Customer Churn Analysis

Aheer Srabon

2025-08-21

Table of contents

1 Uni-variate & Distributional Analysis

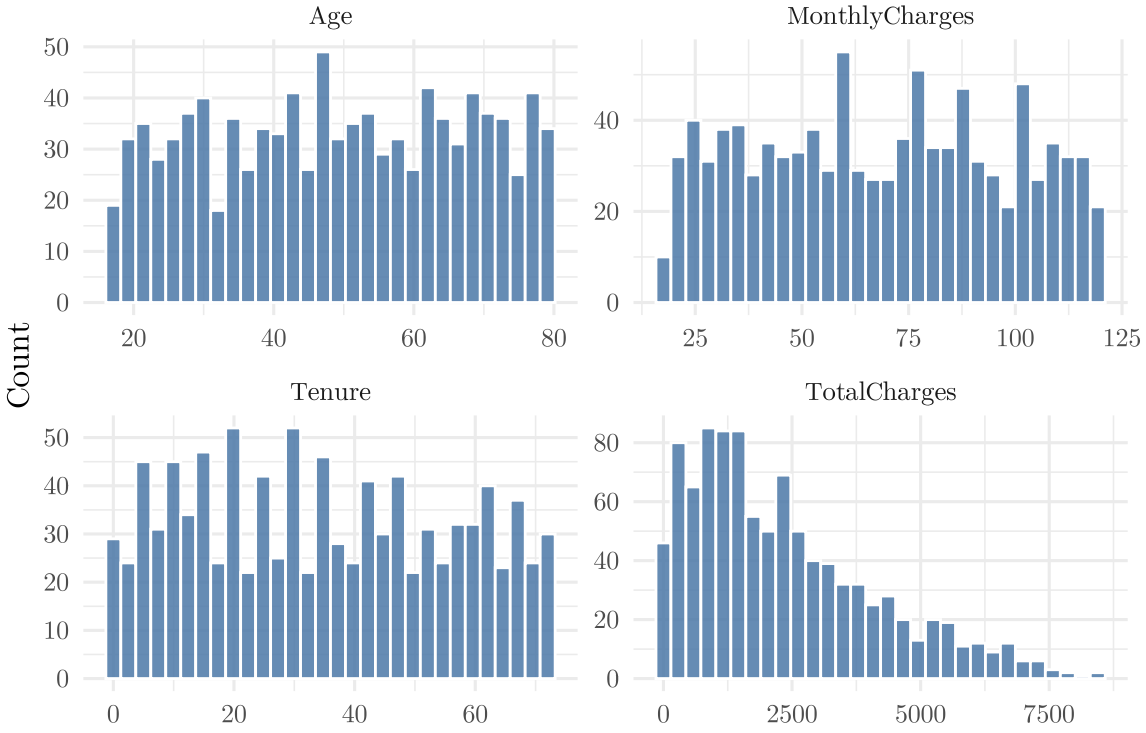
Table 1.1: Summary of the Dataset

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

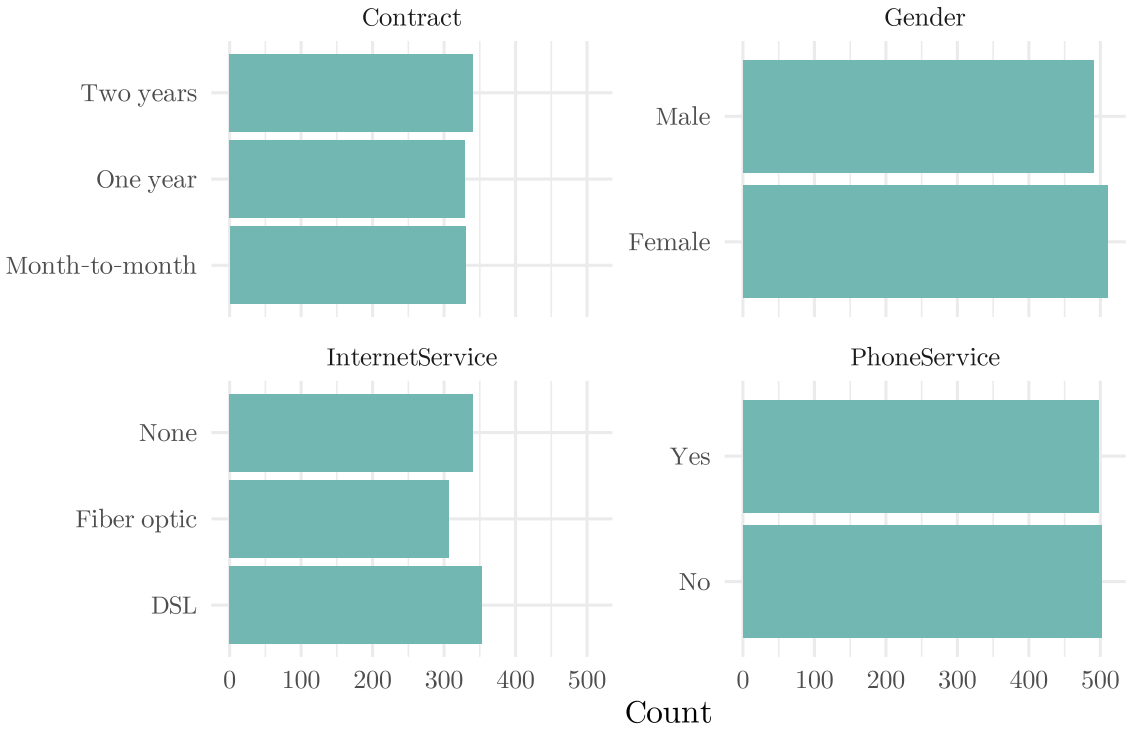
Table 1.2: Random 10 samples from the dataset

Gender	Age	Tenure	PhoneService	InternetService	Contract	MonthlyCharges	TotalCharges	Churn
Female	78	5	Yes	None	One year	30.28	151.40	No
Female	26	66	Yes	Fiber optic	Month-to-month	85.26	5627.16	Yes
Female	80	28	Yes	Fiber optic	Month-to-month	109.59	3068.52	Yes
Male	57	51	Yes	DSL	One year	116.50	5941.50	No
Male	48	55	Yes	DSL	One year	30.23	1662.65	No
Male	23	22	No	DSL	One year	81.16	1785.52	No
Male	38	20	Yes	None	Two years	83.85	1677.00	No
Female	66	21	No	None	Two years	20.45	429.45	No
Female	69	64	Yes	Fiber optic	Month-to-month	66.81	4275.84	Yes
Female	78	22	No	None	One year	81.51	1793.22	No

