SYRIA-TEL CUSTOMER CHURN

- BEDAN KIBUNJA CHEGE
- DSFT 09
- MORINGA SCHOOL

BUSINESS UNDERSTANDING

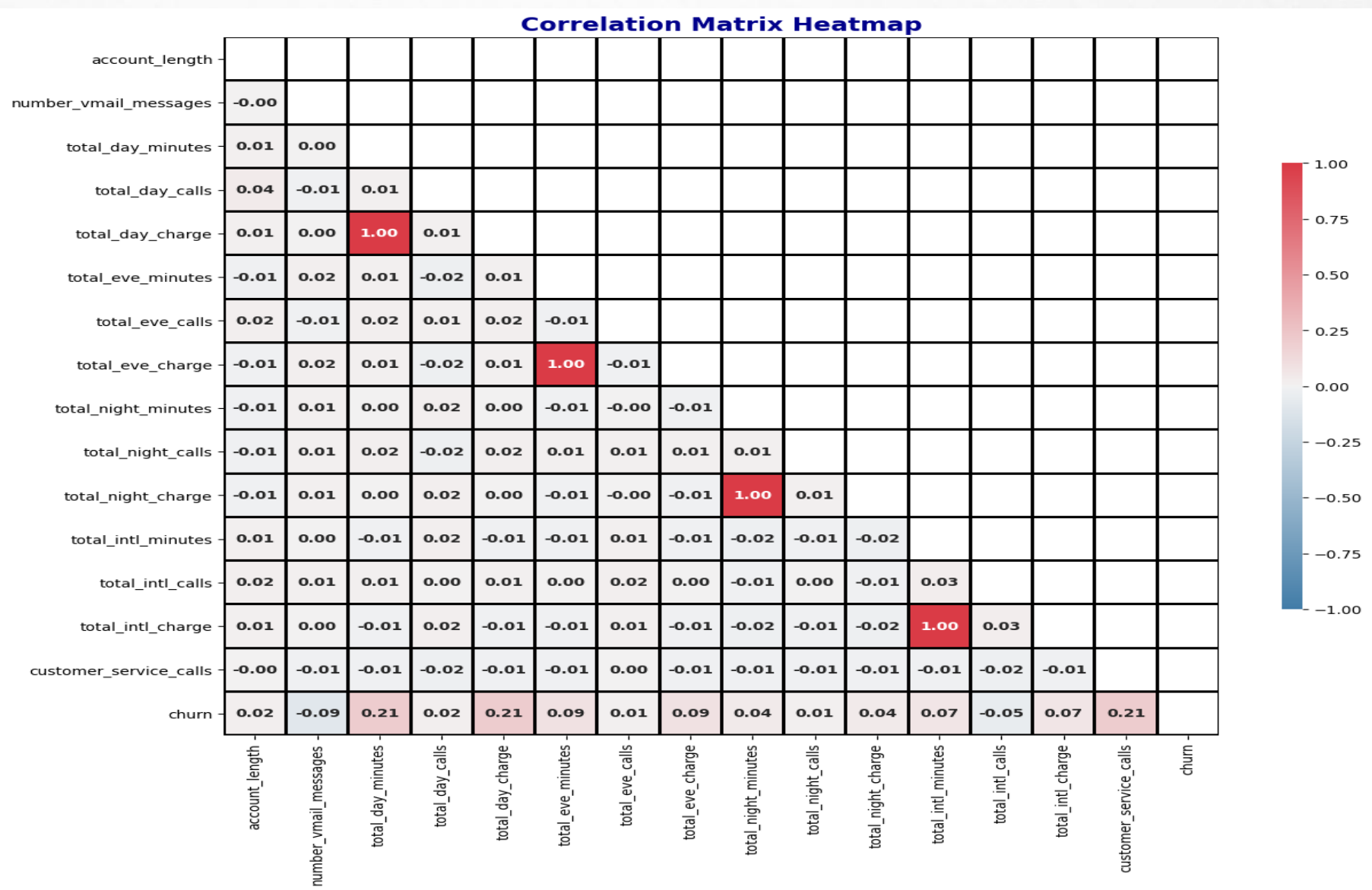
- **PROJECT OVERVIEW: CUSTOMER CHURN PREDICTION FOR SYRIA-TEL**
- **OBJECTIVE:** DEVELOP A PREDICTIVE MODEL TO IDENTIFY CUSTOMERS LIKELY TO CHURN.
- **IMPORTANCE:**
  - RETAINING CUSTOMERS IS MORE COST-EFFECTIVE THAN ACQUIRING NEW ONES.
  - REDUCING CHURN CAN LOWER REVENUE LOSS AND BOOST CUSTOMER SATISFACTION.
- **GOAL:**
  - BUILD A BINARY CLASSIFICATION MODEL TO PREDICT CUSTOMER CHURN.
  - ANALYZE CUSTOMER DATA TO UNCOVER INSIGHTS AND REDUCE CHURN IMPACT.

