

2025

Fundamentos de la Ciencia de Datos

Trabajo Práctico Especial

Grupo 13

Arouxet, Justo

Yunis, Clara

1. Introducción

En este informe vamos a analizar la seguridad vial en la red de autopistas de la Ciudad Autónoma de Buenos Aires (CABA) en el año 2022. Nuestro objetivo no es meramente cuantificar la cantidad de incidentes, sino aplicar las herramientas de la Ciencia de Datos que aprendimos para identificar patrones, comprender la dinámica de los siniestros viales, hacer predicciones permitiendo a las autoridades dirigir recursos (patrullas, mantenimiento, señalización, campañas de consentimiento) a los puntos de mayor vulnerabilidad, con el fin de reducir la tasa de lesionados y fallecidos, como también la frecuencia de siniestros.

Los datos están disponibles en el portal de datos abiertos del Gobierno de la Ciudad de Buenos Aires (Buenos Aires Data). Su principal fuente de información es el Observatorio de Movilidad y Seguridad Vial de la Ciudad de Buenos Aires (OMSV), que basa sus estadísticas en sumarios de la Policía de la Ciudad por los delitos de lesiones culposas y homicidios culposos viales.

Además, los informes públicos del observatorio señalan que colaboran con otros organismos como hospitales, el Sistema de Atención Médica de Emergencias (SAME), y la empresa Autopistas Urbanas S.A. (AUSA), para completar y validar la información.

Concluimos que este dataset es el trabajo colectivo de varias instituciones (policía, hospitales, fiscalías, ministerios tanto de justicia y seguridad o transporte).

Nuestro trabajo se estructura de la siguiente manera:

1. Análisis Exploratorio de Datos.
2. Planteo de Hipótesis.
3. Validación de Hipótesis.
4. Conclusión Final.

2. Análisis exploratorio de los datos

El dataset registra los siniestros viales en las Autopistas de la Ciudad Autónoma de Buenos Aires (CABA) durante el periodo de junio a diciembre de 2022. Debemos mencionar que la información perteneciente al mes de julio no está incluida.

Contamos con 477 observaciones y 14 columnas.

Comenzamos analizando las variables, sus tipos, que representan y qué información aportan:

- **FECHA:** Fecha del siniestro en formato YYYY-MM-DD (date)
- **HORA:** Hora del siniestro en formato 24 horas. (int)
- **AUTOPISTA:** Nombre de la autopista donde ocurrió el siniestro. (string)
- **BANDA y/o RAMAL:** Indica si el siniestro ocurrió en la banda ascendente, descendente o en un ramal (tramo de carretera que forma parte de un enlace y se desvía del tronco principal)
- **PK:** Punto kilométrico donde ocurrió el siniestro. (string)
- **CONDICIONES METEOROLÓGICAS:** Estado del clima al momento del siniestro (Ej.: BUENO, LLUVIOSO, (NIEBLA, BRUMA O HUMO)).
- **SUPERFICIE DE LA VÍA:** Estado de la superficie de la vía (Ej.: SECA, MOJADA/HUMEDA, OTRA).
- **LESIONADOS:** Cantidad de personas lesionadas en el siniestro. (int)
- **FALLECIDOS:** Cantidad de personas fallecidas en el siniestro. (int)
- **TIPO DE SINIESTRO:** Clasificación del siniestro (Ej.: Colisión, Obstáculo fijo, Siniestro sin colisión, etc.).
- **MOTO:** Cantidad de motocicletas involucradas en el siniestro. (int)
- **LVIANO:** Cantidad de vehículos livianos involucrados en el siniestro. (int)
- **BUS:** Cantidad de autobuses involucrados en el siniestro. (int)
- **CAMION:** Cantidad de camiones involucrados en el siniestro. (int)

Dada la definición específica de cada variable, podemos dividir el análisis en 3 esferas:

El foco geográfico y temporal (el dónde y el cuándo). El dataset nos proporciona datos geográficos (**AUTOPISTA, PK, BANDA y/o RAMAL**) y temporales (**FECHA, HORA**) que nos permiten identificar dónde y cuándo están ocurriendo los siniestros.

Factores contribuyentes (el por qué y el cómo). Buscamos comprender los factores de riesgo a través de variables que describen el entorno (**CONDICIONES METEOROLÓGICAS, SUPERFICIE DE LA VÍA**) y la dinámica del accidente (**TIPO DE SINIESTRO**) nos permite distinguir y clasificar los siniestros para en un futuro proponer medidas de prevención específicas o comprender el estado y las carencias de la/s carretera/s.

Consecuencias y vulnerabilidad (La gravedad). Las variables de resultado (**LESIONADOS, FALLECIDOS**) definen la gravedad del problema y la vulnerabilidad de los usuarios. El objetivo principal de la seguridad vial es reducir los indicadores de fatalidad y lesiones. Queremos que la Ciencia de Datos en este proyecto apunte a modelar la probabilidad de que un siniestro resulte en lesiones o fallecidos basándose en los sucesos anteriores. El dataset también nos permite identificar vulnerabilidad por tipo de vehículo (**MOTO, LVIANO, BUS, CAMION**).

A continuación vamos a hacer “zoom” a algunas variables de cada esfera que nos parecieron interesantes:

AUTOPISTA: Esta variable nos da el nombre de donde ocurrió el siniestro. Si hacemos una tabla de frecuencias podemos ver que la AU. 25 DE MAYO lidera la tabla, seguido de la AU DELLEPIANE y AU. PERITO MORENO. Esto nos da un puntapié para analizar qué está ocurriendo en esas autopistas ya que estas 3 tienen una gran diferencia con respecto a las otras (casi el doble de siniestros).

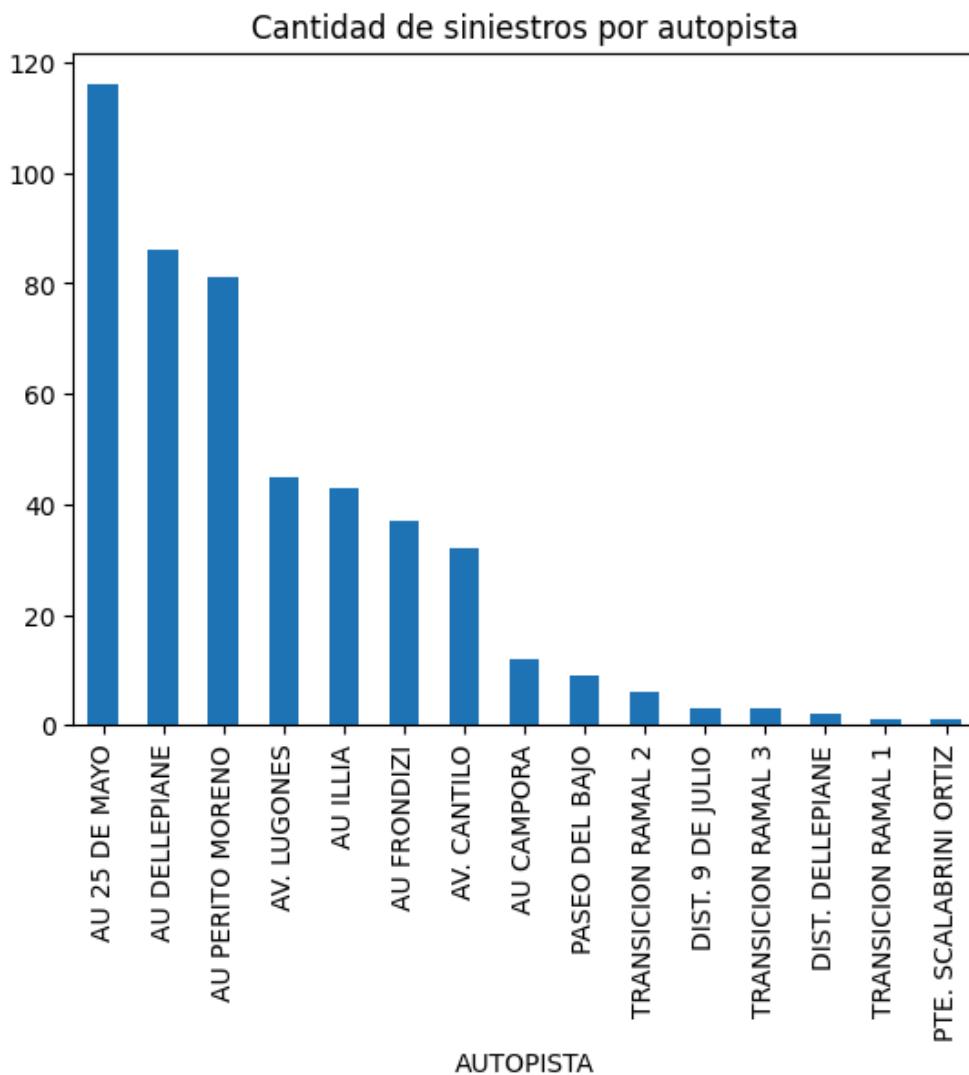


Gráfico de barras de la frecuencia de siniestros por la variable 'AUTOPISTA'

FALLECIDOS: Al analizar esta variable, observamos que muchos de sus campos están en 0. Si bien esto limita las operaciones que podemos realizar, es una noticia positiva, ya que indica una baja cantidad de decesos en los siniestros.

Cantidad de siniestros por cantidad de fallecidos

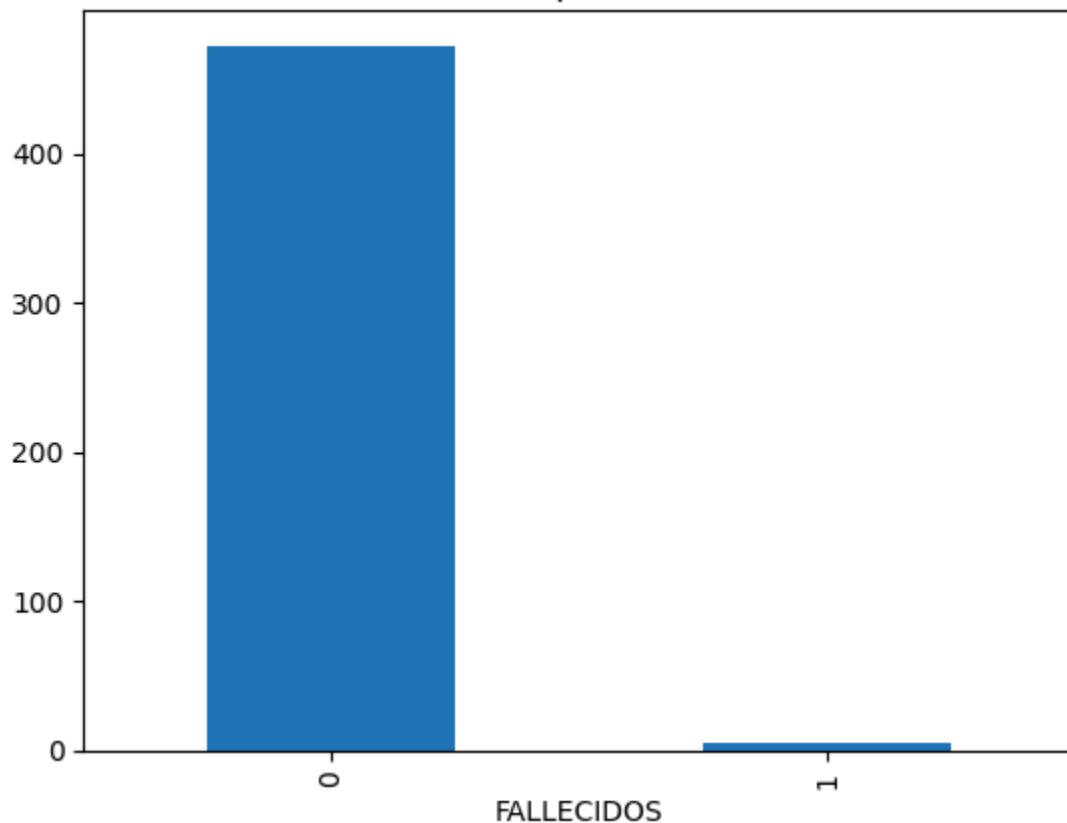


Gráfico de barras de la cantidad de 'FALLECIDOS' por siniestro

LESIONADOS: Si analizamos la media de lesionados de nuestro dataset, podemos ver que es de 0.454 aproximadamente. Esto nos da una idea de que en la mayoría de los siniestros podría no haber lesionados. Nos llamó la atención que el valor máximo que toma la variable es 14. Decidimos investigar al respecto y descubrimos que corresponde a un triple choque en Autopista Dellepiane entre un vehículo liviano, un colectivo y un camión. Con esto corroboramos que no se trata de un error a la hora de cargar los datos si no de un hecho fatal.

