Churn:AD&D - fictional telecom

AD&D

Matt Curcio
2/2023
github.com/mccurcio/Churn_Analysis_Modeling

Summary

- **Loss of AD&D customers is a constant problem.**

- **Churn: 21 Factors x 5000 observations**
  (Oct 5, 2022 to Nov. 5, 2022)
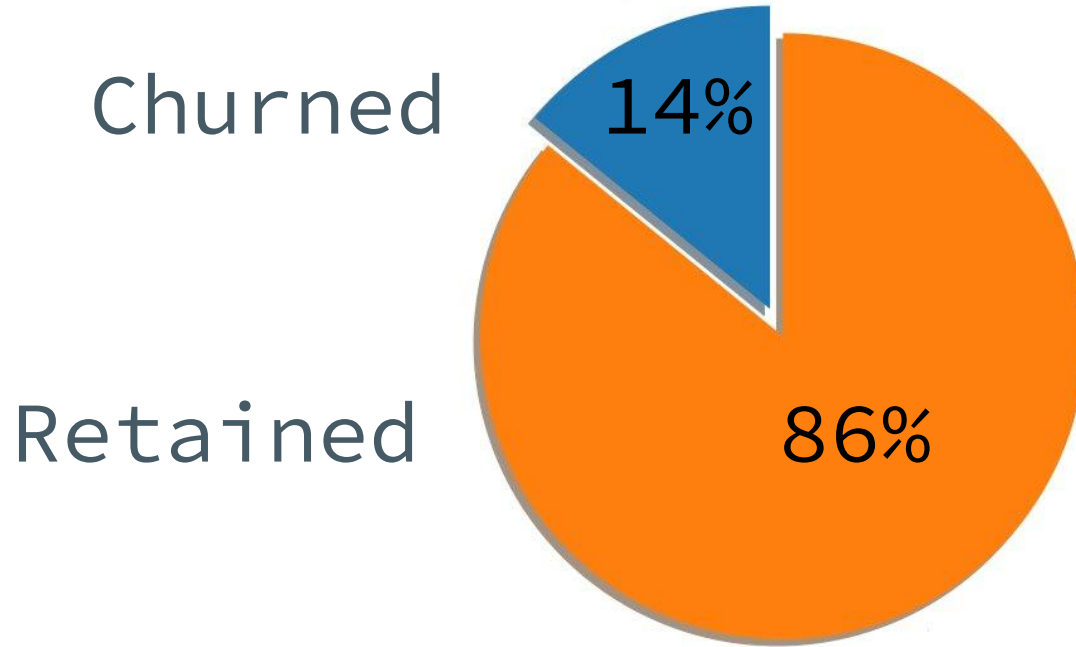
Aims:

- **What Performance Metrics are obtained with**
  Logistic Regression model?

