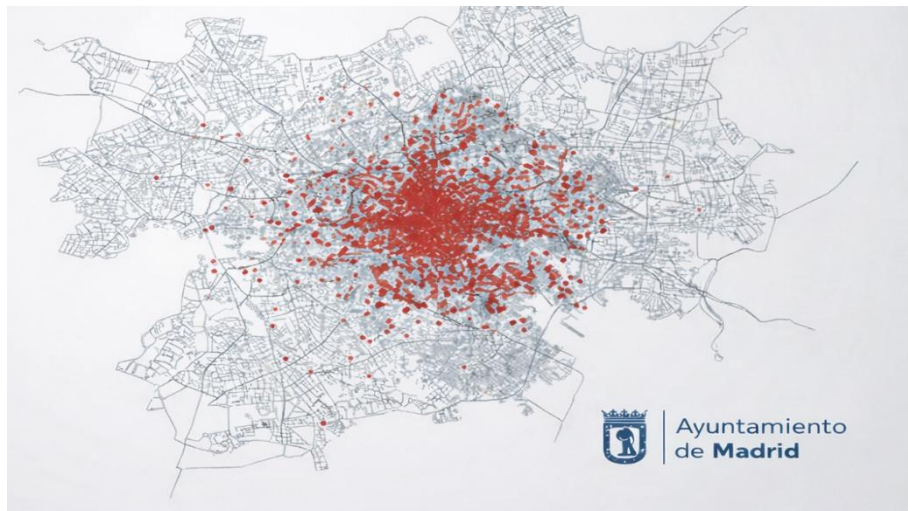


The Bridge School

Proyecto EDA: Accidentalidad en Madrid

Años 2020-2024



Autores:

César Sánchez Parra

Lucía Fuentes González

Adrián Quindimil Rengel

Contenido

1. Introducción	2
1.1. Objetivos del análisis	2
1.2. Hipótesis de partida	2
2. Descripción, preprocesado y limpieza de los datos	3
3. Análisis Exploratorio de Datos	4
3.1 Análisis univariante	4
3.1.1 Variables temporales	4
3.1.2 Variables geográficas	5
3.1.3 Variables demográficas	6
3.1.5 Gravedad del accidente	9
3.2 Análisis bivalente	9
3.2.1 Gravedad del accidente	10
3.2.2 Consumo de alcohol del conductor	12
3.2.3 Análisis geográfico	14
3.3 Análisis multivariante	14
3.3.1 Variables geográficas	14
3.3.2 Variables demográficas	16
3.3.3 Gravedad accidente	18
4. Conclusiones	20
4.1 Verificación de hipótesis y principales hallazgos (año 2024)	20
4.2 Conclusiones generales de 2020 a 2024	21
4.3 Recomendaciones de prevención	22
5. Anexo	22

MEMORIA TÉCNICA: ACCIDENTES INTRAURBANOS EN MADRID

1. Introducción

Desde el año 2010, la Policía Municipal de Madrid ha registrado información detallada sobre todas las personas implicadas en los accidentes intraurbanos de la ciudad. Con el objetivo de reducir tanto el número de accidentes como la gravedad de los mismos, el Ayuntamiento de Madrid ha encargado a un equipo de científicos de datos (nosotros) la elaboración de un informe de análisis de la siniestralidad urbana entre los años 2020 y 2024, con el fin de orientar la implementación progresiva y sistemática de medidas de seguridad vial.

En este informe se presentan las hipótesis de partida, el análisis exploratorio de los datos y las conclusiones derivadas de los *datasets* correspondientes a los últimos cinco años, proporcionando una visión integral de los factores asociados a la siniestralidad y su evolución temporal, geográfica y demográfica.

1.1. Objetivos del análisis

El **objetivo general** de este EDA es determinar qué variables influyen de manera más significativa en la siniestralidad y la lesividad de los accidentes intraurbanos en la ciudad de Madrid, así como analizar la relación entre ambas, con el fin de aportar evidencia que permita reducir la accidentalidad y mejorar la seguridad vial urbana.

Los **objetivos específicos** son:

1. Analizar los accidentes en Madrid desde una perspectiva temporal, geográfica y demográfica, evaluando la influencia de factores asociados y su relación con la lesividad, para identificar periodos, zonas y condiciones de mayor riesgo.
2. Establecer posibles relaciones bivariantes y multivariantes entre variables, para identificar patrones y asociaciones relevantes para la seguridad vial.
3. Proporcionar recomendaciones basadas en los resultados, priorizando intervenciones en los periodos, zonas y situaciones de mayor riesgo.

1.2. Hipótesis de partida

- Los meses, días de la semana y horas con mayor número de accidentes podrían ser enero, los lunes y las 8:00 de la mañana, respectivamente.
- La franja horaria de mayor siniestralidad podría corresponder a la mañana (06:00–12:00).
- Los distritos céntricos de Madrid presentan un mayor número de accidentes en comparación con los distritos periféricos.
- Los jóvenes de 18 a 29 años podrían ser el grupo etario más implicado en accidentes.
- Los hombres presentan mayor frecuencia de accidentes que las mujeres, y la mayoría de los implicados en los siniestros son conductores.
- El turismo es el tipo de vehículo más implicado, y la colisión lateral podría ser el tipo de accidente más frecuente.
- La lesividad de los implicados podría estar estadísticamente relacionada con el rango de edad.

- El tipo de accidente podría estar relacionado con el estado de ebriedad del conductor.
- La mayoría de los accidentes intraurbanos no requieren asistencia sanitaria, indicando que predominan los accidentes leves o sin lesiones visibles.
- Los puntos negros urbanos (tramos concretos de calles y autopistas dentro de la ciudad) podrían concentrar un número desproporcionado de accidentes y deberían ser priorizados para medidas de prevención.

2. Descripción, preprocesado y limpieza de los datos

Todos los *datasets* contienen 19 campos y un número de registros que varía de 32433 (año 2020) y 49340 (año 2024). La descripción, tipo e importancia de cada una de las 19 variables incluidas se detallan en la **Tabla 1**.

Tabla 1. Descripción, tipificación e importancia de las variables. Escala de importancia: 0 = Variable directora, 1 = agrupación importante, 2 = agrupación interesante y 3 = agrupación secundaria.

Columna/variable	Descripción	Tipo final	Imp.	Comentarios
num_expediente	ID de accidente	Cadena	0	Varios registros para el mismo número o ID
fecha	Día del accidente	Fecha	1	Falta convertir
hora	Hora del accidente	Fecha	1	Falta convertir
localizacion	Nombre de la calle	Categórica nominal	2	Calle o cruce
numero	Número o código de la calle	Cadena	3	Mezclas números y cadenas
cod_distrito	Número de distrito	Numérica Ordinal	1	21 números enteros
distrito	Nombre del distrito	Categórica nominal	1	Hay 21 distritos
tipo_accidente	Tipo de accidente	Categórica nominal	1	13 tipos de accidente
estado_meteorológico	Descripción del clima	Categórica nominal	2	6 estados meteorológicos
tipo_vehiculo	Tipo de vehículo	Categórica nominal	2	32 tipos de vehículo
tipo_persona	Tipo de persona implicada	Categórica nominal	1	4 tipos de persona
rango_edad	Edad dividida por rangos	Categórica	1	17 rangos
sexo	Género de la persona implicada	Binaria o Categórica nominal	2	Hombre, mujer y desconocido
cod_lesividad	Describe la gravedad del accidente	Numérica ordinal	0	Valores 1-7 + 14
lesividad	Describe la gravedad del accidente	Categórica ordinal	0	Variable 7 tipos de lesividad
coordenada_x_utm	Coordenadas eje X (Este)	Numérica continua	1	Útil para mapas
coordenada_y_utm	Coordenadas eje Y (Norte)	Numérica continua	1	Útil para mapas
positiva_acohol	Resultado de test de alcoholemia	Binaria	1	N (No) / S (Si)
positiva_droga	Resultado de test de drogas	Binaria	3	Valores: 1.0 o NaN

Las variables principales de los datasets son num_expediente y cod_lesividad/lesividad, ya que permiten definir el número de accidentes por registro y la gravedad de los accidentes, respectivamente. El resto de variables son mayoritariamente categóricas, a excepción de cod_distrito y las coordenadas X e Y, que son numéricas, así como las variables positiva_alcohol y positiva_droga, que son binarias.

Una vez identificadas las variables, se procedió a la limpieza y preprocesado de los datos siguiendo los siguientes pasos:

1. Eliminación de registros duplicados, conservando únicamente el primero (*Keep = "first"*).
2. Unificación de formato y tratamiento de valores nulos en las variables categóricas: estado_meteorológico, tipo_persona, tipo_vehículo, cod_lesividad, lesividad, positiva_alcohol y positiva_droga.
3. Modificación de variables categóricas: positiva_alcohol y rango_edad, para garantizar consistencia en los códigos y etiquetas.
4. Creación de nuevas variables: fecha_hora y franja_horaria, que facilitan el análisis temporal de los accidentes, y grado_lesividad, que simplifica el análisis por gravedad de accidente.
5. Eliminación de columnas innecesarias o redundantes, como cod_distrito, cod_lesividad, lesividad, fecha, hora y positiva_droga, para simplificar el dataset sin pérdida de información relevante.

3. Análisis Exploratorio de Datos

A continuación, se detalla el análisis uni-, bi- y multivariante de los campos incluidos en los datasets de accidentalidad intraurbana en Madrid correspondientes al año 2024.

3.1 Análisis univariante

Siguiendo un enfoque sistemático, las variables del dataset se han agrupado según el tipo de análisis realizado: variables temporales, geográficas, demográficas, factores asociados al accidente y gravedad del accidente.

3.1.1 Variables temporales

La **Figura 1** muestra la frecuencia de accidentes a diferentes escalas temporales. En la **Figura 1a** se observa que los meses con mayor número de accidentes en 2024 fueron octubre, noviembre y diciembre. Este incremento podría estar relacionado con el inicio del año académico, el aumento de la actividad urbana y el empeoramiento de las condiciones meteorológicas durante los meses de otoño. Por el contrario, durante los meses de verano se registra una disminución del número de accidentes, siendo agosto el mes con menor siniestralidad (5,33 % del total anual). Esta reducción podría explicarse por una menor densidad poblacional y de tráfico, asociada al periodo vacacional. En cuanto al día de la semana, el viernes es el día en el que se registra un mayor número de accidentes, aunque con diferencias poco pronunciadas respecto al resto de días laborables. Durante los fines de semana, y especialmente los domingos, se observa una menor frecuencia de accidentes, lo que podría atribuirse a una reducción del volumen de tráfico (**Figura 1b**). Las **Figuras 1c** y **1d** muestran la distribución de los accidentes por hora del día y por franjas horarias, respectivamente. Destaca que la franja de tarde (12:00–18:00) concentra el mayor número de accidentes, identificándose las 18:00 horas como la hora punta de

sinistralidad. Este patrón podría estar relacionado con los desplazamientos asociados al final de la jornada laboral y al aumento de la congestión viaria. En general, los resultados indican que los meses de otoño, los días laborables (especialmente los viernes) y la franja horaria de la tarde constituyen los periodos de mayor sinistralidad, por lo que deberían ser considerados prioritarios para la implementación de medidas preventivas y de control del tráfico.