Cantidad de lesionados por siniestro

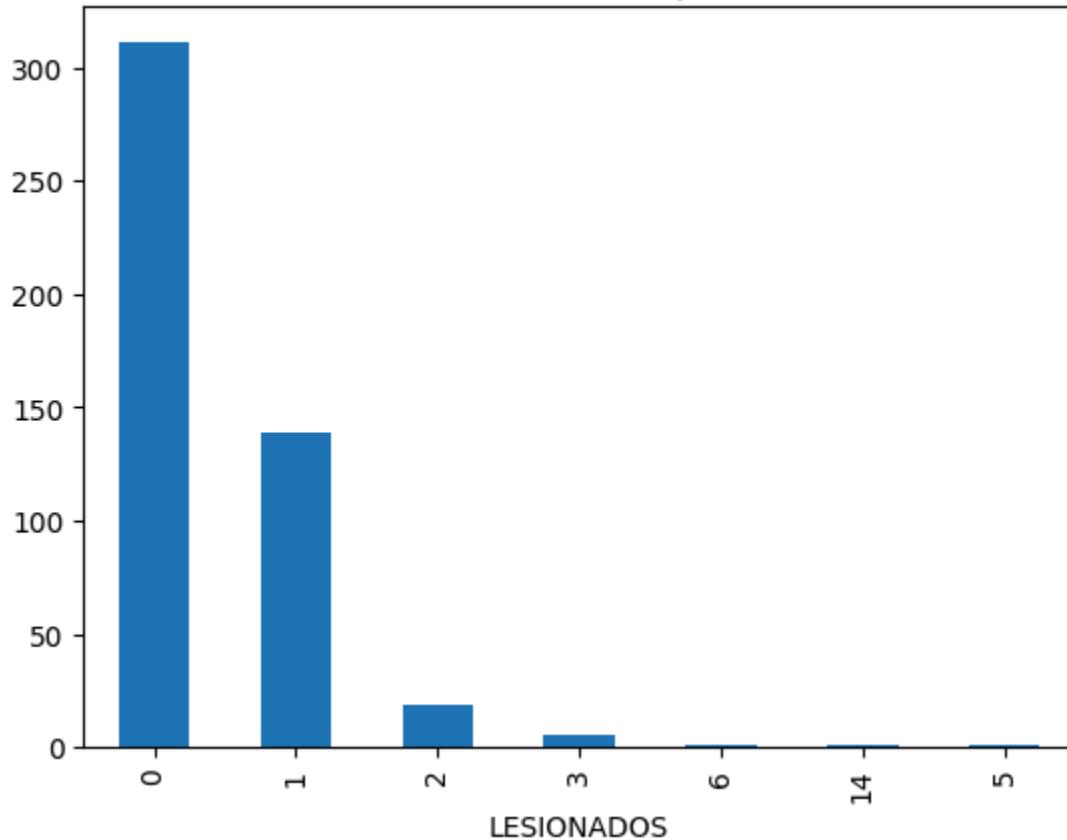


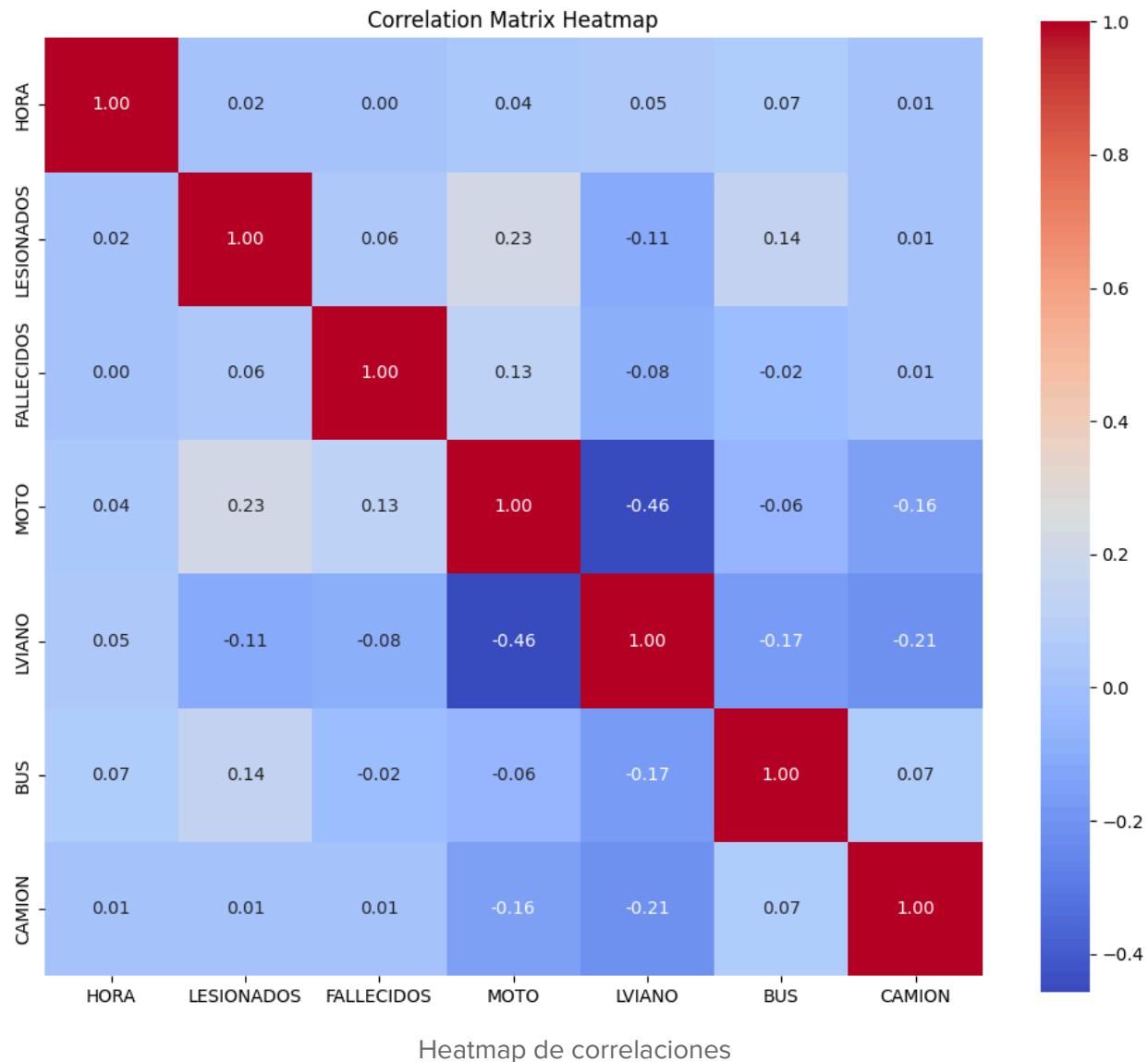
Gráfico de barras de la cantidad de 'LESIONADOS' por siniestro

Luego, observamos la existencia de ceros y valores nulos:

- BANDA y/o RAMAL: presenta 1 valor nulo con el símbolo "-".
- PK: presenta 18 valores nulos representados por el símbolo "-".

Consideramos que el KM donde ocurrió el siniestro no fue registrado. Lo reemplazamos con NaN para un posterior análisis.

Estudiamos las correlaciones entre las variables de tipo numérico con un heatmap:



Vemos que las únicas variables con un coeficiente de correlación interesante son MOTO y LIVIANO. El mismo nos indica que existe una correlación negativa entre ambas, esto se traduce a que cuando hay presencia de autos livianos hay menos presencia de motos involucradas en el siniestro.

3. Hipótesis planteadas y resolución

3.1. Hipótesis 1

3.1.1. Definición de la hipótesis

La siniestralidad en la AU 25 de Mayo no es uniforme, hay alta concentración de siniestros en ciertos tramos críticos.

Planteamos esta hipótesis ya que observamos en los datos que la AU 25 de mayo concentra la mayor cantidad de siniestros del dataset. Sospechamos que puede haber zonas que requieren de alta concentración y habilidad por parte de los conductores.

