

2025

Fundamentos de la Ciencia de Datos

Trabajo Práctico Especial

Grupo 01

Rojas, Tomas Alejandro

Campo Zambotti, Agustin



1. Introducción

Introducción y contexto del problema

El presente trabajo se centra en el análisis de siniestros viales ocurridos en las autopistas de la Ciudad de Buenos Aires. La problemática de los accidentes de tránsito constituye un desafío relevante en términos de seguridad vial y gestión urbana, ya que impacta directamente en la salud pública, la movilidad y la calidad de vida de los ciudadanos. Comprender cómo, cuándo y dónde se producen estos siniestros resulta fundamental para diseñar estrategias que contribuyan a reducir su frecuencia y severidad.

Conjunto de datos utilizado

El dataset con el que trabajamos recopila información sobre accidentes registrados en distintas autopistas porteñas. Incluye variables relacionadas con la hora del siniestro, el tipo de vehículo involucrado, la severidad (lesionados y fallecidos), y características del trazado (bandas y ramales). Para enriquecer el análisis, se consultaron fuentes externas como portales informativos (InfoBae, La Nación, entre otros) y documentación oficial de **AUSA (Autopistas Urbanas S.A.)**, lo que permitió contextualizar los datos y comprender mejor las particularidades de cada autopista.

Problemas identificados y correcciones realizadas

Durante la exploración inicial se detectaron algunas inconsistencias y valores faltantes en columnas clave (por ejemplo, hora del accidente, autopista y banda/ramal). Estos casos fueron depurados para garantizar la calidad del análisis. Asimismo, se aplicaron transformaciones a ciertas variables, convirtiéndolas en dicotómicas (0/1) para facilitar la construcción de perfiles y el uso de técnicas estadísticas y de clustering.

2. Análisis exploratorio de los datos

El dataset reúne información sobre siniestros viales ocurridos en autopistas de la Ciudad de Buenos Aires durante 2022. Incluye variables como hora del accidente, autopista y ramal, tipo de vehículo involucrado y severidad (lesionados y fallecidos).

En la exploración inicial observamos que los accidentes se concentran en franjas horarias de alta circulación, que autopistas como la AU Dellepiane y la AU 25 de Mayo presentan mayor frecuencia de siniestros, y que las motos tienen una participación destacada. También detectamos que la mayoría de los casos involucran lesionados, aunque los fallecidos son menos frecuentes.

Durante la limpieza se corrigieron valores que parecían ser “nulos” y no lo eran, y se transformaron variables en dicotómicas para facilitar comparaciones. Estas primeras observaciones motivaron la formulación de hipótesis en distintos niveles: univariadas (patrones simples), bivariadas (relaciones entre dos factores) y multivariadas (perfiles combinados mediante clustering).



Fig. 1: Vista aérea de la Autopista 25 de Mayo, uno de los corredores urbanos incluidos en el análisis. La imagen contextualiza el espacio físico donde se registran los siniestros viales.

3. Hipótesis planteadas y resolución

3.1. Los siniestros ocurren más frecuentemente entre las 12 y las 19 debido a las salidas escolares y laborales.

3.1.1. Definición de la hipótesis

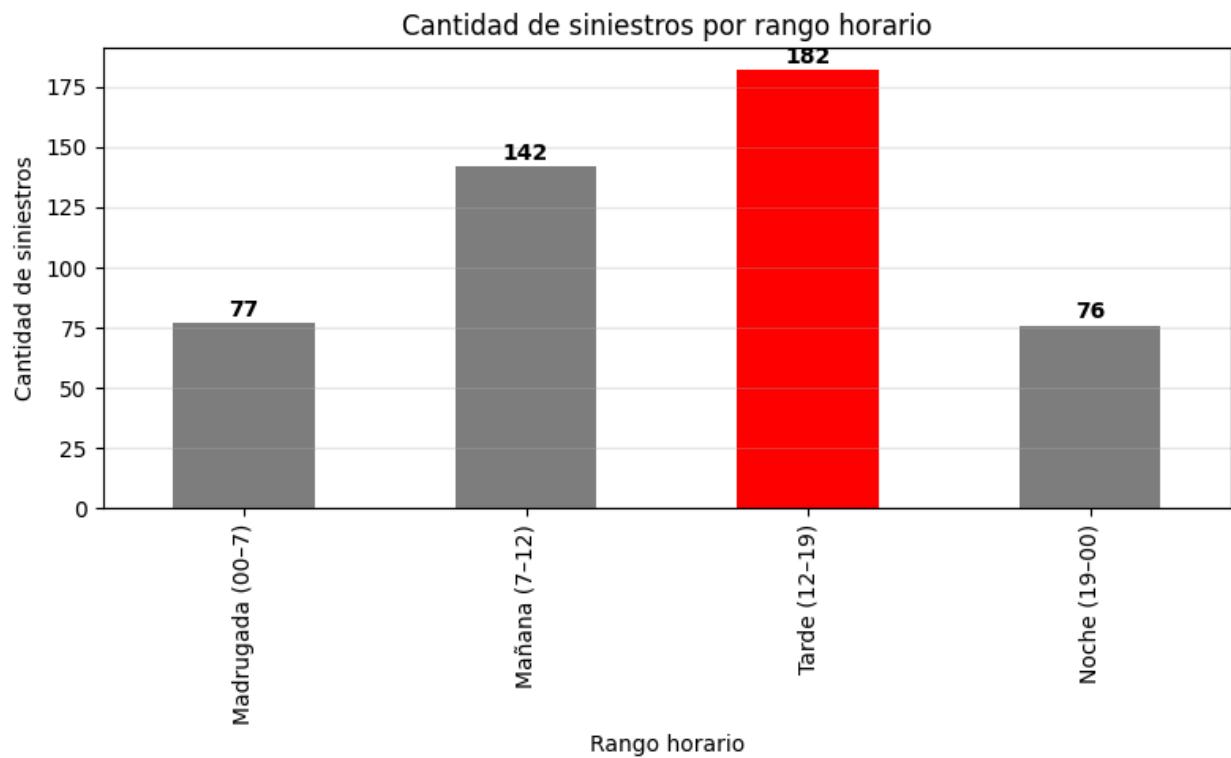
Hipótesis Nula (H_0): La distribución de los siniestros viales es *uniforme* a lo largo de todas las horas del día, y cualquier pico observado en la franja de 12:00 a 19:00 no es estadísticamente significativo.

Hipótesis Alternativa (H_1): La franja horaria comprendida entre las 12:00 y las 19:00 registra una frecuencia de siniestros viales significativamente mayor que las otras franjas horarias del día (*no uniformidad*).

3.1.2. Estrategia de abordaje

Para el análisis, decidimos agrupar los horarios en cuatro franjas principales: *Madrugada (00-07)*, *Mañana (07-12)*, *Tarde (12-19)* y *Noche (19-00)*.

Definimos estos intervalos —especialmente extendiendo la tarde hasta las 19 hs— tras observar que la gran mayoría de los siniestros ocurrieron en los meses de **primavera y verano**. Dado que en esta época del año los días son más largos y la luz solar persiste hasta más tarde, nos pareció adecuado ajustar las franjas para reflejar esa actividad extendida. Ademas, tomamos esta división basandonos en los movimientos típicos de la ciudad, como los horarios de entrada y salida laboral o escolar. Luego, contamos cuántos siniestros cayeron en cada grupo para ver dónde se concentraban.



3.1.3. Resultados obtenidos y discusión

Como se observa en el gráfico, la frecuencia de siniestros es efectivamente más alta por la tarde, alcanzando los 182 casos, seguida por la mañana con 142.

Si bien el rango de 12 a 19 hs es extenso (7 horas), la diferencia con el resto de los horarios es lo suficientemente clara como para **rechazar la Hipótesis Nula (H_0) y aceptar la Hipótesis Alternativa (H_1)**.

Este resultado confirma que el horario vespertino es el de mayor siniestralidad. Esto cobra mayor sentido considerando que los datos corresponden mayoritariamente a primavera-verano: en esta época, la luz solar y la actividad urbana se extienden hasta más tarde, prolongando el riesgo más allá del horario laboral tradicional.

3.2. Los siniestros más frecuentes suelen ocurrir en la autopista 25 de mayo.

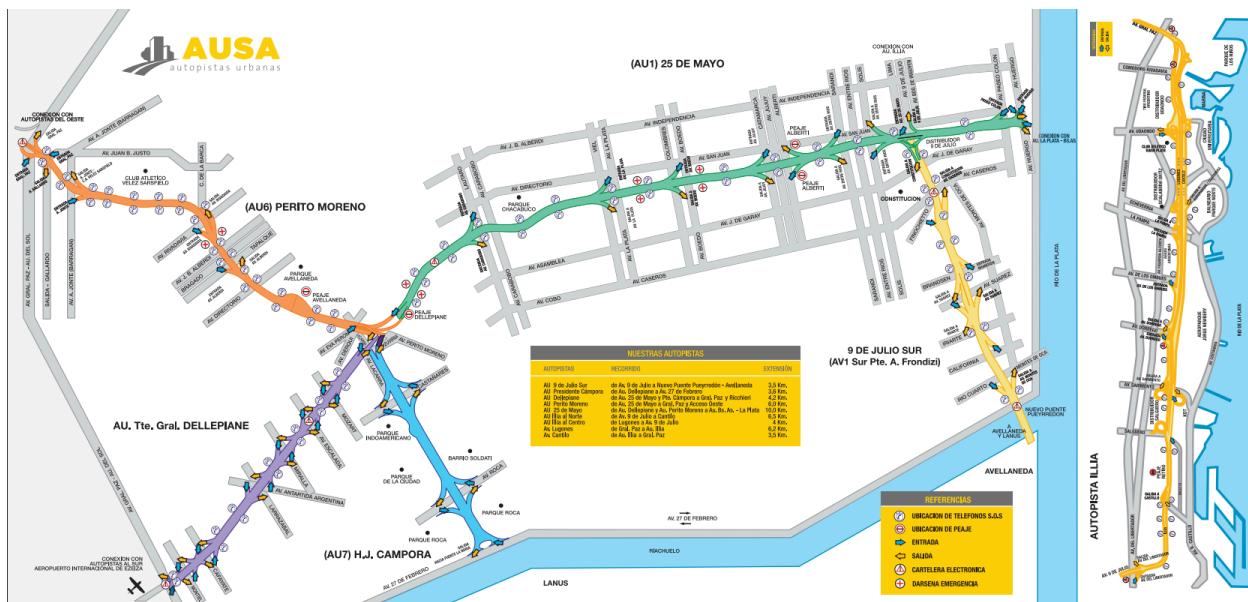
3.2.1. Definición de la hipótesis

Hipótesis Nula (H0): La frecuencia de siniestros viales se distribuye de manera **uniforme** entre todas las autopistas analizadas. (Es decir, todas las autopistas tienen la misma probabilidad de tener un accidente).

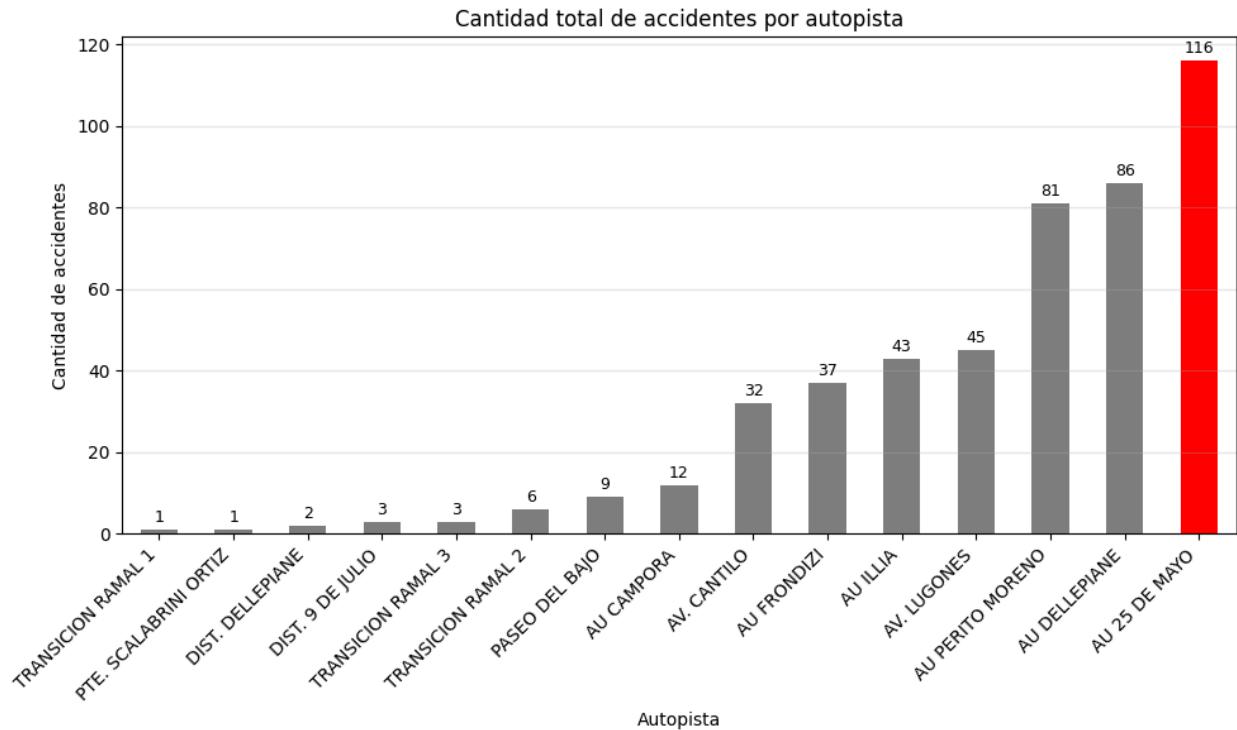
Hipótesis Alternativa (H1): La frecuencia de siniestros viales **NO** se distribuye de manera uniforme. Al menos una autopista tiene una frecuencia de siniestros significativamente diferente de la esperada.

3.2.2. Estrategia de abordaje

Como paso inicial, recurrimos a la cartografía oficial de la ciudad (obtenida de la misma fuente de datos) para contextualizar la ubicación, extensión y complejidad de las principales arterias viales. Esto nos permitió entender mejor la distribución espacial y los puntos de conexión de cada autopista.



Posteriormente, agrupamos los siniestros según la vía donde ocurrieron (autopistas, ramales y transiciones) y calculamos la frecuencia absoluta en cada una para identificar cuál presentaba la mayor concentración de incidentes.



3.2.3. Resultados obtenidos y discusión

Los datos obtenidos nos permiten **rechazar la Hipótesis Nula (H_0)** y **aceptar la Hipótesis Alternativa (H_1)**: la distribución de siniestros no es uniforme.

Como se observa en el gráfico, la **Autopista 25 de Mayo** registra el pico más alto, alcanzando los **116 siniestros** y superando claramente a la segunda (Au Dellepiane).

Este resultado se alinea con las características físicas de la vía que observamos en el mapa: la AU 25 de Mayo no solo es la más extensa en longitud (aprox. 10 km dentro del radio analizado), sino que presenta una alta densidad de accesos, bajadas y transiciones. Esta complejidad estructural, sumada a su extensión, explica por qué concentra la mayor frecuencia de accidentes.

