In [1]:
```python
import pandas as pd
from pandas.plotting import scatter_matrix
import matplotlib.pyplot as plt
import pickle
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
import seaborn as sns
from sklearn.linear_model import LinearRegression
import statsmodels.formula.api as smf
import numpy as np
```

In [2]:
```python
data =pd.read_csv("Telecom_customer_churn.csv")
data
```

Out[2]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMe
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	No	No	Month-to-month	Yes	Electronic
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	No	No	No	One year	No	Mailed
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No	No	No	Month-to-month	Yes	Mailed
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	Yes	No	No	One year	No	Bank tra (autor
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No	No	No	Month-to-month	Yes	Electronic
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
7038	6840-RESVB	Male	0	Yes	Yes	24	Yes	Yes	DSL	Yes	...	Yes	Yes	Yes	Yes	One year	Yes	Mailed
7039	2234-XADUH	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	No	...	Yes	No	Yes	Yes	One year	Yes	Credi (autor
7040	4801-JZAZL	Female	0	Yes	Yes	11	No	No phone service	DSL	Yes	...	No	No	No	No	Month-to-month	Yes	Electronic
7041	8361-LTMKD	Male	1	Yes	No	4	Yes	Yes	Fiber optic	No	...	No	No	No	No	Month-to-month	Yes	Mailed
7042	3186-AJIEK	Male	0	No	No	66	Yes	No	Fiber optic	Yes	...	Yes	Yes	Yes	Yes	Two year	Yes	Bank tra (autor

7043 rows × 21 columns

In [3]:
```python
data.shape
```

Out[3]:
```
(7043, 21)
```

In [4]:
```python
data.info()
```

```
<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
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

In [5]:
```python
data.isnull().sum()
```

Out[5]:
```
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 [6]:
```python
data.describe()
```

Out[6]:

	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

In [7]:
```python
data.drop(columns ="customerID" , axis =1 , inplace = True)
data
```

Out[7]:

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMet
0	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No	No	No	No	Month-to-month	Yes	Electronic ch
1	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed ch
2	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-month	Yes	Mailed ch
3	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank tran (automa
4	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic ch
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
7038	Male	0	Yes	Yes	24	Yes	Yes	DSL	Yes	No	Yes	Yes	Yes	Yes	One year	Yes	Mailed ch
7039	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	One year	Yes	Credit (automa
7040	Female	0	Yes	Yes	11	No	No phone service	DSL	Yes	No	No	No	No	No	Month-to-month	Yes	Electronic ch
7041	Male	1	Yes	No	4	Yes	Yes	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Mailed ch
7042	Male	0	No	No	66	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes	Yes	Two year	Yes	Bank tran (automa

7043 rows × 20 columns

In [8]:
```python
data.TotalCharges = pd.to_numeric(data.TotalCharges , errors ='coerce')
```

