Reducing Telecom Customer Churn

orange™

Michael Church Carson, Hadyan Fahreza, Solomon Gomez, Danyal Hamid, Vahid Hedley, Atrin Morteza Ghasemi, and Reza Soleimani

The Team

Business Analysts

Michael Church Carson
GitHub: MWCC8

Vahid Hedley
GitHub: VahidHedley
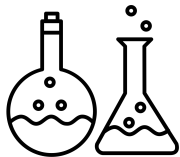
Data Scientists

Hadyan Fahreza
GitHub: hifahreza

Atrin Morteza Ghasemi
GitHub: Atrin-Morteza-Ghasemi

Reza Soleimani
GitHub: rsoleimani

Data Analysts

Solomon Gomez
GitHub: solomongomez

Danyal Hamid
GitHub: danyalhamid1996

Business Context

➔ Orange is a French multinational telecommunications company with approximately **266 million customers**.

➔ To maintain its revenue, Orange needs to **minimize** the rate at which its current **customers are leaving**

➔ Observed **churn rate** is approximately **17%**

➔ Need to **predict churn** so we can learn how to reduce it.
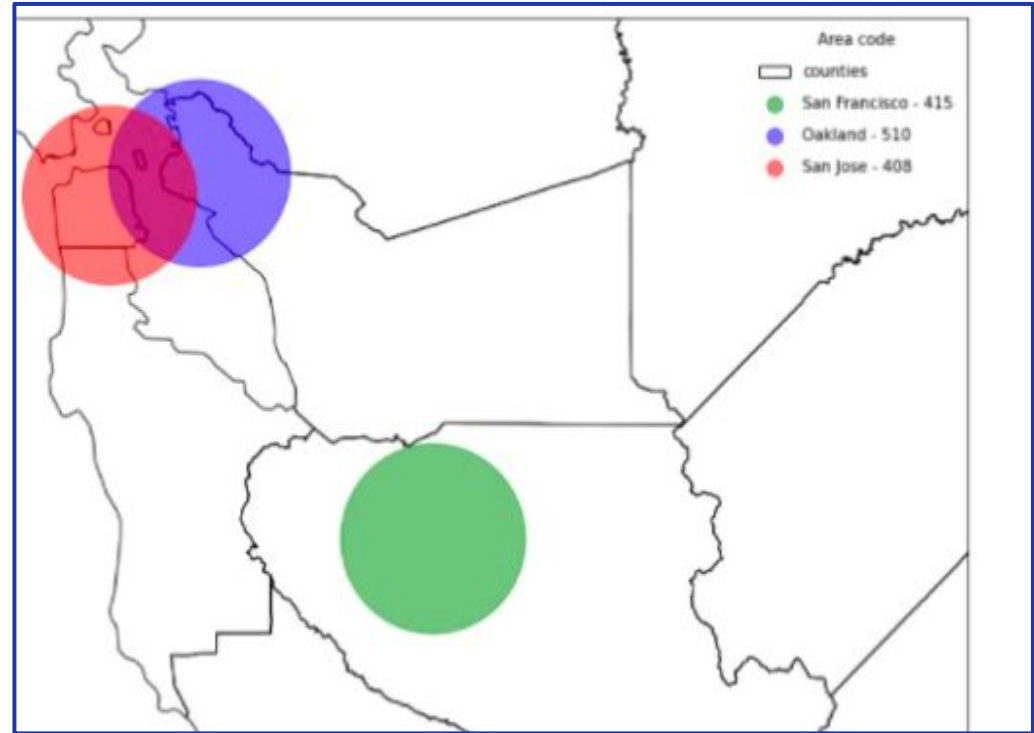

WHY DO CUSTOMERS LEAVE?

