



## TP-1

Laboratorio de datos 2023 (comisión: G. Solovey)

### Objetivo

El objetivo de este TP es que apliquen todo lo que aprendieron hasta ahora de descripción, visualización, exploración y manipulación de datos al análisis de dos datasets. Los dos son datasets públicos, uno contiene el registro de usos de bicicletas públicas de CABA y el otro datos climáticos diarios.

### Datasets

#### Dataset de uso de EcoBici

---

El dataset de uso del sistema Ecobici de la Ciudad de Buenos Aires está disponible [acá](#). En particular, vamos a trabajar con datos del año 2022. El dataset que está en la página tiene casi 3 millones de registros, pero para este TP preparamos un dataset reducido de 10000 observaciones. Ese dataset reducido se encuentra [acá](#).

#### Importación y preprocesado de datos en R

---

En primer lugar seleccionar solo los viajes de entre 5 minutos y 1 hora de duración. Luego reconozcan qué significa cada variable. En la página donde está el dataset original encontrarán parte de esta información. Verifiquen que las variables tienen sentido, que tengan datos para todos los días del 2023, piensen en qué unidades se miden las variables numéricas, etc.

#### Dataset de clima

---

Importen datos meteorológicos y climáticos de MeteoStat de [esta página](#). En particular, consideren datos de la estación meteorológica de Aeroparque (ID 87582), desde el 1ro de enero de 2022 hasta el 31 de diciembre de 2022.

#### Importación y preprocesado de datos en R

---

Descarguen ese dataset en formato csv e importen los datos en R. Verifiquen que tienen datos de todos los días del 2023. ¿Qué representa cada variable?

### Análisis descriptivo

En esta etapa tienen que trabajar con cada dataset por separado. Tienen que hacer un análisis descriptivo de las variables para familiarizarse con los datos y reconocer sus características y variabilidad.

En cuanto al dataset de uso de bicicletas, ¿cómo se codifica el género de los usuarios? ¿lo usa más algún grupo? ¿cuáles son los viajes mas comunes (de dónde salen y dónde llegan?). Describan la variable que indica la duración de viaje (globalmente y por separado para los 5 viajes más frecuentes). En cuanto al dataset de datos climáticos, ¿qué variables tienen? ¿Cómo varía la temperatura a lo largo del año? ¿Tiene sentido? ¿Cómo es la cantidad de lluvias según el mes?

Estos son sólo algunas ideas, ustedes pueden tener otras. Para cada caso piensen cómo describirían lo que quieren mostrar (si una tabla, un gráfico, qué grafico, etc).

## Análisis Exploratorio

Supongan que quieren entender patrones de uso del sistema de Ecobici. Su variable de interés es el número de usos del sistema por día y quieren entender qué factores podrían explicar ese número. Algunas preguntas disparadoras para el análisis exploratorio son:

- ¿Cómo varía el uso de la EcoBici a lo largo del año?
- ¿El uso es diferente en días de semana vs. fin de semana? (para crear esta variable, investiguen el uso de la función `weekdays()`).
- ¿Hay algún día atípico en el uso de la EcoBici? ¿le pueden encontrar una explicación?
- ¿Si hace mucho calor se usa más o se usa menos (o depende)?
- ¿Y si llueve?

Piensen más preguntas. Es libre, pueden explorar por dónde se les ocurra. Realicen visualizaciones pertinentes, muestren datos de resumen cuando les parezca adecuado, etc.

## Reporte para entregar

- En cada sección (descripción, exploración, etc) escriban texto que acompañe el informe, las figuras y tablas que muestren.
- Deben entregar un archivo con el proyecto de R que contenga un archivo `qmd` y el `html` que se genera.