Initial Data Statistics

Initial Data Analysis - Churn Vs Retained
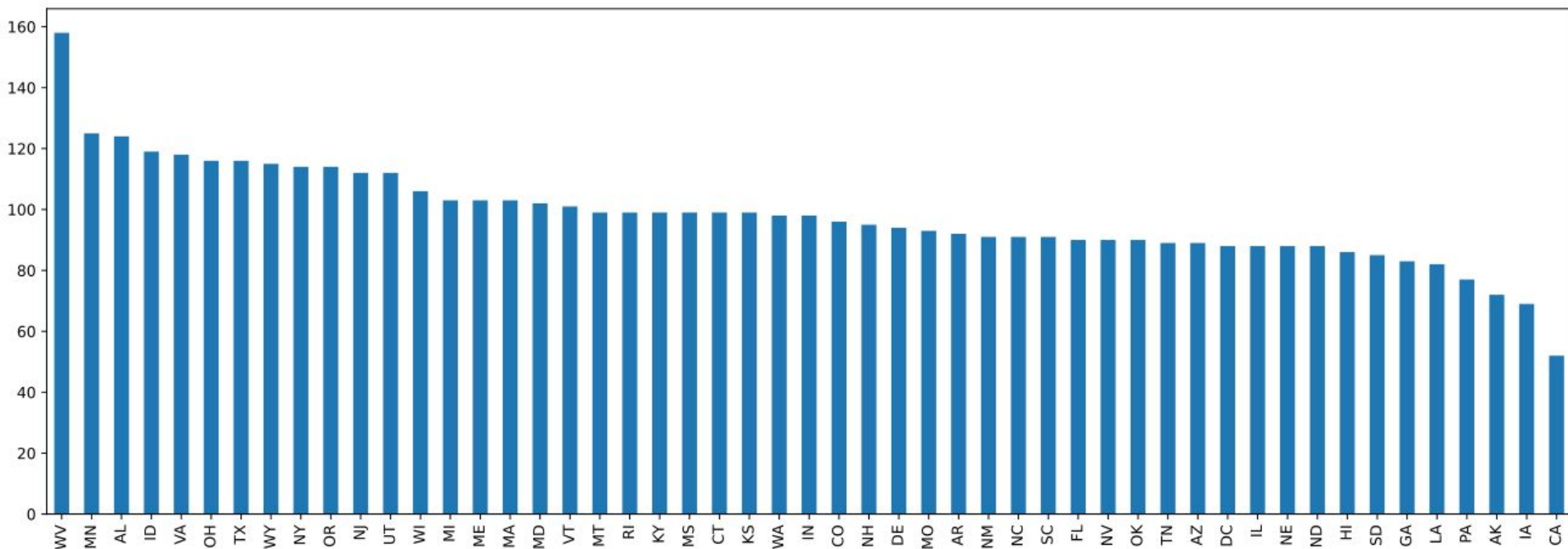
Proportion of customer churned and retained, ~1:7.