Data and Hypothesis

- Phone plan feature and usage data on **2666** of Orange's American customers, including charges, call type, number of service calls, and state of residence.
- Target = **Churn**, binary (0 or 1)
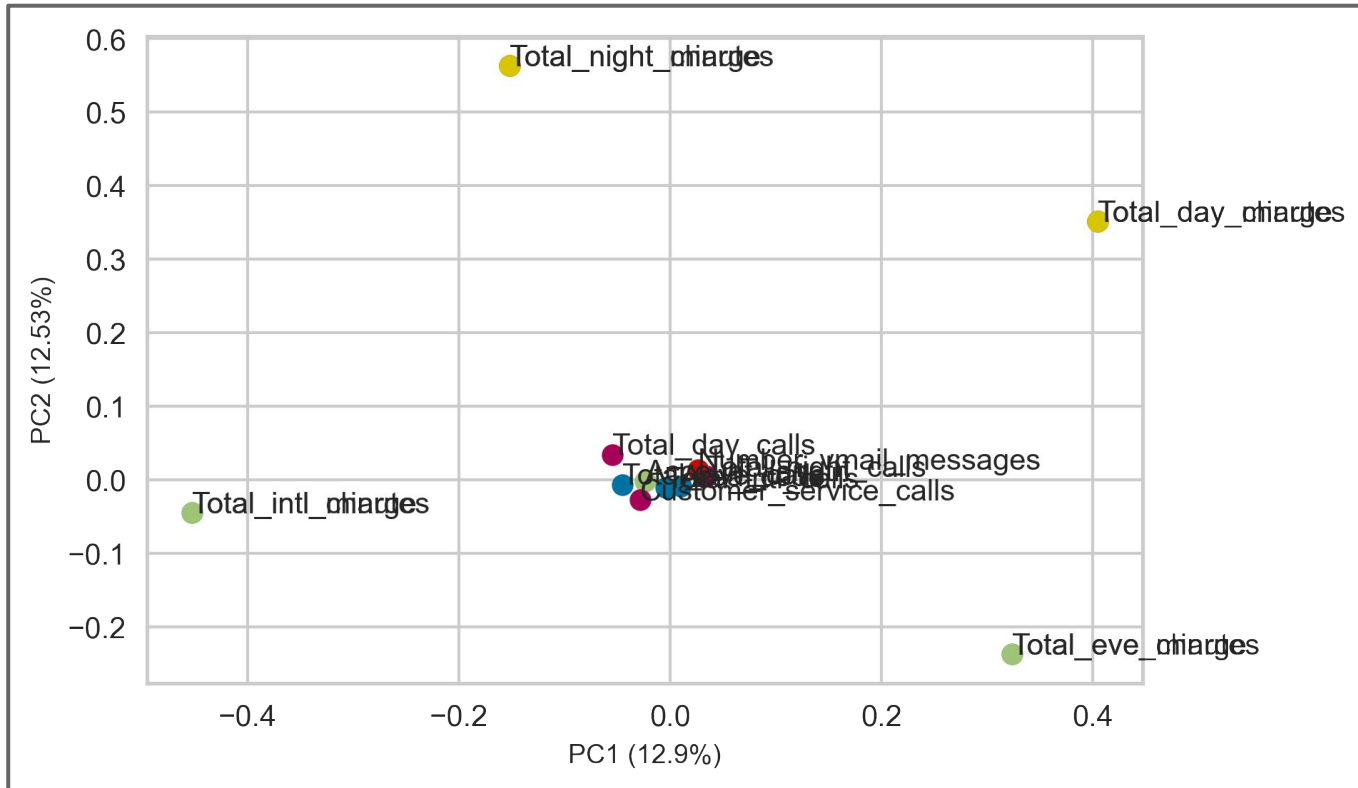
Hypothesis:

- Customers who **call the service agents** more often will be more likely to **churn**
- Customers who **pay more** will be more likely to **churn**
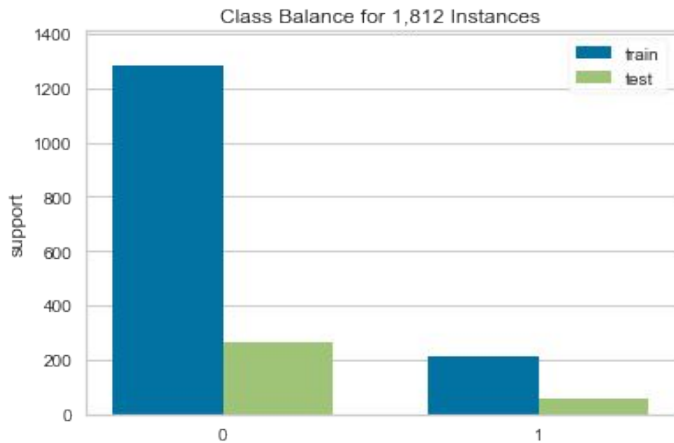
Principal Component Analysis
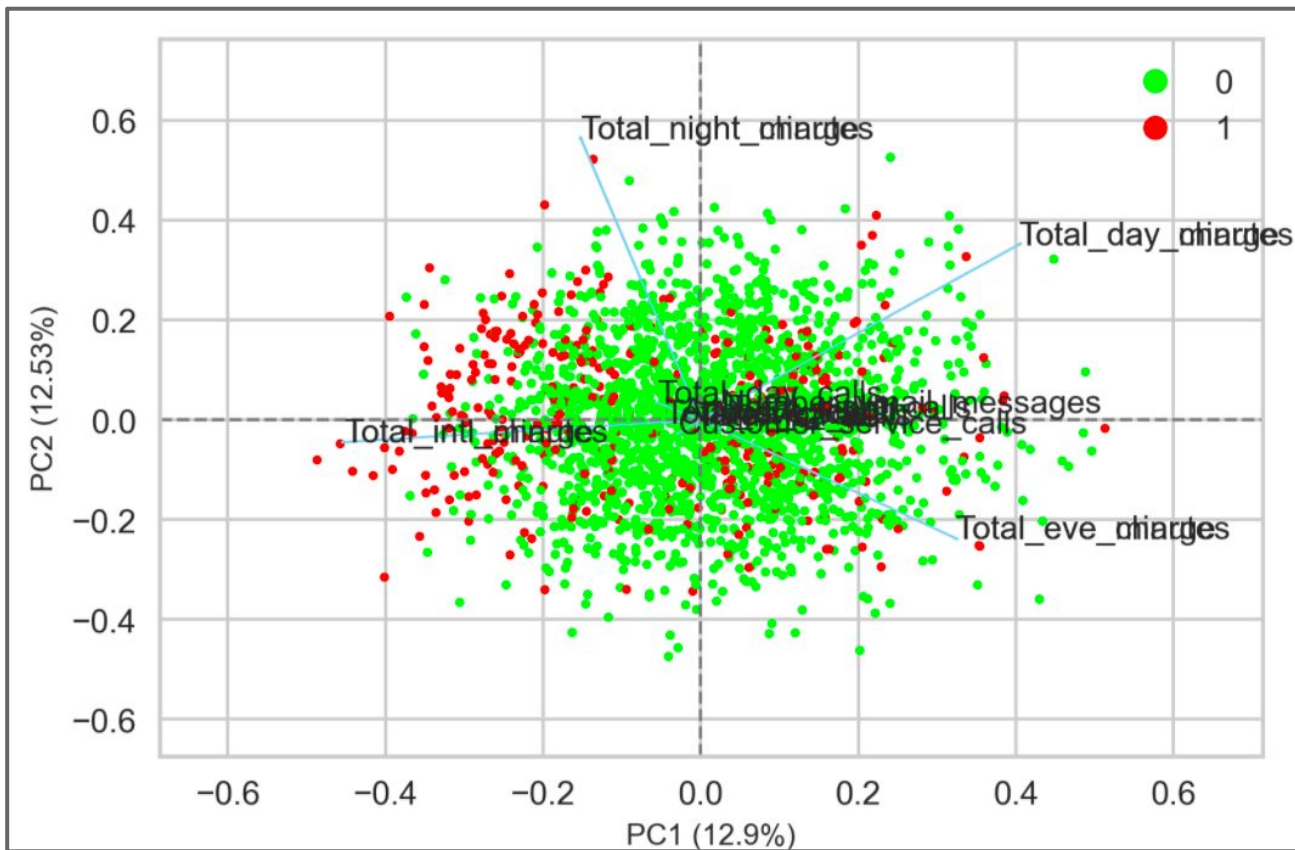
Principal Component Analysis
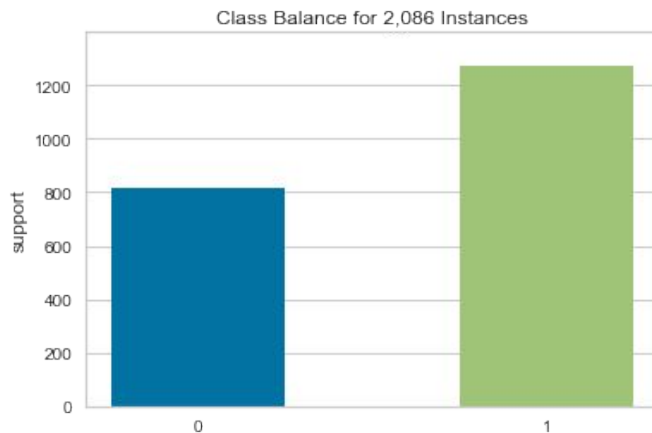
Dealing with Imbalanced Classes