DATA UNDERSTANDING

- **DATA CHARACTERISTICS:**
- **NUMERICAL FEATURES:** REPRESENT CUSTOMER USAGE AND INTERACTIONS; CRITICAL FOR CHURN PATTERN IDENTIFICATION.
- **CATEGORICAL FEATURES:** INCLUDES STATE, INTERNATIONAL PLAN, VOICE MAIL PLAN; USEFUL FOR CHURN INSIGHTS.
- **CLASS IMBALANCE:** CHURN IS LESS FREQUENT THAN NON-CHURN; NEEDS ADDRESSING TO PREVENT BIASED PREDICTIONS.
- **PREPROCESSING CONSIDERATIONS:**
- **MULTICOLLINEARITY:** REMOVED FEATURES TO ENSURE INDEPENDENT CONTRIBUTIONS.
- **FEATURE ENGINEERING:** ADDITIONAL FEATURES OR TRANSFORMATIONS MAY BE REQUIRED.
- **NORMALIZATION & SCALING:** APPLIED TO BALANCE FEATURE CONTRIBUTIONS.

DATA CLEANING & PREPARATION

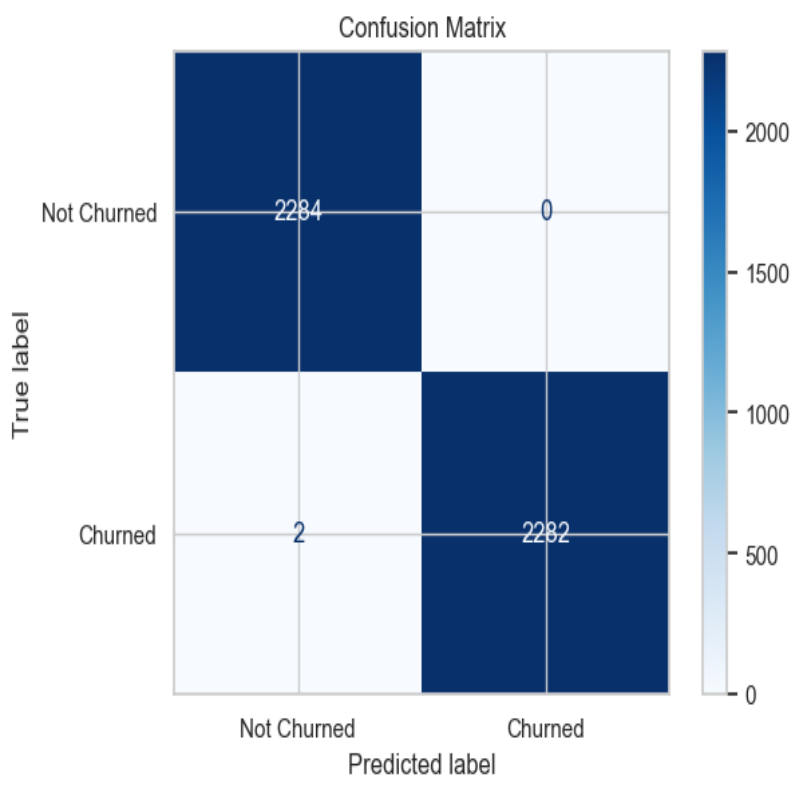
- **NO MISSING VALUES:** COMPLETE DATASET WITH NO MISSING ENTRIES.
- Area code Column: Only three distinct area codes observed across 51 states. Indicates potential data reliability issues. Decision: Drop area code from analysis.
- Phone number Column: Does not contribute to churn prediction. Decision: Remove from analysis.
- Perfect Correlations Identified: Between total_x_minutes and total_x_charge (e.g., total_day_minutes with total_day_charge).
- Action: Drop total_x_minutes columns to address multicollinearity. Retain total_x_charge columns to avoid redundancy.


Correlation Matrix Heatmap
account_length
number_vmail_messages
total_day_minutes
total_day_calls
total_day_charge
total_eve_minutes
total_eve_calls
total_eve_charge
total_night_minutes
total_night_calls
total_night_charge
total_intl_minutes
total_intl_calls
total_intl_charge
customer_service_calls
churn
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

