```
import pandas as pd
import numpy as np
pd.set_option('display.max_columns',None)
df = pd.read_csv('DatasetTelcoChurn.csv')
df
```

```
      customerID  gender  SeniorCitizen Partner Dependents  tenure  \
0     7590-VHVEG  Female              0     Yes         No       1
1     5575-GNVDE    Male              0      No         No      34
2     3668-QPYBK    Male              0      No         No       2
3     7795-CFOCW    Male              0      No         No      45
4     9237-HQITU  Female              0      No         No       2
...          ...     ...            ...     ...        ...     ...
7038  6840-RESVB    Male              0     Yes        Yes      24
7039  2234-XADUH  Female              0     Yes        Yes      72
7040  4801-JZAZL  Female              0     Yes        Yes      11
7041  8361-LTMKD    Male              1     Yes         No       4
7042  3186-AJIEK    Male              0      No         No      66

     PhoneService     MultipleLines InternetService OnlineSecurity  \
0              No  No phone service             DSL             No
1             Yes                No             DSL            Yes
2             Yes                No             DSL            Yes
3              No  No phone service             DSL            Yes
4             Yes                No     Fiber optic             No
...           ...               ...             ...            ...
7038          Yes               Yes             DSL            Yes
7039          Yes               Yes     Fiber optic             No
7040           No  No phone service             DSL            Yes
7041          Yes               Yes     Fiber optic             No
7042          Yes                No     Fiber optic            Yes

     OnlineBackup DeviceProtection TechSupport StreamingTV
StreamingMovies  \
0             Yes               No          No          No
No
1              No              Yes          No          No
No
2             Yes               No          No          No
No
3              No              Yes         Yes          No
No
4              No               No          No          No
No
...           ...              ...         ...         ...
...
7038           No              Yes         Yes         Yes
Yes
7039          Yes              Yes          No         Yes
Yes
```

```
7040           No               No          No          No
No
7041           No               No          No          No
No
7042           No              Yes         Yes         Yes
Yes

            Contract PaperlessBilling              PaymentMethod  \
0     Month-to-month              Yes           Electronic check
1           One year               No               Mailed check
2     Month-to-month              Yes               Mailed check
3           One year               No  Bank transfer (automatic)
4     Month-to-month              Yes           Electronic check
...              ...              ...                        ...
7038        One year              Yes               Mailed check
7039        One year              Yes    Credit card (automatic)
7040  Month-to-month              Yes           Electronic check
7041  Month-to-month              Yes               Mailed check
7042        Two year              Yes  Bank transfer (automatic)

      MonthlyCharges TotalCharges Churn
0              29.85        29.85    No
1              56.95       1889.5    No
2              53.85       108.15   Yes
3              42.30      1840.75    No
4              70.70       151.65   Yes
...              ...          ...   ...
7038           84.80       1990.5    No
7039          103.20       7362.9    No
7040           29.60       346.45    No
7041           74.40        306.6   Yes
7042          105.65       6844.5    No

[7043 rows x 21 columns]
```

Missing Values Checking & Handling

```
df.isna().sum()
```

```
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
```

```
df.isnull().sum()
```

```
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
```

```
df.dtypes
```

```
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
```

Tidak ditemukan data NA maupun Null, artinya tidak ada missing value. Namun, tipe data TotalCharges berupa object sehingga diduga terdapat nonnumerical data.

```
#TotalCharges berkaitan dengan tenure dan monthlycharges
#cek value 0 pada tenure dan monthlycharges
print('Cek Value = 0 pada tenure')
print((df['tenure'].values == 0).sum())
print('Cek Value = 0 pada MonthlyCharges')
print((df['MonthlyCharges'].values == 0).sum())
```

```
Cek Value = 0 pada tenure
11
Cek Value = 0 pada MonthlyCharges
0
```

```
#ditemukan 11 data dengan value 0 pada tenure
#cek jumlah NaN dan Null pada TotalCharges
print('Jumlah NaN TotalCharges')
print(df['TotalCharges'].isna().sum())
print('')
print('Jumlah Null TotalCharges')
print(df['TotalCharges'].isnull().sum())
```

```
Jumlah NaN TotalCharges
0

Jumlah Null TotalCharges
0
```

```
#NaN dan Null tidak ditemukan, cek empty value
print('Cek Empty Value " " pada TotalCharges')
print((df['TotalCharges'].values == ' ').sum())
```

```
Cek Empty Value " " pada TotalCharges
11
```

```
#ditemukan empty value ' ', replace dengan NaN
df = df.replace(' ', np.nan)
df['TotalCharges'].isna().sum()
```

```
11
```

```
#replace NaN dengan 0
df=df.fillna(value=0)
df['TotalCharges'].isna().sum()
```

```
0
```