Initial Dataset Classes

- **Train set) 0: 1282 | 1: 210**
- **Test set) 0: 266 | 1: 54**

After Applying SMOTEENN

- **Train set) 0: 1272 | 1: 814**

Class Balance for 1,812 Instances

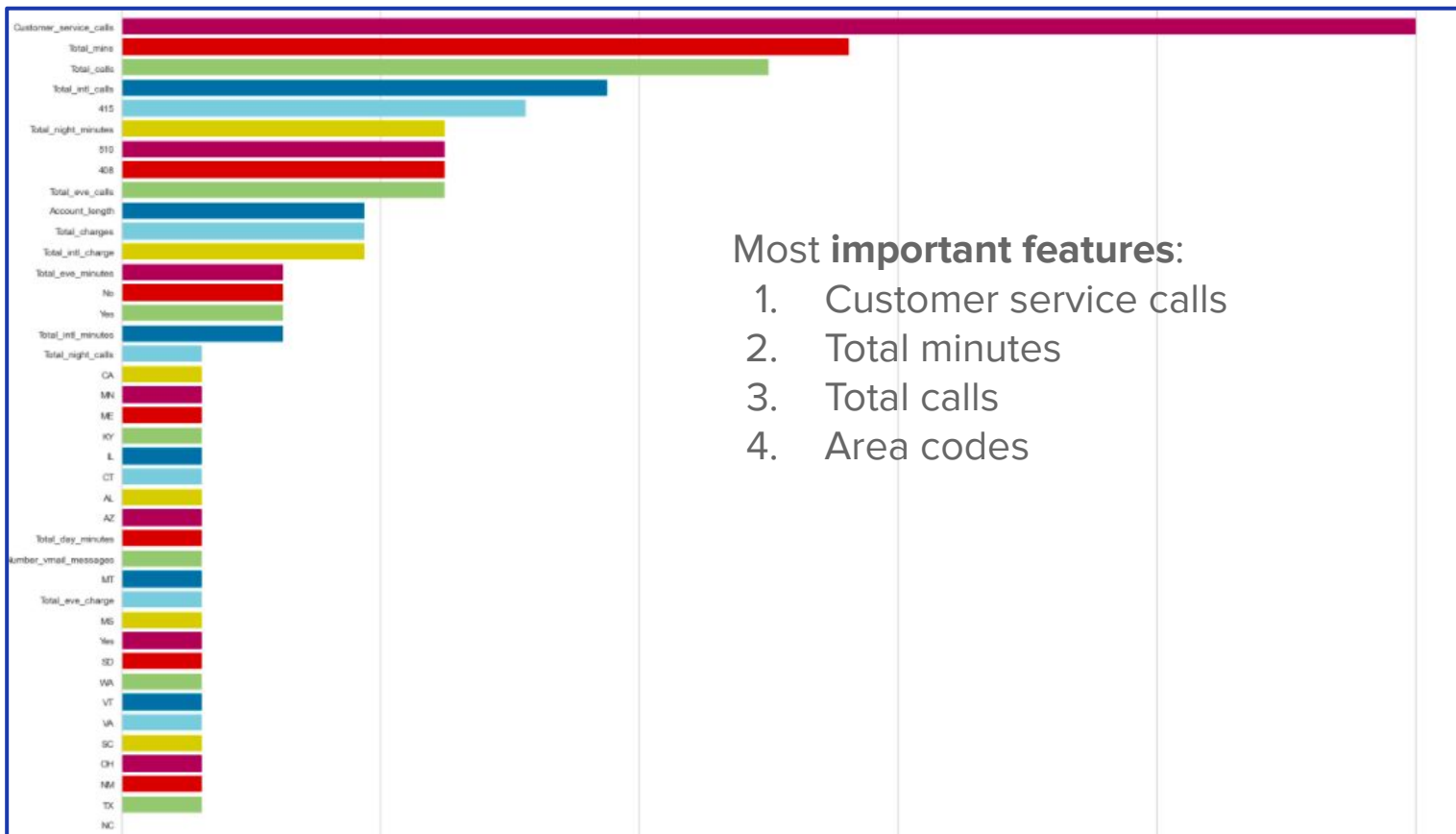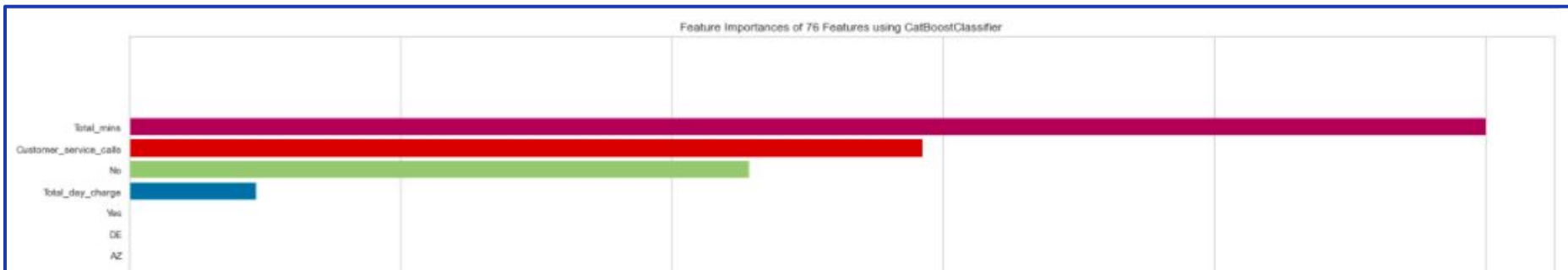Class Balance for 2,086 Instances