1.1 Numeric Distributions (Histograms)



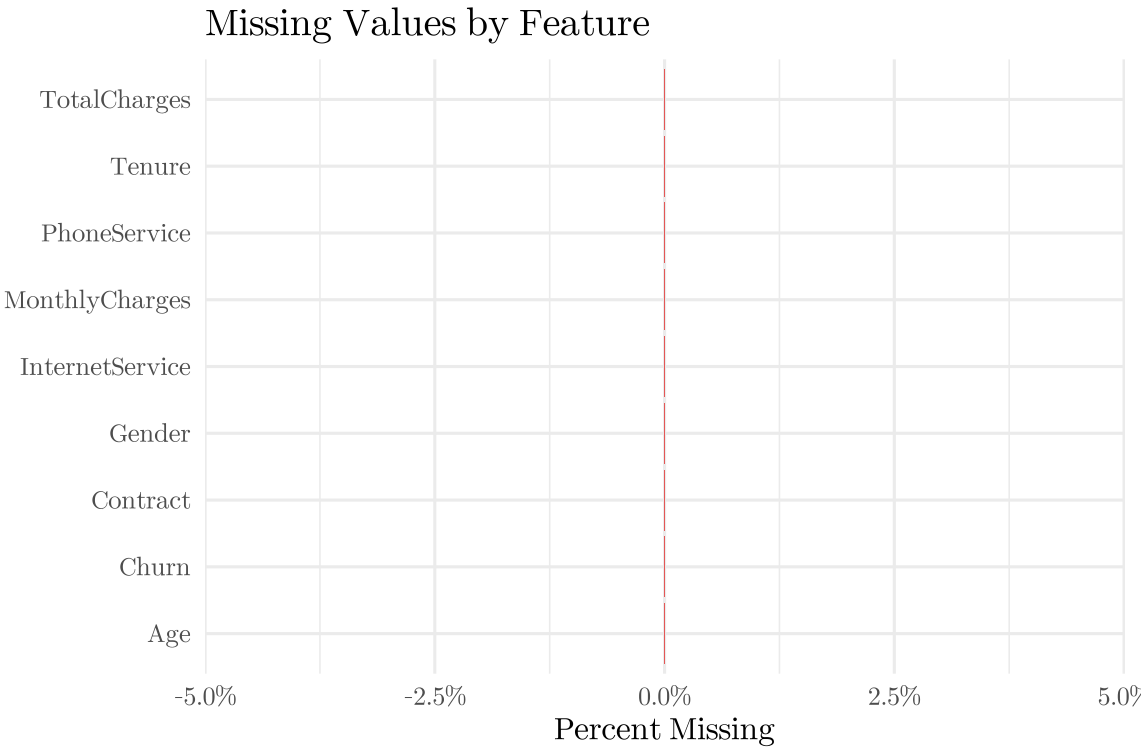
1.2 Categorical Counts (Phone/Internet/Contract)



1.3 Categorical Proportions by Churn

1.4 Missingness Overview

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



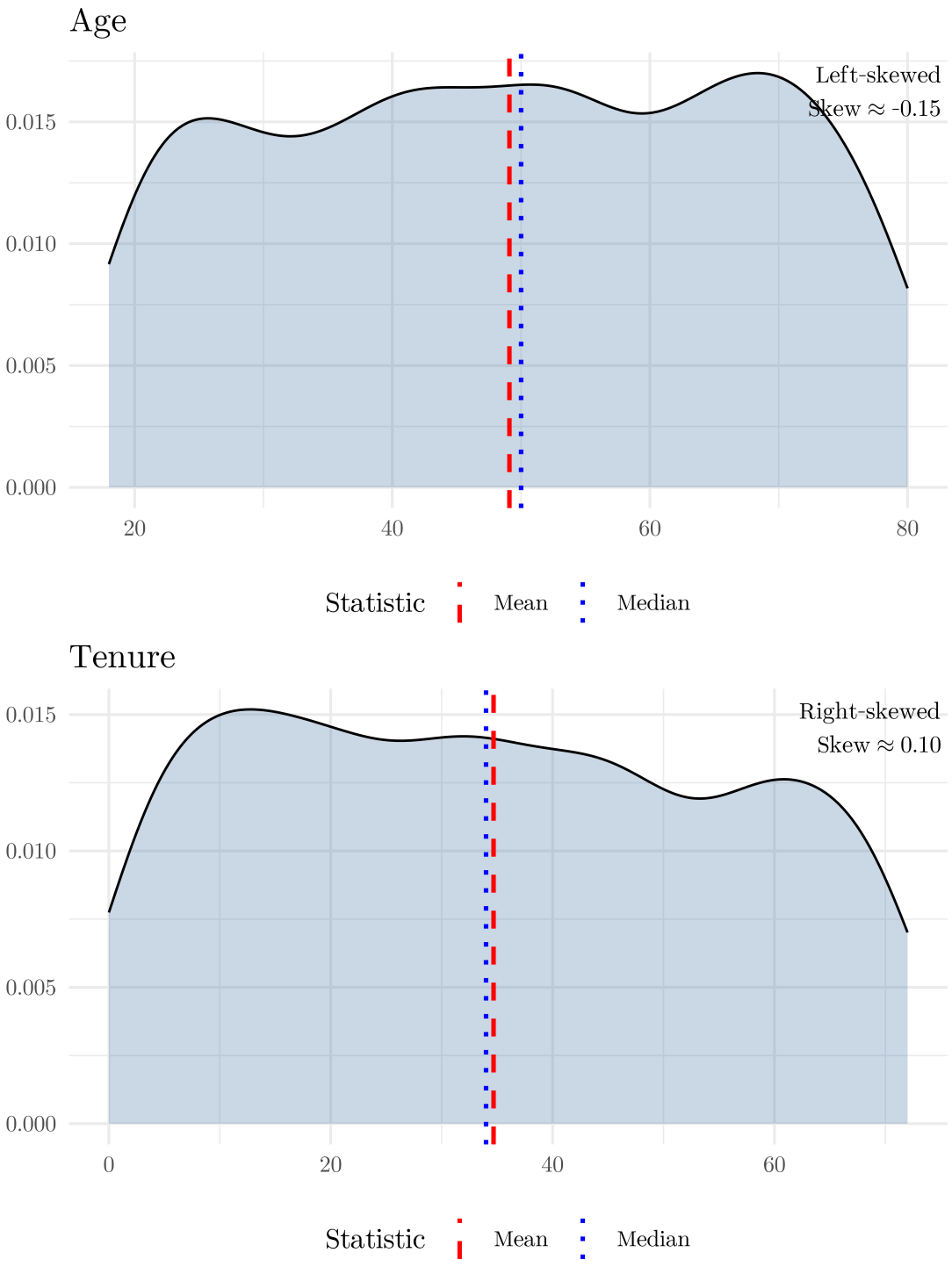
1.5 Skewness

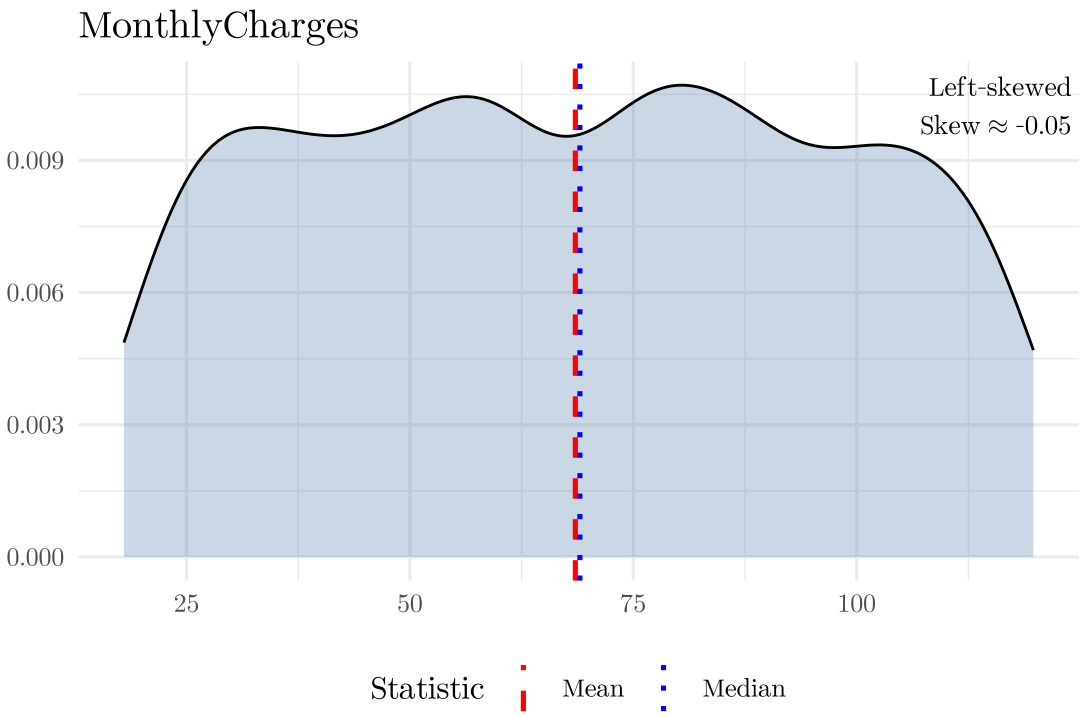
1.5.1 Quick skew check (numeric)

