**Primera entrega Trabajo Final - Data Science**

**Presentación de problema**

El problema que se presenta en esta entidad financiera es la ocurrencia de transacciones fraudulentas que ocasionan grandes pérdidas de dinero.

En este dataset, el comportamiento fraudulento de los agentes tiene el fin de obtener ganancias tomando el control de las cuentas de los clientes e intentar vaciar los fondos transfiriéndolos a otra cuenta y luego cobrando del sistema. Un ejemplo de un intento ilegal en este dataset es un intento de transferir más de 200.000 USD en una sola transacción.

**Objetivo**

* Predecir el mayor número de transacciones fraudulentas en base a patrones dados.
* Identificar los patrones con más ocurrencia en los fraudes.
* Disminuir el número de fraudes, para así evitar pérdidas de dinero.

**Conformación del equipo de trabajo**

* Yanina Acosta
* Darwin Moreno
* Santiago Wencelblat
* Daniela Maidana
* Agustín Iglesias

**Indicación de la fuente del dataset**

El dataset se encuentra en el siguiente [link](https://www.kaggle.com/datasets/vardhansiramdasu/fraudulent-transactions-prediction).

A continuación, se detalla la explicación de cada columna:

* step: mapea una unidad de tiempo. En este caso, 1 step es 1 hora.
* type: son los tipos de transacciones de la operación específica (CASH-IN, CASH-OUT, DEBIT, PAYMENT, TRANSFER)
* amount: monto total de la transacción (USD)
* nameOrig: cliente que inició la transacción.
* oldbalanceOrg: saldo inicial antes de la transacción.
* newbalanceOrig: nuevo saldo después de la transacción.
* nameDest: cliente destinatario de la transacción.
* oldbalanceDest: destinatario del saldo inicial antes de la transacción.
* newbalanceDest: destinatario del saldo inicial después de la transacción.
* isFraud: transacciones realizadas por agentes fraudulentos.
* isFlaggedFraud: es la señalización del fraude. La entidad financiera tiene como objetivo controlar las transferencias masivas de una cuenta a otra y señala los intentos ilegales. Un intento ilegal en este conjunto de datos es un intento de transferir más de 200.000 en una sola transacción.
* Indicación de la fuente del dataset y los criterios de selección (Data Acquisition).
* Generación del primer Data Wrangling y EDA, apuntado a sus datos (insights) univariado, bivariado y multivariado.
* Análisis de componentes principales.
* Contar la historia de sus datos
* Filtros aplicados a los datos. Distribución. Dataset final para analizar.
* Palttear objetivos u objetivo para esos datos.