Churned 14%

Retained 86%

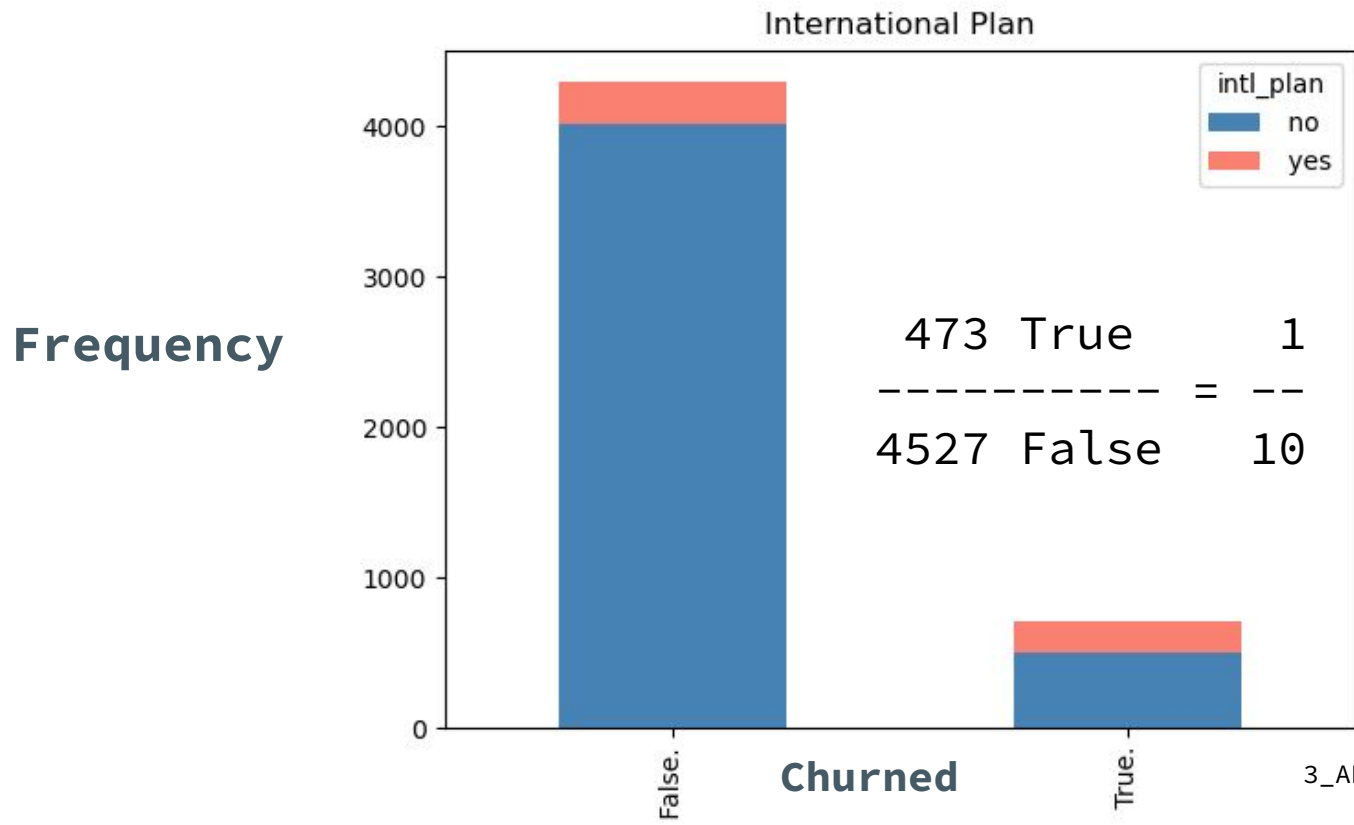# Initial Data Analysis - States

– **All 50 states + D.C. represented**

Initial Data Analysis - International Plans

Frequency

Churned

$$\frac{473 \text{ True}}{4527 \text{ False}} = \frac{1}{10}$$

Initial Data Analysis – VoiceMail Plans

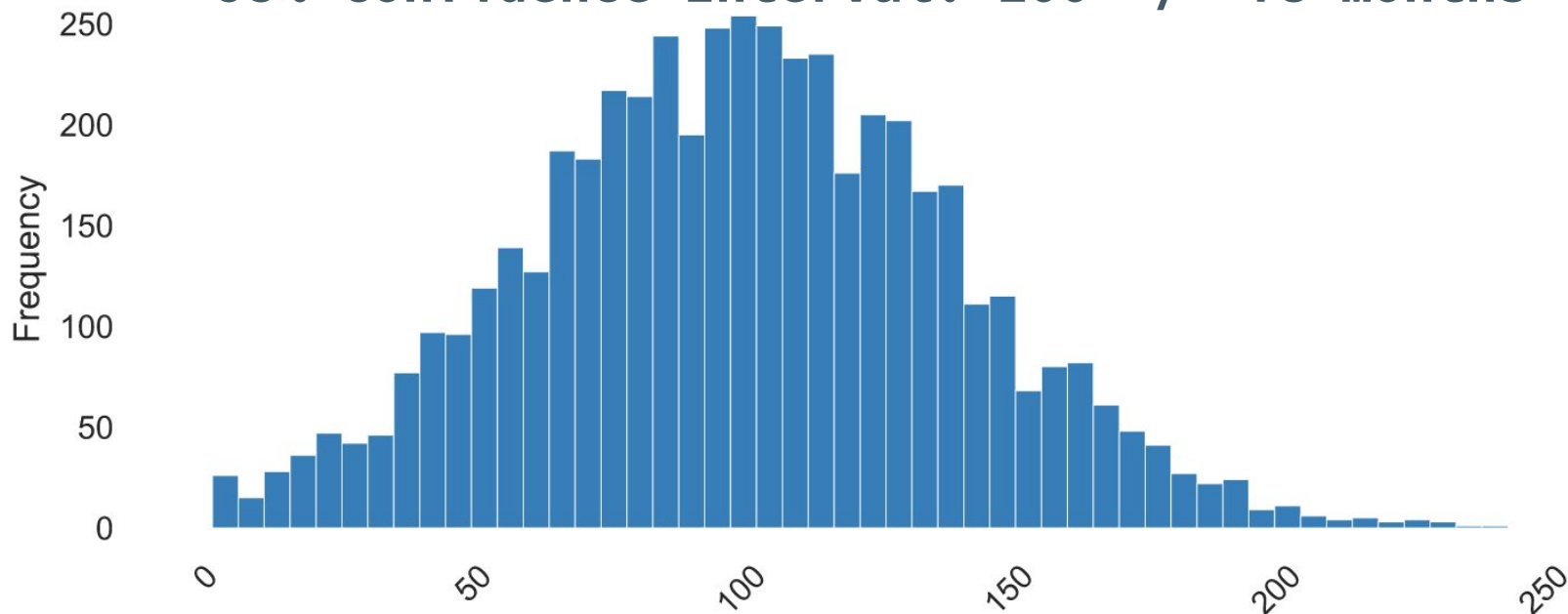

Voice Mail Plan