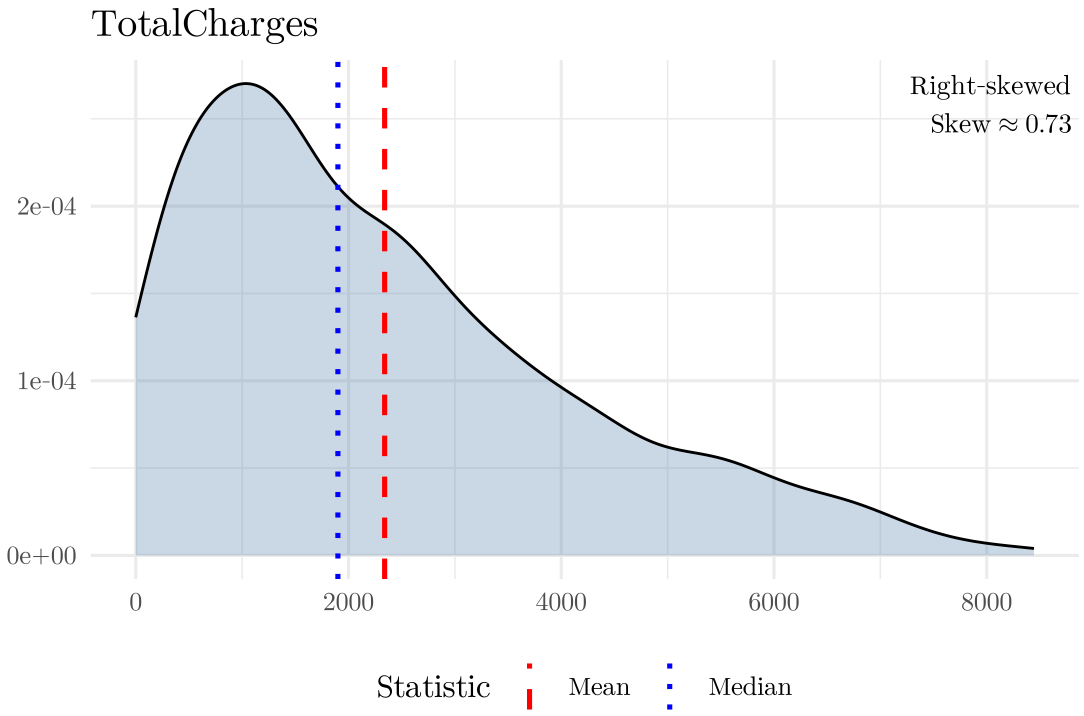
Table 1.3: Skewness

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





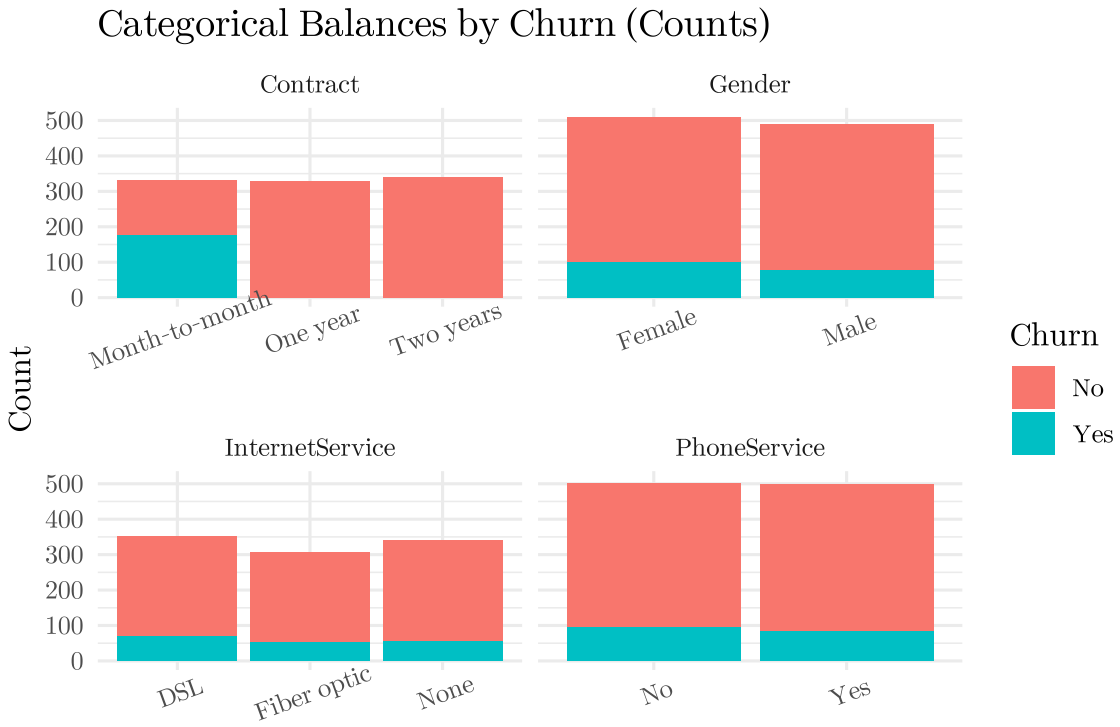


1.6 Churn vs Non-Churn Proportions

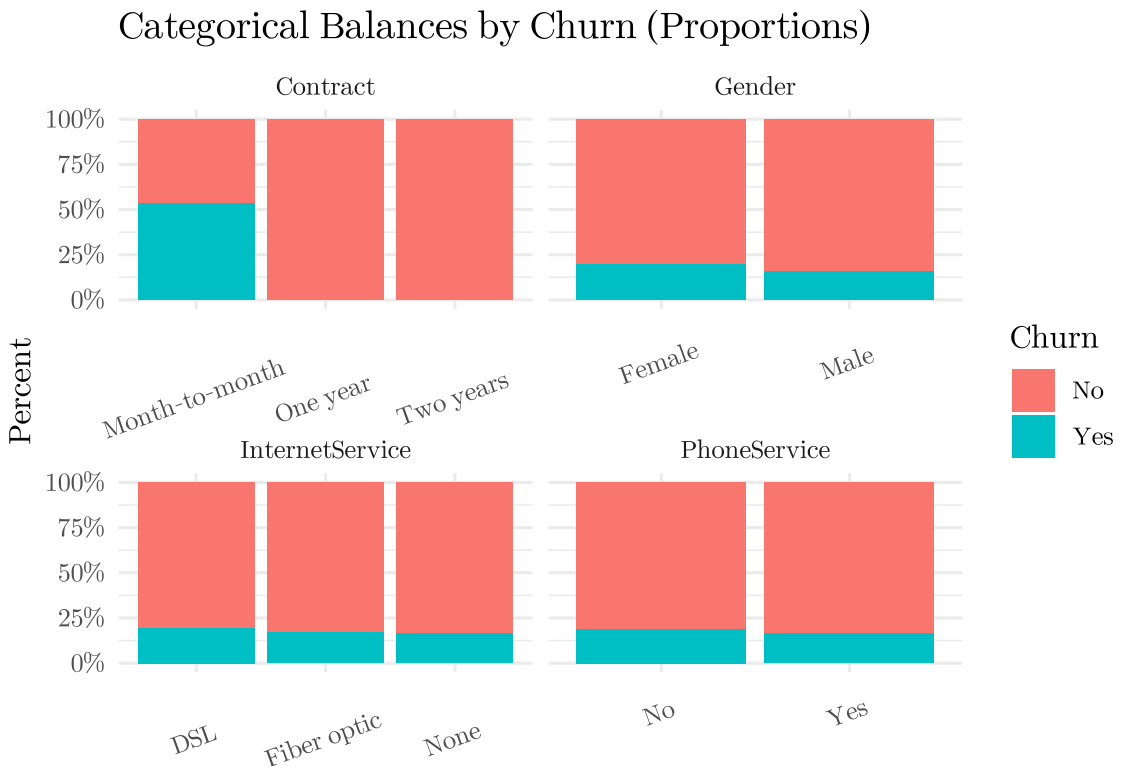
Table 1.4: Categorical features of interest

Churn	Feature	Level
No	Contract	Two years
No	Gender	Female
No	PhoneService	No
No	Gender	Female
No	PhoneService	Yes

1.6.1 Stacked bar (counts by category)



1.6.2 Normalized bar (proportions within each category)



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

