Telecom Churn Prediction

IRONHACK Mid Bootcamp Project

Mojgan Shabani

About The Project

The Kaggle project

The Data Source for this project has been provided from an open-source data(cell2cell) by Teradata center for customer relationship management at Duke University.

https://www.kaggle.com/jpacse/datasets-for-churn-telecom

Overview

Customer **acquisition** and **retention** is a key concern for **telecommunications** industry.

Since the **cost of retaining** a good customer is **much lower than acquiring** a new one, it is very profit-effective to input valuable resource on the Retention Campaign.

Goal

The primary goal of **churn analysis** is usually to create a list of **customers** that are **likely** to be **cancelled** in the **near future**.

Input Data Review & Categorization

The dataset provides a sample size of around 50.000 data records of a Communication Service Provider in the US.

It includes around 60 features that can be categorized as follow:

Demographical Data

- Occupation
- Marital Status
- District
- Age
- Income Group
- Credit Ratings
- Service Area
- ….

Behavioral Data

- Revenue Generated
- Monthly Consumption
- Different Service Usage
- Number of Handsets
- Tenure
- Changes in Revenue
- Changes in Consumption
- …

Customer Experience

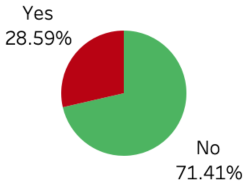
- Dropped Calls
- Blocked Calls
- Referrals
- Retention Offers
- Assisted Calls

Data Exploration

28,6%

Churn Rate

28,59% of the customer base was tagged as churned.



World Healthy Operating Range:
24%-30%

56$ vs 55$

Monthly Average Revenue

The average revenue of churn customers is registered as 55$, which is less than 2% below average. This shows that they are not among only low-value customers

506 vs 456 min

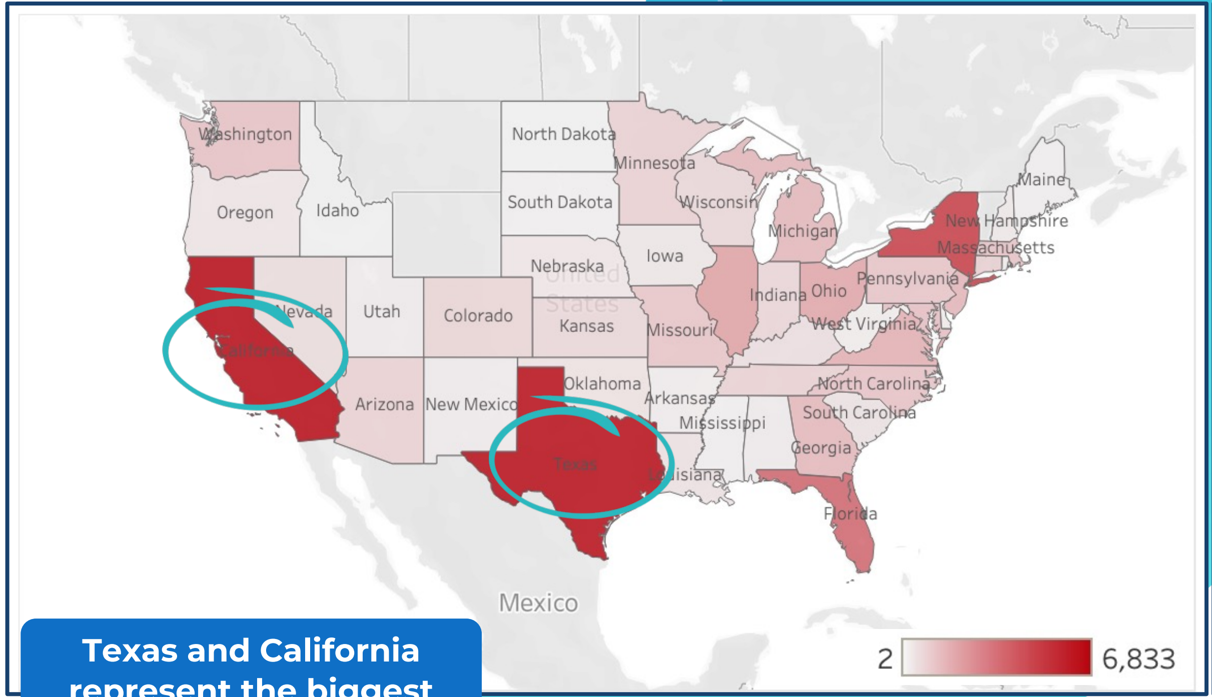
Average Monthly Minute of Usage

In comparison, the Churn Customer's average usage is 456$, which is significantly lower than the total average.
Comparing Average Revenue and Usage shows the churn Customers have paid more for each minute.

