

UNIVERSIDAD DE CONCEPCIÓN

DEPARTAMENTO DE INGENIERÍA INFORMÁTICA Y
CIENCIAS DE LA COMPUTACIÓN

ENTREGA N°2: EXTRACCIÓN Y EXPLORACIÓN DE DATOS
ANÁLISIS DE DATOS

Profesor: Mabel Angélica Vidal Miranda

Estudiantes:

- Cristóbal Figueroa Burgos
- Vicente Ríos Adasme
- Joaquín San Martín Vargas
- Mario Salgado Orellana

5 de Mayo 2025



1. Descripción Detallada de los Datos

El análisis se basa fundamentalmente en la **Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) 2023**, administrada por el **Instituto Nacional de Estadísticas (INE)**. Esta encuesta es la principal fuente de información a nivel nacional sobre la seguridad en zonas urbanas, y su propósito es caracterizar la **victimización** personal y de hogares, la **percepción** de inseguridad en la población, la tasa de **denuncia** de delitos y la **evaluación** ciudadana de instituciones clave como Carabineros, PDI y Fiscalía.

Los datos brutos se obtuvieron del archivo `base-usuario-20-enusc-2023.csv`, el cual contiene aproximadamente 146.000 registros (a nivel de hogar/persona) y más de 1.300 variables. La estructura incluye identificadores geográficos, respuestas a preguntas cerradas (muchas codificadas numéricamente, por ejemplo, 1 = “Sí”, 2 = “No”) y variables sociodemográficas. Dada la codificación numérica, el documento `diccionario-de-variables-enusc-2023.pdf` es indispensable para la correcta interpretación.

La estructura del dataset se caracteriza por:

- **Variables tipo float64 e int64**, con algunas categóricas tipo `object`.
- Una gran cantidad de variables con valores nulos (algunas con más del 97%), debido a la lógica condicional del cuestionario.
- **Cada variable representa una pregunta del cuestionario**, a menudo repetida por región, comuna, barrio u otros contextos.

Existen variables sobre **victimización delictual**, incluyendo:

- Robo con violencia o intimidación
- Robo con sorpresa
- Robo desde vehículo
- Robo con fuerza en vivienda
- Hurto
- Estafa
- Amenazas
- Hackeo
- Virus o malware
- Ciberacoso
- Suplantación de identidad

Estas variables están organizadas por tipo de delito. En el caso del **robo con violencia o intimidación (RVI)**, se consideran variables como:

- `RVI_QUIEN`, `RVI_ACOMP`, `RVI_MES`
- `RVI_DONDE_OCURRIO`, `RVI_DONDE_COMUNA`, `RVI_DONDE_ESP`



- RVI_ROBADO__1 a RVI_ROBADO__13, RVI_VALOR
- Variables sobre denuncia, violencia, relación con el agresor, armas, etc.

Variables clave consideradas para este análisis:

- **Geográficas:** enc_rpc
- **Sociodemográficas:** rph_nse, rph_edad, rphsexo
- **Victimización:** VH_DC, SCREEN_ROB_*, HACK
- **Percepción de inseguridad:** P_AUMENTO_COM, P_AUMENTO_BARRIO, P_INSEG_*
- **Otros:** *_DENUNCIAS, *_MOTIV_NO_DEN, EV_CONFIA_*, P_MOD_ACTIVIDADES_*

2. Proceso de Extracción y Limpieza de los Datos

Los datos fueron importados usando **pandas** desde un archivo CSV con 146.294 registros y 1.301 columnas. Se realizó un proceso de **selección de variables** clave para el análisis.