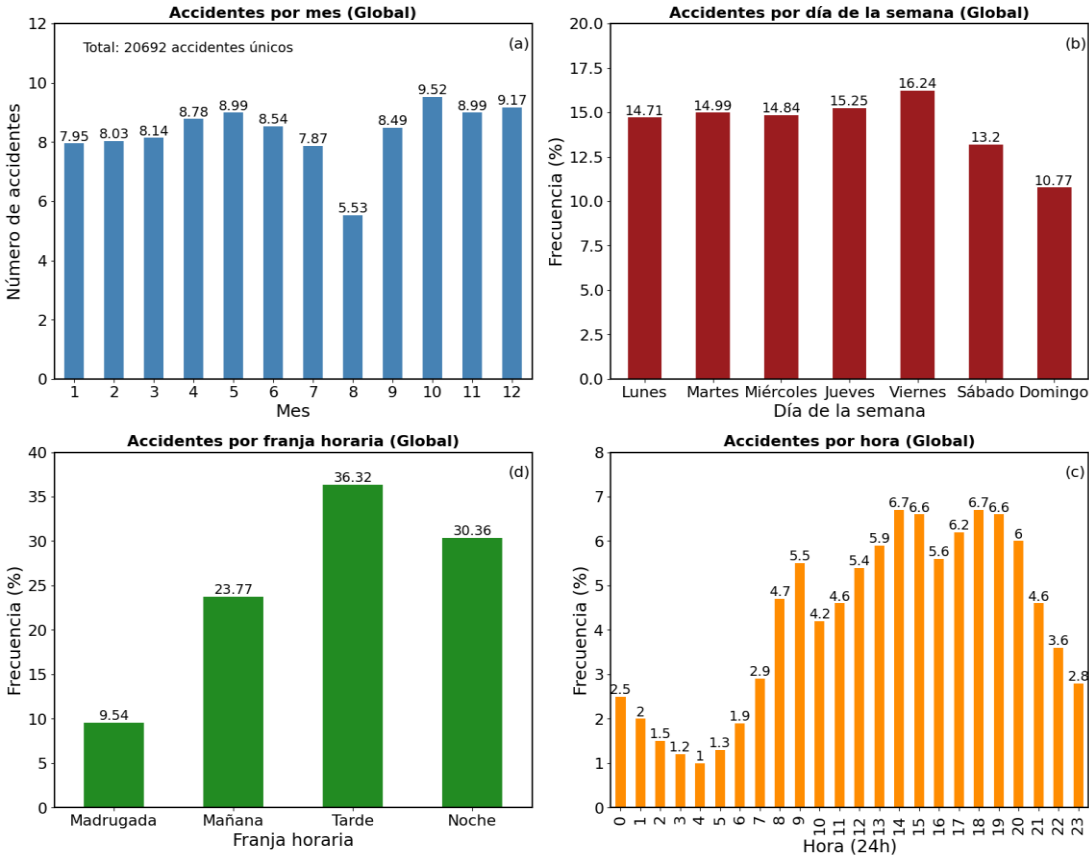


Figura 1. Frecuencia de accidentes intraurbanos a diferentes escalas temporales: a) mes, b) día de la semana, c) hora del día y d) franja horaria.

3.1.2 Variables geográficas

La distribución de los accidentes por distrito, así como las cinco ubicaciones con mayor sinistralidad, se muestran en la **Figura 2**. Se observa que Puente de Vallecas lidera el ranking, concentrando el 7,4 % de los accidentes registrados en 2024 (1.530 en total), seguido de Chamartín (7,1 %) y Carabanchel (6,96 %). Otros distritos céntricos, como Centro, Salamanca y Retiro, también presentan un número elevado de accidentes, lo que podría estar relacionado con un mayor volumen de circulación y actividad urbana. Por otra parte, cabe destacar que los cinco puntos negros identificados se localizan en tramos específicos de autopistas, concretamente en la A-2, A-42 y M-23, y se encuentran definidos mediante su referencia kilométrica. En el caso de la A-2, el punto situado en el kilómetro +00500E, dentro del distrito de Ciudad Lineal, concentra 35 accidentes únicos, lo que pone de manifiesto una recurrencia significativa de la sinistralidad en una localización concreta. Los puntos negros detectados en las autopistas A-2 y A-42 pueden considerarse zonas de peligrosidad recurrente, posiblemente asociadas a factores como incorporaciones complejas, cambios de rasante, trazados curvos o deficiencias en la señalización. Estos tramos deberían constituir una prioridad para la implementación de medidas

de seguridad vial, dado que concentran un número elevado de accidentes únicos en localizaciones muy específicas.

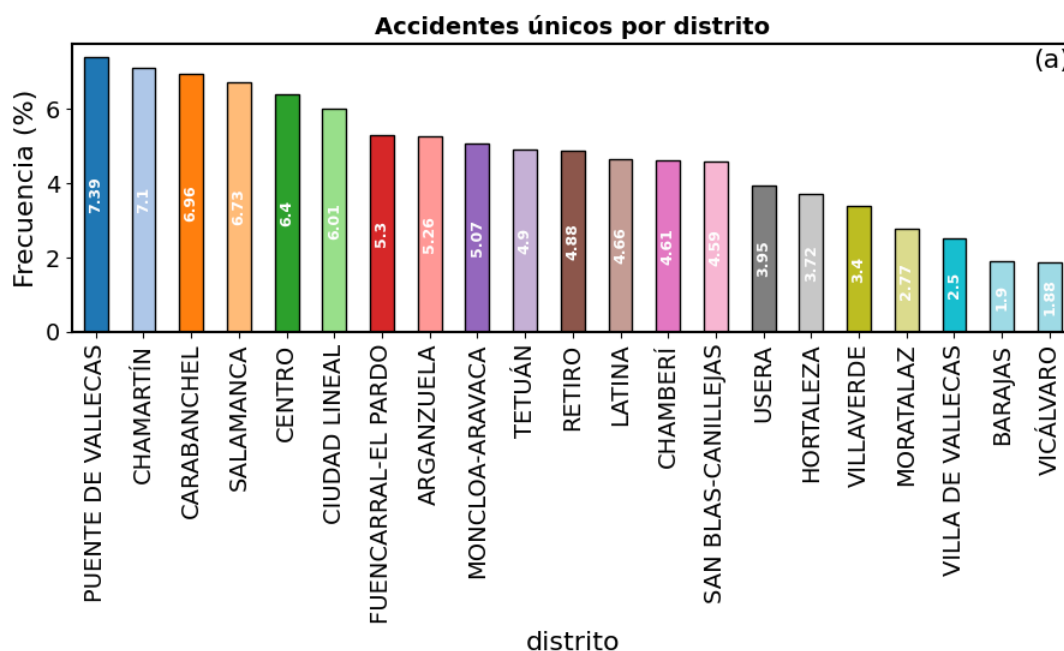


Figura 2. Frecuencia de accidentes en función de variables geográficas: frecuencias relativas por distrito.

3.1.3 Variables demográficas

La **Figura 3** muestra la distribución de las personas implicadas en accidentes según rango de edad, tipo de persona y sexo del conductor. En relación con la edad, el histograma no presenta una distribución normal, sino una distribución relativamente plana o de tipo meseta. En concreto, las frecuencias correspondientes a los rangos de edad comprendidos entre 25 y 54 años se sitúan entre el 10,5 % y el 11,5 %, acumulando un total de 30673 personas implicadas (64,9 %). Este intervalo coincide con las edades de mayor actividad laboral, movilidad urbana y uso intensivo de vehículos. No obstante, cabe destacar que el rango de edad que concentra un mayor número de accidentes es el comprendido entre 45 y 49 años, lo que podría estar relacionado con factores como una elevada exposición al tráfico y, en menor medida, una posible disminución progresiva de los reflejos. Por otro lado, los menores de 18 años representan aproximadamente el 2 % del total de personas implicadas, mientras que los mayores de 60 años suponen en torno al 7,1 %. Aunque su participación relativa es menor, estos grupos presentan una mayor vulnerabilidad, especialmente en calidad de peatones o pasajeros.

La **Figura 3b** muestra la distribución de las personas implicadas según el tipo de persona. Como era de esperar, los conductores constituyen el grupo mayoritario (80 %), seguidos de los pasajeros (16 %) y los peatones (4 %). El menor porcentaje de pasajeros en comparación con los conductores podría indicar que una parte significativa de los desplazamientos se realiza sin acompañantes o que existe una mayor probabilidad de verse implicado en un accidente al viajar solo. Finalmente, en cuanto al sexo de los conductores, se observa un claro predominio del género masculino, con una frecuencia cercana al 75 %. Esta diferencia notable respecto a las conductoras podría estar asociada a una mayor exposición al tráfico (medida en kilómetros recorridos) y, potencialmente, a diferencias en los patrones y hábitos de conducción, si bien este aspecto requeriría un análisis específico para ser confirmado.

