Customer Churn Analysis

Aheer Srabon

2025-08-17

Table of contents

1 Numeric Distributions (Histograms) 2

2 Numeric Densities by Churn 3

3 Numeric Box Plots by Churn 4

4 Categorical Counts (Phone/Internet/Contract) 5

5 Categorical Proportions by Churn 6

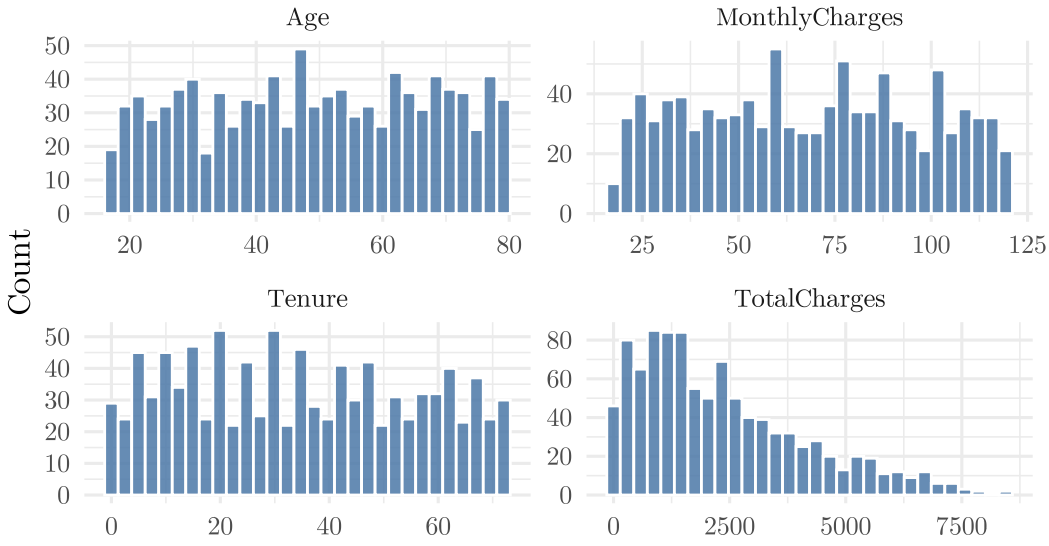
6 Missingness Overview 7

7 Skewness 8
- 7.1 Quick skew check (numeric) . . . 8
- 7.2 Skew check . . . 8

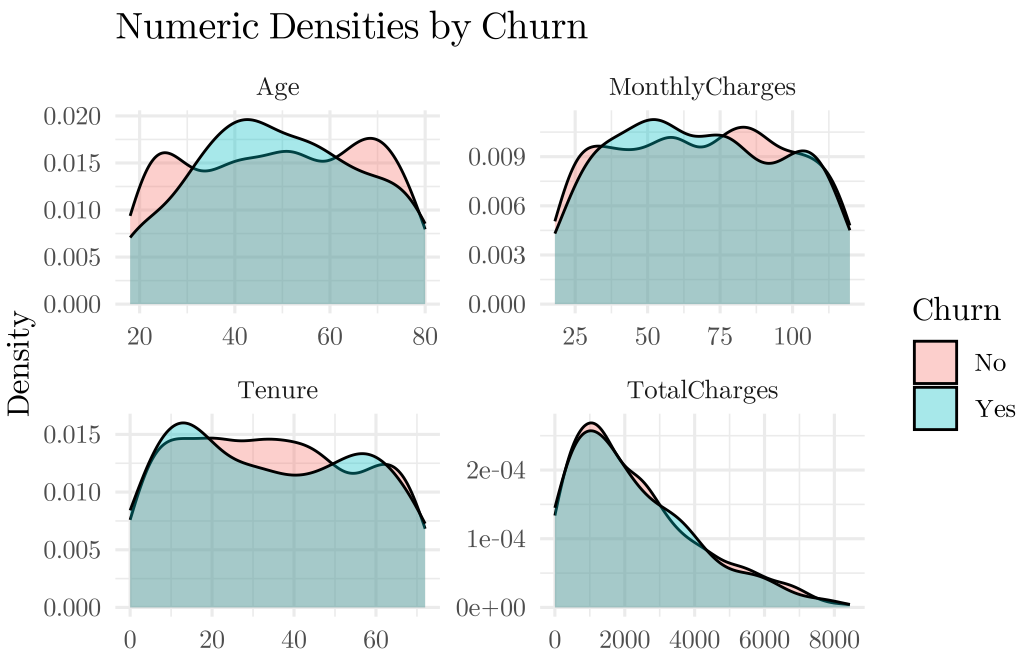
8 Churn vs Non-Churn Proportions 11
- 8.1 Stacked bar (counts by category) . . . 11
- 8.2 Normalized bar (proportions within each category) . . . 12

