# dlookr: EDA

# Overview

## Data Structures

| Divisions | Metrics | Values |
|---|---|---|
| size | observations | 6,499 |
| size | variables | 21 |
| size | values | 136,479 |
| size | memory size (MB) | 1 |
| duplicated | duplicate observation | 0 |
| missing | complete observation | 6,499 |
| missing | missing observation | 0 |
| missing | missing variables | 0 |
| missing | missing values | 0 |

## Data Types

| Divisions | Metrics | Values |
|---|---|---|
| data type | numerics | 4 |
| data type | integers | 0 |
| data type | factors/ordered | 0 |
| data type | characters | 17 |
| data type | Dates | 0 |
| data type | POSIXcts | 0 |
| data type | others | 0 |

## Job Informations

| Divisions | Metrics | Values |
|---|---|---|
| dataset | dataset | . |
| dataset | dataset type | spec_tbl_df |
| dataset | target | Total_Charges |
| job | samples | 6,499 / 6,499 (100%) |
| job | created | 2023-10-10 21:20:08.060742 |
| job | created by | dlookr |

# Univariate Analysis

# Descriptive Statistics

Descriptive statistics and visualization of individual variables

| | Variables | Missing (%) | Distincts (Ratio) | Zeros | Negatives | Outliers |
|---|---|---|---|---|---|---|
| ▶ | **CustomerID** <br> character | 0 (0) | 6,499 (1) | - | - | - |
| ▶ | **Gender** <br> character | 0 (0) | 2 (0) | - | - | - |
| ▶ | **Senior Citizen** <br> numeric | 0 (0) | 2 (0) | 5,443 | 0 | 1,056 |
| ▶ | **Partner** <br> character | 0 (0) | 2 (0) | - | - | - |
| ▶ | **Dependents** <br> character | 0 (0) | 2 (0) | - | - | - |
| ▶ | **Tenure** <br> numeric | 0 (0) | 73 (0.011) | 9 | 0 | 0 |
| ▶ | **Phone Service** <br> character | 0 (0) | 2 (0) | - | - | - |
| ▶ | **Multiple Lines** <br> character | 0 (0) | 3 (0) | - | - | - |
| ▶ | **Internet Service** <br> character | 0 (0) | 3 (0) | - | - | - |
| ▶ | **Online Security** <br> character | 0 (0) | 3 (0) | - | - | - |

1–10 of 21 rows                                        Previous   **1**   2   3   Next

# Normality Test

Normality test results of numeric variables

| | Variables | Mean | Min | Q1 | Median | Q3 | |
|---|---|---|---|---|---|---|---|
| ▶ | Senior Citizen | 0.16 | 0 | 0 | 0 | 0 | |
| ▶ | Tenure | 32.37 | 0 | 9 | 29 | 55 | |
| ▶ | Monthly_Charges | 64.73 | 18.25 | 35.4 | 70.35 | 89.85 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ▶ Total_Charges | 2282.94 | 18.8 | 399.8 | 1399.35 | 3778.1 | 8 |

◀ ▬▬▬▬▬▬▬▬▬ ▶

# Bivariate Analysis

## Compare Numerical Variables

Relationship between two numerical variables

| First Variable | Second Variable | Correlation Coefficient |
|---|---|---|
| ▶ Senior Citizen | Tenure | 0.01573 |
| ▶ Senior Citizen | Monthly_Charges | 0.22010 |
| ▶ Senior Citizen | Total_Charges | 0.10255 |
| ▶ Tenure | Monthly_Charges | 0.24895 |
| ▶ Tenure | Total_Charges | 0.82456 |
| ▶ Monthly_Charges | Total_Charges | 0.65105 |

## Compare Categorical Variables

The number of categorical variables is less than 2.