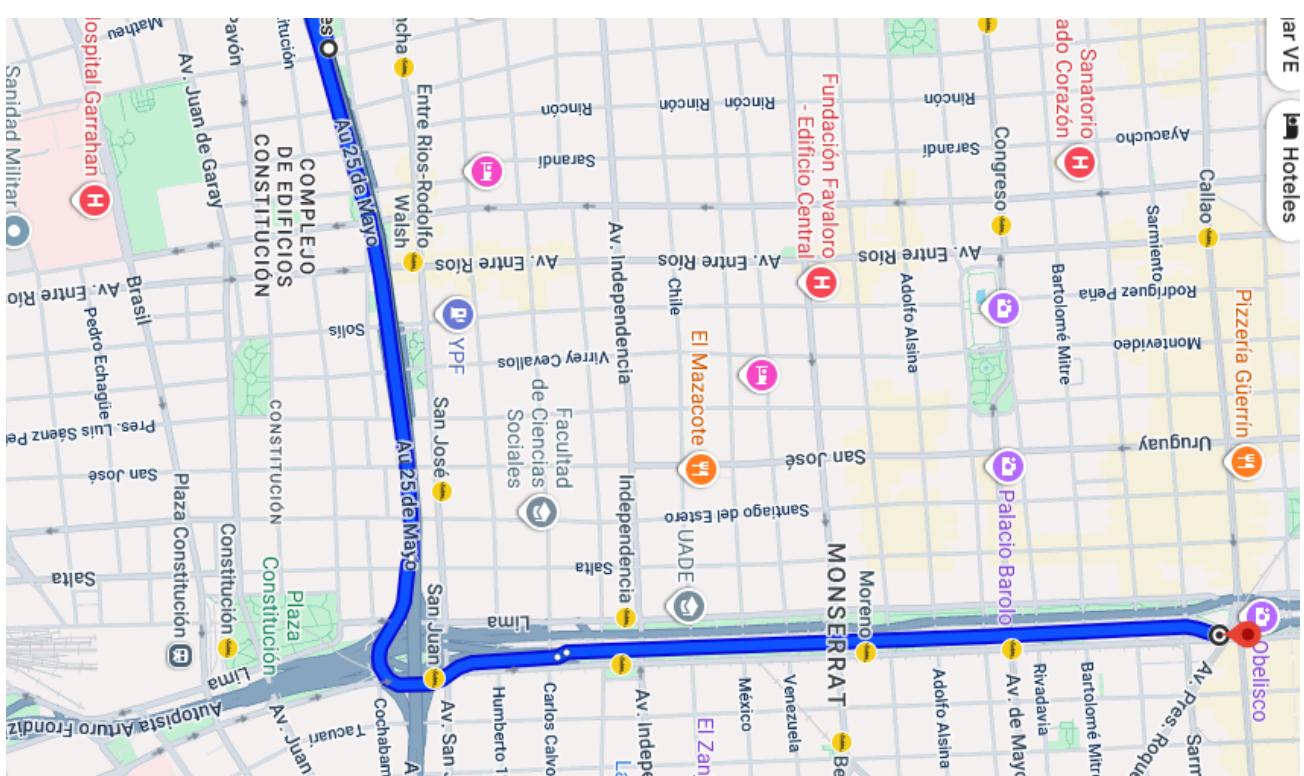
3.1.2. Estrategia de abordaje

Para abordar la hipótesis, decidimos separar el análisis en 3 zonas críticas: el empalme inicial (PK 0-3000), la zona de salidas a barrios (PK 3000-6000) y el complejo de bifurcación final (PK 6000-10000).

Hotspot: PK 0-3000 (El "Embudo" Inicial)

Este tramo es el inicio de la autopista. Incluye la conexión con la **Av. 9 de Julio y el Paseo del Bajo**, y las primeras salidas importantes. Finaliza en la Av Boedo.

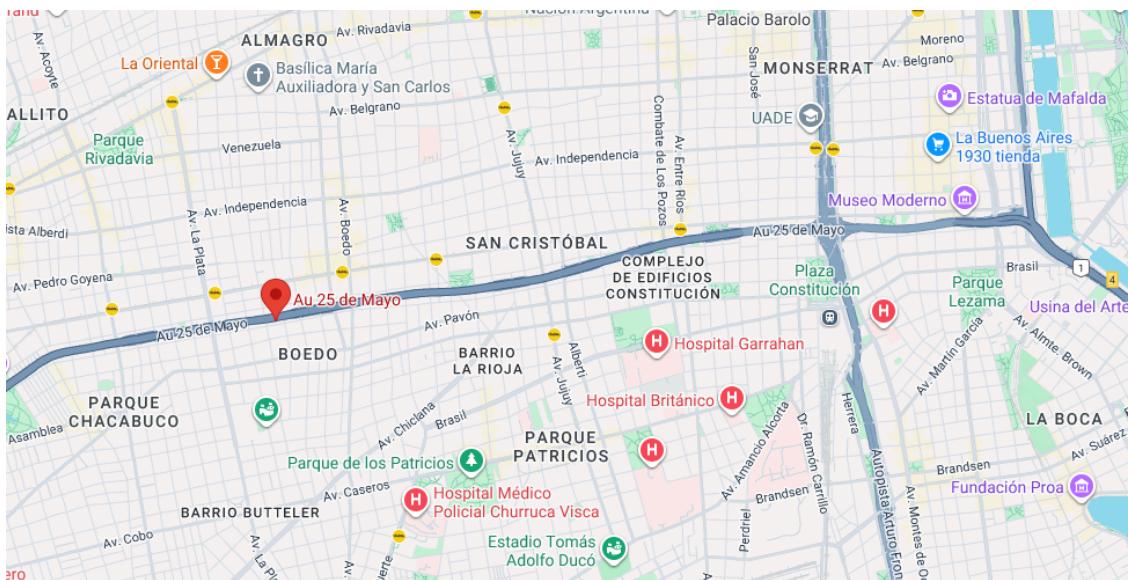
Es una zona de "embudo". El tráfico proveniente de múltiples avenidas debe incorporarse y posicionarse rápidamente.



PK 0-3000 AU 25 DE MAYO

Hotspot: PK 3000-6000 (Salidas de Barrios)

Este tramo cruza varios barrios como los barrios de **Boedo** y **San Cristóbal**, e incluye las salidas de **Av. Boedo** y **Av. Jujuy**.

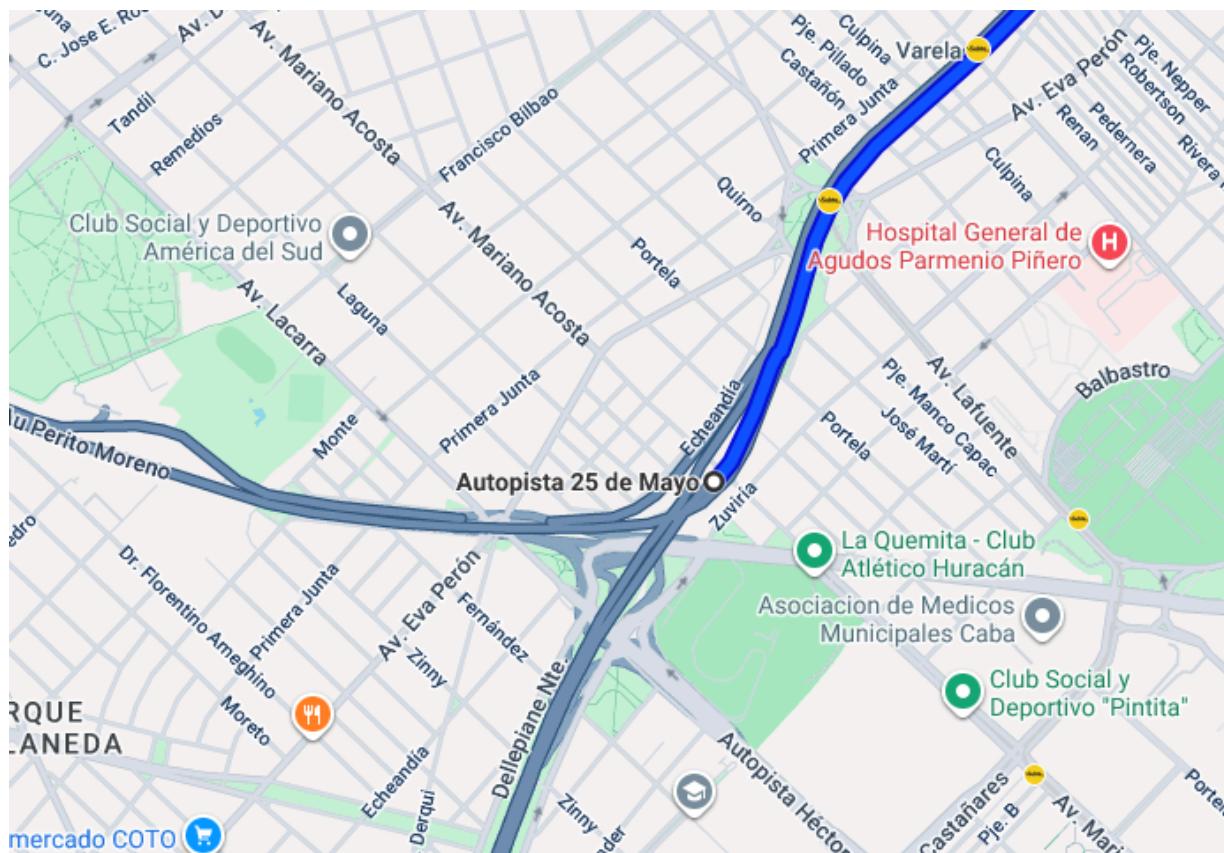


PK 3000-6000 AU 25 DE MAYO

Hotspot: PK 6000-10000 (El "Trébol" de Bifurcación)

Esta es la zona final de la autopista, donde se produce la bifurcación (o "trébol") con la **AU Perito Moreno** y la **AU Dellepiane**).

Esta es, quizás, la zona más compleja. Los conductores deben realizar maniobras de entrecruzamiento en una distancia corta para tomar la rama correcta (hacia **Gral. Paz, Liniers, o hacia Ezeiza**). El alto volumen de tráfico y la complejidad de la decisión de carril la convierten en la zona de mayor siniestralidad sostenida.



PK 6000-10000 AU 25 DE MAYO

3.1.3. Resultados obtenidos y discusión

Siniestros en tramo PK 0-3000: 34 (29.57%)

Siniestros en tramo PK 3000-6000: 36 (31.30%)

Siniestros en tramo PK 6000-10000: 45 (39.13%)