Frequency

1323 True
——————————
3677 False
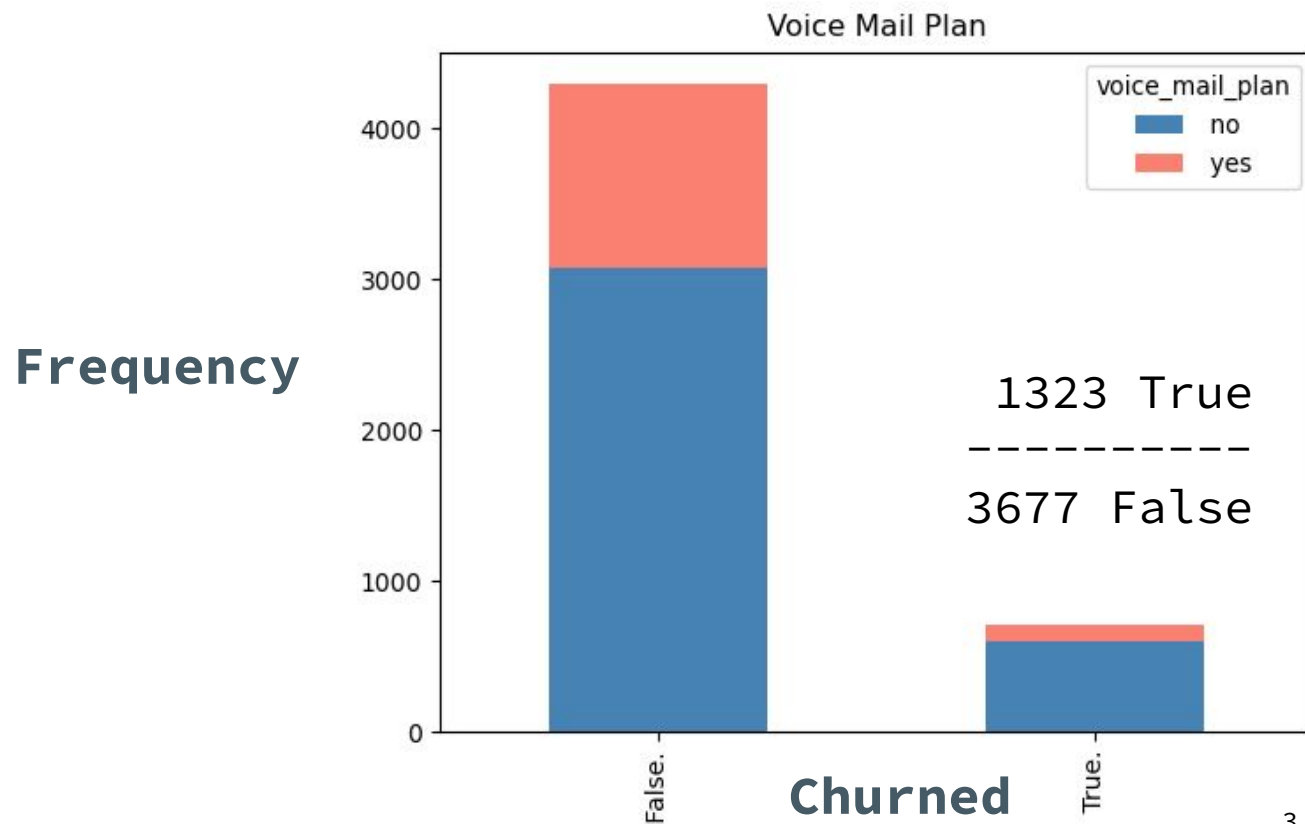
Churned

Exploratory Data Analysis

Customer Account Length (Months)

95% Confidence Interval: 100 +/- 78 months

Histogram

Frequency

Histogram with fixed size bins (bins=50)

Correlation Matrix

Daily Calls ≡ Daily Min.
Eve. Calls ≡ Eve. Min.
Night Calls ≡ Night Min.
Int. Call ≡ Int. Min.