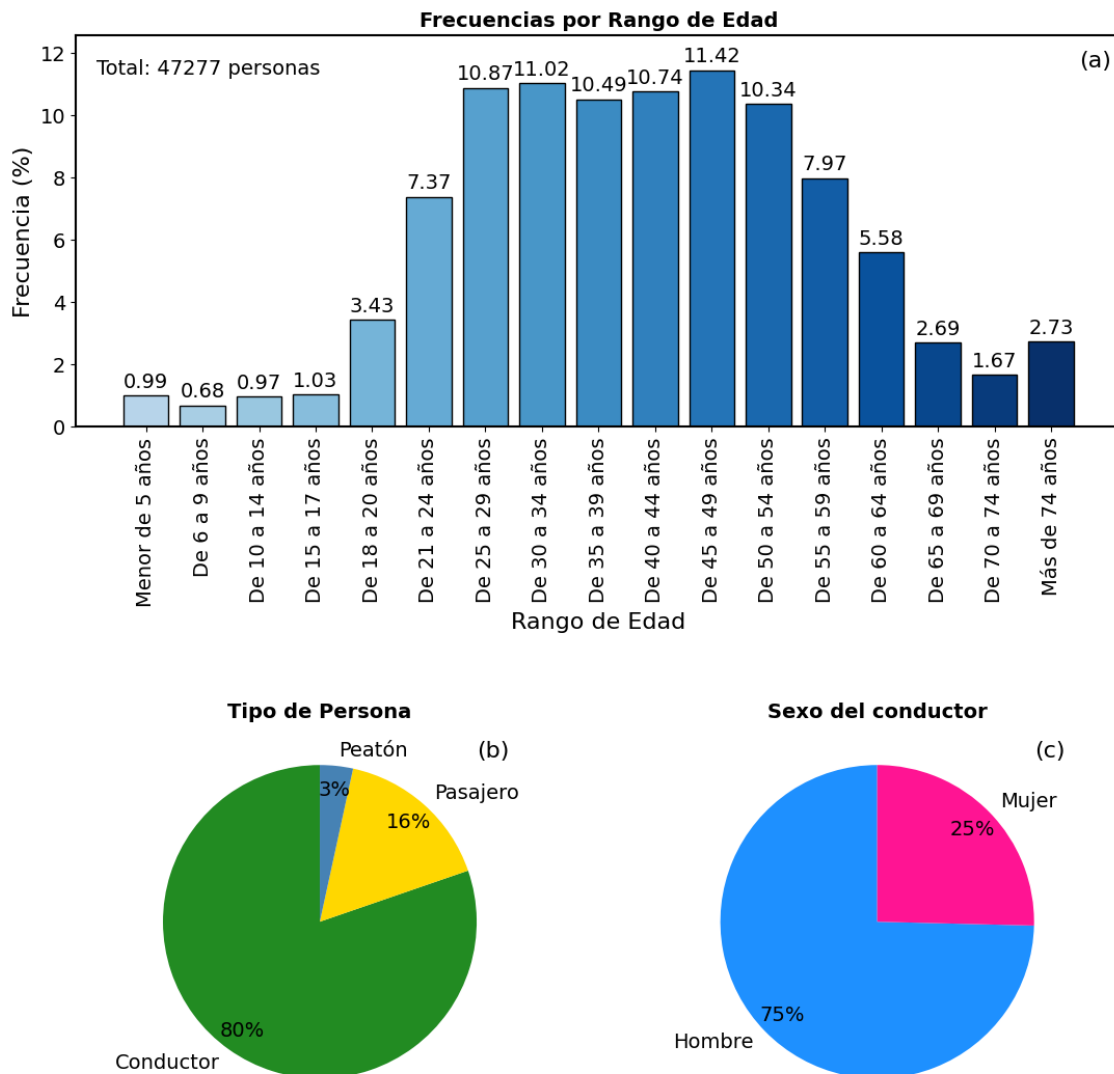


Figura 3. Frecuencia de personas implicadas según (a) rango de edad, (b) tipo de persona y (c) sexo del conductor.

3.1.4 Factores asociados al accidente

La **Figura 4** muestra la frecuencia de accidentes en función de distintos factores asociados al accidente, concretamente el tipo de vehículo, el resultado del test de alcoholemia, el estado meteorológico y el tipo de accidente.

La proporción de accidentes según el tipo de vehículo (cinco categorías más frecuentes) se ilustra en la **Figura 4a** mediante un diagrama de tipo *treemap*. Como era de esperar, el turismo es el vehículo más frecuentemente implicado, lo que refleja que la movilidad urbana particular constituye el principal escenario de riesgo. En segundo y tercer lugar se sitúan las motocicletas de distinta cilindrada (hasta y más de 125 cc), con más de 4.400 accidentes en conjunto, lo que representa una proporción considerable. Este resultado sugiere que, aunque menos numerosas que los turismos, las motocicletas presentan una alta exposición al riesgo, por lo que deben ser consideradas de forma prioritaria en las políticas de seguridad vial. Aunque en menor medida, los vehículos de transporte, como furgonetas y camiones rígidos, también aparecen entre los

cinco tipos de vehículo más implicados, lo que indica que la logística urbana contribuye de manera relevante a la siniestralidad, especialmente en zonas de carga, reparto y maniobras.

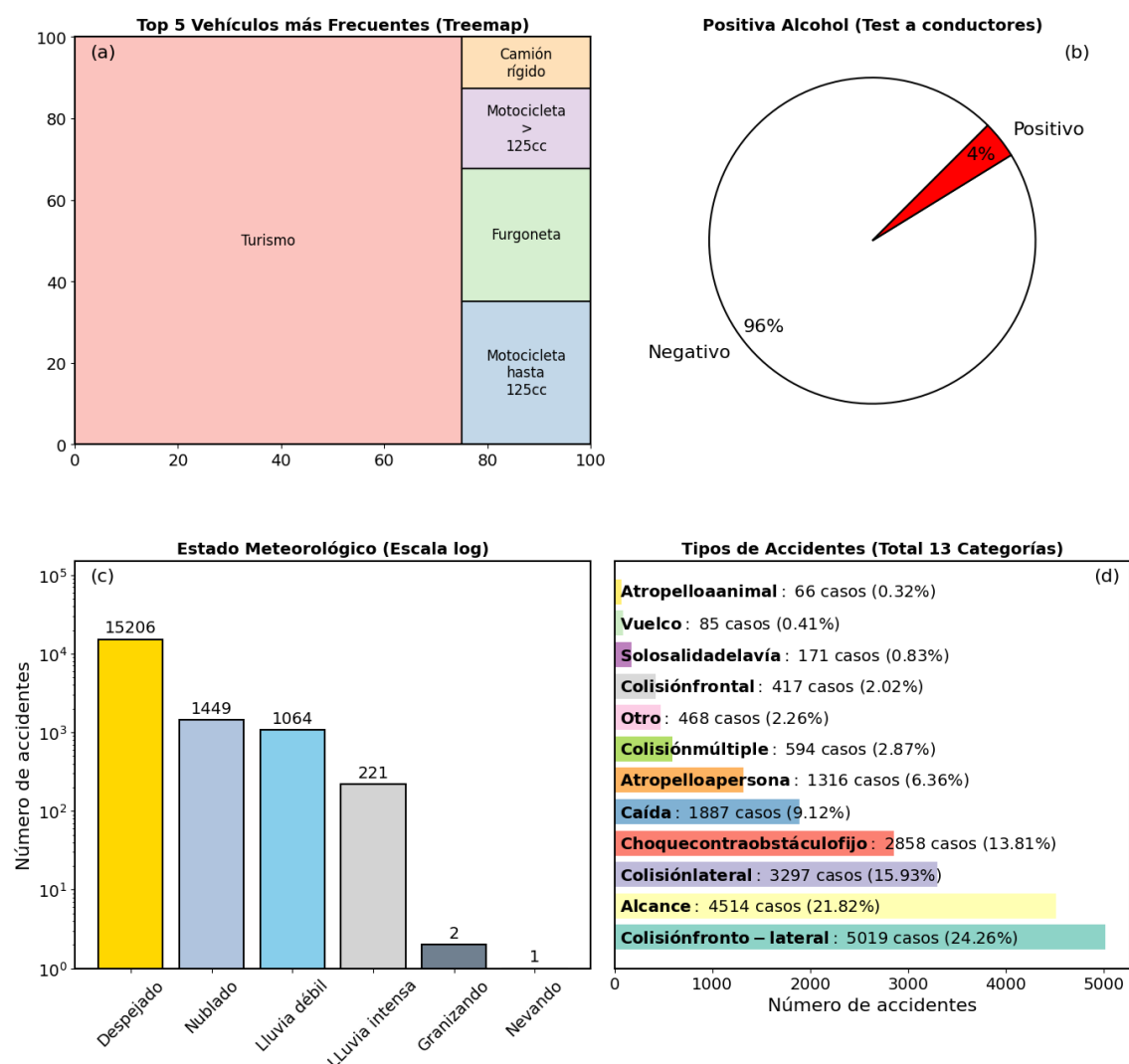


Figura 4. Frecuencia de accidentes intraurbanos en función del (a) tipo de vehículo, (b) resultado del test de alcoholemia, (c) estado meteorológico y (d) tipo de accidente.

En relación con el test de alcoholemia (**Figura 4b**), el 96 % de los conductores implicados dio resultado negativo, lo que indica que la mayoría de los accidentes intraurbanos no están directamente asociados al consumo de alcohol. No obstante, el 4 % de conductores con resultado positivo constituye una proporción no despreciable, cuya reducción resulta prioritaria, dado que el consumo de alcohol suele estar asociado a una mayor gravedad de las lesiones. Respecto a las condiciones meteorológicas (**Figura 4c**), la mayoría de los accidentes (15.206 expedientes, 85 %) se producen con cielo despejado, lo cual es coherente con el hecho de que esta es la condición meteorológica más habitual en Madrid. El 15 % restante de los accidentes se concentra en condiciones adversas, principalmente lluvia (débil e intensa, en torno al 10 %) y niebla (5 %). Aunque menos frecuentes, estas condiciones pueden representar un factor de riesgo adicional, especialmente para usuarios vulnerables como motociclistas y peatones. Finalmente, según la **Figura 4d**, los tipos de accidente más frecuentes son la colisión fronto-lateral, el alcance y la colisión lateral, que en conjunto acumulan aproximadamente el 62 % del

total de accidentes. Dada la especial vulnerabilidad de determinados usuarios, resulta destacable el porcentaje de atropellos (9,12 %, 1.887 casos), una tipología de accidente cuya reducción debería constituir un objetivo prioritario en las estrategias de seguridad vial.

3.1.5 Gravedad del accidente

La **Figura 5** muestra la frecuencia de accidentes según la lesividad de las personas implicadas, siendo la escala de la **Figura 5a** logarítmica para facilitar la visualización de las categorías menos frecuentes.

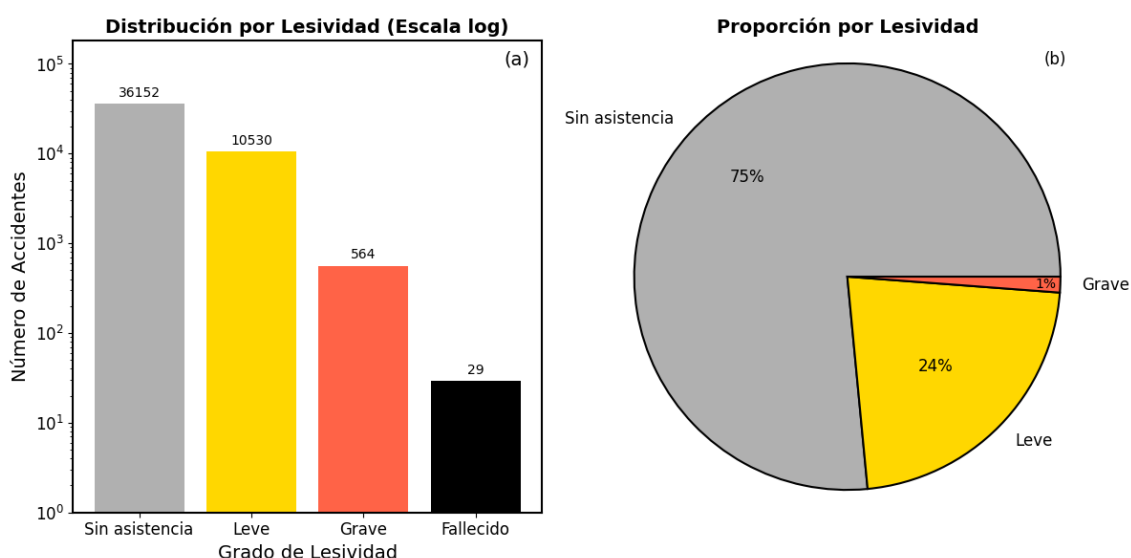


Figura 5. Frecuencias de accidentes (a) absolutas y (b) relativas por grado de lesividad.