In [9]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 20 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            7043 non-null   object
 1   SeniorCitizen     7043 non-null   int64
 2   Partner           7043 non-null   object
 3   Dependents        7043 non-null   object
 4   tenure            7043 non-null   int64
 5   PhoneService      7043 non-null   object
 6   MultipleLines     7043 non-null   object
 7   InternetService   7043 non-null   object
 8   OnlineSecurity    7043 non-null   object
 9   OnlineBackup      7043 non-null   object
 10  DeviceProtection  7043 non-null   object
 11  TechSupport       7043 non-null   object
 12  StreamingTV       7043 non-null   object
 13  StreamingMovies   7043 non-null   object
 14  Contract          7043 non-null   object
 15  PaperlessBilling  7043 non-null   object
 16  PaymentMethod     7043 non-null   object
 17  MonthlyCharges    7043 non-null   float64
 18  TotalCharges      7032 non-null   float64
 19  Churn             7043 non-null   object
dtypes: float64(2), int64(2), object(16)
memory usage: 1.1+ MB
```

In [10]:
```python
data =data.dropna()
```

In [11]:
```python
data.replace("No phone service","No" , inplace =True)
data.replace("No internet service","No" , inplace =True)
```

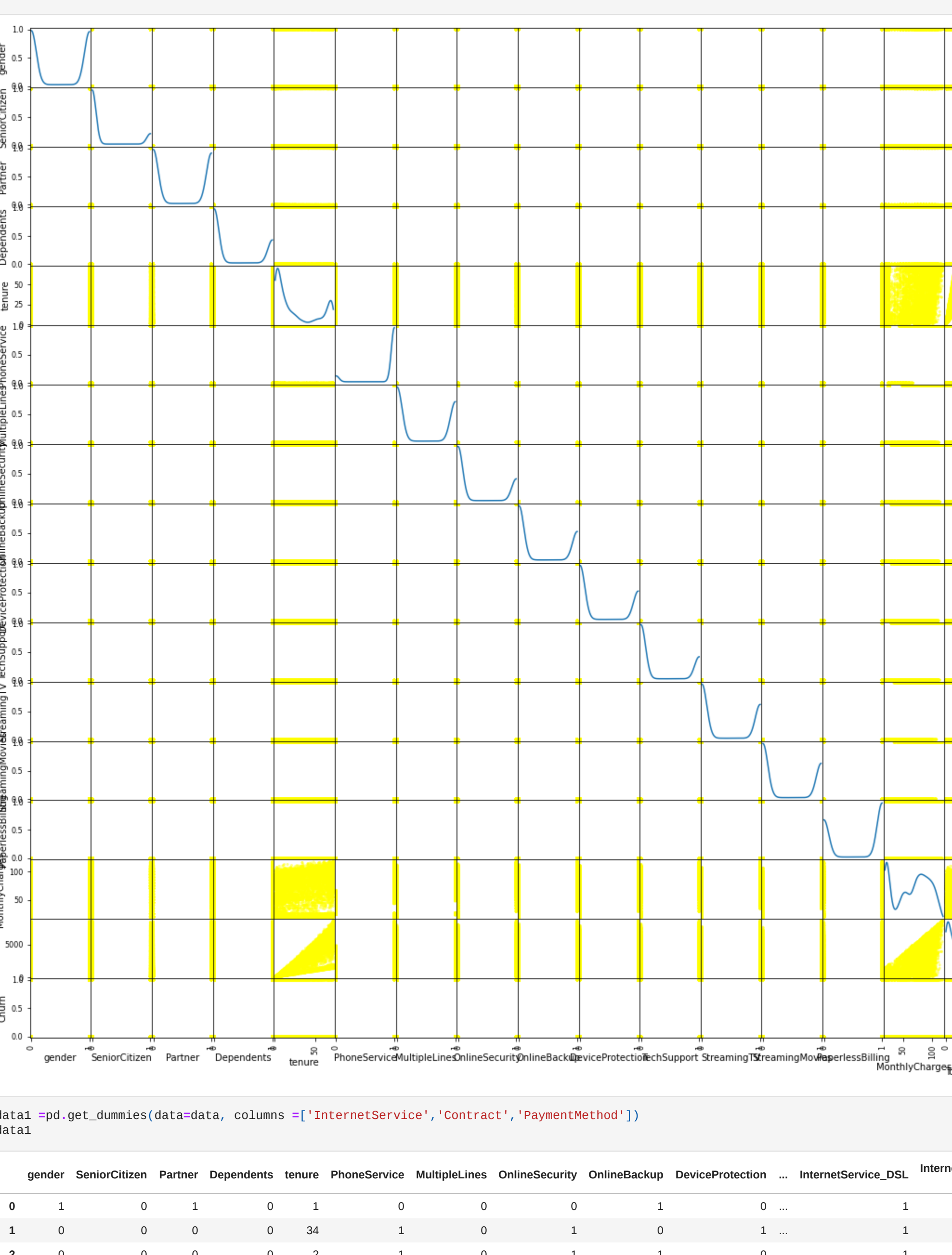
In [12]:
```python
data.replace({'Partner':{'No':0,'Yes':1},'Dependents':{'No':0,'Yes':1},'PhoneService':{'No':0,'Yes':1},'MultipleLines':{'No':0,'Yes':1},'OnlineSecurity':{"No":0,"Yes":1},
              'OnlineBackup':{'No':0,'Yes':1},'DeviceProtection':{'No':0,'Yes':1},'TechSupport':{'No':0,'Yes':1},'StreamingTV':{'No':0,'Yes':1},
          'StreamingMovies':{'No':0,'Yes':1},'PaperlessBilling':{'No':0,'Yes':1},'Churn':{'No':0,'Yes':1},'gender':{'Male':0,'Female':1}},inplace =True)
```

In [13]:
```python
data.head()
```

Out[13]:

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod
0	1	0	1	0	1	0	0	DSL	0	1	0	0	0	0	Month-to-month	1	Electronic check
1	0	0	0	0	34	1	0	DSL	1	0	1	0	0	0	One year	0	Mailed check
2	0	0	0	0	2	1	0	DSL	1	1	0	0	0	0	Month-to-month	1	Mailed check
3	0	0	0	0	45	0	0	DSL	1	0	1	1	0	0	One year	0	Bank transfer (automatic)
4	1	0	0	0	2	1	0	Fiber optic	0	0	0	0	0	0	Month-to-month	1	Electronic check

In [14]:
```python
scatter_matrix(data , figsize=(20,20) ,diagonal="kde" ,color="yellow")
plt.show()
```



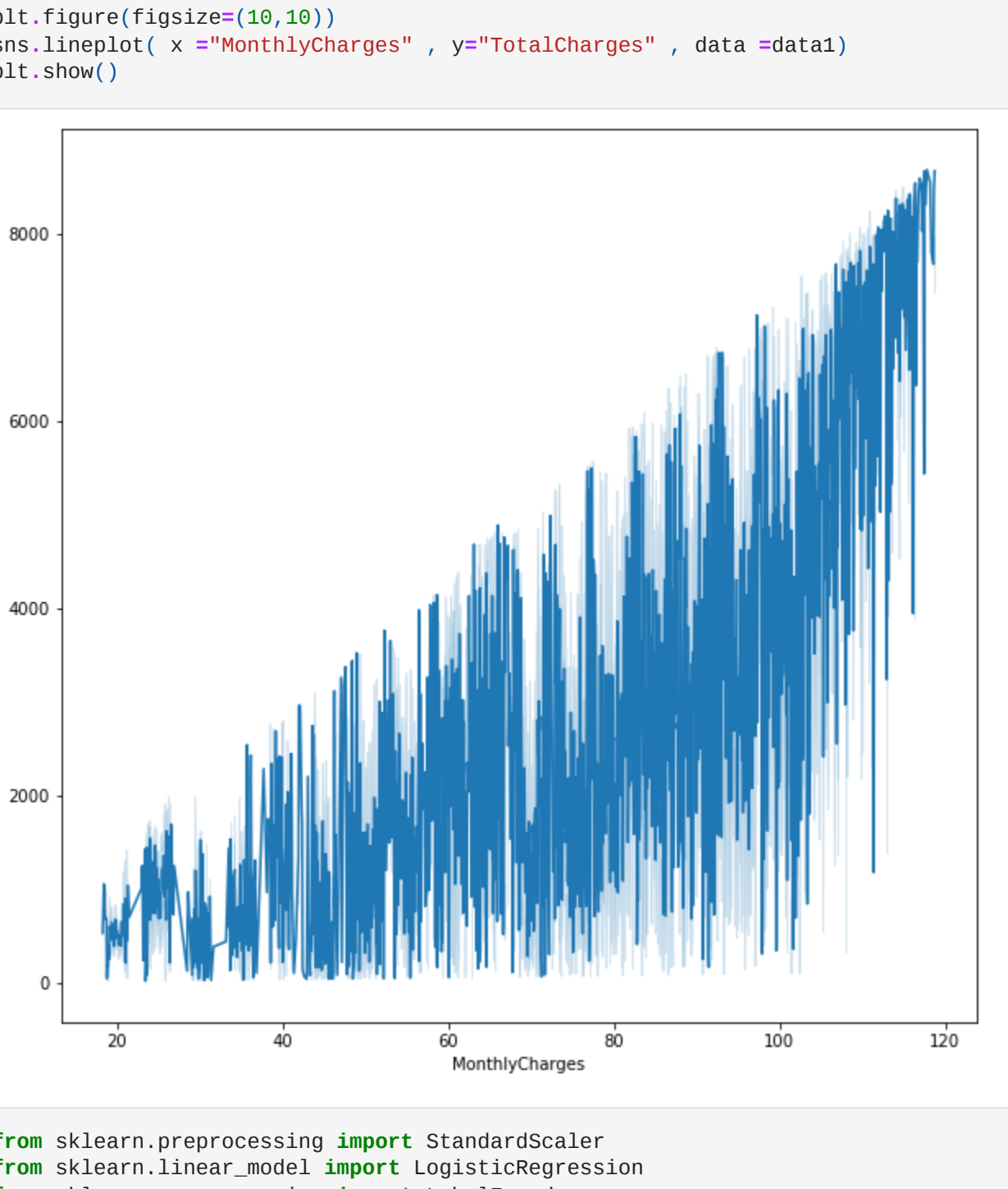
In [15]:
```python
data1 =pd.get_dummies(data=data, columns =['InternetService','Contract','PaymentMethod'])
data1
```

Out[15]:

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	OnlineSecurity	OnlineBackup	DeviceProtection	...	InternetService_DSL	InternetService_Fiber optic	InternetService_No	Contract_Month-to-month	Contract_One year