Categorical Data Encoding

```
#converting data types
df['TotalCharges'] = df['TotalCharges'].astype('float64')
df.dtypes
```

```
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        float64
Churn                object
dtype: object
```

```
df=df.drop(columns = "customerID")
df.head()
```

```
   gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0  Female              0     Yes         No       1           No
1    Male              0      No         No      34          Yes
2    Male              0      No         No       2          Yes
3    Male              0      No         No      45           No
4  Female              0      No         No       2          Yes

      MultipleLines InternetService OnlineSecurity OnlineBackup  \
0  No phone service             DSL             No          Yes
1                No             DSL            Yes           No
2                No             DSL            Yes          Yes
3  No phone service             DSL            Yes           No
4                No     Fiber optic             No           No
```

```
  DeviceProtection TechSupport StreamingTV StreamingMovies
Contract  \
0               No          No          No              No  Month-to-
month
1              Yes          No          No              No        One
year
2               No          No          No              No  Month-to-
month
3              Yes         Yes          No              No        One
year
4               No          No          No              No  Month-to-
month

  PaperlessBilling              PaymentMethod  MonthlyCharges
TotalCharges  \
0              Yes           Electronic check           29.85
29.85
1               No               Mailed check           56.95
1889.50
2              Yes               Mailed check           53.85
108.15
3               No  Bank transfer (automatic)           42.30
1840.75
4              Yes           Electronic check           70.70
151.65

  Churn
0    No
1    No
2   Yes
3    No
4   Yes

df_dummy=pd.get_dummies(df)
df_dummy

      SeniorCitizen  tenure  MonthlyCharges  TotalCharges
gender_Female  \
0                 0       1           29.85         29.85
1
1                 0      34           56.95       1889.50
0
2                 0       2           53.85        108.15
0
3                 0      45           42.30       1840.75
0
4                 0       2           70.70        151.65
1
...             ...     ...             ...           ...            .
..
```

```
7038              0      24           84.80       1990.50
0
7039              0      72          103.20       7362.90
1
7040              0      11           29.60        346.45
1
7041              1       4           74.40        306.60
0
7042              0      66          105.65       6844.50
0

      gender_Male  Partner_No  Partner_Yes  Dependents_No
Dependents_Yes  \
0               0           0            1              1
0
1               1           1            0              1
0
2               1           1            0              1
0
3               1           1            0              1
0
4               0           1            0              1
0
...           ...         ...          ...            ...
...
7038            1           0            1              0
1
7039            0           0            1              0
1
7040            0           0            1              0
1
7041            1           0            1              1
0
7042            1           1            0              1
0

      PhoneService_No  PhoneService_Yes  MultipleLines_No  \
0                   1                 0                 0
1                   0                 1                 1
2                   0                 1                 1
3                   1                 0                 0
4                   0                 1                 1
...               ...               ...               ...
7038                0                 1                 0
7039                0                 1                 0
7040                1                 0                 0
7041                0                 1                 0
7042                0                 1                 1

      MultipleLines_No phone service  MultipleLines_Yes
```

```
InternetService_DSL  \
0                                  1                  0
1
1                                  0                  0
1
2                                  0                  0
1
3                                  1                  0
1
4                                  0                  0
0
...                              ...                ...
...
7038                               0                  1
1
7039                               0                  1
0
7040                               1                  0
1
7041                               0                  1
0
7042                               0                  0
0

      InternetService_Fiber optic  InternetService_No
OnlineSecurity_No  \
0                                  0                  0
1
1                                  0                  0
0
2                                  0                  0
0
3                                  0                  0
0
4                                  1                  0
1
...                              ...                ...                .
..
7038                               0                  0
0
7039                               1                  0
1
7040                               0                  0
0
7041                               1                  0
1
7042                               1                  0
0

      OnlineSecurity_No internet service  OnlineSecurity_Yes
```

```
OnlineBackup_No  \
0                                    0                   0
0
1                                    0                   1
1
2                                    0                   1
0
3                                    0                   1
1
4                                    0                   0
1
...                                ...                 ...
...
7038                                 0                   1
1
7039                                 0                   0
0
7040                                 0                   1
1
7041                                 0                   0
1
7042                                 0                   1
1

      OnlineBackup_No internet service  OnlineBackup_Yes
DeviceProtection_No  \
0                                   0                 1
1
1                                   0                 0
0
2                                   0                 1
1
3                                   0                 0
0
4                                   0                 0
1
...                               ...               ...
...
7038                                0                 0
0
7039                                0                 1
0
7040                                0                 0
1
7041                                0                 0
1
7042                                0                 0
0

      DeviceProtection_No internet service  DeviceProtection_Yes  \
```

```
0                                    0                   0
1                                    0                   1
2                                    0                   0
3                                    0                   1
4                                    0                   0
...                                ...                 ...
7038                                 0                   1
7039                                 0                   1
7040                                 0                   0
7041                                 0                   0
7042                                 0                   1

      TechSupport_No  TechSupport_No internet service  TechSupport_Yes
\
0                  1                                0                0

1                  1                                0                0

2                  1                                0                0

3                  0                                0                1

4                  1                                0                0

...              ...                              ...              ...

7038               0                                0                1

7039               1                                0                0

7040               1                                0                0

7041               1                                0                0

7042               0                                0                1


      StreamingTV_No  StreamingTV_No internet service  StreamingTV_Yes
\
0                  1                                0                0

1                  1                                0                0

2                  1                                0                0

3                  1                                0                0

4                  1                                0                0
```

```
...              ...                                ...              ...

7038               0                                  0                1

7039               0                                  0                1

7040               1                                  0                0

7041               1                                  0                0

7042               0                                  0                1


      StreamingMovies_No  StreamingMovies_No internet service  \
0                      1                                    0
1                      1                                    0
2                      1                                    0
3                      1                                    0
4                      1                                    0
...                  ...                                  ...
7038                   0                                    0
7039                   0                                    0
7040                   1                                    0
7041                   1                                    0
7042                   0                                    0

      StreamingMovies_Yes  Contract_Month-to-month  Contract_One year
\
0                       0                        1                  0

1                       0                        0                  1

2                       0                        1                  0

3                       0                        0                  1

4                       0                        1                  0

...                   ...                      ...                ...

7038                    1                        0                  1

7039                    1                        0                  1

7040                    0                        1                  0

7041                    0                        1                  0
```

```
7042                    1                         0                   0


      Contract_Two year  PaperlessBilling_No  PaperlessBilling_Yes  \
0                     0                    0                     1
1                     0                    1                     0
2                     0                    0                     1
3                     0                    1                     0
4                     0                    0                     1
...                 ...                  ...                   ...
7038                  0                    0                     1
7039                  0                    0                     1
7040                  0                    0                     1
7041                  0                    0                     1
7042                  1                    0                     1

      PaymentMethod_Bank transfer (automatic)  \
0                                           0
1                                           0
2                                           0
3                                           1
4                                           0
...                                       ...
7038                                        0
7039                                        0
7040                                        0
7041                                        0
7042                                        1

      PaymentMethod_Credit card (automatic)  PaymentMethod_Electronic
check  \
0                                         0
1
1                                         0
0
2                                         0
0
3                                         0
0
4                                         0
1
...                                     ...
...
7038                                      0
0
7039                                      1
0
7040                                      0
1
```

```
7041                                        0
0
7042                                        0
0
```