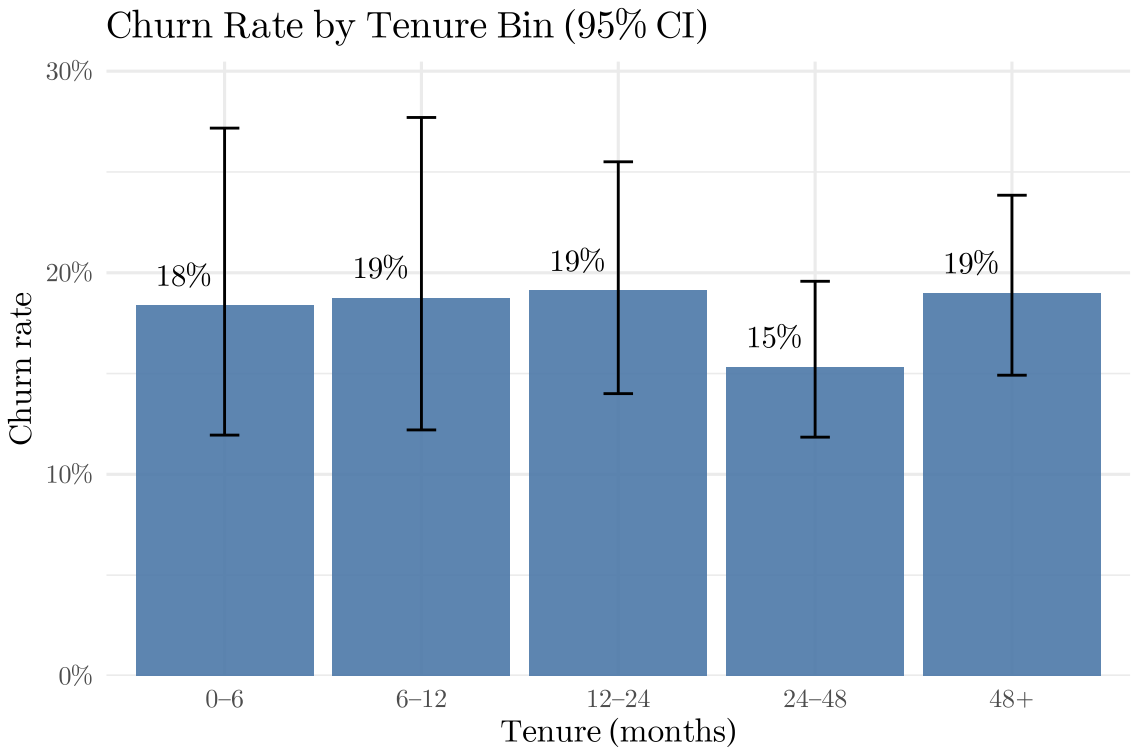
Gender	Age	Contract	Tenure Bin
Female	71	One year	48+
Male	73	Month-to-month	48+
Female	64	Month-to-month	48+
Female	75	One year	0–6
Male	55	One year	0–6

- `n` = number of customers
- `churn_yes` = number of churned customers
- `p` = `churn_yes` / `n` = churn proportion

Table 2.2: Tenure statistics

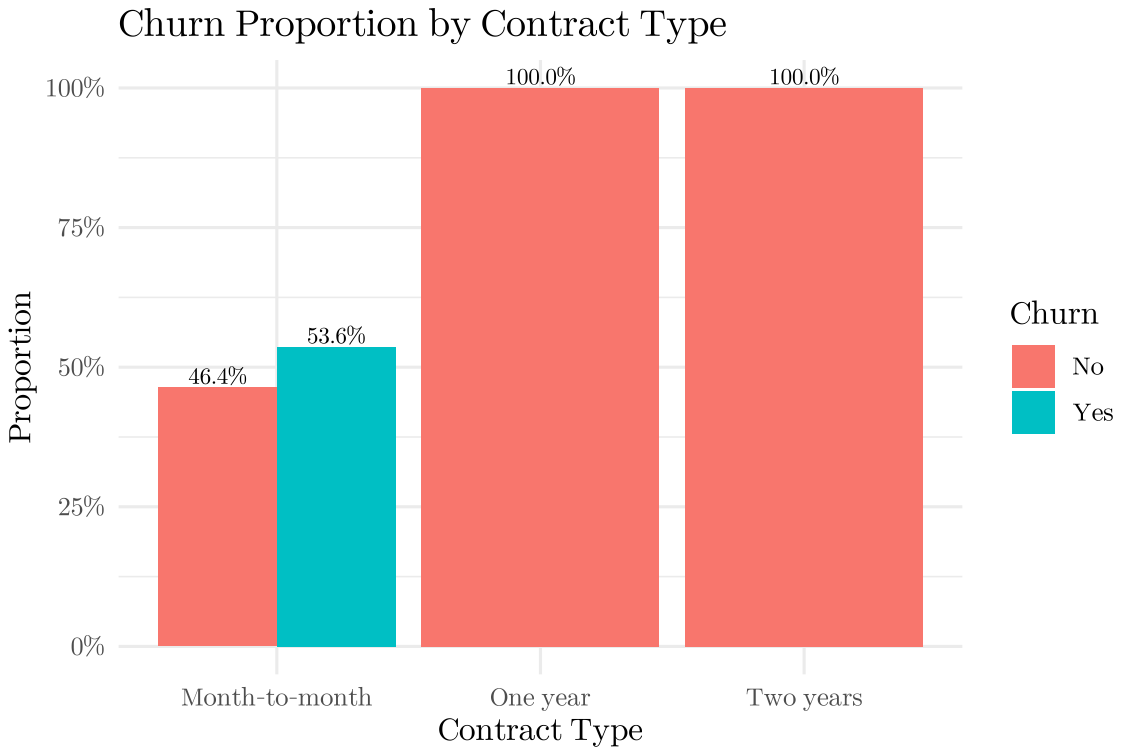
level	churn_yes	n	p	p_lo	p_hi
0–6	18	98	0.18	0.12	0.27
6–12	18	96	0.19	0.12	0.28
12–24	34	178	0.19	0.14	0.26
24–48	51	333	0.15	0.12	0.20
48+	56	295	0.19	0.15	0.24

