# Welcome to the office!

# Seminar #1

***Names:*** *Benoît Fontannaz, Julien Tedeschi,*

## Abstract

For the first session, we will learn how to manage data form a telecom company. The work consists to learn more about the loyal client in order to attract some new one. In other words, we will focus on the life time value of loyal consumers.

To achieve that, we have to learn how to deal with the datasets in python. The first part of a project is always to familiarize yourself with data, it is call EDA (Exploratory Data Analysis). It will also summarize main data and shows us some potential problem we could have. EDA will be helpful to identify the better technique to for each project and it is a basic start of data analysis and gives us some important information. We can compare this step as a kick-off meeting in the business world.

The part after the EDA, we be to clean the data and organize then with “Python”. The aim will be for us to could learn some main properties to the actual loyal client to give some type for Telecom to find new loyal clients on long term.

## Introduction

Service in sell must often use data analysis on clients in order to keep them or attract new one. Their business is based on client how stay in the firm as the cost to retain a customer is less than attracting a new one. In order to keep most clients, companies use service to keep or recaptured leaving clients as this is less expensive than finding a new client. New recruits could be interesting for the company if there are worthwhile, in order word if their expected lifelong value is high enough. Basic information, about actual loyal client, will help us to determine which type of new recruits the firm as to look for.

Our goal is to figure the best way to generalize the data, and have a good visual representation. Great machine learning program can discriminate information between each other to fine some different people with same habit or other thing in common. During our work, we have to keep in mind the general idea for a machine learning algorithm:

* Intuitive
* Problem of optimization
* Geometrical representation in 2D
* Pros and cons of the method

During the rapport, we added images as it is easier to understand an idea with images instated of words.

## Data

Pas expliqué le loading du data set

At the begin, the code cleans data by keeping only active loyal clients. As the information about those who leave are not important for this analyse. Basic data, about each one, will be store in a table as the follow one (not all columns of the table are in this picture, to simplify the lecture of the rapport):

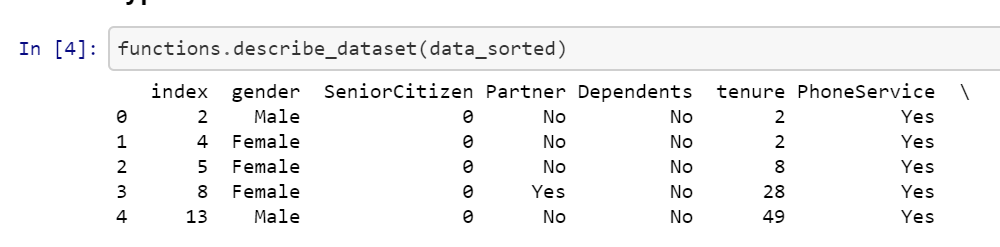


Figure 1: Table of basic information

Each line represents a unique client and in columns, there are all information about them. The datatype of each rows is important as the code will works with in order to found similarities between clients. If some errors are made at the begins the code will provide with wrong conclusion.

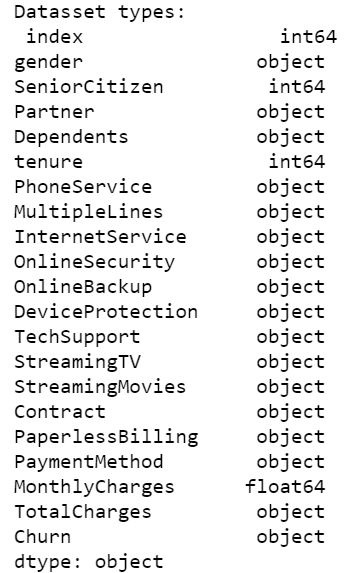
As we can see on each row of the table have a dataset type. There are two type of them:

Figure 2: Dataset type

1. Object
2. Number

One is a quantitative representation (number) and the other one a qualitative (object). This matter when the code will compare some data together.