Therefore Remove
Calls or Min.
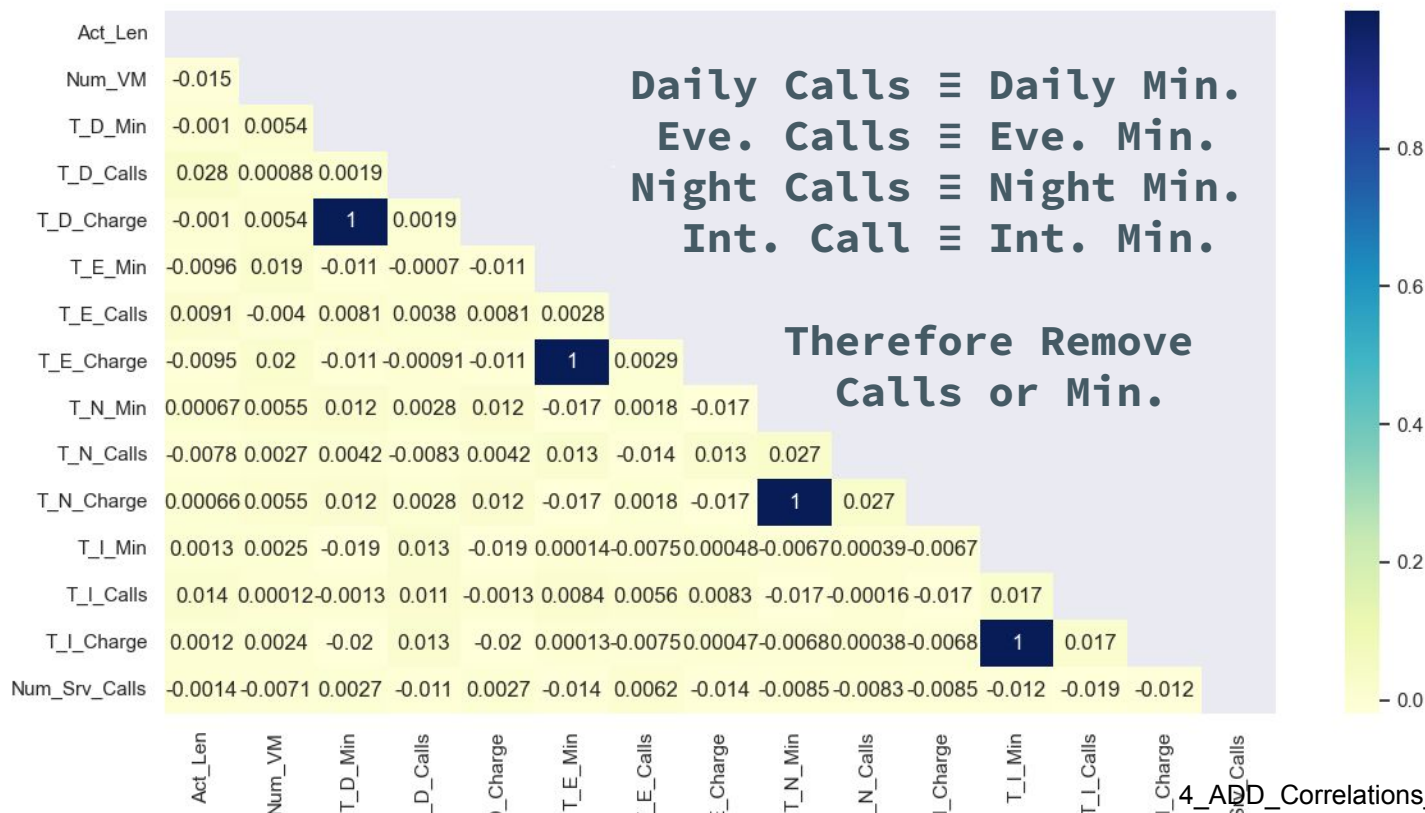
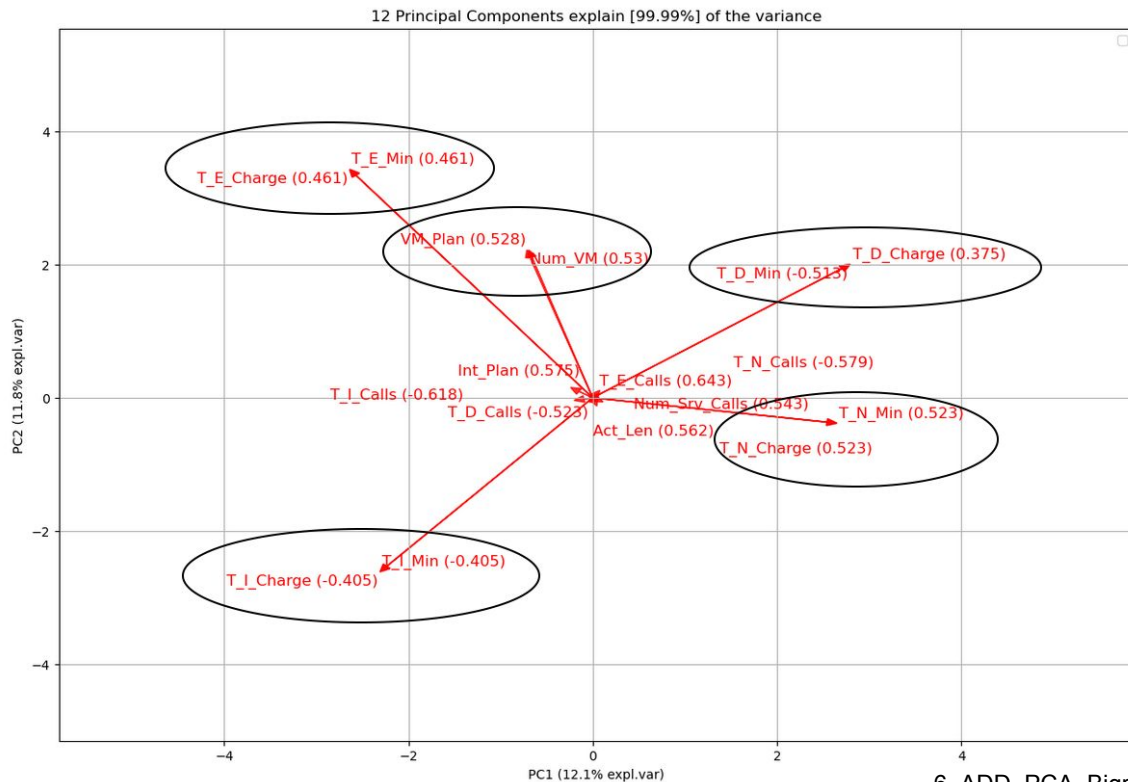# EDA Using -Pandas Profiling- Package