1 Numeric Distributions (Histograms)

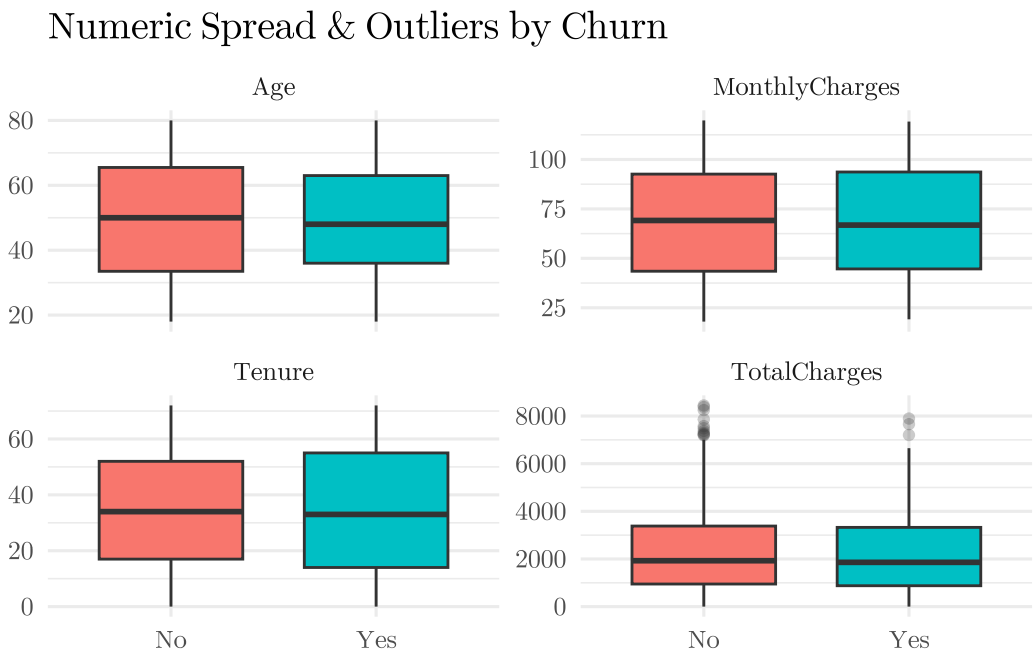
```
[1] "X"              "CustomerID"      "Gender"          "Age"
[5] "Tenure"         "PhoneService"    "InternetService" "Contract"
[9] "MonthlyCharges" "TotalCharges"    "Churn"
```