Distribución de Siniestros por PK en AU 25 de Mayo

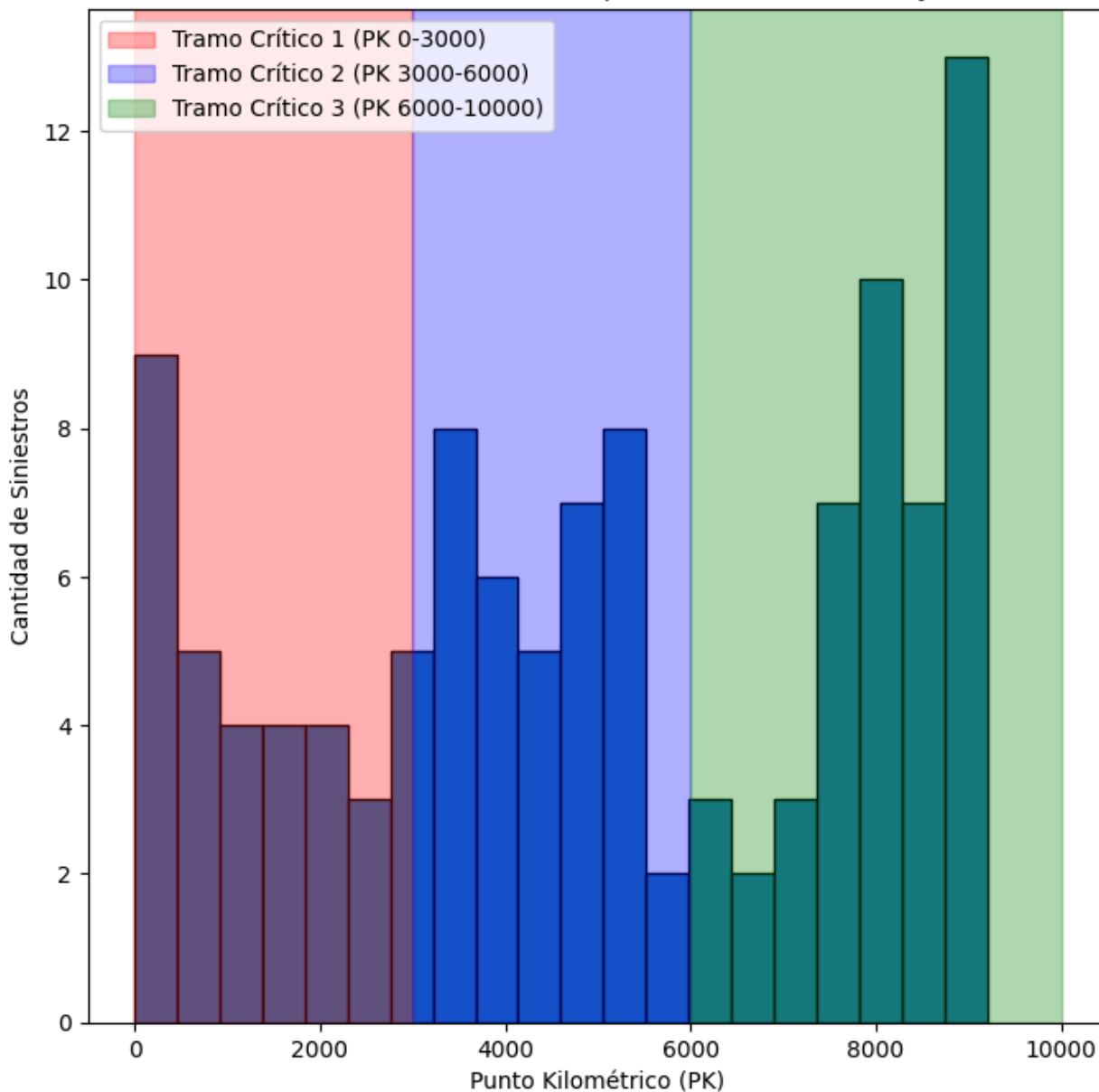


Gráfico de la distribución de siniestros por PK en la AU 25 DE MAYO

Realizamos un Test Chi-Cuadrado para ver si las **frecuencias observadas** difieren de las **esperadas**. Además se cumple que las categorías son mutuamente excluyentes y las observaciones son independientes

Hipótesis nula (H_0): los siniestros se distribuyen uniformemente entre los tres tramos.

Hipótesis alternativa (H_1): los siniestros no se distribuyen uniformemente (hay tramos más peligrosos).

Si el p-valor es menor a 0.05, Rechazamos H_0 por ende hay una concentración significativa en ciertos tramos.

Resultado:

Chi-cuadrado = 1.7913, p-valor = 0.4083

No hay diferencias significativas entre los tramos.

Conclusión: Aunque a simple vista parece que hay más siniestros en el último tramo (6000-10000), el test estadístico indica que esa diferencia no es lo suficientemente grande como para considerarla significativa.

Según los datos disponibles, **no podemos afirmar con evidencia estadística que la siniestralidad sea distinta entre los tres tramos analizados de la AU 25 de Mayo.** Por lo tanto **se rechaza la hipótesis** ya que la distribución de siniestros podría ser más uniforme de lo que parecía inicialmente.

Aunque la idea inicial no se pudo confirmar completamente, lo importante es que **necesitamos analizar a fondo la Autopista 25 de Mayo.**

Deberíamos investigar las posibles causas de este problema. Por ejemplo, el diseño de la autopista: cómo son las curvas, la inclinación de las pendientes, el ancho de los carriles y si se ve bien en distintos tramos. También es clave ver el estado de la infraestructura, como la calidad del asfalto, las señales (si se leen bien y dónde están), la iluminación de noche y si hay barreras de seguridad adecuadas.

Por otro lado, tenemos que prestar atención a los conductores y su comportamiento. Esto incluye estudiar si se respetan los límites de velocidad, si usan el celular al volante, el cansancio, la agresividad al manejar y la educación vial. Las condiciones del tiempo también son un factor importante a analizar, pero del lado de la infraestructura.

3.2. Hipótesis 2

3.2.1. Definición de la hipótesis

La tasa de siniestros por hora es significativamente mayor durante las horas pico (7-10, 17-20) que durante las horas no pico, ya que hay mayor cantidad de vehículos circulando y los conductores están más apurados/estresados por desplazamientos laborales o escolares.

Creemos que si logramos detectar en qué franja horaria hubo más siniestros, se pueden tomar medidas preventivas específicas y optimizar la asignación de recursos. Además nos pareció interesante analizar la influencia de las horas pico en las autopistas.

3.2.2. Estrategia de abordaje

Primero, realizamos un histograma para obtener una idea visual de cómo se distribuyen los siniestros a lo largo del día.

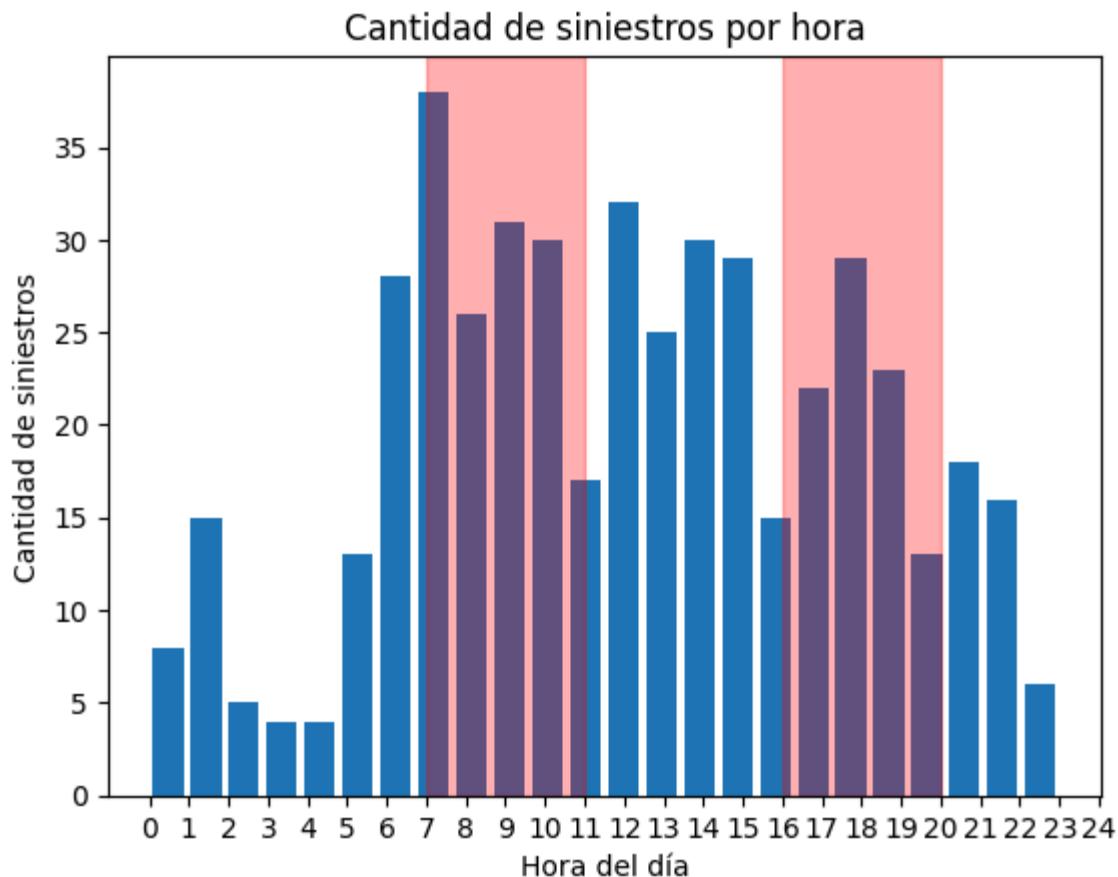


Gráfico de barras de la cantidad de siniestros por hora del día

Resulta interesante ver este histograma, ya que podemos ver en el que a la hora de comenzar el día (7 hs - 10 hs) el pico de siniestros es más alto que a la tarde. También podemos ver que hay un bache a la hora del almuerzo (13 hs). Luego las horas nocturnas parecen “seguras”.

Creamos una columna para detectar cuando una observación ocurrió en hora pico o no.

Hora pico: 244 siniestros

No es hora pico: 233 siniestros



Gráfico de barras de la cantidad de siniestros en hora pico y no pico

Observamos rápidamente que definitivamente ocurren más siniestros en hora pico que en otra franja horaria del día, pero ¿Qué tan significativa es esta diferencia?

Para probar nuestra hipótesis, creamos dos grupos: siniestros hora pico y siniestros no hora pico.

Como no podemos asumir normalidad, pero probamos homocedasticidad con Levene, decidimos utilizar un test Mann Whitney.

Planteamos:

H₀: La mediana de la tasa de siniestros por hora es igual en horas pico y no pico.

H₁: La mediana de la tasa de siniestros por hora es **mayor** en horas pico que en horas no pico

3.2.3. Resultados obtenidos y discusión

Como resultado, obtuvimos:

Mann-Whitney: stat=99.0, p=0.047399764404639215

Dado que el **p-valor (0.0474)** es menor que nuestro nivel de significancia, rechazamos la hipótesis nula (**H₀**).

Esto significa que existe evidencia estadística suficiente para afirmar que la **tasa de siniestros por hora es significativamente más alta** durante las horas pico que durante las horas no pico.

Esta conclusión permite reforzar controles en los horarios críticos, por ejemplo, controles de velocidad o mejoras en la señalización. Además, permite optimizar recursos como organismos de emergencia, ambulancias o personal de tránsito para garantizar una mejor cobertura en las franjas horarias de mayor riesgo.

3.3. Hipótesis 3

3.3.1. Definición de la hipótesis

Hipótesis: El **Índice de Peligrosidad (siniestros/km totales)** está *inversamente* relacionado con la segregación del tráfico y *directamente* relacionado con la complejidad de la vía y su convergencia con otras autopistas. Se conjeta que las autopistas con tráfico segregado (Quiere decir que tiene limitaciones o carriles exclusivos) tendrán el índice más bajo, mientras que las autopistas "conectoras" de alta complejidad tendrán el índice más alto.

Nos motiva plantear un ranking justo de peligrosidad. Este permite tomar medidas (reasignar recursos de patrullaje y mantenimiento) como también tomar dimensión real del peligro en determinada autopista. Un siniestro en una autopista de aproximadamente 4 km (ej. AU Frondizi) no "pesa" lo mismo que un siniestro en una de 10 km (ej. AU Illia). Las autopistas cortas suelen actuar como "conectores" o "distribuidores" (ej: Dellepiane conecta Gral. Paz y AU Campora con la AU 25 de Mayo). Esta función de "conector" obliga a los conductores a realizar maniobras complejas (múltiples cambios de carril, fusiones de tráfico) en una distancia muy corta, elevando el riesgo por kilómetro.

