CEIA Análisis de Datos

Trabajo Práctico Integrador

**Integrantes:**

Oksana Bokhonok

Ezequiel Guinsburg

## Informe de avance 1

**Fecha:** 31/03/2022

Se trabajó principalmente en lo siguiente:

* Librería de funciones de transformación:
  + Drop\_nans: Elimina las columnas que superan un porcentaje determinado de datos faltantes.
  + Date2columns: Genera una columna con el año y la otra con el mes y la proporción del día, eliminando la columna “Date” original.
  + Geoloc: Convierte el nombre de la ciudad en coordenadas de Latitud y Longitud. Elimina la columna “Location” original.
  + Normal: Normaliza los datos de las columnas respecto a su media y su desvío estándar.
* Código de visualización para el análisis de la información: Se implementa en una notebook lo siguiente:
  + Vista de las primeras filas del dataset (head)
  + Descripción estadística de los valores numéricos (describe)
  + Información de las columnas del dataset
  + Vista de la cantidad y proporción de datos faltantes por columna
  + Vista de las distribuciones de cada feature.
  + Vista de matriz de correlación.
  + Box-plot de los features.
  + Se observan los datos de los años incompletos para futuro análisis.
* Tabla descriptiva, columna a columna, de las transformaciones que se definen como necesarias en el pre-procesamiento.