Pandas Assignment

```
Name : Kesavan R
Reg No : 19MIM10086
College : Vellore Institute of Technology
```

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

In [ ]:
```
#Dataset link : https://www.kaggle.com/kashnitsky/mlcourse?select=telecom_churn.c
```

In [2]:
```
df = pd.read_csv('telecom_churn.csv')
```

In [3]:
```
df
```

Out[3]:

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Tota ev call
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	9
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	10
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	11
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	8
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	12
...	...	...	...	...	...	...	...	...	...	...	.
3328	AZ	192	415	No	Yes	36	156.2	77	26.55	215.5	12
3329	WV	68	415	No	No	0	231.1	57	39.29	153.4	5
3330	RI	28	510	No	No	0	180.8	109	30.74	288.8	5
3331	CT	184	510	Yes	No	0	213.8	105	36.35	159.6	8
3332	TN	74	415	No	Yes	25	234.4	113	39.85	265.9	8

3333 rows × 20 columns

In [4]:
```
df.head()
```

Out[4]:

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122

In [5]:
```
df.tail()
```

Out[5]:

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Tota ev call
3328	AZ	192	415	No	Yes	36	156.2	77	26.55	215.5	12
3329	WV	68	415	No	No	0	231.1	57	39.29	153.4	5
3330	RI	28	510	No	No	0	180.8	109	30.74	288.8	5
3331	CT	184	510	Yes	No	0	213.8	105	36.35	159.6	8
3332	TN	74	415	No	Yes	25	234.4	113	39.85	265.9	8

In [6]:
```
df1 = df.head()
```

In [7]:
```
df1
```

Out[7]:

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122

In [8]:
```
df2 = df.tail()
```

In [10]:
```
mer = pd.merge(df1,df2,on="State")
```

In [11]:
```
mer
```

Out[11]:

State	Account length_x	Area code_x	International plan_x	Voice mail plan_x	Number vmail messages_x	Total day minutes_x	Total day calls_x	Total day charge_x	Tota minut

0 rows × 39 columns

In [16]:
```
df2 = pd.DataFrame({'col1':[2,1,2,3],'col2':[3,24,5,3],'col3':[20,45,23,23]})
```

In [17]:
```
df3 = pd.DataFrame({'col4':[2,13,2,3],'col5':[20,45,23,23],'col6': [3,24,5,3]})
```

In [18]:
```
join = df3.join(df2)
```

In [19]:
```
join
```

Out[19]:

	col4	col5	col6	col1	col2	col3
0	2	20	3	2	3	20
1	13	45	24	1	24	45
2	2	23	5	2	5	23
3	3	23	3	3	3	23

In [23]:
```
concat = pd.concat([join,df2])
```

In [24]:
```
concat
```

Out[24]:

	col4	col5	col6	col1	col2	col3
0	2.0	20.0	3.0	2	3	20
1	13.0	45.0	24.0	1	24	45
2	2.0	23.0	5.0	2	5	23
3	3.0	23.0	3.0	3	3	23
0	NaN	NaN	NaN	2	3	20
1	NaN	NaN	NaN	1	24	45
2	NaN	NaN	NaN	2	5	23
3	NaN	NaN	NaN	3	3	23

In [25]:
```
#Checkin the null values
```

In [26]:
```
df.isnull()
```

Out[26]:

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Tot ev cal
0	False	False	False	False	False	False	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	False	False	Fal
4	False	False	False	False	False	False	False	False	False	False	Fal
...	...	...	...	...	...	...	...	...	...	...	
3328	False	False	False	False	False	False	False	False	False	False	Fal
3329	False	False	False	False	False	False	False	False	False	False	Fal
3330	False	False	False	False	False	False	False	False	False	False	Fal
3331	False	False	False	False	False	False	False	False	False	False	Fal
3332	False	False	False	False	False	False	False	False	False	False	Fal

3333 rows × 20 columns

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

Out[27]:
```
State                     0
Account length            0
Area code                 0
International plan        0
Voice mail plan           0
Number vmail messages     0
Total day minutes         0
Total day calls           0
Total day charge          0
Total eve minutes         0
Total eve calls           0
Total eve charge          0
Total night minutes       0
Total night calls         0
Total night charge        0
Total intl minutes        0
Total intl calls          0
Total intl charge         0
Customer service calls    0
Churn                     0
dtype: int64
```

In [29]:
```
df.describe()
```

Out[29]:

	Account length	Area code	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes
count	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000
mean	101.064806	437.182418	8.099010	179.775098	100.435644	30.562307	200.980348
std	39.822106	42.371290	13.688365	54.467389	20.069084	9.259435	50.713844
min	1.000000	408.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	74.000000	408.000000	0.000000	143.700000	87.000000	24.430000	166.600000
50%	101.000000	415.000000	0.000000	179.400000	101.000000	30.500000	201.400000
75%	127.000000	510.000000	20.000000	216.400000	114.000000	36.790000	235.300000
max	243.000000	510.000000	51.000000	350.800000	165.000000	59.640000	363.700000

