

Customer Churn Analysis

Aheer Srabon

2025-08-31

Table of contents

1	Uni-variate & Distributional Analysis	2
1.1	Numeric Distributions (Histograms)	2
1.2	Categorical Counts (Phone/Internet/Contract)	3
1.3	Categorical Proportions by Churn	3
1.4	Missingness Overview	4
1.5	Skewness	4
1.5.1	Quick skew check (numeric)	4
1.5.2	Skew check	5
1.6	Churn vs Non-Churn Proportions	6
1.6.1	Stacked bar (counts by category)	7
1.6.2	Normalized bar (proportions within each category)	8
2	Bivariate (Churn vs. Features)	9
2.1	Churn rate by Tenure bins (0-6, 6-12, 12-24, 24-48, 48+)	9
2.2	Contract Type vs. Churn	10
2.3	Internet Service vs. Churn	11
2.4	Numeric Densities by Churn	11
2.5	Numeric Box Plots by Churn	12
2.6	Numeric Violin Plots by Churn	12
3	Multivariate Relationships	13
3.1	Age x Tenure x Churn	13
3.2	MonthlyCharges x Contract x Churn	14
3.3	InternetService x Contract (churn risk heatmap)	15
4	Feature Engineering for EDA	16
4.1	Customer lifetime value proxy	16
4.2	At-risk cohorts (cohort flagging)	16
5	Correlation & Association	18
5.1	Numeric-only correlation matrix	18
5.2	Chi-square + Cramer's V for categorical vs. Churn	18
5.3	Mutual information (any type \rightarrow Churn)	18
6	Survival-style Analysis	20

1 Uni-variate & Distributional Analysis

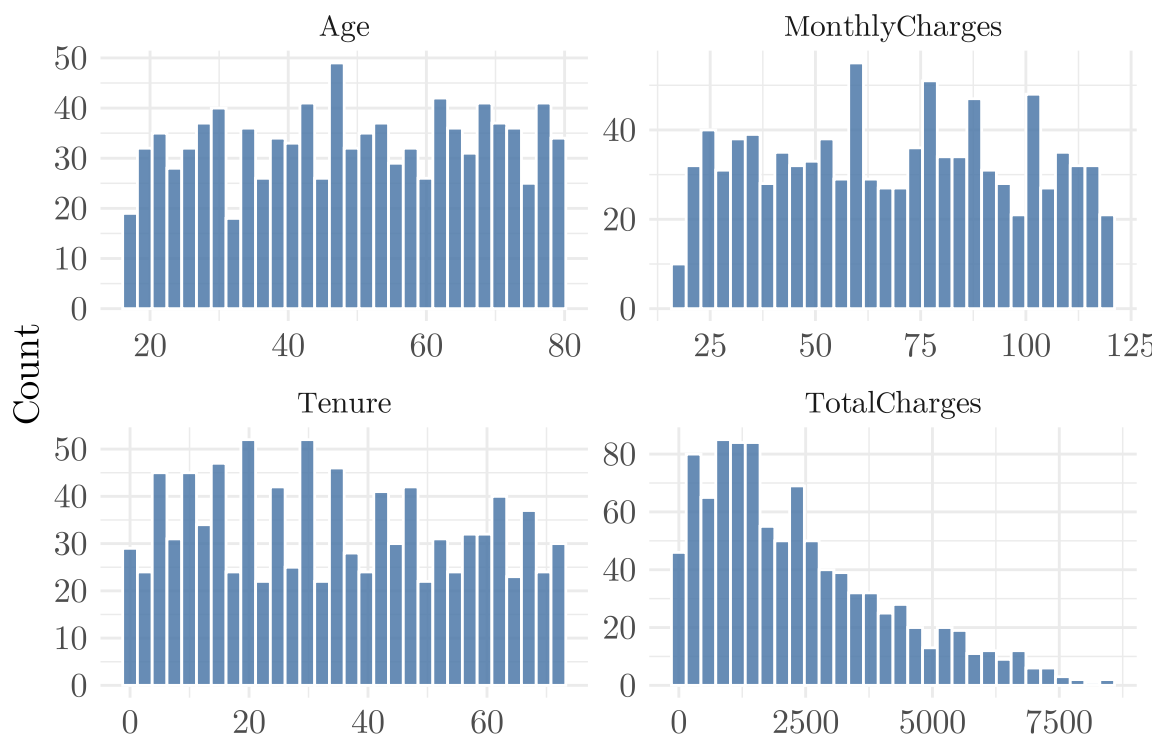
Table 1.1: Summary of the Dataset

Variable	Mean	Median	Sd	Min	Q1	Q2	Q3	Q4	Max
Age	49.09	50.00	18.17	18	34.00	50.00	65.00	80.00	80.00
Tenure	34.68	34.00	21.04	0	16.00	34.00	52.25	72.00	72.00
MonthlyCharges	68.51	69.02	29.07	18	43.74	69.02	92.65	119.77	119.77
TotalCharges	2339.68	1900.12	1808.26	0	938.64	1900.12	3380.87	8444.03	8444.03

Table 1.2: Random 10 samples from the dataset

Gender	Age	Tenure	PhoneService	InternetService	Contract	MonthlyCharges	TotalCharges	Churn
Male	51	59	Yes	DSL	One year	20.30	1197.70	No
Male	60	37	No	None	One year	89.91	3326.67	No
Male	74	4	Yes	DSL	Two years	23.14	92.56	No
Male	34	35	No	DSL	Two years	82.41	2884.35	No
Female	19	57	No	None	Month-to-month	108.82	6202.74	Yes
Female	49	20	Yes	None	Two years	102.72	2054.40	No
Female	79	45	Yes	Fiber optic	Two years	119.24	5365.80	No
Female	53	57	Yes	Fiber optic	Month-to-month	81.52	4646.64	No
Male	45	26	No	Fiber optic	One year	85.50	2223.00	No
Male	34	14	No	DSL	Month-to-month	79.29	1110.06	No

1.1 Numeric Distributions (Histograms)



1.2 Categorical Counts (Phone/Internet/Contract)

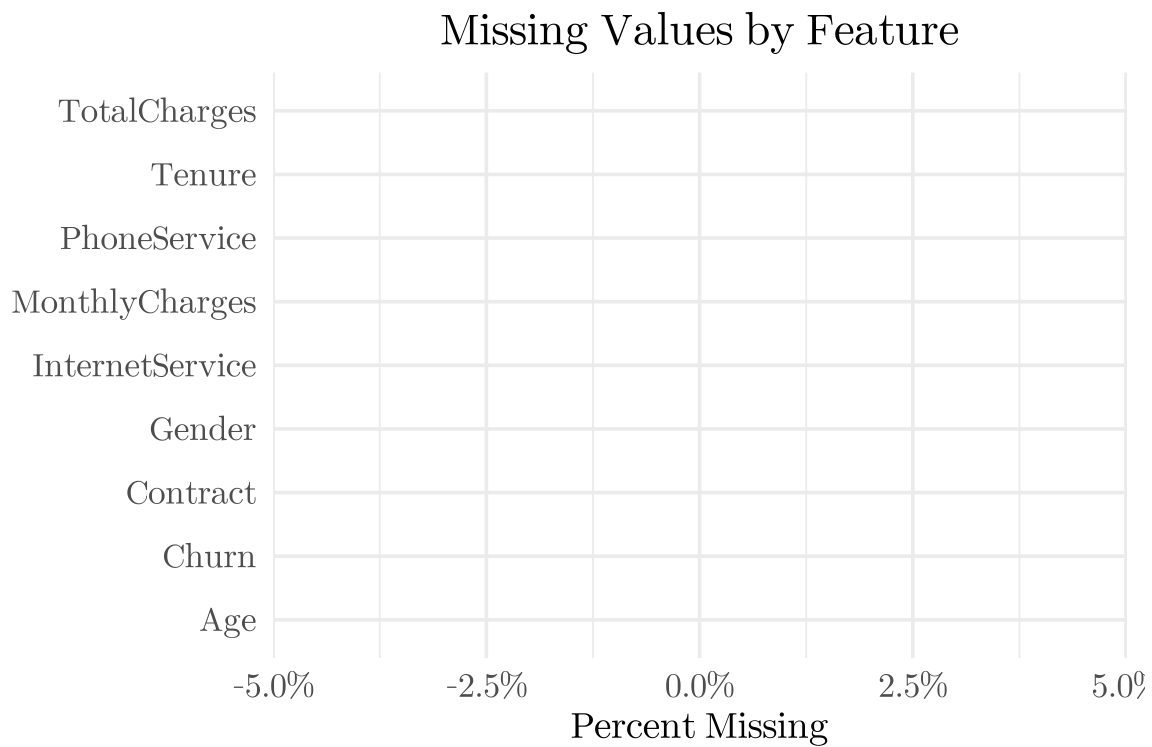


1.3 Categorical Proportions by Churn



1.4 Missingness Overview

As can be seen from the following table, there is no missing data.