	PaymentMethod_Mailed check	Churn_No	Churn_Yes
0	0	1	0
1	1	1	0
2	1	0	1
3	0	1	0
4	0	0	1
...	...	...	...
7038	1	1	0
7039	0	1	0
7040	0	1	0
7041	1	0	1
7042	0	1	0

[7043 rows x 47 columns]

Anomalies and Outliers

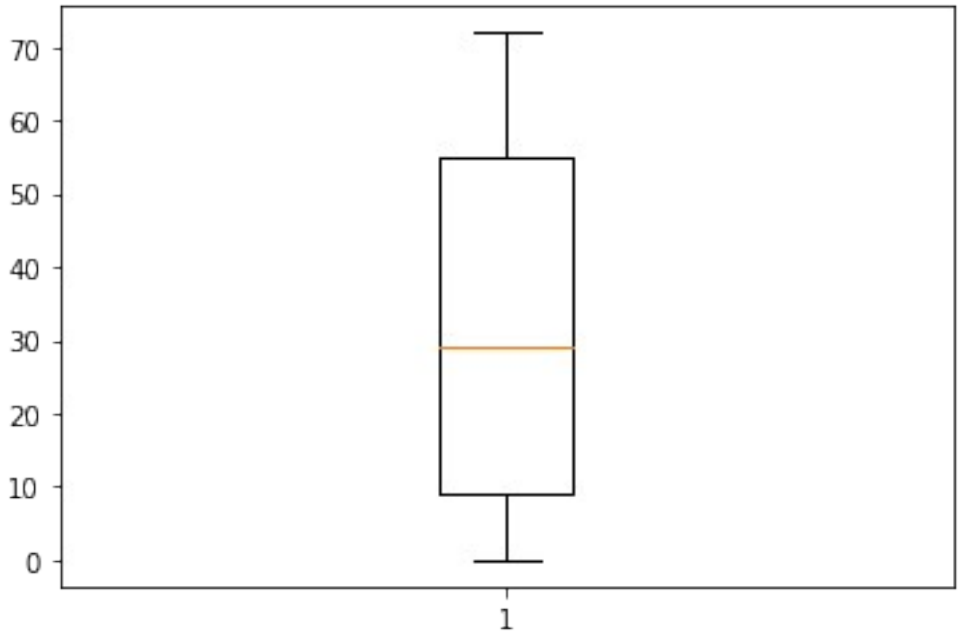
```
df.shape
```

```
(7043, 20)
```

```
import matplotlib.pyplot as plt
```

