SyriaTel
Customer Churn Analysis

Flatiron Online Data Science Course
Phase 3 Project

Oliver Zimmer, Aug 22nd 2022

Outline

- Business Problem
- Data
- Results
- Conclusions & Actionable Insights
- Further studies

Business Problem

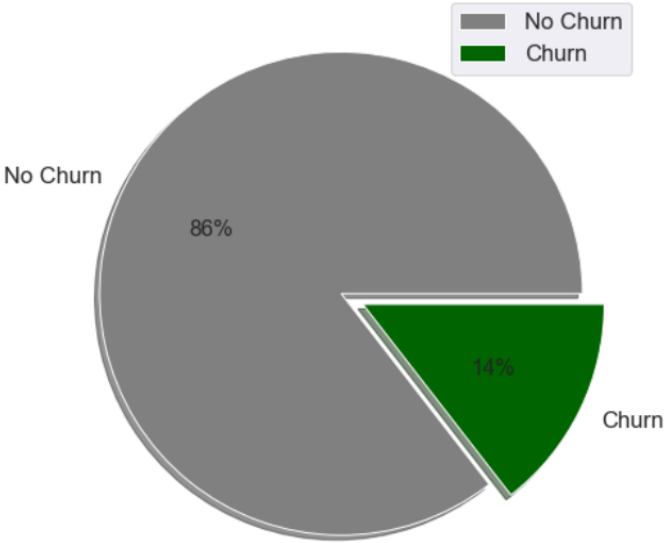
The **management of SyriaTel** is interested in predicting customer churn for its US telecommunications business.

Applying **machine learning theory and algorithms** enables identification of the most important business areas to focus on in order to:

1. reduce churn
2. further improve the relationships with loyal customers

Data

- The SyriaTel dataset consists of 3,333 entries with information on its US customer base
- Currently, the proportion of customer churn is approx. 14%



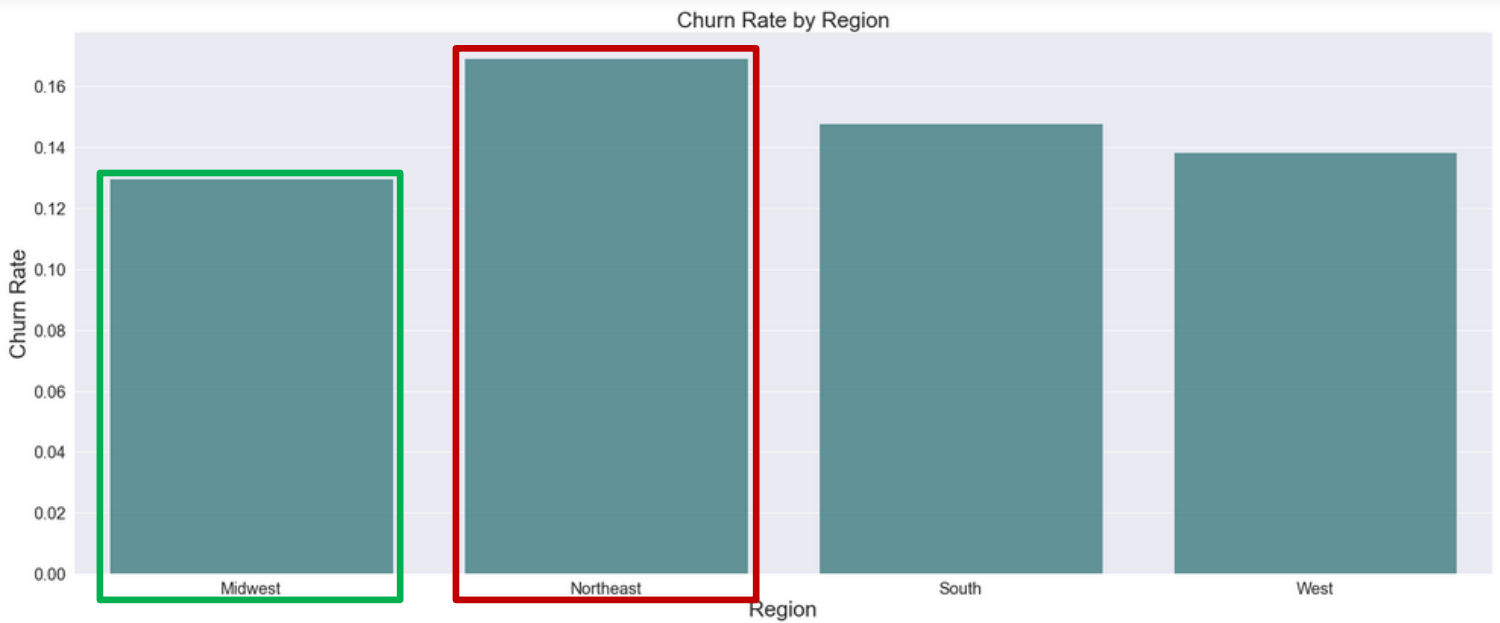
Results

The following key features were explored throughout this project:

1. Region
2. Total Charge
3. Customer Service Calls
4. International Plan

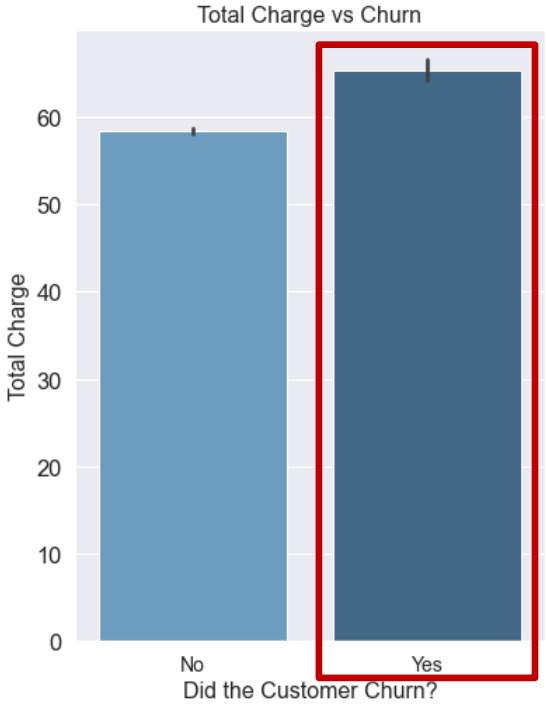
1. Region

- Is customer churn dependent on the location of the customer?
    - **Northeastern** region of the US: **highest** rate of customer churn
    - **Midwestern** region of the US : **lowest** rate of customer churn



2. Total Charge

- How does total charge of phone services affect customer churn?



- Level of Total charge has a **clear effect on the likelihood of customer churn**

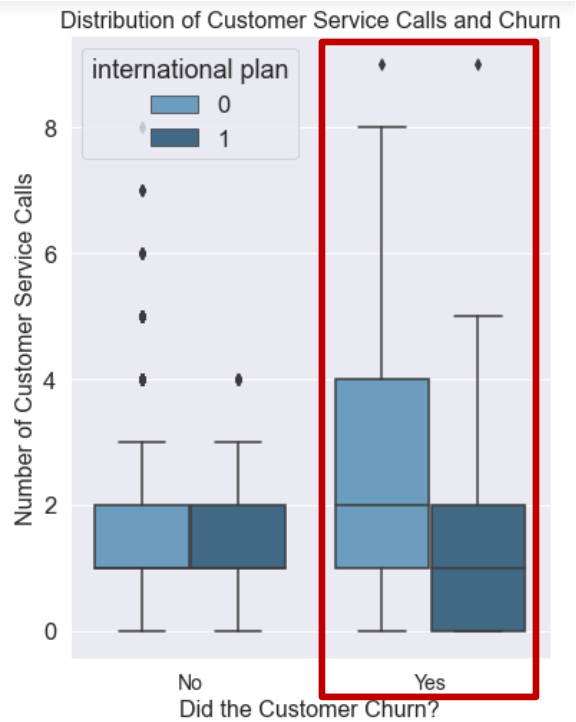
3. Customer Service Calls

- What is the effect of customer service calls on customer churn?

- The **more customer service calls** a customer makes, the **more likely it is for the customer to churn**
- Customer churn **increases considerably after 3 customer service calls**

4. International Plan

- Does the availability of an international plan affect customer churn?



- Customer churn is **less likely in case of the presence of an international plan**

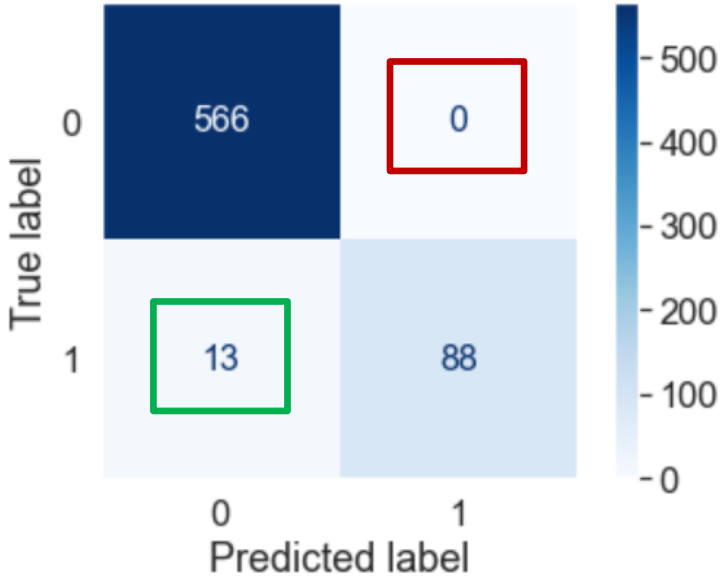
Predictive models

Model type	Description	Test Score
Random Forest	Bagging	98.1%
XGBoost	Base	98.1%
Gradient Boost	Base	97.9%
…	…	…
Logistic Regression	Base	77.9%

- Various different models have been built and analyzed for their predictive abilities
- **Best performing model** has been identified for the data available

Best performing model

Random Forest with Bagging



0 False Negative results

- In **zero cases** does this model **incorrectly predict that a churning customer will stay**
- This implies **direct savings of customer replacement costs** (i.e. foregone revenue, advertising, administrative, etc.)

13 False Positive results

- In **only 13 cases** does the model predict a customer will **churn when in reality the customer will stay**
- In this case actions might be taken to keep customers, thereby only **further improving customer satisfaction**

Conclusions & Actionable Insights

1. Focus on **lower pricing** or cost benefits such as discounts

2. Ensure **higher quality customer service**, reaching a benign solution for the customer after three customer service calls in order to avoid an increased potential for churn

3. Focus analytical efforts on the **international plan** as it has been shown that it has a strong influence on whether or not a customer is retained.
    - Possible aspects to investigate are **availability and pricing**

Further Studies

1. **Gridsearch (Gradient Boost or XGBoost)**
   - Find the best combination or parameters

2. **Expand research and analysis by more variables**
   - E.g. Number of Voicemail messages or in-depth analysis of call lenghts

3. **Model refinements**
   - Analyze feature importance

Thank You!

Email: kontakt@oliverzimmer.eu
GitHub: @senseize
LinkedIn: linkedin.com/in/username/oliver-zimmer-cfa-8824881ab/