3.3. En los días lluviosos o con neblina se producen más siniestros.

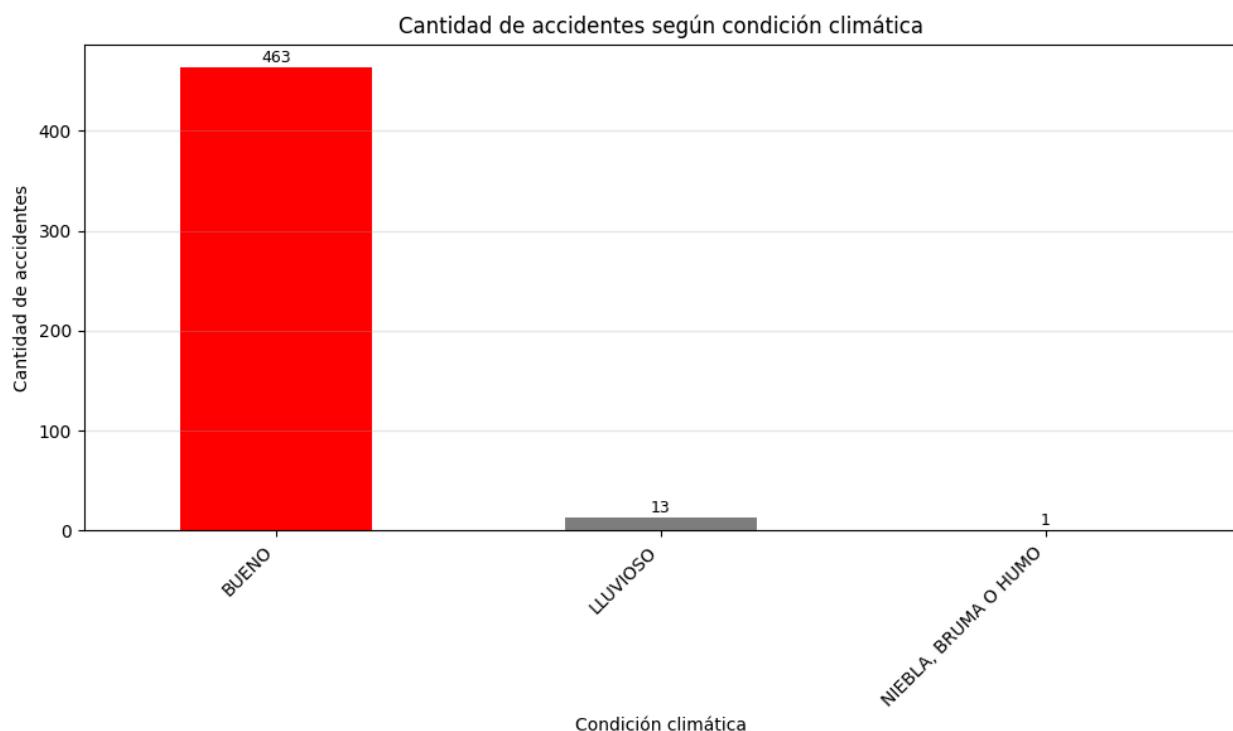
3.3.1. Definición de la hipótesis

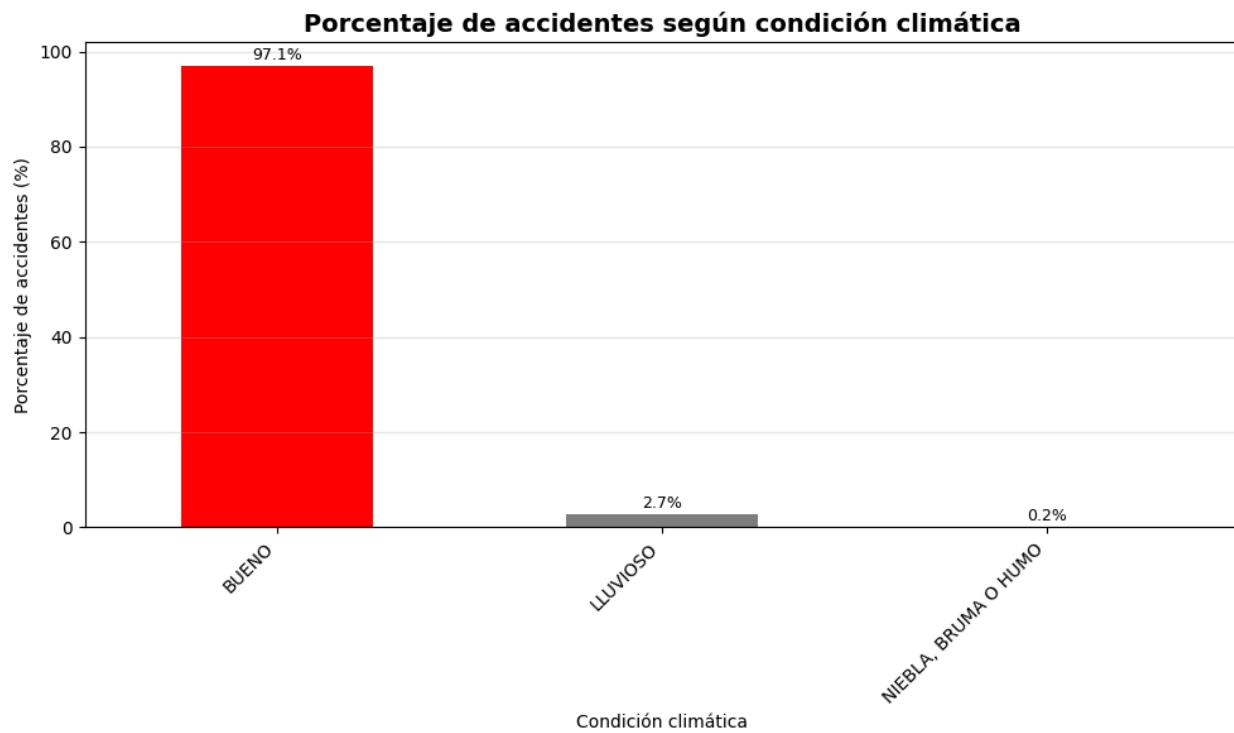
Hipótesis Nula (H_0): Las condiciones climáticas (como "lluvia" o "neblina") **no tienen un efecto significativo** en la frecuencia de los siniestros viales.

Hipótesis Alternativa (H_1): La frecuencia de siniestros viales es **significativamente mayor** en días que registran condiciones climáticas como lluvia o neblina, en comparación con los días que presentan condiciones climáticas normales o despejadas.

3.3.2. Estrategia de abordaje

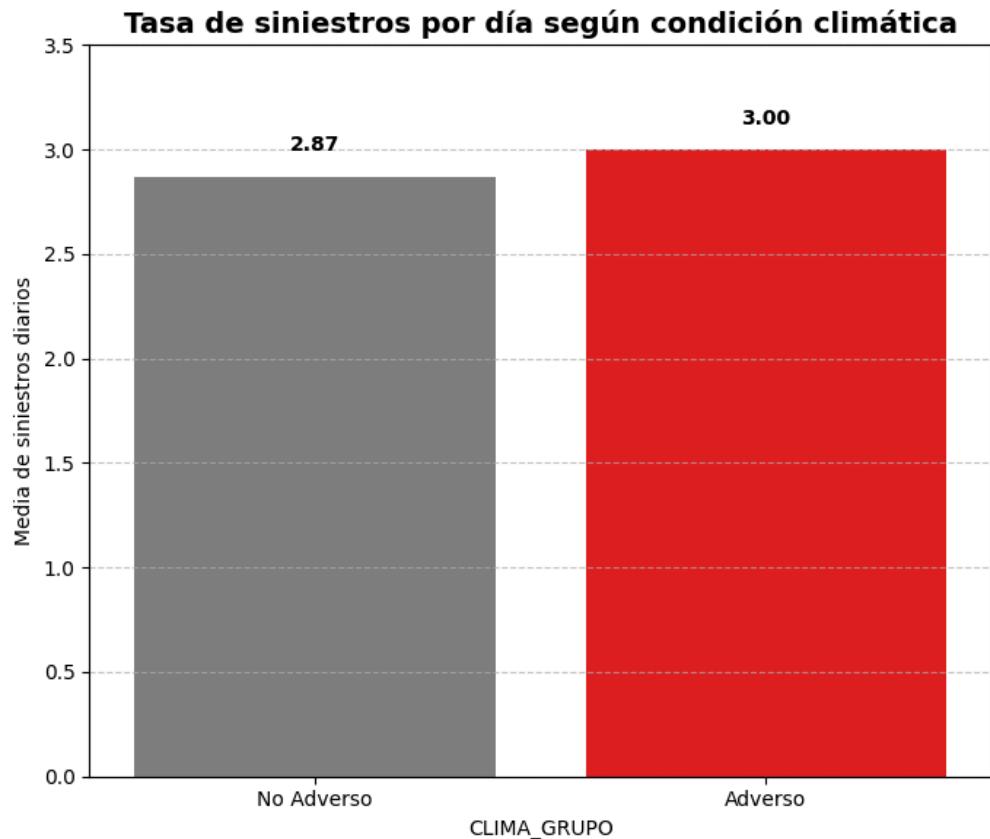
Para evaluar esta hipótesis, **procedimos a agrupar los siniestros según la condición climática registrada**. Observamos que las categorías disponibles fueron *Bueno*, *Lluvioso* y una única ocurrencia clasificada como *Niebla, Bruma y Humo*. Si bien esta última podría corresponder a un error de imputación, decidimos mantenerla en el análisis por su posible relevancia contextual (por ejemplo, cortes de calles o quema de neumáticos).





En una primera aproximación, **observamos que la mayoría de los siniestros se producen bajo condiciones favorables**. Sin embargo, comprendimos que este razonamiento no considera que en un mismo día con clima “bueno” pueden registrarse múltiples siniestros, al igual que en días con lluvia o neblina. Por ello, **optamos por analizar la proporción de siniestros por día**, distinguiendo entre días **adversos y no adversos**, lo cual nos permitió realizar una comparación más justa de la frecuencia relativa de ocurrencias.

Procedimos a construir un gráfico comparativo de las **medias de siniestros diarios** en días clasificados como *Adversos* (lluvia o niebla) y *No Adversos* (clima bueno). El objetivo fue evaluar si existían diferencias significativas en la tasa de ocurrencia de siniestros entre ambos grupos.

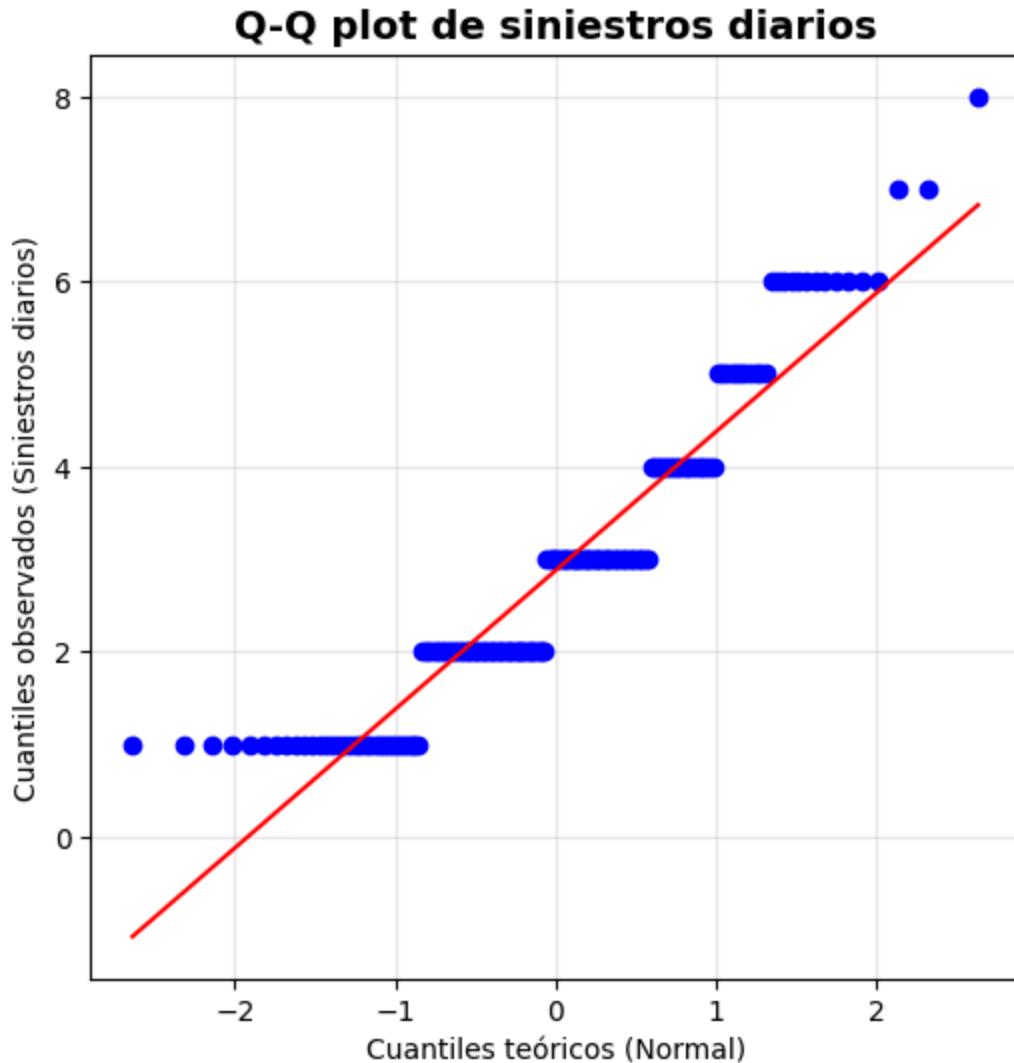


A simple vista, observamos que la diferencia entre ambos grupos es mínima (0.13 siniestros por día), lo que sugiere que no habría un efecto fuerte de las condiciones adversas sobre la tasa de siniestros.

Seguidamente de ver la **media de siniestros diarios** según el clima vamos a complementar esta visualización con pruebas estadísticas de normalidad, homogeneidad de varianzas y comparación de medias, para verificar si podemos realizar algún test frente a estas medias.

Procedimos a analizar la **normalidad de la variable TOTAL_SINIESTROS**, es decir, la cantidad de siniestros registrados por día en todo el período.

- **Test de Shapiro-Wilk:** estadístico = 0.895, **p-valor = 0.000** → la distribución de siniestros diarios **no es normal**.
- **Q-Q plot:** los puntos se desviaron de la línea diagonal, confirmando gráficamente la falta de normalidad en la distribución.



Este resultado nos indica que no podemos aplicar directamente un test t de Student, ya que el supuesto de normalidad no se cumple.

Aun así, estudiamos también, la varianza entre los grupos *adverso* y *no adverso*, veamos que tan homocedasticos son y qué sucede:

- **Test de Levene:** estadístico = 3.687, **p-valor = 0.057** → no se rechaza la hipótesis de igualdad de varianzas.

Esto significa que las varianzas de siniestros diarios en ambos grupos *pueden considerarse homogéneas*.

Comparación de medias

Ahora bien, dado que la normalidad no se cumple, no podemos aplicar un test t (parametrico), entonces optamos por un test no paramétrico:

- **Mann-Whitney U:** estadístico = 472.000, **p-valor = 0.504** → no hay diferencia significativa en la distribución de siniestros diarios entre días adversos y no adversos.

3.3.3. Resultados obtenidos y discusión

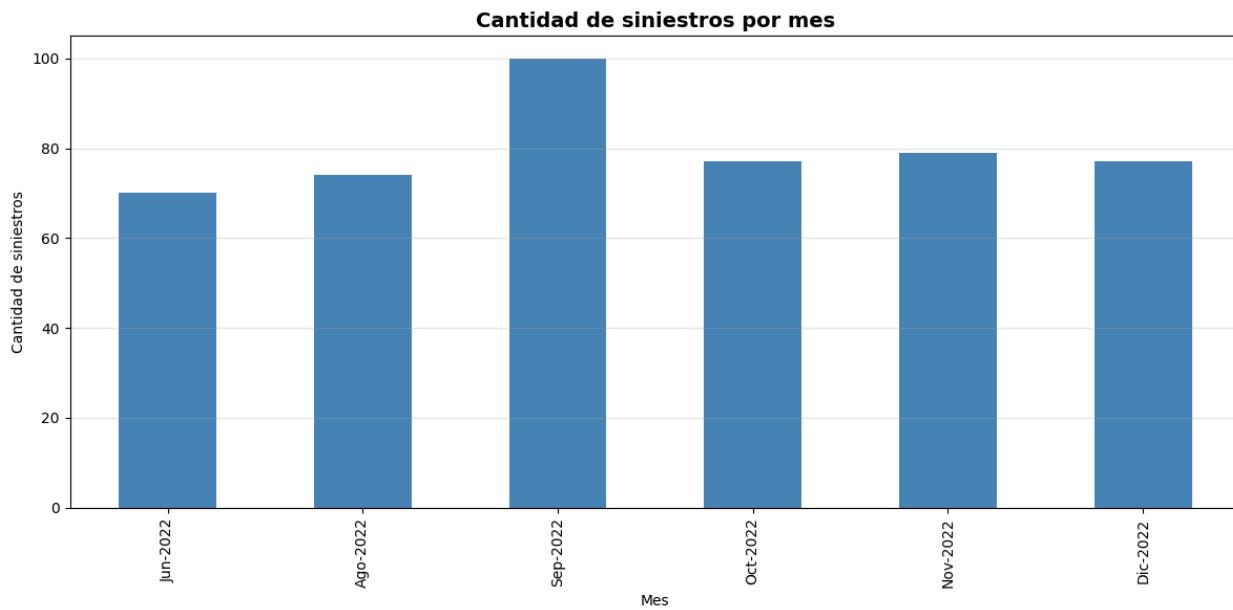
Cuando miramos el gráfico de las medias, vimos que los días con clima adverso (lluvia o niebla) tuvieron en promedio **3 siniestros diarios**, mientras que los días con clima bueno tuvieron **2.87 siniestros diarios**. Esa diferencia es muy pequeña, y ya nos daba la impresión de que no había un efecto fuerte del clima adverso.

Para confirmarlo, probamos con los tests estadísticos:

- **Normalidad (Shapiro-Wilk y Q-Q plot):** la distribución de la cantidad de siniestros por día no es normal. Esto nos obliga a usar un test no paramétrico.
- **Varianzas (Levene):** las varianzas entre los grupos adverso y no adverso se pueden considerar homogeneas.
- **Mann-Whitney U:** el resultado fue un p-valor de **0.504**, lo que significa que no hay diferencia significativa entre los dos grupos.

En otras palabras, tanto el gráfico como los tests nos dicen lo mismo: **no hay evidencia de que los días con lluvia o niebla tengan más siniestros que los días con clima bueno** en este conjunto de datos.

Además, hay que tener en cuenta el contexto: el **97,1% de los siniestros se dieron en días buenos**, y todos los registros están entre junio y diciembre de 2022, meses en los que en Buenos Aires suele llover poco. Eso explica por qué tenemos tan pocos días adversos (solo 5), y por qué no aparece una diferencia clara. Si tuviéramos datos de los meses más lluviosos (marzo a mayo), probablemente el análisis sería distinto.



👉 Así, la conclusión es que **aceptamos la hipótesis nula**: en este período y con estos datos, el clima adverso no tuvo un efecto significativo en la cantidad de siniestros diarios.

3.4. En los siniestros donde participaron motos, hay más cantidad de lesionados.

3.4.1. Definición de la hipótesis

Hipótesis Nula (H_0): La participación de una motocicleta **no está asociada** con un aumento en el promedio de lesionados; la cantidad es igual o menor que en los siniestros sin motos.

Hipótesis Alternativa (H_1): Los siniestros viales que involucran motocicletas **registran un promedio mayor** de lesionados, en comparación con los siniestros donde no participan motos.

3.4.2. Estrategia de abordaje

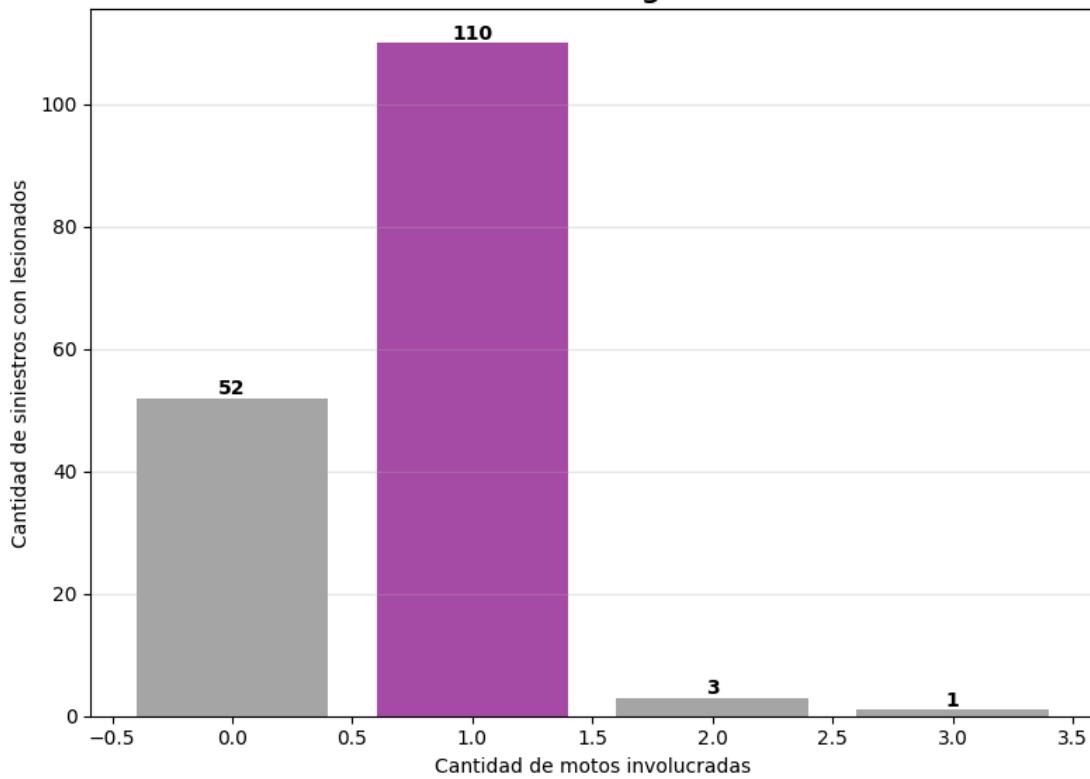
Para evaluar esta hipótesis, decidimos analizar la relación entre la cantidad de motocicletas involucradas en un siniestro y la presencia de lesionados. La estrategia consistió en:

1. **Construir un gráfico de barras** con el eje x representando la cantidad de motos involucradas (0, 1, 2, 3) y el eje y la frecuencia de siniestros con lesionados.
2. **Complementar con porcentajes relativos**, mostrando:
 - La proporción de siniestros con lesionados dentro de cada categoría de motos.
 - El porcentaje que representan esos siniestros sobre el total de siniestros registrados.
3. **Interpretar los resultados** comparando siniestros **con** y **sin** motos, para determinar si la hipótesis alternativa se sostiene.

A continuación dejamos los graficos tanto con las frecuencias absolutas como relativas. En **(1)**, frecuencias absolutas, en **(2)** y **(3)** denotamos las frecuencias relativas pero hacemos la distinción de que en **(2)** *hay al menos un lesionado*, mientras que en **(3)** tomamos el total de lesionados (así sean cero).

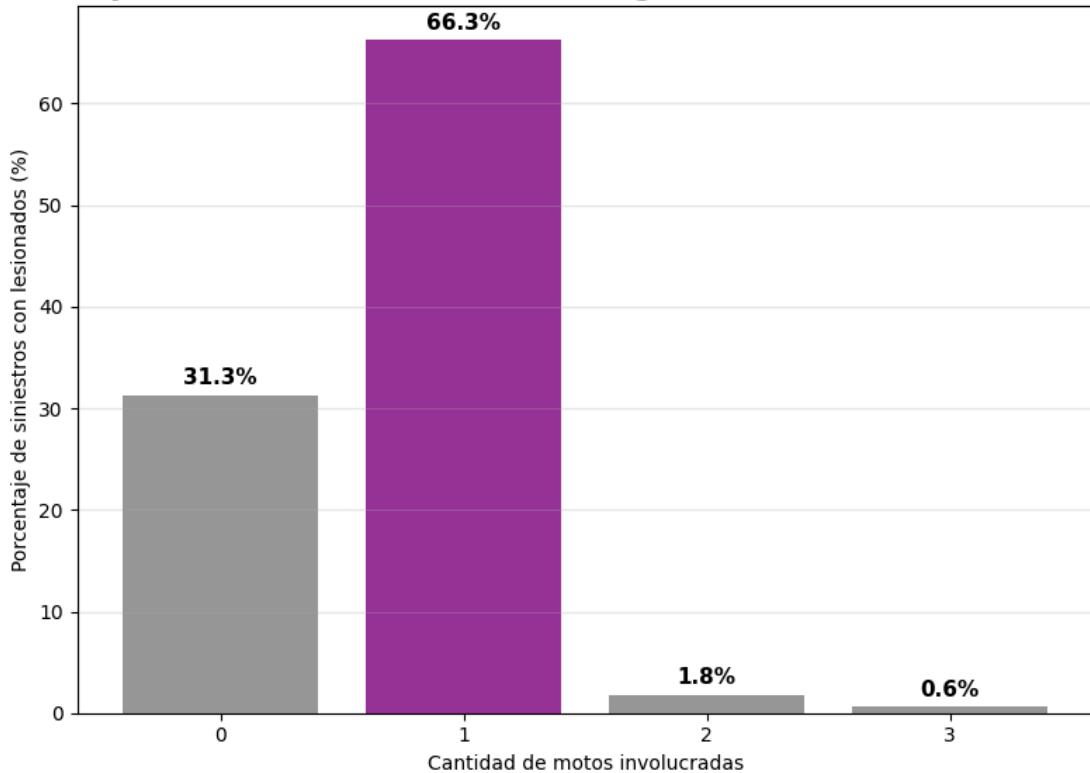
(1)

Frecuencia de siniestros con lesionados según cantidad de motos involucradas



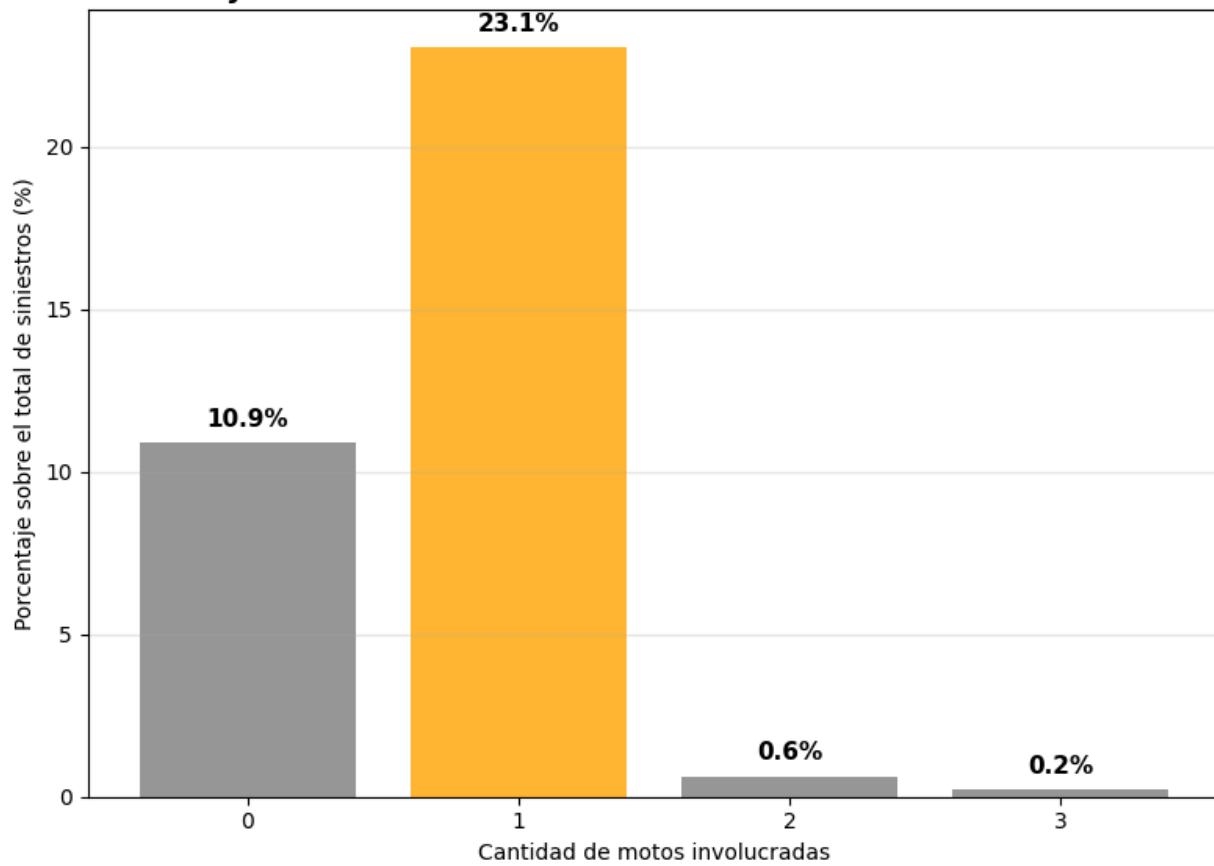
(2)

Porcentaje de siniestros con lesionados según cantidad de motos involucradas



(3)

Porcentaje de siniestros con lesionados sobre el total de siniestros



3.4.3. Resultados obtenidos y discusión

Los resultados muestran una tendencia clara:

- Los siniestros con **una moto involucrada** tienen una probabilidad mucho mayor de generar lesionados que los siniestros sin motos.
- La diferencia es contundente: 66.3% vs 31.3%.
- Los casos con 2 o más motos son muy pocos, por lo que no podemos sacar conclusiones sólidas allí.

Esto confirma lo que suele observarse en la práctica: *los motociclistas son más vulnerables en el tránsito, y cuando ocurre un siniestro, la probabilidad de lesiones es mucho más alta que en otros vehículos.*

Con base en los gráficos y las proporciones analizadas, **rechazamos la hipótesis nula y aceptamos la hipótesis alternativa**. Es decir, los siniestros que involucran motocicletas registran un promedio mayor de lesionados en comparación con los siniestros sin motos.