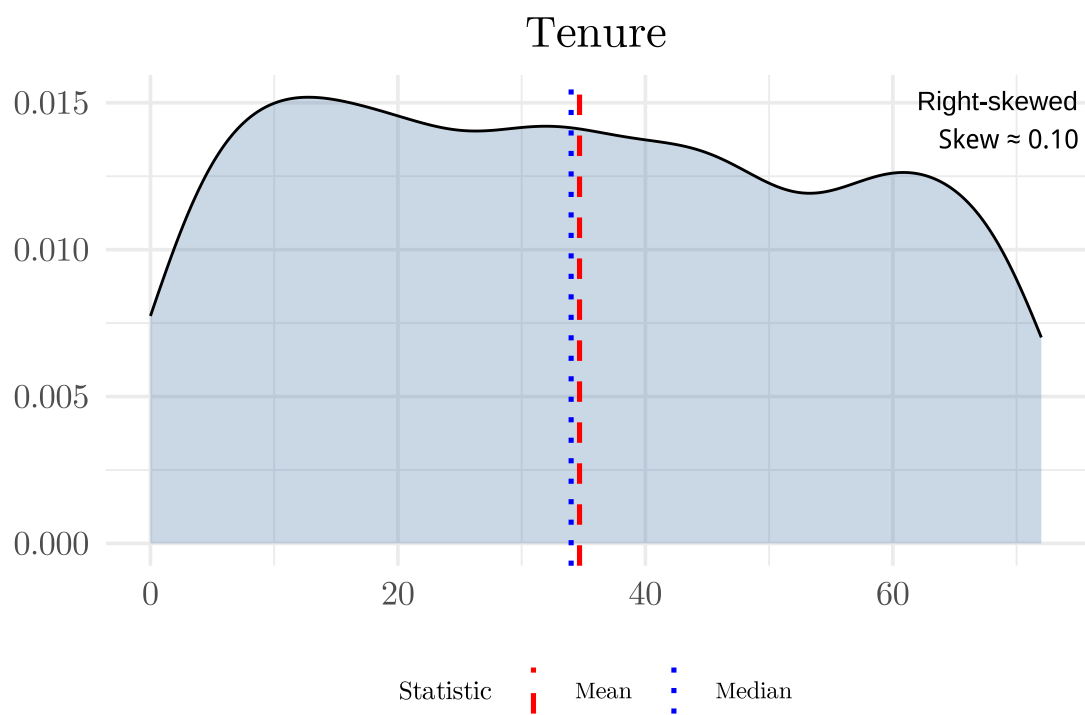
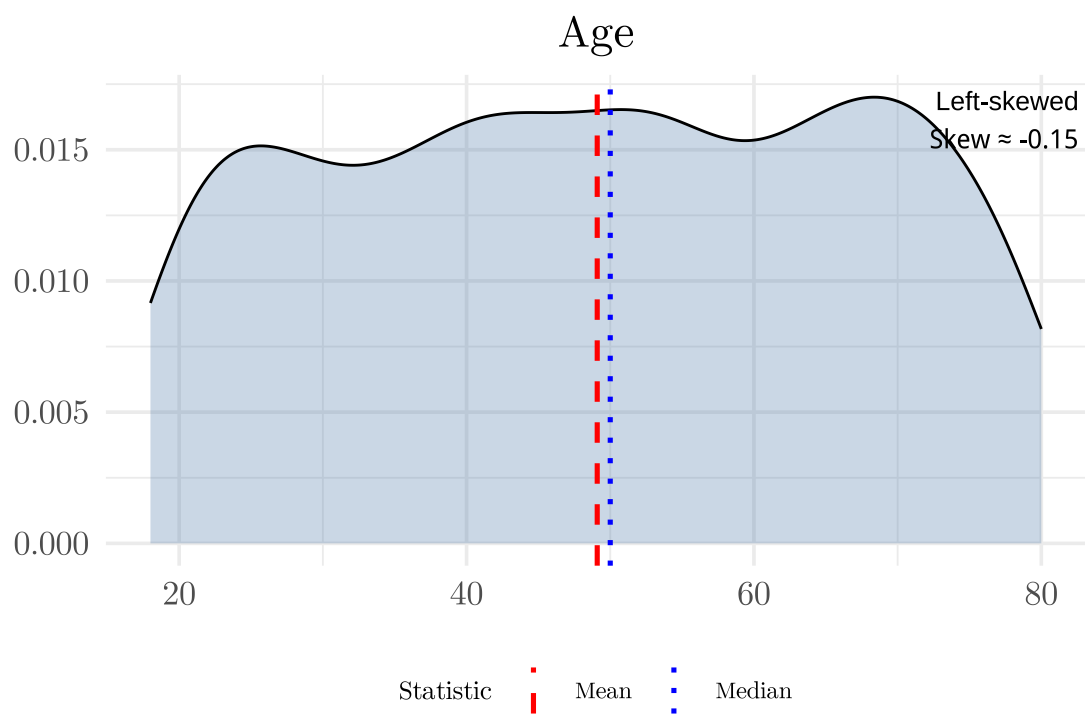
1.5 Skewness