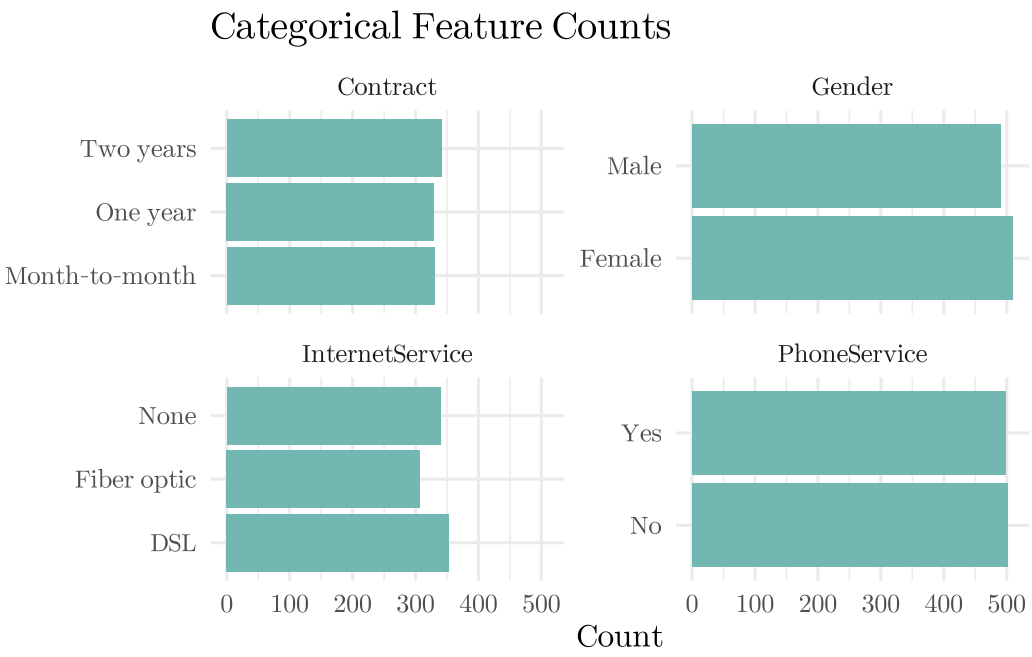
2 Numeric Densities by Churn



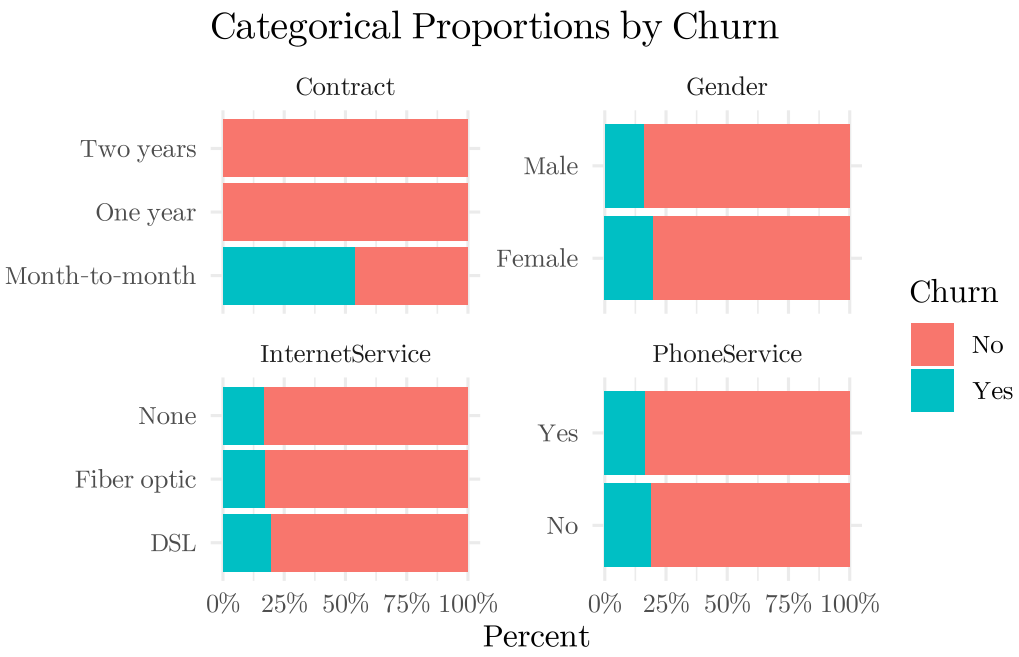
3 Numeric Box Plots by Churn



4 Categorical Counts (Phone/Internet/Contract)



5 Categorical Proportions by Churn



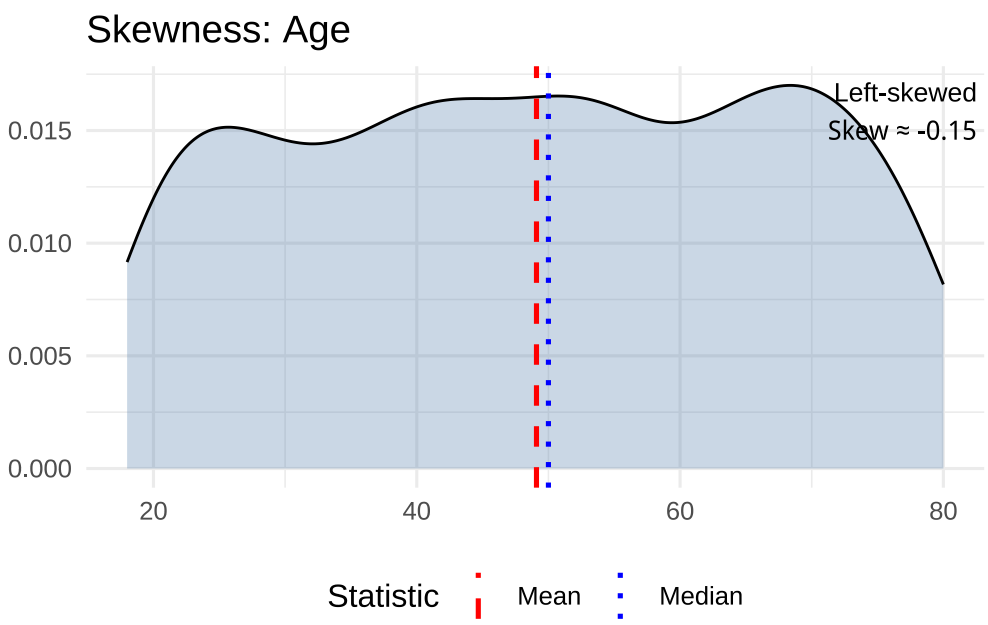
