



## Introducción

Uno de los grandes desafíos de las fintech a nivel global tiene que ver con facilitar el acceso a productos financieros.

En el equipo de Data trabajamos con soluciones de *Machine Learning* para tomar decisiones automatizadas al momento de brindar un crédito: En menos de dos minutos, un cliente debe tener depositado el dinero en su cuenta bancaria desde que inicia el proceso de solicitud.

## Sobre el test

La idea es que puedas explorar un conjunto de datos utilizando **Python**, y cuando termines de realizar los ejercicios, subas el código a un repositorio privado en **Github**. Vamos a necesitar que nos envíes un usuario tuyo para darte acceso. **Tenés 48hs para realizarlo**, contando desde el momento que te enviamos los ejercicios.

## Dataset

- Los datos corresponden a créditos otorgados por una compañía de lending.
- Cada fila en el dataset es un crédito.
- El objetivo es **predecir clientes que van a incumplir con el compromiso de pago**.
- La falta de pago se identifica mediante el campo **loan\_status**:
  - Fully Paid: Pagó el crédito.
  - Charged Off: No cumplió con el pago del crédito.

## Ejercicios

### 1. Análisis exploratorio de datos

Explorar el conjunto de datos para descubrir relaciones entre las variables y la falta de pago.

### 2. Desarrollo de un modelo predictivo

Entrenar un modelo predictivo utilizando el conjunto de datos. Identificar las variables más importantes y calcular una métrica de performance para el modelo.

### 3. Optimización de hiper parámetros

Si es necesario, hacer optimización de hiper parámetros con alguna técnica.



#### 4. Procesamiento de lenguaje natural

Hacer algún procesamiento del texto de las variables `emp_title` y `title` para poder analizarlas y potencialmente agregarlas a un modelo predictivo.

Es esperable que la solución que entregues no tenga el nivel de profundidad que se vería en un entorno productivo, nos interesa conocer tus conocimientos sobre los diferentes puntos.