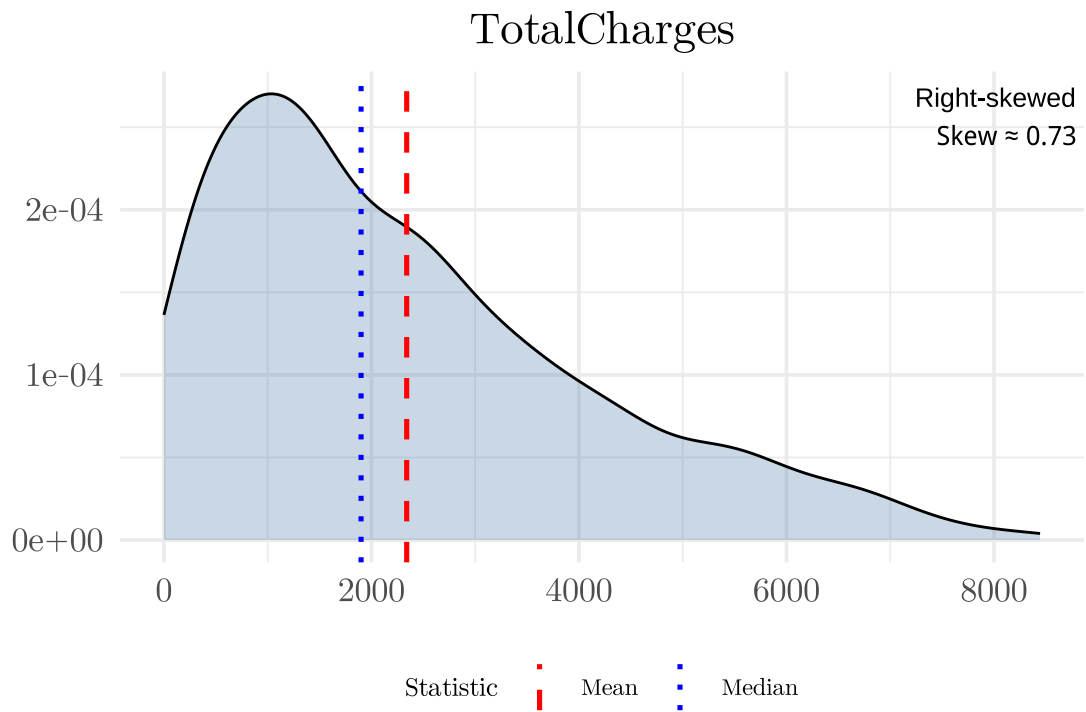
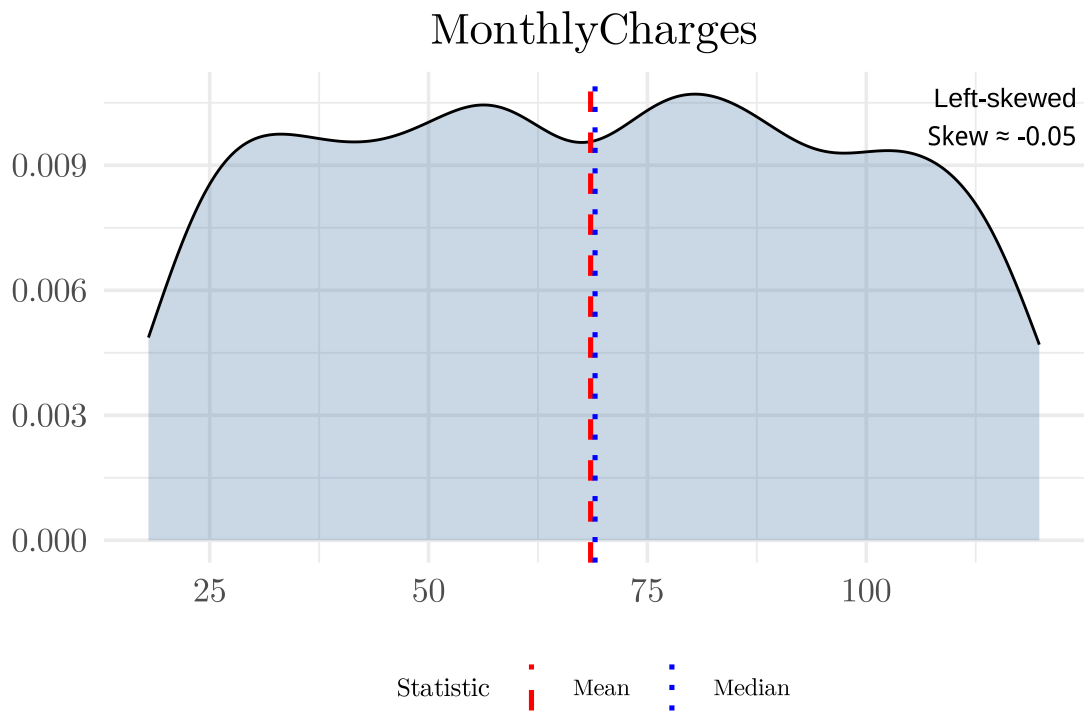
1.5.1 Quick skew check (numeric)

Table 1.3: Skewness

feature	n	mean	median	sd	p99	skew_hint
TotalCharges	1000	2339.68	1900.12	1808.26	7247.88	Right-skewed
Age	1000	49.09	50.00	18.17	80.00	Left-skewed
Tenure	1000	34.68	34.00	21.04	72.00	Right-skewed
MonthlyCharges	1000	68.51	69.02	29.07	118.93	Left-skewed

1.5.2 Skew check





1.6 Churn vs Non-Churn Proportions

Table 1.4: Categorical features of interest

Churn	Feature	Level
No	Gender	Male

Table 1.4: Categorical features of interest (continued)

Churn	Feature	Level
No	Gender	Female
No	Gender	Male
No	PhoneService	No
No	PhoneService	No

1.6.1 Stacked bar (counts by category)

Categorical Balances by Churn (Counts)



1.6.2 Normalized bar (proportions within each category)

Categorical Balances by Churn (Proportions)



2 Bivariate (Churn vs. Features)

2.1 Churn rate by **Tenure** bins (0-6, 6-12, 12-24, 24-48, 48+)

Tenure is how long a customer has been with the company (in months). Following is the hypothesis, Customer who are new (short tenure) are more likely to churn; customers who have stayed longer are “stickier”

To test this,

1. Group customers into bins of tenure (0-6 months, 6-12, etc.).
2. Compute churn rate = (# churned customers) / (total customers) in each bin.
3. Plot churn rate across bins to visualize patterns.

If churn is very high in 0-6 months bin, that means the company has an onboarding / early experience problem.

Table 2.1: Data Frame with Tenure Bin tags

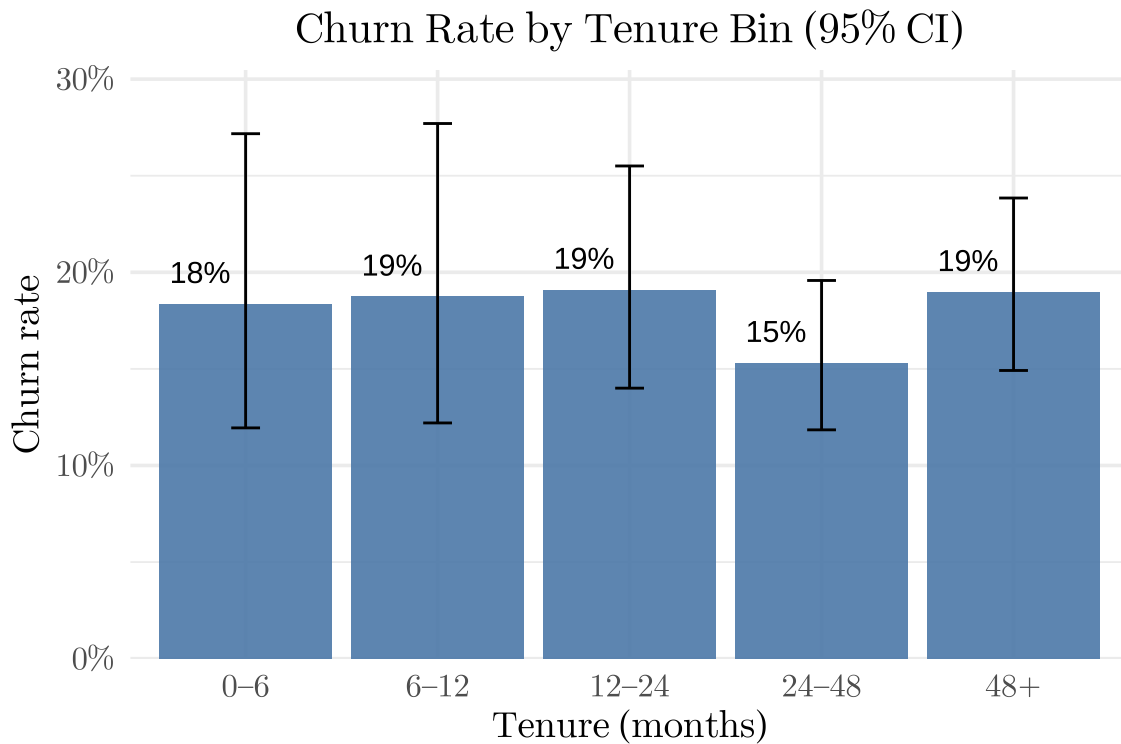
Gender	Age	Contract	Tenure Bin
Female	57	Month-to-month	24-48
Male	38	Two years	12-24
Female	27	One year	48+
Male	26	One year	6-12
Female	46	Month-to-month	6-12

- n = number of customers
- churn_yes = number of churned customers
- $p = \text{churn_yes} / n$ = churn proportion