6 Missingness Overview

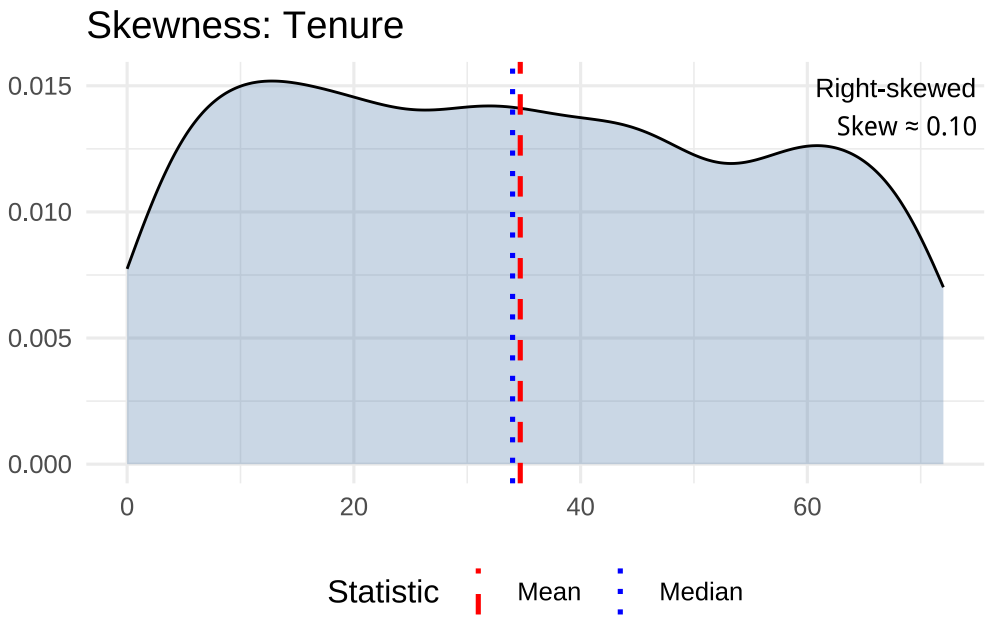
7 Skewness

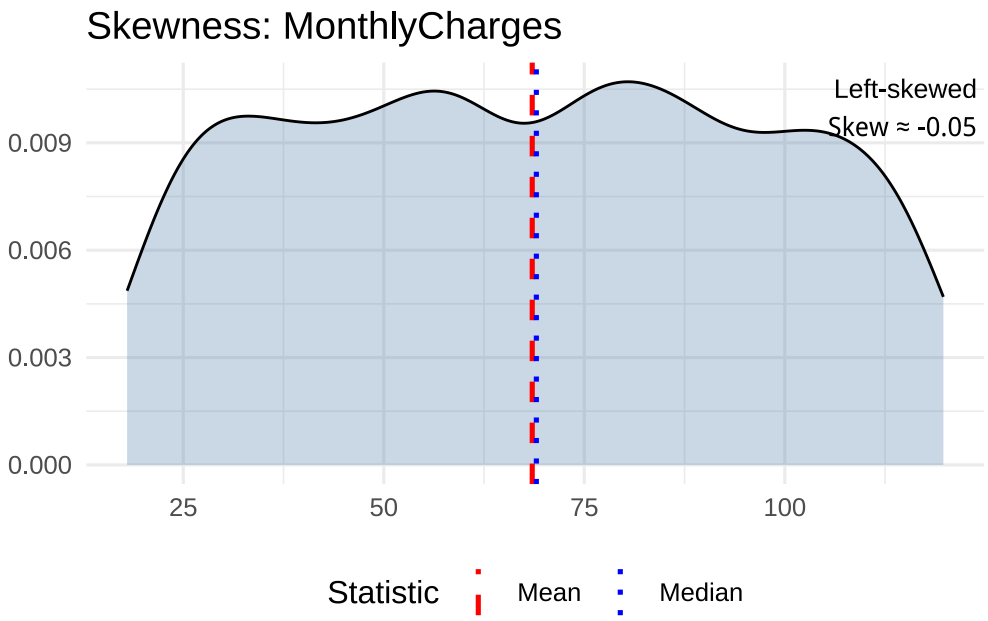
7.1 Quick skew check (numeric)

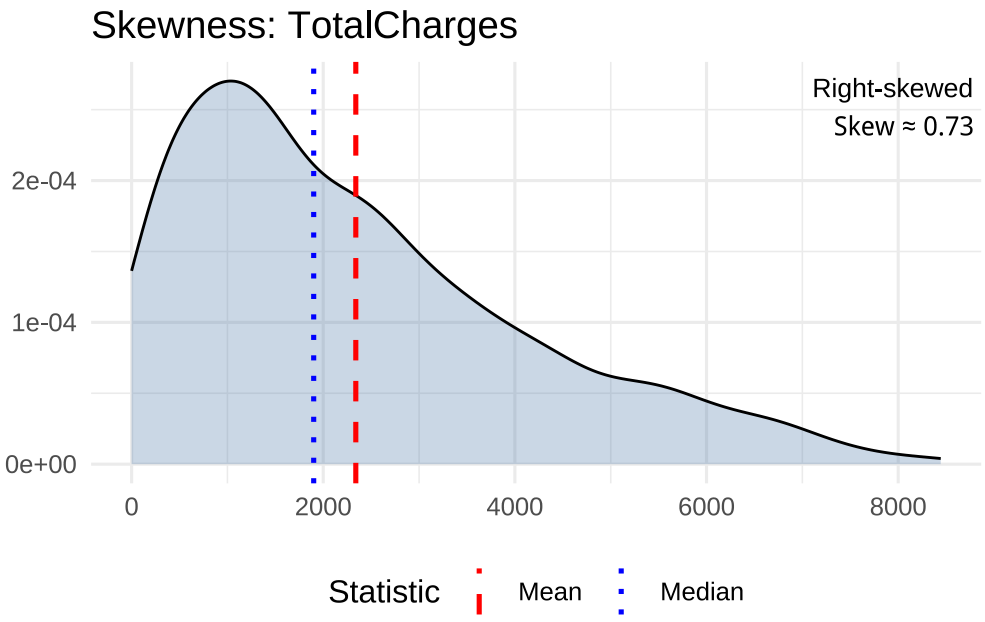
```
# A tibble: 4 x 7
  feature            n   mean median     sd   p99 skew_hint
  <chr>          <int>  <dbl>  <dbl>  <dbl> <dbl> <chr>
1 TotalCharges    1000 2340.  1900.  1808.  7248. Right-skewed
2 Age             1000   49.1   50     18.2   80  Left-skewed
3 Tenure          1000   34.7   34     21.0   72  Right-skewed
4 MonthlyCharges  1000   68.5   69.0   29.1  119. Left-skewed
```