Modelling

1. Logistic Regression

2. Decision Tree Classifier

3. Random Forest Classifier

4. Gradient Boosting Classifier

5. Light Gradient Boosting Machine Classifier

6. Extreme Gradient Boosting Classifier

7. Support Vector Classifier

8. **AdaBoost**

9. TPOTClassifier

10. **CatBoostClassifier**

11. ANN - Sequential Model

12. Causal Inference
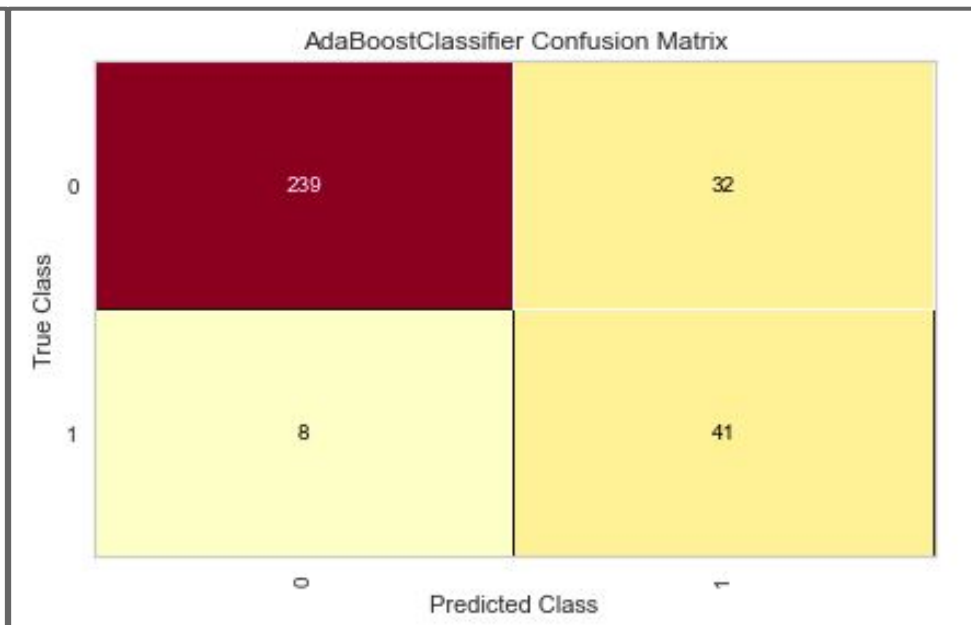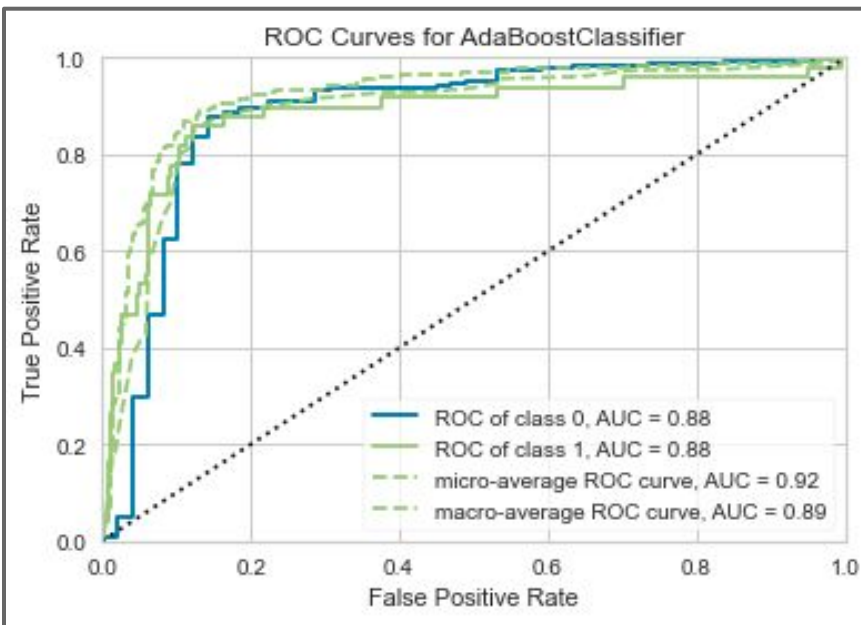
Results - Feature importance AdaBoost Classifier