Table 2.2: Tenure statistics

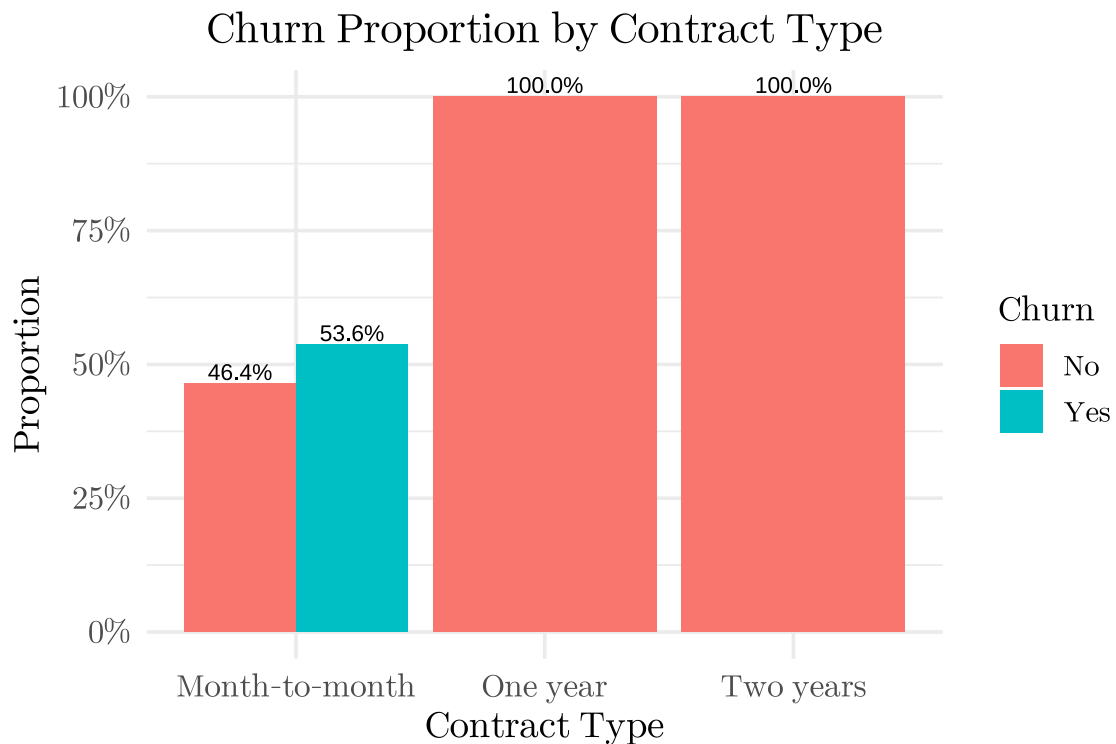
level	churn_yes	n	p	p_lo	p_hi
0-6	18	98	0.18	0.12	0.27
6-12	18	96	0.19	0.12	0.28
12-24	34	178	0.19	0.14	0.26
24-48	51	333	0.15	0.12	0.20
48+	56	295	0.19	0.15	0.24

Because churn rate is an estimate from data, we add 95% confidence interval using the Wilson method.



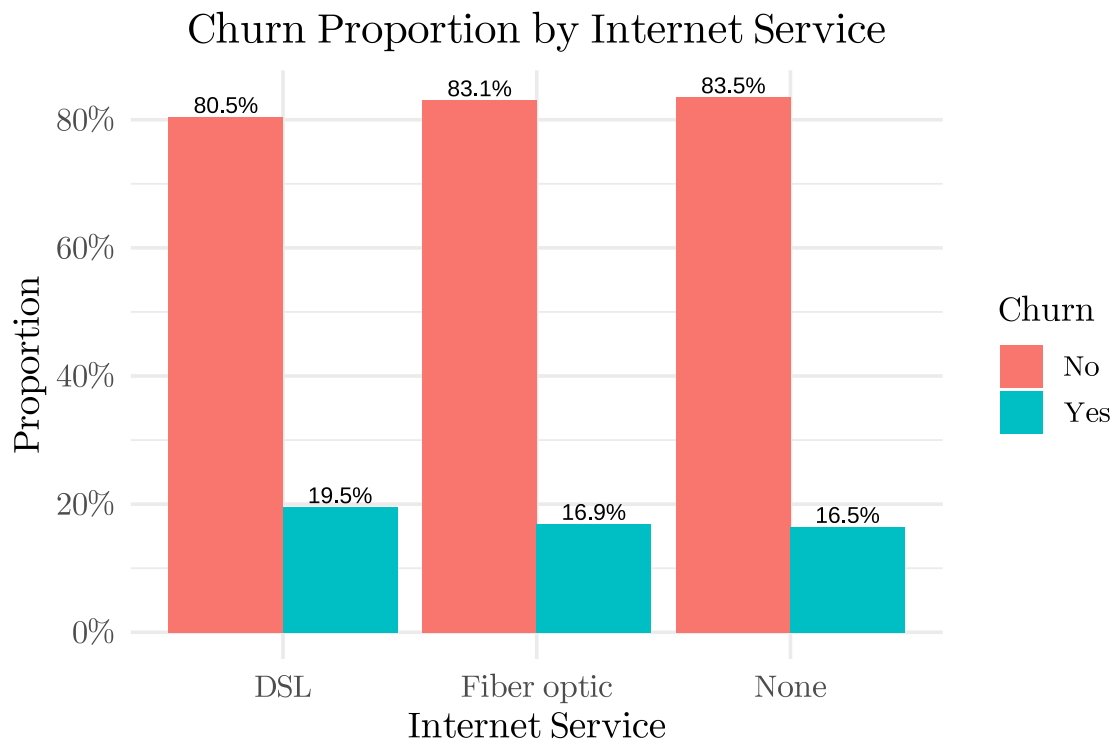
2.2 Contract Type vs. Churn

Goal: Show how churn differs by contract type. Hypothesis: Month-to-month has higher churn than One/Two year contracts.

