

## Preprocesamiento y Análisis Exploratorio de Datos

### Taller final

Este taller tiene como objetivo aplicar diversas técnicas de preprocesamiento a un conjunto de datos reales relacionado con incidentes en los cuales se han robado bicicletas. Cada registro incluye información detallada sobre el incidente, como la fecha y hora de ocurrencia, el tipo principal de delito, la ubicación geográfica, detalles sobre la bicicleta involucrada, el estado del incidente y otros atributos relevantes.

Este conjunto de datos podría utilizarse para realizar análisis exploratorios, identificar patrones temporales y espaciales, así como para desarrollar estrategias de prevención y seguridad en torno a los incidentes de robo de bicicletas en la ubicación especificada.

Diccionario de las variables:

1. **X, Y:** Coordenadas geográficas.
2. **OBJECTID:** Identificador único del registro.
3. **event\_unique\_id:** Identificador único del evento.
4. **Primary\_Offence:** Tipo principal de delito.
5. **Occurrence\_Date:** Fecha y hora en que ocurrió el delito.
6. **Occurrence\_Year, Occurrence\_Month, Occurrence\_DayOfWeek, Occurrence\_DayOfMonth, Occurrence\_DayOfYear, Occurrence\_Hour:** Componentes de la fecha y hora de ocurrencia del delito.
7. **Report\_Date:** Fecha y hora en que se informó el delito.
8. **Report\_Year, Report\_Month, Report\_DayOfWeek, Report\_DayOfMonth, Report\_DayOfYear, Report\_Hour:** Componentes de la fecha y hora en que se informó el delito.
9. **Division:** División policial.
10. **City:** Ciudad.
11. **Hood\_ID:** Identificador de vecindario.
12. **NeighbourhoodName:** Nombre del vecindario.
13. **LocaQon\_Type:** Tipo de ubicación.
14. **Premises\_Type:** Tipo de propiedad.
15. **Bike\_Make, Bike\_Model, Bike\_Type, Bike\_Speed, Bike\_Colour, Cost\_of\_Bike:** Detalles relacionados con la bicicleta.
16. **Status:** Estado del incidente.
17. **Longitude, Latitude:** Coordenadas geográficas del lugar.
18. **ObjectID2:** Otro identificador único.

## Objetivos del Taller final

Las tareas para realizar en este taller incluyen:

### 1. Carga de datos y exploración inicial:

- Importar el conjunto de datos
- Identificar y describir los tipos de variables presentes en el conjunto de datos.
- Realizar un análisis para detectar y eliminar datos duplicados si existen.

### 2. Preprocesamiento:

- Manejar los valores faltantes aplicando estrategias adecuadas de imputación, según el tipo de variable.
- Analizar posibles valores atípicos (outliers) en las variables clave y tratarlos si es necesario.
- Normalizar o estandarizar las variables numéricas para asegurar que estén en la misma escala.
- Determinar si la variable "status" está desbalanceada. Si es así, aplicar algún método de balanceo. Justificar

### 3. Análisis temporal:

- Convertir las variables de fecha y hora al formato adecuado para análisis.
- Identificar patrones temporales en los robos de bicicletas y crear visualizaciones que ayuden a detectar tendencias.

### 4. Análisis multivariado

- Explorar la relación entre las variables para identificar características relevantes que ayuden a detectar robos de bicicletas de manera más efectiva.

### 5. Responder las siguientes preguntas:

#### 1. Frecuencia y Patrones Temporales

- ¿Cómo varía la frecuencia de robos de bicicletas según los años, meses, días de la semana y horas del día?
- ¿Existen patrones estacionales en los robos de bicicletas?

#### 2. Ubicaciones más Afectadas

- ¿Cuáles son las zonas de la ciudad con mayor incidencia de robos de bicicletas?
- ¿Hay algún lugar en particular que sea más propenso a sufrir robos de bicicletas?

#### 3. Características de las Bicicletas Robadas

- ¿Qué marcas y modelos de bicicletas son los más robados?
- ¿Algunos colores de bicicletas son más propensos a ser robados?
- ¿Existe una relación entre el costo de la bicicleta y la gravedad del robo?

#### 4. Análisis del Tiempo entre Ocurrencia y Reporte del Robo

- ¿Cuál es el tiempo promedio que transcurre entre el robo de la bicicleta y su reporte?
- ¿Existen diferencias significativas en este tiempo según la hora del día, el día de la semana o la ubicación?
- ¿Cuánto tiempo pasa entre la fecha de ocurrencia del robo y la fecha de reporte? ¿Cómo afecta esto la recuperación de la bicicleta?

**Archivo de entrega:** Documento en formato “.pdf” con las respuestas a todas estas preguntas que incluya el enlace al archivo “.ipynb” correspondiente con el soporte del análisis realizado.

**Permitidos grupos de 3 integrantes.**

**Fecha máxima de entrega 03 de noviembre a las 23:59**

**Nota:** recuerden que cualquier consulta pueden contactarme por correo y/o coordinar reunión por zoom para aclarar cualquier inquietud del taller.