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



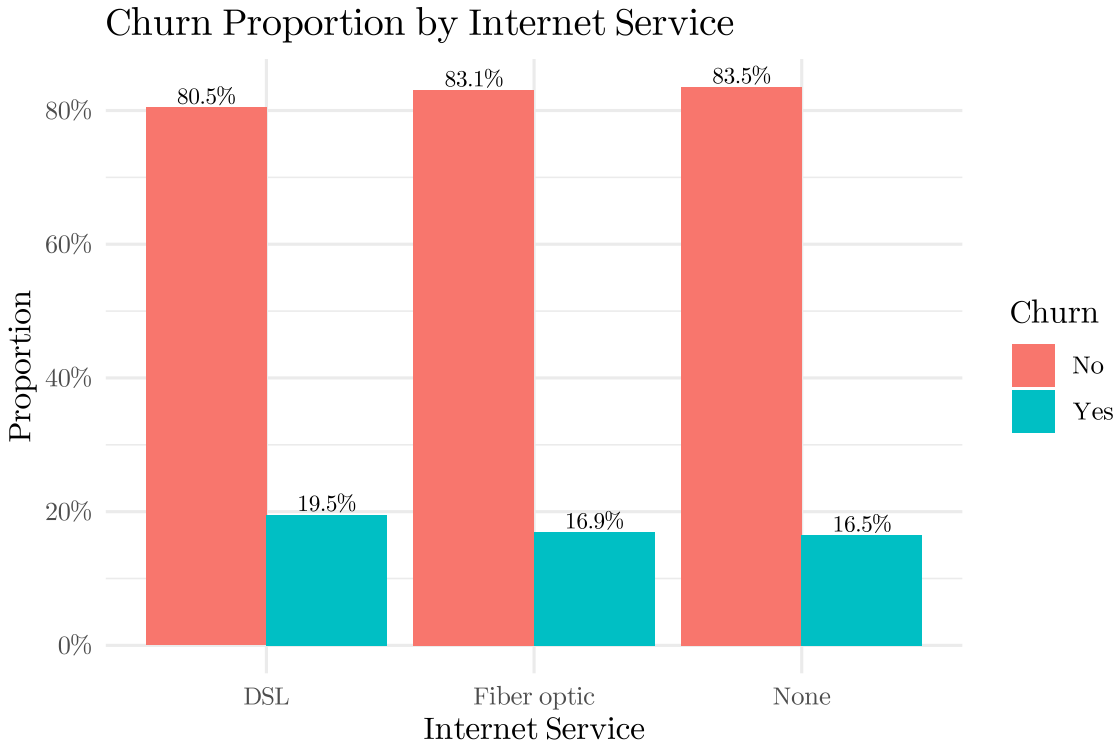
2.2 Contract Type vs. Churn

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

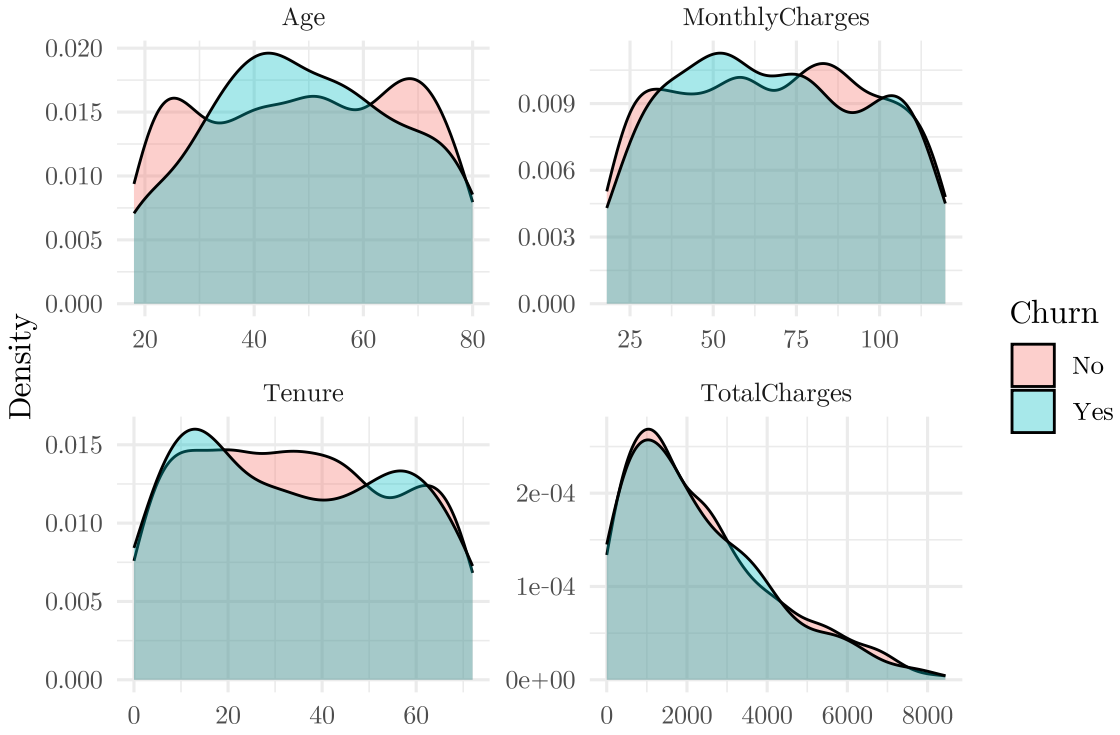


2.3 Internet Service vs. Churn

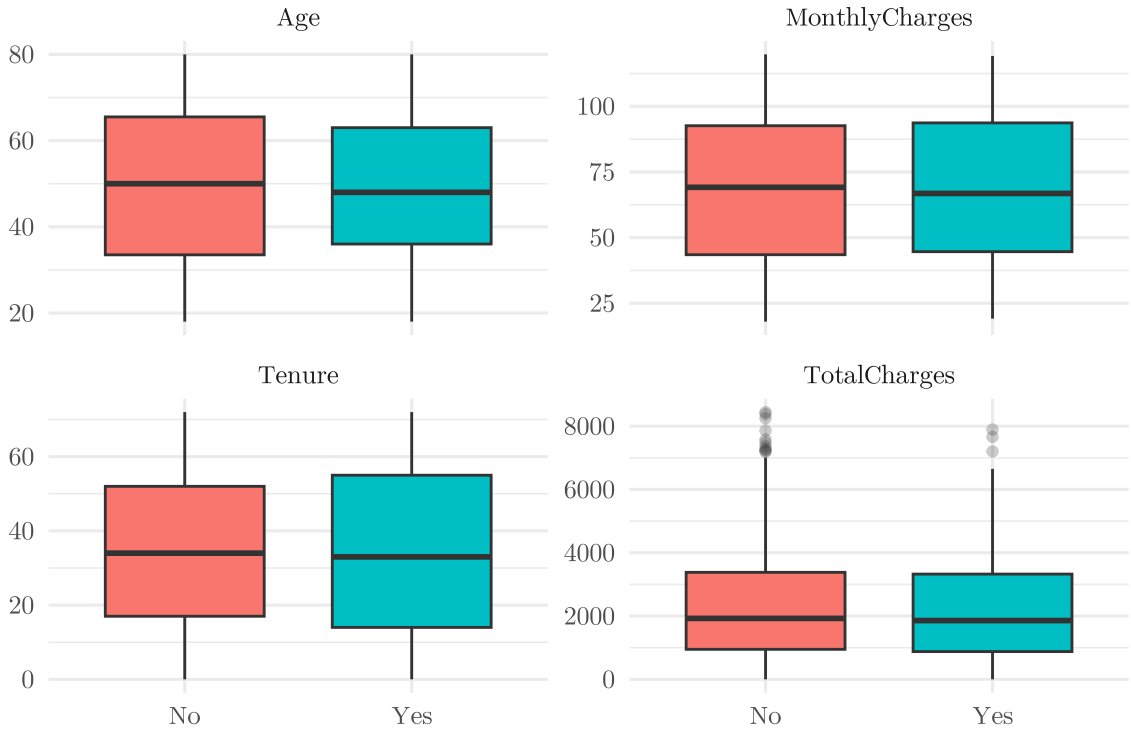
Goal: Compare churn across DSL / Fiber optic / No Internet.