Se observa que en la mayoría de los accidentes las personas implicadas no requirieron asistencia sanitaria, lo que indica que la mayoría de los accidentes registrados fueron muy leves o sin consecuencias físicas aparentes. No obstante, esto no excluye la posibilidad de impactos psicológicos o daños materiales derivados del accidente. Por otra parte, menos de una quinta parte de los implicados (24 %) presentó lesiones leves, mientras que alrededor del 1 % sufrió lesiones graves, y se registraron 29 fallecimientos. Aunque estas cifras son numéricamente bajas, representan casos críticos desde el punto de vista de la seguridad vial, y su análisis es fundamental para la planificación de medidas preventivas, especialmente en las zonas o tipos de accidente donde se concentran.

3.2 Análisis bivalente

La estructura seguida para el análisis bivalente ha sido primero realizar una visualización mediante un mapa de calor cruzando dos variables y a continuación el test Chi-cuadrado, para afirmar o negar la existencia de una relación estadísticamente significativa entre ambas variables.

En esta sección se cruza el grado de lesividad y el resultado del test de alcoholemia con distintas variables, y para un análisis espacial, se cruza tipo de accidente y distrito. Otros cruces se presentan en el Anexo.

3.2.1 Gravedad del accidente

La lesividad de las personas implicadas en un accidente es un factor clave en la siniestralidad vial. En este apartado se realiza el análisis bivalente del grado de lesividad de los usuarios con distintas variables.

Se analiza la distribución del grado de lesividad según si la persona es un conductor, un pasajero o un peatón. Para los siguientes dos análisis, la categoría “Fallecido” se ha excluido porque su frecuencia es extremadamente baja (solo 29 casos), lo que invalida el test Chi-cuadrado al generar más del 20% de valores de frecuencias esperadas inferiores a 5, o generar valores menores que 1. De esta forma se garantiza que las frecuencias esperadas sean adecuadas y se mejora la robustez del contraste estadístico.

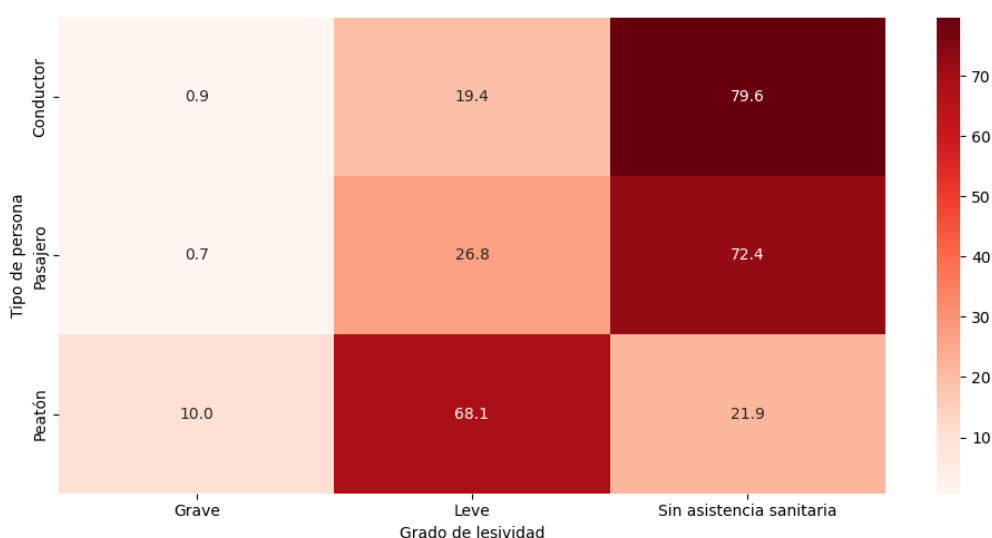


Figura 6. Grado de lesividad por tipo de persona (no fallecida)

La **Figura 6** muestra que los conductores y los pasajeros tienen una distribución de grado de lesividad similar a la global. Los peatones en cambio destacan por su alta proporción en lesiones leves (un 68,1 %) y con una proporción de lesiones graves superior a la de los pasajeros y conductores.

El test Chi-cuadrado afirma que existe una relación entre el grado de lesividad y el tipo de persona implicada en el accidente.

A continuación, se analiza la distribución del grado de lesividad según el rango de edad de las personas adultas implicadas.

En el análisis univariante se ha concluido que los rangos de 25 a 54 años, concentran la mayor proporción de conductores en general. Por tanto, es coherente que en el mapa de calor estos rangos acumulen los valores más altos en las tres categorías, como se observa en la **Figura 7**. En particular, la mayor concentración de lesiones leves se encuentra dentro del rango de 25 a 34 años. Por otro lado, los mayores de 65 años presentan porcentajes bajos en todas las categorías, salvo los mayores de 74 años con un grado de lesividad grave. Además, a partir de los 60 años, las proporciones con lesiones graves son superiores a las proporciones de las heridas leves y a las proporciones de los que no han necesitado asistencia sanitaria. Los jóvenes de entre 18 y 24 años, tienen la menor proporción de lesiones graves, pero están presentes.



Figura 7. Rango de edad por grado de lesividad

Se confirma que existe una relación estadísticamente significativa entre el rango de edad y el grado de lesividad mediante el test Chi-cuadrado (p -valor $<< 0.0001$), por lo que se concluye que la edad influye en la gravedad de las lesiones sufridas en un accidente.

Se sigue analizando cómo se distribuye los tipos de accidentes más significativos según el grado de lesividad, recuperando la categoría de fallecidos.

La **Figura 8** muestra que las personas atropelladas representan una proporción dentro de los fallecidos muy elevada, un 61,5 %. Dentro de los heridos graves también representan una proporción elevada, junto con las personas implicadas en accidentes por colisión fronto-lateral. Como la colisión fronto-lateral fue el tipo de accidente más común, es coherente que también muestre porcentajes altos en lesividad leve y en las personas que no han necesitado asistencia médica. Lo mismo ocurre con los accidentes por alcance. También se observa que los accidentes por colisión múltiple y choques contra obstáculos fijos tienen porcentajes bajos en todas las categorías, y que las caídas tienen una presencia relevante en lesiones leves y graves.

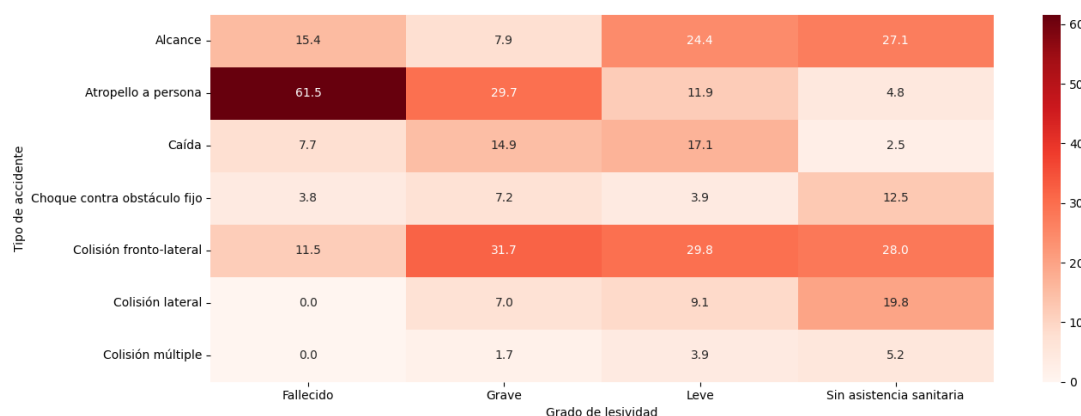


Figura 8. Tipo de accidente por grado de lesividad

El test chi-cuadrado confirma que esta relación es estadísticamente significativa (p -valor $<< 0.0001$), lo que indica una fuerte relación entre el tipo de accidente y el grado de lesividad.

Por último, se cruza las variables de alcohol y lesividad, sin embargo, el resultado del test Chi-cuadrado muestra que no hay suficiente evidencia estadística para afirmar que existe relación entre el resultado de alcohol y el grado de lesividad del conductor. Aún así es un factor importante a tener en cuenta, aunque solo un 4 % diera positivo en los test de alcoholemia. En consecuencia, en el siguiente apartado se realiza el análisis bivalente del resultado de alcohol con distintas variables, considerando un único accidente por expediente según el siguiente criterio: si dentro de un expediente hay involucrado más de un conductor y alguno dio positivo, cuenta como positivo y si todos dieron negativo, se considera negativo.

3.2.2 Consumo de alcohol del conductor

Se analiza la distribución de los tipos de accidentes según el resultado del test de alcoholemia. En la **Figura 9**, solo se muestra la distribución para resultados positivos, pues para resultados negativos los tipos de accidentes se distribuyen según las frecuencias observadas en el análisis univariante. Se observa una gran concentración en choques contra obstáculos fijos. En cuanto al resto de accidentes, siguen una distribución similar a la global.

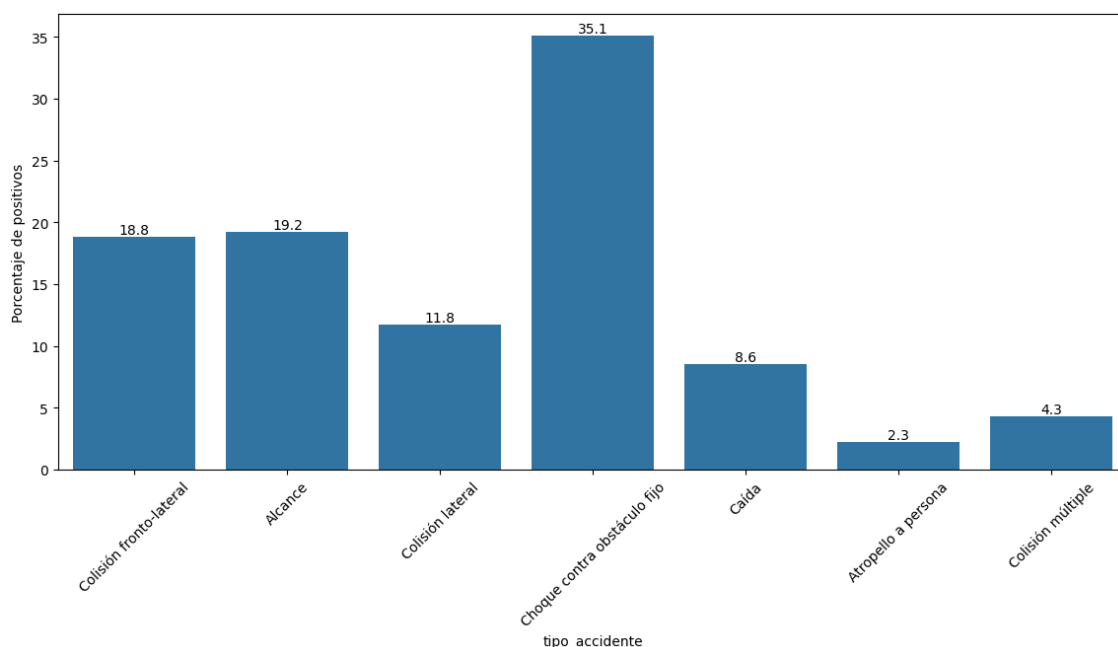


Figura 9. Tipo de accidente por alcohol positivo del conductor

El test Chi-cuadrado confirma que hay una relación estadísticamente significativa entre el consumo de alcohol y el tipo de accidente ($p\text{-valor} < 1e-100$).