Overview   **Alerts 14**   Reproduction

Alerts

`state` has a high cardinality: 51 distinct values   **High cardinality**

`phone_number` has a high cardinality: 5000 distinct values   **High cardinality**

`total_eve_charge` has a high cardinality: 1660 distinct values   **High cardinality**

`total_day_minutes` is highly overall correlated with `total_day_charge`   **High correlation**

`total_day_charge` is highly overall correlated with `total_day_minutes`   **High correlation**

`total_night_minutes` is highly overall correlated with `total_night_charge`   **High correlation**

`total_night_charge` is highly overall correlated with `total_night_minutes`   **High correlation**

`total_intl_minutes` is highly overall correlated with `total_intl_charge`   **High correlation**

`total_intl_charge` is highly overall correlated with `total_intl_minutes`   **High correlation**

`intl_plan` is highly imbalanced (54.8%)   **Imbalance**

`phone_number` is uniformly distributed   **Uniform**

`phone_number` has unique values   **Unique**

`number_vmail_messages` has 3678 (73.6%) zeros   **Zeros**

`number_customer_service_calls` has 1023 (20.5%) zeros   **Zeros**

Source:
assets/churn_eda_profiling_report.html

Principal Components – Erdogant P.C. Package

12 Principal Components explain [99.99%] of the variance

Logistic Regression Model

Classification of Test Results:

1. Overall Accuracy Score: 85.13%

2. Precision: 84.50%

3. Recall: 85.13%

4. F1-score: 84.80%