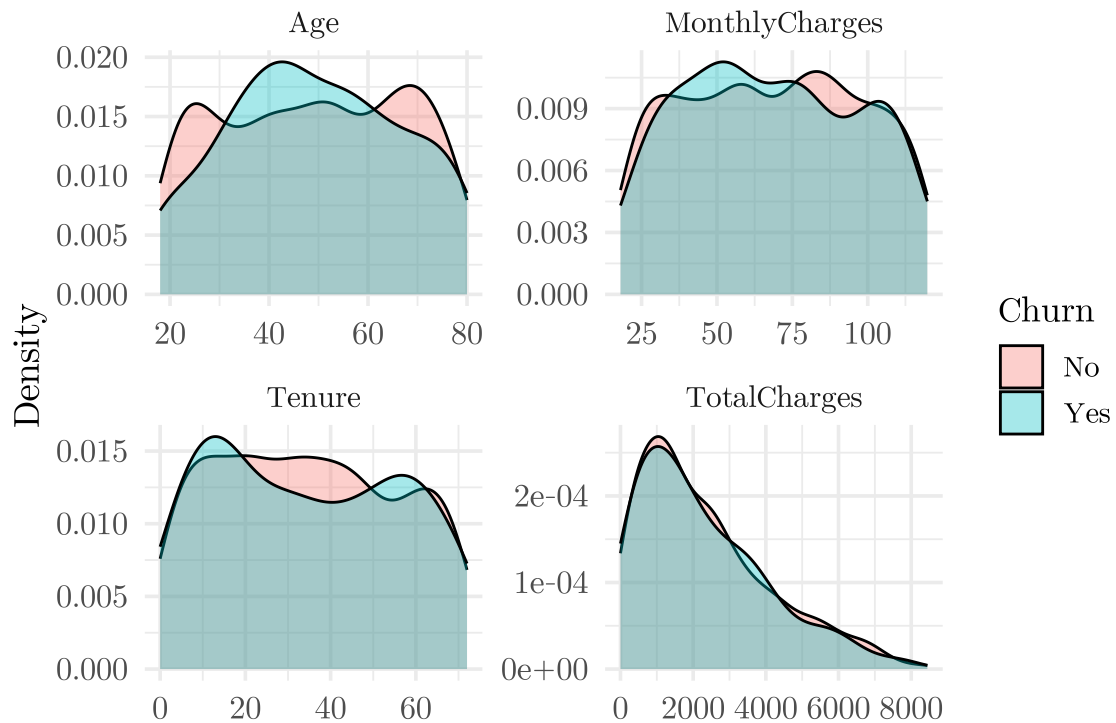


2.3 Internet Service vs. Churn

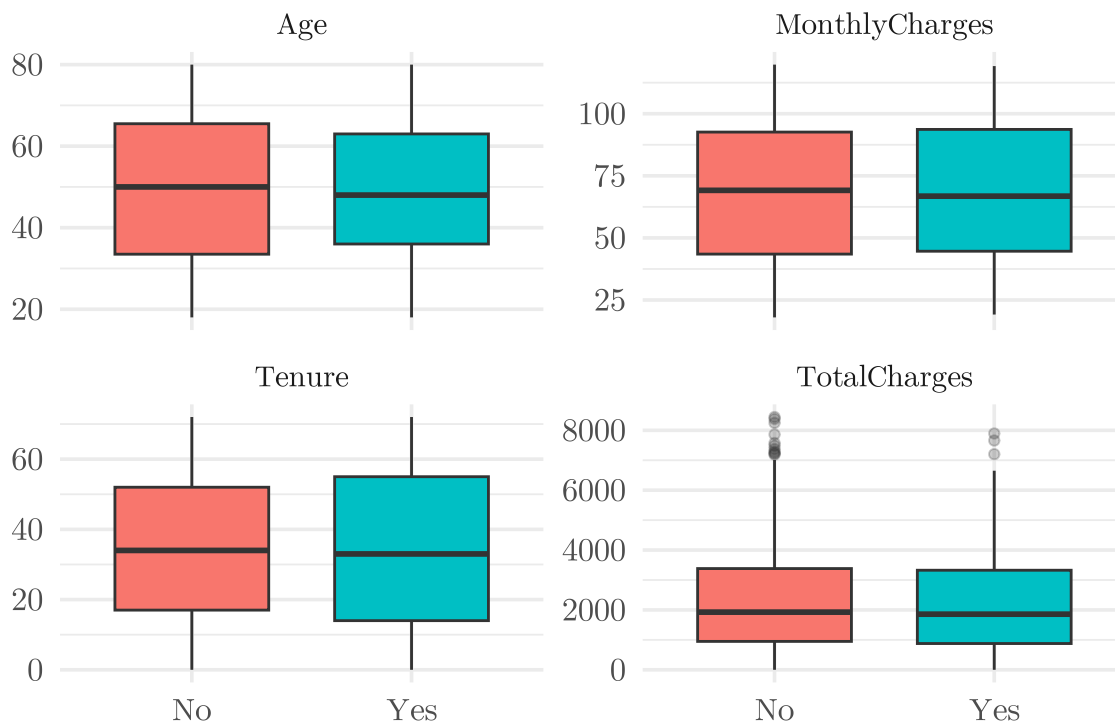
Goal: Compare churn across DSL / Fiber optic / No Internet.



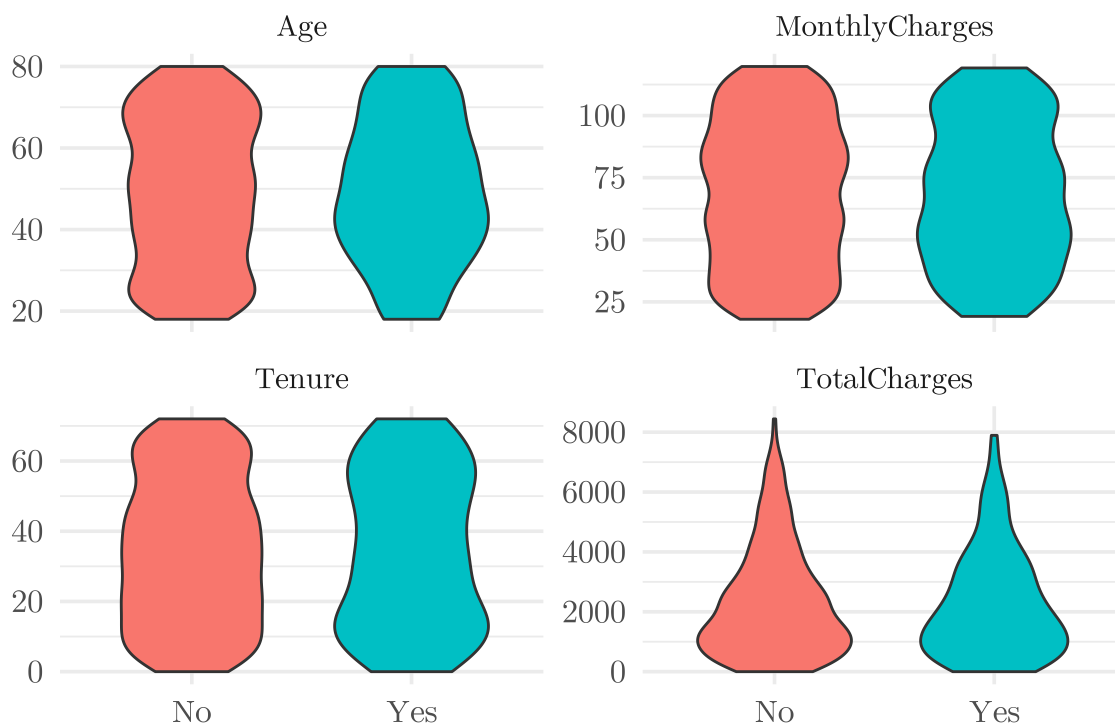
2.4 Numeric Densities by Churn



2.5 Numeric Box Plots by Churn



2.6 Numeric Violin Plots by Churn



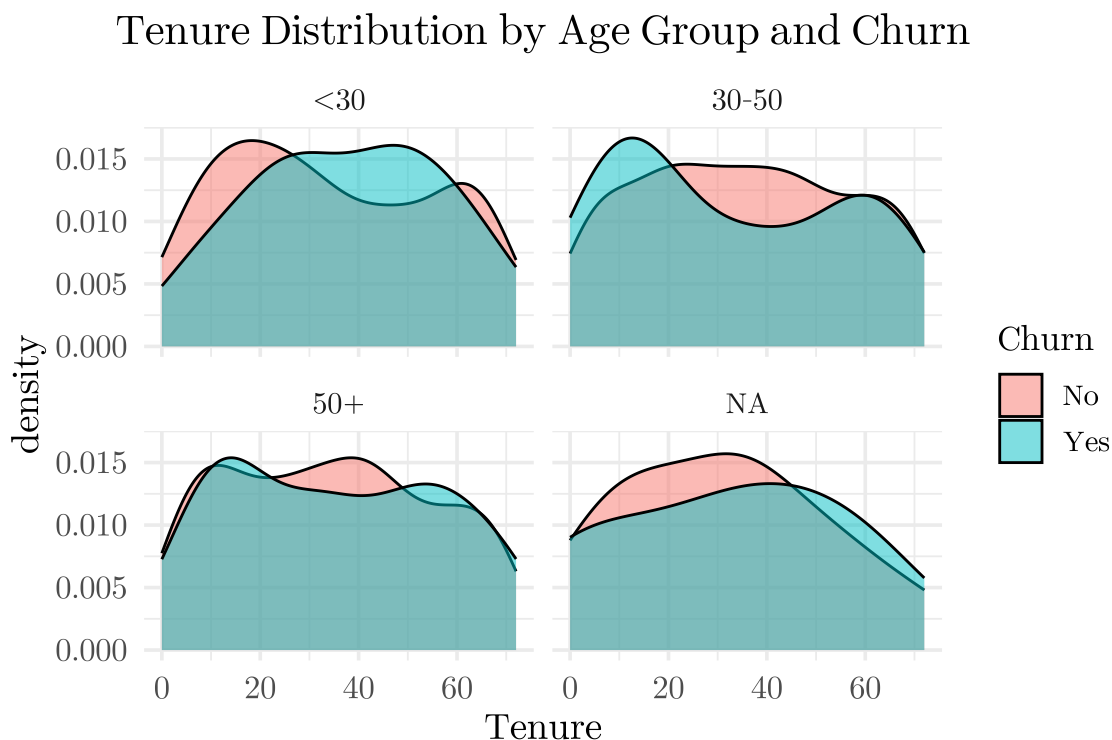
3 Multivariate Relationships

3.1 Age x Tenure x Churn

Table 3.1: Data for ‘Tenure Distribution by Age Group and Churn’