Se sigue analizando la distribución de los rangos de edad de los accidentes con alcohol positivo y negativo del conductor, de nuevo por accidentes únicos. En la **Figura 10** se observa que, de nuevo las personas de entre 25 y 54 años acumulan los valores más altos tanto en positivos como en negativos. Hasta los 39 años, los valores de positivos son mayores que los negativos y el máximo de positivos se encuentra en el rango de 25 a 29. A partir de los 55 años, los positivos descienden de forma progresiva. Por tanto, se observa que hay una mayor presencia de alcohol en los jóvenes-adultos.

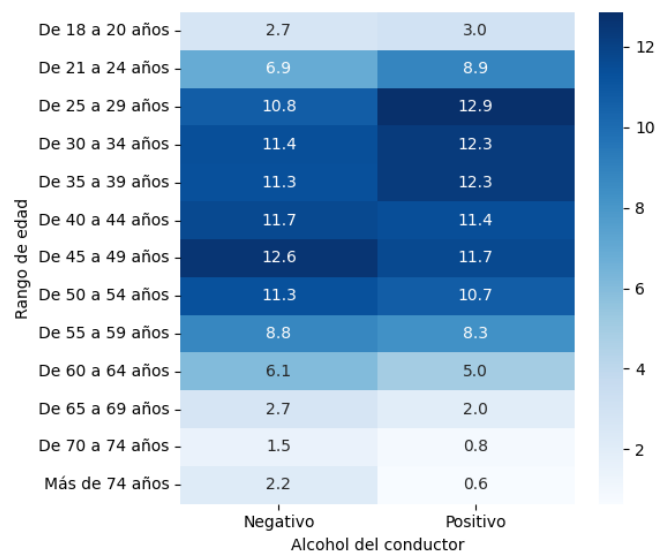


Figura 10. Rango de edad por alcohol del conductor

El test Chi-cuadrado revela una relación estadísticamente significativa entre el rango de edad y el resultado de alcohol en accidentes (p -valor < 0.0001), lo que indica que el patrón de edades en accidentes con alcohol positivo es diferente al de los negativos.

Por último, se analiza la distribución de las franjas horarias del día en los positivos y negativos de alcohol de conductores.

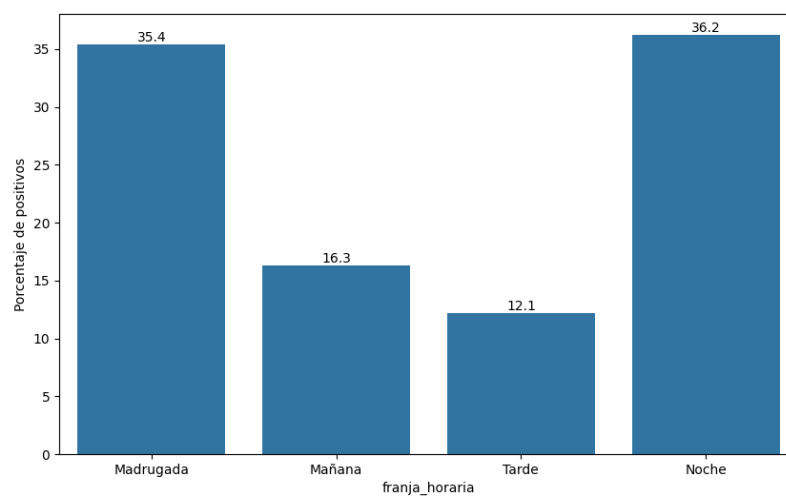


Figura 11. Franja horaria por alcohol positivo del conductor

Debido a que los resultados negativos siguen la misma distribución que las frecuencias globales de las franjas horarias, en la **Figura 11**, solo se muestran los resultados positivos. Se observa que los accidentes con alcohol positivo se concentran en Madrugada y Noche, representando un 71,6 %.

Al realizar el test se confirma que existe una relación estadísticamente significativa entre la franja horaria y el resultado de alcohol del accidente. Como conclusión, la prevalencia de alcohol positivo en conductores varía significativamente según la hora del día.

3.2.3 *Análisis geográfico*

Se analiza cómo varían los tipos de accidentes más significativos entre los distintos distritos, considerando solo expedientes únicos, mediante un mapa de calor (para una visualización completa, véase la **Figura 19** del Anexo), que gracias a su escala de colores es posible identificar patrones territoriales, contrastes entre distritos y tipos de accidente que destacan por su frecuencia. Se observa que, en prácticamente todos los distritos, los accidentes que predominan son las colisiones fronto-lateral o por alcance. Cabe recordar que estos tipos están en el top 2 de tipos de accidentes con unas frecuencias relativas porcentuales de 24,26 % y 21,82 % respectivamente. Las colisiones laterales, con una proporción de 15,93 %, también están muy presentes en casi todos los distritos. También se observa que en los distritos menos céntricos hay una proporción relativamente elevada de choques contra obstáculos fijos. Por otro lado, el Centro destaca en atropellos a personas o caídas, respecto al resto de distritos, lo que podría estar asociado una alta interacción peatonal.

El análisis revela que los tipos de accidentes significativos varían sustancialmente entre distritos. Esta variación no es aleatoria: el test chi-cuadrado confirma una relación estadísticamente significativa entre distrito y tipo de accidente (p -valor < $1e-100$). Esto respalda la necesidad de estrategias de prevención diferenciadas por zona, adaptadas al perfil de siniestralidad local.

3.3 Análisis multivariante

El presente análisis multivariante tiene como objetivo comprender cómo interactúan simultáneamente distintos factores geográficos, demográficos y de gravedad en los accidentes de tráfico registrados en la ciudad de Madrid durante el año 2024. A diferencia de los análisis univariantes o bivariantes, este enfoque permite identificar patrones complejos que solo emergen cuando se estudian varias variables de forma conjunta, proporcionando una visión más profunda y estructural del fenómeno de la accidentalidad urbana.

El análisis multivariante desarrollado en este estudio tiene un carácter fundamentalmente **descriptivo-exploratorio**. Su propósito es identificar **patrones, concentraciones espaciales y perfiles de riesgo** derivados de la combinación simultánea de varias variables, sin pretender establecer relaciones causales ni evaluar formalmente la significancia estadística de sus interacciones. Bajo este enfoque, se plantean las siguientes preguntas de investigación:

- ¿Qué distritos presentan mayor densidad de accidentes por km² en Madrid?
- ¿Cómo varía el grado de lesividad en función del distrito y del tipo de accidente?
- ¿Cómo varía el grado de lesividad en función del sexo y la edad de las personas implicadas en accidentes de tráfico?

3.3.1 *Variables geográficas*

En este primer bloque se examina la localización espacial de los accidentes mediante sus coordenadas UTM (X e Y), coloreadas por distrito. Esta representación permite visualizar la forma aproximada de cada distrito, su extensión y su estructura territorial sin necesidad de un mapa explícito. Además, la densidad visual de puntos revela áreas con mayor concentración de siniestros. También se calcula la densidad de accidentes por kilómetro cuadrado.

Tabla 2. Densidad accidentes por km2

distrito	n_accidentes_unicos	area_km2	densidad_accidentes_km2
SALAMANCA	1393	5.38	258.921933
CENTRO	1327	5.23	253.728489
CHAMARTÍN	1470	5.99	245.409015
CHAMBERÍ	954	4.69	203.411514
TETUÁN	1014	5.38	188.475836
RETIRO	1010	5.47	184.643510
ARGANZUELA	1088	6.46	168.421053
CIUDAD LINEAL	1244	11.43	108.836395
USERA	818	7.75	105.548387
PUENTE DE VALLECAS	1530	14.89	102.753526
CARABANCHEL	1440	14.76	97.560976
MORATALAZ	574	6.34	90.536278
SAN BLAS-CANILLEJAS	949	22.36	42.441860
LATINA	965	25.41	37.977174
VILLAVERDE	704	20.29	34.696895
HORTALEZA	769	27.41	28.055454
VICÁLVARO	388	14.40	26.944444
MONCLOA-ARAVACA	1050	46.65	22.508039
VILLA DE VALLECAS	517	48.34	10.695077
BARAJAS	394	41.94	9.394373
FUENCARRAL-EL PARDO	1097	237.81	4.612926

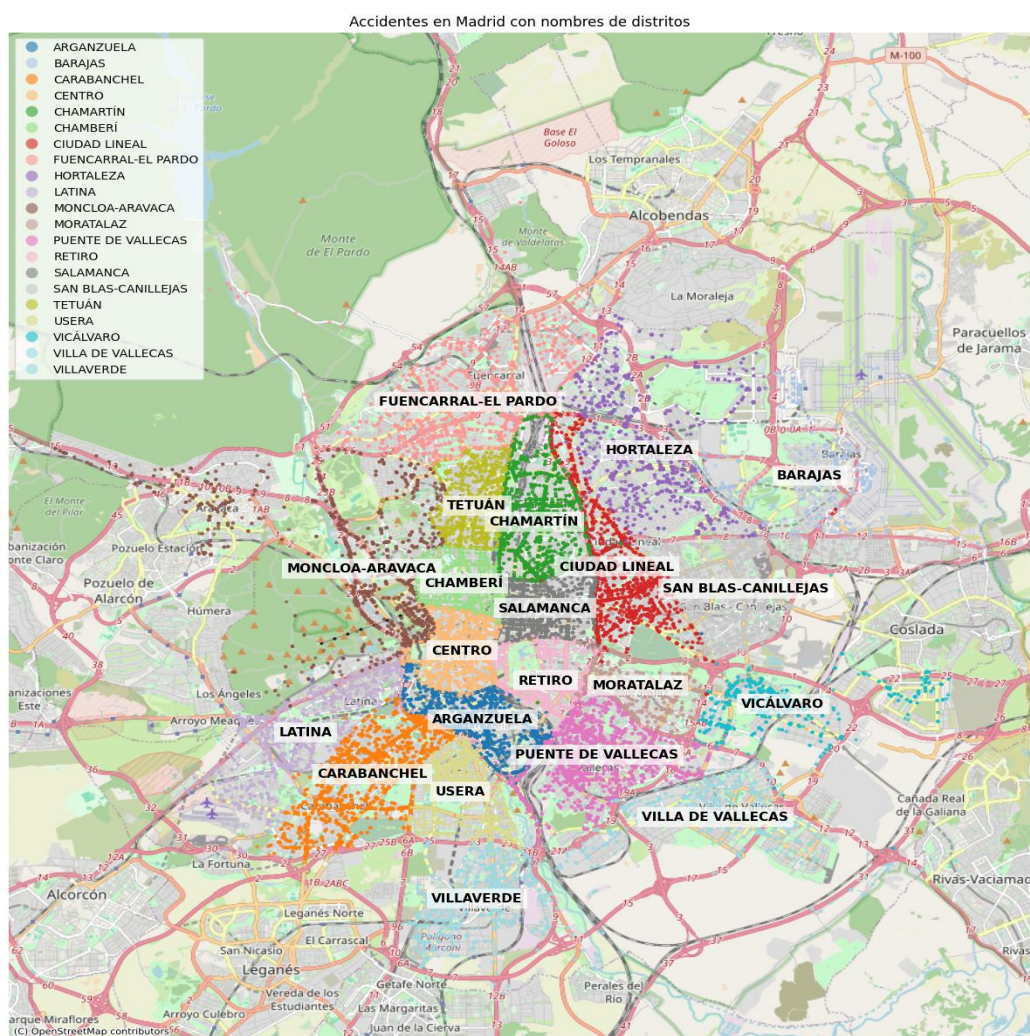


Figura 12. Accidentes Madrid 2024 por distritos

¿Qué distritos presentan mayor densidad de accidentes por km² en Madrid?