Most **important features**:
1. Customer service calls
2. Total minutes
3. Total calls
4. Area codes

Results - Feature importance CatBoost Classifier

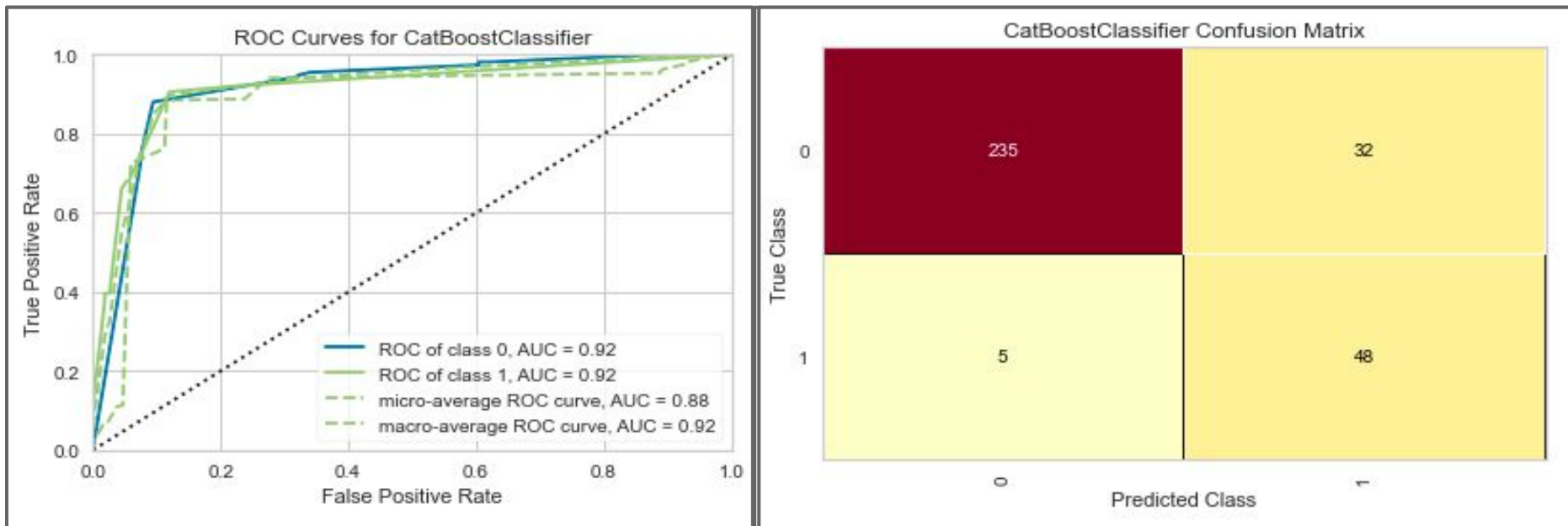Feature Importances of 76 Features using CatBoostClassifier

Most **important features**:
1. Total minutes
2. Customer service calls
3. Voicemail feature
4. Total day charges