Debido a la lógica condicional del cuestionario, muchas variables contenían valores nulos. Para tratarlos se aplicaron estrategias como:

- Reemplazo de valores especiales (88, 99, 85)
- Eliminación de columnas o registros irrelevantes o vacíos
- Eliminación de valores bordes, analizando al individuo en específico

Caso Ejemplo: Delito de Hackeo (HACK)

Para el análisis del delito de **hacking**, se detectaron muchos NaN en variables relacionadas. En vez de eliminar registros, se completaron con valores por defecto:

```
1 df_hack.fillna({
2     "HACK_DENUNCIAS_N": 0,
3     "HACK_PERSONAL_N": 0,
4     "HACK_VALOR_H": 99,
5     "HACK_DENUNCIA_COMO": 99,
6     "HACK_SATISF_DEN": 99,
7     "HACK_ACCION_POLICIAL": 99,
8     "HACK_CONTACTO_MP": 99,
9     "HACK_CONT_MP_MOT__1": 0,
10    "HACK_CONT_MP_MOT__2": 0,
11    "HACK_CONT_MP_MOT__3": 0,
12    "HACK_CONT_MP_MOT__4": 0,
13    "HACK_CONT_MP_MOT__5": 0,
14    "HACK_CONT_MP_MOT__88": 0,
15    "HACK_CONT_MP_MOT__99": 0,
16    "HACK_MOTIV_NO_DEN": 99
17 }, inplace=True)
```

Listing 1: Código de limpieza para variables HACK

Esta estrategia permitió preservar registros válidos y mantener una base completa para el análisis de delitos cibernéticos.



3. Análisis Exploratorio Inicial

3.1. Objetivo

El EDA inicial tuvo como fin explorar las distribuciones de las variables de interés, identificar patrones y visualizar relaciones clave entre percepción, victimización y otros factores, utilizando Matplotlib y Seaborn.

3.2. Visualizaciones y Hallazgos Preliminares

3.2.1. Percepción de inseguridad

Como primer paso del análisis, se exploró la percepción de inseguridad a nivel país. En la **Figura 1** se muestra la distribución de respuestas a la pregunta sobre si ha aumentado la inseguridad en el país.

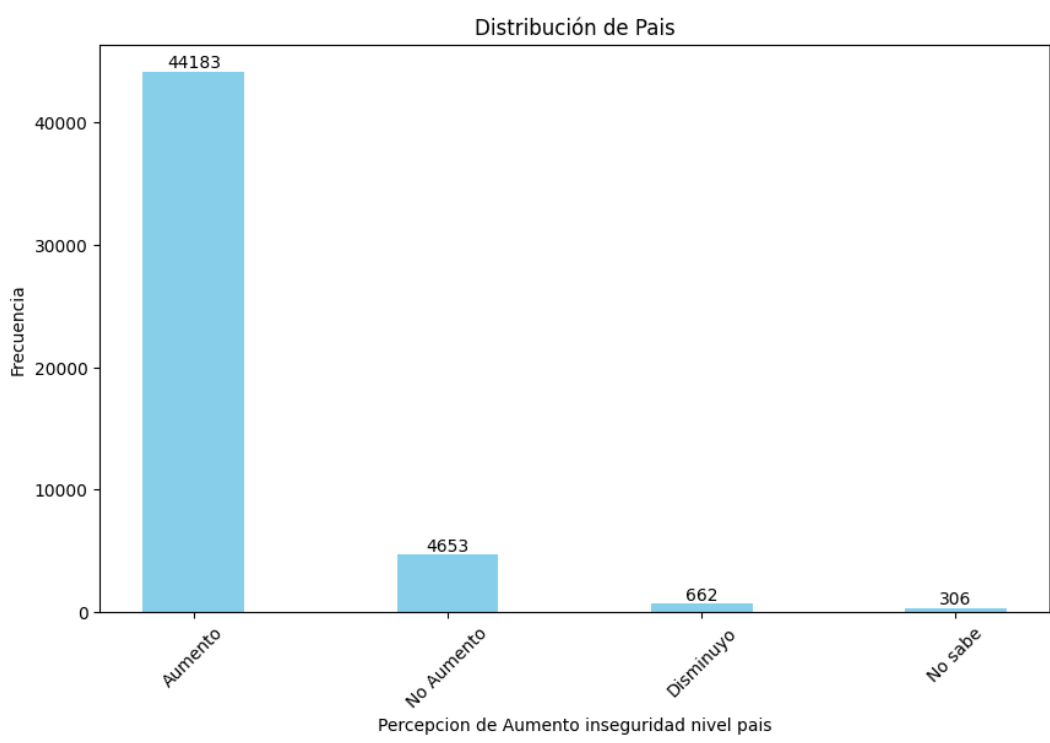


Figura 1: Percepción de aumento de la inseguridad a nivel nacional.

Como se puede observar del gráfico, revela que una **inmensa mayoría** de los encuestados percibe que aumento de la inseguridad (44.183), seguido por un grupo **significativamente menor** que considera que la inseguridad no aumentado, las percepciones de que la inseguridad ha disminuido (662) o que no se sabe (306) son muy poco frecuentes.

Este resultado inicial sugiere una percepción generalizada de deterioro en la seguridad a nivel nacional. Para comprender mejor cómo se experimenta esta inseguridad en la vida cotidiana, se profundizó el análisis



considerando la **percepción de inseguridad durante el día según el grupo etario** como se puede apreciar en la Figura 2.

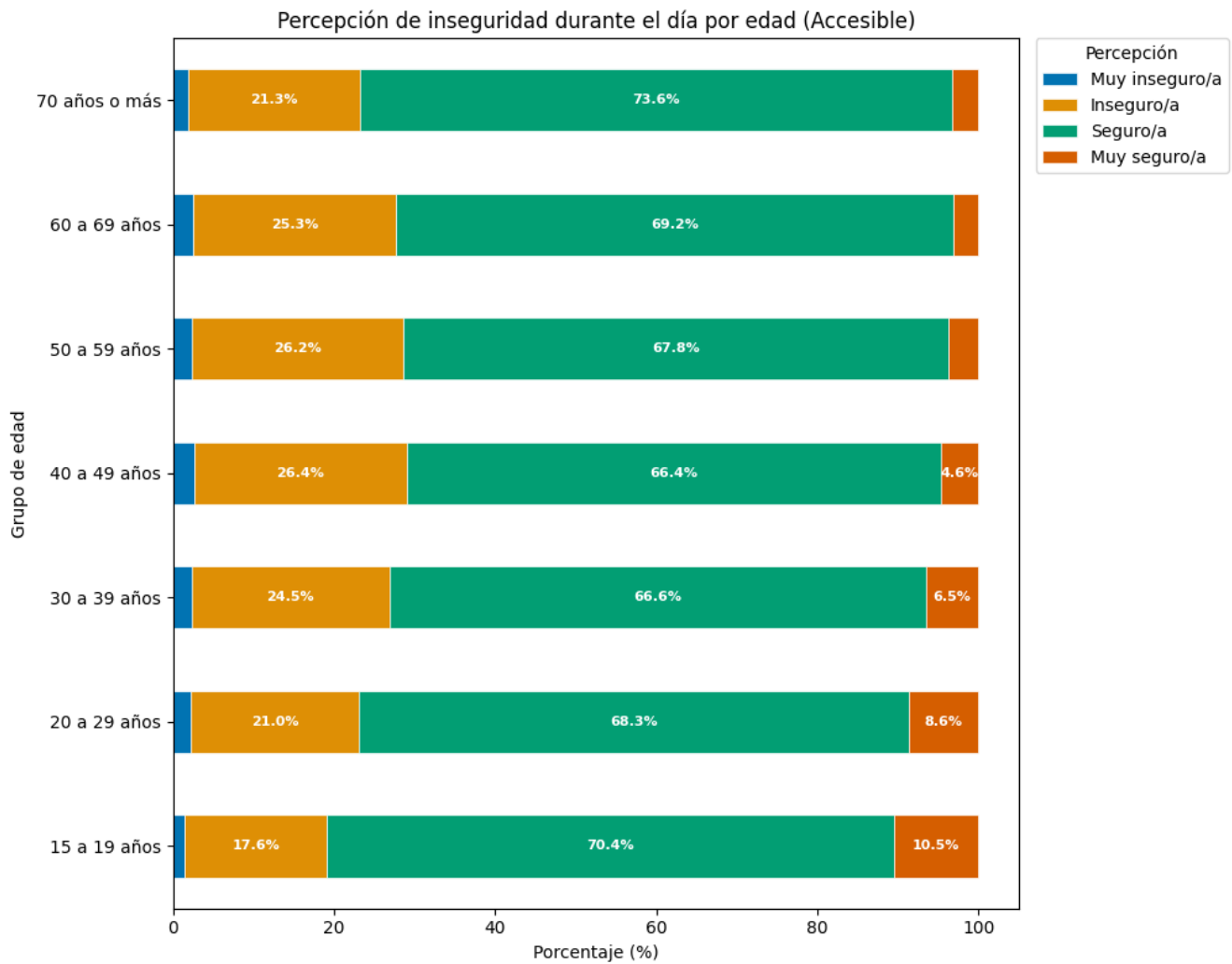


Figura 2: Percepción de inseguridad durante el día por edad.



Este gráfico nos permite observar lo siguiente:

- La mayoría de las personas se sienten "Seguro/a" durante el día en todos los grupos de edad (color verde), con porcentajes que varían entre 66.4 % y 73.6 %,
- El grupo de mayor edad (70 años o más) se siente más seguro durante el día (73.6 %) seguido por los más jóvenes (15-19 años) con 70.4 %.
- Los grupos de edad intermedia (30-59 años) reportan los niveles más bajos de seguridad, aunque siguen siendo altos (66.4 %-67.8 %).
- La percepción de "Inseguro/a" (color naranja) es mayor en los grupos de edad de 40 a 49 años (26.4 %), seguidos por los de 50 a 59 años (26.2 %) y 60 a 69 años (25.3 %),
- Los jóvenes de 15 a 19 años reportan el nivel más bajo de inseguridad (17.6 %) comparado con otros grupos de edad.
- La categoría "Muy seguro/a" (color marrón) es más prevalente en los jóvenes: 10.5 % en el grupo de 15-19 años y 8.6 % en el grupo de 20-29 años
- y por ultimo, la percepción de sentirse "Muy inseguro/a" (color azul) es consistentemente baja en todos los grupos de edad, oscilando entre 1.5 % y 2.6 %.

Esta información sugiere que hay una relación entre la edad y la percepción de seguridad, con los extremos de la escala de edad (más jóvenes y más mayores) sintiéndose generalmente más seguros que los grupos de edad media, quienes parecen mostrar mayor preocupación por la inseguridad durante el día.



3.2.2. cómo afecta esta percepción

Otro Punto que es interesante analizar no es solo ver que nos dice la gente, sino cómo ese sentir afecta su vida. Por ejemplo, quisimos ver si el miedo en el barrio cambia las rutinas. Así que cruzamos la percepción de si la delincuencia aumentó en el barrio (P_AUMENTO_BARRIO) con si la gente ha dejado de salir de noche por temor (P_MOD_ACTIVIDADES_8).

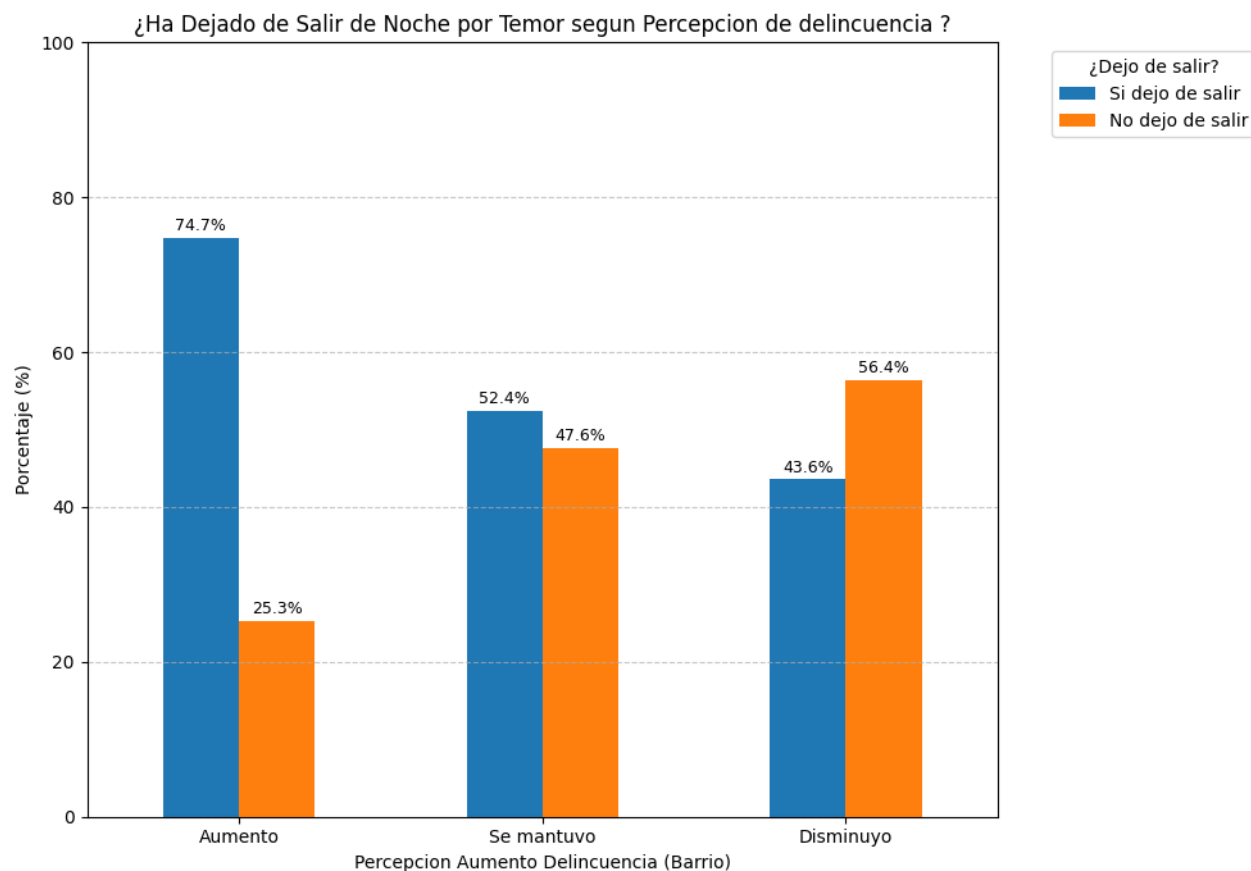


Figura 3: Impacto de la percepción delictual barrial en el hábito de salir de noche.

Lo que muestra la Figura 3 es bien claro:

- Cuando la gente siente que la delincuencia en su barrio **“Aumentó”**, Casi **8 de cada 10 personas** (un 78 % más o menos) prefieren no salir de noche por miedo.
- Si la percepción es que **“Se mantuvo”** igual, el impacto sigue siendo fuerte: cerca de la **mitad** (un 49 % aproximadamente) también deja de salir.
- Y cuando creen que la delincuencia **“Disminuyó”**, aunque baja, igual un **30 %** (unas 3 de cada 10 personas) toma la precaución de no salir de noche.

El miedo que se siente en el barrio le pega directo a la vida de las personas y a cómo usan los espacios públicos, como la calle de noche. Este es solo un ejemplo de cómo con los datos de la ENUSC podemos



conectar lo que la gente piensa y siente con lo que realmente hace. La base de datos da para muchos otros cruces así de interesantes para entender mejor todo el panorama de la seguridad.

3.2.3. Análisis del delito de Hackeo

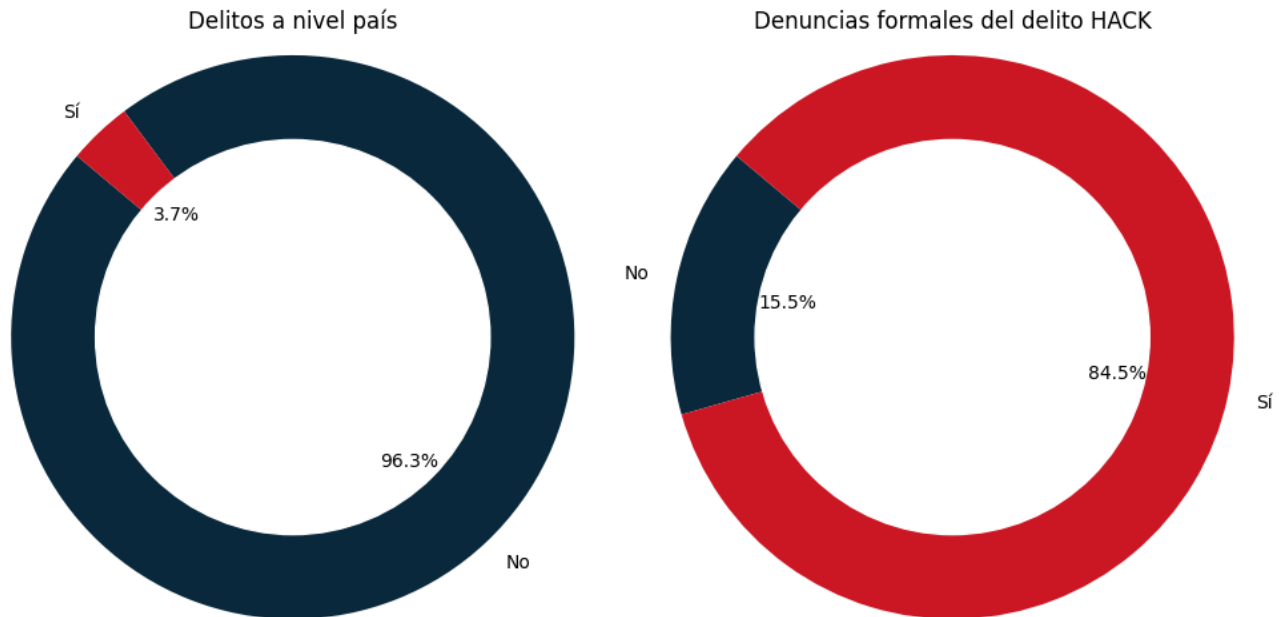


Figura 4: Percepción de inseguridad durante el día por edad.

Delitos a nivel país

El primer gráfico nos muestra que solo el 3.7% de las personas encuestadas declararon ser víctimas del delito de hackeo en el país, mientras que la mayoría con un 96.3% fueron víctimas de otros delitos.

Denuncias formales del delito HACK

En el segundo gráfico vemos más en detalle el delito de hackeo, donde en este se ven las denuncias formales del mismo, con un 84.5% que sí realizó la denuncia formal contra un 15.5% que no lo hizo.

Interpretación

Podemos decir que, pese a que el delito de hackeo representa un porcentaje pequeño del total de delitos, genera una mayor reacción en quienes lo sufren para realizar la denuncia formal.

Podría haber varias causas que atribuyan esto como:

- Tener un mayor registro de evidencia, que pueden ser mensajes, videos, y demás registros digitales.
- Una mayor concientización digital sobre la gravedad de este delito.
- Alta visibilidad social, ya que es un tema ampliamente comentado. Es común conocer a alguien, ya sea un amigo, un familiar o incluso un famoso, que haya sido víctima de este delito.



Por ello, aunque el delito de hackeo no tenga una alta incidencia en nuestro país, este tiene una característica distintiva, que quienes lo sufren tienden a realizar la denuncia formal.

Esto se puede ver contrastado en gran medida con otros tipos de delitos, donde un gran porcentaje de las víctimas optan por no denunciar.

Conclusión

A partir del análisis exploratorio realizado sobre la ENUSC 2023, se identificaron patrones claros que confirman la existencia de una brecha relevante entre la percepción de inseguridad y la tasa de victimización. En la mayoría de los casos, aun cuando las personas no han sido víctimas de delitos, perciben un aumento en la delincuencia, fenómeno que se repite en diversas regiones del país.

También se observaron diferencias según el grupo etario y tipo de delito, así como efectos concretos del miedo percibido en la vida cotidiana. Estos hallazgos refuerzan la importancia de considerar no solo hechos objetivos (como haber sido víctima de un delito), sino también aquellos hechos que capturan la percepción, sensación, opinión o sentimiento de las personas.

En resumen, se evidencia el valor del ENUSC 2023 para estudiar la inseguridad desde una perspectiva multidimensional, integrando tanto variables objetivas como subjetivas.