El análisis geográfico combina dos enfoques complementarios:

1. **Visualización mediante scatterplot**, donde la superposición de puntos permite identificar zonas con alta concentración aparente de accidentes.
2. **Cálculo de densidad por km²**, que corrige el efecto del tamaño del distrito y permite determinar si la concentración observada es proporcional a su superficie o si, por el contrario, refleja una mayor intensidad real de siniestros.

Al normalizar por área, se confirma que los distritos con **mayor densidad de accidentes por km²** son:

1. **Salamanca**
2. **Centro**
3. **Chamartín**
4. **Chamberí**
5. **Tetuán**

Estos distritos presentan una combinación de **alta actividad urbana, intenso flujo peatonal y vehicular y elevada densidad poblacional**, lo que explica que concentren un número de accidentes significativamente mayor en relación con su superficie.

3.3.2 Variables demográficas

El tercer bloque examina cómo se distribuye la lesividad entre hombres y mujeres a lo largo del ciclo vital, analizando simultáneamente el sexo y el rango de edad de las personas implicadas en accidentes de tráfico. Este enfoque permite identificar no solo diferencias globales entre hombres y mujeres, sino también cómo estas diferencias se modifican en función de la etapa vital, revelando patrones de vulnerabilidad específicos para cada grupo demográfico.

El análisis pone especial atención en la interacción entre ambas variables, ya que la exposición al riesgo y la gravedad de las lesiones no se distribuyen de manera uniforme a lo largo de la vida.

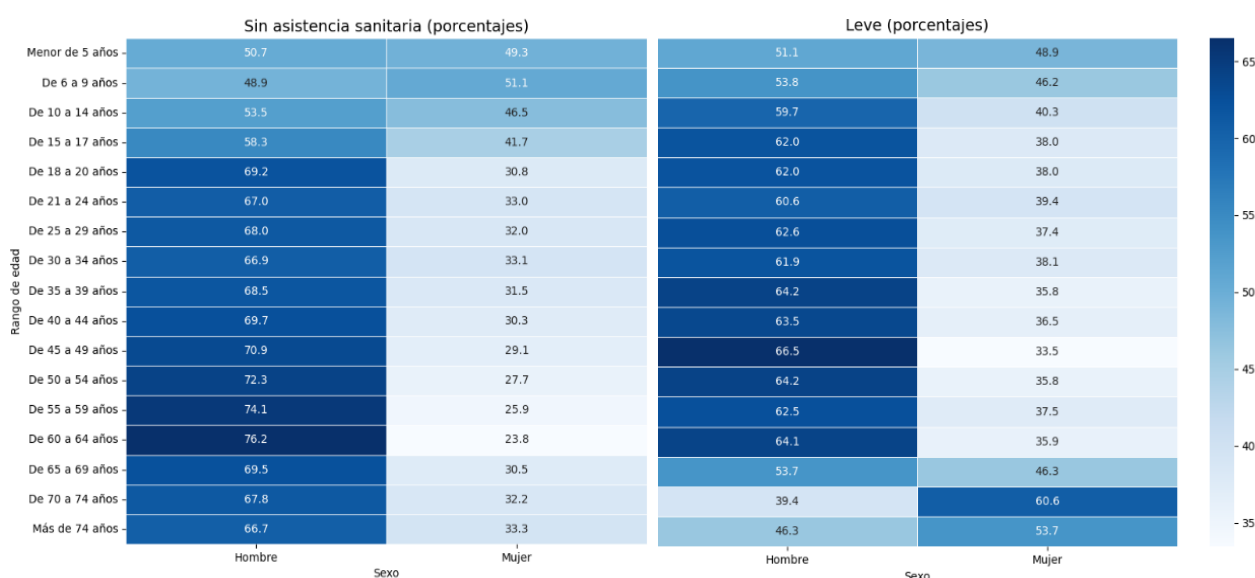


Figura 13. Rango de edad y sexo por grado de lesividad: a) Sin asistencia sanitaria y b) leve

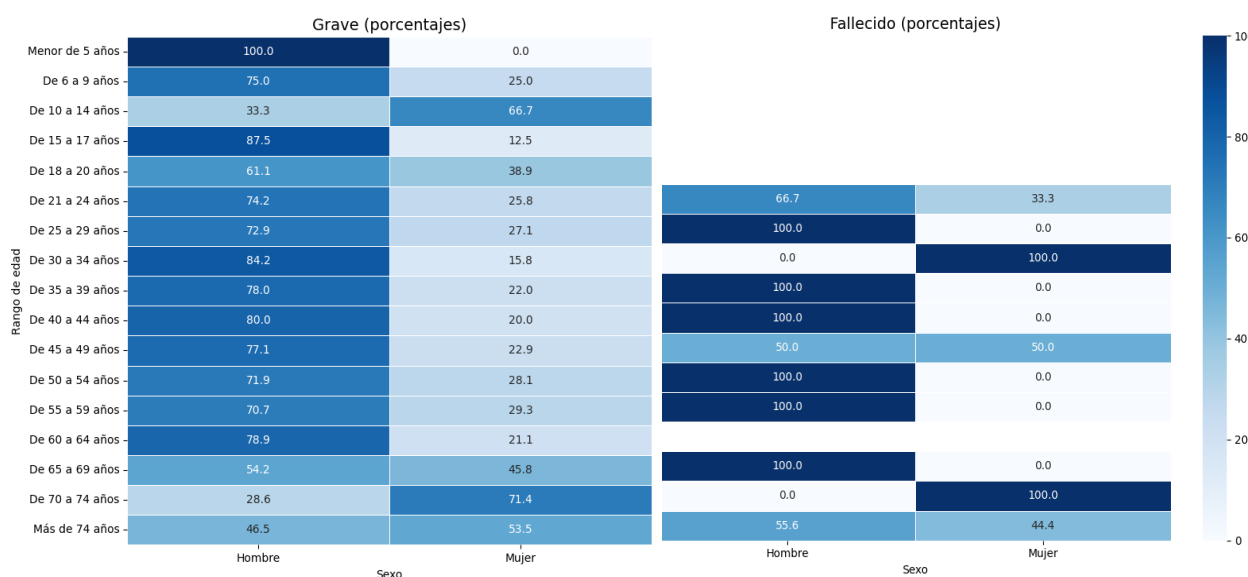


Figura 14. Rango de edad y sexo por grado de lesividad: a) grave y b) fallecido.

El análisis conjunto del sexo y el rango de edad de las personas implicadas en accidentes de tráfico permite identificar patrones diferenciados de lesividad a lo largo del ciclo vital. La distribución de las categorías de gravedad revela comportamientos específicos tanto en hombres como en mujeres, así como variaciones relevantes en edades avanzadas.

¿Cómo varía el grado de lesividad en función del sexo y la edad de las personas implicadas en accidentes de tráfico?

Sin asistencia sanitaria

En la categoría “*Sin asistencia sanitaria*” se observa un **predominio masculino constante** a partir de los 15 años. Este patrón se intensifica en edades medias y avanzadas:

- A partir de los **50 años**, los hombres representan **más del 70 %** de los casos sin asistencia.

Este comportamiento sugiere que los hombres tienden a verse implicados en accidentes que no requieren atención médica, posiblemente debido a una **mayor exposición vial**, diferencias en los patrones de movilidad o una **menor percepción del riesgo**.

Lesiones leves

Las *lesiones leves* también muestran un predominio masculino en la mayoría de los grupos de edad, especialmente entre los **35 y 59 años**, donde los hombres superan el **64 %** de los casos.

Sin embargo, en edades avanzadas el patrón se invierte:

- En el grupo **70–74 años**, las mujeres representan **más del 60%** de los casos.
- En mayores de **74 años**, las mujeres alcanzan el **53,7 %**.

Este cambio puede reflejar una **mayor vulnerabilidad femenina en edades avanzadas**, diferencias en la fragilidad física o una mayor propensión a reportar lesiones leves.

Lesiones graves

En la categoría de *lesiones graves* se mantiene un **fuerte predominio masculino** en la mayoría de los grupos de edad, con valores que oscilan entre el **70 % y el 87 %** en edades adultas.

No obstante, existen **excepciones relevantes**:

- **10–14 años:** predominio femenino (**66,7 %**).
- **70–74 años:** predominio femenino (**71,4 %**).
- **Más de 74 años:** predominio femenino (**53,5 %**).

Estos resultados indican que las **mujeres mayores presentan una mayor probabilidad de sufrir lesiones graves** en caso de accidente, posiblemente asociada a factores de fragilidad o condiciones de salud previas.

Fallecidos

Aunque el número de casos es reducido, se observan patrones extremos:

- En edades medias (**25–59 años**), los fallecidos son **casi exclusivamente hombres** (100 % en varios grupos).
- En mayores de **74 años**, la distribución se equilibra (55,6 % hombres, 44,4 % mujeres).
- En el grupo **70–74 años**, los fallecidos registrados son **exclusivamente mujeres** (100 %).

Estos resultados sugieren que, aunque la mortalidad es mayoritariamente masculina en edades activas, en edades avanzadas la vulnerabilidad femenina aumenta de forma notable.

3.3.3 Gravedad accidente

El tercer bloque del análisis multivariante aborda de manera conjunta la relación entre el distrito en el que ocurre el siniestro, el tipo de accidente y el grado de lesividad resultante. Este enfoque permite identificar patrones de severidad que no emergen cuando estas variables se analizan de forma aislada, revelando cómo el contexto territorial y la tipología del siniestro condicionan la gravedad final de las lesiones.