# Multivariate Analysis
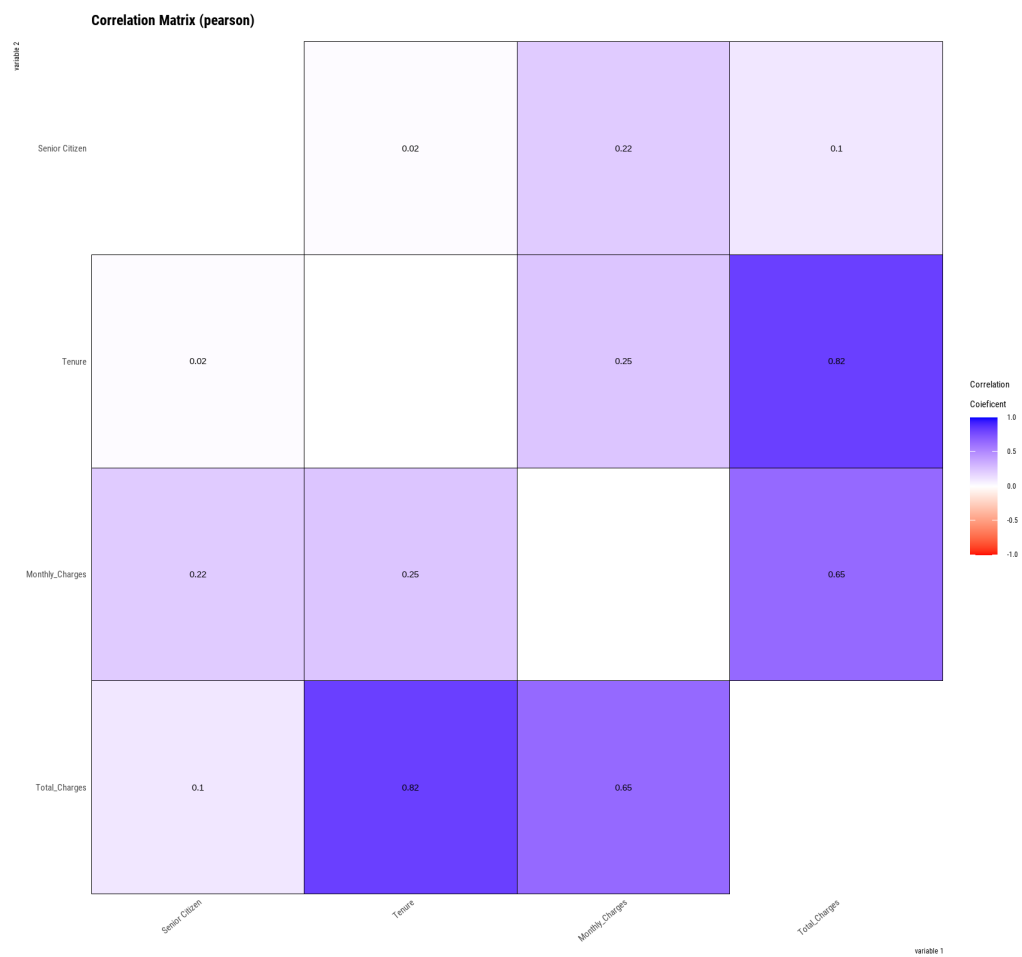
## Correlation Analysis

### Correlation Matrix

Correlation coefficient matrix of numeric variables

| | | Second Variable | | |
|---|---|---|---|---|
| First Variable | Senior Citizen | Tenure | Monthly_Charges | Total_Charges |
| **Senior Citizen** | | 0.016 | 0.220 | 0.103 |
| **Tenure** | 0.016 | | 0.249 | 0.825 |
| **Monthly_Charges** | 0.220 | 0.249 | | 0.651 |

| | | | |
|---|---|---|---|
| **Total_Charges** | 0.103 | 0.825 | 0.651 |

# Correlation Plot

Correlation coefficient matrix plot



# Target based Analysis
## Target Variable and Numeric Variables

Relationship between two numerical variables by level of target variable

| | Variable | Target Variable |
|---|---|---|
| ▶ | Senior Citizen | Total_Charges |
| ▶ | Tenure | Total_Charges |
| ▶ | Monthly_Charges | Total_Charges |

# Target Variable and Categorical Variables

There are no categorical variables except for the target variable.

---

Created by [dlookr package](https://choonghyunryu.github.io/dlookr/) (https://choonghyunryu.github.io/dlookr/)