

Visualización de datos

PRACTICA PARTE I: selección del conjunto de datos

Eduardo Mora González

1. INFORMACIÓN DEL CONJUNTO DE DATOS

El conjunto de datos seleccionado para la creación de la visualización es "*Crime Data from 2020 to Present*", el cual documenta incidentes de crimen en la Ciudad de Los Ángeles desde el año 2020 hasta el 24 de noviembre de 2023.

Este conjunto de datos fue obtenido de la página oficial del gobierno de los Estados Unidos, disponible en <https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

A continuación, se proporciona información relevante sobre el conjunto de datos:

- **Cobertura Temporal:** Este conjunto abarca incidentes de crimen en la Ciudad de Los Ángeles desde el año 2020 hasta el 24 de noviembre de 2023.
- **Origen de los Datos:** La información se transcribe a partir de informes originales de crímenes escritos en papel, lo que podría introducir algunas inexactitudes en la información.
- **Ubicación de Datos Faltantes:** Algunos campos de ubicación con datos faltantes se indican como (0°, 0°).
- **Privacidad:** Para preservar la privacidad de las personas, las direcciones se proporcionan solo hasta el bloque más cercano.
- **Precisión de los Datos:** La precisión de los datos se afirma ser tan precisa como la información en la base de datos. Se alienta a los usuarios a señalar preguntas o inquietudes en los comentarios.

Este conjunto de datos consta de 25 variables y abarca un total de 829.778 registros. A continuación, se detalla la descripción de las variables:

1. **DR_NO:** Número de informe para el incidente del crimen.
2. **Date Rptd:** Fecha en la que se reportó el crimen.
3. **DATE OCC:** Fecha en la que ocurrió el crimen.
4. **TIME OCC:** Hora en que ocurrió el crimen.
5. **AREA:** Código numérico que representa el área donde ocurrió el crimen.

6. **AREA NAME:** Nombre del área donde ocurrió el crimen.
7. **Rpt Dist No:** Número de distrito de informe.
8. **Part 1-2:** Código que indica el tipo de crimen (Parte 1: delito leve o Parte 2: delito grave).
9. **Crm Cd:** Código del crimen.
10. **Crm Cd Desc:** Descripción del crimen basada en el código del crimen.
11. **Mocodes:** Códigos de modus operandi (método de operación).
12. **Vict Age:** Edad de la víctima.
13. **Vict Sex:** Género de la víctima.
14. **Vict Descent:** Descendencia (etnia) de la víctima.
15. **Premis Cd:** Código de instalaciones que indica el tipo de ubicación del crimen.
16. **Premis Desc:** Descripción de las instalaciones donde ocurrió el crimen.
17. **Weapon Used Cd:** Código que indica el tipo de arma utilizada.
18. **Weapon Desc:** Descripción del arma utilizada.
19. **Status:** Código que indica el estado de la investigación.
20. **Status Desc:** Descripción del estado de la investigación.
21. **Crm Cd 1-4:** Códigos adicionales de crimen.
22. **LOCATION:** Dirección o ubicación donde ocurrió el crimen.
23. **Cross Street:** Calle transversal cerca de la ubicación del crimen.
24. **LAT:** Coordenadas de latitud de la ubicación.
25. **LON:** Coordenadas de longitud de la ubicación.

2. JUSTIFICACIÓN

La selección del conjunto de *datos "Crime Data from 2020 to Present"* responde a motivaciones personales, particularmente relacionadas con mi formación como estudiante de derecho.

En este contexto, mi objetivo principal consiste en llevar a cabo un análisis detallado y visualización de este conjunto de datos, creando representaciones gráficas que ofrezcan información valiosa. Este tipo de información no solo tiene el potencial de influir en las decisiones cotidianas de las personas, sino que también puede fomentar la participación en iniciativas comunitarias centradas en la seguridad.

Además, la exploración más profunda de variables clave, como la naturaleza específica de los crímenes, los diferentes tipos de crímenes y los patrones emergentes, puede resultar instrumental para la prevención de futuros casos. Este enfoque personal busca no solo comprender sino también contribuir a la prevención y mejora de la seguridad pública, integrando la perspectiva legal en el análisis de datos criminales.

3. RELEVANCIA DEL CONJUNTO DE DATOS

La relevancia del conjunto de datos "*Crime Data from 2020 to Present*" en su contexto es significativa por diversas razones:

- Como se ha comentado anteriormente, el conjunto de datos abarca incidentes de crimen desde el año 2020 hasta el 24 de noviembre de 2023. Esta temporalidad reciente garantiza la actualidad de la información, permitiendo un análisis que refleje la situación del crimen en la Ciudad de Los Ángeles hasta el momento más reciente.
- La naturaleza detallada del conjunto de datos ofrece una comprensión más profunda de la dinámica del crimen en áreas específicas de Los Ángeles. Esto es relevante para diversos colectivos, como residentes, autoridades locales, y profesionales del ámbito legal y de seguridad. La información puede ser crucial para la toma de decisiones informadas y el diseño de estrategias de seguridad adaptadas a necesidades particulares.
- Aunque la descripción del conjunto de datos no especifica la inclusión explícita de variables relacionadas con la perspectiva de género, la presencia de información como la edad, género y descendencia de la víctima puede proporcionar elementos para analizar patrones de crimen desde esta perspectiva.

En resumen, la relevancia del conjunto de datos se destaca por su actualidad, su importancia para diversos colectivos en la ciudad y el potencial para abordar aspectos específicos.

4. COMPLEJIDAD DEL CONJUNTO DE DATOS

A continuación, se proporcionan detalles sobre la complejidad del conjunto de datos "*Crime Data from 2020 to Present*":

- Como se ha mencionado anteriormente, el conjunto de datos consta de 25 variables y abarca un total de 829.778 registros.
- El conjunto de datos incluye una amplia gama de variables, abordando aspectos como la fecha y hora del crimen, detalles sobre la ubicación, características de las víctimas, descripciones del crimen, información sobre armas, estado de la investigación, entre otros.
- El conjunto de datos combina tanto datos categóricos como cuantitativos. Variables como "TIME OCC" y "Vict Age" contienen datos cuantitativos, mientras que variables como "AREA NAME" y "Vict Sex" son categóricas.

- Además de las variables típicas asociadas con incidentes criminales, el conjunto de datos también incluye información adicional como códigos de modus operandi ("Mocodes") y códigos de distrito de informe ("Rpt Dist No").

5. ORIGINALIDAD

Este conjunto de datos se presenta como una actualización continua de los informes de crimen en Los Ángeles, sugiriendo una evolución con respecto a conjuntos de datos previos. Aunque no he identificado visualizaciones específicas previas basadas en estos datos, se vislumbran diversas oportunidades para la visualización y el análisis.

Por ejemplo, se podrían emplear mapas de calor para representar la densidad de crímenes en distintas áreas, gráficos de barras o líneas para ilustrar las tendencias delictivas a lo largo del tiempo, e incluso gráficos de dispersión para explorar relaciones entre diversos tipos de crímenes.

En cuanto al enriquecimiento inicial del conjunto de datos, no se contempla de manera explícita, ya que, a primera vista, presenta una variedad de variables que permiten realizar análisis interrelacionados. No obstante, durante el desarrollo de la segunda parte del proyecto, si se percibe que el análisis resulta limitado, se podría considerar la posibilidad de combinar este conjunto de datos con otros relacionados, como datos demográficos o socioeconómicos, con el fin de obtener nuevas perspectivas y profundizar en el análisis.

6. CUESTIONES QUE RESPONDER CON LA VISUALIZACIÓN DE LOS DATOS

Basándome en el contexto del conjunto de datos sobre crímenes en Los Ángeles y en las oportunidades identificadas para la visualización me planteo algunas preguntas clave que podrían abordarse mediante la representación gráfica de los datos. Estas preguntas podrían incluir:

1. Distribución Geográfica de Crímenes:
 - ¿Cuál es la distribución geográfica de los crímenes en Los Ángeles?
 - ¿Existen áreas de la ciudad con mayor densidad de crímenes?
2. Tendencias Temporales:
 - ¿Cómo han evolucionado los crímenes a lo largo del tiempo desde 2020 hasta la fecha actual?
 - ¿Existen patrones estacionales o tendencias notables?

3. Tipos de Crímenes Más Frecuentes:

- ¿Cuáles son los tipos de crímenes más frecuentes en la ciudad?
- ¿Existen cambios significativos en la prevalencia de ciertos tipos de crímenes?

4. Relación entre Variables:

- ¿Hay alguna correlación entre variables como el tipo de crimen, la ubicación y la hora del día?
- ¿Cómo se distribuyen las edades, géneros y descendencias de las víctimas en relación con ciertos tipos de crímenes?

5. Análisis de Modus Operandi:

- ¿Cuáles son los métodos de operación más comunes utilizados en la comisión de crímenes?
- ¿Existen patrones discernibles en la forma en que se llevan a cabo ciertos tipos de crímenes?