

# Elements of Machine Learning 2024

Primer Proyecto

23.febrero.2024

El proyecto consiste en elegir alguna de las dos opciones de datos abiertos:

1) Los datos del sistema de transporte ECOBICI

- <https://www.ecobici.cdmx.gob.mx/>
- <https://www.ecobici.cdmx.gob.mx/es/informacion-del-servicio/open-data>.

2) Datos de stocks históricos entre compañías SP500

- <https://www.kaggle.com/datasets/suddharshan/historical-stock-price-of-10-popular-companies>
- <https://www.kaggle.com/code/thomasrahman/s-p-500-5-stock-data-kmeans-clustering>.

La idea es utilizar las herramienta vistas en el curso (y las que aún no hemos visto que ustedes deseen incorporar), para obtener información valiosa a partir de los datos.

## 1 Análisis

El proyecto consiste en elaborar un análisis de datos. Para ello, deberá

- Redactar un informe detallado de su análisis. Tomar en cuenta lo siguiente:
  - breve descripción de los datos
  - exploración
  - selección de variables importantes
  - imputación o remoción de datos faltantes
  - técnicas de visualización
  - análisis (indicar con detalle la metodología)
  - resultados
  - conclusiones
  - referencias
- Entregar copias del código utilizado (en un archivo comprimido).
- Preparar una presentación de sus hipótesis más importantes, así como de resultados y conclusiones. El tiempo de esta presentación sería de 30 minutos, y 10 minutos adicionales para preguntas.

Las presentaciones se realizarán el día viernes 5 de abril (el orden será asignado de forma aleatoria).

## 2 Observaciones

- Pueden colaborar y compartir ideas en equipos, pero cada uno por separado debe hacer su análisis, reporte y presentación.
- Pueden utilizar algoritmos o técnicas no vistas aún en el curso, desde que se indiquen las referencias.
- Pueden usar código o partes de código ajeno (disponible en la web), desde que se indique la referencia correspondiente.

## 3 Fechas de entrega

- Presentaciones: viernes 5 de abril.
  - Entrega de informe, código y presentación: viernes 12 de abril.
-