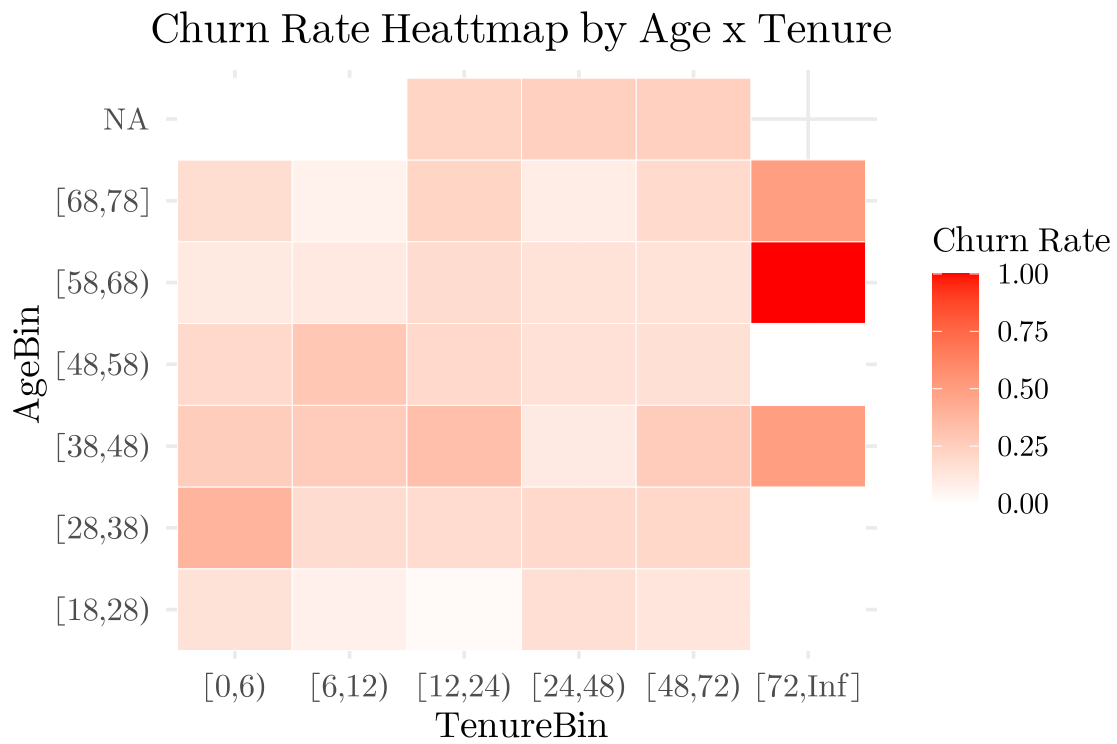
Gender	Age	Tenure	PhoneService	InternetService	Contract	MonthlyCharges	TotalCharges	Churn	AgeGroup
Female	23	45	Yes	None	Month-to-month	48.32	2174.40	No	<30
Female	25	64	Yes	DSL	One year	24.08	1541.12	No	<30
Female	70	4	Yes	None	One year	68.27	273.08	No	50+
Female	21	70	Yes	DSL	One year	60.05	4203.50	No	<30
Male	45	26	No	Fiber optic	One year	85.50	2223.00	No	30-50

Following plot shows distributions of Tenure across age groups and split by Churn,

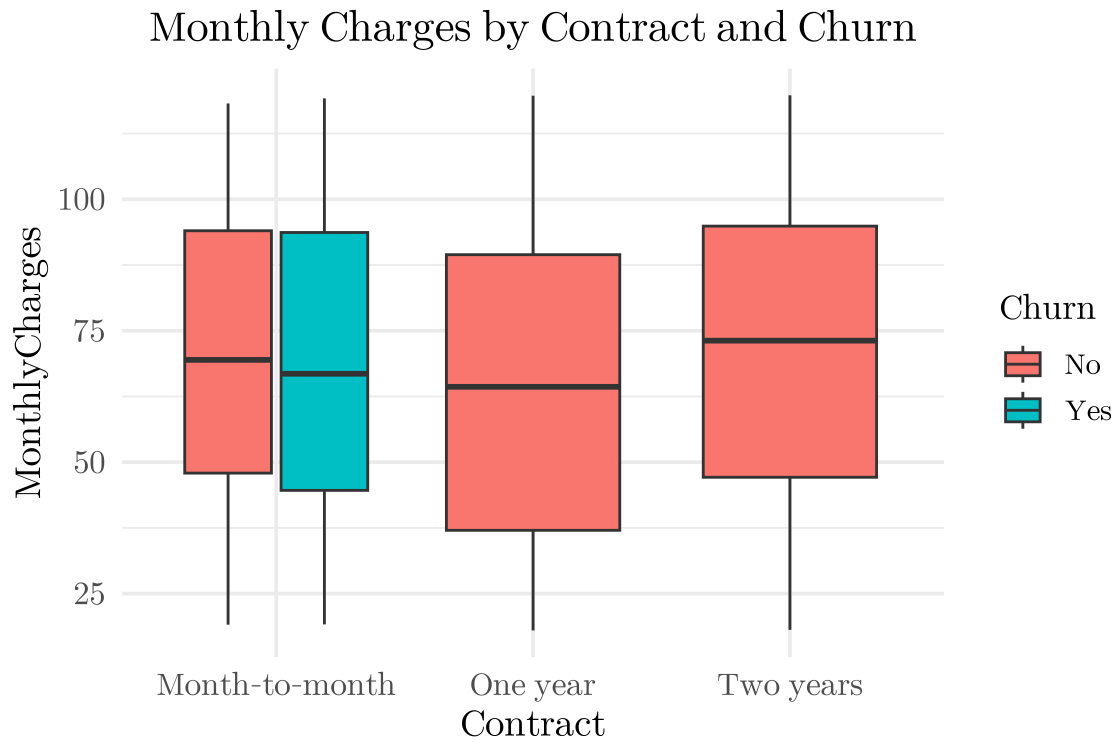


For extra clarity, following is a Heatmap,

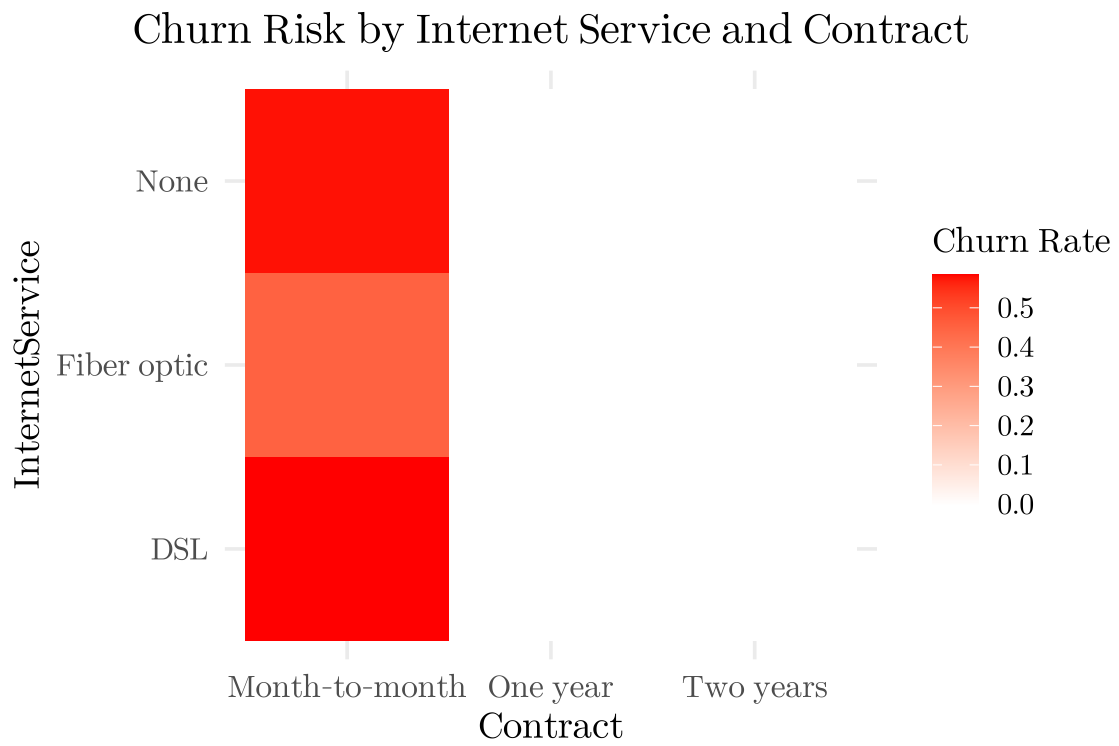
AgeBin	TenureBin	ChurnRate
[68,78]	[12,24)	0.22
NA	[12,24)	0.22
[28,38)	[72,Inf]	0.00
[58,68)	[12,24)	0.18
[38,48)	[24,48)	0.11