Threats to Validity - AdaBoost Classifier

Threats to Validity - CatBoost Classifier

Causal Inference - Results

Treatment

Customer Service Calls

Most important feature
(AdaBoost)

Package

CausalML

Binary Approach | Churn Prediction

1. **No Service Calls*** → **13.92%**
2. **Any Service Calls** → **14.95%**

Bucketed Approach | Churn Prediction

1. **No Service Calls*** → **13.89%**
2. **1-3 Service Calls** → **10.24%**
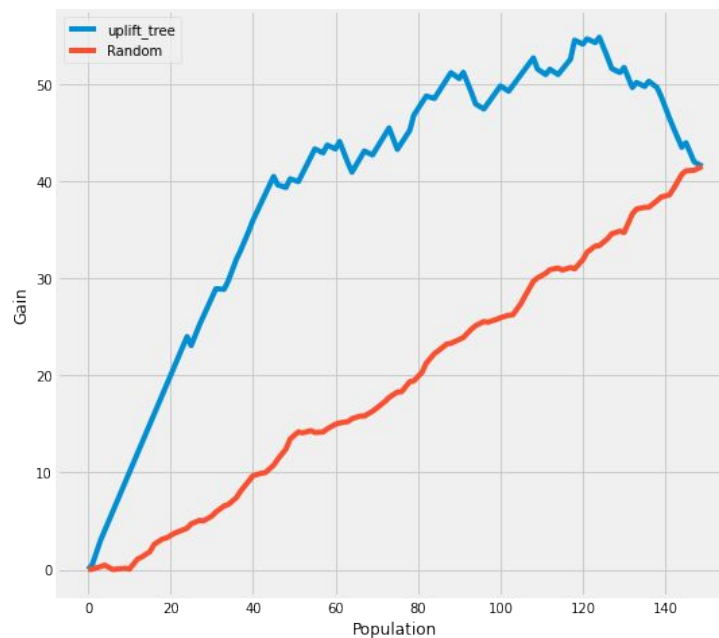3. **4-9 Service Calls** → **55.74%**

Causal Inference - Uplift Curves

No Service Calls vs **Any Service Calls**
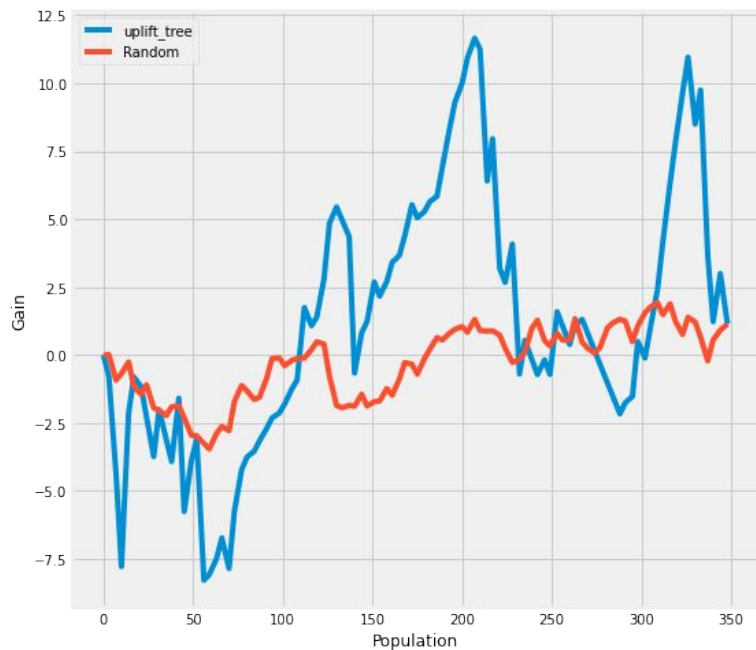
No Service Calls vs **4-9 Service Calls**

Recommendations

➔ Orange should invest in **customer service**:

    ◆ Orange should improve the quality of its customer service

    ◆ Ensure that customers do not have to call back more than 3 times. The rate of attrition increases significantly when customers have to call back several times.

➔ Customers who do not have **voicemail** are more likely to churn

    ◆ Orange should offer voicemail as a standard feature.

➔ Customers who have **a lot of minutes** are more likely to churn.

    ◆ Orange should offer a reduced rate per minute after customers pass a certain number of total calling minutes.

Questions?