

## **M2.859 Visualización de datos**

### **M2.859 – A7 – PRA1 - Selección del conjunto de datos**

**Estudiante:** Erika Paola Martínez Soria

**Inicio:**28/11/22

**Entrega:**14/12/22

**Calificación:**24/12/22

**Dedicación:**20 %

#### **Descripción y enunciado**

Esta actividad, primera parte de la práctica final, consiste en la selección por parte del estudiante de un conjunto de datos de su interés que será usado en el proyecto de creación de la visualización de datos, de acuerdo con unos criterios establecidos. Básicamente, la temática es libre, pero se valorarán los aspectos siguientes:

#### **1. [10%] Justificad brevemente vuestra selección, ya sea por motivos personales o profesionales.**

El dataset escogido es Major Crime Indicators (MCI) de la ciudad de Toronto proveniente del portal de datos del servicio de policía de Toronto.

Este es el link donde se puede acceder a información sobre los datos y la descarga:

<https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-1/about>

Mi background es en geografía y estudios urbanos, así que estoy interesada en poder analizar y visualizar eventos como crímenes desde diferentes perspectivas como lugar, tiempo y tipo de crimen. Además, tengo familiaridad con la ciudad de Toronto y conozco que ciertos tipos de crimen han incrementado mientras otros han disminuido. Por lo que sería importante entender en más detalle cual ha sido la trayectoria y posibles causas de cada uno de estos.

#### **2. [10%] La relevancia del conjunto de datos en su contexto. ¿Son datos actuales? ¿Tratan un tema importante por algún colectivo concreto? ¿Se ha tenido en cuenta la perspectiva de género?**

El dataset Major Crime Indicators (MCI) está compuesto por ocurrencias de delitos desde 2014 hasta el 30 de junio de 2022. Por lo que podemos concluir, que el dataset está bastante actualizado. Este contiene las categorías de delitos como: asalto, allanamiento de morada, robo de auto, robo a propiedad y hurto a persona (excluye violaciones sexuales).

Existe interés por parte de investigadores, la academia y los departamentos de policía de entender la intersección entre la geografía y la criminología. Por lo tanto, el análisis y visualización de este dataset contribuirá a la mejor comprensión de esta relación, al encontrar patrones o tendencias en diferentes tipos de crímenes de acuerdo a su locación.

Lamentablemente, los datos de las víctimas no contienen el elemento de género. La ciudad de Toronto maneja con mucho cuidado los datos de género y raza, usualmente no haciéndolos públicos.

**3. [25%] La complejidad (medida, variables disponibles, tipos de datos, etc.). ¿Tiene del orden de centenares o miles de registros? ¿Tiene del orden de decenas de variables? ¿Combina datos categóricos y cuantitativos? ¿Incluye otros tipos de datos? Evitad los conjuntos excesivamente simples.**

El dataset tiene 25 campos (excluyendo el id e índice) con variables numéricas, categóricas, de tiempo y geográficas. El número de registros es alrededor de 300 000.

#	Name	Type
1	Index	Number
2	event_unique_id	Text
3	Division	Text
4	occurrencedate	Date or Time
5	reporteddate	Date or Time
6	location_type	Text
7	premises_type	Text
8	ucr_code	Number
9	ucr_ext	Number
10	offence	Text
11	reportedyear	Number
12	reportedmonth	Text
13	reportedday	Number
14	reporteddayofyear	Number
15	reporteddayofweek	Text
16	reportedhour	Number
17	occurrenceyear	Number
18	occurrencemonth	Text
19	occurrenceday	Number
20	occurrencedayofyear	Number
21	occurrencedayofweek	Text
22	occurrencehour	Number
23	mci_category	Text
24	Hood_ID	Text
25	Neighbourhood	Text
26	Longitude	Number
27	Latitude	Number

#### Major Crime Indicators (MCI) - Data Field Descriptions

Field	Field Name	Description
1	Index	Unique Identifier
2	event_unique_id	Offence Number
3	Division	Police Division where Offence Occurred
4	occurrence_date	Date of Offence
5	reporteddate	Date Offence was Reported
6	location_type	Location Type of Offence
7	premises_type	Premises Type of Offence
8	ucr_code	UCR Code for Offence
9	ucr_ext	UCR Extension for Offence
10	Offence	Title of Offence
11	reportedyear	Year Offence was Reported
12	reportedmonth	Month Offence was Reported
13	reportedday	Day of the Month Offence was Reported
14	reporteddayofyear	Day of the Year Offence was Reported
15	reporteddayofweek	Day of the Week Offence was Reported
16	reportedhour	Hour Offence was Reported
17	occurrenceyear	Year Offence Occurred
18	occurrencemonth	Month Offence Occurred
19	occurrenceday	Day of the Month Offence Occurred

20	occurrencedayofyear	Day of the Year Offence Occurred
21	occurrencedayofweek	Day of the Week Offence Occurred
22	occurrencehour	Hour Offence Occurred
23	MCI	MCI Category of Occurrence
24	Hood_ID	Identifier of Neighbourhood
25	Neighbourhood	Name of Neighbourhood
26	Long	Longitude Coordinates (Offset to nearest intersection)
27	Lat	Latitude Coordinates (Offset to nearest intersection)

**4. [25%] La originalidad. No repetid los conjuntos de datos clásicos. Podéis, por ejemplo, combinar o mejorar visualizaciones existentes. ¿hay otras visualizaciones basadas en este conjunto de datos? ¿Es una evolución o actualización de un conjunto anterior? ¿Habéis enriquecido un conjunto de datos ya existente?**

El departamento de policía de Toronto ofrece visualizaciones básicas basadas en este conjunto de datos. Me entusiasma poder explorar los datos de diferentes formas y con técnicas de visualización avanzadas aprendidas en esta asignatura como por ejemplo el word cloud o el gráfico de enjambre. Además, los datos pueden ser complementados con datos socioeconómicos de los barrios de la ciudad, al igual que con datos del entorno construido (zonificación) que incluirían áreas de comercio, vivienda, centros educativos, etc. y como su presencia se podría relacionar con los delitos.

**5. [30%] Las cuestiones que responderéis con la visualización de datos, ¿Tienen en cuenta los puntos anteriores? ¿Están bien planteadas? ¿Son adecuadas por el conjunto de datos elegido?**

Algunas gráficos podrían aclarar preguntas como:

- Gráfico circular con el porcentaje de cada tipo de delito en la ciudad de Toronto.
- Gráfico de barras con los porcentajes de delitos por cada barrio.
- Tendencia según años y meses de los diferentes delitos.

- Gráfico de enjambre con los barrios como eje X y los días de un cierto mes como eje Y.
- Word cloud de los barrios donde se comete mayor robo de vehículos.
- Gráfico de mariposas comparando el número de negocios con el número de delitos en cierto vecindario.
- Mapa de densidad de puntos con casos de hurto a personas.

Estas y más visualizaciones responden adecuadamente a las preguntas de como los delitos en la ciudad de Toronto se relacionan con su geografía, con el tiempo, con variables socioeconómicas y con el entorno construido. Así, se demuestra que se puede utilizar al máximo los diferentes atributos que el dataset contiene representándolos con gráficas de gran variedad.

Recordad que debéis librar un documento PDF en el registro de evaluación continuada que contenga todos los puntos mencionados anteriormente.

**NOTA:** Si se considera que esta parte no se ha resuelto de forma adecuada, el profesor os puede proponer otro conjunto de datos. Una C+ en esta actividad es un indicador de que podréis tener problemas después en la hora de crear la visualización de datos en la segunda parte de la práctica. Si obtienes una A o una B podéis considerar que vuestra propuesta de datos ha sido aceptada para utilizarla a la segunda parte de la práctica.

[guia-visualizacion \(1\).pdf](#)