7.2 Skew check




Skewness: Tenure
0.015
0.010
0.005
0.000
0
20
40
60
Right-skewed
Skew ≈ 0.10
Statistic
Mean
Median


Skewness: MonthlyCharges
0.009
0.006
0.003
0.000
25
50
75
100
Left-skewed
Skew ≈ -0.05
Statistic
Mean
Median

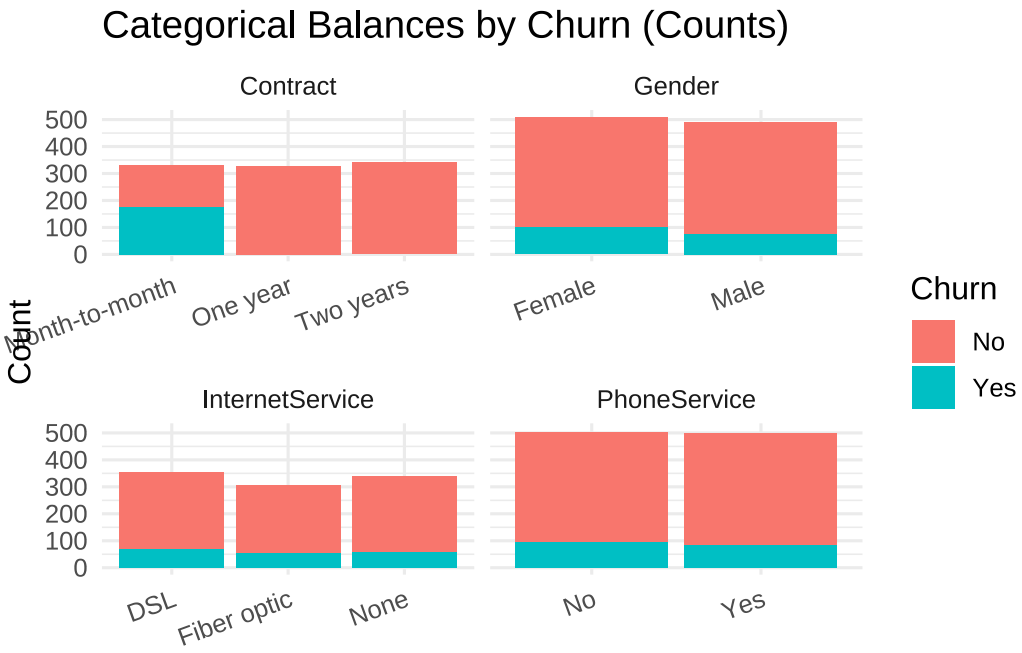

Skewness: TotalCharges
Right-skewed
Skew ≈ 0.73
2e-04
1e-04
0e+00
0
2000
4000
6000
8000
Statistic
Mean
Median

8 Churn vs Non-Churn Proportions

```
# Select the categorical features of interest
cat_features <- c("Gender", "PhoneService", "InternetService", "Contract")

cat_long_churn <- df_use %>%
  select(Churn, all_of(cat_features)) %>%
  pivot_longer(-Churn, names_to = "feature", values_to = "level") %>%
  drop_na(level, Churn)
```

8.1 Stacked bar (counts by category)



8.2 Normalized bar (proportions within each category)