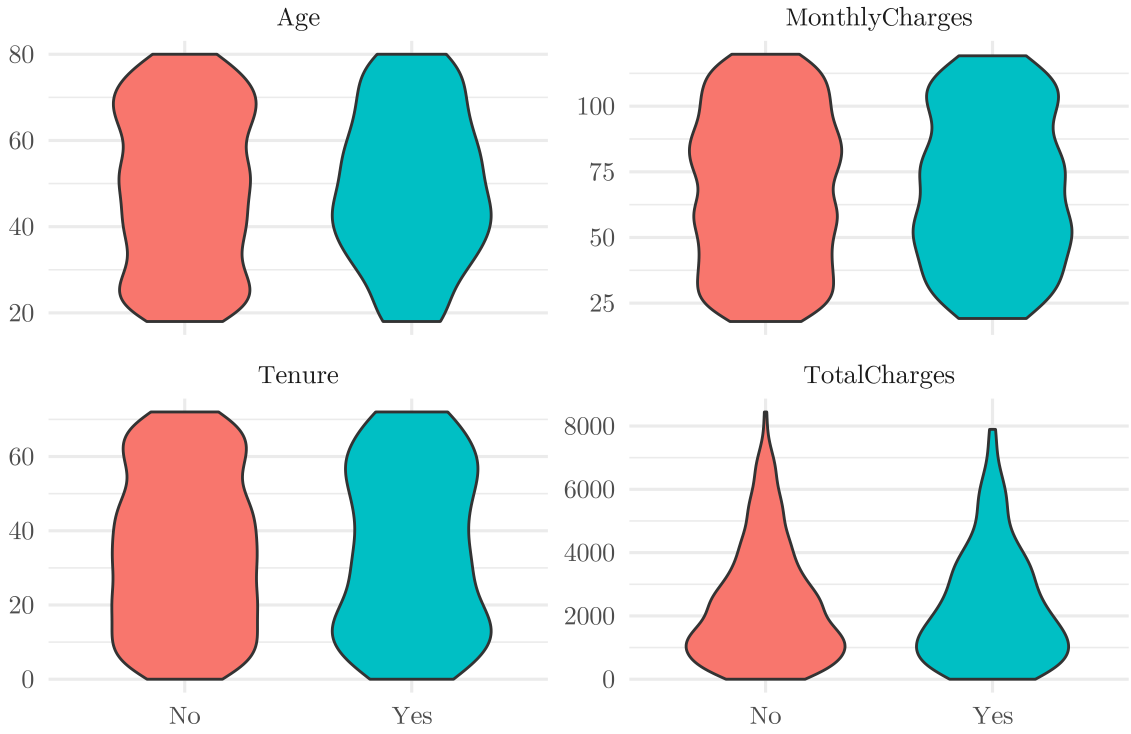
2.4 Numeric Densities by Churn



2.5 Numeric Box Plots by Churn



2.6 Numeric Violin Plots by Churn



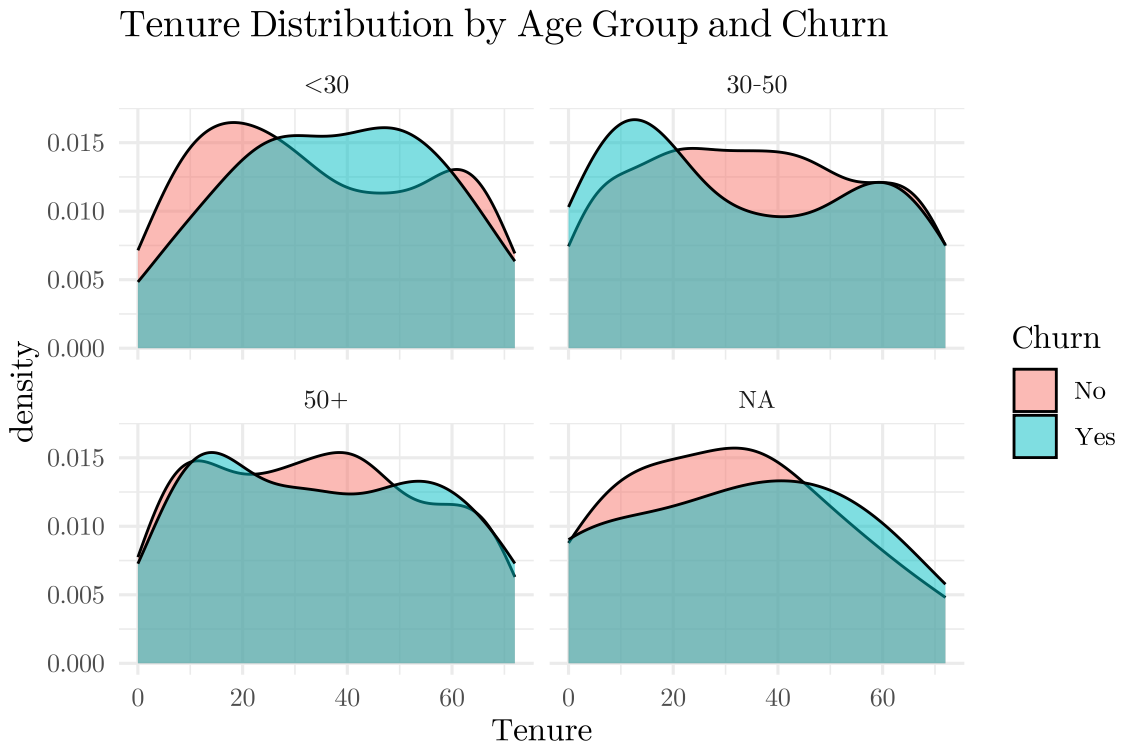
3 Multivariate Relationships

3.1 Age x Tenure x Churn

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

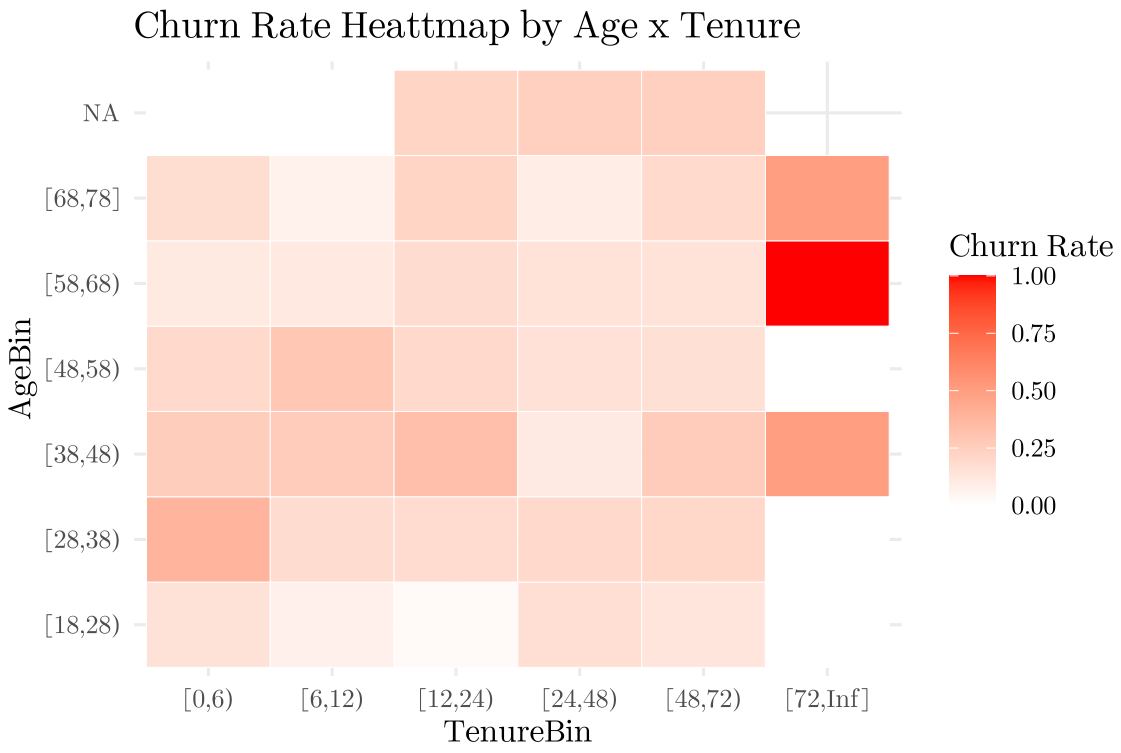
Gender	Age	Tenure	PhoneService	InternetService	Contract	MonthlyCharges	TotalCharges	Churn	AgeGroup
Female	35	37	No	DSL	Two years	63.09	2334.33	No	30-50
