# Predicting customer churn with supervised machine learning

Leo Evancie

January 2022

### Business Problem

It's well established that customer acquisition is typically more costly than customer retention. This is of particular importance to companies who bill customers on a periodic basis. In order to keep revenue streams healthy, these companies need to do everything they can to keep their customers. But are all customers equally likely to "churn" (i.e., stop being customers)? Or are there clues in their profile and behavior that may help us predict, and hopefully prevent, their dropoff?

### Dataset

For this project, we are using an IBM sample dataset (via Kaggle) consisting of over 7,000 telecom customers. We have some basic demographic information (e.g., gender, senior citizen status), their tenure as a customer, which of a wide range of services they've opted into (e.g., multiple phone lines, paperless billing), their monthly fee and payment method, their total charges, and whether or not they churned.

### Anticipated Data Science Approach

Since we are trying to predict churn, which entails a discrete set of possible outcomes (in this case "churned" or "not churned"), we will explore classification rather than regression models. We will try a baseline model and a selection of more sophisticated models to find the best candidate for predicting our target variable. In this context, a false positive would mean that we predict that a certain customer will churn, when in fact they did not. A false negative would mean we predicted that a customer will not churn, when in fact they did. For our business purpose, a false negative would be the more concerning outcome. It's far worse to be blindsided by a churned customer than to provide extra retention efforts to a customer who wasn't going to churn in the first place. As such, we will use *recall* as our key metric to evaluate and compare model candidates. The fewer false negatives we have, the higher the recall:

Recall = True Positive / (True Positives + False Negatives)

### Deliverables

- Original, cleaned, and preprocessed datasets (.csv)
- Jupyter notebooks:
    - Data wrangling
    - Preprocessing
    - Modeling
- Exploratory Data Analysis via Tableau
- Text file containing model summary and metrics (.txt)
- Final written report (.pdf)
- Presentation slide deck (.pptx)