3.2 MonthlyCharges x Contract x Churn



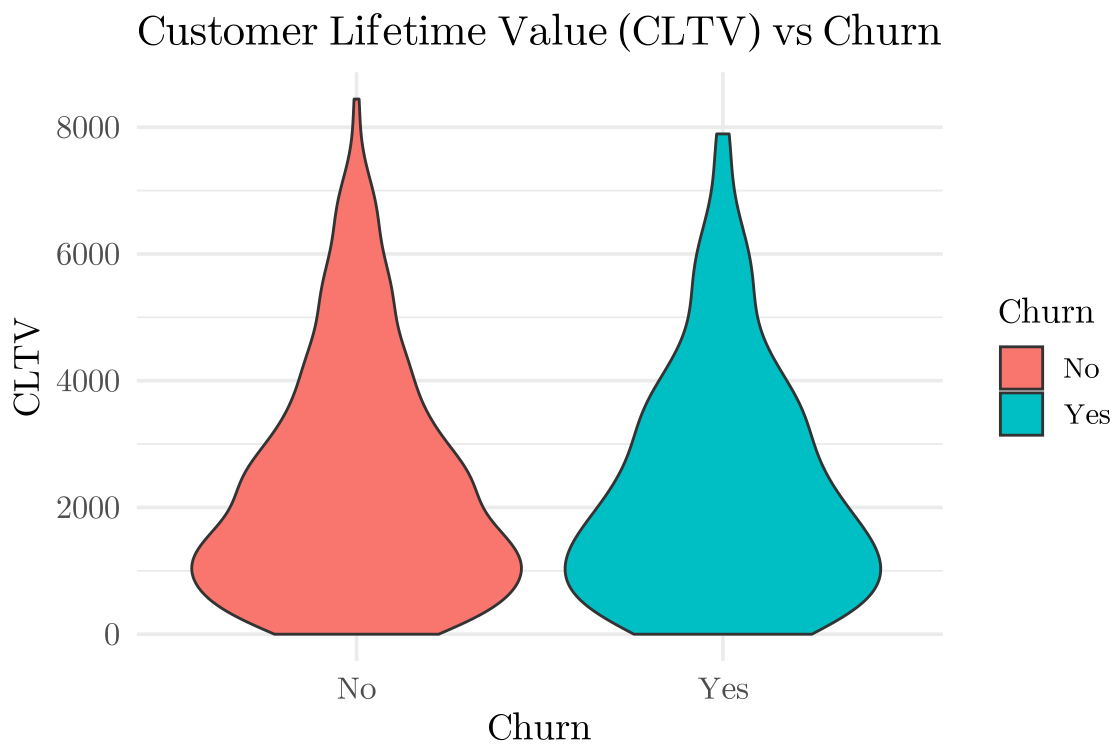
3.3 InternetService x Contract (churn risk heatmap)



4 Feature Engineering for EDA

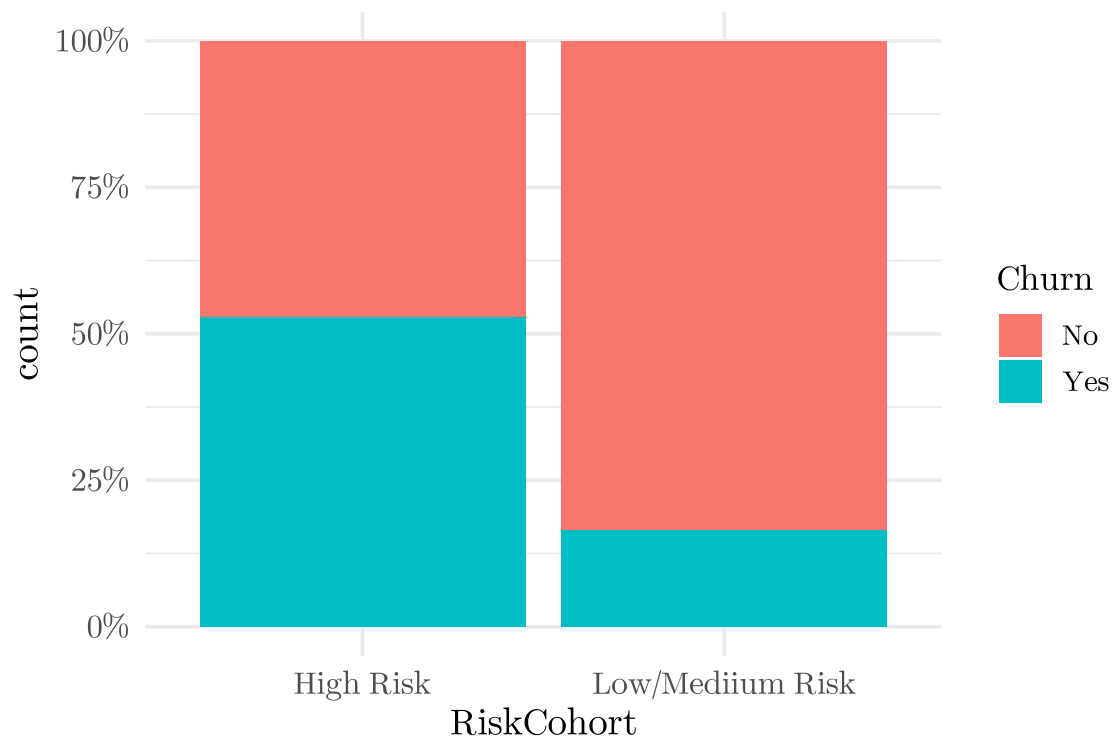
4.1 Customer lifetime value proxy

Gender	Age	Tenure	PhoneService	InternetService	Contract	MonthlyCharges	TotalCharges	Churn	CLTV
Male	18	40	Yes	None	Month-to-month	57.11	2284.40	No	2284.40
Female	50	50	Yes	Fiber optic	Month-to-month	49.19	2459.50	No	2459.50
Male	73	59	No	DSL	One year	78.71	4643.89	No	4643.89
Female	73	13	Yes	Fiber optic	Month-to-month	119.18	1549.34	Yes	1549.34
Male	37	6	Yes	Fiber optic	Month-to-month	113.31	679.86	No	679.86



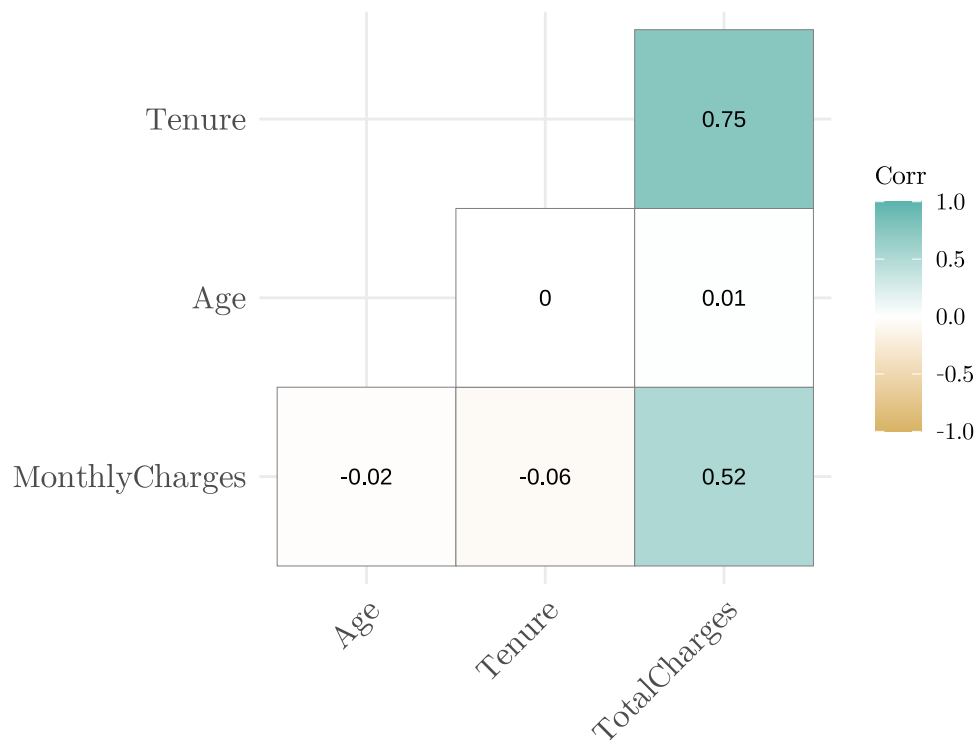
4.2 At-risk cohorts (cohort flagging)

Create a derived feature for risk segments,



5 Correlation & Association

5.1 Numeric-only correlation matrix



5.2 Chi-square + Cramer's V for categorical vs. Churn

Table 5.1: Cramér's V - Effect size measure

variable	chisq_stat	df	p_value	cramers_v
Contract	436.65	2	0.00	0.66
Gender	2.34	1	0.13	0.05
PhoneService	0.88	1	0.35	0.03
InternetService	1.30	2	0.52	0.04

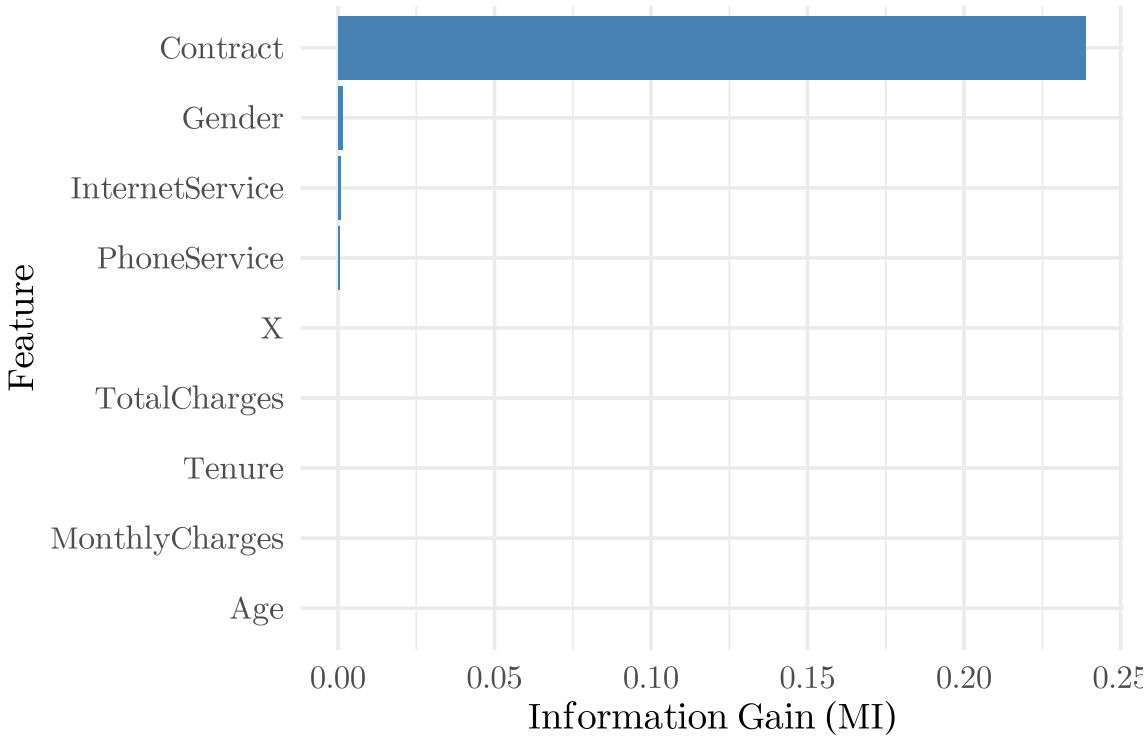
Which shows that, to predict churn, Contract is a strong (because of low `p_value`) predictor. Next up are, Gender, PhoneService, and InternetService.

5.3 Mutual information (any type \rightarrow Churn)

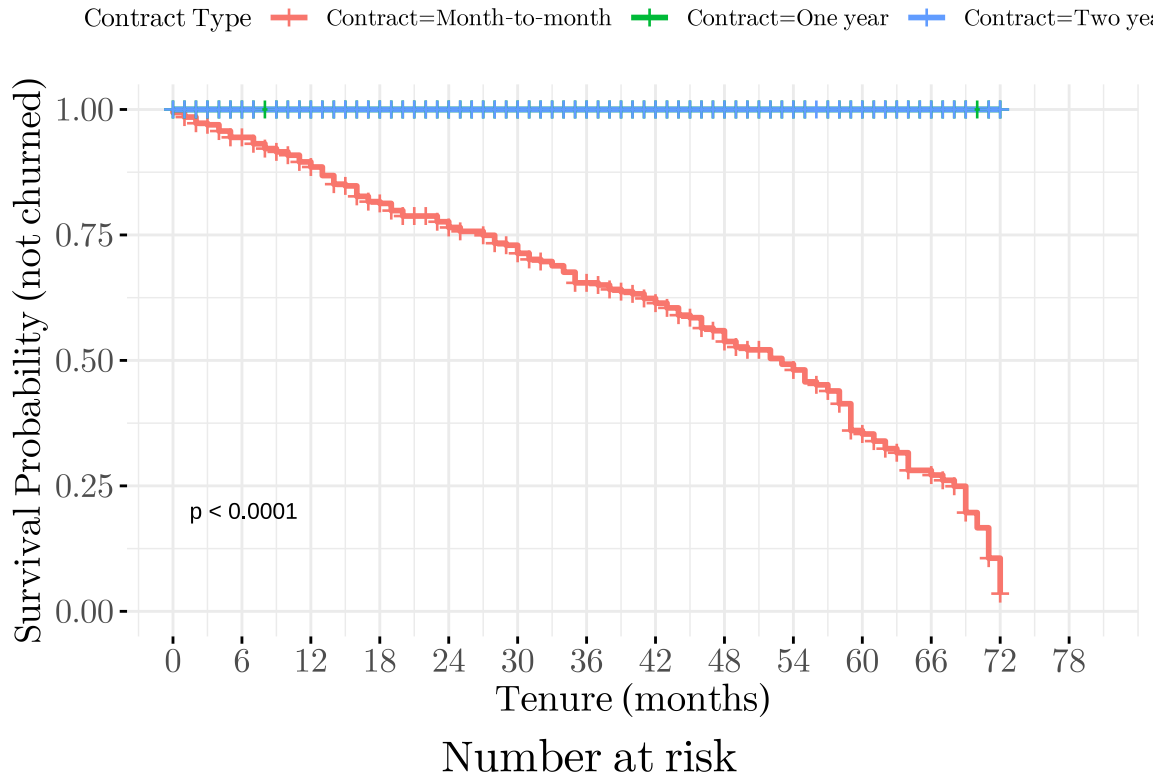
Mutual Information is a great way to rank all features (numeric + categorical) together by how much they help explain churn. It complements correlation, Chi-square, and Cramér's V by giving a unified importance measure.

Table 5.2: Importance of variables

attributes	importance
Contract	0.238949
Gender	0.001304
InternetService	0.000644
PhoneService	0.000519



6 Survival-style Analysis



Contract=Month-to-month 33 29 26 23 20 18 15 13 10 8 6 5 2 30 6 0

Contract=One year 32 30 27 25 22 18 21 14 12 9 10 7 8 15 9 35 5 0

Contract=Two years 34 31 32 34 22 20 21 7 21 3 19 9 7 5 1 26 3 0