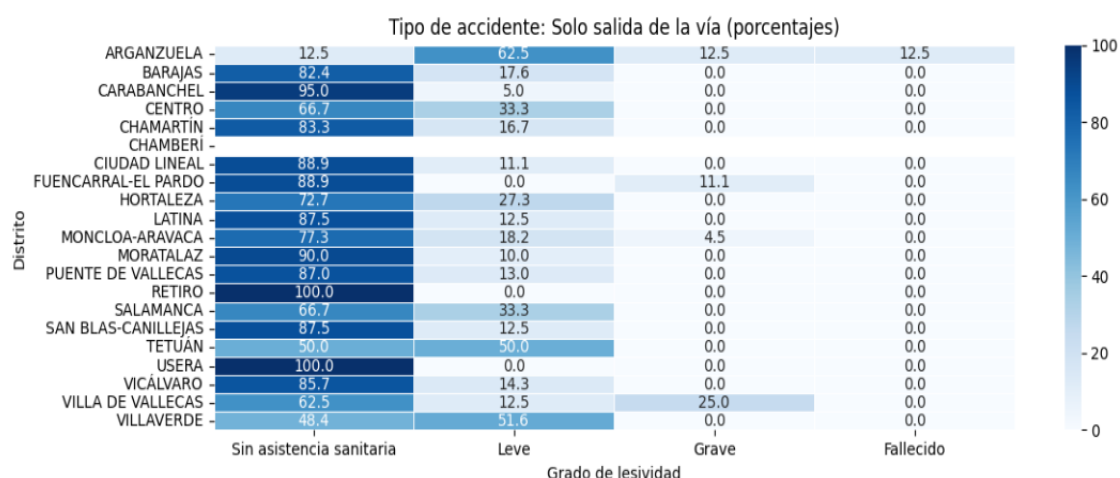


Figura 15. Solo salida vía - Grado de lesividad - Distrito

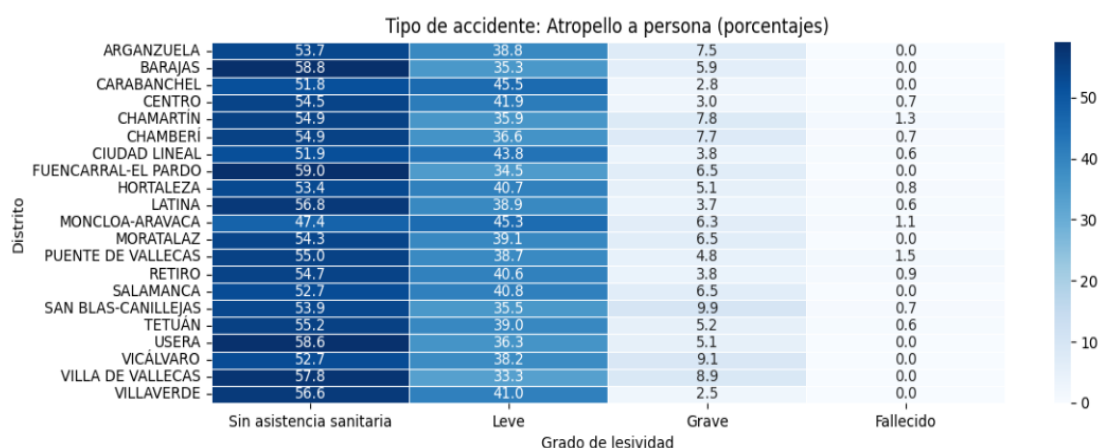


Figura 16. Atropello a persona - Grado de lesividad - Distrito

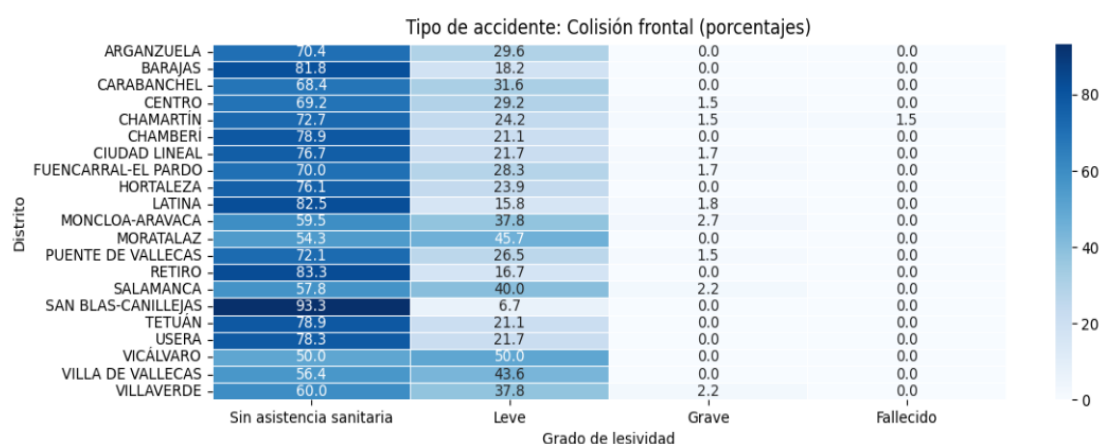


Figura 17. Colisión frontal - Grado de Lesividad - Distrito

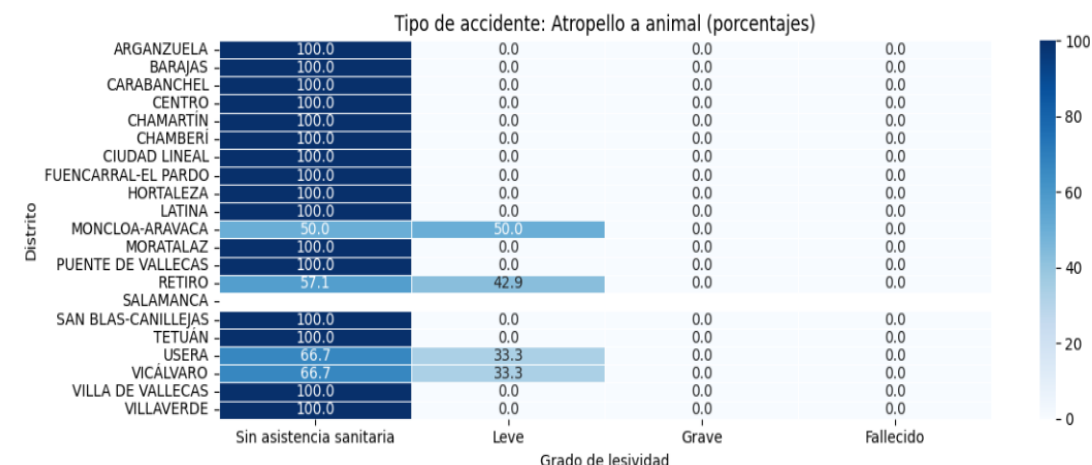


Figura 18. Atropello a animal - Grado de lesividad - Distrito

¿Cómo varía el grado de lesividad en función del distrito y del tipo de accidente?

En términos generales, se observa que:

- Predominio de accidentes leves o sin asistencia sanitaria: En la mayoría de los tipos de accidente, los porcentajes correspondientes a las categorías “Sin asistencia sanitaria” y “Leve” superan el 90 % en numerosos distritos. Esto indica que, aunque los siniestros

son frecuentes, la mayor parte no deriva en lesiones graves. Este patrón se mantiene de forma consistente en colisiones laterales, alcances y choques contra objeto fijo.

- Tipos de accidente con mayor gravedad: Existen, sin embargo, ciertos tipos de accidente que presentan una distribución de lesividad más severa. En particular:
 - **Atropello a persona** Muestra porcentajes significativamente más altos en las categorías “*Grave*” y “*Fallecido*”, especialmente en algunos distritos concretos. Esto sugiere un mayor riesgo asociado a este tipo de siniestro.
 - **Colisión frontal** También presenta una mayor proporción de lesiones graves y fallecimientos en comparación con otros tipos de accidente, lo que refuerza su carácter de alta severidad.
- Mayor gravedad en distritos periféricos en accidentes de salida de vía: En los accidentes clasificados como “*Salida de la vía*”, se aprecia un patrón territorial claro: distritos periféricos como **Fuencarral-El Pardo, Villa de Vallecas y Arganzuela** concentran una mayor proporción de heridos graves. Este comportamiento puede estar relacionado con características viales específicas, como mayor presencia de vías rápidas o menor densidad urbana.
- Tipos de accidente con lesividad prácticamente nula: Algunos tipos de siniestro presentan perfiles de lesividad muy bajos. Por ejemplo:
 - **Atropello a animal** En la mayoría de los distritos, este tipo de accidente muestra un 100 % de casos en la categoría “*Sin asistencia sanitaria*”, lo que indica que rara vez implica lesiones humanas.

4. Conclusiones

4.1 Verificación de hipótesis y principales hallazgos (año 2024)

Tras el análisis exploratorio, se contrastan las hipótesis planteadas al inicio con los resultados obtenidos.

En el análisis temporal, se concluye que los meses con mayor número de accidentes fueron octubre, noviembre y diciembre, y no enero como se había planteado. En cuanto al día de la semana, aunque los lunes presentan una frecuencia relevante, son los viernes los que presentan la mayor proporción de accidentes. Asimismo, la franja horaria de mayor siniestralidad corresponde a la tarde (12:00-18:00), siendo las 18:00 la hora punta. Por tanto, la hipótesis inicial no queda respaldada.

En el análisis geográfico, los datos muestran que los tres distritos con mayor número de accidentes fueron periféricos. No obstante, algunos de los céntricos también presentan una proporción elevada de accidentes. También se han identificado tramos concretos de las autovías A-2, A-42 y M-23 con concentraciones significativas de accidentes. Como conclusión, las hipótesis se confirman parcialmente.

En relación al análisis demográfico, se observa que los jóvenes acumulan un número elevado de accidentes, especialmente en el rango de 25 a 29 años, sin embargo, el grupo con mayor número de accidentes corresponde al rango de 45 a 49 años. Por otro lado, la hipótesis de que los hombres presentan mayor frecuencia de accidentes que las mujeres, y que la mayoría de los implicados en los siniestros son conductores, queda respaldada por los datos.

En cuanto al análisis de los factores que influyen en un accidente, se observa que el turismo es el tipo de vehículo más frecuente. Por otra parte, aunque la colisión lateral presenta un número elevado de casos, el tipo de accidente más frecuente es la colisión fronto-lateral. Además, se respalda la hipótesis de que el estado de ebriedad del conductor está relacionado con el tipo de accidente, mediante el test Chi-cuadrado.

Sobre la gravedad del accidente, predominan los accidentes leves o sin lesiones visibles, lo que confirma la hipótesis de partida. Además, el test Chi-cuadrado confirma que está estadísticamente relacionado con el rango de edad.

Adicionalmente al contraste de las hipótesis planteadas, los resultados del análisis indican que la distribución de los tipos de accidente no es homogénea entre los distintos distritos. También se observa que entre los distritos con mayor densidad de accidentes por km², se encuentran algunos de los céntricos como Salamanca, Centro, Chamartín, Chamberí y Tetuán. Por otro lado, se confirma que la ebriedad del conductor está influenciada por la franja horaria y por la edad, de acuerdo con los resultados del test Chi-cuadrado. Además, que el grado de lesividad de los implicados está estadísticamente relacionado con el tipo de accidente, y que su distribución varía según el tipo de persona (conductor, pasajero o peatón), en función del sexo y la edad de las personas implicadas, y en función del distrito y del tipo de accidente.

