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	2
data type	characters	15
data type	Dates	0
data type	POSIXcts	0
data type	others	0

Job Informations

Divisions	Metrics	Values
dataset	dataset	.
dataset	dataset type	tbl_df
dataset	target	Total Charges
job	samples	6,499 / 6,499 (100%)
job	created	2023-10-02 23:30:47.831508
job	created by	dlookr

Univariate Analysis
Descriptive Statistics

Descriptive statistics and visualization of individual variables

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

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
▶	Total Charges	2281.71	18.8	399.8	1397.1	3778.1

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	0.10278
▶	Tenure	Monthly Charges	0.24895
▶	Tenure	Total Charges	0.82519
▶	Monthly Charges	Total Charges	0.65150

Compare Categorical Variables

Relationship between two categorical variables

	First Variable	Second Variable	Degree of Freedom	Statistic	P-Value
▶	Paperless Billing	Payment Method	8	414.26	0.00000

Multivariate Analysis
Correlation Analysis

Correlation Matrix

Correlation coefficient matrix of numeric variables

First Variable	Second 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		0.652
Total Charges	0.103	0.825	0.652	

Correlation Plot

Correlation coefficient matrix plot

Correlation Matrix (pearson)
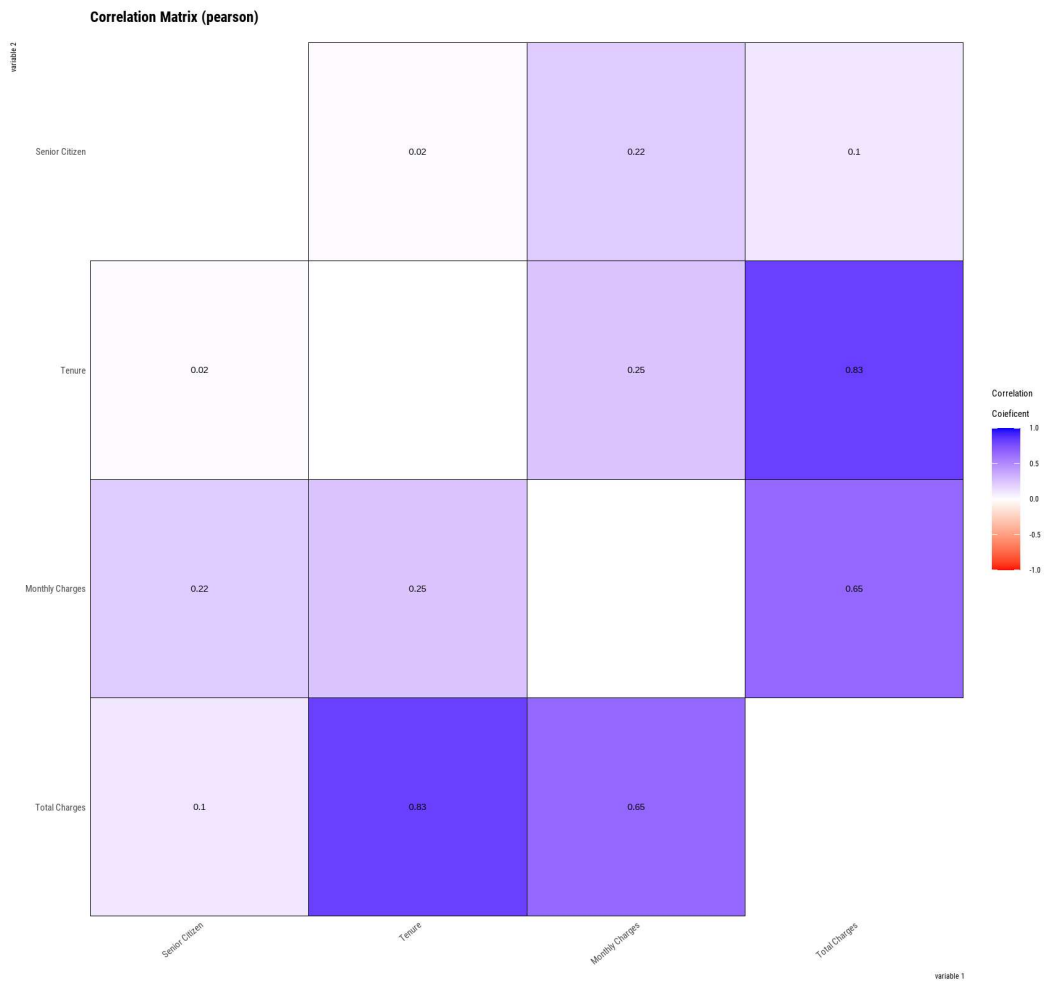
Target based Analysis

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

Relationship between two categorical variables by level of target variable

Variable	Target Variable
▶ Paperless Billing	Total Charges
▶ Payment Method	Total Charges

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