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

In [2]:
```
dict1={"name":["ankit","manish",None,"aymen"],
       "age":[32,30,26,np.nan],
       "city":["Hdr",None,"Channei","Blr"]}
df=pd.DataFrame(dict1)
df
```

Out[2]:

	name	age	city
0	ankit	32.0	Hdr
1	manish	30.0	None
2	None	26.0	Channei
3	aymen	NaN	Blr

In [3]:
```
df.dtypes
```

Out[3]:
```
name     object
age     float64
city     object
dtype: object
```

In [4]:
```
dict1={"name":["mrityunjay","manish",None,"nitish"],
       "age":[32,30,"26",np.nan],
       "city":["Hdr",None,"Channei","Blr"]}
df1=pd.DataFrame(dict1)
df1
```

Out[4]:

	name	age	city
0	mrityunjay	32	Hdr
1	manish	30	None
2	None	26	Channei
3	nitish	NaN	Blr

In [5]:
```
df1.dtypes
```

Out[5]:
```
name    object
age     object
city    object
dtype: object
```

In [6]:
```
dict1={"name":["singh","manish",None,"aymen"],
       "age":[32,30,"26",26],
       "city":["Hdr",None,"Channei","Blr"]}
```

```
df2=pd.DataFrame(dict1)
df2
```

Out[6]:

	name	age	city
0	singh	32	Hdr
1	manish	30	None
2	None	26	Channei
3	aymen	26	Blr

In [7]:
```
df2["age"]=df2["age"].astype("int")
df2.dtypes
```

Out[7]:
```
name    object
age      int32
city    object
dtype: object
```

In [8]:
```
# handal missing value
dict1={"name":["ashish","rakesh",None,"manish"],
      "age":[32,30,"26",np.nan],
      "city":["Hdr",None,"Channei","Blr"]}
df1=pd.DataFrame(dict1)
df1.isnull().sum()
```

Out[8]:
```
name    1
age     1
city    1
dtype: int64
```

In [9]:
```
df1.isna().sum()
```

Out[9]:
```
name    1
age     1
city    1
dtype: int64
```

In [10]:
```
df1.isna().sum()/len(df1)
```

Out[10]:
```
name    0.25
age     0.25
city    0.25
dtype: float64
```

In [11]:
```
df1.isna().sum()*100/len(df1)
```

Out[11]:
```
name    25.0
age     25.0
city    25.0
dtype: float64
```

In [12]:
```
#fill the value
df.fillna(20)
```

Out[12]:

	name	age	city
0	ankit	32.0	Hdr
1	manish	30.0	20
2	20	26.0	Channei
3	aymen	20.0	Blr

In [13]:
```
df.fillna("Channei")
```

Out[13]:

	name	age	city
0	ankit	32.0	Hdr
1	manish	30.0	Channei
2	Channei	26.0	Channei
3	aymen	Channei	Blr

In [14]:
```
df["city"]=df["city"].fillna("panjab")
df
```

Out[14]:

	name	age	city
0	ankit	32.0	Hdr
1	manish	30.0	panjab
2	None	26.0	Channei
3	aymen	NaN	Blr

In [15]:
```
dict1={"name":["manish",None,"asish","aymen"],
       "age":[32,30,26,np.nan],
       "city":["Hdr",None,"Channei","Blr"]}
df=pd.DataFrame(dict1)
df
```

Out[15]:

	name	age	city
0	manish	32.0	Hdr
1	None	30.0	None
2	asish	26.0	Channei
3	aymen	NaN	Blr

In [16]:
```
# bfill, ffill, backfill, pad
df.fillna(method="bfill")
```

```
C:\Users\Mrityunjay\AppData\Local\Temp\ipykernel_10520\1595065152.py:2: FutureWarnin
g: DataFrame.fillna with 'method' is deprecated and will raise in a future version.
Use obj.ffill() or obj.bfill() instead.
  df.fillna(method="bfill")
```

Out[16]:

	name	age	city
0	manish	32.0	Hdr
1	asish	30.0	Channei
2	asish	26.0	Channei
3	aymen	NaN	Blr

In [17]:
```
df.fillna(method="backfill")
```

```
C:\Users\Mrityunjay\AppData\Local\Temp\ipykernel_10520\28332659.py:1: FutureWarning:
DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use
obj.ffill() or obj.bfill() instead.
  df.fillna(method="backfill")
```

Out[17]:

	name	age	city
0	manish	32.0	Hdr
1	asish	30.0	Channei
2	asish	26.0	Channei
3	aymen	NaN	Blr

In [18]:
```
df.fillna(method="ffill")
```

```
C:\Users\Mrityunjay\AppData\Local\Temp\ipykernel_10520\3944122520.py:1: FutureWarnin
g: DataFrame.fillna with 'method' is deprecated and will raise in a future version.
Use obj.ffill() or obj.bfill() instead.
  df.fillna(method="ffill")
```

Out[18]:

	name	age	city
0	manish	32.0	Hdr
1	manish	30.0	Hdr
2	asish	26.0	Channei
3	aymen	26.0	Blr

In [19]:
```
df.fillna(method="pad")
```

```
C:\Users\Mrityunjay\AppData\Local\Temp\ipykernel_10520\4167762737.py:1: FutureWarnin
g: DataFrame.fillna with 'method' is deprecated and will raise in a future version.
Use obj.ffill() or obj.bfill() instead.
  df.fillna(method="pad")
```