DATA ANALYSIS

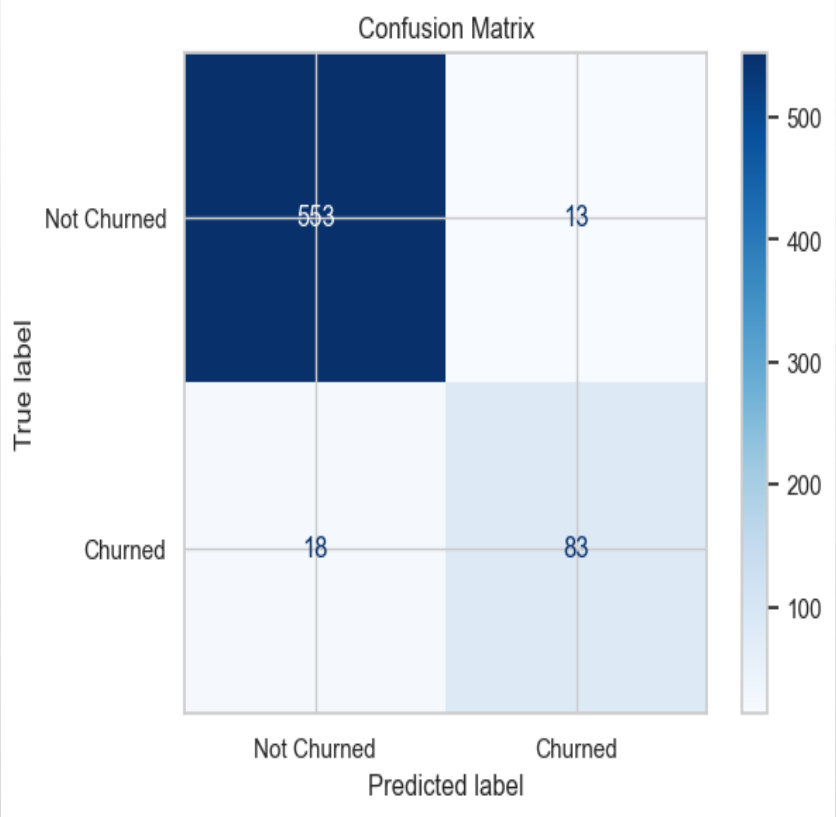
- MODEL EVALUATION: WE TESTED MULTIPLE MODELS TO DETERMINE WHICH WOULD BEST FIT OUR DATA:
- LOGISTIC REGRESSION: SERVED AS OUR BASELINE MODEL. WE TRAINED IT ON THE TRAINING DATA AND ASSESSED ITS PERFORMANCE.
- DECISION TREE: WE EXPLORED THIS MODEL WHILE ADDRESSING OVERFITTING CONCERNS BY TUNING ITS HYPERPARAMETERS.
- XGBOOST: AS A MORE ADVANCED MODEL, XGBOOST WAS EVALUATED FOR ITS PERFORMANCE AND ABILITY TO HANDLE THE DATASET'S COMPLEXITY. THE GOAL WAS TO IDENTIFY THE MODEL THAT OFFERED THE BEST PERFORMANCE METRICS, SUCH AS ACCURACY, RECALL, AND PRECISION, AND WAS WELL-SUITED FOR PREDICTING CUSTOMER CHURN.

XGBOOST



- **TRAIN DATA**
- EXTREMELY HIGH ACCURACY: THE OVERALL ACCURACY OF THE RESNET50 MODEL IS EXCEPTIONALLY HIGH AT APPROXIMATELY 99.98%. THIS INDICATES THAT THE MODEL IS MAKING CORRECT PREDICTIONS FOR A VAST MAJORITY OF INSTANCES.
- PERFECT PRECISION: THE PRECISION IS 1.0, WHICH IS THE HIGHEST POSSIBLE VALUE. THIS MEANS THAT WHENEVER THE MODEL PREDICTS A POSITIVE INSTANCE (A CUSTOMER CHURNING), IT IS ALWAYS CORRECT.
- VERY HIGH RECALL: THE RECALL IS ALSO VERY HIGH AT 99.91%. THIS MEANS THAT THE MODEL IS CAPTURING A SIGNIFICANT NUMBER OF ACTUAL POSITIVE INSTANCES (CUSTOMERS WHO CHURNED).

XGBOOST



- TEST DATA
- USING THE DATA PROVIDED, WE WERE ABLE TO CREATE A MODEL WITH 82% RECALL, MEANING OF THE CUSTOMERS WHO ARE GOING TO LEAVE, WE ARE ABLE TO IDENTIFY 82% OF THEM. WE WERE ABLE TO DO THIS WHILE MAINTAINING A HIGH ACCURACY OF 95%.

RECOMMENDATION

- IT'S EVIDENT THAT MANY STATES APPEAR IN THE TOP 20 MOST IMPORTANT FEATURES. BASED ON THESE INSIGHTS, WE RECOMMEND THE FOLLOWING ACTIONS:
- INVESTIGATE THE NEEDS OF INTERNATIONAL PLAN CUSTOMERS: CONSIDER WHETHER THERE'S A SHIFT TOWARD ONLINE COMMUNICATION TOOLS FOR INTERNATIONAL USERS, SUCH AS SKYPE, DISCORD, GOOGLE CHAT, OR FACETIME. EXPLORE THE POSSIBILITY OF OFFERING MORE ROBUST DATA PLANS TO ACCOMMODATE THIS TREND.
- AUDIT CUSTOMER SERVICE CALLS: ENSURE THAT CUSTOMERS ARE RECEIVING ADEQUATE SUPPORT. IF THE MODEL IDENTIFIES A CUSTOMER AT RISK OF CHURNING, IT MIGHT BE BENEFICIAL TO REVIEW THEIR RECENT INTERACTIONS WITH CUSTOMER SERVICE TO ADDRESS ANY UNRESOLVED ISSUES.
- REVIEW RATE COMPETITIVENESS: GIVEN THAT TOTAL DAY CHARGES ARE A SIGNIFICANT FACTOR IN OUR MODEL, IT’S IMPORTANT TO VERIFY THAT PRICING IS COMPETITIVE, PARTICULARLY IN STATES WITH HIGHER CHURN RATES.