Business Analysis Internship

Task 1

Understanding the Dataset

Name: Adeel Khan

Role: Business Analysis Intern

Company: Saiket Systems

❖ Task Objective

The objective of this task is to understand the given telecom customer dataset.

The dataset is explored to identify its structure, columns, data types, and missing values to ensure it is ready for analysis.

❖ Tools Used

- Python
- VS Code
- Pandas library

❖ Steps Performed

Step 1: Loading the dataset

The dataset was loaded using the pandas library to make it available for analysis.

Step 2: Viewing initial records

The first 10 rows of the dataset were displayed to understand the type of data and columns present.

Step 3: Checking dataset size

The number of rows and columns was checked to understand how many customer records and features are included.

Step 4: Identifying column names and data types

Column names and their data types were reviewed to distinguish between numerical and categorical data.

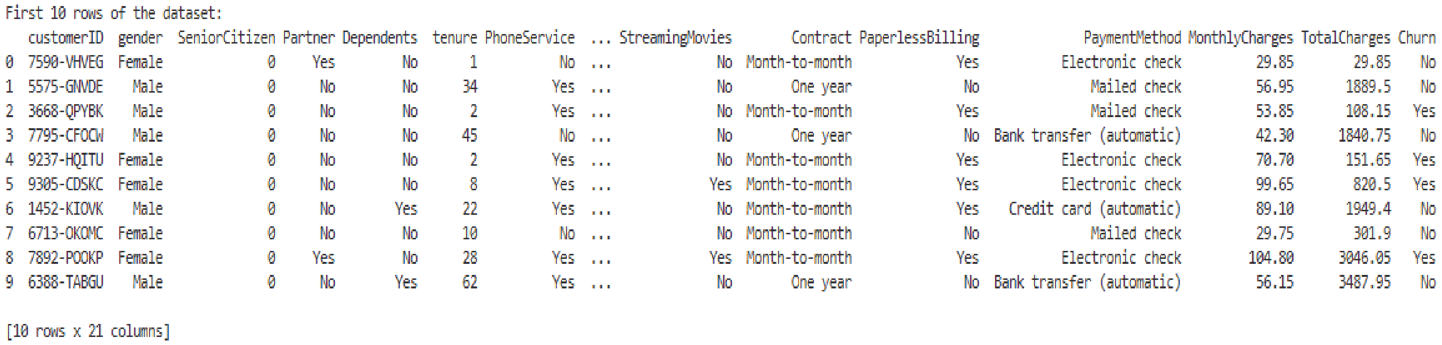
Step 5: Checking missing values

Each column was checked for missing values to ensure data completeness.

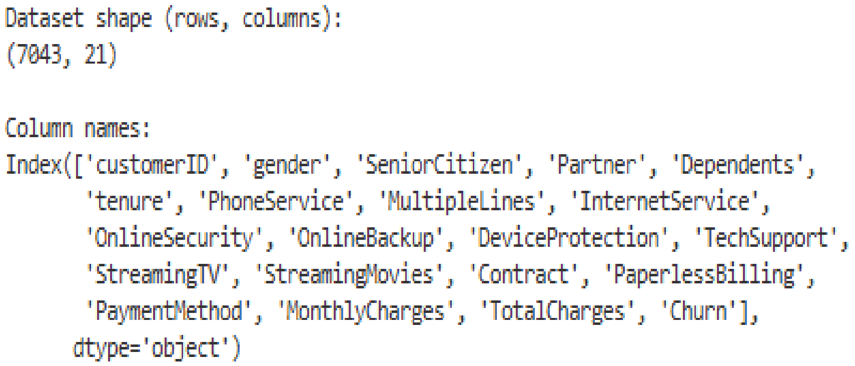
Step 6: Dataset summary

A summary of the dataset was generated to review overall structure and memory usage.