0	1	0	1	0	1	0	0	0	1	0	...	1	0	0	1	0
1	0	0	0	0	34	1	0	1	0	1	...	1	0	0	0	1
2	0	0	0	0	2	1	0	1	1	0	...	1	0	0	1	0
3	0	0	0	0	45	0	0	1	0	1	...	1	0	0	0	1
4	1	0	0	0	2	1	0	0	0	0	...	0	1	0	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7038	0	0	1	1	24	1	1	1	0	1	...	1	0	0	0	1
7039	1	0	1	1	72	1	1	0	1	1	...	0	1	0	0	1
7040	1	0	1	1	11	0	0	1	0	0	...	1	0	0	1	0
7041	0	1	1	0	4	1	1	0	0	0	...	0	1	0	1	0
7042	0	0	0	0	66	1	0	1	0	1	...	0	1	0	0	0

7032 rows × 27 columns

In [16]:
```python
plt.figure(figsize=(10,10))
sns.lineplot( x ="MonthlyCharges" , y="TotalCharges" , data =data1)
plt.show()
```



In [17]:
```python
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,confusion_matrix,roc_curve,roc_auc_score
from sklearn.metrics import classification_report
from sklearn.preprocessing import MinMaxScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
```

In [18]:
```python
scale_colum =["tenure","MonthlyCharges","TotalCharges",]
scaler = MinMaxScaler()
data1[scale_colum] =scaler.fit_transform(data1[scale_colum])
```

In [19]:
```python
x =data1.drop("Churn" ,axis =1)
y =data1["Churn"]
```

In [20]:
```python
x_train,x_test,y_train,y_test= train_test_split(x,y, test_size=0.2 , random_state =5)
print(x_train,y_train)
```