Out[19]:

	name	age	city
0	manish	32.0	Hdr
1	manish	30.0	Hdr
2	asish	26.0	Channei
3	aymen	26.0	Blr

In [20]:
```
import warnings
warnings.filterwarnings("ignore")
```

In [21]:
```
df.fillna(method="bfill")
```

Out[21]:

	name	age	city
0	manish	32.0	Hdr
1	asish	30.0	Channei
2	asish	26.0	Channei
3	aymen	NaN	Blr

In [22]:
```
df.fillna(method="ffill")
```

Out[22]:

	name	age	city
0	manish	32.0	Hdr
1	manish	30.0	Hdr
2	asish	26.0	Channei
3	aymen	26.0	Blr

In [23]:
```
df.fillna(method="backfill")
```

Out[23]:

	name	age	city
0	manish	32.0	Hdr
1	asish	30.0	Channei
2	asish	26.0	Channei
3	aymen	NaN	Blr

In [24]:
```
df.fillna(method="pad")
```

Out[24]:

	name	age	city
0	manish	32.0	Hdr
1	manish	30.0	Hdr
2	asish	26.0	Channei
3	aymen	26.0	Blr

In [25]:
```
df.fillna(method="bfill",axis=1)
```

Out[25]:

	name	age	city
0	manish	32.0	Hdr
1	30.0	30.0	NaN
2	asish	26.0	Channei
3	aymen	Blr	Blr

In [26]:
```
df.fillna(method="bfill",axis=0)       #by default
```

Out[26]:

	name	age	city
0	manish	32.0	Hdr
1	asish	30.0	Channei
2	asish	26.0	Channei
3	aymen	NaN	Blr

In [27]:
```
df.fillna(method="ffill")
```

Out[27]:

	name	age	city
0	manish	32.0	Hdr
1	manish	30.0	Hdr
2	asish	26.0	Channei
3	aymen	26.0	Blr

In [28]:
```
df
```

Out[28]:

	name	age	city
0	manish	32.0	Hdr
1	None	30.0	None
2	asish	26.0	Channei
3	aymen	NaN	Blr

KKN Imputer

In [30]:
```
from sklearn.impute import KNNImputer
k=KNNImputer()
df1["age"]=k.fit_transform(round(df1[["age"]],2))
df1
```

Out[30]:

	name	age	city
0	ashish	32.000000	Hdr
1	rakesh	30.000000	None
2	None	26.000000	Channei
3	manish	29.333333	Blr

In [31]:
```
#read the data frame
df=pd.read_csv("telecom_churn_data.csv")
df
```

Out[31]:

	year	customer_id	phone_no	gender	age	no_of_days_subscribed	multi_screen	ma
0	2015	100198	409-8743	Female	36	62	no	
1	2015	100643	340-5930	Female	39	149	no	
2	2015	100756	372-3750	Female	65	126	no	
3	2015	101595	331-4902	Female	24	131	no	
4	2015	101653	351-8398	Female	40	191	no	
...	...	...	...	...	...	...	...	
1995	2015	997132	385-7387	Female	54	75	no	
1996	2015	998086	383-9255	Male	45	127	no	
1997	2015	998474	353-2080	NaN	53	94	no	
1998	2015	998934	359-7788	Male	40	94	no	
1999	2015	999961	414-1496	Male	37	73	no	

2000 rows × 16 columns

In [32]:
```
#check the null columns
df.isnull().sum()
```

Out[32]:
```
year                      0
customer_id               0
phone_no                  0
gender                   24
age                       0
no_of_days_subscribed     0
multi_screen              0
mail_subscribed           0
weekly_mins_watched       0
minimum_daily_mins        0
maximum_daily_mins        0
weekly_max_night_mins     0
videos_watched            0
maximum_days_inactive    28
customer_support_calls    0
churn                    35
dtype: int64
```

In [33]:
```
#mode
g_mode=df["gender"].mode()
g_mode
```

Out[33]:
```
0    Male
Name: gender, dtype: object
```

In [34]:
```
#null value replace with values
```

```
df["gender"]=df["gender"].fillna(g_mode[0])
```

In [35]:
```
mean_=df["maximum_days_inactive"].mean()
mean_
```

Out[35]:
```
3.2505070993914806
```

In [36]:
```
#null value replace with values
df["maximum_days_inactive"]=df["maximum_days_inactive"].fillna(mean_)
```

In [37]:
```
c_mode=df["churn"].mode()
c_mode
```

Out[37]:
```
0    0.0
Name: churn, dtype: float64
```

In [38]:
```
#null value replace with values
df["churn"]=df["churn"].fillna(c_mode[0])
```

In [39]:
```
# After fill null value check it.
df.isnull().sum()
```

Out[39]:
```
year                      0
customer_id               0
phone_no                  0
gender                    0
age                       0
no_of_days_subscribed     0
multi_screen              0
mail_subscribed           0
weekly_mins_watched       0
minimum_daily_mins        0
maximum_daily_mins        0
weekly_max_night_mins     0
videos_watched            0
maximum_days_inactive     0
customer_support_calls    0
churn                     0
dtype: int64
```

In [ ]:
```

```