

EDA REPORT

CHURN



Report Overview

This report was created for the EDA of **churn** data. It helps explore data to understand the data and find scenarios for performing the analysis.

Contents

Overview	2
Data Structures	2
Job Informations	2
Univariate Analysis	3
Descriptive Statistics	3
Numerical Variables	3
Categorical Variables	5
Normality Test	7
Bivariate Analysis	12
Compare Numerical Variables	12
Compare Categorical Variables	19
Multivariate Analysis	20
Correlation Analysis	20
Correlation Coefficient Matrix	20
Correlation Plot	21

Overview

Data Structures

division	metrics	value	division	metrics	value
size	observations	6,499	data type	numerics	2
size	variables	21	data type	integers	2
size	values	136,479	data type	factors/ordered	0
size	memory size (MB)	1	data type	characters	17
duplicated	duplicate observation	0	data type	Dates	0
missing	complete observation	6,490	data type	POSIXcts	0
missing	missing observation	9	data type	others	0
missing	missing variables	1			
missing	missing values	9			

Table 1: Data structures and types

Job Informations

division	metrics	value
dataset	dataset	churn
dataset	dataset type	data.frame
dataset	target	not defied
job	samples	6,499 / 6,499 (100%)
job	created	2024-06-03 12:05:59.629789
job	created by	dlookr

Table 2: Job informations

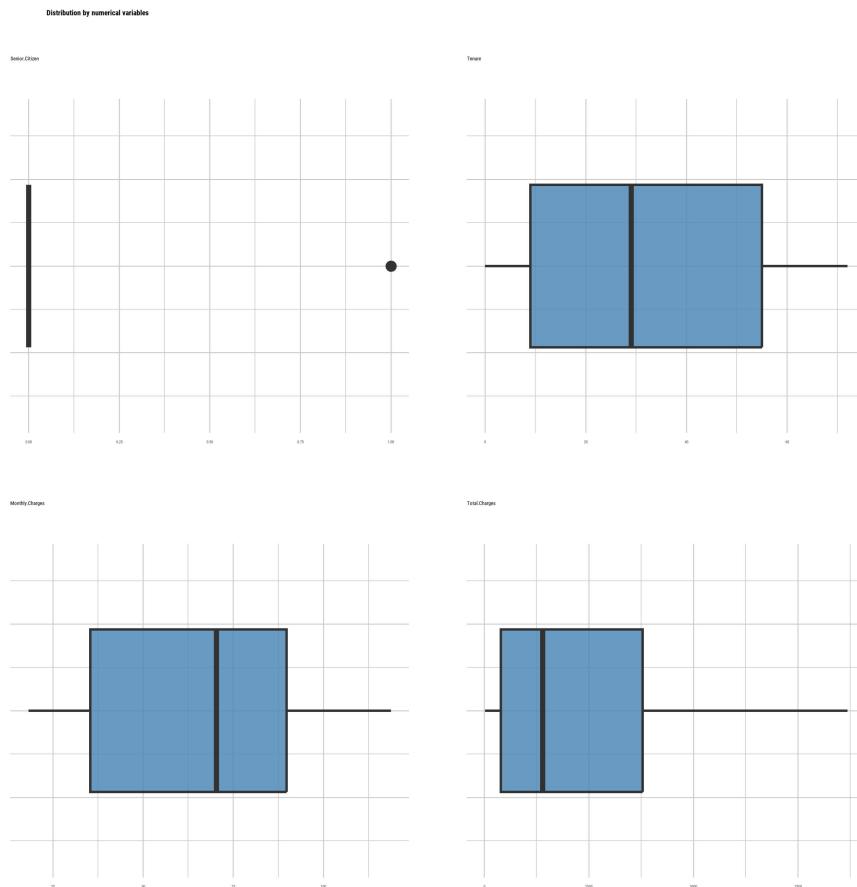
Univariate Analysis

Descriptive Statistics

Numerical Variables

variables	missing	mean	sd	min	Q1	median	Q3	max
Senior.Citizen	0	0.16	0.37	0.00	0.0	0.00	0.00	1.00
Tenure	0	32.37	24.58	0.00	9.0	29.00	55.00	72.00
Monthly.Charges	0	64.73	30.14	18.25	35.4	70.35	89.85	118.75
Total.Charges	9	2,282.94	2,270.03	18.80	399.3	1,397.10	3,786.61	8,684.80

Table 3: Descriptive statistics of numerical variables



variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
Senior.Citizen	integer	2	1.83	1.35	5,443	0	1,056
Tenure	integer	73	0.24	-1.39	9	0	0
Monthly.Charges	numeric	1,554	-0.22	-1.26	0	0	0
Total.Charges	numeric	6,047	0.96	-0.23	0	0	0