```
      gender  SeniorCitizen  Partner  Dependents    tenure  PhoneService  \
5664       1              1        0           0  0.126761             1
101        1              0        1           1  0.000000             1
2621       0              0        1           0  0.985915             1
392        1              1        0           0  0.014085             1
1327       0              0        1           0  0.816901             1
...      ...            ...      ...         ...       ...           ...
3051       1              0        1           1  0.985915             1
1730       0              0        0           0  0.000000             1
4086       1              0        1           0  1.000000             1
2259       1              0        0           0  0.098592             1
2920       1              0        1           0  1.000000             1

      MultipleLines  OnlineSecurity  OnlineBackup  DeviceProtection  ...  \
5664              0               0             0                 1  ...
101               0               0             0                 0  ...
2621              0               0             1                 1  ...
392               0               0             0                 0  ...
1327              1               0             0                 1  ...
...             ...             ...           ...               ...  ...
3051              1               1             1                 1  ...
1730              1               0             0                 0  ...
4086              1               1             1                 1  ...
2259              0               0             0                 0  ...
2920              0               1             1                 1  ...

      InternetService_DSL  InternetService_Fiber optic  InternetService_No  \
5664                    0                            1                   0
101                     0                            0                   1
2621                    1                            0                   0
392                     1                            0                   0
1327                    0                            1                   0
...                   ...                          ...                 ...
3051                    1                            0                   0
1730                    1                            0                   0
4086                    1                            0                   0
2259                    0                            0                   1
2920                    1                            0                   0

      Contract_Month-to-month  Contract_One year  Contract_Two year  \
5664                        1                  0                  0
101                         1                  0                  0
2621                        0                  0                  1
392                         1                  0                  0
1327                        0                  1                  0
...                       ...                ...                ...
3051                        0                  0                  1
1730                        1                  0                  0
4086                        0                  0                  1
2259                        0                  0                  1
2920                        0                  0                  1

      PaymentMethod_Bank transfer (automatic)  \
5664                                        0
101                                         0
2621                                        0
392                                         0
1327                                        1
...                                       ...
3051                                        1
1730                                        0
4086                                        0
2259                                        0
2920                                        0

      PaymentMethod_Credit card (automatic)  PaymentMethod_Electronic check  \
5664                                      1                               0
101                                       0                               1
2621                                      1                               0
392                                       0                               1
1327                                      0                               0
...                                     ...                             ...
3051                                      0                               0
1730                                      0                               1
4086                                      1                               0
2259                                      0                               0
2920                                      0                               0

      PaymentMethod_Mailed check
5664                           0
101                            0
2621                           0
392                            0
1327                           0
...                          ...
3051                           0
1730                           0
4086                           0
2259                           1
2920                           1

[5625 rows x 26 columns] 5664    1
101     0
2621    0
392     0
1327    1
       ..
3051    0
1730    1
4086    0
2259    0
2920    0
Name: Churn, Length: 5625, dtype: int64
```

In [21]:
```python
lr =LogisticRegression()
lr.fit(x_train,y_train)
```

Out[21]:
```
LogisticRegression()
```

In [22]:
```python
y_pred =lr.predict(x_test)
y_pred
```

Out[22]:
```
array([0, 0, 0, ..., 1, 1, 0], dtype=int64)
```

In [23]:
```python
print("Accuracy =",accuracy_score(y_test,y_pred))
print("Confusion matrix =",confusion_matrix(y_test,y_pred))
```

```
Accuracy = 0.7945984363894811
Confusion matrix = [[890 109]
 [180 228]]
```

In [24]:
```python
print(classification_report(y_test,y_pred))
```

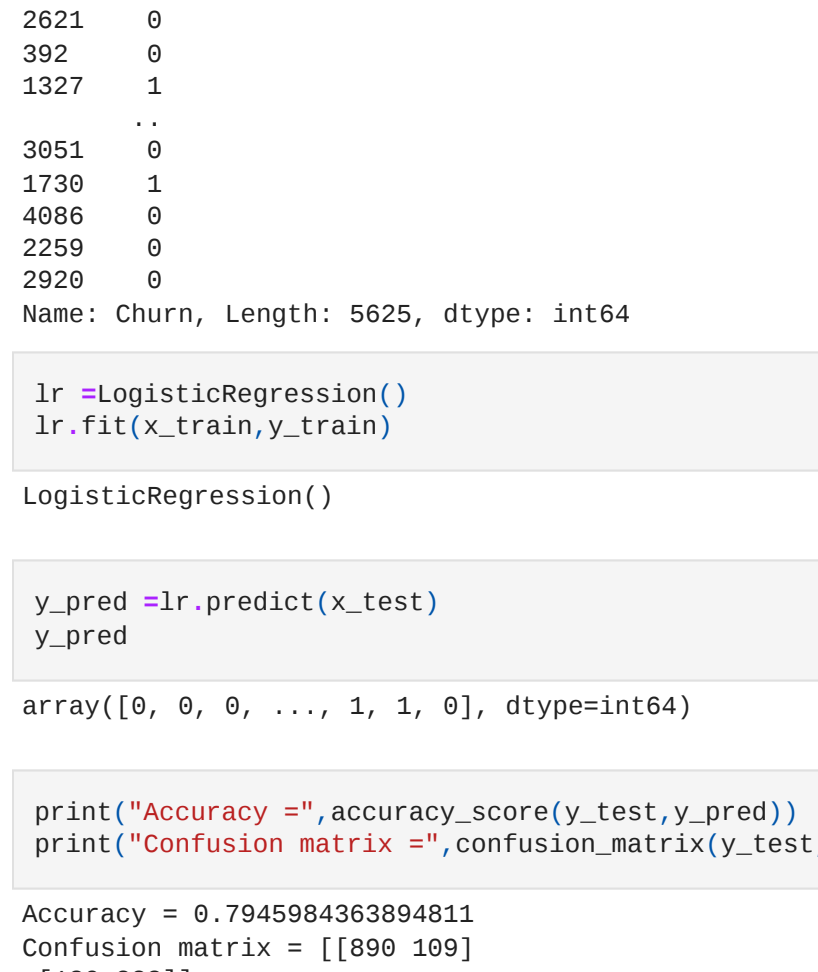
```
              precision    recall  f1-score   support

