Too many customers leaving

What id churn?

Churn is the number of customers have left or unsubscribed to the service.

How Will We Predict Customer Churn?

The basic layer for predicting future customer churn is data from the past.

We look at data from customers that already have churned (response) and their characteristics / behaviour (predictors) before the churn happened. By fitting a statistical model that relates the predictors to the response, we will try to predict the response for existing customers.

A screenshot of a cell phone

Description automatically generated

Use Case / Business Case  
Step one is actually understanding the business or use case with the desired outcome. Only by understanding the final objective we can build a model that is actually of use. In our case the objective is reducing customer churn by identifying potential churn candidates beforehand, and take proactive actions to make them stay.

Data collection & cleaning  
With understanding the context it is possible to identify the right data sources, cleansing the data sets and preparing for feature selection or engineering. It sounds quite simple, but this is likely the hardest part. The predicting model is only as good as the data source. And especially Startups or small companies have often trouble to find enough data to train the model adequately.

Feature selection & engineering  
With the third step we decide which features we want to include in our model and prepare the cleansed data to be used for the machine learning algorithm to predict customer churn.

Modelling  
With the prepared data we are ready to feed our model. But to make good predictions, we firstly need to find the right model (selection) and secondly need to evaluate that the algorithm actually works. While this usually takes a few iterations, we will keep this quite simple and stop as soon as the results fit our needs.

Insights and Actions  
Last but not least we have to evaluate and interpret the outcomes. What does it mean and what actions can we derive from the results? Because predicting customer churn is only half of the part and many people forget that by just predicting, they can still leave. In our case we actually want to make them stop leaving.

**The Dataset**

One of the most valuable assets a company has is data. As data is rarely shared publicly, we take an available dataset you can find on [IBMs](https://www.ibm.com/communities/analytics/watson-analytics-blog/guide-to-sample-datasets/) website as well as on other pages like [Kaggle](http://www.kaggle.com/): Telcom Customer Churn Dataset. The raw dataset contains more than 7000 entries. All entries have several features and of course a column stating if the customer has churned or not.  
To better understand the data we will first load it into pandas and explore it with the help of some very basic commands.

import numpy as np