En conclusión, los resultados evidencian que la accidentalidad intraurbana en la ciudad de Madrid presenta patrones temporales, geográficos y demográficos relevantes, así como relaciones significativas con la lesividad y el contexto del accidente.

4.2 Conclusiones generales de 2020 a 2024

Del análisis temporal se concluye, que eventos externos, como lo fueron el COVID y la Filomena, impactan de forma directa en la siniestralidad, disminuyendo el número de accidentes. En ausencia de estos eventos, se observa un patrón estacional con picos de mayo a junio, y de octubre a noviembre, y una bajada drástica en agosto, mes vacacional.

Asimismo, se observa una estabilidad estructural en la distribución territorial, siendo los distritos céntricos los que concentran mayor siniestralidad por su alta densidad poblacional y de tráfico.

Sobre los tipos de accidentes, la colisión fronto-lateral, el alcance y la colisión lateral son los tres más comunes, en ese orden, durante los cinco años y con tendencia ascendente. Estos tres tipos representan más del 60 % de los accidentes únicos, lo que sugiere que la mayoría de siniestros ocurren en contextos de interacción entre vehículos en movimiento. Por otra parte, ha habido un descenso progresivo en accidentes por choque contra obstáculo fijo y sostenido en caídas.

En cuanto a las personas implicadas en un accidente, los conductores presentan la mayor siniestralidad con más de un 80 %. Por otro lado, desde 2020 el porcentaje de pasajeros ha crecido ligeramente y en cuanto a los peatones, aunque presentan un peso relativo bajo, son los más vulnerables.

Del análisis demográfico, se concluye que los conductores jóvenes y adultos son los más implicados en accidentes. Por otro lado, los grupos de a partir de los 60 años, mantienen porcentajes bajos pero constantes, lo cual es relevante para políticas de conducción sénior, revisión médica y adaptación de infraestructura. Asimismo, se observa un descenso progresivo en el grupo de 25 a 29 años, que podría estar relacionado con cambios en hábitos de movilidad

(más uso de transporte público, VTC o bicicleta) o con factores socioeconómicos que retrasan el acceso al vehículo propio.

En relación a la lesividad, las personas que no han requerido asistencia sanitaria representan la mayoría de los casos cada año, con un crecimiento sostenido, excluyendo 2021. La gran cantidad de casos en esta categoría, podría indicar un subregistro de lesiones leves, de las cuales ha habido un descenso, salvo en 2021. Los fallecimientos y las lesiones graves presentan bajas incidencias, pero críticas.

Por último, se observa que en los últimos años hay un ligero descenso de positivos en pruebas de alcohol, lo que podría reflejar mayor control policial, campañas de concienciación o cambios de hábitos de consumo y movilidad nocturna.

4.3 Recomendaciones de prevención

Debido a la distribución geográfica de accidentes mantenida a lo largo de los años, se recomienda mayor intervención territorial prioritaria en distritos con mayor densidad de accidentes y en puntos negros recurrentes. Asimismo, es aconsejable un mayor refuerzo de control y vigilancia en franjas de tarde, noche y madrugada, donde se concentran los accidentes más graves. La vulnerabilidad de los peatones y personas mayores, indica una necesidad de mejoras en señalización de tráfico. También es importante la prevención del alcohol al volante, intensificando controles nocturnos y campañas dirigidas a conductores jóvenes-adultos. Por último, las estrategias deberían enfocarse en mantener vigilancia sobre los casos graves y mortales, como es el caso de los atropellos y colisiones frontales.

5. Anexo

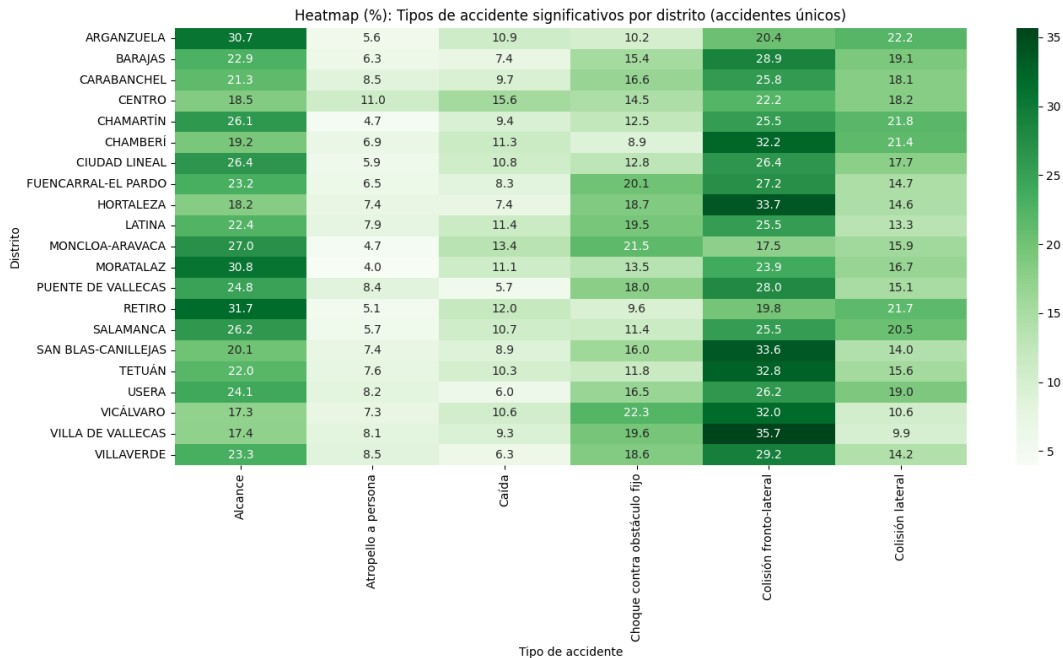


Figura 19. Tipo de accidente por distrito

Franja horaria por grado de lesividad

Se analiza cómo se distribuyen los accidentes según su gravedad en cada franja horaria.

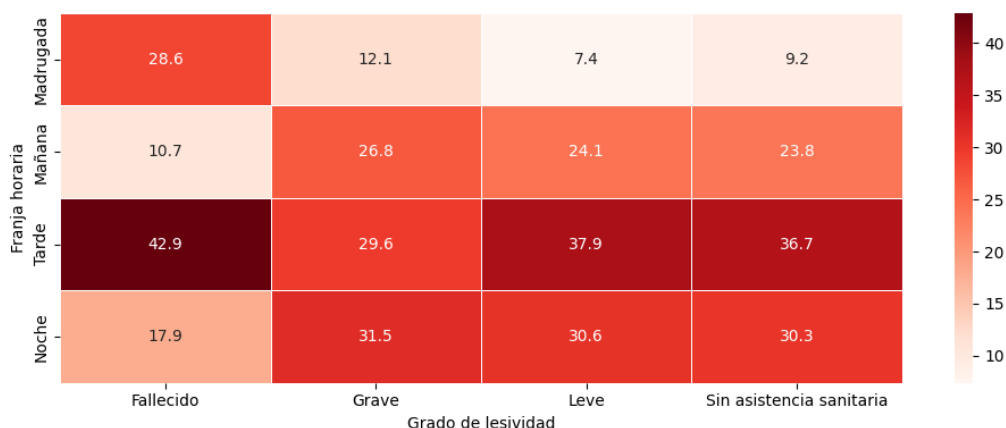


Figura 20. Franja horaria por grado de lesividad

Se observa que los fallecidos se concentran en horarios de madrugada (28,6 %) y especialmente de tarde (42.9 %). En el resto de grados de lesividad, los accidentes de mañana, tarde y noche están más distribuidos, todos presentando porcentajes altos y en particular, los accidentes de tarde en heridas leves (37,9 %) y sin asistencia sanitaria (36,7 %), y los de noche los siguen, pero en menor medida (30,6 % y 30,3% respectivamente). En conclusión, los accidentes más graves se concentran en la tarde, la madrugada y la noche, mientras que los leves y sin asistencia sanitaria predominan en la noche y en especial, en la tarde.

Se confirma mediante el test Chi-cuadrado la relación entre la franja horaria y el grado de lesividad en un accidente ($p\text{-valor} \ll 0.00001$).

Tipo de accidente por estado meteorológico

Se analiza la posible relación entre el tipo de accidente con el resultado del alcohol del conductor y con el estado meteorológico por accidentes únicos. Para ello se considera los tipos de accidentes más comunes, y en el caso del estado meteorológico igual.

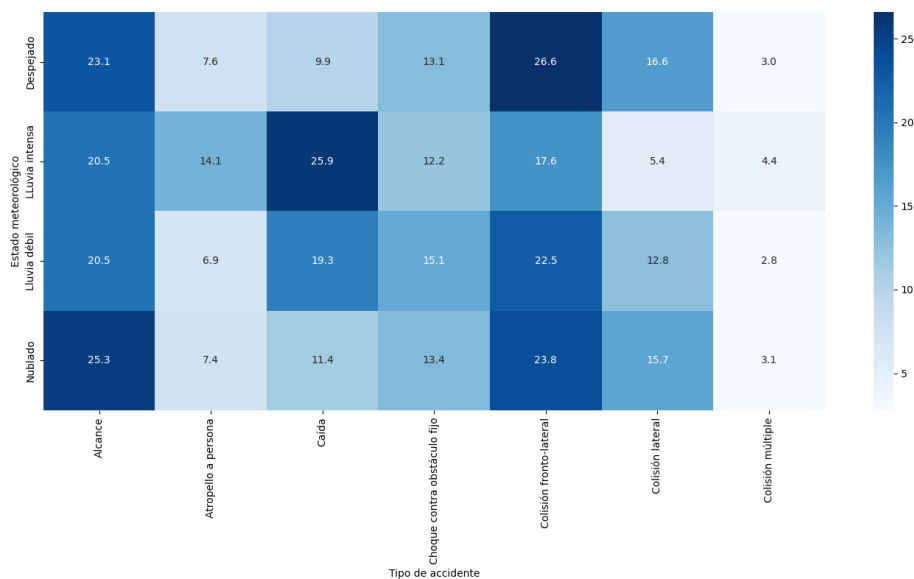


Figura 21. Tipo de accidente por estado meteorológico

Se analiza la distribución de los tipos de accidentes según los siguientes estados meteorológicos: despejado, nublado, lluvia débil o lluvia intensa. En la **Figura 21**, se observa que, en los cuatro estados, los tipos de accidentes más presentes son por alcance y colisión fronto-lateral. En los días nublados el porcentaje de alcance, caídas y colisiones laterales es mayor que en los días despejados. En las caídas, la mayor proporción se encuentra en días de lluvia débil y en días de lluvia intensa, superando a los días despejados. Además, los días de lluvia intensa presentan la mayor proporción de atropello a personas.

El test Chi-cuadrado confirma que el tipo de accidente está relacionado tanto con el consumo del alcohol como con el estado meteorológico.