3.3.2. Estrategia de abordaje

Para empezar miramos en nuestros datos cuántos siniestros tenemos por cada autopista para filtrar aquellas autopistas que cuenten tal vez con pocos siniestros. Definimos que las autopistas observadas van a ser: **AU ILLIA, AU 25 DE MAYO, PASEO DEL BAJO, AV. LUGONES, AU PERITO MORENO, AU DELLEPIANE, AU CAMPORA, AU FRONDIZI Y AV CANTILLO.**

Luego, buscamos cual es la longitud total de cada una de ellas y las anotamos en una tabla como la siguiente

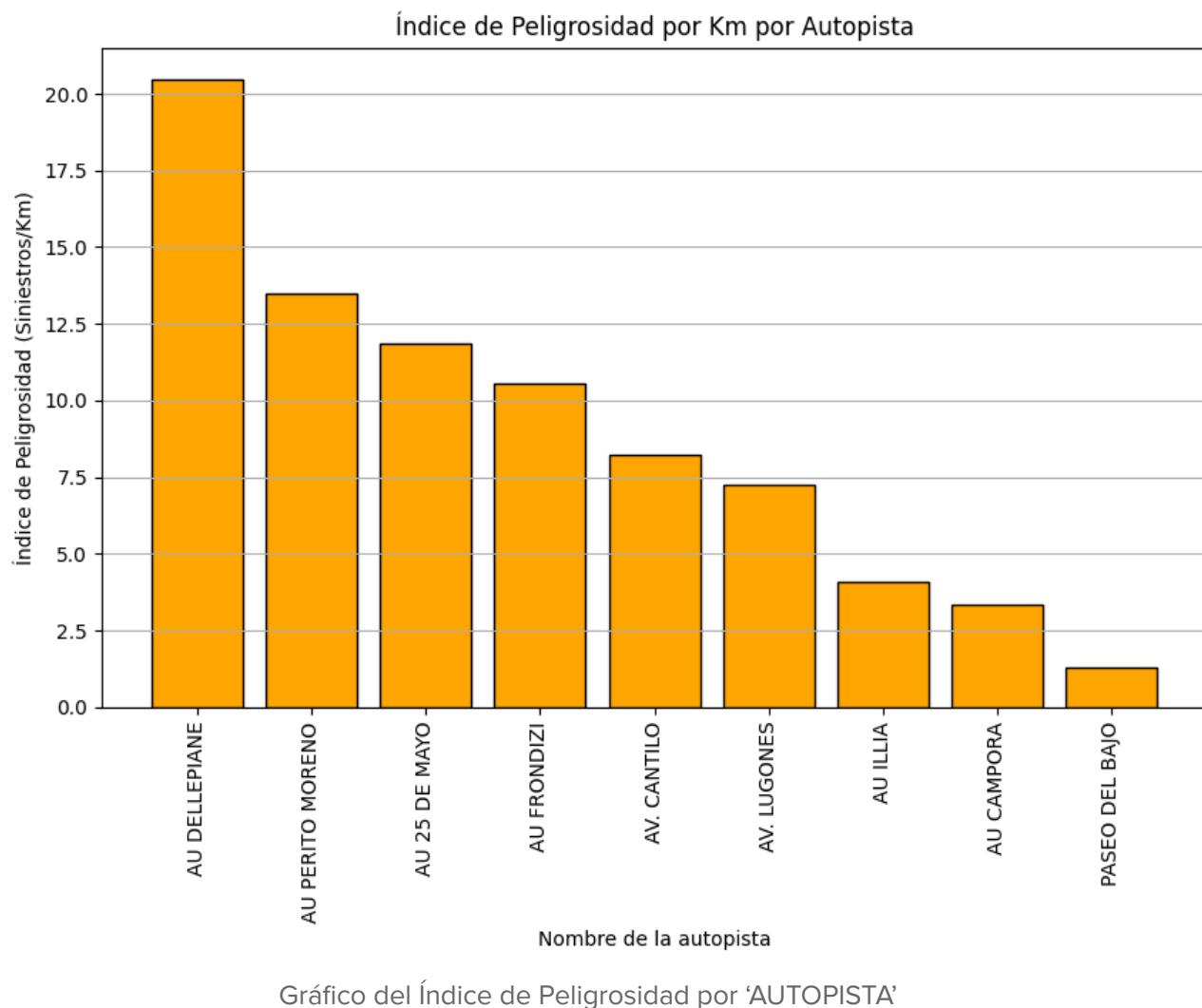
Autopista	Longitud
AU ILLIA	10.5 km
AU 25 DE MAYO	10 km
PASEO DEL BAJO	7.1 km
AV. LUGONES	6.2 km
AU PERITO MORENO	6.0 km
AU DELLEPIANE	4.2 km
AU CAMPORA	3.6 km

AU FRONDIZI	3.5 km
AV. CANTILLO	3.5 km

Contamos cuántos siniestros ocurrieron en cada una de las autopistas seleccionadas y calculamos el **Índice de Peligrosidad** para cada una de las mismas, que es el resultado de dividir la cantidad de siniestros por la longitud de la propia autopista. Con este índice (basado en los datos) vamos a medir que tan peligrosa puede resultar una autopista.

AUTOPISTA	IDP
AU DELLEPIANE	20.476190
AU PERITO MORENO	13.500000
AU 25 DE MAYO	11.836735
AU FRONDIZI	10.571429
AV. CANTILLO	8.205128
AV. LUGONES	7.258065
AU ILLIA	4.095238
AU CAMPORA	3.333333
PASEO DEL BAJO	1.267606

Para visualizar mejor el resultado, utilizamos un gráfico de barras:



3.3.3. Resultados obtenidos y discusión

Con estos resultados se verifica que:

La AU Dellepiane es un caso atípico extremo, siendo **casi el doble** de peligrosa por kilómetro que la autopista que le sigue en el ranking (AU Perito Moreno) y **19 veces** más peligrosa que el Paseo del Bajo.

Por otro lado, el diseño de la infraestructura parecería funcionar, en determinados casos. **El Paseo del Bajo** (diseñado para separar camiones de autos) es, por un margen enorme, la vía más segura de toda la red. Al mismo tiempo, las autopistas que actúan como "embudos" o "conectores" (Dellepiane, Perito Moreno), que obligan a maniobras complejas de cambio de carril y fusión de tráfico, son las que lideran el ranking de peligrosidad.

Por otro lado la AU 25 de Mayo sigue siendo protagonista de nuestro informe, ya que es una de las autopistas más largas de CABA (por lo tanto no puede definirse sólo como Conectora o de vía rápida y segregada) y a la vez está tercera en el ranking de peligrosidad que armamos. Casualmente en la hipótesis 1 vimos que ocurren más accidentes en el último tramo (conector) que conectaba otras autopistas (entre ellas la AU DELLEPIANE, primera en el ranking)

Con los resultados obtenidos y las conclusiones a las que llegamos podemos decir que se **comprueba la hipótesis inicial**

3.4. Hipótesis 4:

3.4.1. Definición de la hipótesis

En los siniestros con motos involucradas tiende a haber más heridos que con otro tipo de vehículos.

A diferencia de otros vehículos, las motos no tienen una estructura de protección alrededor del conductor, por lo que se encuentran más expuestos. En caso de choque o caída, el impacto es directamente al cuerpo y aumenta la probabilidad de sufrir lesiones.

Saber si las motos están asociadas a un mayor número de heridos es de vital importancia para el enfoque de las políticas públicas en la prevención y la educación vial a motociclistas, ya sea promoviendo el uso de casco o mejorando la infraestructura (carriles exclusivos, mejor señalización).

3.4.2. Estrategia de abordaje

Como estrategia de abordaje, dividimos los datos en dos grupos: moto y no moto.

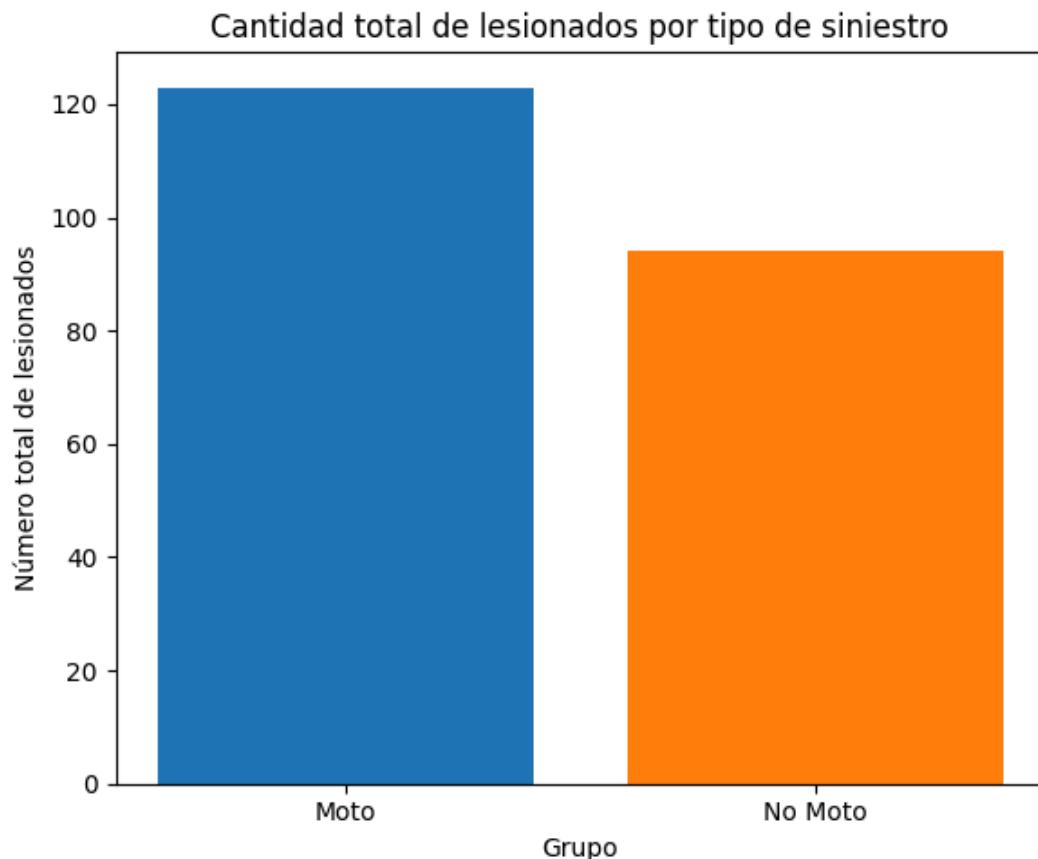


Gráfico de la cantidad de lesionados en siniestros que involucran motos

Así podemos ver la cantidad de lesionados en cada grupo, con una cantidad claramente superior en el grupo moto.

Planteamos:

H0: La cantidad de lesionados en siniestros con motos es menor o igual que en los siniestros sin motos.

H1: La cantidad de lesionados en siniestros con motos es mayor que en los siniestros sin motos.

Como la distribución de los datos no es normal, pero se cumple el supuesto de homocedasticidad hicimos un U-Test.

3.4.3. Resultados obtenidos y discusión

Como resultado, obtuvimos:

Estadístico U: 38385.0 | p-valor: 1.0243526689761333e-25

Conclusión: se rechaza H0, **hay evidencia estadísticamente significativa para afirmar que los siniestros con motos involucradas tienen una mayor cantidad de heridos.**

Con esta conclusión, decidimos analizar en qué zonas ocurrían estos siniestros. Según las fuentes que recolectamos, uno de los lugares con mayor cantidad de accidentes (y lesionados) con motos involucradas es sobre la autopista 25 de Mayo en el kilómetro 4.5, ubicado entre las calles Treinta y Tres Orientales y Bocayuva, en Boedo, ya que en la zona mencionada, hay una dársena de emergencia que es utilizada de forma indebida por motociclistas en su mayoría. Esto permite generar una discusión sobre el enfoque de las propuestas de educación vial para motociclistas y las posibles mejoras en la infraestructura, como la implementación de carriles exclusivos, con el objetivo de prevenir futuros accidentes.