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
CustomerID	0002-ORFBO	6,499	1	0.02	1
CustomerID	0003-MKNFE	6,499	1	0.02	1
CustomerID	0004-TLHLJ	6,499	1	0.02	1
CustomerID	0011-IGKFF	6,499	1	0.02	1
CustomerID	0013-EXCHZ	6,499	1	0.02	1
CustomerID	0013-MHZWF	6,499	1	0.02	1
CustomerID	0013-SMEOE	6,499	1	0.02	1
CustomerID	0014-BMAQU	6,499	1	0.02	1
CustomerID	0015-UOCOJ	6,499	1	0.02	1
CustomerID	0016-QLJIS	6,499	1	0.02	1
Gender	Male	6,499	3,290	50.62	1
Gender	Female	6,499	3,209	49.38	2
Partner	No	6,499	3,359	51.68	1
Partner	Yes	6,499	3,140	48.32	2
Dependents	No	6,499	4,561	70.18	1
Dependents	Yes	6,499	1,938	29.82	2
Phone.Service	Yes	6,499	5,884	90.54	1
Phone.Service	No	6,499	615	9.46	2
Multiple.Lines	No	6,499	3,138	48.28	1
Multiple.Lines	Yes	6,499	2,746	42.25	2
Multiple.Lines	No phone service	6,499	615	9.46	3
Internet.Service	Fiber optic	6,499	2,860	44.01	1
Internet.Service	DSL	6,499	2,217	34.11	2
Internet.Service	No	6,499	1,422	21.88	3
Online.Security	No	6,499	3,208	49.36	1

Table 4: Top rank levels of categorical variables

variables	levels	observations	frequency	frequency(%)	rank
variables	levels	observations	frequency	frequency(%)	rank
26	Online.Security	Yes	6,499	1,869	28.76
27	Online.Security	No internet service	6,499	1,422	21.88
28	Online.Backup	No	6,499	2,855	43.93
29	Online.Backup	Yes	6,499	2,222	34.19
30	Online.Backup	No internet service	6,499	1,422	21.88
31	Device.Protection	No	6,499	2,843	43.75
32	Device.Protection	Yes	6,499	2,234	34.37
33	Device.Protection	No internet service	6,499	1,422	21.88
34	Tech.Support	No	6,499	3,209	49.38
35	Tech.Support	Yes	6,499	1,868	28.74
36	Tech.Support	No internet service	6,499	1,422	21.88
37	Streaming.TV	No	6,499	2,589	39.84
38	Streaming.TV	Yes	6,499	2,488	38.28
39	Streaming.TV	No internet service	6,499	1,422	21.88
40	Streaming.Movies	No	6,499	2,555	39.31
41	Streaming.Movies	Yes	6,499	2,522	38.81
42	Streaming.Movies	No internet service	6,499	1,422	21.88
43	Contract	Month-to-month	6,499	3,576	55.02
44	Contract	Two year	6,499	1,565	24.08
45	Contract	One year	6,499	1,358	20.90
46	Paperless.Billing	Yes	6,499	3,836	59.02
47	Paperless.Billing	No	6,499	2,663	40.98
48	Payment.Method	Electronic check	6,499	2,182	33.57
49	Payment.Method	Mailed check	6,499	1,485	22.85
50	Payment.Method	Bank transfer (automatic)	6,499	1,426	21.94
51	Payment.Method	Credit card (automatic)	6,499	1,406	21.63
52	Churn	No	6,499	4,784	73.61
53	Churn	Yes	6,499	1,715	26.39

Table 4: Top rank levels of categorical variables (continued)

The number of categorical(factor/ordered) variables is 0.

Normality Test

described_variables	min	Q1	median	Q3	max	skewness	kurtosis	balance
Senior.Citizen	0.0	0.0	0.0	0.0	1.0	1.8	1.4	Right-Skewed
Tenure	0.0	9.0	29.0	55.0	72.0	0.2	-1.4	Balanced
Monthly.Charges	18.2	35.4	70.3	89.8	118.8	-0.2	-1.3	Balanced
Total.Charges	18.8	399.3	1397.1	3786.6	8684.8	1.0	-0.2	Balanced

Table 5: Descriptive statistics of numerical variables

Senior.Citizen

(unique) sample size must be greater than 3

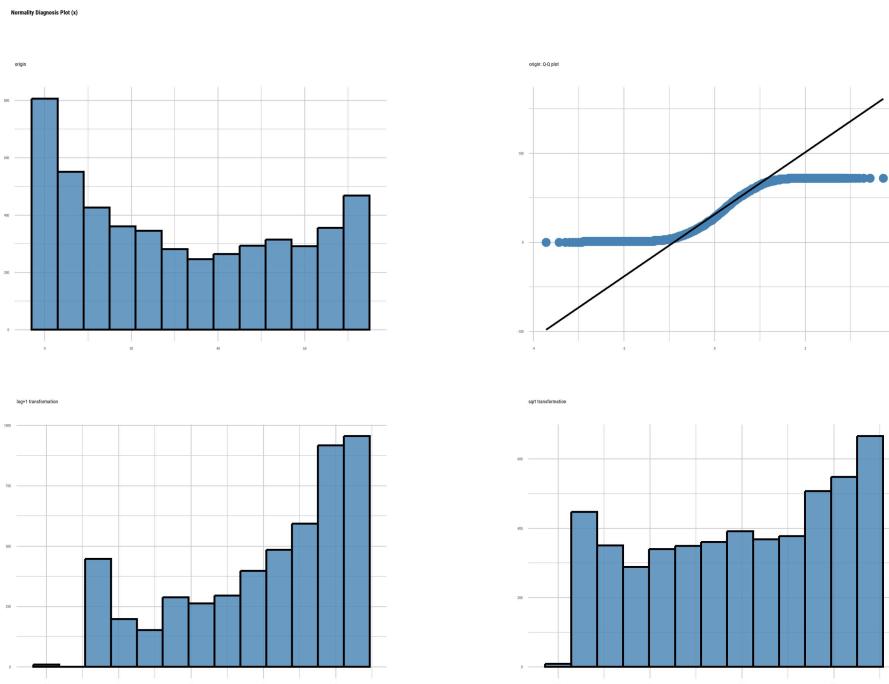
Tenure

statistic	p_value	remark
0.90143	5.7413e-49	5000 samples

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.2649	1.6236
log+1 transformation	-0.7461	2.2925
sqrt transformation	-0.2110	1.7022

Table 6: skewness and kurtosis



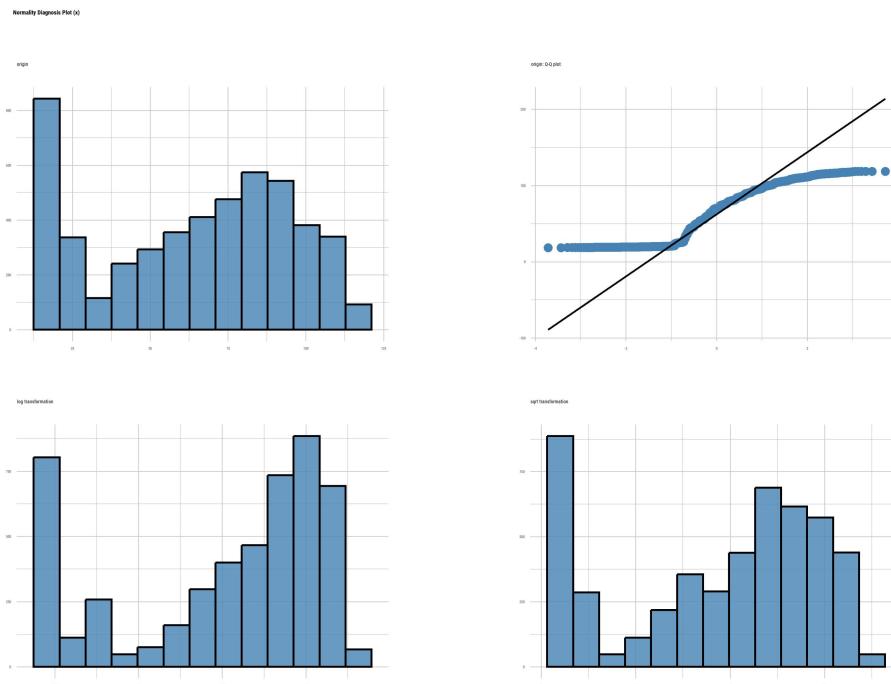
Monthly.Charges

statistic	p_value	remark
0.92108	2.6124e-45	5000 samples

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.1975	1.7384
log transformation	-0.7064	2.0034
sqrt transformation	-0.4631	1.8202

Table 6: skewness and kurtosis



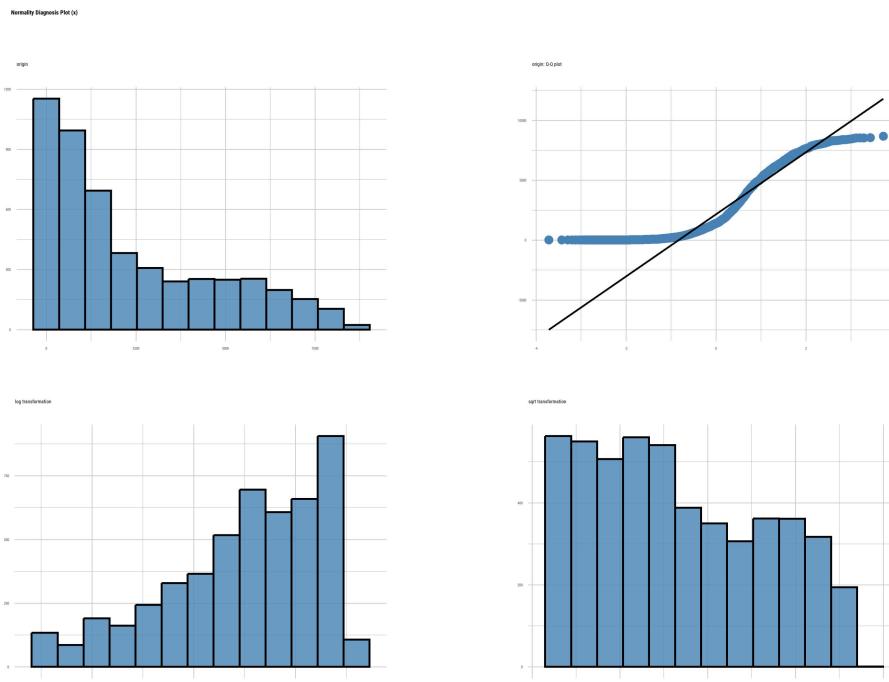
Total.Charges

statistic	p_value	remark
0.85993	4.7494e-55	5000 samples

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.9382	2.6789
log transformation	-0.7667	2.8040
sqrt transformation	0.3089	1.9213

Table 6: skewness and kurtosis



Bivariate Analysis

Compare Numerical Variables

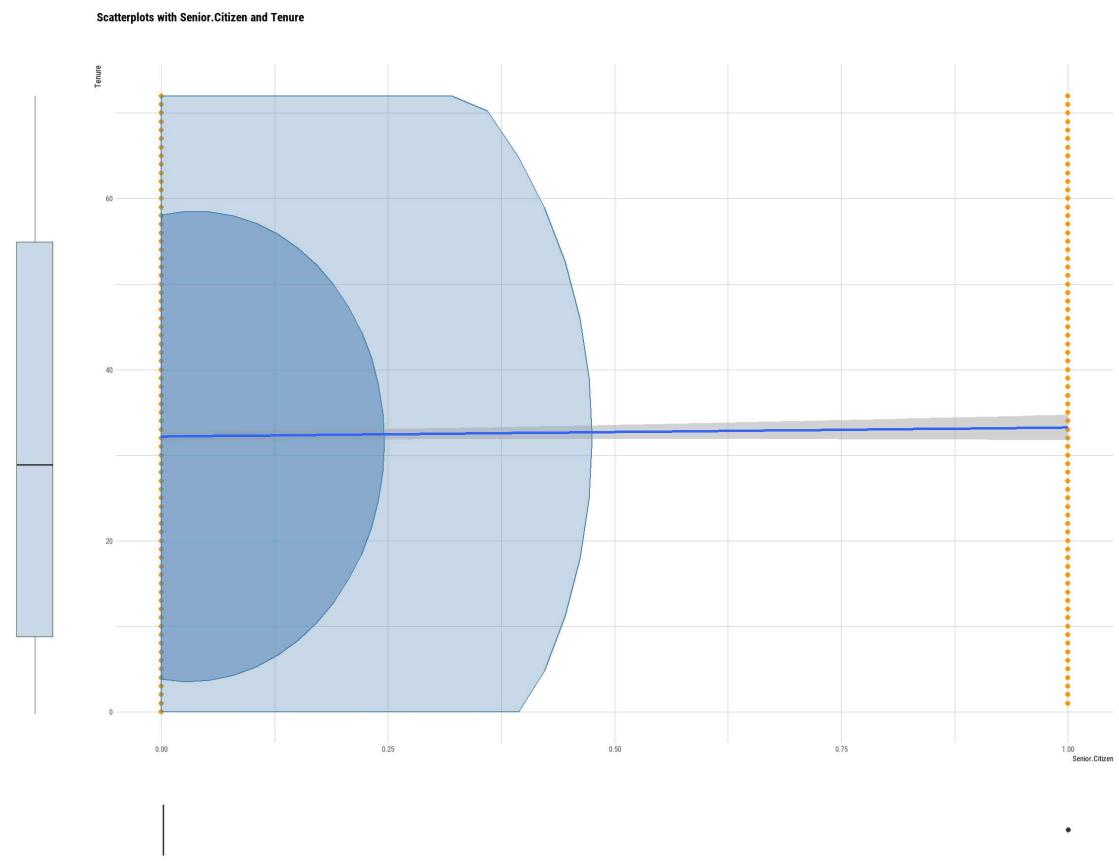
first variable	second variable	correlation coefficient
Senior.Citizen	Tenure	0.01573
Senior.Citizen	Monthly.Charges	0.22010
Senior.Citizen	Total.Charges	0.10256
Tenure	Monthly.Charges	0.24895
Tenure	Total.Charges	0.82555
Monthly.Charges	Total.Charges	0.65169

Table 7: Correlation coefficient

'Senior.Citizen' vs 'Tenure'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Senior.Citizen	Tenure	0.0002475	9.36e-05	0.3689077	1.608412	0.2047606	1

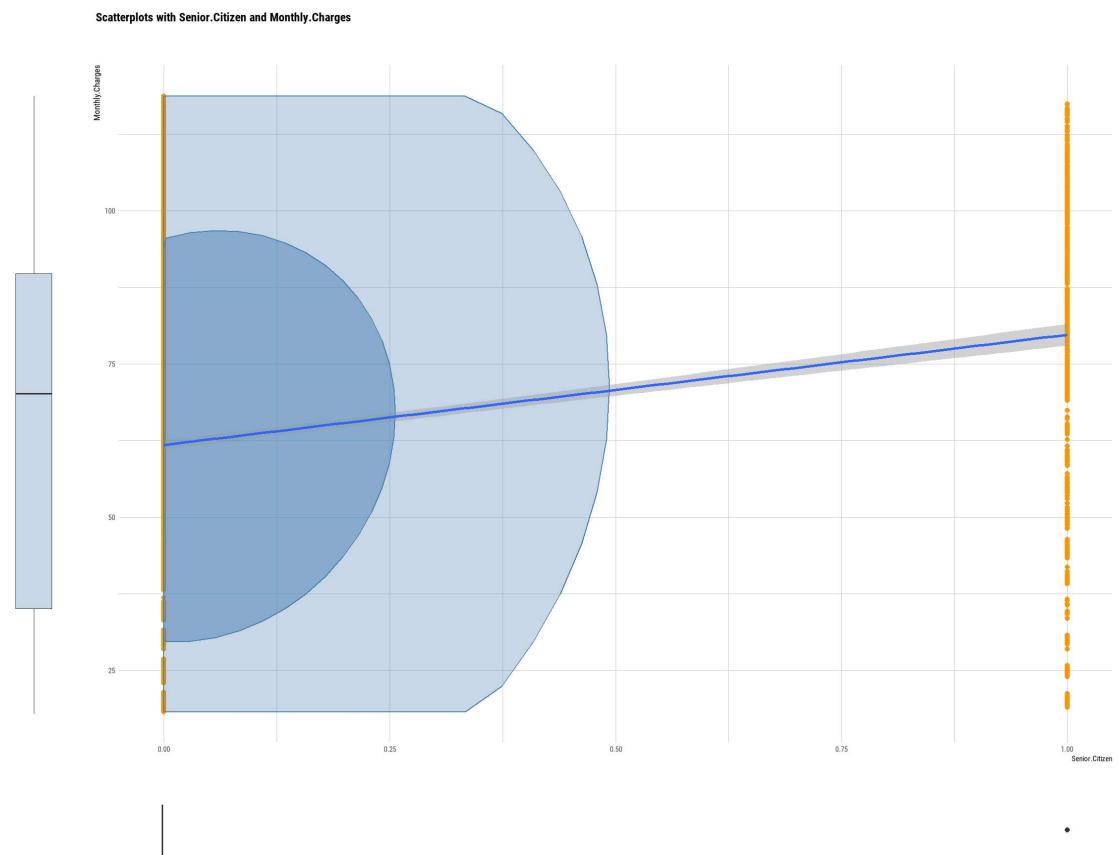
Table 7: Summary of linear model



'Senior.Citizen' vs 'Monthly.Charges'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Senior.Citizen	Monthly.Charges	0.048442	0.0482956	0.359906	330.7501	0	1

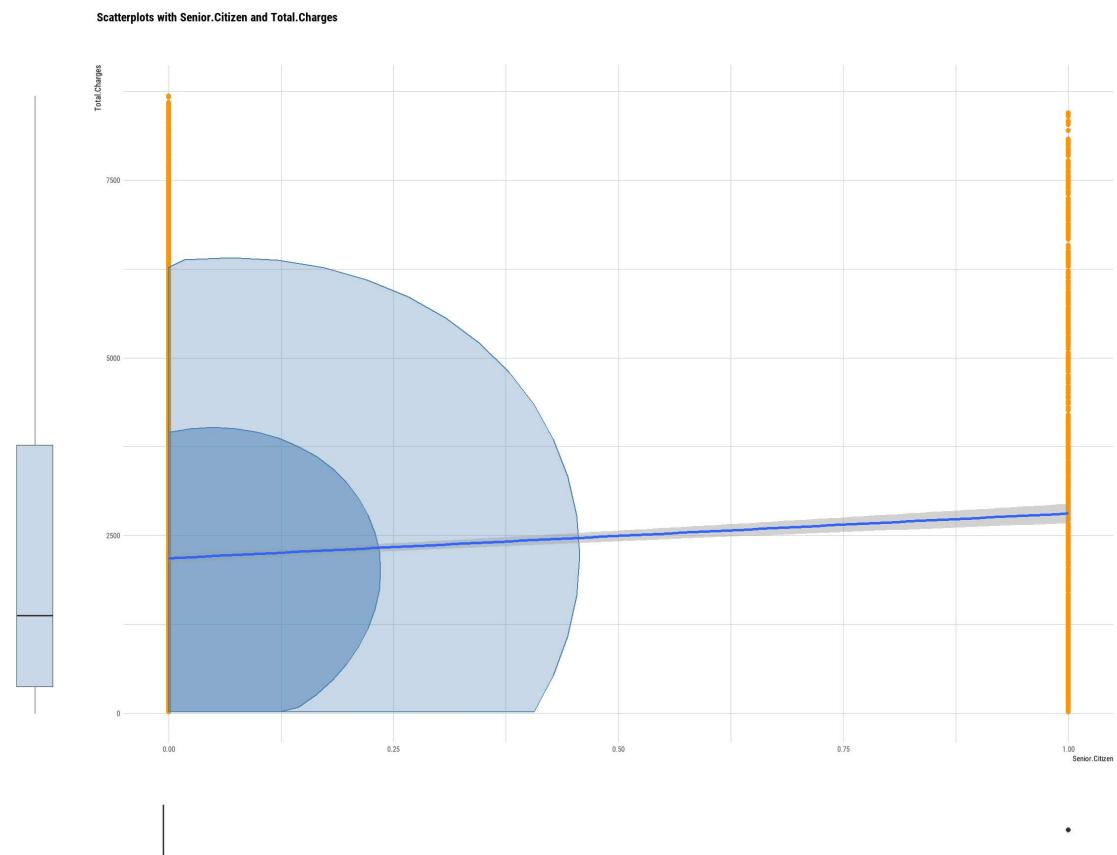
Table 7: Summary of linear model



'Senior.Citizen' vs 'Total.Charges'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Senior.Citizen	Total.Charges	0.0105191	0.0103666	0.3672127	68.97319	0	1

Table 7: Summary of linear model

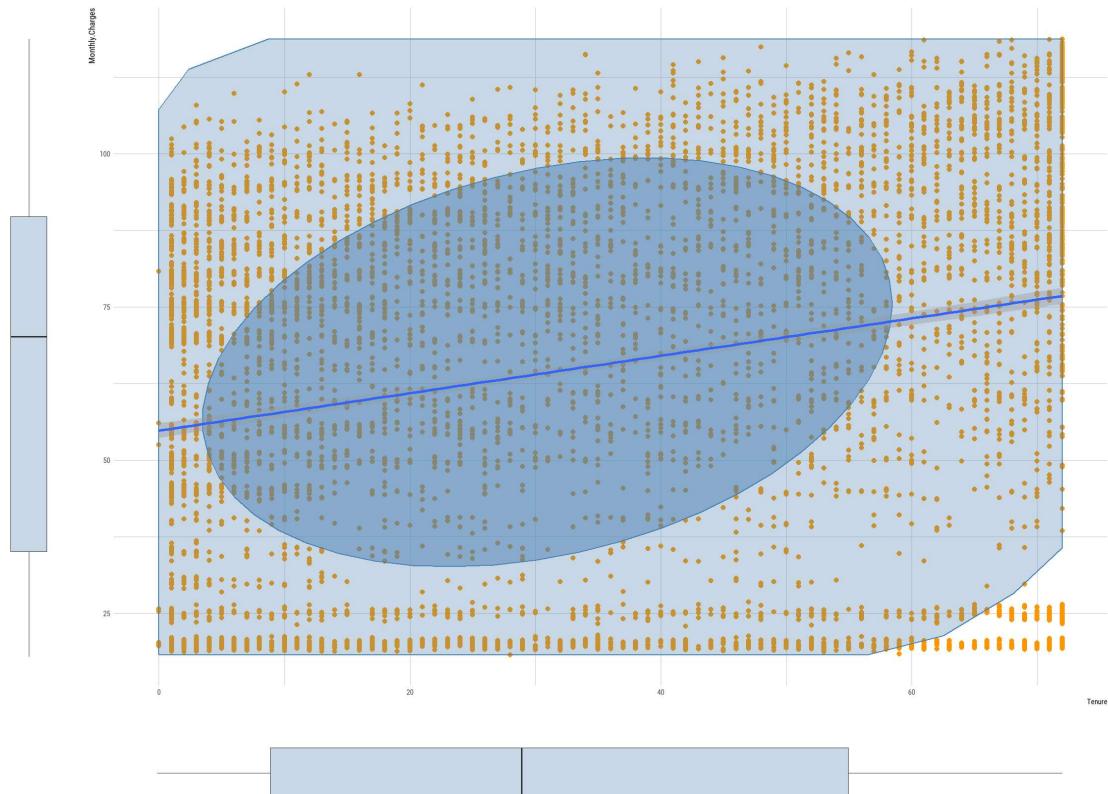


'Tenure' vs 'Monthly.Charges'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Tenure	Monthly.Charges	0.0619755	0.0618312	23.81252	429.2586	0	1

Table 7: Summary of linear model

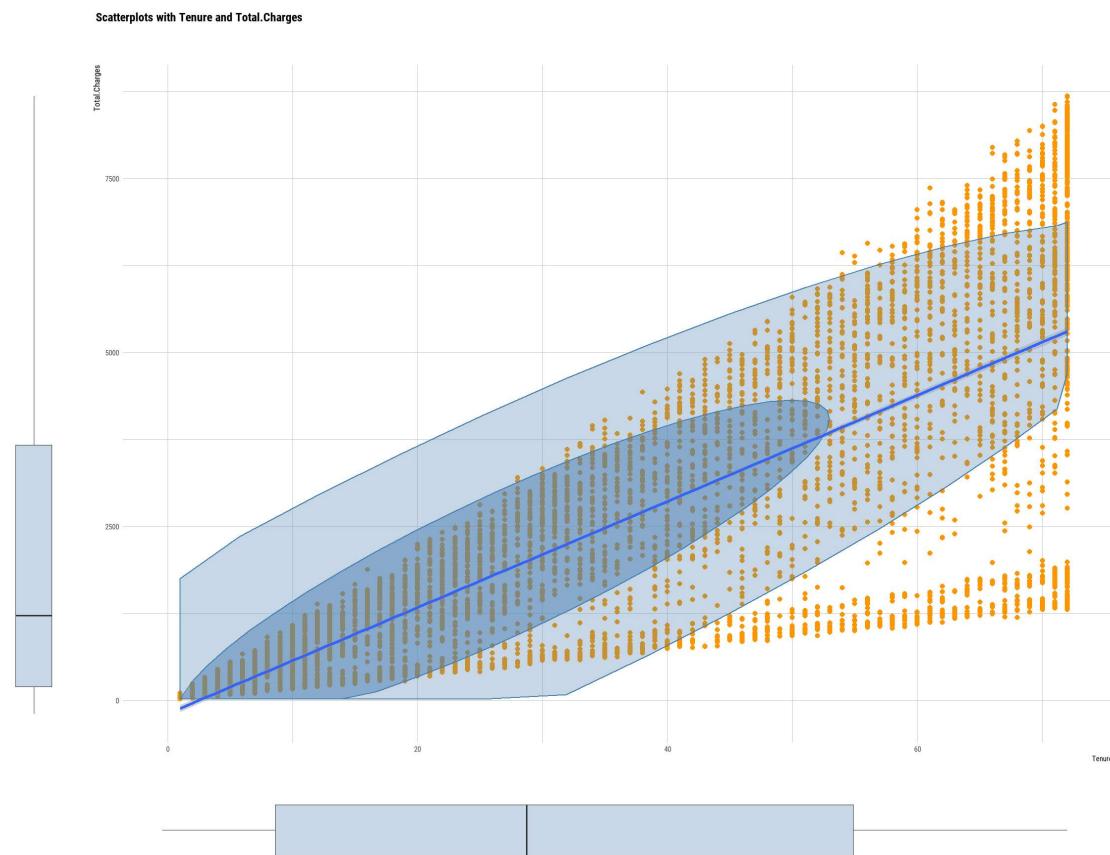
Scatterplots with Tenure and Monthly.Charges



'Tenure' vs 'Total.Charges'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Tenure	Total.Charges	0.6815394	0.6814903	13.8677	13885.01	0	1

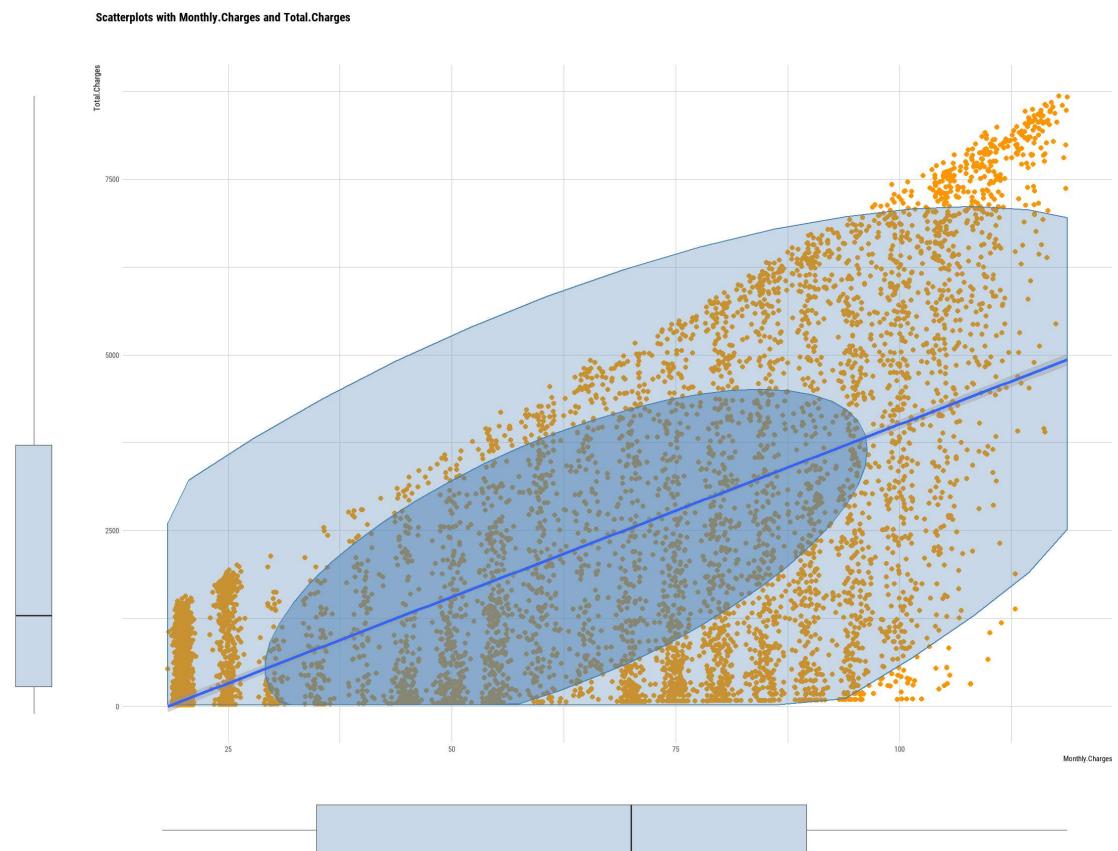
Table 7: Summary of linear model



'Monthly.Charges' vs 'Total.Charges'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Monthly.Charges	Total.Charges	0.4246978	0.4246091	22.85304	4789.55	0	1

Table 7: Summary of linear model



Compare Categorical Variables

The number of categorical variables is less than 2.

Multivariate Analysis

Correlation Analysis

Correlation Coefficient Matrix

first variable	second variable			
	Senior.Citizen	Tenure	Monthly.Charges	Total.Charges
Senior.Citizen	NA	0.016	0.220	0.103
Tenure	0.016	NA	0.249	0.826
Monthly.Charges	0.220	0.249	NA	0.652
Total.Charges	0.103	0.826	0.652	NA

Table 8: Matrix table of correlation coefficient

Correlation Plot