Observación metodológica

Al tratar dos variables cuantitativas, podríamos haber considerado la construcción de un *scatter plot* (gráfico de dispersión) para analizar si los puntos mostraban alguna relación lineal entre ellos. Sin embargo, esta técnica resulta útil principalmente cuando se trabaja con **variables continuas**, ya que permite visualizar tendencias o patrones de correlación. En nuestro caso, las variables disponibles eran **discretas** (cantidad de motos involucradas y presencia de lesionados), por lo que un scatter plot sólo habría mostrado agrupaciones por clases, sin aportar información adicional relevante. Por este motivo, optamos por utilizar gráficos de barras y porcentajes, que se ajustan mejor a la naturaleza de los datos y permiten una interpretación más clara.

3.5. Los accidentes en la autopista 'Dellepiane' suelen darse más en los horarios pico que informan los peajes.

3.5.1. Definición de la hipótesis

Hipótesis Nula (H0): En la Autopista Dellepiane, **no existe una diferencia significativa** en la frecuencia de siniestros entre las "horas pico" y las "horas valle". La ocurrencia de accidentes es independiente de esta franja horaria.

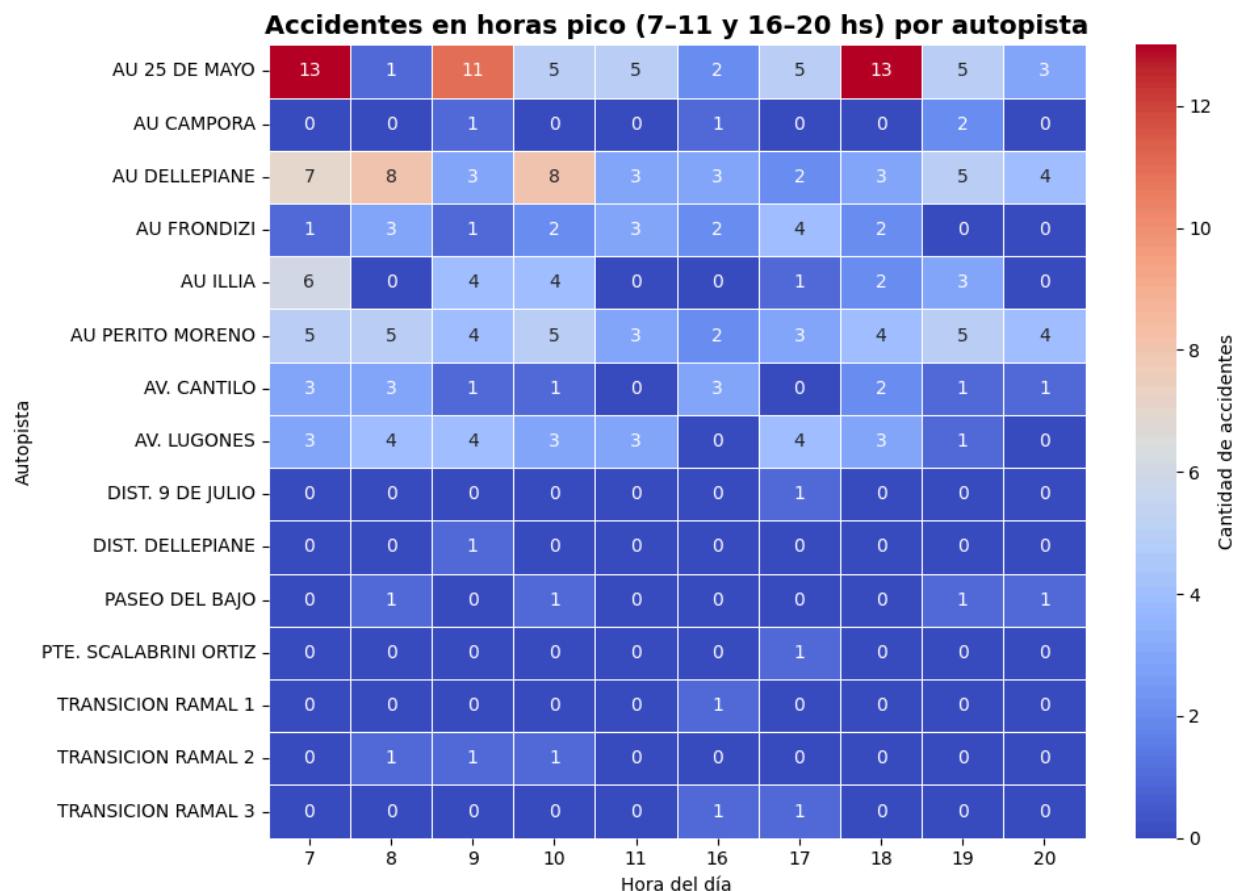
Hipótesis Alternativa (H1): En la Autopista Dellepiane, la frecuencia de siniestros viales es **significativamente mayor** durante las "**horas pico**" en comparación con las "**horas valle**".

3.5.2. Estrategia de abordaje

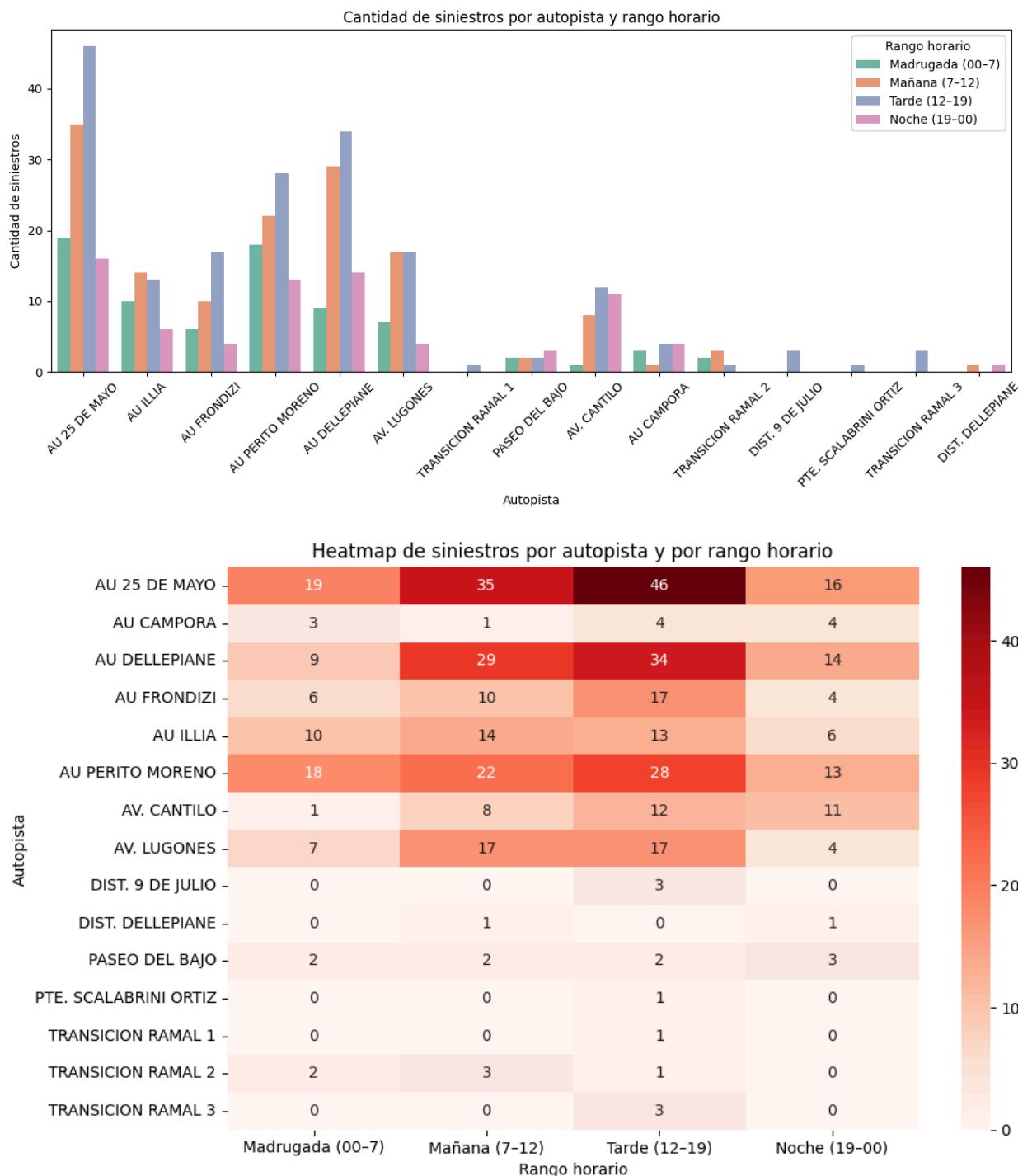
Para poner a prueba esta hipótesis, analizamos la ocurrencia de siniestros en los rangos horarios definidos como pico por los peajes:

- **Mañana:** 7 a 11 hs
- **Tarde:** 16 a 20 hs

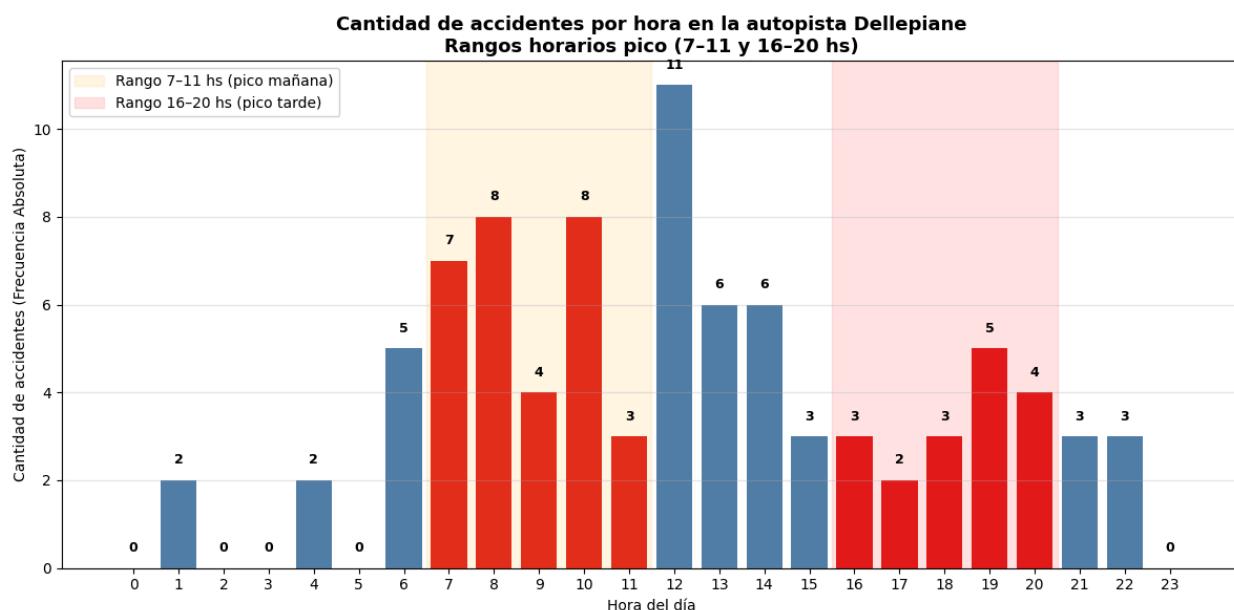
Primero construimos un **heatmap** que muestra la cantidad de accidentes por autopista y hora dentro de los rangos pico. La intensidad del color permite identificar rápidamente las autopistas con mayor concentración de siniestros. Allí se destacan tres: **AU Dellepiane, AU Perito Moreno y AU 25 de Mayo**, que nunca presentan valores nulos en esos horarios. Esto nos permitió pensar en dos grupos: autopistas con alta concentración de accidentes en horas pico y autopistas con baja incidencia.



O si preferimos verlo en rangos horarios como los que hemos definido para hipótesis anteriormente:



Luego, para centrarnos en el foco de la hipótesis, elaboramos un **gráfico de barras** con la distribución de accidentes por hora en la autopista **Dellepiane**. En este gráfico se marcaron con franjas los rangos pico, y se diferenciaron las barras correspondientes a esas horas para visualizar mejor la comparación con el resto del día.



3.5.3. Resultados obtenidos y discusión

Al observar el gráfico de barras de la autopista Dellepiane, notamos que:

- Si bien en el **pico de la mañana (7-11 hs)** se concentran varios accidentes, el **máximo absoluto ocurre a las 12 hs**, fuera del rango pico definido por los peajes.
- El tramo de **12 a 15 hs** presenta más accidentes que el pico de la tarde (16–20 hs).
- Esto confirma que la mayor ocurrencia de siniestros no siempre coincide con las horas pico oficiales.

En consecuencia, los resultados nos llevan a **aceptar la hipótesis nula**: en la autopista Dellepiane no se observa una diferencia significativa que indique que los accidentes se concentran exclusivamente en las horas pico definidas por los peajes.

¿Por qué elegimos hacer foco en esta autopista?

La **Autopista Dellepiane** conecta la zona sur de la Ciudad de Buenos Aires con el centro y la Autopista 25 de Mayo. Es utilizada diariamente por miles de vehículos, especialmente en los horarios pico de ingreso y salida laboral. Según datos de el *Diario “El destape”*, más de **200 mil personas** circulan por este trazado a diario, lo que la convierte en una de las autopistas con mayor densidad de tránsito en la ciudad

En los últimos años, la Dellepiane ha sido escenario de varios siniestros viales de magnitud:

- En noviembre de 2024 se produjo un **choque en cadena que involucró ocho vehículos y dejó al menos 18 heridos**, con tránsito interrumpido durante más de 40 minutos.
- En noviembre de 2025, un **grave accidente en el empalme con la Av. General Paz** provocó el incendio de un vehículo y dejó un conductor con politraumatismos.

3.6. La hora de los siniestros influye en el perfil multivariado de los mismos

3.6.1. Definición de la hipótesis

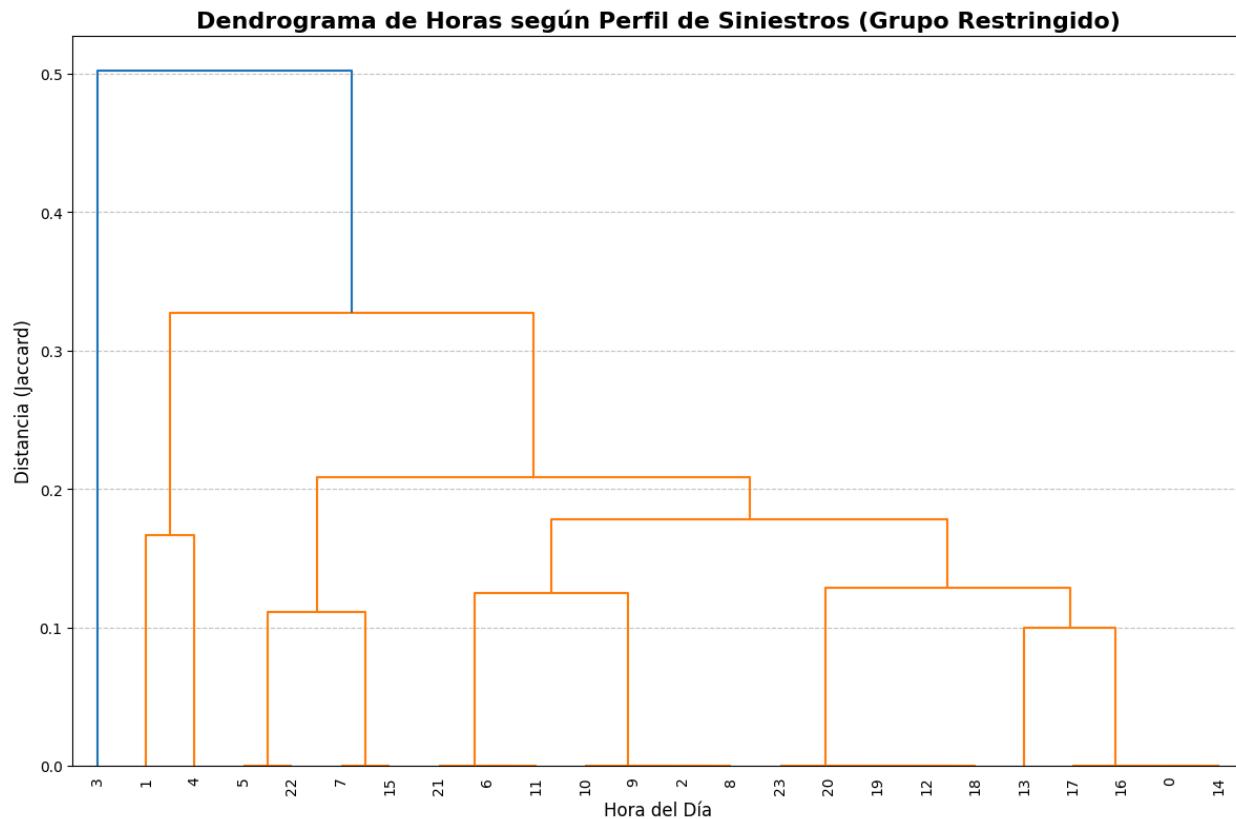
Hipótesis nula (H0): Los perfiles de siniestros viales son iguales entre las distintas horas del día; la hora no influye en la distribución de características.

Hipótesis alternativa (H1): Los perfiles de siniestros viales no son iguales entre las distintas horas; existen agrupamientos de horas con características comunes.

3.6.2. Estrategia de abordaje

Para poner a prueba esta hipótesis, diseñamos un análisis multivariado en varias etapas:

1. **Selección de variables relevantes:** Transformé en dicotómicas (0/1) las siguientes características de los siniestros:
 - **AU_25_DE_MAYO:** si el accidente ocurrió en esa autopista.
 - **LESIONADOS / FALLECIDOS:** si hubo al menos un lesionado o fallecido.
 - **MOTO, LVIANO, BUS, CAMION:** si participó al menos un vehículo de cada tipo.
 - **BANDA_DESCENDENTE,** **BANDA_ASCENDENTE,** **BANDA_RAMAL_DE_ENLACE:** codificación de las tres trazas más frecuentes.
2. **Construcción de perfiles horarios:**
 - Agrupé los siniestros por **hora del día (0–23)**.
 - Para cada hora, marqué 1 si al menos un accidente con esa característica ocurrió en esa franja, y 0 si no.
 - Esto generó una matriz de **24 horas × 10 variables binarias**, que resume la presencia/ausencia de rasgos clave en cada hora.
3. **Clustering jerárquico:**
 - Calculé la **distancia Jaccard** entre los perfiles horarios, ideal para comparar vectores binarios.
 - Apliqué el método de enlace promedio (*average linkage*) para construir el dendrograma.
 - Visualicé el resultado en un **dendrograma de horas**, donde las ramas muestran qué horas comparten patrones similares.



Interpretación de los perfiles binarios de las horas 6, 11 y 21, a modo de ejemplo

Tomando de la matriz binaria, se aprecia que las horas **6, 11 y 21** presentan exactamente el mismo patrón de variables activas e inactivas. Esto explica por qué en el dendrograma aparecen con **distancia cero**: sus perfiles son idénticos.

Variables activas (valor = 1 en las tres horas)

- **AU_25_DE_MAYO:** todos los siniestros registrados en estas horas ocurrieron en la Autopista 25 de Mayo.
- **LESIONADOS:** en las tres horas hubo al menos un accidente con lesionados.
- **MOTO:** participaron motocicletas.
- **LVIANO:** participaron vehículos livianos.
- **CAMION:** participaron camiones.
- **BANDA_DESCENDENTE y BANDA_ASCENDENTE:** se registraron siniestros tanto en sentido descendente como ascendente.

Variables inactivas (valor = 0 en las tres horas)

- **FALLECIDOS:** no se registraron fallecidos en estas horas.
- **BUS:** no hubo participación de colectivos.
- **BANDA_RAMAL_DE_ENLACE:** no se registraron siniestros en ramales de enlace.

Discusión

Este patrón común indica que las horas 6, 11 y 21 comparten un **perfil multivariado idéntico**: accidentes en la Autopista 25 de Mayo, con lesionados, involucrando motos, vehículos livianos y camiones, y distribuidos en ambas bandas principales, pero sin fallecidos ni colectivos.

La coincidencia exacta en todas las variables refuerza la idea de que la **hora del día influye en el tipo de siniestro registrado**, ya que distintos momentos pueden compartir perfiles específicos y repetidos. En este caso, tres franjas horarias separadas (mañana, mediodía y noche) reproducen el mismo patrón, lo que evidencia que ciertos perfiles de accidentes se repiten en distintos momentos del día.

3.6.3. Resultados obtenidos y discusión

El dendrograma evidencia que las horas del día **no se distribuyen de manera aleatoria**, sino que forman agrupamientos consistentes:

- **Horas con distancia cero:** Por ejemplo, las horas **6, 11 y 21** aparecen con perfiles idénticos. Al revisar la matriz binaria, se observa que comparten exactamente las mismas características activas e inactivas, lo que explica su agrupamiento perfecto.
- **Bloques horarios diferenciados:** Se identifican conjuntos de horas que tienden a agruparse por similitud en el tipo de vehículos involucrados y la severidad de los siniestros.
- **Interpretación:** Esto significa que la hora del día no solo influye en la frecuencia de accidentes (como vimos en hipótesis anteriores), sino también en el **perfil cualitativo** de los siniestros: qué vehículos participan, si hay lesionados o fallecidos, y en qué trazado ocurren.

¿Y qué pasa con la hora 3 que quedó aislada?

- Que quede aislada significa que su perfil **no se parece lo suficiente a ningún otro** como para fusionarse en los clusters principales.
- Puede ser que en esa hora ocurrieron accidentes con un patrón único (ej. sin motos, sin camiones, sin lesionados, o en un ramal poco frecuente).
- En términos de clustering, es un **outlier**: una hora cuyo vector binario es distinto y no encuentra “pareja” cercana.
- Esto no es un error, sino un hallazgo: muestra que no todas las horas se comportan igual y que hay franjas con perfiles singulares.

En consecuencia, los resultados permiten **rechazar la hipótesis nula y aceptar la hipótesis alternativa**: los perfiles de siniestros viales dependen de la hora del día y se organizan en agrupamientos consistentes. La hora, por lo tanto, **predice parcialmente la combinación de factores involucrados en los accidentes**.

4. Conclusiones

El análisis de los siniestros viales en autopistas de Buenos Aires nos permitió confirmar que factores como la hora del día, la autopista involucrada y el tipo de vehículo influyen en la ocurrencia y características de los accidentes. A través de hipótesis univariadas, bivariadas y multivariadas pudimos demostrar que los siniestros no se distribuyen de manera aleatoria, sino que responden a patrones consistentes: mayor frecuencia en la tarde, concentración en la AU 25 de Mayo, vulnerabilidad de las motos y perfiles horarios diferenciados.

Si bien algunas hipótesis, como la influencia del clima adverso, no mostraron efectos significativos en este período, el trabajo evidenció la importancia de analizar múltiples dimensiones para comprender mejor el fenómeno.

Finalmente, queremos destacar que gracias a las herramientas metodológicas y estadísticas que se nos proporcionaron en la materia pudimos llevar adelante esta investigación, desarrollar la curiosidad por explorar los datos y encontrar respuestas fundamentadas. Esto nos permitió no solo validar hipótesis, sino también adquirir una mirada crítica y analítica sobre la problemática de la seguridad vial.

A modo de conclusion y breve visualizacion dejamos una tabla con las hipótesis descriptas rápidamente y sus resultados:

Nº	Hipótesis planteada	Hipótesis nula (H_0)	Hipótesis alternativa (H_1)	Resultado
3.1	Los siniestros ocurren más frecuentemente entre las 12 y las 19	La distribución de siniestros es igual en todas las horas del día	La franja 12–19 tiene más siniestros	Se rechaza H_0. La tarde concentra más siniestros
3.2	Los siniestros más frecuentes ocurren en la AU 25 de Mayo	La frecuencia de siniestros es igual en todas las autopistas	Al menos una autopista concentra más accidentes	Se rechaza H_0. AU 25 de Mayo registra el mayor número de siniestros
3.3	En días lluviosos o con neblina se producen más siniestros	El clima no influye en la frecuencia de siniestros	El clima adverso aumenta la frecuencia	Se acepta H_0. No hubo diferencia significativa en este período
3.4	En siniestros con motos hay más lesionados	La proporción de lesionados es igual en siniestros con y sin motos	Los siniestros con motos tienen más lesionados	Se rechaza H_0. Las motos duplican la probabilidad de lesionados
3.5	En la AU Dellepiane los siniestros se concentran en horas pico	La frecuencia de siniestros es igual en horas pico y valle	Los accidentes son más frecuentes en horas pico	Se acepta H_0. No se concentran exclusivamente en horas pico
3.6	La hora influye en el perfil multivariado de los siniestros	Los perfiles de siniestros son iguales entre todas las horas	Existen diferencias: agrupamientos de horas con perfiles comunes	Se rechaza H_0. Se identificaron clusters y perfiles idénticos

Objetivos y propuestas futuras

A partir de los hallazgos, consideramos que el análisis puede servir como base para pensar en medidas de prevención y seguridad más específicas, tales como:

- **Instalación de cámaras de control de velocidad** en tramos críticos de las autopistas.
- **Carteles de limitación de velocidad más claros y visibles**, especialmente en accesos y ramales.
- **Mayor presencia policial y controles viales** en horarios de alta siniestralidad.
- **Campañas de concientización dirigidas a motociclistas**, dado que se identificó su mayor vulnerabilidad.
- **Monitoreo inteligente del tránsito** mediante sistemas de detección temprana de congestión y condiciones de riesgo.

Estas propuestas buscan vincular el análisis estadístico con acciones concretas que contribuyan a reducir la frecuencia y severidad de los siniestros, reforzando el valor práctico del trabajo realizado.

Referencias

- InfoBae. *Noticias sobre tránsito y siniestros viales en la Ciudad de Buenos Aires.* Disponible en: <https://www.infobae.com>
- La Nación. *Cobertura de accidentes y movilidad urbana en autopistas porteñas.* Disponible en: <https://www.lanacion.com.ar>
- Autopistas Urbanas S.A. (AUSA). *Información oficial sobre la red de autopistas de la Ciudad de Buenos Aires.* Disponible en: <https://www.auysa.com.ar>
- El Destape. *Datos sobre la circulación diaria en la Autopista Dellepiane.* Disponible en: <https://www.eldestapeweb.com>
- Clarín. *Cobertura de accidentes recientes en autopistas porteñas.* Disponible en: <https://www.clarin.com>
- Google. *Gemini: modelo de inteligencia artificial multimodal.* Disponible en: <https://deepmind.google/technologies/gemini>
- Microsoft. *Copilot: asistente de inteligencia artificial para productividad y análisis de datos.* Disponible en: <https://copilot.microsoft.com>
- Cátedra: Fundamentos de Ciencia de Datos, Facultad de Ciencias Exactas. *Filminas y materiales de clase sobre estadística aplicada y análisis exploratorio.* Documento interno, 2025.
- Los collabs de nacho ❤️