In [30]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3333 entries, 0 to 3332
Data columns (total 20 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   State                   3333 non-null   object 
 1   Account length          3333 non-null   int64  
 2   Area code               3333 non-null   int64  
 3   International plan      3333 non-null   object 
 4   Voice mail plan         3333 non-null   object 
 5   Number vmail messages   3333 non-null   int64  
 6   Total day minutes       3333 non-null   float64
 7   Total day calls         3333 non-null   int64  
 8   Total day charge        3333 non-null   float64
 9   Total eve minutes       3333 non-null   float64
 10  Total eve calls         3333 non-null   int64  
 11  Total eve charge        3333 non-null   float64
 12  Total night minutes     3333 non-null   float64
 13  Total night calls       3333 non-null   int64  
 14  Total night charge      3333 non-null   float64
 15  Total intl minutes      3333 non-null   float64
 16  Total intl calls        3333 non-null   int64  
 17  Total intl charge       3333 non-null   float64
 18  Customer service calls  3333 non-null   int64  
 19  Churn                   3333 non-null   bool   
dtypes: bool(1), float64(8), int64(8), object(3)
memory usage: 498.1+ KB
```

In [31]:
```
df.columns
```

Out[31]:
```
Index(['State', 'Account length', 'Area code', 'International plan',
       'Voice mail plan', 'Number vmail messages', 'Total day minutes',
       'Total day calls', 'Total day charge', 'Total eve minutes',
       'Total eve calls', 'Total eve charge', 'Total night minutes',
       'Total night calls', 'Total night charge', 'Total intl minutes',
       'Total intl calls', 'Total intl charge', 'Customer service calls',
       'Churn'],
      dtype='object')
```

In [32]:
```
df.columns
```

Out[32]:
```
Index(['State', 'Account length', 'Area code', 'International plan',
       'Voice mail plan', 'Number vmail messages', 'Total day minutes',
       'Total day calls', 'Total day charge', 'Total eve minutes',
       'Total eve calls', 'Total eve charge', 'Total night minutes',
       'Total night calls', 'Total night charge', 'Total intl minutes',
       'Total intl calls', 'Total intl charge', 'Customer service calls',
       'Churn'],
      dtype='object')
```

In [33]:
```
df.loc[2:6]
```

Out[33]:

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122
5	AL	118	510	Yes	No	0	223.4	98	37.98	220.6	101
6	MA	121	510	No	Yes	24	218.2	88	37.09	348.5	108

In [34]:
```
df.iloc[6]
```

Out[34]:
```
State                     MA
Account length           121
Area code                510
International plan        No
Voice mail plan          Yes
Number vmail messages     24
Total day minutes       218.2
Total day calls            88
Total day charge        37.09
Total eve minutes       348.5
Total eve calls           108
Total eve charge        29.62
Total night minutes     212.6
Total night calls         118
Total night charge       9.57
Total intl minutes        7.5
Total intl calls            7
Total intl charge        2.03
Customer service calls      3
Churn                   False
Name: 6, dtype: object
```

In [35]:
```
df.set_index('State',inplace=True)
```

In [36]:
```
df
```

Out[36]:

State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Tc cha
KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99	16
OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16
NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10
OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5
OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122	12
...	...	...	...	...	...	...	...	...	...	...	
AZ	192	415	No	Yes	36	156.2	77	26.55	215.5	126	18
WV	68	415	No	No	0	231.1	57	39.29	153.4	55	13
RI	28	510	No	No	0	180.8	109	30.74	288.8	58	24
CT	184	510	Yes	No	0	213.8	105	36.35	159.6	84	13
TN	74	415	No	Yes	25	234.4	113	39.85	265.9	82	22

3333 rows × 19 columns

In [37]:
```
df = df.rename(columns={"International plan":"world plan"})
```

In [38]:
```
df
```

Out[38]:

	Account length	Area code	world plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Tota nigh minute
ıte												
KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99	16.78	244.
)H	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16.62	254.
NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10.30	162.
)H	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5.26	196.
)K	75	415	Yes	No	0	166.7	113	28.34	148.3	122	12.61	186.
...	...	...	...	...	...	...	...	...	...	...	...	.
AZ	192	415	No	Yes	36	156.2	77	26.55	215.5	126	18.32	279.
/V	68	415	No	No	0	231.1	57	39.29	153.4	55	13.04	191.
RI	28	510	No	No	0	180.8	109	30.74	288.8	58	24.55	191.
CT	184	510	Yes	No	0	213.8	105	36.35	159.6	84	13.57	139.
TN	74	415	No	Yes	25	234.4	113	39.85	265.9	82	22.60	241.

3 rows × 19 columns

In [41]:
```
df.shape
```

Out[41]:
```
(3333, 19)
```

In [47]:
```
df['Area code']=df['Area code'].fillna(df['Area code'].mean())
```

In [48]:
```
df['Account length']=df['Account length'].fillna(df['Account length'].mean())
```

In [ ]:

In [ ]: