

# Customer Churn Analysis

Aheer Srabon

2025-08-17

## Table of contents

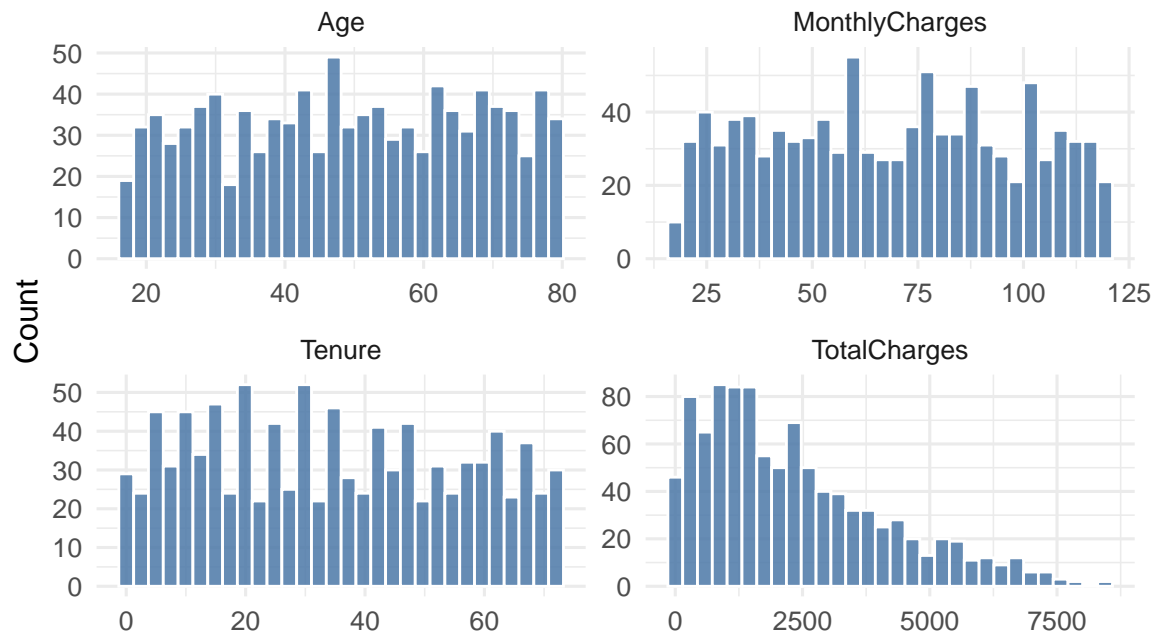
1	Numeric Distributions (Histograms)	1
2	Numeric Densities by Churn	2
3	Numeric Box Plots by Churn	3
4	Categorical Counts (Phone/Internet/Contract)	3
5	Categorical Proportions by Churn	4
6	Missingness Overview	4
7	Skewness	5
7.1	Quick skew check (numeric) . . . . .	5
7.2	Skew check . . . . .	6
8	Churn vs Non-Churn Proportions	7
8.1	Stacked bar (counts by category) . . . . .	8
8.2	Normalized bar (proportions within each category) . . . . .	9

## 1 Numeric Distributions (Histograms)

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

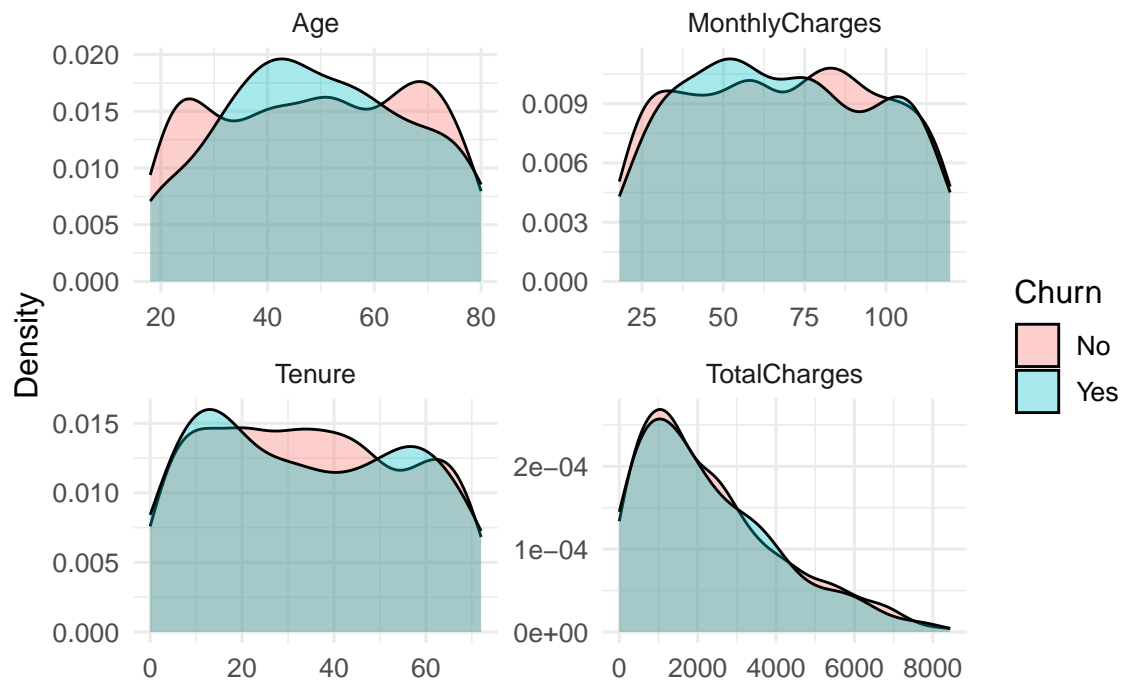
## Numeric Feature Distributions

### Age, Tenure, MonthlyCharges, TotalCharges



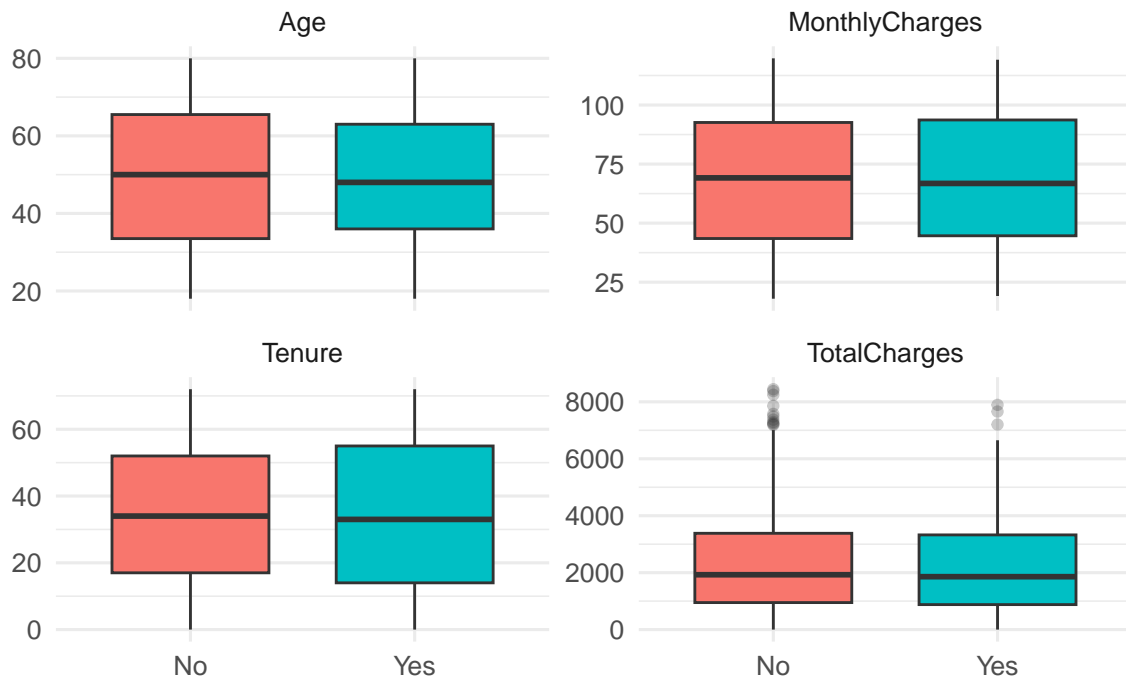
## 2 Numeric Densities by Churn

### Numeric Densities by Churn



### 3 Numeric Box Plots by Churn

#### Numeric Spread & Outliers by Churn

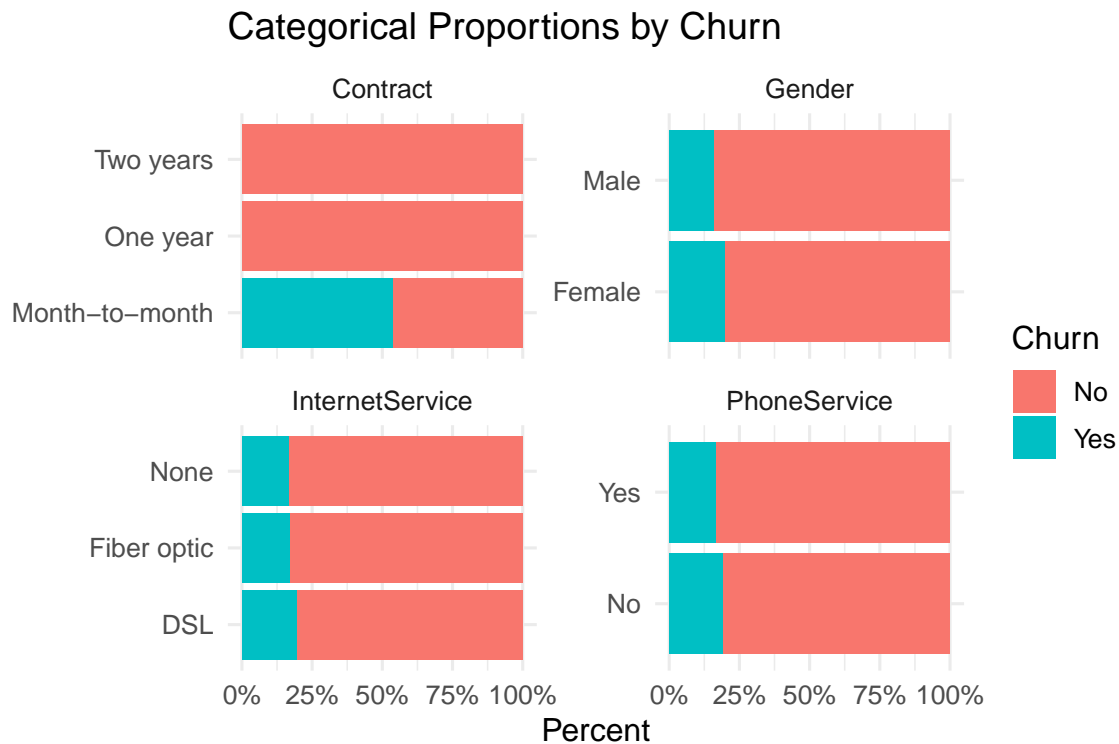


### 4 Categorical Counts (Phone/Internet/Contract)

#### Categorical Feature Counts



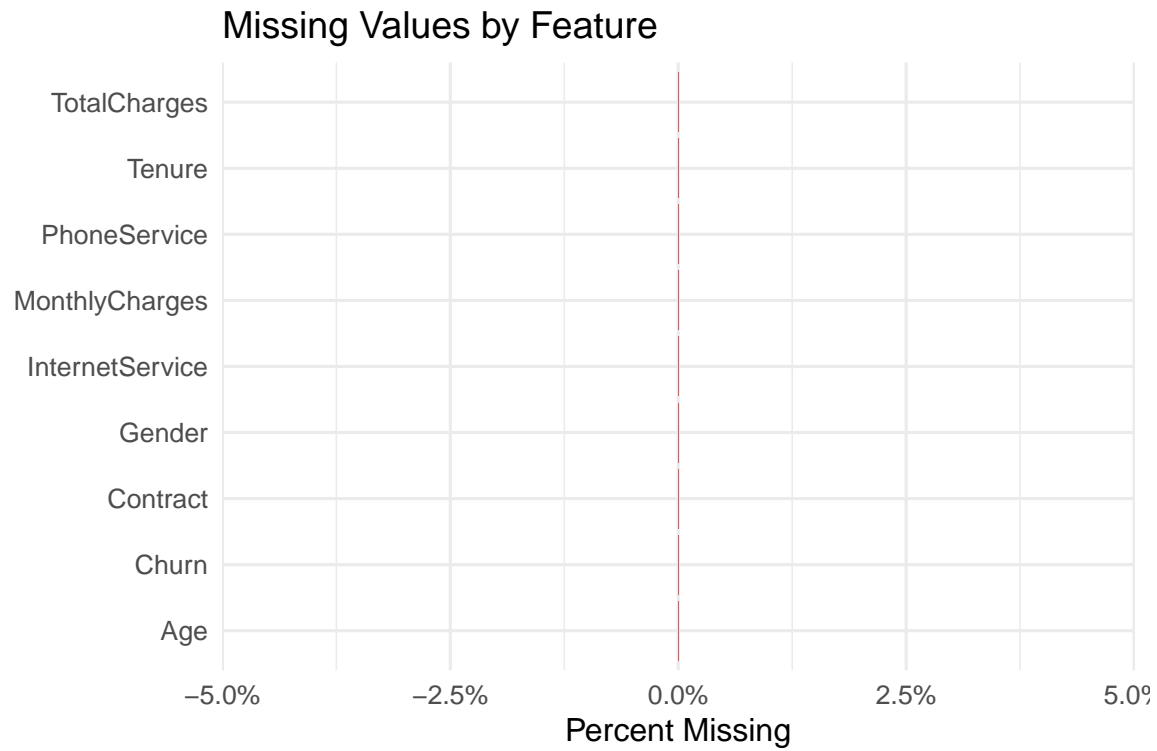
## 5 Categorical Proportions by Churn



## 6 Missingness Overview

```
missing_df <- df_use %>%
  summarise(across(everything(), ~ sum(is.na(.)))) %>%
  pivot_longer(everything(), names_to = "feature", values_to = "n_missing") %>%
  mutate(pct = n_missing / nrow(df_use))

ggplot(missing_df, aes(reorder(feature, pct), pct)) +
  geom_col(fill = "#E45756") +
  coord_flip() +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Missing Values by Feature", x = NULL, y = "Percent Missing")
```



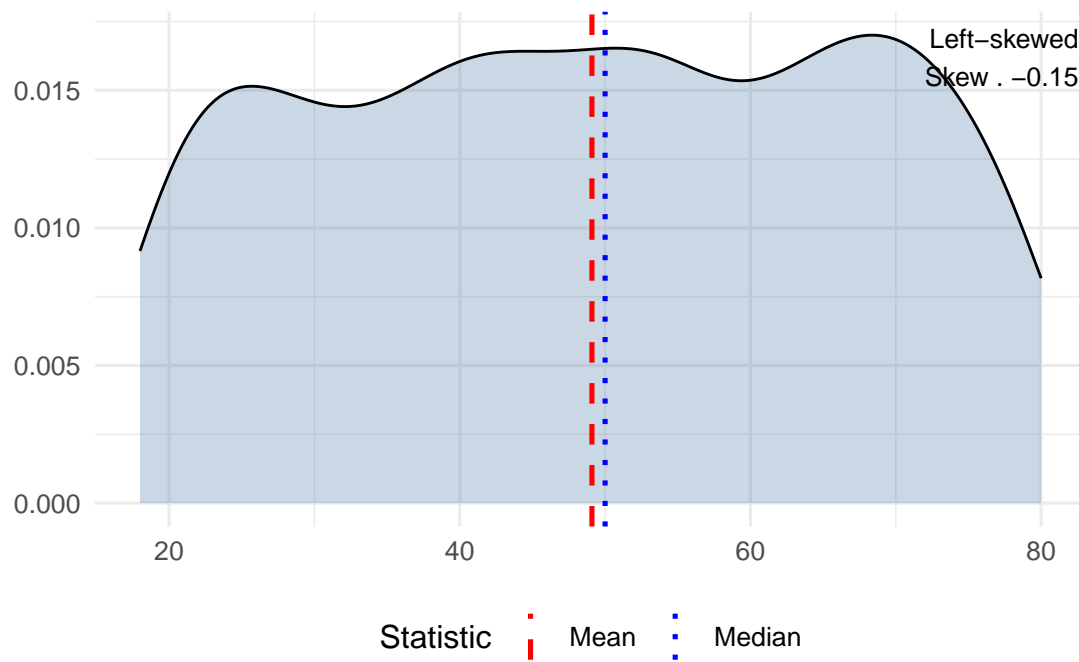
## 7 Skewness

### 7.1 Quick skew check (numeric)

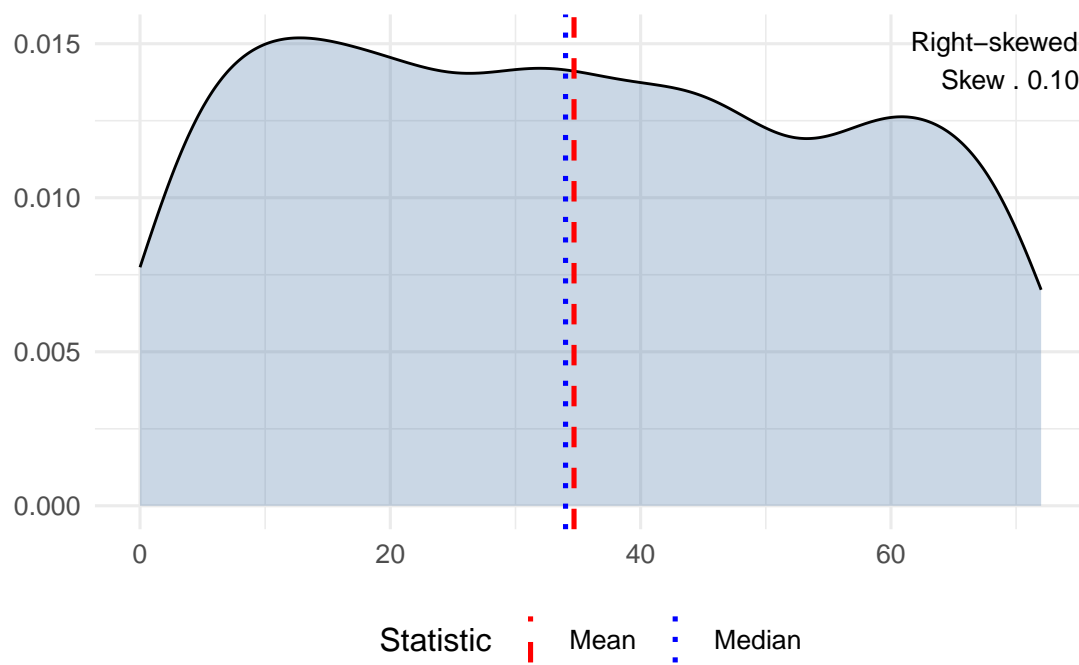
feature	n	mean	median	sd	p99	skew_hint
TotalCharges	1000	2339.68434	1900.125	1808.26360	7247.8838	Right-skewed
Age	1000	49.09300	50.000	18.16751	80.0000	Left-skewed
Tenure	1000	34.67800	34.000	21.03880	72.0000	Right-skewed
MonthlyCharges	1000	68.51068	69.020	29.07392	118.9306	Left-skewed

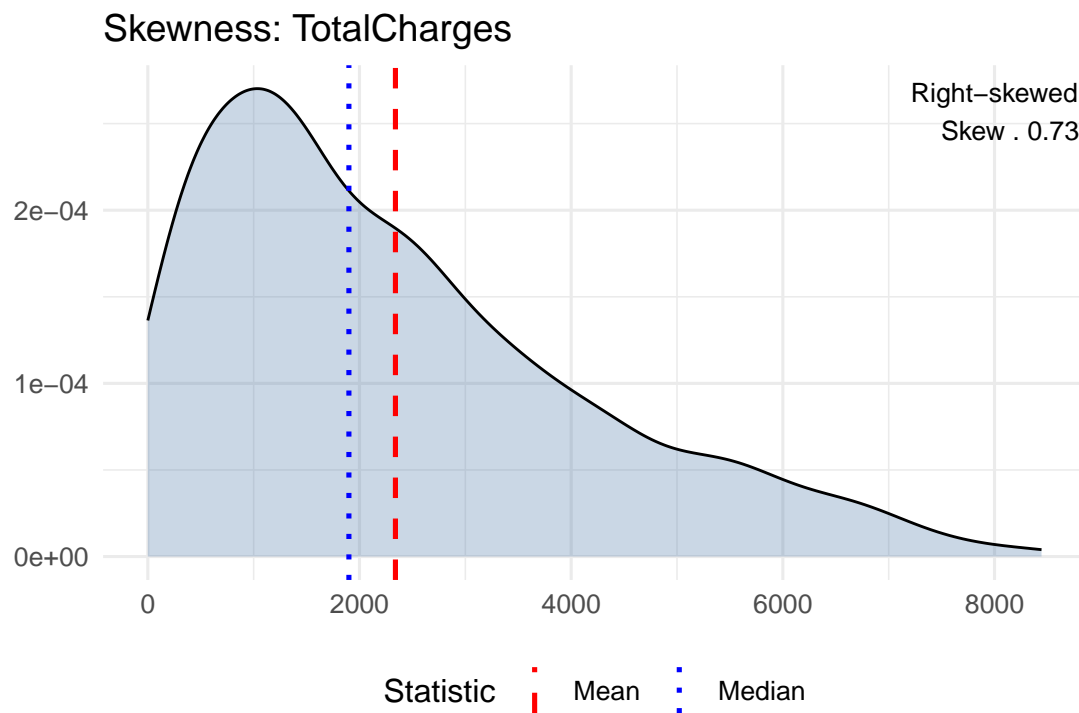
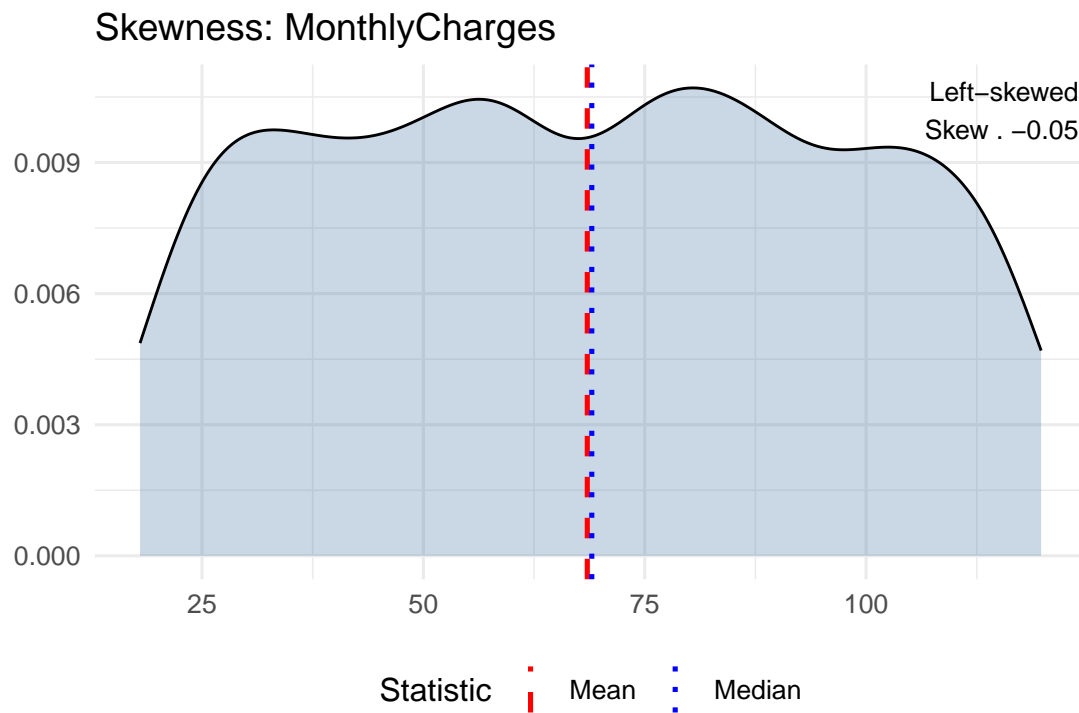
## 7.2 Skew check

### Skewness: Age



### Skewness: Tenure





## 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)

### Categorical Balances by Churn (Counts)





## 8.2 Normalized bar (proportions within each category)

Categorical Balances by Churn (Proportions)

