

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [108]:

```
1 import pandas as pd
2 data = pd.read_csv("Telco Customer Churn.csv")
3
4 print("Display all first of 5 rows :")
5 display(data.head())
6 print("The shape of data in (nrows,ncols)")
7 print(data.shape)
```

Display all first of 5 rows :

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	... DeviceProtection	TechSupp
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No

5 rows × 21 columns

A/C ENG 2:51 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

The shape of data in (nrows,ncols)
(7043, 21)

Declare required libraries

```
In [109]: 1 import numpy as np  
2 import pandas as pd  
3  
4 import matplotlib.pyplot as plt  
5 import seaborn as sns
```

Show data

```
In [110]: 1 print("How to extract Index of Dataframe ? \n\t", data.index)  
2 print("\nHow to extract Column of Dataframe in the list type ? \n\t", list(data.columns))  
3 print(f"\nWhat is min max value of MonthlyCharges in the data ? \n\t \\\n\t from min : {np.min(data['MonthlyCharges'])} to max : {np.max(data['MonthlyCharges'])}")  
4 print(f"\nWhat is sum and median value of tenure in the data ? \n\t \\\n\t sum : {data['tenure'].sum()} median : {data['tenure'].median()}")  
5 print(f"\nWhat is sum and median value of TotalCharges in the data ? \n\t \\\n\t sum : {data['MonthlyCharges'].sum()} median : {data['MonthlyCharges'].median()} std : {data['MonthlyCharges'].std()}")  
6  
7 How to extract Index of Dataframe ?  
RangeIndex(start=0, stop=7043, step=1)  
  
How to extract Column of Dataframe in the list type ?  
['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn']  
  
What is min max value of MonthlyCharges in the data ?
```

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

What is min max value of MonthlyCharges in the data ?
from min : 18.25 to max : 118.75

What is sum and median value of tenure in the data ?
sum : 227990 median : 29.0

What is sum and median value of TotalCharges in the data ?
sum : 456116.6 median : 70.35 std : 30.09004709767854

In [111]:

```
1 print("How many cases of Churn ?")
2 print(data["Churn"].value_counts())
3 print("\n")
4
5 print("Can we see the statistics table of the whole data ?")
6 display(data.describe())
7 print("\n")
8
9 print("Is there any missing value at all columns ?")
10 display(data.isnull().sum())
11 print("\n")
```

How many cases of Churn ?
No 5174
Yes 1869
Name: Churn, dtype: int64

Can we see the statistics table of the whole data ?

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Can we see the statistics table of the whole data ?

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

Is there any missing value at all columns ?

customerID	0
gender	0
SeniorCitizen	0
Partner	0
Dependents	0
tenure	0
PhoneService	0
MultipleLines	0
InternetService	0
OnlineSecurity	0
OnlineBackup	0
DeviceProtection	0
TechSupport	0
StreamingTV	0

A/C ENG 2:53 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Is there any missing value at all columns ?

```
customerID      0
gender          0
SeniorCitizen   0
Partner         0
Dependents     0
tenure          0
PhoneService    0
MultipleLines   0
InternetService 0
OnlineSecurity  0
OnlineBackup    0
DeviceProtection 0
TechSupport     0
StreamingTV     0
StreamingMovies 0
Contract        0
PaperlessBilling 0
PaymentMethod   0
MonthlyCharges  0
Totalcharges    0
Churn           0
dtype: int64
```

In [112]:

```
1 print("How many customer use PhoneService in data")
2 print(data["PhoneService"].value_counts())
3 print("\n")
4
5 print("How we can get only the rows from index 100 to 105 ?")
6 display(data.loc[100:105, :])
7 print("\n")
```

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

In [112]:

```
1 print("How many customer use PhoneService in data")
2 print(data["PhoneService"].value_counts())
3 print("\n")
4
5 print("How we can get only the rows from index 100 to 105 ?")
6 display(data.loc[100:105, :])
7 print("\n")
8
9 print("How we can reset index of the above results in a new table ? ")
10 df = data.loc[10:15, :]
11 df = df.reset_index(drop = True)
12 display(df)
13 print("\n")
14
15 print("How we can get the rows from index 100 to 105 of columns customerID, MonthlyCharges, TotalCharges and Churn ?")
16 display(data.loc[10:15, ["customerID", "MonthlyCharges", "TotalCharges", "Churn"]])
17 print("\n")
18
19 print("How we can get only the rows from index 110 of columns customerID, MonthlyCharges, TotalCharges and Churn ?")
20 display(data.loc[110, ["customerID", "MonthlyCharges", "TotalCharges", "Churn"]])
```

How many customer use PhoneService in data

PhoneService	Count
No	682
Yes	6361

Name: PhoneService, dtype: int64

How we can get only the rows from index 100 to 105 ?

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	Churn
100	Male	0	No	No	1	Yes	No	No	No	No	No	No	No	One year	Yes	Bank transfer (automatic)	Yes
101	Female	0	Yes	Yes	1	Yes	No	No	No	No	No	No	No	One year	Yes	Credit card (automatic)	Yes

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

How we can get only the rows from index 100 to 105 ?