Data Exploration

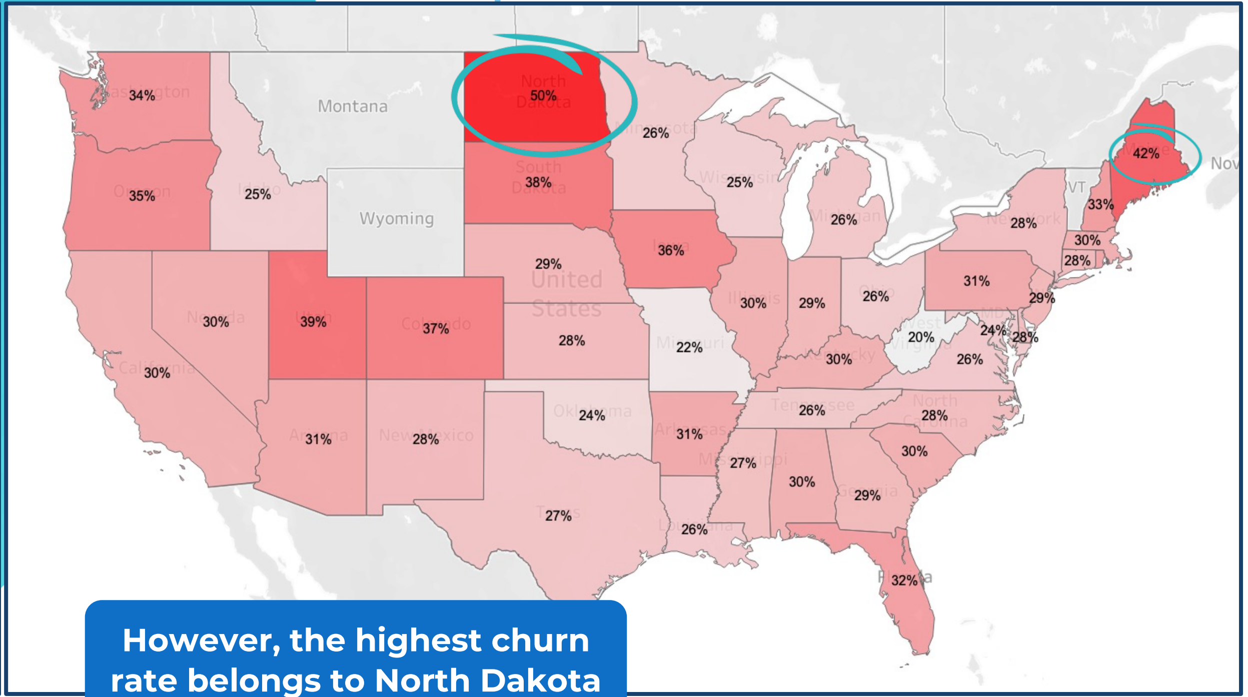
Geographical Exploration of input data:

Geographical Distribution of Customers



Texas and California represent the biggest share

Churn Rate per State



However, the highest churn rate belongs to North Dakota (50%) and Maine (42%)

01.

Customer Segmentation

In this study, the segmentation of the customer in order to identify high-value customers has been practiced as well. Around 4,63% of the customers have been considered as High Value (HVC):

Tenure

Revenue	Low	Moderate	High
High	2.496	**1.860**	**413**
Moderate	7.013	4.904	844
Low	16.719	12.879	1.948

HVC

Revenue	Low	Moderate	High
High	5,09%	**3,79%**	**0,84%**
Moderate	14,29%	9,99%	1,72%
Low	34,07%	26,24%	3,97%

Tenure

02.

Customer Segmentation

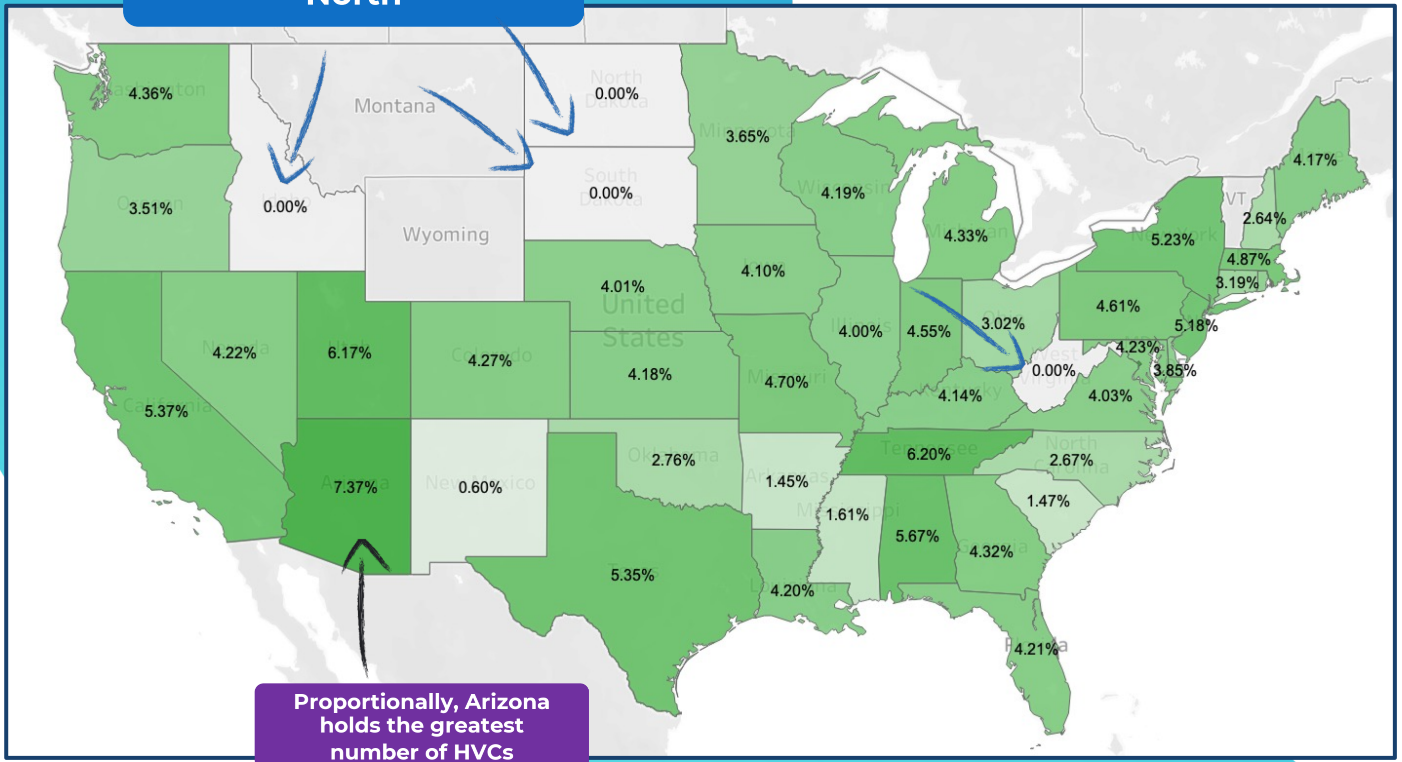
The map shows the percentage of HVC relative to the total customers in each state.

Revenue \ Tenure	Low	Moderate	High
High	782	473	83
Moderate	1.990	1.324	193
Low	4.687	3.904	595

Churn vs Segmentation

556 HVCs lost

Several state have no HVC. Especially in the North



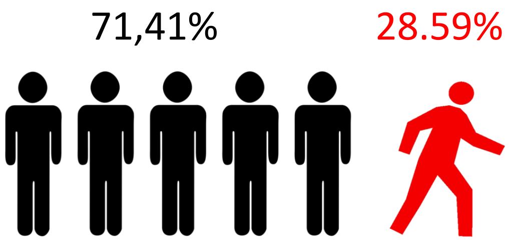
Proportionally, Arizona holds the greatest number of HVCs