Female	38	59	No	Fiber optic	Month-to-month	32.94	1943.46	No	30-50
Male	25	9	No	DSL	Two years	111.68	1005.12	No	<30
Male	59	20	No	DSL	One year	41.57	831.40	No	50+
Male	43	49	No	DSL	Month-to-month	86.45	4236.05	Yes	30-50

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

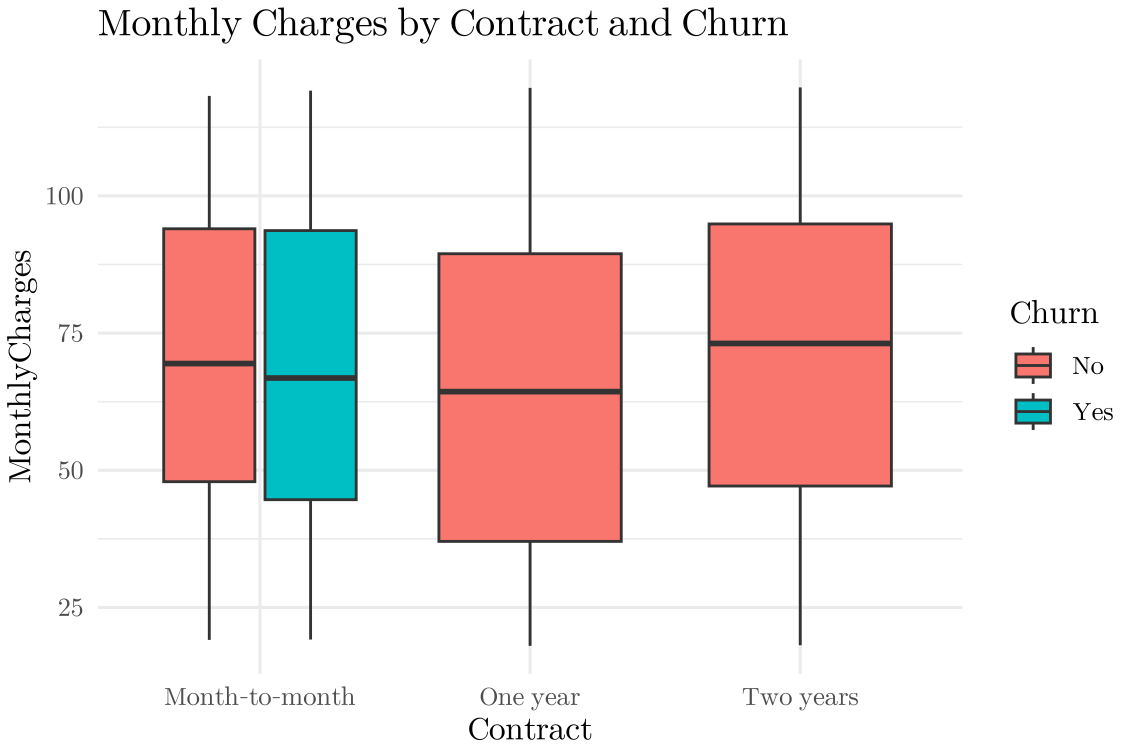


For extra clarity, following is a Heatmap,

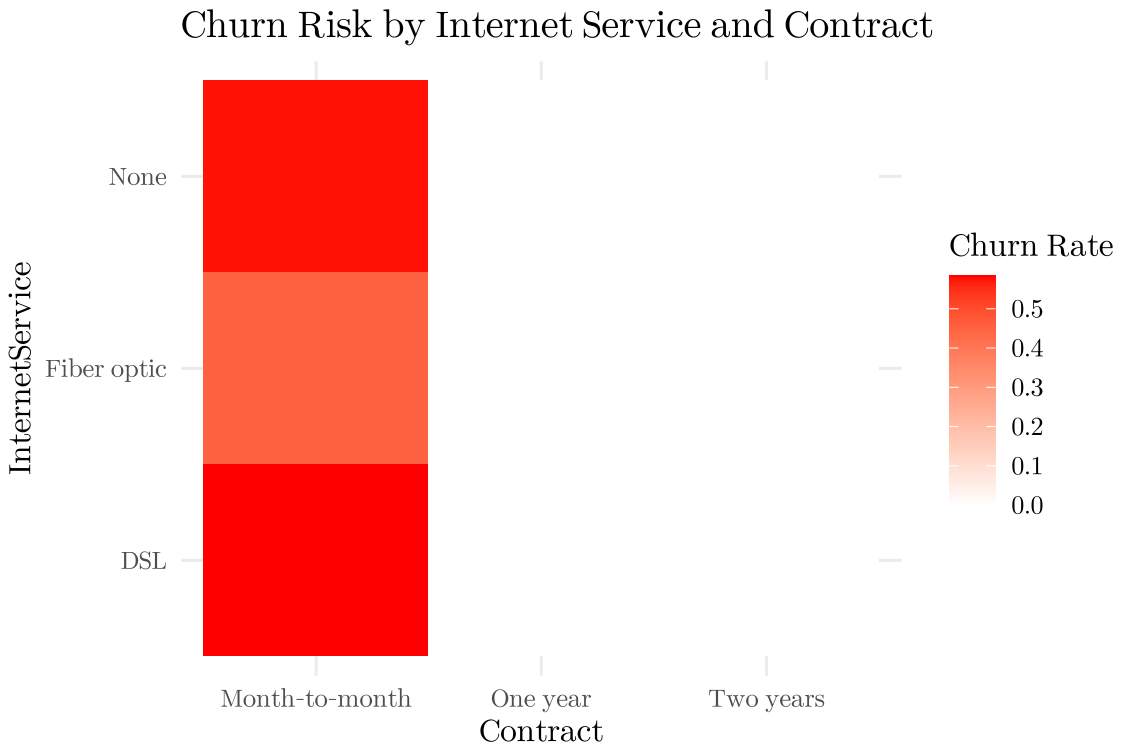
AgeBin	TenureBin	ChurnRate
[38,48)	[0,6)	0.26
[68,78]	[0,6)	0.17
NA	[0,6)	0.00
[28,38)	[48,72)	0.21
NA	[12,24)	0.22



