', '45', '80', '70'], dtype=object)
```

[5]:
```
data2["Python"].replace('missing',np.nan,inplace=True)
```

[6]:
```
data2
```

[6]:

	Name	Python	Machine Learning	Age	AI	NLP	Cv
0	jay	90	?	23	90	?	23
1	raj	NaN	89	NaN	missing	89	NaN
2	kumar	78	78	?	78	78	?
3	suraj	NaN	NaN	missing	NaN	NaN	missing
4	pawan	NaN	NaN	21	missing	NaN	21
5	viraj	65	56	22	65	56	22
6	vijay	45	67	?	45	67	?
7	sumit	80	45	24	80	45	24
8	akash	70	67	missing	70	67	missing
9	sujit	NaN	NaN	22	NaN	NaN	22

[13]:
```
data2["Machine Learning"].unique()
```

[13]:
```
array(['?', '89', '78', nan, '56', '67', '45'], dtype=object)
```

[14]:
```
data2["Machine Learning"].replace('?',np.nan,inplace=True)

data2["Age"].replace(['?','missing'],np.nan,inplace=True)

data2["AI"].replace('missing',np.nan,inplace=True)

data2["NLP"].replace('?',np.nan,inplace=True)

data2["Cv"].replace(['missing','?'],np.nan,inplace=True)
```

[15]:
```
data2
```

[15]:

	Name	Python	Machine Learning	Age	AI	NLP	Cv
0	jay	90	NaN	23	90	NaN	23
1	raj	NaN	89	NaN	NaN	89	NaN
2	kumar	78	78	NaN	78	78	NaN
3	suraj	NaN	NaN	NaN	NaN	NaN	NaN
4	pawan	NaN	NaN	21	NaN	NaN	21
5	viraj	65	56	22	65	56	22
6	vijay	45	67	NaN	45	67	NaN
7	sumit	80	45	24	80	45	24
8	akash	70	67	NaN	70	67	NaN
9	sujit	NaN	NaN	22	NaN	NaN	22

[17]:
```
data2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
```

```
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Name              10 non-null     object
 1   Python            6 non-null      object
 2   Machine Learning  6 non-null      object
 3   Age               5 non-null      object
 4   AI                6 non-null      object
 5   NLP               6 non-null      object
 6   Cv                5 non-null      object
dtypes: object(7)
memory usage: 688.0+ bytes
```

[27]:
```
data2[["Python","Machine Learning","Age","AI","NLP","Cv"]]=(
    data2[["Python","Machine Learning","Age","AI","NLP","Cv"]]
    .astype("float64"))
```

[28]:
```
data2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Name              10 non-null     object
 1   Python            6 non-null      float64
 2   Machine Learning  6 non-null      float64
 3   Age               5 non-null      float64
 4   AI                6 non-null      float64
 5   NLP               6 non-null      float64
 6   Cv                5 non-null      float64
dtypes: float64(6), object(1)
memory usage: 688.0+ bytes
```

[30]:
```
Pythonmean,MLmean,Agemean,AImean,NLPmean,Cvmean=data2[["Python","Machine␣
↪Learning","Age","AI","NLP","Cv"]].mean()
```

[55]:
```
data2[["Python","Machine␣
↪Learning","Age","AI","NLP","Cv"]]=(data2[["Python","Machine␣
↪Learning","Age","AI","NLP","Cv"]]
                                                    .fillna({"Python":
↪Pythonmean,"Machine Learning": MLmean,
                                                             "Age":
↪Agemean,"AI":AImean,"NLP":NLPmean,"Cv":Cvmean}))
```

[56]:
```
data2
```

[56]:

	Name	Python	Machine Learning	Age	AI	NLP	Cv
0	jay	90.000000	67.0	23.0	90.000000	67.0	23.0
1	raj	71.333333	89.0	22.4	71.333333	89.0	22.4
2	kumar	78.000000	78.0	22.4	78.000000	78.0	22.4
3	suraj	71.333333	67.0	22.4	71.333333	67.0	22.4
4	pawan	71.333333	67.0	21.0	71.333333	67.0	21.0
5	viraj	65.000000	56.0	22.0	65.000000	56.0	22.0
6	vijay	45.000000	67.0	22.4	45.000000	67.0	22.4
7	sumit	80.000000	45.0	24.0	80.000000	45.0	24.0
8	akash	70.000000	67.0	22.4	70.000000	67.0	22.4
9	sujit	71.333333	67.0	22.0	71.333333	67.0	22.0

[54]:
```
data2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Name              10 non-null     object
 1   Python            10 non-null     float64
 2   Machine Learning  10 non-null     float64
 3   Age               10 non-null     float64
 4   AI                10 non-null     float64
 5   NLP               10 non-null     float64
 6   Cv                10 non-null     float64
dtypes: float64(6), object(1)
memory usage: 688.0+ bytes
```

[ ]:
```
```