3.5. Hipótesis 5:

3.5.1. Definición de la hipótesis

La frecuencia de los siniestros viales varía significativamente según la hora del día y el día de la semana

Variable	Tipo	Rol en el análisis
FECHA	temporal	Extraer día de la semana
HORA	Numérica discreta	Agrupar en franjas horarias
AUTOPISTA	Categórica	Analizar las diferencias entre las autopistas

El propósito de esta hipótesis es identificar momentos y lugares de mayor riesgo, integrando las variables de tiempo (día y hora) con la localización (autopista).

Comprender estas variaciones permite orientar **acciones preventivas** como mayor presencia policial, mejoras en iluminación o señalización, y campañas específicas en los horarios más críticos.

En términos de gestión pública, este tipo de análisis aporta información estratégica para **distribuir recursos viales** y diseñar políticas adaptadas al comportamiento temporal del riesgo.

3.5.2. Estrategia de abordaje

Para empezar decidimos discretizar la columna HORA y construimos la columna **franja horaria** con cuatro categorías:

- Madrugada: De 1 a 6
- Mañana: De 6 a 12
- Tarde: De 12 a 20
- Noche: De 20 a 23

Luego calculamos la frecuencia de siniestros por franja horaria y día de la semana, extraído de la columna FECHA. Lo visualizamos en el siguiente Heatmap

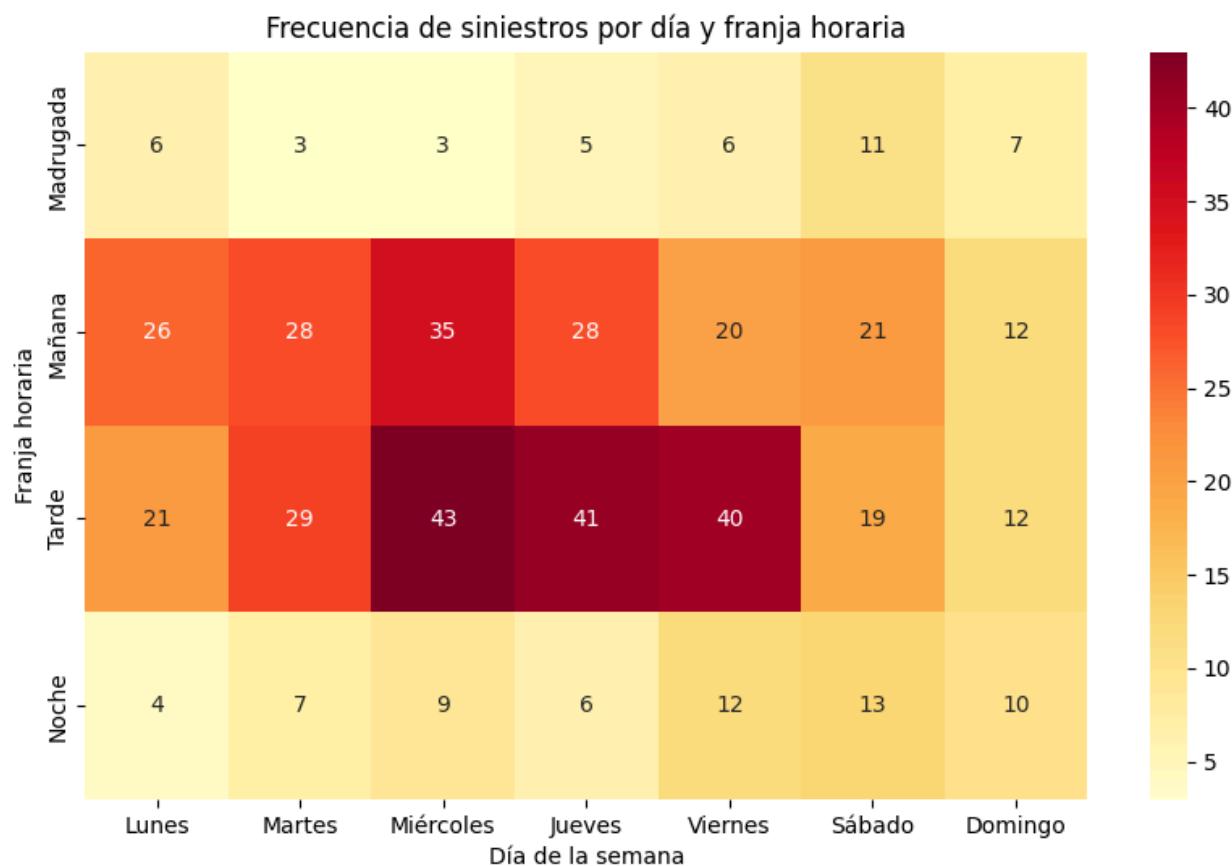


Gráfico de la frecuencia de siniestros por día y franja horaria

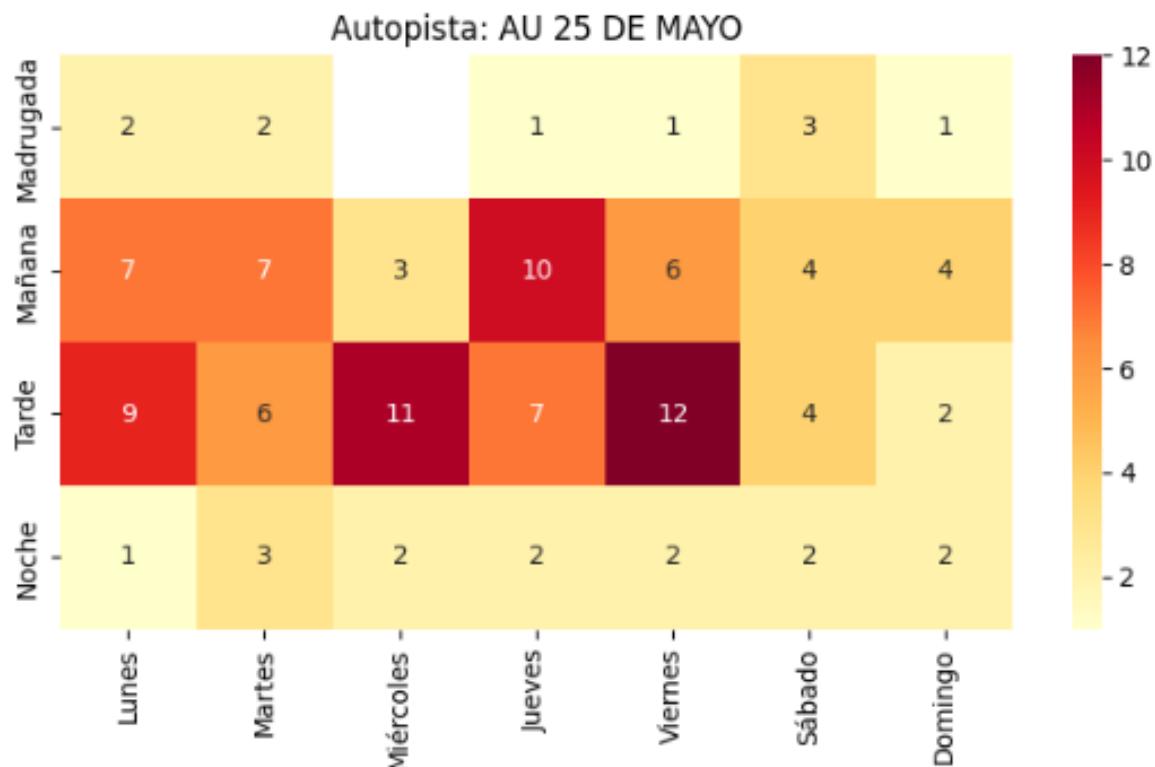
Vemos que hay una alta frecuencia de accidentes tanto a la mañana como a la tarde, lo que es consistente con nuestro análisis de mayor frecuencia de accidentes durante las horas pico. Es

interesante ver que en los días de mitad de semana (Miércoles, Jueves y Viernes) ocurrieron más accidentes en el 2022.

Para hacer un mejor análisis sobre lo que está pasando en las autopistas de CABA extendimos la tabla para incluir la variable **AUTOPISTA** (Día x Franja x Autopista). Cabe aclarar que vamos a analizar solo las primeras 5 autopistas con más siniestros registrados que son las siguientes:

Nombre de la autopista	Total de siniestros registrados
25 DE MAYO	116
AU DELLEPIANE	86
AU PERITO MORENO	81
AV. LUGONES	45
AU ILLIA	43

Al incorporar la variable **AUTOPISTA**, para cada una se observa que

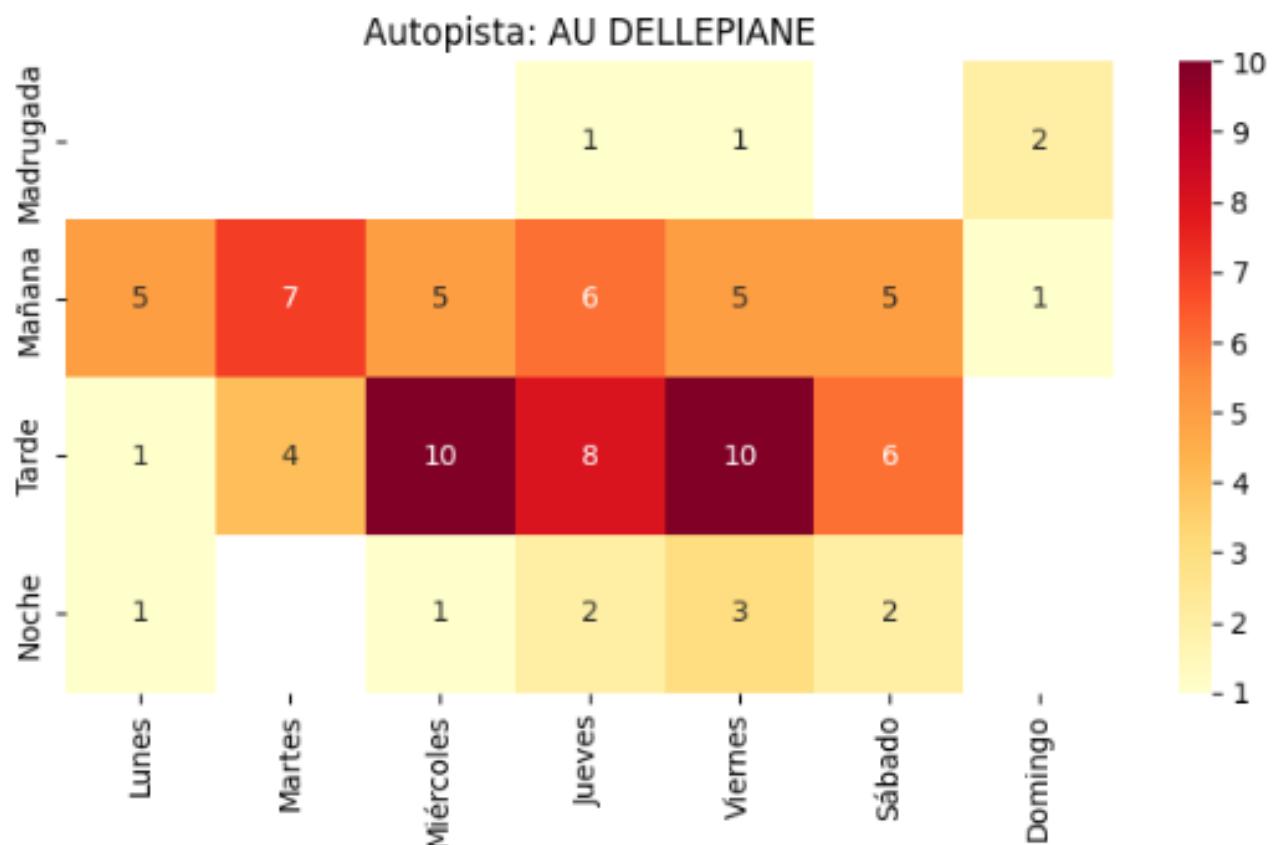


AU 25 de Mayo es la autopista con mayor actividad y un patrón marcadamente diurno

- Se observa una **alta concentración de siniestros en la mañana y la tarde**, especialmente entre **miércoles y viernes**.
- Los valores alcanzan picos de **10 a 12 siniestros** en la franja diurna.
- La Madrugada y la Noche presentan muy poca actividad

Conclusión: AU 25 de Mayo concentra sus siniestros en horarios laborales y días hábiles, coherente con su rol como corredor principal de ingreso/egreso al centro porteño.

Analicemos la **Autopista Dellepiane** donde enlaza con la Autopista Buenos Aires-La Plata y el Paseo del Bajo.

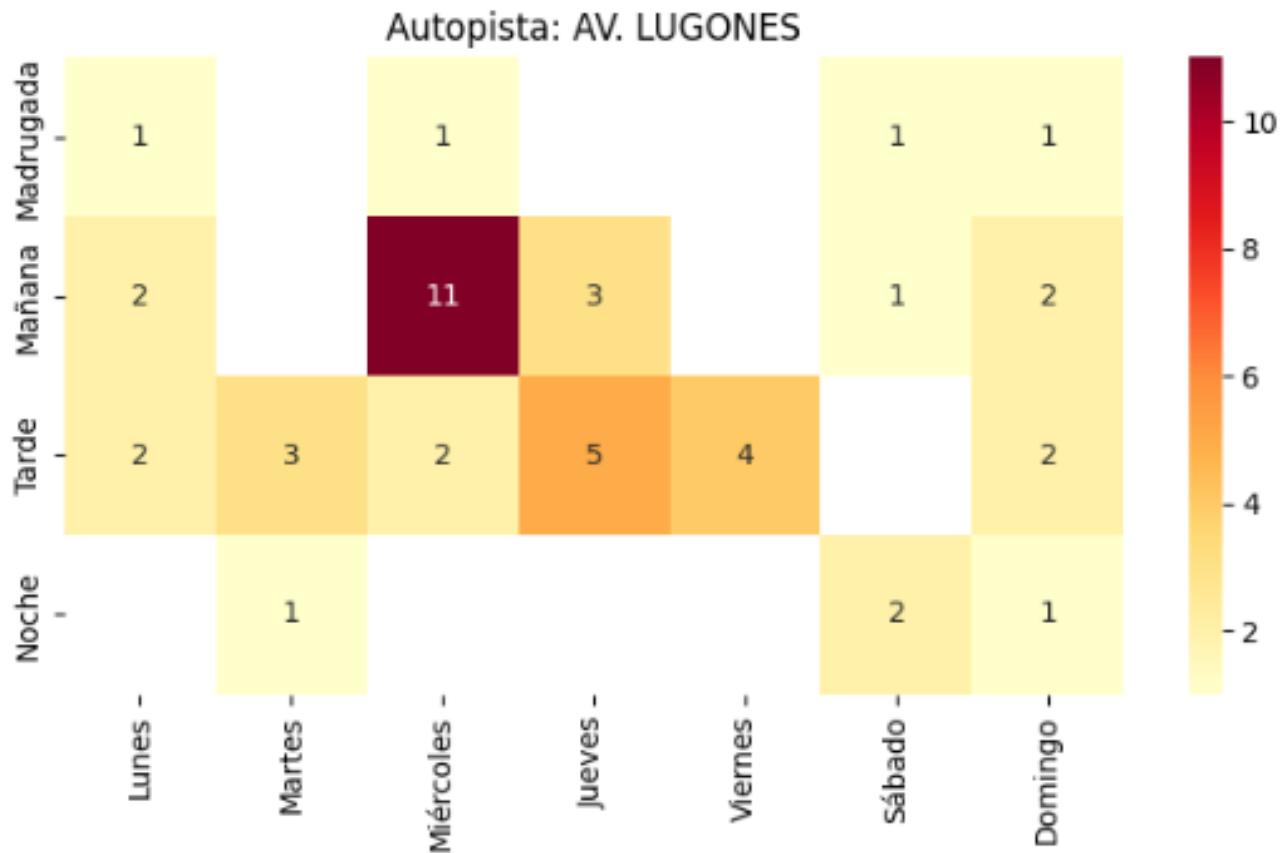


AU Dellepiane tiene patrones muy marcados en la franja de la tarde.

- La tarde es la franja claramente más riesgosa, con picos de **8 a 10 siniestros** entre miércoles y viernes.
- La mañana también muestra actividad, pero de menor magnitud.
- Casi nula actividad nocturna.

La autopista Dellepiane tiene una fuerte concentración de siniestros en la franja tarde, posiblemente por el tráfico desde/hacia el Aeropuerto Internacional Ezeiza y alto flujo de transporte de carga.

Luego, nos pareció relevante analizar la autopista Lugones:



Notamos un comportamiento extraño los miércoles a la mañana en esa autopista, ya que vemos un pico muy fuerte en esta franja horaria. Analizando un poco más, solo podemos identificar que todos ocurrieron por la banda descendente de la autopista.

Continuamos realizando algunos test no paramétricos:

¿La distribución de siniestros es independiente del día y la hora? o, dicho de otra forma, ¿ocurren siniestros en los días y horas con la misma probabilidad?

Para contestar esta pregunta, utilizamos el Test Chi-Cuadrado que evalúa si existe asociación entre dos variables categóricas: **Día de la semana y Franja Horaria**

Este test va a evaluar si la distribución de siniestros es aleatoria o si existen diferencias significativas entre los distintos momentos del tiempo, es decir, si existen días y horarios más riesgosos.

3.5.3. Resultados obtenidos y discusión

Resultados del test:

Chi² = 38.92, p-valor = 0.0029

Como **p < 0.05**, se rechaza la hipótesis nula y **se concluye que la frecuencia de siniestros varía significativamente según el día y la hora**. Coincide con lo que veíamos anteriormente en los gráficos.

Los resultados indican que la frecuencia de siniestros viales varía de manera significativa según el momento del día y el día de la semana ($p = 0.0024$) En términos prácticos, esto sugiere que hay momentos con mayor cantidad de siniestros.

El patrón identificado coincide con la dinámica urbana de CABA: los picos de ocurrencia se concentran en horas pico y fines de semana, cuando aumenta la exposición vehicular o se intensifican conductas de riesgo (fatiga, velocidad, consumo de alcohol).

Conclusión: la hipótesis fue confirmada.

Se comprobó una variación significativa en la frecuencia de los siniestros según el momento del tiempo.

3.6. Hipótesis 6:

3.6.1. Definición de la hipótesis

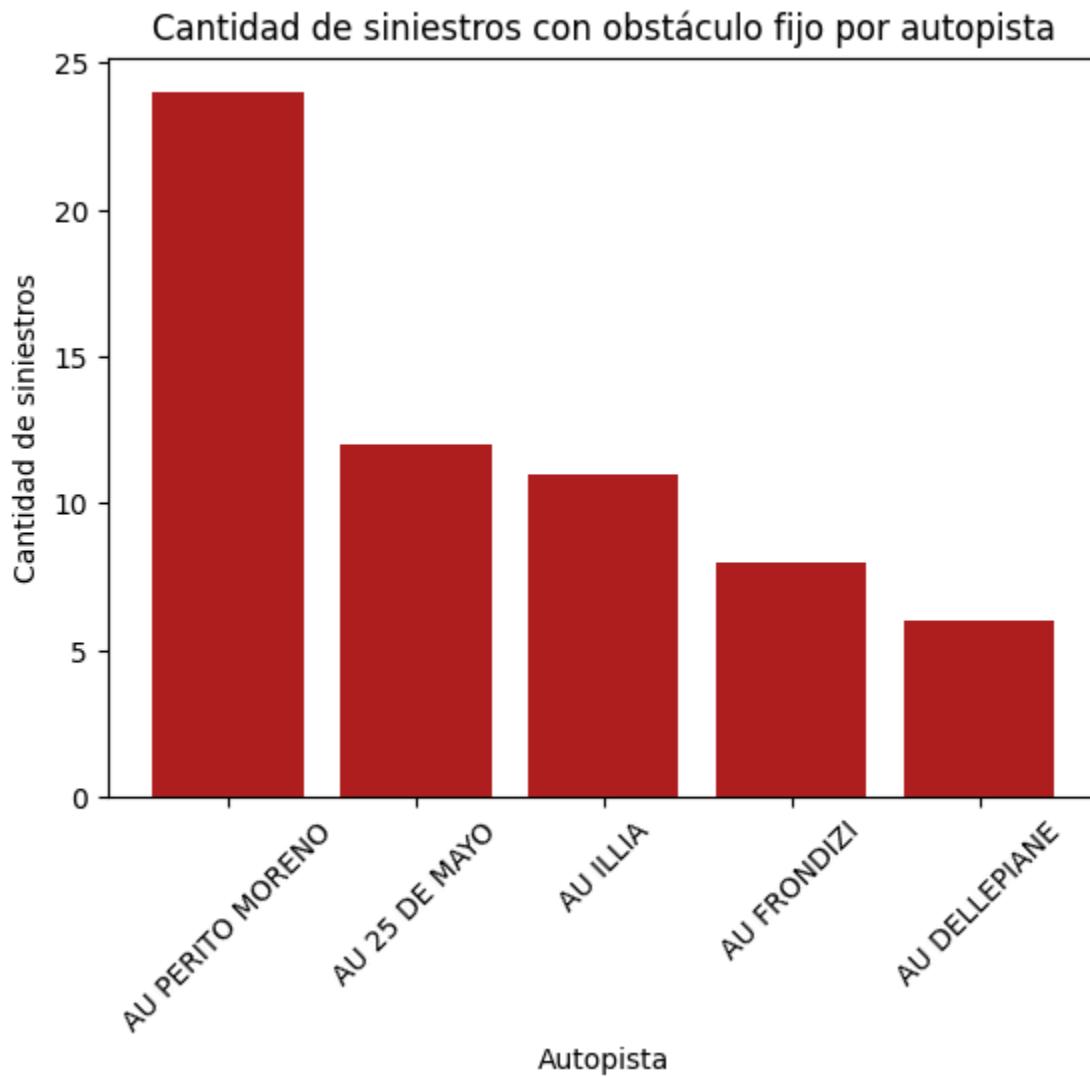
La AU Perito Moreno presenta un riesgo significativamente mayor de colisiones con obstáculos fijos, ya que hay una frecuencia relativamente alta de este tipo de colisiones comparado con otras autopistas.