3.2 MonthlyCharges x Contract x Churn



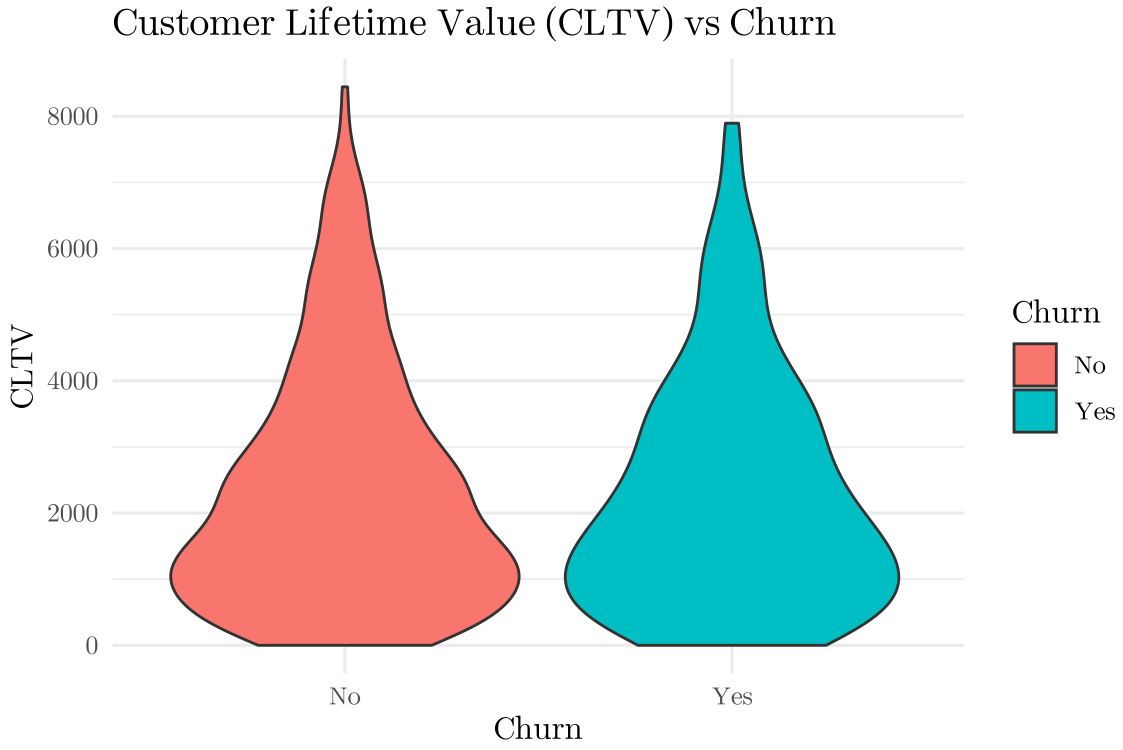
3.3 InternetService x Contract (churn risk heatmap)



4 Feature Engineering for EDA

4.1 Customer lifetime value proxy

Gender	Age	Tenure	PhoneService	InternetService	Contract	MonthlyCharges	TotalCharges	Churn	CLTV
Female	20	8	No	None	Month-to-month	106.34	850.72	No	850.72
Male	55	49	Yes	Fiber optic	One year	101.69	4982.81	No	4982.81
Male	54	40	No	Fiber optic	Month-to-month	59.94	2397.60	No	2397.60
Female	68	12	No	None	One year	22.65	271.80	No	271.80
Male	64	10	Yes	None	Two years	100.15	1001.50	No	1001.50



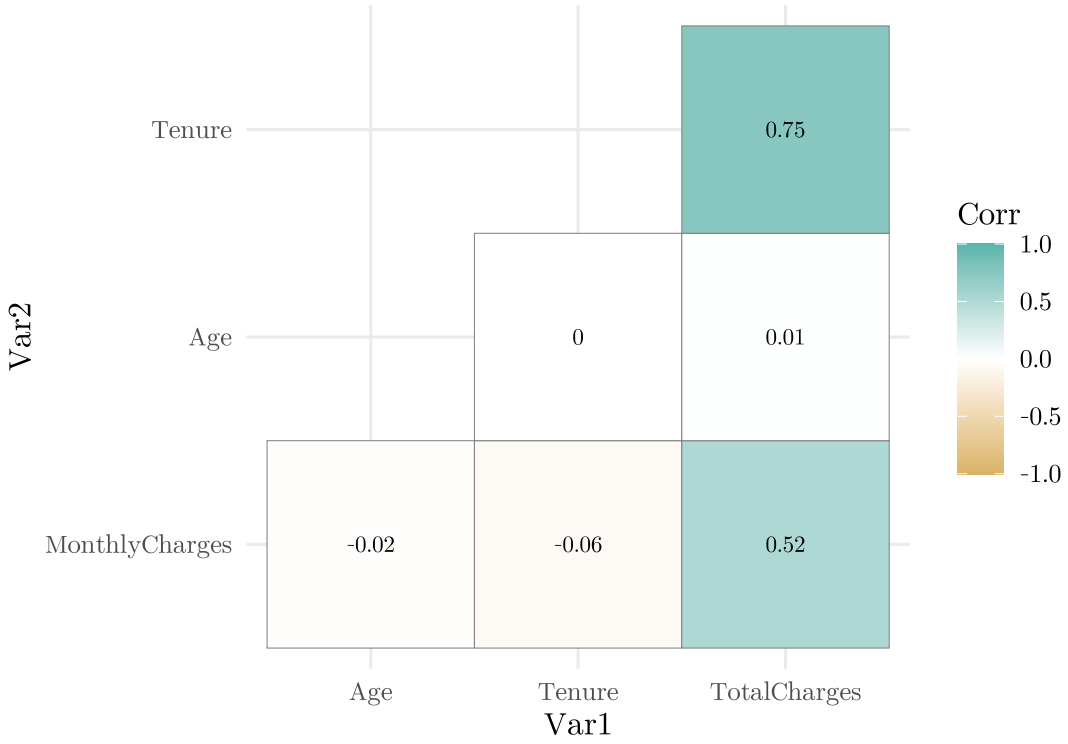
4.2 At-risk cohorts (cohort flagging)

Create a derived feature for risk segments,


Churn Rate by Risk Cohorts
100%
75%
50%
25%
0%
count
High Risk
Low/Mediium Risk
RiskCohort
Churn
No
Yes

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 → 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