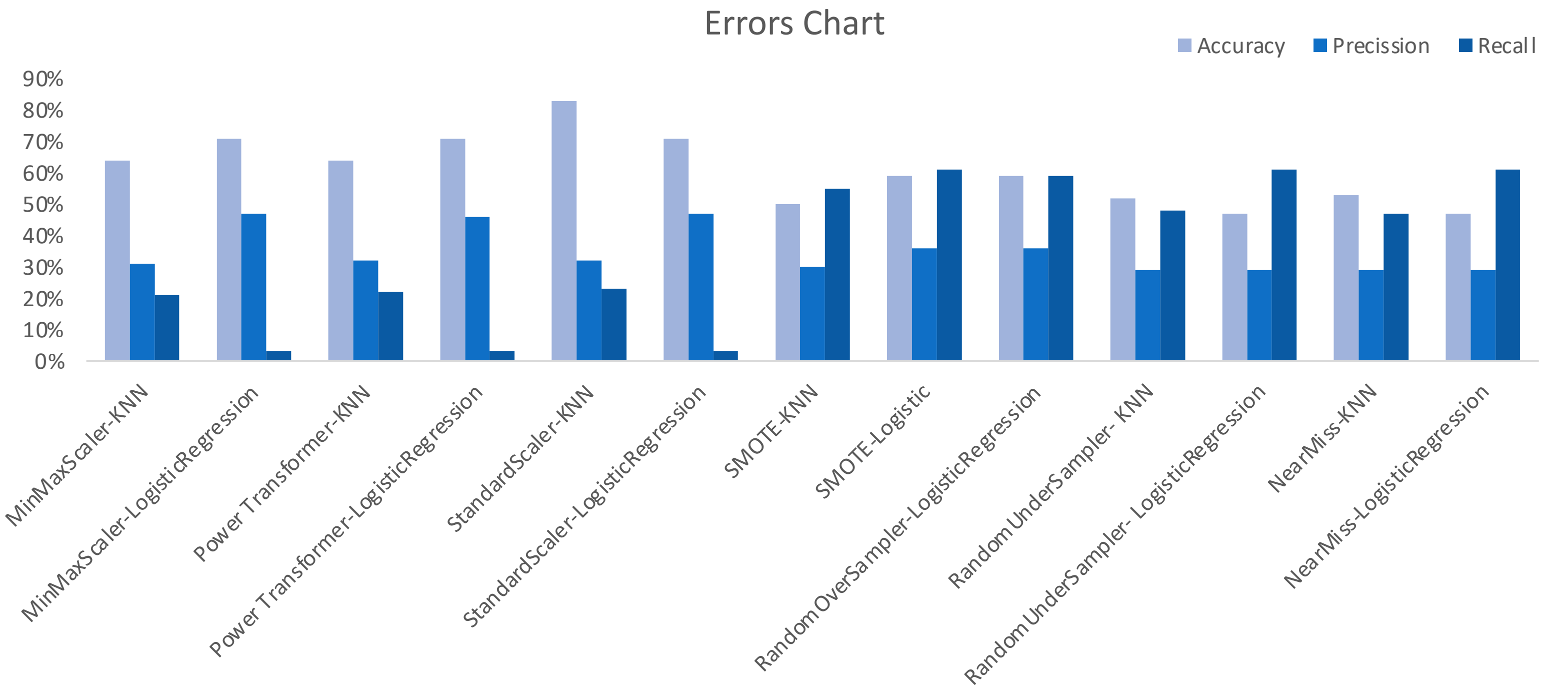
Churn Prediction

Different normalization models have been performed, and "Power Transform provided the best outcome."
Because of the imbalanced data, the result of KNN and logistic regression was not good enough.
So, for prediction, over and under sampling for Both KNN and logistic regression have been performed:

- Over Sampling:
  1. RandomOverSampler
  2. SMOTE
- Under Sampling:
  1. NearMiss
  2. RandomUnderSampler

Huge class imbalanced

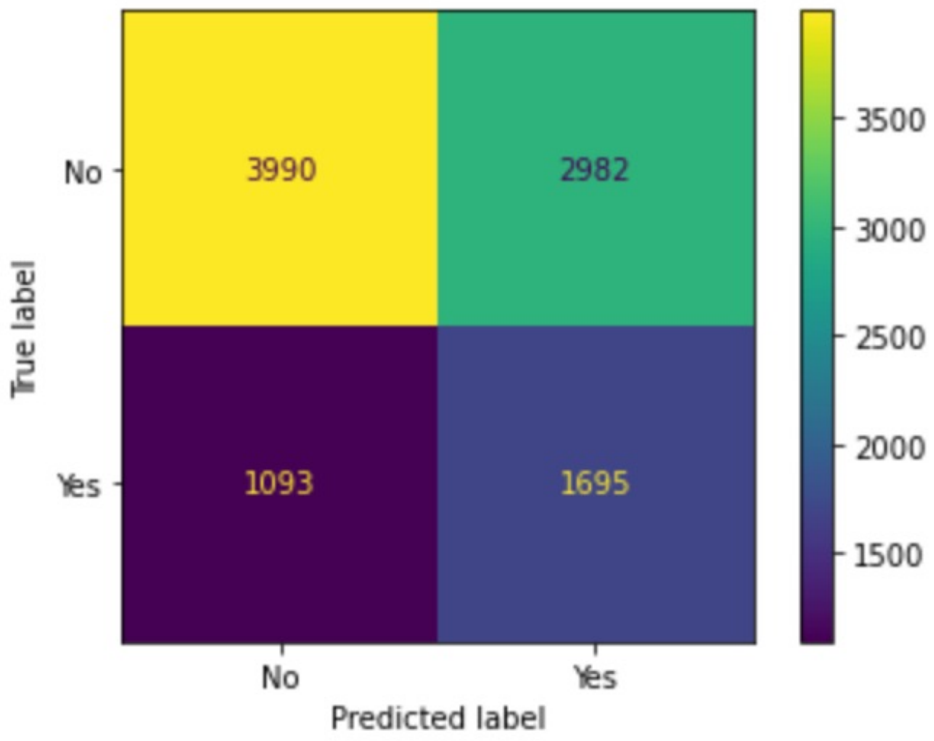




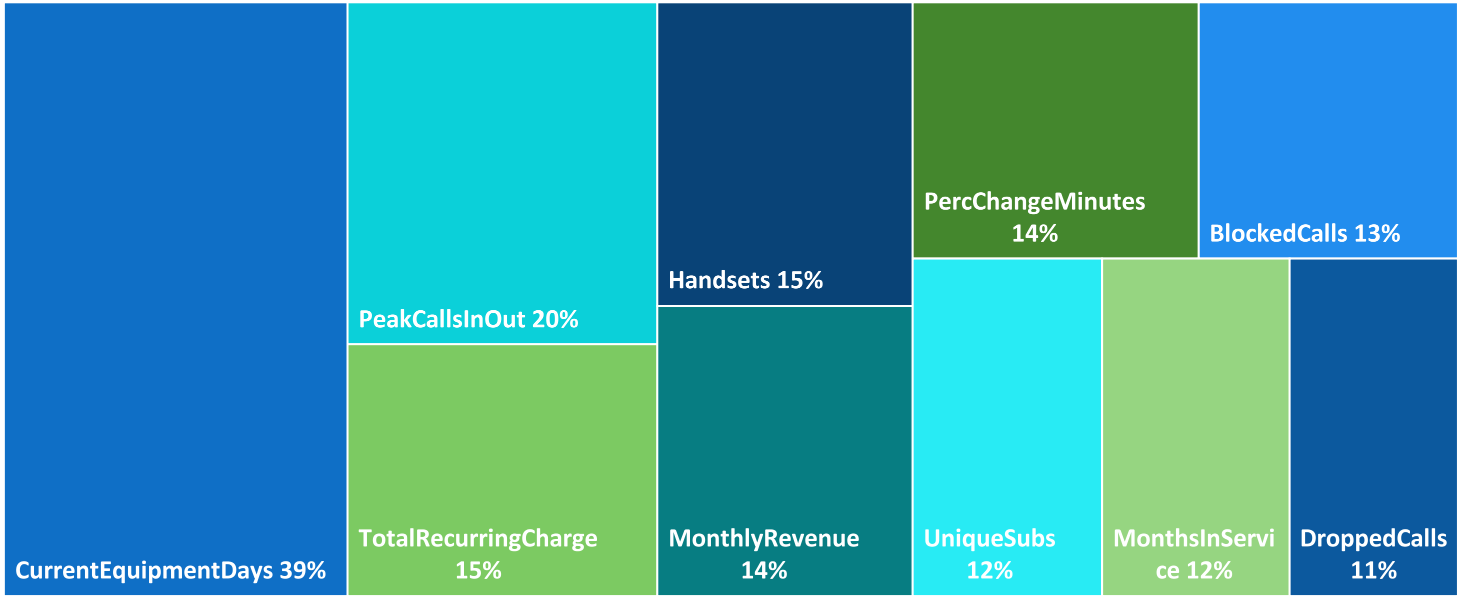
Churn Prediction

The best one was a logistic regression with Over sampling of SMOTE.

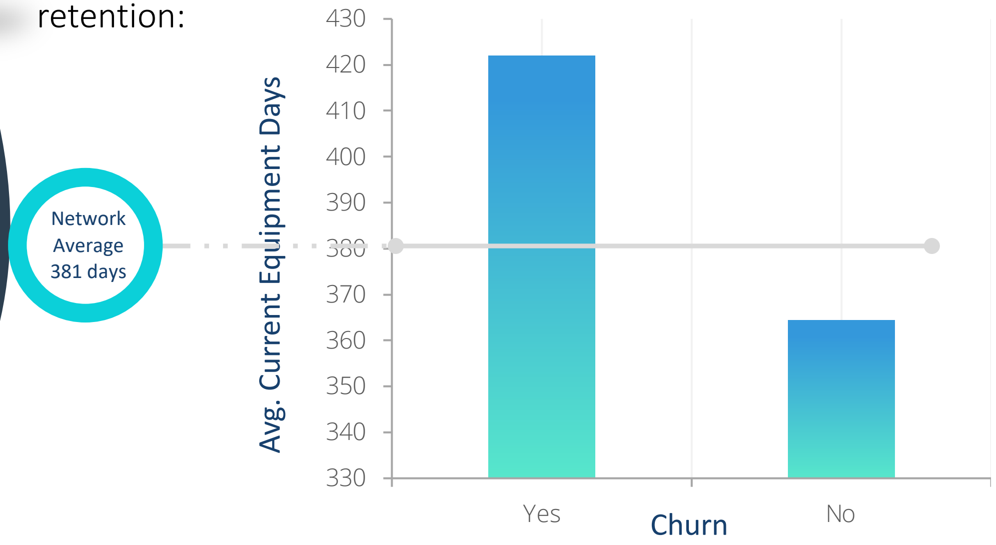
And based On this Model, Churn Can be predicted by 61% of success



Coefficient Calculation – What Features is the Most…



Based on coefficient of the model the Current Equipment is the most effective feature on Churn or retention:



As per these figures, it can be observed that the longer the customer holds on to the device there more likely it is to churn the customer.

As it is very typical for the Communication Service Providers to offer Devices through installment and permanency, the hypothesis to justify this correlation could be upon termination of the installment they decide to leave, however there not enough data to test it.

What Features are the most important

89,7 vs 81,2

Avg. Peak Calls In Out

The churn users demonstrate significant reduction in usage time prior to leave.

47,2$ vs 44,1$

Avg. Recurring Charges

The commitment and recurring income from potential churn users tends to drop below network average of 46.30 min.

1,8 vs 1,7

Avg. Handsets

¡¡ This feature is not clear as the meta data is missing !!

56,7$ vs 55,2$

Avg. Monthly Revenue

Same as recurring charges however as notable the monthly revenue shows reduction.

-5,4% vs -24,4%

Avg. % Changes Minutes

The drop in the minute of use is more prominent in user with tendency to churn while in the coefficient it is ranked 6th.

4,0% vs 3,9%

Avg. Blocked Calls

The drop in experience doesn't seems to be the driver in churn as both type of customers blocked calls are in par.

Suggestion:

- After one year of the customer contract, I would suggest a new device to the customer with a high tendency to churn(based on prediction)
- I would monitor customer usage every month, and if we have a high reduction in use, I would suggest a retention offer to them

Thank You!

Mojgan Shabani