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport
100	6380-ARCEH	Male	0	No	No	1	Yes	No	No	No internet service	...	No internet service	No in s
101	3679-XASPY	Female	0	Yes	Yes	1	Yes	No	No	No internet service	...	No internet service	No in s
102	7123-WQUHX	Male	0	No	No	38	Yes	Yes	Fiber optic	No	Yes
103	5386-THSLQ	Female	1	Yes	No	66	No	No phone service	DSL	No	Yes
104	3192-NQECA	Male	0	Yes	No	68	Yes	Yes	Fiber optic	No	Yes
105	6180-YBIQI	Male	0	No	No	5	No	No phone service	DSL	No	No

6 rows × 21 columns

How we can reset index of the above results in a new table ?

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport
0	9763-GRSKD	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No
1	7469-LKBCI	Male	0	No	No	16	Yes	No	No	No internet service	...	No internet service	No inter ser
2	8091-TTVAX	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	Yes
3	0280-XJGEX	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes

A/C ENG 2:54 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

How we can reset index of the above results in a new table ?

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	... DeviceProtection	TechSupp
0	9763-GRSKD	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	...	No
1	7469-LKBCI	Male	0	No	No	16	Yes	No	No	No internet service	...	No internet service
2	8091-TTVAX	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	...	Yes
3	0280-XJGEX	Male	0	No	No	49	Yes	Yes	Fiber optic	No	...	Yes
4	5129-JLPIS	Male	0	No	No	25	Yes	No	Fiber optic	Yes	...	Yes
5	3655-SNQYZ	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	...	Yes

6 rows x 21 columns

How we can get the rows from index 100 to 105 of columns customerID, MonthlyCharges, TotalCharges and Churn ?

	customerID	MonthlyCharges	TotalCharges	Churn
10	9763-GRSKD	49.95	587.45	No
11	7469-LKBCI	18.95	326.8	No
12	8091-TTVAX	100.35	5681.1	No
13	0280-XJGEX	103.70	5036.3	Yes
14	5129-JLPIS	105.50	2686.05	No

localhost:8888/notebooks/TNHien/Khai%20thac%20d%20li%20lu%20/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

How we can get the rows from index 100 to 105 of columns customerID, MonthlyCharges, TotalCharges and Churn ?

	customerID	MonthlyCharges	TotalCharges	Churn
10	9763-GRSKD	49.95	587.45	No
11	7469-LKBCI	18.95	326.8	No
12	8091-TTVAX	100.35	5681.1	No
13	0280-XJGEX	103.70	5036.3	Yes
14	5129-JLPIS	105.50	2686.05	No
15	3655-SNQYZ	113.25	7895.15	No

How we can get only the rows from index 110 of columns customerID, MonthlyCharges, TotalCharges and Churn ?

```
customerID      0486-HECZI
MonthlyCharges    96.75
TotalCharges      5238.9
Churn            Yes
Name: 110, dtype: object
```

In [113]:

```
1 print("How we can see the type of all columns in data ?")
2 display(data.info())
3 print("\n")
4
5 print("How we can change TotalCharges from object to float64")
6 print("Original Type of TotalCharges :", data["TotalCharges"].dtypes)
7 data["TotalCharges"] = pd.to_numeric(data['TotalCharges'], errors='coerce')
8 print("New Type of TotalCharges :", data["TotalCharges"].dtypes)
9 print("\n")
10
11 print("Is there any missing value at all columns ?")
```

The screenshot shows a Jupyter Notebook interface running on a Windows operating system. The notebook is titled "19110315_TrinhNgocHien_DM-Lab01.ipynb". The code in the cell (In [113]) is as follows:

```
1 print("How we can see the type of all columns in data ?")
2 display(data.info())
3 print("\n")
4
5 print("How we can change TotalCharges from object to float64")
6 print("Original Type of TotalCharges : ", data["TotalCharges"].dtypes)
7 data["TotalCharges"] = pd.to_numeric(data['TotalCharges'], errors='coerce')
8 print("New Type of TotalCharges : ", data["TotalCharges"].dtypes)
9 print("\n")
10
11 print("Is there any missing value at all columns ?")
12 display(data.isnull().sum())
13 print("\n")
14
15 print("How we can extract the categorical and numeric columns ?")
16 CatFeatures = [col for col in data.columns if data[col].dtypes in ["object", "bool"]]
17 NumFeatures = [col for col in data.columns if data[col].dtypes in ["int64", "float64"]]
18 print("Categorical Features :\n", CatFeatures)
19 print("Numeric Features :\n", NumFeatures)
20 print("\n")
21
22 print("How we can show the all statistics of Numeric Features ?")
23 display(data.describe())
24 print("\n")
25
26 print("How we can show the all statistics of Categorical Features ?")
27 display(data[CatFeatures].describe(include='all'))
28 print("\n")
29
30 print("How we can get data from describe table ?")
31 NumStats = data[NumFeatures].describe(include='all')
32 CatStats = data[CatFeatures].describe(include='all')
33 min_max_TotalCharges = NumStats.loc[['min', 'max'], "TotalCharges"]
34 Contract_top_freq = CatStats.loc[['top', 'freq'], "Contract"]
35 print("Min and Max of TotalCharges: \n", min_max_TotalCharges)
36 print("\n")
```

The browser tab bar shows various open tabs including Facebook, YouTube, Python, Lab01 - DataFrame Tutorials, Google search results, and the current Jupyter notebook. The taskbar at the bottom includes icons for File Explorer, Task View, Start, and other system functions.

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
30 print("How we can get data from describe table ?")
31 NumStats = data[NumFeatures].describe(include='all')
32 CatStats = data[CatFeatures].describe(include='all')
33 min_max_TotalCharges = NumStats.loc[['min','max'], "TotalCharges"]
34 Contract_top_freq = CatStats.loc[['top', "freq"], "Contract"]
35 print("Min and Max of TotalCharges: \n", min_max_TotalCharges)
36 print("\n")
37 print("Top and Frequency of Contract: \n", Contract_top_freq)
```

How we can see the type of all columns in data ?
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):

#	Column	Non-Null Count	Dtype
0	customerID	7043	non-null object
1	gender	7043	non-null object
2	SeniorCitizen	7043	non-null int64
3	Partner	7043	non-null object
4	Dependents	7043	non-null object
5	tenure	7043	non-null int64
6	PhoneService	7043	non-null object
7	MultipleLines	7043	non-null object
8	InternetService	7043	non-null object
9	OnlineSecurity	7043	non-null object
10	OnlineBackup	7043	non-null object
11	DeviceProtection	7043	non-null object
12	TechSupport	7043	non-null object
13	StreamingTV	7043	non-null object
14	StreamingMovies	7043	non-null object
15	Contract	7043	non-null object
16	PaperlessBilling	7043	non-null object
17	PaymentMethod	7043	non-null object
18	MonthlyCharges	7043	non-null float64
19	TotalCharges	7043	non-null object
20	Churn	7043	non-null object

A/C ENG 2:55 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

20 Churn 7043 non-null object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB

None

How we can change TotalCharges from object to float64
Original Type of TotalCharges : object
New Type of TotalCharges : float64

Is there any missing value at all columns ?

customerID	0
gender	0
SeniorCitizen	0
Partner	0
Dependents	0
tenure	0
PhoneService	0
Multiplelines	0
InternetService	0
OnlineSecurity	0
OnlineBackup	0
DeviceProtection	0
TechSupport	0
StreamingTV	0
StreamingMovies	0
Contract	0
PaperlessBilling	0
PaymentMethod	0
MonthlyCharges	0
Totalcharges	11
Churn	0
dtype:	int64

A/C ENG 2:57 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

How we can extract the categorical and numeric columns ?

Categorical Features :

```
[ 'customerID', 'gender', 'Partner', 'Dependents', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'Churn' ]
```

Numeric Features :

```
[ 'SeniorCitizen', 'tenure', 'MonthlyCharges', 'TotalCharges' ]
```

How we can show the all statistics of Numeric Features ?

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges
count	7043.000000	7043.000000	7043.000000	7032.000000
mean	0.162147	32.371149	64.761692	2283.300441
std	0.369612	24.559481	30.090047	2266.771362
min	0.000000	0.000000	18.250000	18.800000
25%	0.000000	9.000000	35.500000	401.450000
50%	0.000000	29.000000	70.350000	1397.475000
75%	0.000000	55.000000	89.850000	3794.737500
max	1.000000	72.000000	118.750000	8684.800000

How we can show the all statistics of Categorical Features ?

	customerID	gender	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	Contract	PaperlessBilling	PaymentMethod	Churn
count	7043	7043	7043	7043	7043	7043	7043	7043	7043	7043	7043	7043	7043	7043	
unique	7043	2	2	2	2	3	3	3	3	3	3	3	3	3	
top	6834-NYDCA	Male	No	No	Yes	No	Fiber optic	No	No	No	No	No	No	No	

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

How we can show the all statistics of Categorical Features ?

	customerID	gender	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	TotalCharges
count	7043	7043	7043	7043	7043	7043	7043	7043	7043	7043	7043	7043
unique	7043	2	2	2	2	3	3	3	3	3	3	3
top	6834-NXDCA	Male	No	No	Yes	No	Fiber optic	No	No	No	No	No
freq	1	3555	3641	4933	6361	3390	3096	3498	3088	3095	3473	3473

How we can get data from describe table ?
Min and Max of TotalCharges:
min 18.8
max 8684.8
Name: TotalCharges, dtype: float64

Top and Frequency of Contract:
top Month-to-month
freq 3875
Name: Contract, dtype: object

Fix missing data

```
In [114]: 1 #Find the index of missing data
2 print("The index of missing data are: ",np.where(data['TotalCharges'].isnull())[0])
3 print("The index 487: ",data['TotalCharges'][487], "\nThe index 489: ",data['TotalCharges'][489])
4 new_data = data
```

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [26]:

```
1 #Find the index of missing data
2 print("The index of missing data are: ",np.where(data['TotalCharges'].isnull())[0])
3 print("The data of the index 488 is: ",data['TotalCharges'][488])
4 print("The index 487: ",data['TotalCharges'][487], "\nThe index 489: ",data['TotalCharges'][489])
5 new_data = data
6
7 #Fix the missing data with the mean of between data
8 new_data = new_data.interpolate()
9 print("The data of the index 488 is: ",new_data['TotalCharges'][488])
10 print("Is there any missing value at all columns in new_data ?")
11 display(new_data.isnull().sum())
12 print("\n")
13
14
15 #Assign new data to old data
16 data = new_data
17
```

The index of missing data are: [488 753 936 1082 1340 3331 3826 4380 5218 6670 6754]
The data of the index 488 is: nan
The index 487: 6201.95
The index 489: 74.35
The data of the index 488 is: 3138.15
Is there any missing value at all columns in new_data ?

customerID	0
gender	0
SeniorCitizen	0
Partner	0
Dependents	0
tenure	0
PhoneService	0

A/C ENG 3:03 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

The index 489: 74.35
The data of the index 488 is: 3138.15
Is there any missing value at all columns in new_data ?

```
customerID      0
gender          0
SeniorCitizen   0
Partner         0
Dependents     0
tenure          0
PhoneService    0
MultipleLines   0
InternetService 0
OnlineSecurity  0
OnlineBackup    0
DeviceProtection 0
TechSupport     0
StreamingTV     0
StreamingMovies 0
Contract        0
PaperlessBilling 0
PaymentMethod   0
MonthlyCharges  0
TotalCharges    0
Churn           0
dtype: int64
```

In [27]:

```
1 print("How we can draw chart for a numeric features ?")
2 feature = "MonthlyCharges"
3 f, (ax_box, ax_hist) = plt.subplots(2, sharex=True, gridspec_kw={"height_ratios": (.15, .85)})
4 f.set_figheight(3)
5 f.set_figwidth(15)
6 sns.boxplot(data[feature], ax=ax_box)
7 sns.histplot(data=data, x=feature, ax=ax_hist)
```

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

In [27]:

```
1 print("How we can draw chart for a numeric features ?")
2 feature = "MonthlyCharges"
3 f, (ax_box, ax_hist) = plt.subplots(2, sharex=True, gridspec_kw={"height_ratios": (.15, .85)})
4 f.set_figheight(3)
5 f.set_figwidth(15)
6 sns.boxplot(data[feature], ax=ax_box)
7 sns.histplot(data=data, x=feature, ax=ax_hist)
8 plt.grid()
9 plt.show()
10
11 print("How we can map Yes/No to True/False in Churn feature ?")
12 MapDict = {"Yes" : True, "No": False}
13 data["Churn_Or_Not"] = data["Churn"].map(MapDict)
14 display(data.head())
15
16 print("How we can draw chart for a numeric feature according to a categorical feature ?")
17 feature = "MonthlyCharges"
18 plt.figure(figsize = (15,3))
19 sns.boxplot(y ='Churn', x = feature, data = data)
20 plt.title(feature)
21 plt.grid()
22 plt.show()
23
24 print("How we can draw chart for two numeric features according to a categorical feature ?")
25 plt.figure(figsize=(15,5))
26 feature_x = "tenure"
27 feature_y = "MonthlyCharges"
28 feature_hue = "Churn"
29 sns.scatterplot(x = feature_x, y= feature_y, hue=feature_hue, data = data, legend='full')
30 plt.grid()
31 plt.show()
```

How we can draw chart for a numeric features ?

C:\Users\Rony\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword argument x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit key word will result in an error or misinterpretation

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

How we can draw chart for a numeric features ?

```
C:\Users\Rony\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword argument: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

warnings.warn(

How we can map Yes/No to True/False in Churn feature ?

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	TechSupport	StreamingTV
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	No
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	Yes
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

5 rows x 22 columns

How we can draw chart for a numeric feature according to a categorical feature ?

MonthlyCharges

A box plot titled "MonthlyCharges" comparing the distribution of monthly charges between customers who have left ("Yes") and those who have not ("No"). The x-axis represents "MonthlyCharges" from 20 to 120. The y-axis represents "Churn" with categories "No" and "Yes". The "No" group (blue) has a median around 70, while the "Yes" group (orange) has a median around 60. Both groups show significant overlap in their charge distributions.

How we can draw chart for two numeric features according to a categorical feature ?

A scatter plot showing the relationship between "MonthlyCharges" (Y-axis, 20-120) and "TotalCharges" (X-axis, 20-120) for two categories of customers based on "Churn": "No" (blue dots) and "Yes" (orange dots). The plot shows that customers who have left ("Yes") tend to have higher total charges than those who have not ("No"), particularly at the higher end of the monthly charges scale.

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 Logout

MonthlyCharges

How we can draw chart for two numeric features according to a categorical feature ?

In [28]:

```
1 print("How we can draw chart for a categorical feature ?")
2 feature = "Churn"
3 plt.figure(figsize=(15,5))
4 plt.subplot(1,2,1)
5 data[feature].value_counts().plot.pie(autopct='%1.1f%%')
6 plt.subplot(1,2,2)
7 sns.countplot(data[feature])
8 plt.show()
9
10 print("How we can draw chart for a categorical feature according to another categorical feature ?")
11 plt.figure(figsize=(15,5))
12 feature_x = "PhoneService"
13 feature_y = "MonthlyCharges"
```

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

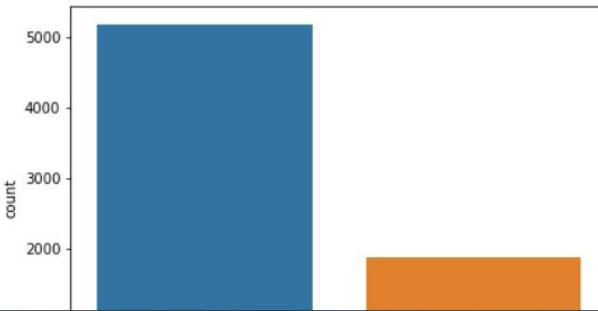
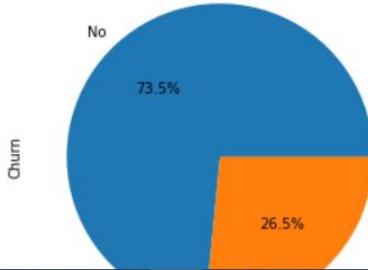
jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

In [28]:

```
1 print("How we can draw chart for a categorical feature ?")
2 feature = "Churn"
3 plt.figure(figsize=(15,5))
4 plt.subplot(1,2,1)
5 data[feature].value_counts().plot.pie(autopct='%1.1f%%')
6 plt.subplot(1,2,2)
7 sns.countplot(data[feature])
8 plt.show()
9
10 print("How we can draw chart for a categorical feature according to another categorical feature ?")
11 plt.figure(figsize=(15,5))
12 feature_x = "PhoneService"
13 feature_y = "MonthlyCharges"
14 sns.stripplot(data[feature_x],data[feature_y])
15 plt.show()
```

How we can draw chart for a categorical feature ?

C:\Users\Rony\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword argument: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(



localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

15 plt.show()

How we can draw chart for a categorical feature ?

```
C:\Users\Rony\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword argument x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.  
warnings.warn(
```

Churn Category	Percentage
No	73.5%
Yes	26.5%

Churn Category	Count
No	5100
Yes	1900

How we can draw chart for a categorical feature according to another categorical feature ?

```
C:\Users\Rony\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword arguments x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.  
warnings.warn(
```

120

Category	Count
No	5100
Yes	1900

A/C

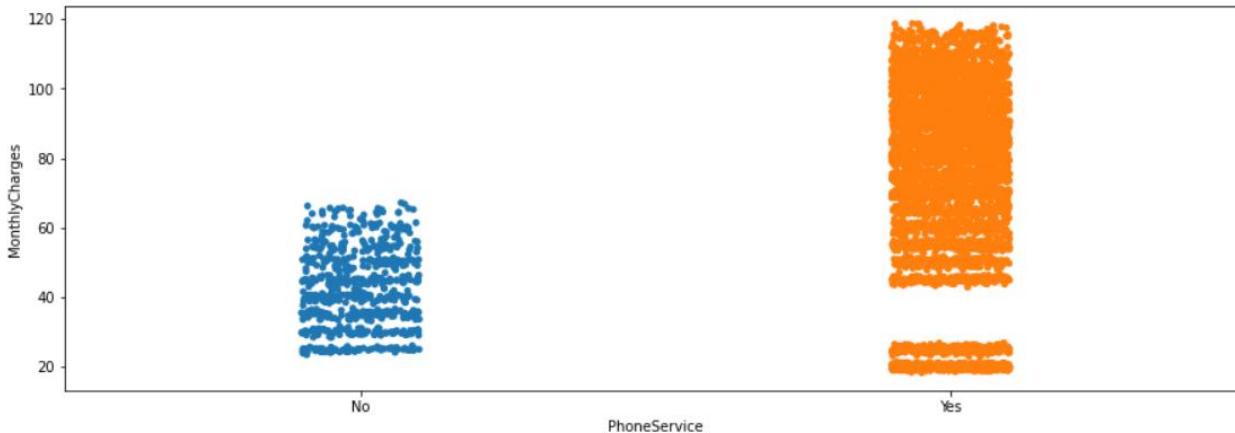
localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 Logout

How we can draw chart for a categorical feature according to another categorical feature ?

```
C:\Users\Rony\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(
```



In [29]: 1 data.columns

```
Out[29]: Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'Totalcharges', 'Churn', 'Churn_Or_Not'], dtype='object')
```

A/C ENG 3:06 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

In [30]:

```
1 print("How we can split data into many data ? ")
2 feature_Customer = ['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'tenure', 'Churn']
3 data_Customer = data[feature_Customer].copy()
4 display(data_Customer.head())
5
6 feature_Service = ['PhoneService', 'MultipleLines', 'InternetService', "OnlineSecurity", "OnlineBackup", "DeviceProtection",
7 data_Service = data[feature_Service].copy()
8 display(data_Service.head())
9
10 feature_Fee = list(set(data.columns) - set(feature_Customer) - set(feature_Service))
11 data_Fee = data[feature_Fee].copy()
12 display(data_Fee.head())
13
14 print("How we can merge two data into one by cols ? ")
15 data_Customer_Service = pd.concat([data_Customer, data_Service], axis = 1)
16 display(data_Customer_Service.head())
```

How we can split data into many data ?

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	Churn
0	7590-VHVEG	Female	0	Yes	No	1	No
1	5575-GNVDE	Male	0	No	No	34	No
2	3668-QPYBK	Male	0	No	No	2	Yes
3	7795-CFOCW	Male	0	No	No	45	No
4	9237-HQITU	Female	0	No	No	2	Yes

	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies
0	No	No phone service	DSL	No	Yes	No	No	No	No
1	Yes	No	DSL	Yes	No	Yes	No	No	No

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

PhoneService MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection TechSupport StreamingTV StreamingMovies

	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies
0	No	No phone service	DSL	No	Yes	No	No	No	No
1	Yes	No	DSL	Yes	No	Yes	No	No	No
2	Yes	No	DSL	Yes	Yes	No	No	No	No
3	No	No phone service	DSL	Yes	No	Yes	Yes	No	No
4	Yes	No	Fiber optic	No	No	No	No	No	No

MonthlyCharges TotalCharges Churn_Or_Not Contract PaperlessBilling PaymentMethod

	MonthlyCharges	TotalCharges	Churn_Or_Not	Contract	PaperlessBilling	PaymentMethod
0	29.85	29.85	False	Month-to-month	Yes	Electronic check
1	56.95	1889.50	False	One year	No	Mailed check
2	53.85	108.15	True	Month-to-month	Yes	Mailed check
3	42.30	1840.75	False	One year	No	Bank transfer (automatic)
4	70.70	151.65	True	Month-to-month	Yes	Electronic check

How we can merge two data into one by cols ?

customerID gender SeniorCitizen Partner Dependents tenure Churn PhoneService MultipleLines InternetService OnlineSecurity OnlineBackup DeviceP

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	Churn	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceP
0	7590-VHVEG	Female	0	Yes	No	1	No	No	No phone service	DSL	No	Yes	
1	5575-GNVDE	Male	0	No	No	34	No	Yes	No	DSL	Yes	No	
2	3668-QPYBK	Male	0	No	No	2	Yes	Yes	No	DSL	Yes	Yes	
3	7795-CFOCW	Male	0	No	No	45	No	No	No phone service	DSL	Yes	No	
4	9237-HQITU	Female	0	No	No	2	Yes	Yes	No	Fiber optic	No	No	

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

How we can merge two data into one by cols ?

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	Churn	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceP
0	7590-VHVEG	Female	0	Yes	No	1	No	No	No phone service	DSL	No	Yes	
1	5575-GNVDE	Male	0	No	No	34	No	Yes	No	DSL	Yes	No	
2	3668-QPYBK	Male	0	No	No	2	Yes	Yes	No	DSL	Yes	Yes	
3	7795-CFOCW	Male	0	No	No	45	No	No	No phone service	DSL	Yes	No	
4	9237-HQITU	Female	0	No	No	2	Yes	Yes	No	Fiber optic	No	No	

```
In [31]: 1 print("How we can filter data by condition ?")
2 print("Filter the data of SeniorCitizen > 0 and PhoneService is 'Yes'")
3 Condition1 = data["SeniorCitizen"] > 0
4 Condition2 = data["PhoneService"] == "Yes"
5 data_over100_PhoneService1 = data[Condition1 & Condition2].copy()
6 display(data_over100_PhoneService1.head())
7 print(data_over100_PhoneService1.shape)
8
9 print("Filter the data of SeniorCitizen < 0 and PhoneService is 'Yes'")
10 value1, value2 = 1 , "Yes"
11 data_less1_PhoneService2 = data.query(`SeniorCitizen` < @value1 and `PhoneService` == @value2)
12 display(data_less1_PhoneService2.head())
13 print(data_less1_PhoneService2.shape)
14
15 print("How we can merge two data into one by rows ?")
16 data_merge = pd.concat([data_over100_PhoneService1, data_less1_PhoneService2])
17 display(data_merge)
18 print(data_merge.shape)
```

A/C ENG 3:09 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved) Logout

In [31]:

```
1 print("How we can filter data by condition ?")
2 print("Filter the data of SeniorCitizen > 0 and PhoneService is 'Yes'")
3 Condition1 = data["SeniorCitizen"] > 0
4 Condition2 = data["PhoneService"] == "Yes"
5 data_over100_PhoneService1 = data[Condition1 & Condition2].copy()
6 display(data_over100_PhoneService1.head())
7 print(data_over100_PhoneService1.shape)
8
9 print("Filter the data of SeniorCitizen < 0 and PhoneService is 'Yes'")
10 value1, value2 = 1 , "Yes"
11 data_less1_PhoneService2 = data.query(`SeniorCitizen` < @value1 and `PhoneService` == @value2)
12 display(data_less1_PhoneService2.head())
13 print(data_less1_PhoneService2.shape)
14
15 print("How we can merge two data into one by rows ?")
16 data_merge = pd.concat([data_over100_PhoneService1, data_less1_PhoneService2])
17 display(data_merge)
18 print(data_merge.shape)
```

How we can filter data by condition ?
Filter the data of SeniorCitizen > 0 and PhoneService is 'Yes'

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	TechSupport	StreamingTV		
30	3841-NFECX	Female		1	Yes	No	71	Yes	Yes	Fiber optic	Yes	...	Yes	No
31	4929-XIHWV	Male		1	Yes	No	2	Yes	No	Fiber optic	No	...	No	Yes
34	3413-BMNZE	Male		1	No	No	1	Yes	No	DSL	No	...	No	No
50	8012-SOUDQ	Female		1	No	No	43	Yes	Yes	Fiber optic	No	...	No	Yes
52	6575-SUVOI	Female		1	Yes	No	25	Yes	Yes	DSL	Yes	...	Yes	Yes

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

5 rows × 22 columns

(1038, 22)
Filter the data of SeniorCitizen < 0 and PhoneService is 'Yes'

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	TechSupport	StreamingTV	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	No	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No
5	9305-CDSKC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	...	No	Yes
6	1452-KIOVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	...	No	Yes

5 rows × 22 columns

(5323, 22)
How we can merge two data into one by rows ?

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	TechSupport	StreamingTV	
30	3841-NFECX	Female	1	Yes	No	71	Yes	Yes	Fiber optic	Yes	...	Yes	Y
31	4929-XIHVW	Male	1	Yes	No	2	Yes	No	Fiber optic	No	...	No	Y
34	3413-BMNZE	Male	1	No	No	1	Yes	No	DSL	No	...	No	Y

jupyter 19110315_TrinhNgocHien_DM-Lab01

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3



(5323, 22)

How we can merge two data into one by rows ?

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	TechSupport	StreamingTV	StreamingMovies
30	3841-NFECX	Female		1	Yes	No	71	Yes	Yes	Fiber optic	Yes	...	Yes
31	4929-XIHWV	Male		1	Yes	No	2	Yes	No	Fiber optic	No	...	No
34	3413-BMNZE	Male		1	No	No	1	Yes	No	DSL	No	...	No
50	8012-SOUDQ	Female		1	No	No	43	Yes	Yes	Fiber optic	No	...	No
52	6575-SUVOI	Female		1	Yes	No	25	Yes	Yes	DSL	Yes	...	Yes
...
7035	8456-QDAVC	Male		0	No	No	19	Yes	No	Fiber optic	No	...	No
7037	2569-WGERO	Female		0	No	No	72	Yes	No	No	No internet service	...	No internet service
7038	6840-RESVB	Male		0	Yes	Yes	24	Yes	Yes	DSL	Yes	...	Yes
7039	2234-XADUH	Female		0	Yes	Yes	72	Yes	Yes	Fiber optic	No	...	No
7042	3186-AJIEK	Male		0	No	No	66	Yes	No	Fiber optic	Yes	...	Yes

6361 rows x 22 columns

(6361, 22)

In [22]:

data_columns



A/C ENG 3:12 PM

localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 Logout

In [32]: 1 data.columns

Out[32]: Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn', 'Churn_Or_Not'], dtype='object')

Group by DataFrame

In [33]: 1 print("How we can group by customerID and Max on tenure, Sum on SeniorCitizen and Average on TotalCharges")
2 customerDF = pd.DataFrame()
3 customerDF['Max on tenure'] = data.groupby('customerID')['tenure'].max()
4 customerDF['Sum on Senior Citizen'] = data.groupby('customerID')['SeniorCitizen'].sum()
5 customerDF['Average on Monthly Charges'] = data.groupby('customerID')['MonthlyCharges'].mean()
6 customerDF['Median on Total Charges'] = data.groupby('customerID')['TotalCharges'].median()
7 display(customerDF)
8
9 print("How we can join data with customerDF on the customer information to create new information about customer")
10 customerDF["customerID"] = customerDF.index
11 customerDF = customerDF.reset_index(drop = True)
12 display(customerDF.head())
13 data = pd.merge(data, customerDF, left_on='customerID', right_on='customerID')
14 display(data.head())

How we can group by customerID and Max on tenure, Sum on SeniorCitizen and Average on TotalCharges

customerID	Max on tenure	Sum on Senior Citizen	Average on Monthly Charges	Median on Total Charges
0002-ORFBO	9	0	65.60	593.30

A/C ENG 3:15 PM



jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)



Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3



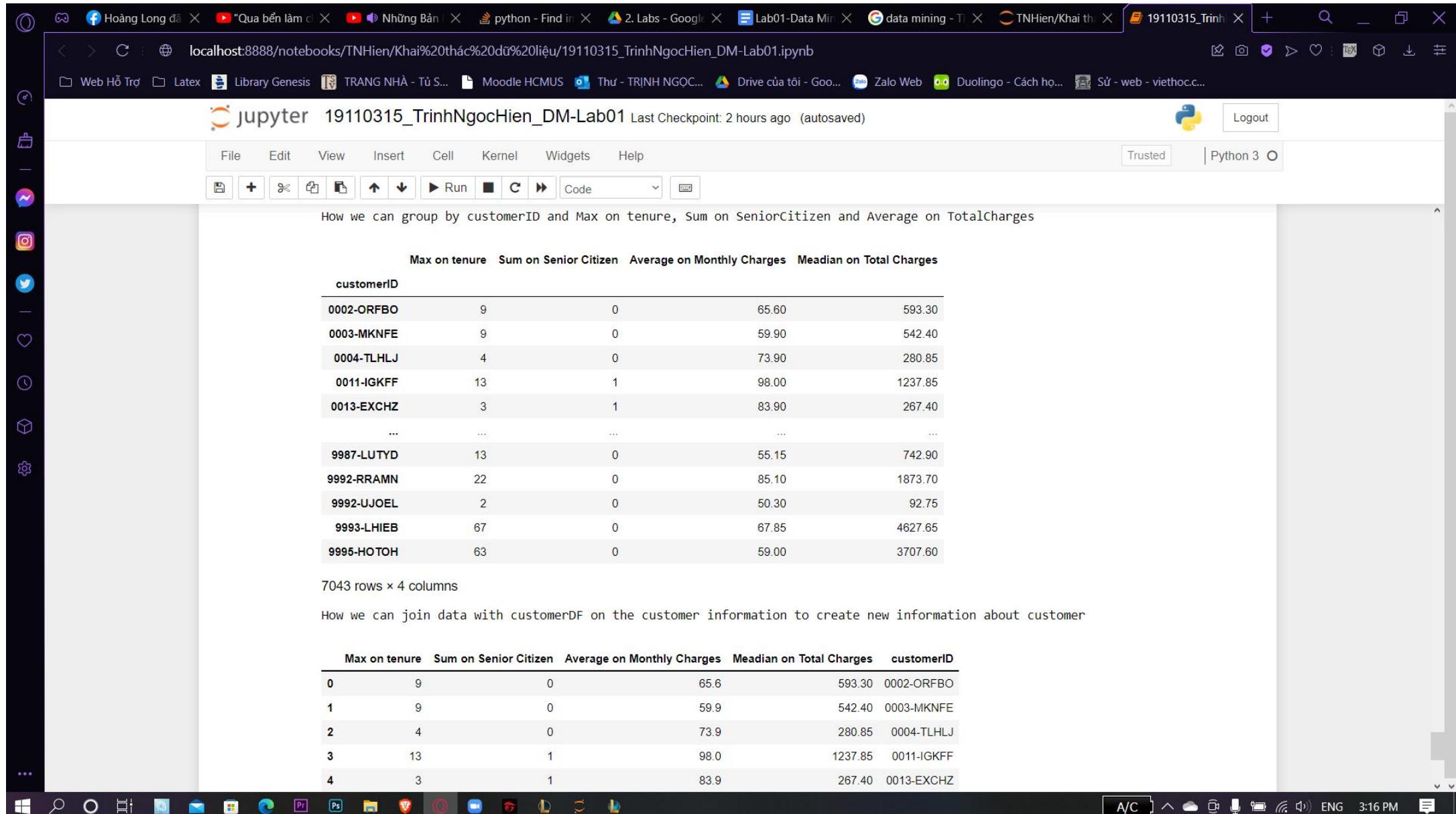
How we can group by customerID and Max on tenure, Sum on SeniorCitizen and Average on TotalCharges

customerID	Max on tenure	Sum on Senior Citizen	Average on Monthly Charges	Median on Total Charges
0002-ORFBO	9	0	65.60	593.30
0003-MKNFE	9	0	59.90	542.40
0004-TLHLJ	4	0	73.90	280.85
0011-IGKFF	13	1	98.00	1237.85
0013-EXCHZ	3	1	83.90	267.40
...
9987-LUTYD	13	0	55.15	742.90
9992-RRAMN	22	0	85.10	1873.70
9992-UJOEL	2	0	50.30	92.75
9993-LHIEB	67	0	67.85	4627.65
9995-HOTOH	63	0	59.00	3707.60

7043 rows × 4 columns

How we can join data with customerDF on the customer information to create new information about customer

	Max on tenure	Sum on Senior Citizen	Average on Monthly Charges	Median on Total Charges	customerID
0	9	0	65.6	593.30	0002-ORFBO
1	9	0	59.9	542.40	0003-MKNFE
2	4	0	73.9	280.85	0004-TLHLJ
3	13	1	98.0	1237.85	0011-IGKFF
4	3	1	83.9	267.40	0013-EXCHZ



localhost:8888/notebooks/TNHien/Khai%20thác%20dữ%20liệu/19110315_TrinhNgocHien_DM-Lab01.ipynb

jupyter 19110315_TrinhNgocHien_DM-Lab01 Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

1043 rows x 4 columns

How we can join data with customerDF on the customer information to create new information about customer

	Max on tenure	Sum on Senior Citizen	Average on Monthly Charges	Median on Total Charges	customerID
0	9	0	65.6	593.30	0002-ORFBO
1	9	0	59.9	542.40	0003-MKNFE
2	4	0	73.9	280.85	0004-TLHLJ
3	13	1	98.0	1237.85	0011-IGKFF
4	3	1	83.9	267.40	0013-EXCHZ

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	PaperlessBilling	PaymentM	
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	Electronic
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Mailed
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	Mailed
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Bank tr (auto)
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	Yes	Electronic

5 rows x 26 columns

In []: 1

A/C ENG 3:18 PM