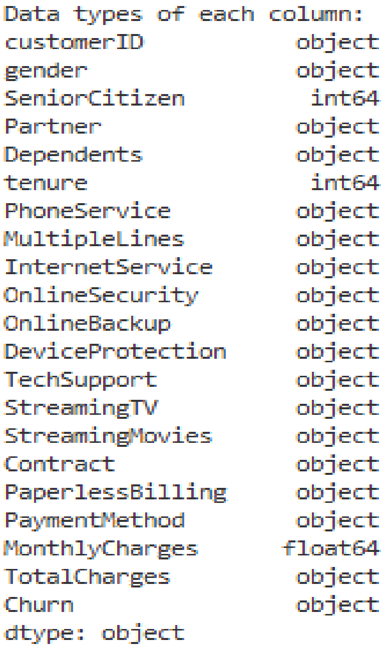
❖ Output



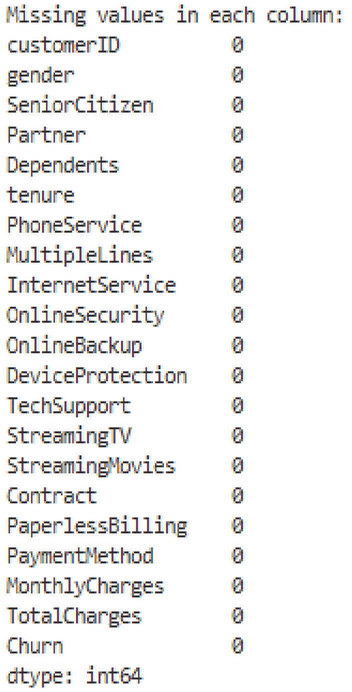
```
First 10 rows of the dataset:
   customerID  gender  SeniorCitizen Partner Dependents  tenure PhoneService  ... StreamingMovies        Contract PaperlessBilling              PaymentMethod MonthlyCharges TotalCharges Churn
0  7590-VHVEG  Female              0     Yes         No       1           No  ...              No  Month-to-month              Yes           Electronic check          29.85        29.85    No
1  5575-GNVDE    Male              0      No         No      34          Yes  ...              No        One year               No               Mailed check          56.95       1889.5    No
2  3668-QPYBK    Male              0      No         No       2          Yes  ...              No  Month-to-month              Yes               Mailed check          53.85       108.15   Yes
3  7795-CFOCW    Male              0      No         No      45           No  ...              No        One year               No  Bank transfer (automatic)          42.30      1840.75    No
4  9237-HQITU  Female              0      No         No       2          Yes  ...              No  Month-to-month              Yes           Electronic check          70.70       151.65   Yes
5  9305-CDSKC  Female              0      No         No       8          Yes  ...             Yes  Month-to-month              Yes           Electronic check          99.65        820.5   Yes
6  1452-KIOVK    Male              0      No        Yes      22          Yes  ...              No  Month-to-month              Yes    Credit card (automatic)          89.10       1949.4    No
7  6713-OKOMC  Female              0      No         No      10           No  ...              No  Month-to-month               No               Mailed check          29.75        301.9    No
8  7892-POOKP  Female              0     Yes         No      28          Yes  ...             Yes  Month-to-month              Yes           Electronic check         104.80      3046.05   Yes
9  6388-TABGU    Male              0      No        Yes      62          Yes  ...              No        One year               No  Bank transfer (automatic)          56.15      3487.95    No

[10 rows x 21 columns]
```



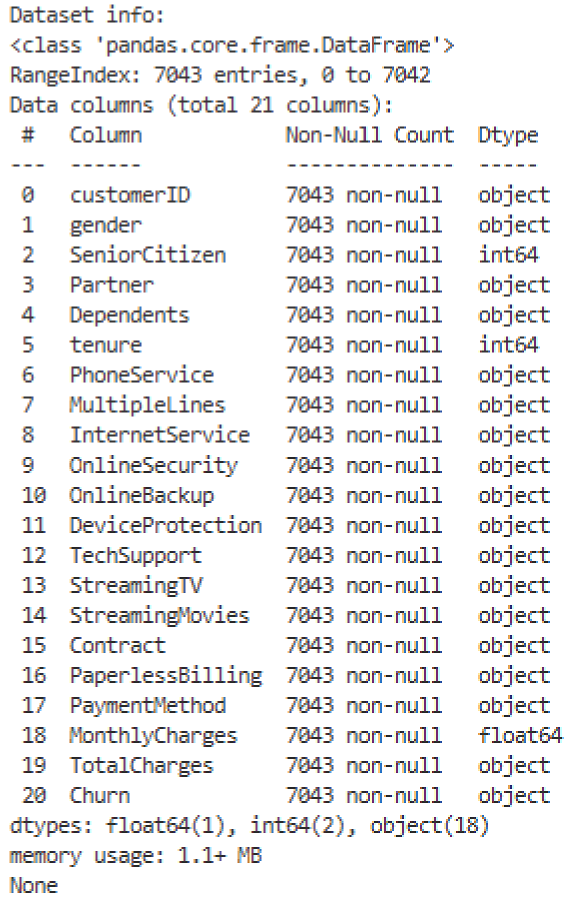
```
Dataset shape (rows, columns):
(7043, 21)

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



```
Data types of each column:
customerID           object
gender               object
SeniorCitizen         int64
Partner              object
Dependents           object
tenure                int64
PhoneService         object
MultipleLines        object
InternetService      object
OnlineSecurity       object
OnlineBackup         object
DeviceProtection     object
TechSupport          object
StreamingTV          object
StreamingMovies      object
Contract             object
PaperlessBilling     object
PaymentMethod        object
MonthlyCharges      float64
TotalCharges         object
Churn                object
dtype: object
```



```
Missing values in each column:
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
```



```
Dataset 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
None
```

❖ Conclusion

From this task, I gained a clear understanding of the dataset structure and content.
The dataset contains customer demographic, service, and billing information.

No major data quality issues were found, making the dataset suitable for further analysis.