           0       0.83      0.89      0.86       999
           1       0.68      0.56      0.61       408

    accuracy                           0.79      1407
   macro avg       0.75      0.72      0.74      1407
weighted avg       0.79      0.79      0.79      1407
```

In [25]:
```python
# ROC Curve
fpr,tpr,threshold =roc_curve(y_test,y_pred)
```

In [26]:
```python
print("Threshold          :",threshold)
print("True Psitive Rate   :",tpr)
print("False Positive Rate :",fpr)
```

```
Threshold           : [2 1 0]
True Psitive Rate   : [0.         0.55882353 1.        ]
False Positive Rate : [0.         0.10910911 1.        ]
```

In [27]:
```python
plt.plot(fpr,tpr,color="r" , label ="ROC")
plt.plot([0,1],[0,1], color ="b" ,linestyle ="--")
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("Receiver Operating Characteristic (ROC) Curve")
plt.legend()
plt.show()
```



In [28]:
```python
# AUC curve

auc_score =roc_auc_score(y_test,y_pred)
print(auc_score)
```

```
0.7248572101513279
```

In [ ]:

In [29]:
```python
# Decision Tree Classifier
dtc =DecisionTreeClassifier()
dtc.fit(x_train,y_train)
```

Out[29]:
```
DecisionTreeClassifier()
```

In [30]:
```python
y_predi =dtc.predict(x_test)
y_predi
```

Out[30]:
```
array([0, 0, 1, ..., 1, 1, 0], dtype=int64)
```

In [31]:
```python
print("Accuracy =",accuracy_score(y_test,y_predi))
print("Confusion matrix =",confusion_matrix(y_test,y_predi))
```

```
Accuracy = 0.7043354655294953
Confusion matrix = [[784 215]
 [201 207]]
```

In [32]:
```python
print(classification_report(y_test,y_predi))
```

```
              precision    recall  f1-score   support

           0       0.80      0.78      0.79       999
           1       0.49      0.51      0.50       408

    accuracy                           0.70      1407
   macro avg       0.64      0.65      0.64      1407
weighted avg       0.71      0.70      0.71      1407
```

In [ ]:

In [33]:
```python
# Random Forest Classifier
rf =RandomForestClassifier()
rf.fit(x_train,y_train)
```

Out[33]:
```
RandomForestClassifier()
```

In [34]:
```python
y_predict =rf.predict(x_test)
y_predict
```

Out[34]:
```
array([0, 0, 0, ..., 1, 1, 1], dtype=int64)
```

In [35]:
```python
print("Accuracy =",accuracy_score(y_test,y_predict))
print("Confusion matrix =",confusion_matrix(y_test,y_predict))
```

```
Accuracy = 0.7718550106609808
Confusion matrix = [[891 108]
 [213 195]]
```

In [36]:
```python
print(classification_report(y_test,y_predict))
```

```
              precision    recall  f1-score   support

           0       0.81      0.89      0.85       999
           1       0.64      0.48      0.55       408

    accuracy                           0.77      1407
   macro avg       0.73      0.68      0.70      1407
weighted avg       0.76      0.77      0.76      1407
```

In [ ]:

In [ ]:

In [ ]: