Churn rate - Telco SA

A data exploratory analysis
for response the business hypothesis.

Summary

Descriptive analysis

- **Multivariate analysis**
  - Pearson correlation in numeric variables
  - Phik correlation
  - Cramer V correlation
- **Bivariate analysis**
  - Categorical barplots
- **Univariate analysis**
  - Distribution plots with numerical variables
  - Highlights
- **Response business hypothesis**
  - Churn customers who have a monthly contract are the majority !
  - Churn customers who are married are the majority!
  - Churn customers who have phone service are the majority!
  - Churn customers who do not have Device Protection are the majority!
  - Churn customers who have a Payment Method such as Electronic Check are the majority!

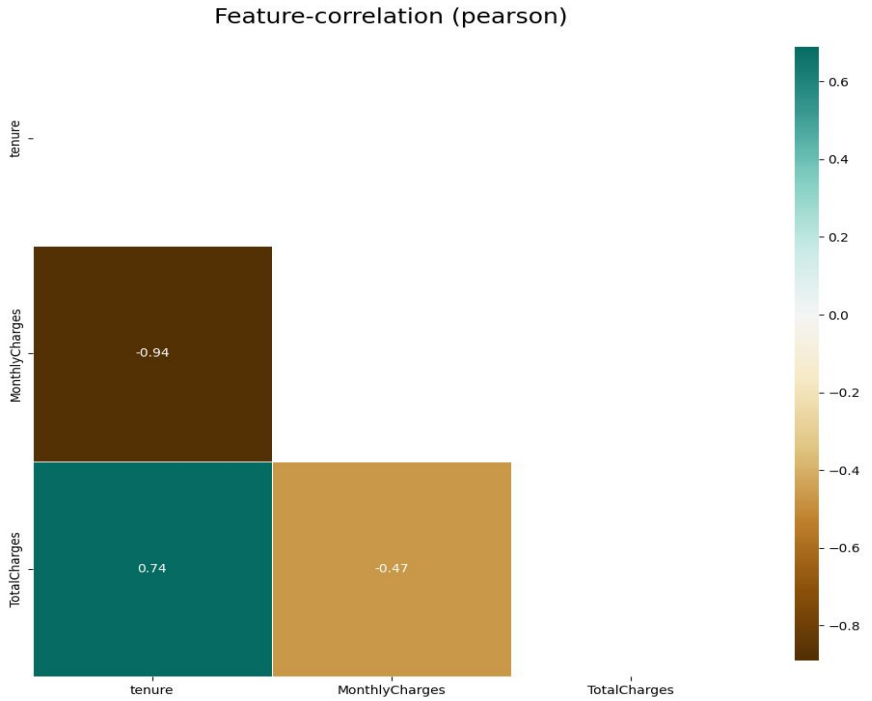
Multivariate analysis

Pearson correlation in numerical variables: In descriptive statistics, Pearson's correlation coefficient, also called "product-momentum correlation coefficient" or simply "Pearson's ρ" measures the degree of correlation between two metric scale variables. This coefficient, normally represented by ρ, only takes on values between -1 and 1.

Pearson correlation in numerical variables:

- TotalCharges is correlated with tenure, the number of tenure impact in total cust

- MonthlyCharges is correlated with TotalCharges, both motive the tenure



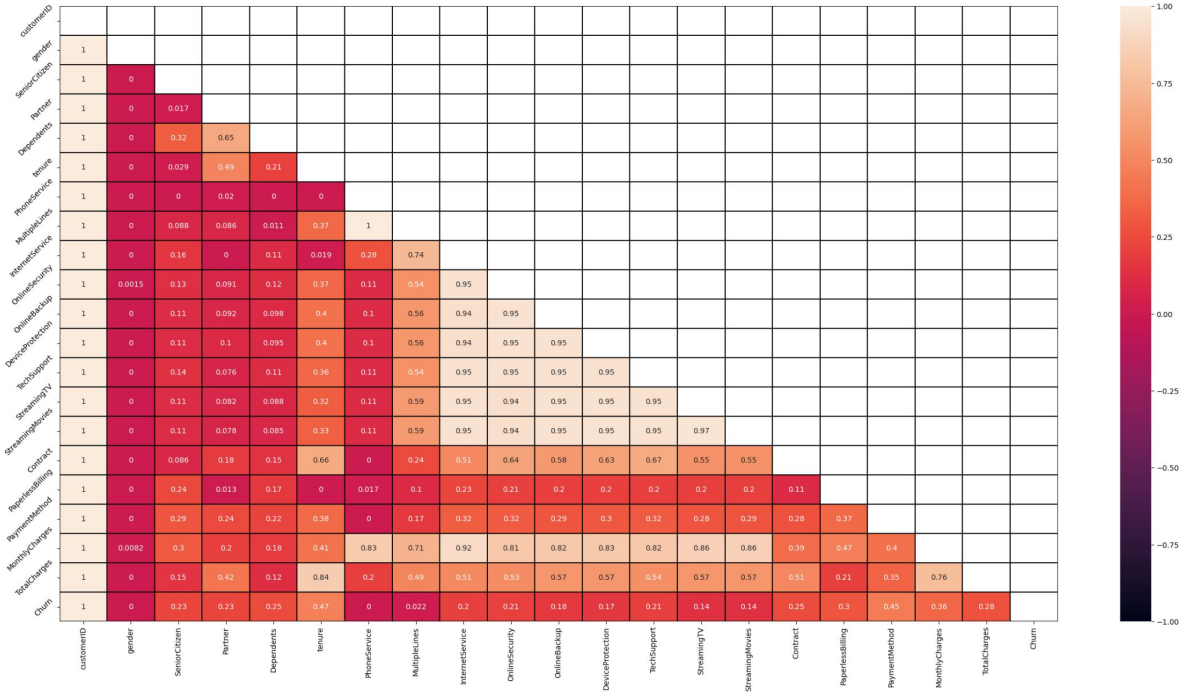
Phik correlation

- The combined features of Phi_K form an advantage over existing coefficients.

- First, it works consistently between categorical, ordinal and interval variables. Second, it captures non-linear dependency.

- Third, it reverts to the Pearson correlation coefficient in case of a bi-variate normal input distribution.

- These are useful features when studying the correlation matrix of variables with mixed types.



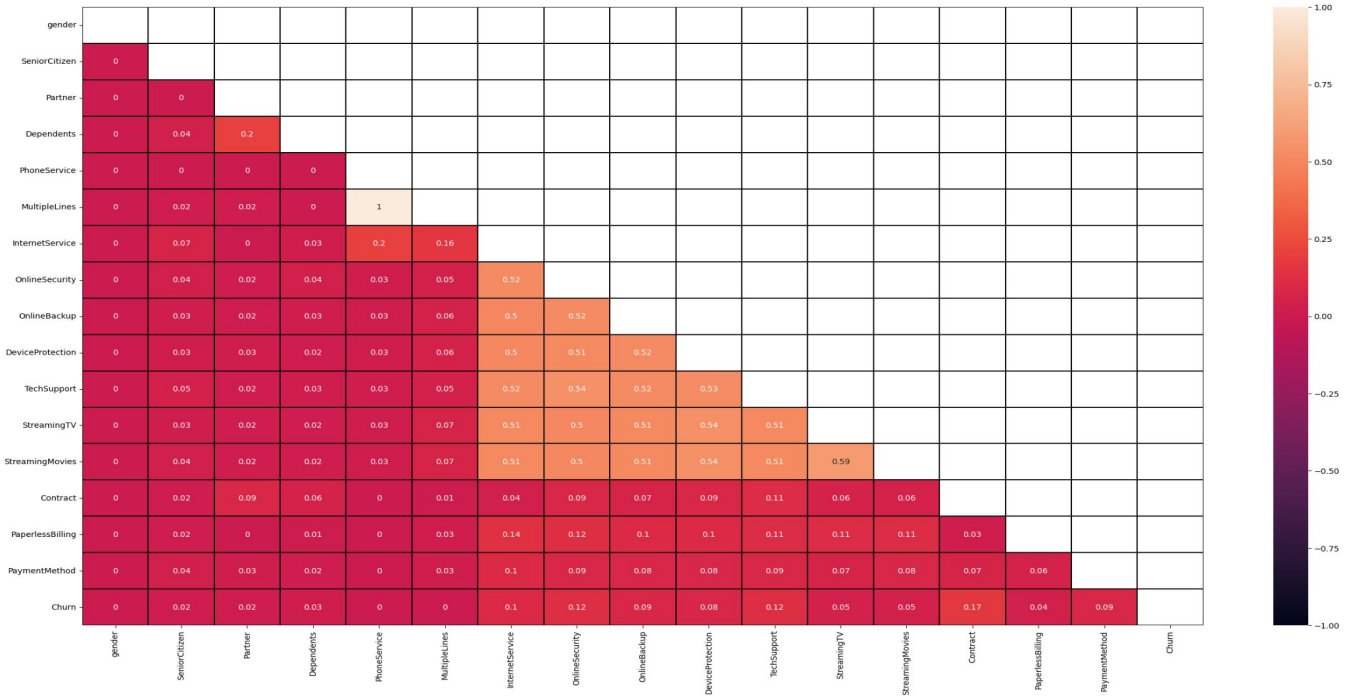
Phik correlation in all variables:

- So what: List of variables contains positive correlation with target (Churn) variable

- tenure;
- paymentmetody;
- monthlyCharges;
- paperlessbilling;
- TotalCharges;
- contract;
- Dependents;
- partner;
- SeniorCitizen;
- OnlineSecurity;
- TechSupport;
- InternetService;
- OnlineBackup;
- DeviceProtection;
- StreamingTV;
- StreamingMovies.

Cramer V analysis

Is a measure of association between two **nominal variables**, giving a value between 0 and +1 (inclusive). It is based on Pearson's chi-squared statistic



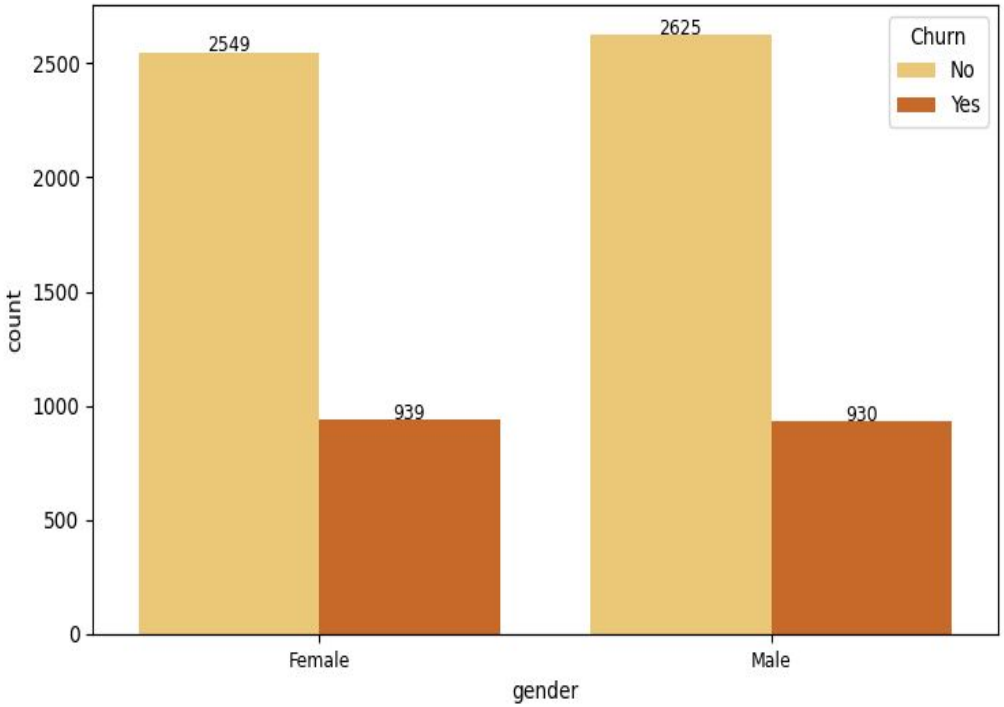
Cramer V analysis

- So what of analysis with focus on Churn variable:

- Contract haves 0.17 cramer v assosiation
- TechSupport and OnlineSecurity haves 0.12 cramer v assosiation
- PaymentMethod haves 0.09 cramer v assosiation

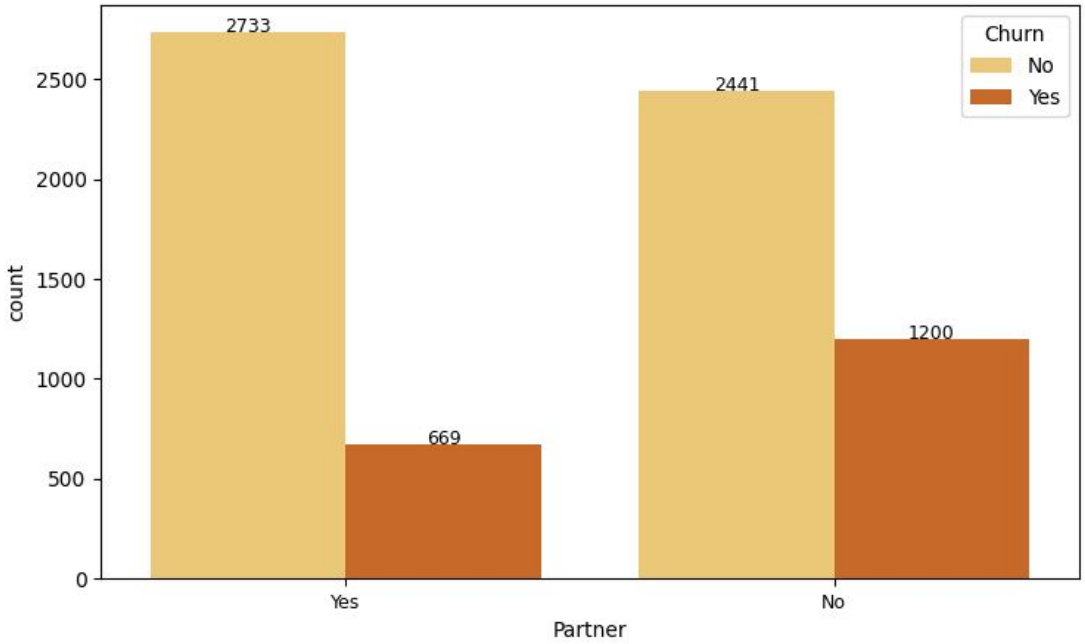
Bivariate analysis

- The churn from gender is balanced !



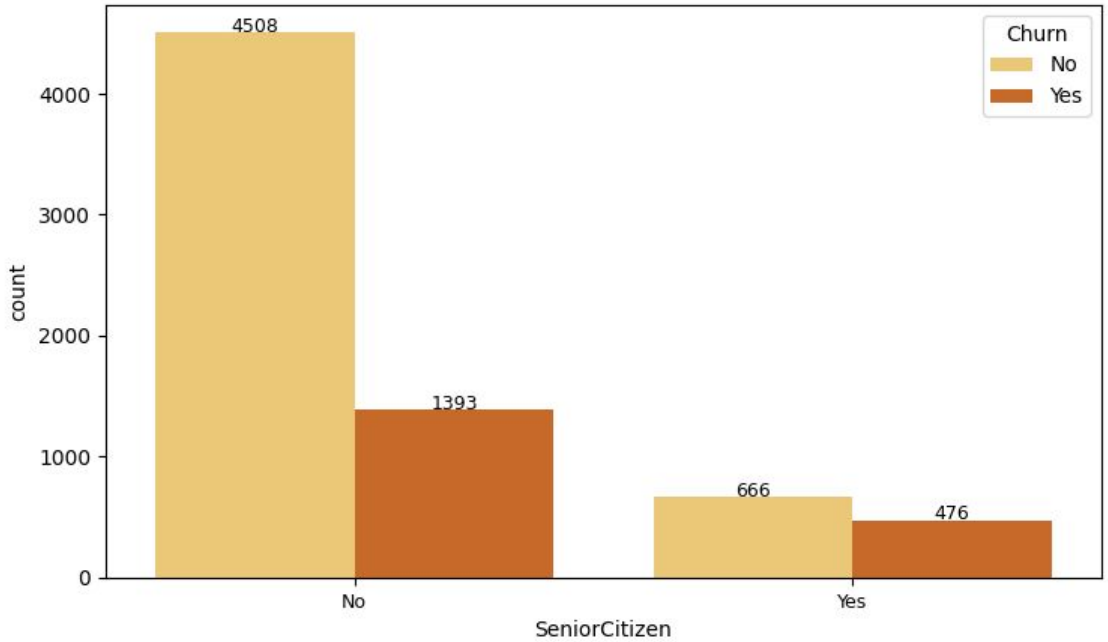
Bivariate analysis

- No partner is most frequente to Churn



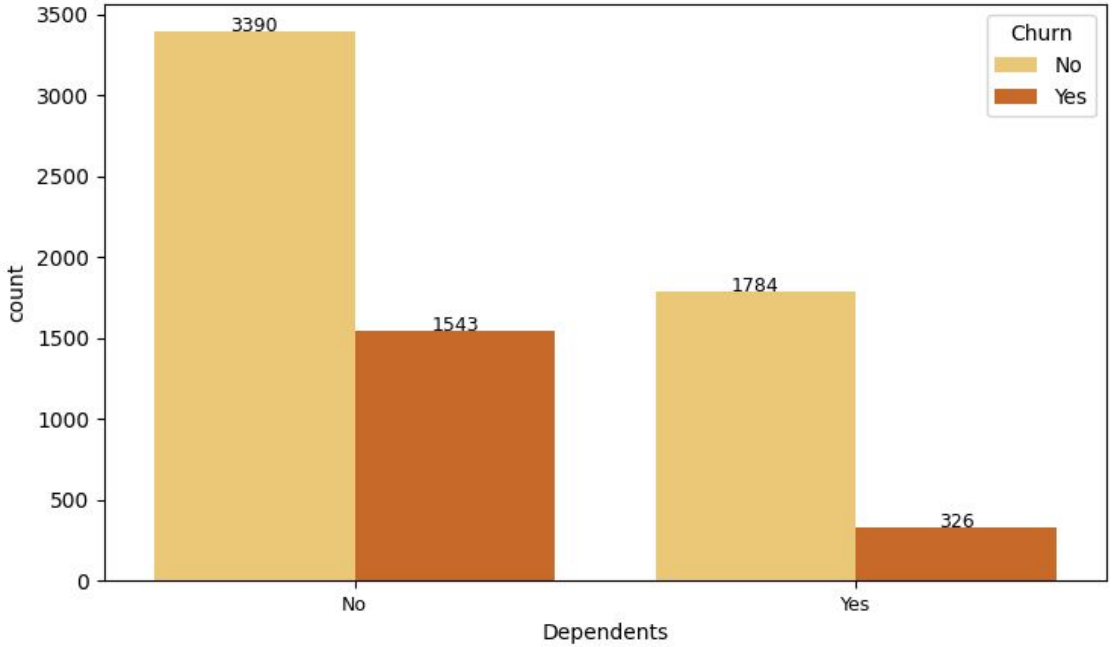
Bivariate analysis

- More frequent churn when peoples not is SeniorCitizen



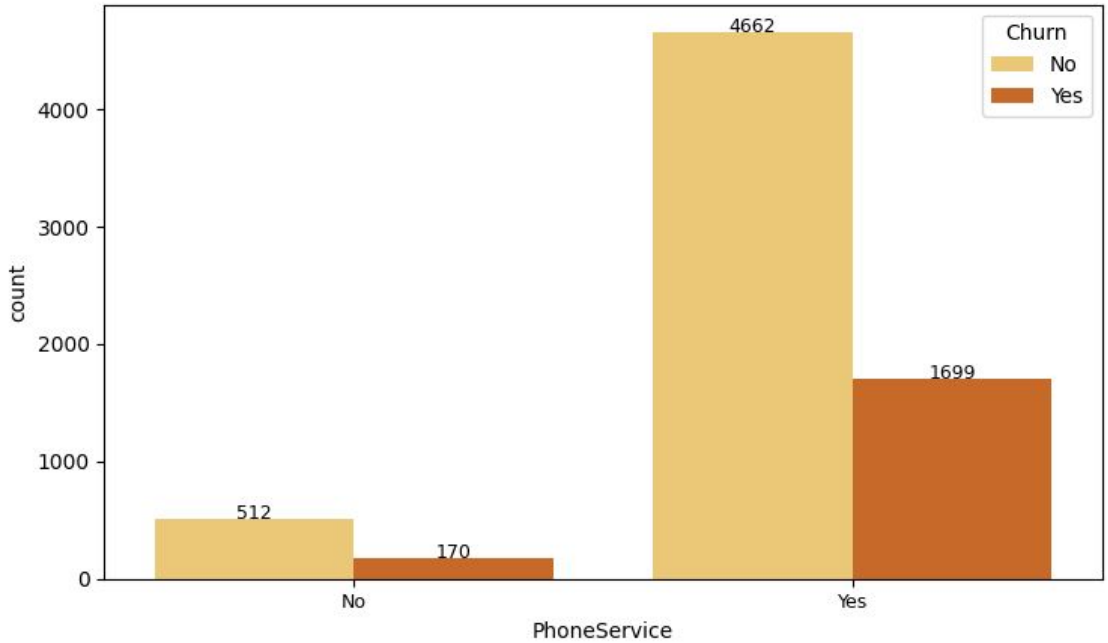
Bivariate analysis

- No Dependents is most frequent to churn



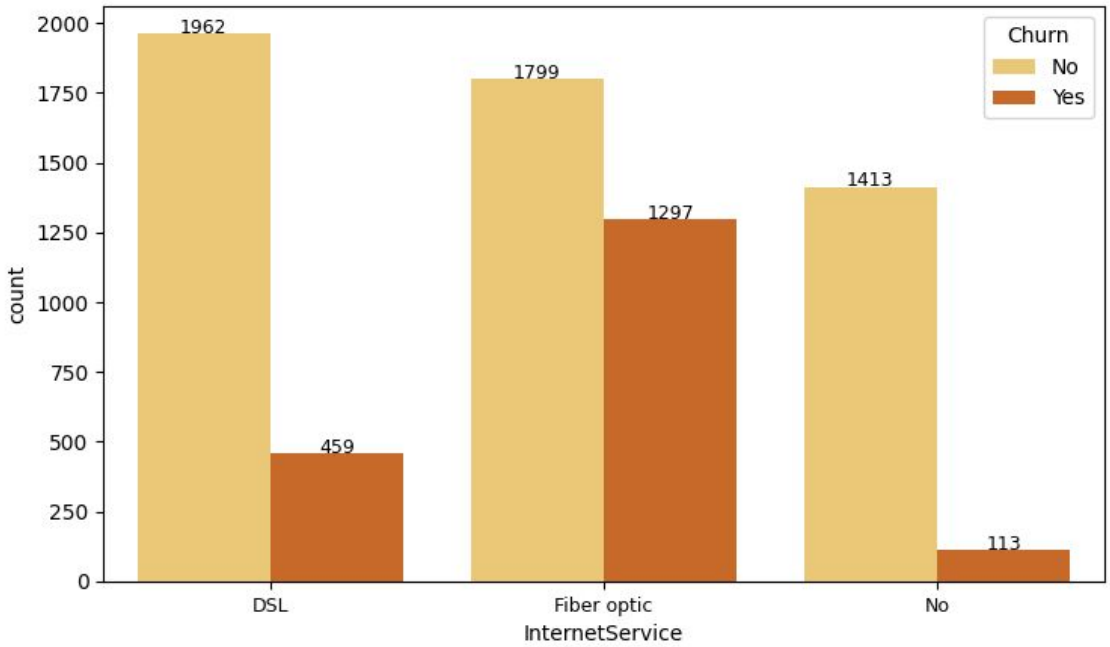
Bivariate analysis

- For those who have PhoneService, Churn is more frequent



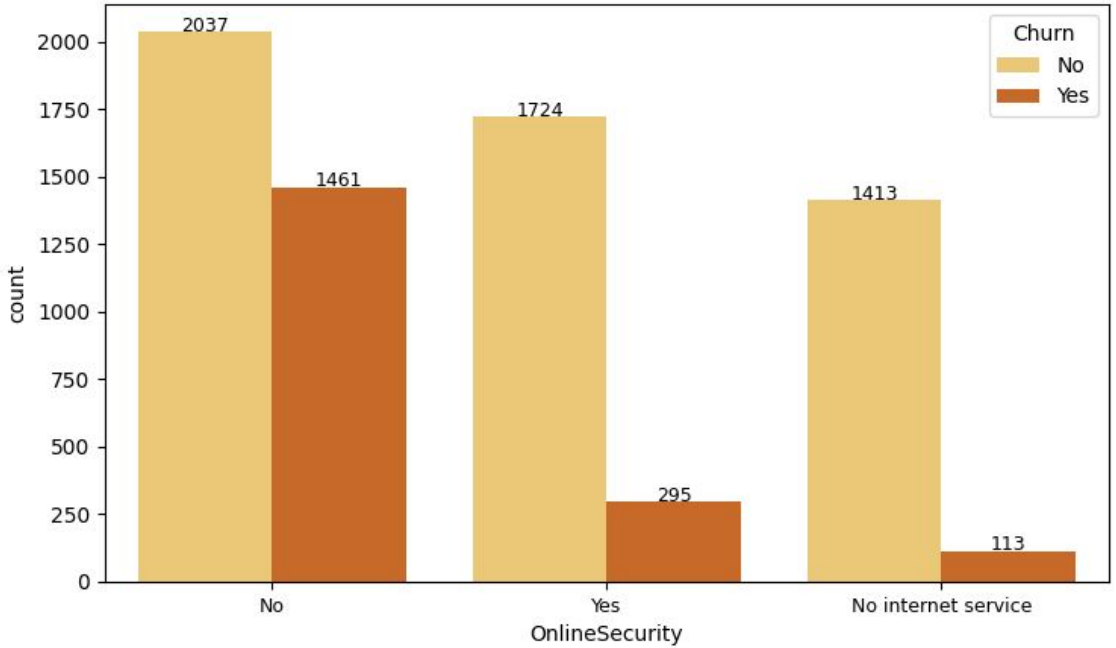
Bivariate analysis

- More frequent churn when InternetService is fiber optic



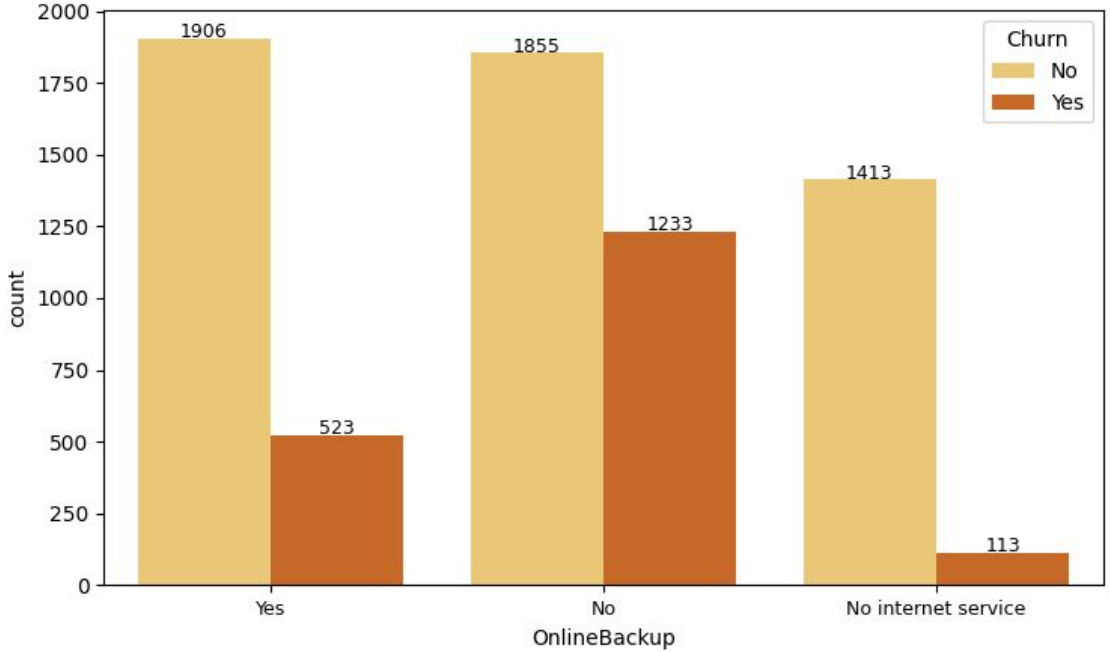
Bivariate analysis

- More frequent churn when peoples not have a OnlineSecurity



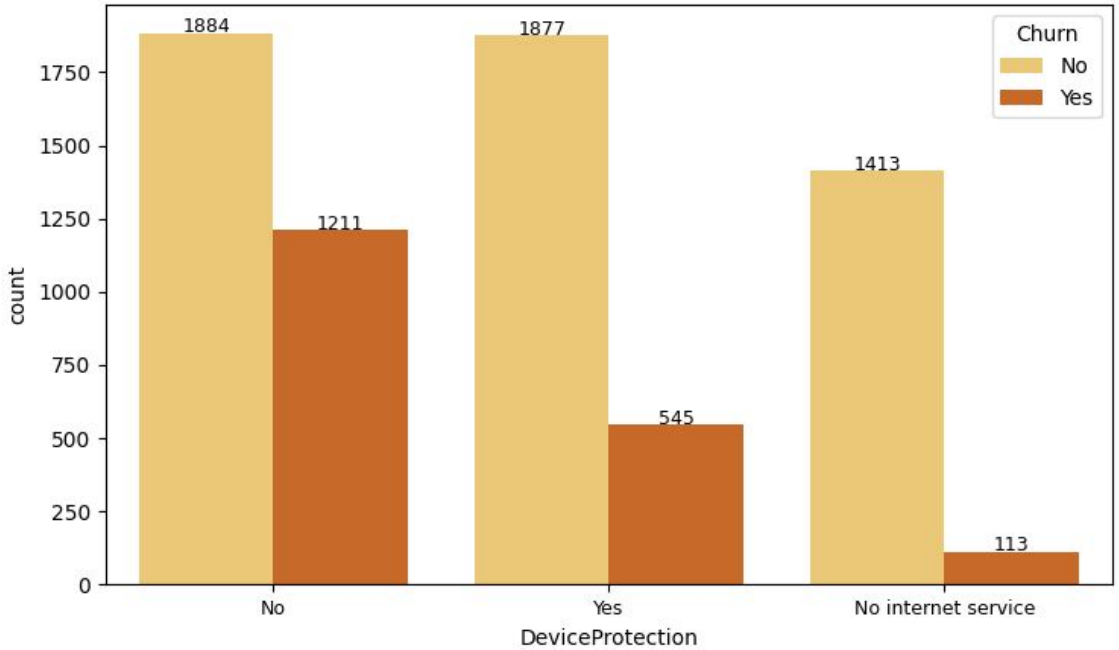
Bivariate analysis

- More frequent churn when peoples not have a OnlineBackup



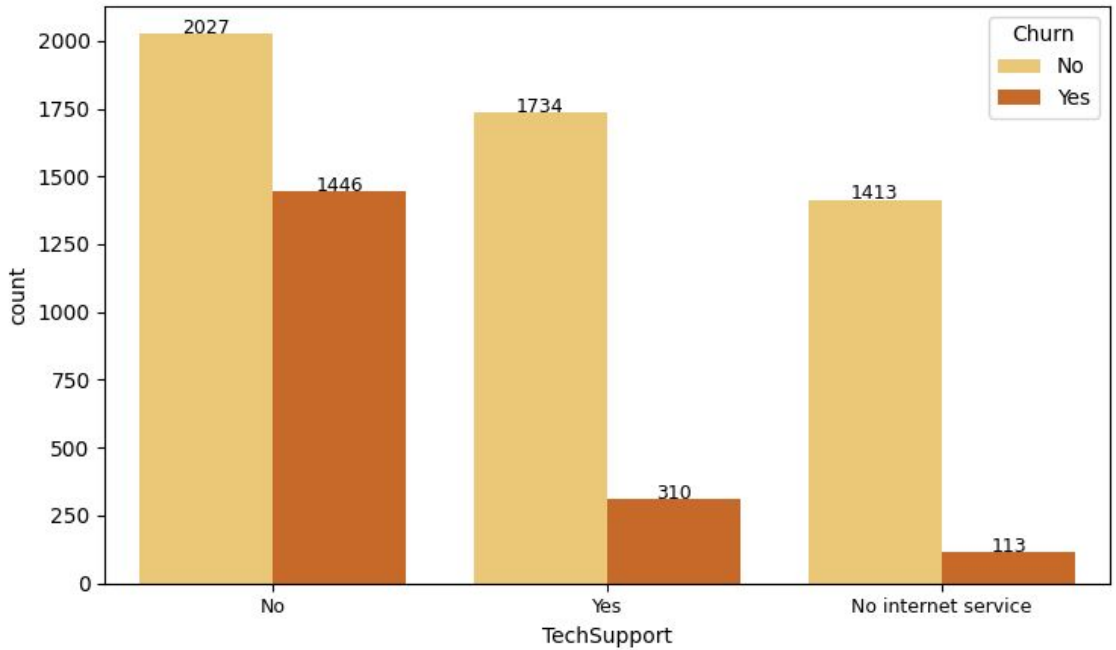
Bivariate analysis

- More frequent churn when peoples not have a DeviceProtection



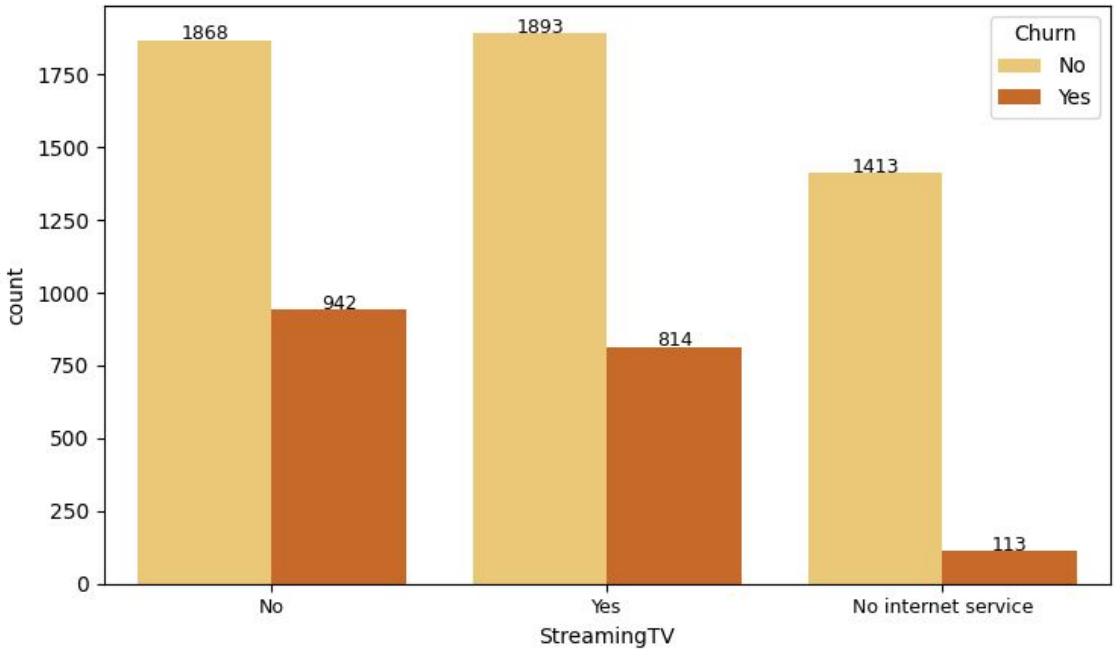
Bivariate analysis

- More frequent churn when peoples not have a TechSupport



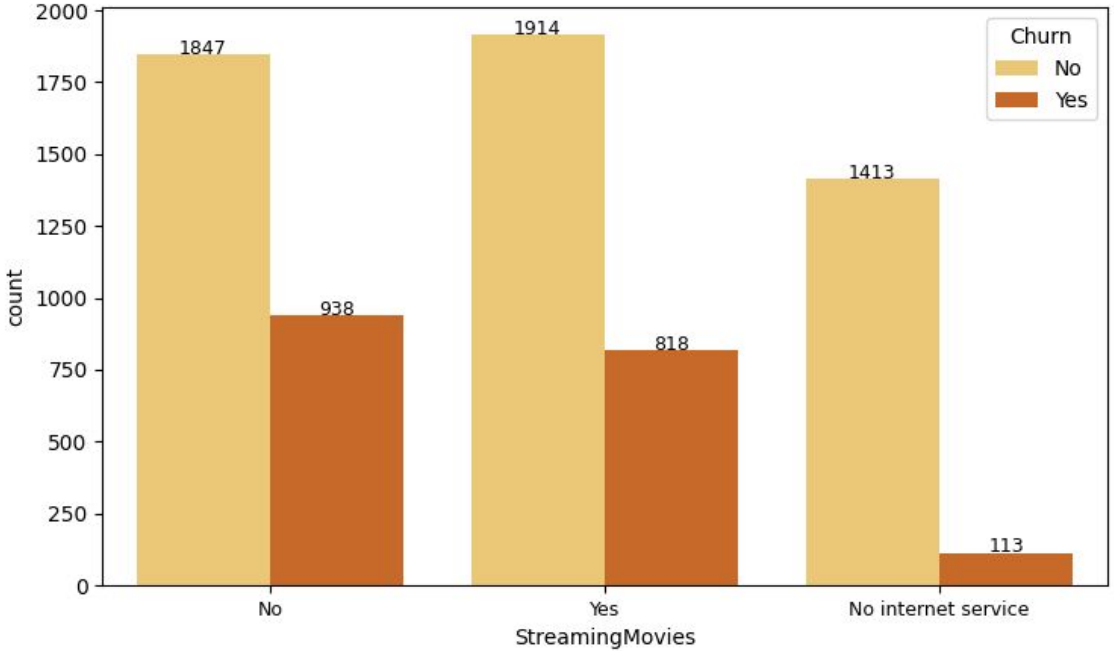
Bivariate analysis

- More frequent churn when peoples not have a StreamingTV



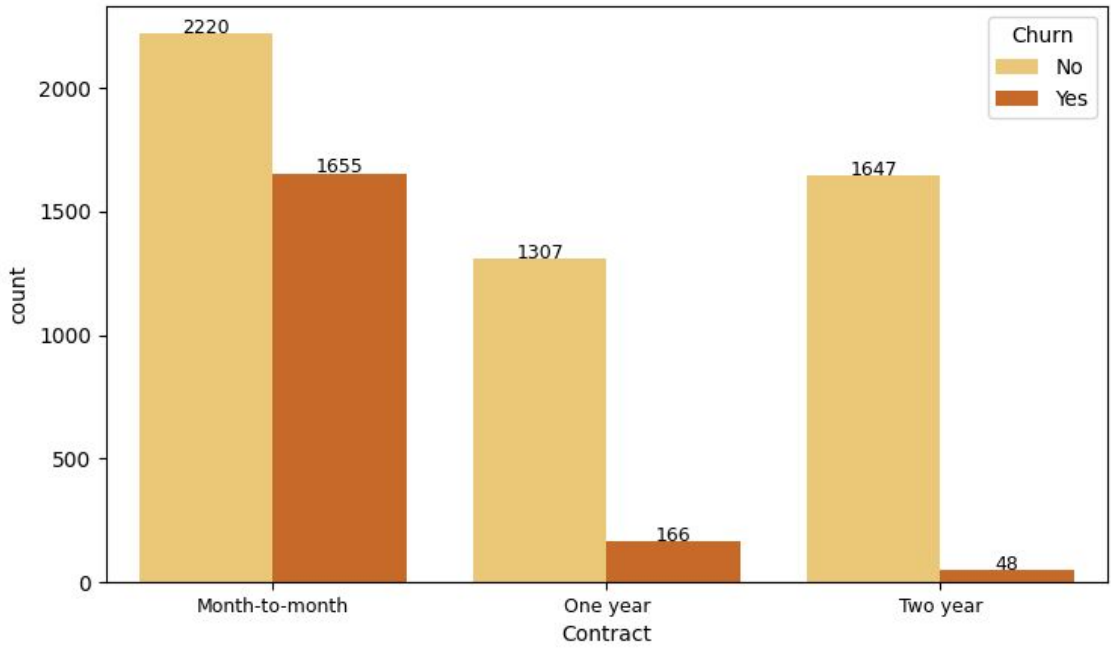
Bivariate analysis

- More frequent churn when peoples not have a StramingMovies



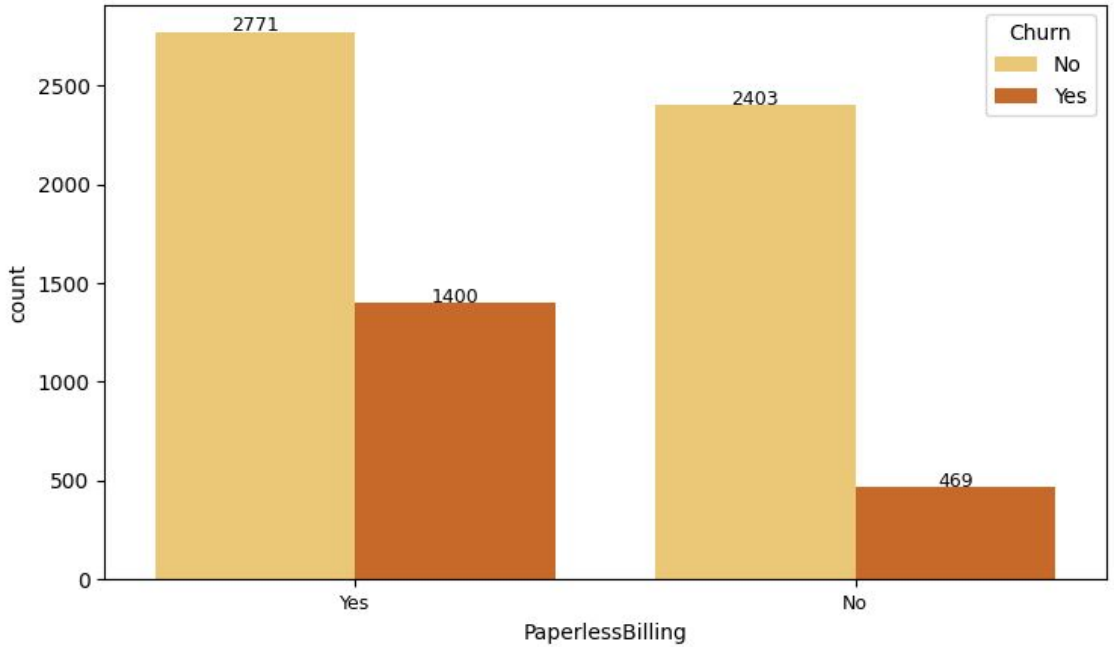
Bivariate analysis

- More frequent churn when peoples have a Contract Month-to-month



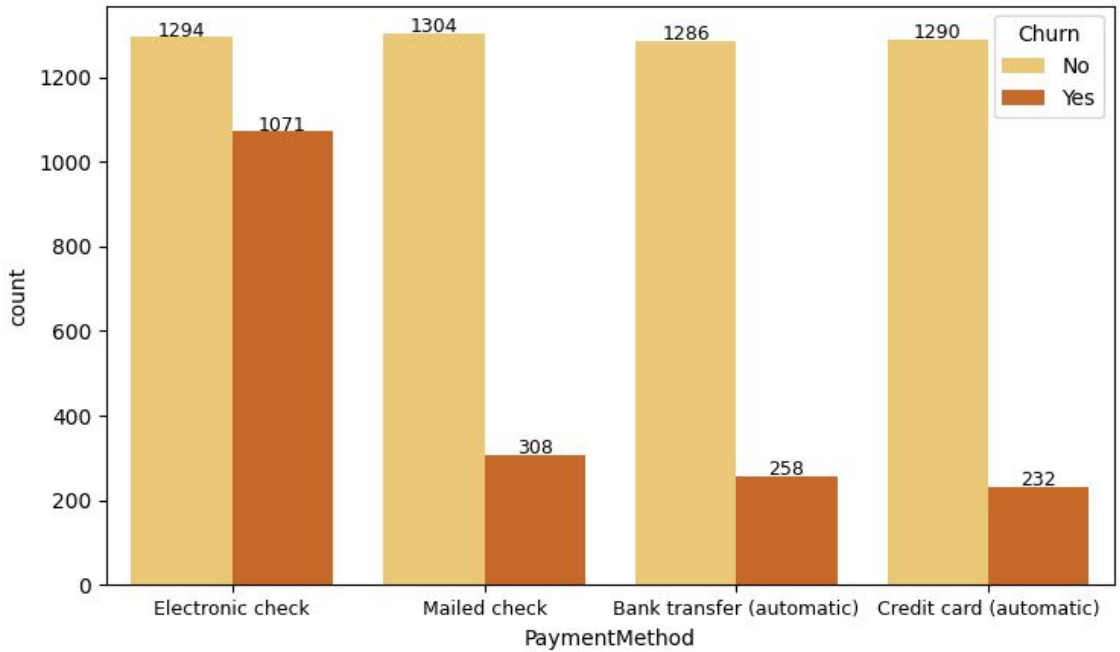
Bivariate analysis

- More frequent churn when peoples have a PaperlessBilling



Bivariate analysis

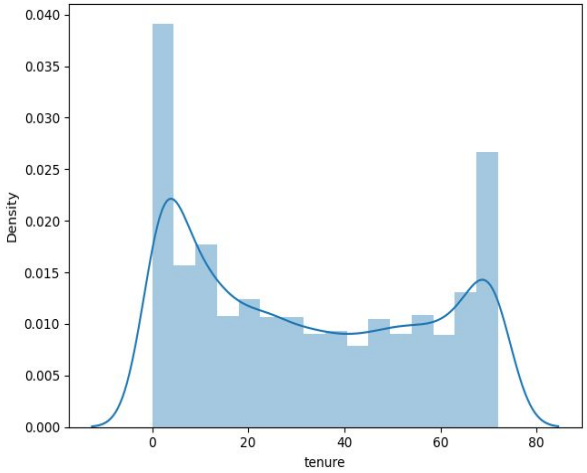
- More frequent churn when peoples have a Electronic check how PaymentMethod

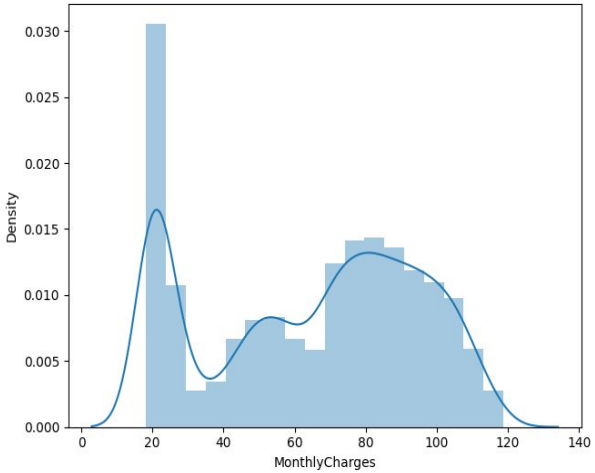


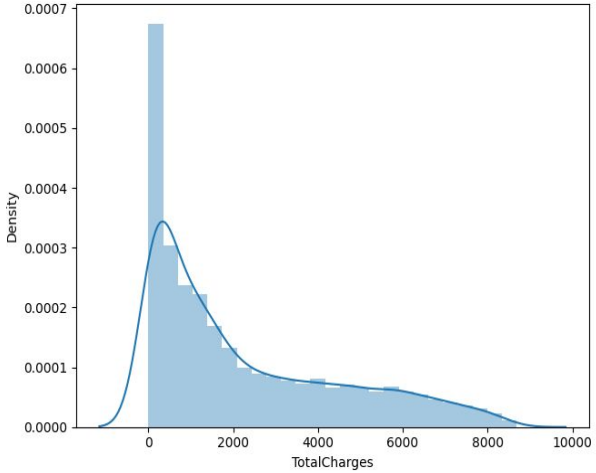
Univariate analysis

- So what numerical variables

- tenure and MonthlyCharges kind of create a bimodal distribution with peaks present at 0 - 70 and 20 - 80 respectively.
- TotalCharges displays a positively exponential distribuition and most frequent total charges are from values around 0 to 1000 dollars







Univariate analysis

- Highlights

- Churn customers are a minority and therefore comprise a smaller number of products in possession
- Churn customers have a higher average cost of monthly charges
- Churn customers have a lower total cost, but their minimum cost exceeds that of non-churn customers

	Churn	No	Yes
tenure	count	5174.000000	1869.000000
	mean	37.569965	17.979133
	std	24.113777	19.531123
	min	0.000000	1.000000
	25%	15.000000	2.000000
	50%	38.000000	10.000000
	75%	61.000000	29.000000
	max	72.000000	72.000000

MonthlyCharges	count	5174.000000	1869.000000
	mean	61.265124	74.441332
	std	31.092648	24.666053
	min	18.250000	18.850000
	25%	25.100000	56.150000
	50%	64.425000	79.650000
	75%	88.400000	94.200000
	max	118.750000	118.350000

TotalCharges	count	5174.000000	1869.000000
	mean	2549.911442	1531.796094
	std	2329.954215	1890.822994
	min	0.000000	18.850000
	25%	572.900000	134.500000
	50%	1679.525000	703.550000
	75%	4262.850000	2331.300000
	max	8672.450000	8684.800000

Response business hypothesis

Churn customers who have a monthly contract are the majority !

- **Yes, true hypothese ! Month-to-month contract represent 88% of churn !**

		Contract
Churn	**Contract**	
No	Month-to-month	0.429068
	Two year	0.318322
	One year	0.252609
Yes	Month-to-month	0.885500
	One year	0.088818
	Two year	0.025682

Response business hypothesis

Churn customers who are married are the majority!

- **No, false hypothese ! the Partner represent 35% of Churn**

		Partner
Churn	**Partner**	
No	Yes	0.528218
	No	0.471782
Yes	No	0.642055
	Yes	0.357945

Response business hypothesis

Churn customers who have phone service are the majority!

- Yes, true ! the customer churn have phone service represent 90%

		PhoneService
Churn	PhoneService	
No	Yes	0.901044
	No	0.098956
Yes	Yes	0.909042
	No	0.090958

Response business hypothesis

Churn customers who do not have Device Protection are the majority!

- **Yes, true ! Churn customers who do not have Device Protection represent 64%**

		DeviceProtection
Churn	**DeviceProtection**	
No	No	0.364128
	Yes	0.362775
	No internet service	0.273096
Yes	No	0.647940
	Yes	0.291600
	No internet service	0.060460

Response business hypothesis

Churn customers who have a Payment Method such as Electronic Check are the majority!

- **Yes, true ! Churn customers who have a Payment Method such as Electronic Check are the majority represent 57%**

		PaymentMethod
Churn	**PaymentMethod**	
No	Mailed check	0.252029
	Electronic check	0.250097
	Credit card (automatic)	0.249324
	Bank transfer (automatic)	0.248550
Yes	Electronic check	0.573034
	Mailed check	0.164794
	Bank transfer (automatic)	0.138042
	Credit card (automatic)	0.124131

Churn rate - Telco SA

End first exploratory analysis.