Queremos identificar los principales puntos de riesgo y posibles causas para esta cantidad de siniestros.

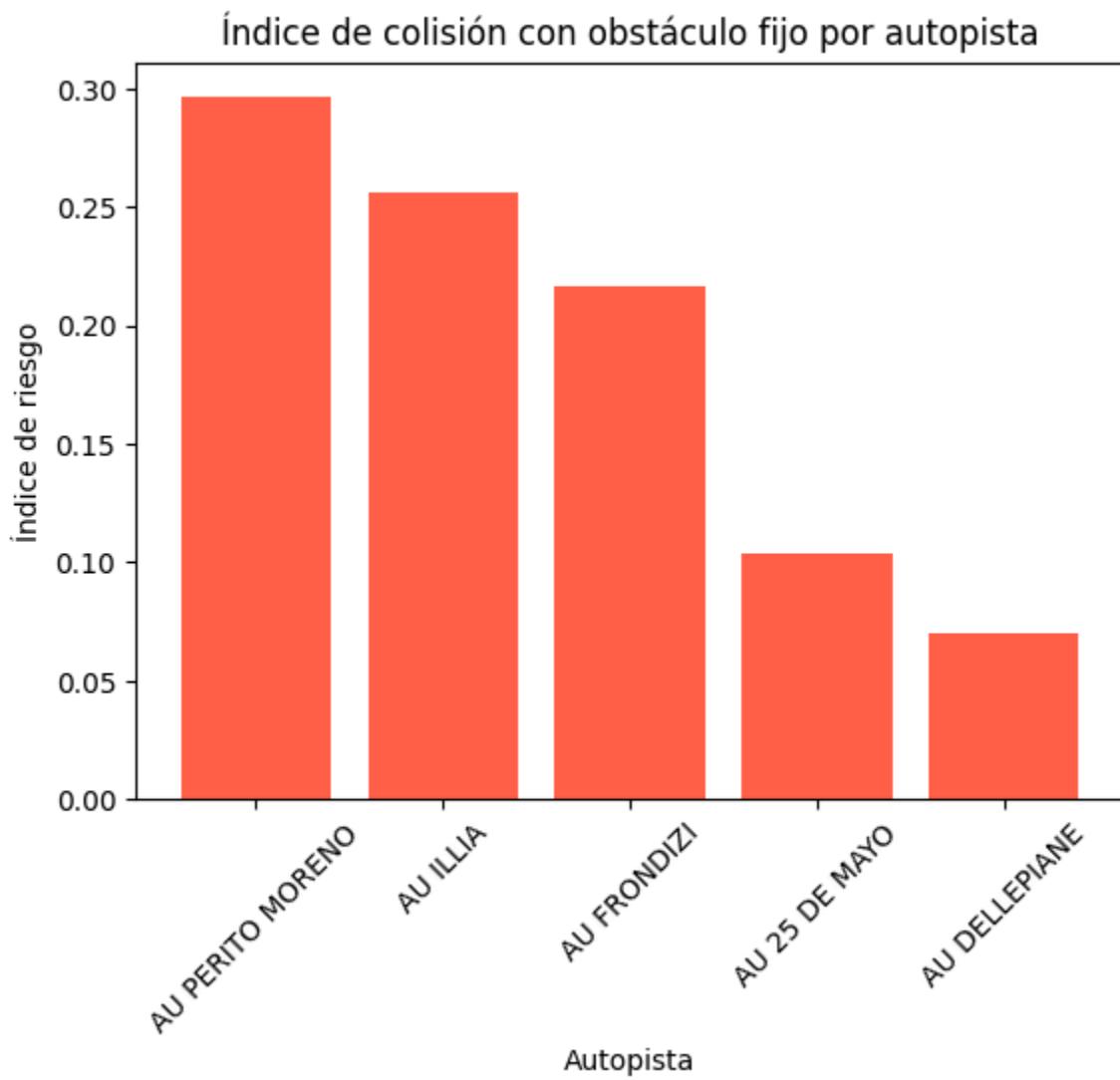
3.6.2. Estrategia de abordaje

Como estrategia de abordaje filtramos los datos y nos quedamos con los siniestros con obstáculo fijo en las autopistas principales para nuestro análisis: AU PERITO MORENO, AU 25 DE MAYO, AU ILLIA, AU FRONDIZI, AU DELLEPIANE.

Como podemos observar en el siguiente gráfico, la autopista Perito Moreno tiene prácticamente el doble de siniestros con obstáculo fijo que el resto de las autopistas.



Como la cantidad de siniestros difiere significativamente entre autopistas, decidimos hacer un índice para realizar una comparación equitativa.



Sigue existiendo una diferencia entre la autopista Perito Moreno y el resto de autopistas, pero no tan grande como parecía inicialmente.

Para continuar el análisis, buscamos en qué punto kilométrico de la autopista Perito Moreno sucedían estos siniestros. Identificamos como problemático el punto entre los kilómetros 5.3 y 5.5.

3.6.3. Resultados obtenidos y discusión

En el tramo anteriormente mencionado, se encuentra el peaje Parque Avellaneda. Lo que sugiere que la reducción de velocidad, la necesidad de alinearse a carriles específicos y la congestión localizada que suele presentarse en este tipo de puntos puede aumentar la probabilidad de maniobras bruscas o pérdidas momentáneas de control, lo que favorece impactos contra elementos fijos de la infraestructura.



Peaje Parque Avellaneda

Analizando la posibilidad de realizar un test formal para contrastar la hipótesis, la cantidad de datos que tenemos disponibles no sería suficiente para aportar evidencia estadísticamente significativa, aunque el patrón identificado sea consistente. Optamos por fundamentar nuestras conclusiones con la evidencia exploratoria y las fuentes encontradas.

Existe un mayor riesgo relativo de colisiones con obstáculos fijos en la AU PERITO MORENO, asociado a factores locales. En particular, al tramo donde se ubica el peaje.

Deberíamos considerar este tramo como prioritario para un análisis más detallado, y eventuales intervenciones orientadas a reducir el riesgo.

4. Conclusiones

El análisis realizado sobre la siniestralidad vial en las autopistas de la Ciudad Autónoma de Buenos Aires durante el 2022 permitió integrar las herramientas que aprendimos de Ciencia de Datos con una problemática urbana concreta. A lo largo del trabajo, identificamos patrones espaciales y temporales que ayudan a comprender mejor la dinámica de los siniestros y que pueden orientar la toma de decisiones con respecto a la prevención, control y mantenimiento.

En primer lugar, la exploración inicial nos mostró que algunas autopistas concentran más siniestros simplemente porque tienen mayor longitud. Por esta razón el uso de índices como el índice de peligrosidad resultó fundamental para realizar comparaciones justas.

Para la hipótesis 1, el análisis estadístico no respaldó una diferencia significativa entre los tramos que definimos. Pero esto no implica ausencia de riesgo, solo que la variación detectada no fue lo suficientemente grande para considerarse estadísticamente distinta.

En el caso de la hipótesis 2, si se encontró evidencia estadística sólida. La frecuencia de siniestros durante las horas pico es mayor que en el resto del día. Esta conclusión permite orientar medidas como el refuerzo de controles, ajustes en señalización y una mejor distribución de recursos de emergencia en las franjas de mayor tránsito.

En cuanto a la hipótesis 3, el análisis del índice de peligrosidad mostró diferencias claras entre autopistas, respaldando la idea de que el diseño de la infraestructura y la función de cada vía influyen en el riesgo. Las autopistas que actúan como conectores (Dellepiane, Perito Moreno) resultaron considerablemente más peligrosas por kilómetro. Mientras que, el Paseo del Bajo, diseñado para separar flujos de tránsito, presentó el índice más bajo.

La hipótesis 4 también fue confirmada. Los siniestros con motos involucradas presentaron una cantidad significativamente mayor de lesionados en comparación con otros vehículos. Este resultado, refuerza la importancia de implementar políticas específicas orientadas a motociclistas, tanto en educación vial como en el diseño de carriles exclusivos o mejoras en zonas conflictivas.

La hipótesis 5 mostró que la variabilidad temporal de los siniestros no es aleatoria: el día de la semana y la franja horaria influyen fuertemente en la frecuencia de incidentes. Las mañanas y tardes laborales concentran gran parte del riesgo, mientras que la madrugada presenta niveles notablemente inferiores. Este análisis está directamente relacionado con nuestra hipótesis 2, de mayor frecuencia de siniestros durante horas pico.

Por último, la hipótesis 6 permitió identificar un punto crítico en la autopista Perito Moreno. Aunque la diferencia no es grande por la cantidad de datos disponibles, si pudimos observar una concentración clara de siniestros en el tramo del peaje Parque Avellaneda. Este análisis, marca este segmento como un área prioritaria para futuras intervenciones.

En conjunto con las herramientas aprendidas, el trabajo nos permitió identificar zonas críticas, horarios de riesgo y factores asociados a la gravedad de los siniestros, que constituyen un insumo valioso para diseñar políticas públicas que reduzcan lesiones, muertes y congestión.

Referencias

Gobierno de la Ciudad de Buenos Aires. (s. f.). *Observatorio de Movilidad y Seguridad Vial*.

https://buenosaires.gob.ar/movilidad/plan-de-seguridad-vial/observatorio-de-movilidad-y-seguridad-vial_Buenos_Aires

Gobierno de la Ciudad de Buenos Aires. (s. f.). *Informes estadísticos y mapas*.

<https://buenosaires.gob.ar/movilidad/plan-de-seguridad-vial/informesestadisticosymapas>

Perfil. (2024, octubre). *Triple choque en autopista Dellepiane: tránsito colapsado y al menos 15 heridos*.

<https://www.perfil.com/noticias/actualidad/triple-choque-en-autopista-dellepiane-transito-colapsado-y-al-menos-15-heridos.phtml>

Gobierno de la Ciudad de Buenos Aires. (2021, 3 de marzo). *Decreto N° 81/21: Anexo. Boletín Oficial de la Ciudad de Buenos Aires*.

<https://documentosboletinoficial.buenosaires.gob.ar/publico/PE-DEC-AJG-AJG-81-21-ANX.pdf>
[Documentos Boletín Oficial+1](#)

Autopistas Urbanas S.A. (AUSA). (s. f.). *Autopistas*. <https://www.ausa.com.ar/Autopistas.html>

La Nación. (2024, 12 de mayo). *En Boedo y San Telmo, los dos puntos con mayor cantidad de siniestros viales a lo largo de las autopistas*.

<https://www.lanacion.com.ar/buenos-aires/en-boedo-y-san-telmo-los-dos-puntos-con-mayor-cantidad-de-siniestros-viales-a-lo-largo-de-las-nid12052024/>