Periksa Outlier kolom tenure

```
plt.boxplot(df['tenure'])
plt.show()
```



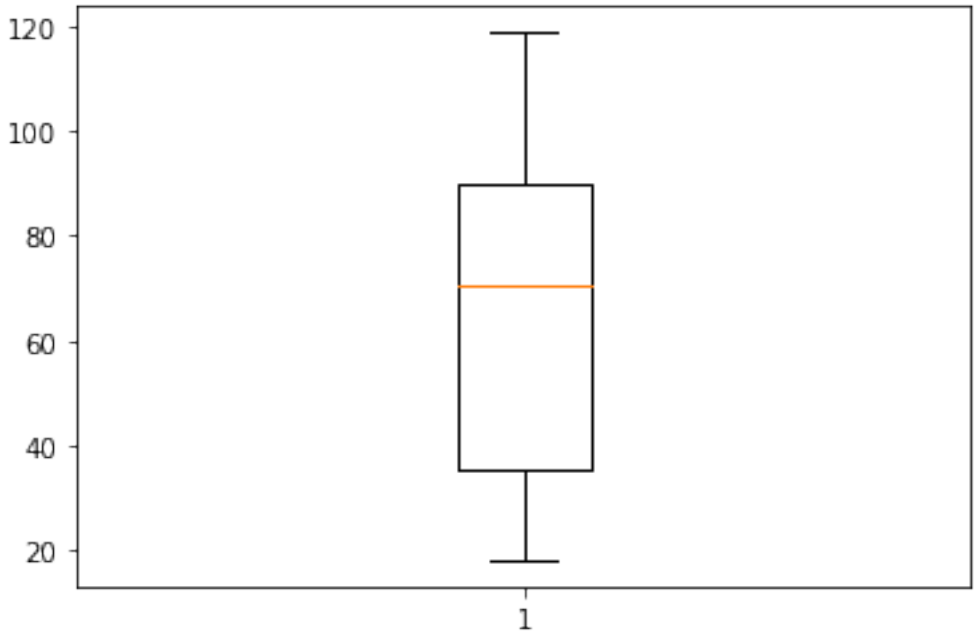
```
Q1_ten = df['tenure'].quantile(0.25)
Q3_ten = df['tenure'].quantile(0.75)
IQR_ten = Q3_ten - Q1_ten
LB_ten = Q1_ten - 1.5*IQR_ten
UB_ten = Q3_ten + 1.5*IQR_ten

Outliers_ten_UB = df[df['tenure']>UB_ten]
Outliers_ten_UB
```

```
Empty DataFrame
Columns: [gender, SeniorCitizen, Partner, Dependents, tenure,
PhoneService, MultipleLines, InternetService, OnlineSecurity,
OnlineBackup, DeviceProtection, TechSupport, StreamingTV,
StreamingMovies, Contract, PaperlessBilling, PaymentMethod,
MonthlyCharges, TotalCharges, Churn]
Index: []
```

```
Outliers_ten_LB = df[df['tenure']<LB_ten]
Outliers_ten_LB
```

```
Empty DataFrame
Columns: [gender, SeniorCitizen, Partner, Dependents, tenure,
PhoneService, MultipleLines, InternetService, OnlineSecurity,
OnlineBackup, DeviceProtection, TechSupport, StreamingTV,
StreamingMovies, Contract, PaperlessBilling, PaymentMethod,
MonthlyCharges, TotalCharges, Churn]
Index: []
```

```
plt.boxplot(df['MonthlyCharges'])
plt.show()
```



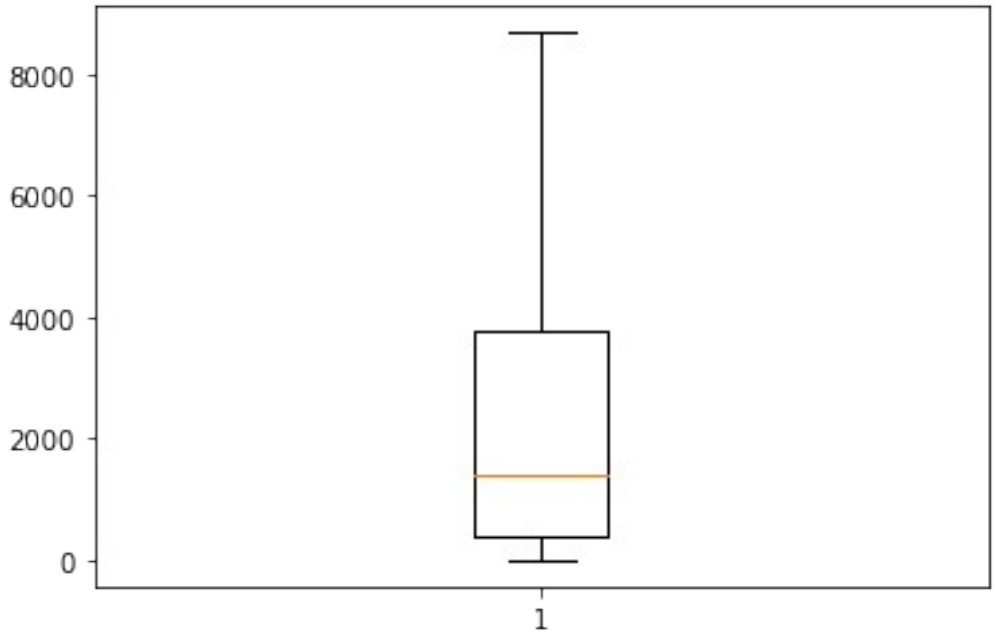
```
Q1_MC = df['MonthlyCharges'].quantile(0.25)
Q3_MC = df['MonthlyCharges'].quantile(0.75)
IQR_MC = Q3_MC - Q1_MC
LB_MC = Q1_MC - 1.5*IQR_MC
UB_MC = Q3_MC + 1.5*IQR_MC

Outliers_MC_UB = df[df['MonthlyCharges']>UB_MC]
Outliers_MC_UB
```

```
Empty DataFrame
Columns: [gender, SeniorCitizen, Partner, Dependents, tenure,
PhoneService, MultipleLines, InternetService, OnlineSecurity,
OnlineBackup, DeviceProtection, TechSupport, StreamingTV,
StreamingMovies, Contract, PaperlessBilling, PaymentMethod,
MonthlyCharges, TotalCharges, Churn]
Index: []
```

```
Outliers_MC_LB = df[df['MonthlyCharges']<LB_MC]
Outliers_MC_LB
```

```
Empty DataFrame
Columns: [gender, SeniorCitizen, Partner, Dependents, tenure,
PhoneService, MultipleLines, InternetService, OnlineSecurity,
OnlineBackup, DeviceProtection, TechSupport, StreamingTV,
StreamingMovies, Contract, PaperlessBilling, PaymentMethod,
MonthlyCharges, TotalCharges, Churn]
Index: []
```

```
plt.boxplot(df['TotalCharges'])
plt.show()
```



```
Q1_TC = df['TotalCharges'].quantile(0.25)
Q3_TC = df['TotalCharges'].quantile(0.75)
IQR_TC = Q3_TC - Q1_TC
LB_TC = Q1_TC - 1.5*IQR_TC
UB_TC = Q3_TC + 1.5*IQR_TC

Outliers_TC_UB = df[df['TotalCharges']>UB_TC]
Outliers_TC_UB
```

```
Empty DataFrame
Columns: [gender, SeniorCitizen, Partner, Dependents, tenure,
PhoneService, MultipleLines, InternetService, OnlineSecurity,
OnlineBackup, DeviceProtection, TechSupport, StreamingTV,
StreamingMovies, Contract, PaperlessBilling, PaymentMethod,
MonthlyCharges, TotalCharges, Churn]
Index: []
```

```
Outliers_TC_LB = df[df['TotalCharges']<LB_TC]
Outliers_TC_LB
```

```
Empty DataFrame
Columns: [gender, SeniorCitizen, Partner, Dependents, tenure,
PhoneService, MultipleLines, InternetService, OnlineSecurity,
OnlineBackup, DeviceProtection, TechSupport, StreamingTV,
StreamingMovies, Contract, PaperlessBilling, PaymentMethod,
MonthlyCharges, TotalCharges, Churn]
Index: []
```