

ESTUDIO SOBRE LA RELACIÓN ENTRE LAS CARACTERÍSTICAS DE UNA MOTOCICLETA Y EL RIESGO VIAL

PROYECTO DESARROLLADO POR: Applus+ IDIADA

PARA: Observatorio Nacional de Seguridad Vial

Dirección General de Tráfico (DGT)

Ministerio del Interior

Expediente: 3DGT6A000172

Realizado por:

Aprobado por:

Diego Martin
ML Software Engineer
Digital Business

Genís Mensa
R+D Manager
Body and Passive Safety

Fecha de entrega: 04/12/2024

Este informe contiene 46 páginas incluida esta portada y los anexos.

RESUMEN EJECUTIVO

De acuerdo con los datos publicados por la Dirección General de Tráfico (DGT) en los últimos años se ha observado una disminución destacable en el número de víctimas mortales y heridos graves en accidentes de tráfico. Este descenso es particularmente significativo si se considera el continuo crecimiento del parque automovilístico en España. Comparativamente, España se posicionó como uno de los países de la Unión Europea con mayor reducción proporcional de fallecidos, aunque a nivel global el objetivo no se cumplió en su totalidad.

Sin embargo, esta evolución favorable no se ha reflejado de la misma manera en el caso de los motoristas. Aunque en el pasado se registró una considerable disminución en el número de víctimas mortales en accidentes de motocicleta, en los últimos años la tendencia había sido contraria, con un aumento en las cifras, a diferencia del comportamiento general del resto de vehículos. Por esta razón, en 2009 la DGT encargó a IDIADA la realización de un estudio piloto enfocado específicamente en el colectivo de motoristas.

El objetivo de este estudio fue desarrollar una metodología que permitiera evaluar el riesgo de los motoristas de sufrir accidentes mortales y graves en función de su nivel de exposición. Dicho estudio analizó perfiles diferentes de motoristas, definidos según la edad y el tipo de motocicleta conducida, identificando así aquellos perfiles con mayor probabilidad de estar expuestos a un accidente mortal.

El presente análisis aplica la metodología propuesta en el estudio previo. Al igual que el estudio anterior, este trabajo tiene como finalidad identificar los perfiles con mayor riesgo de sufrir accidentes graves no mortales en función de su exposición.

INDICE DE CONTENIDOS

1. INTRODUCCIÓN Y OBJETIVOS	7
1.1. ANTECEDENTES.....	7
1.2. OBJETIVOS.....	11
2. METODOLOGÍA DE TRABAJO.....	12
2.1. METODOLOGÍA DE ANÁLISIS DE DATOS.....	12
2.1.1. PRIMERA FASE: EVALUACIÓN DE LA INTEGRIDAD DE DATOS	12
2.1.2. SEGUNDA FASE: INTEGRACIÓN DE BASES DE DATOS EXTERNAS	13
3. DATOS	22
3.1. ANÁLISIS DE DATOS RECIBIDOS.....	22
3.2. PARQUE NACIONAL DE MOTOCICLETAS (CON KM_ANUALES DE ITV)	24
3.2.1. DISTRIBUCIÓN SEGÚN EL TIPO DE MOTOCICLETA.....	24
3.2.2. DISTRIBUCIÓN SEGÚN LA ANTIGÜEDAD DE LA MOTOCICLETA	25
3.2.3. DISTRIBUCIÓN SEGÚN LA CILINDRADA DE LA MOTOCICLETA	26
3.2.4. DISTRIBUCIÓN SEGÚN EL KILOMETRAJE MEDIO ANUAL RECORRIDO POR LA MOTOCICLETA.....	27
3.3. ACCIDENTES DE MOTOCICLETA EN ESPAÑA.....	28
3.3.1. DISTRIBUCIÓN SEGÚN EL TIPO DE MOTOCICLETA.....	28
3.3.2. DISTRIBUCIÓN SEGÚN LA ANTIGÜEDAD DE LA MOTOCICLETA	29
3.3.3. DISTRIBUCIÓN SEGÚN LA CILINDRADA DE LA MOTOCICLETA	30
3.3.4. DISTRIBUCIÓN SEGÚN EL KILOMETRAJE MEDIO ANUAL RECORRIDO POR LA MOTOCICLETA.....	31
4. RESULTADOS	33
4.1. RIESGO EN FUNCIÓN DE SU EXPOSICIÓN.....	33
4.1.1. TASA DE HERIDOS DE GRAVEDAD VS KILOMETRAJE ANUAL.....	33
4.1.2. TASA DE MORTALIDAD VS KILOMETRAJE ANUAL	34
4.2. INDICE DE RIESGO EN FUNCIÓN DE SU EXPOSICIÓN (IREx)	35
4.2.1. PERFILES DE MOTORISTAS CON UN IREx DE HERIDOS GRAVES MÁS ALTO EN ESPAÑA.....	36
4.2.2. PERFILES DE MOTORISTAS CON UN IREx DE MORTALIDAD MÁS ALTO EN ESPAÑA.....	37
4.2.3. IREx CORRESPONDIENTES A MOTOCICLETAS TIPO SPORT	38
4.2.4. IREx CORRESPONDIENTES A MOTOCICLETAS TIPO NAKED	39
5. CONCLUSIONES	40
ANEXO 1: CLASIFICACIÓN DE LOS TIPOS DE MOTOCICLETAS.....	41
ANEXO 2: EXPOSICIÓN AL RIESGO DE LOS MOTORISTAS EN ESPAÑA.....	43

INDICE DE FIGURAS

Figura 1. Víctimas mortales de accidentes de tráfico en la UE en 2022, según usuario y vehículo involucrado en el accidente. Fuente: CARE

Figura 2. Proporción de motociclistas fallecidos respecto al número total de víctimas mortales, por países de la UE27 y la AELC (2022). Fuente: CARE

Figura 3. Evolución de la distribución de personas fallecidas y heridas hospitalizadas por medio de desplazamiento. España, 2013-2022. Fuente: DGT

Figura 4. Evolución de las personas fallecidas por medio de desplazamiento. Base 2013 = 100. España, 2013-2022

Figura 5: Distribución niveles de confianza clasificación

Figura 6: Diagrama proceso ETL datos parque+accidentadas

Figura 7: Distribución del parque nacional de motocicletas durante 2024 según el tipo

Figura 8: Distribución del parque nacional de motocicletas durante 2024 según la antigüedad

Figura 9: Distribución del parque nacional de motocicletas durante 2024 según la cilindrada

Figura 10: Distribución del parque nacional de motocicletas durante 2024 según el kilometraje medio anual recorrido por la motocicleta

Figura 11: Distribución de los accidentes de motocicleta durante 2024 según el tipo de motocicleta

Figura 12: Distribución de los accidentes de motocicleta durante 2024 según la antigüedad de la motocicleta

Figura 13: Distribución de los accidentes durante 2024 según la cilindrada de la motocicleta

Figura 14: Distribución de los accidentes de motocicleta durante 2024 según el kilometraje medio anual recorrido por la motocicleta

Figura 15: Tasa de heridos graves según kilometraje. Cada punto representa la intersección de categoría (color), edad del conductor (forma) y cilindrada (tamaño).

Figura 16: Tasa de víctimas mortales según kilometraje

Figura 17: Perfiles de motoristas con un IREx de heridos graves más alto en España

Figura 18: Perfiles de motoristas con un IREx de mortalidad más alto en España

Figura 19: IREx heridos graves de los perfiles correspondientes a motocicletas tipo Sport

Figura 20: IREx heridos graves de los perfiles correspondientes a motocicletas tipo Naked

Figura 21: Clasificación llevada de los tipos de motocicletas

INDICE DE TABLAS

Tabla 1: Código de valores para la interpretación de la tabla 2

Tabla 2: Tabla resumen de los motoristas heridos graves y muertos en España durante 2022

1. INTRODUCCIÓN Y OBJETIVOS

1.1. ANTECEDENTES

Según los datos de la Asociación Europea de Fabricantes de Automóviles (ACEA), el parque vehicular en 2022 en la Unión Europea (UE, 27 países) constaba de un total de 437 millones de vehículos motorizados. De este total, 252 millones son turismos, 30 millones son vehículos comerciales ligeros, 6 millones son vehículos comerciales medianos y pesados, 720.000 son autobuses, 15 millones son motocicletas, 11 millones son ciclomotores y el resto otro tipo de vehículos (tractores, grúas móviles, vehículos de bomberos, etc.).

En ese mismo 2022, más de 20.600 personas perdieron la vida en accidentes de tráfico en los países de la Unión Europea. Ante esta situación, la Comisión Europea ha establecido un objetivo ambicioso: reducir a la mitad el número de víctimas mortales en carretera para el año 2030, tomando como referencia los datos de 2019. Hasta ahora, se ha logrado un avance significativo, con una disminución del 9% en las muertes por accidentes de tráfico en la UE-27 entre 2019 y 2022. Esta tendencia positiva se observa en casi todos los países miembros. Si ampliamos la perspectiva temporal, desde 2012 la reducción alcanza el 22%.

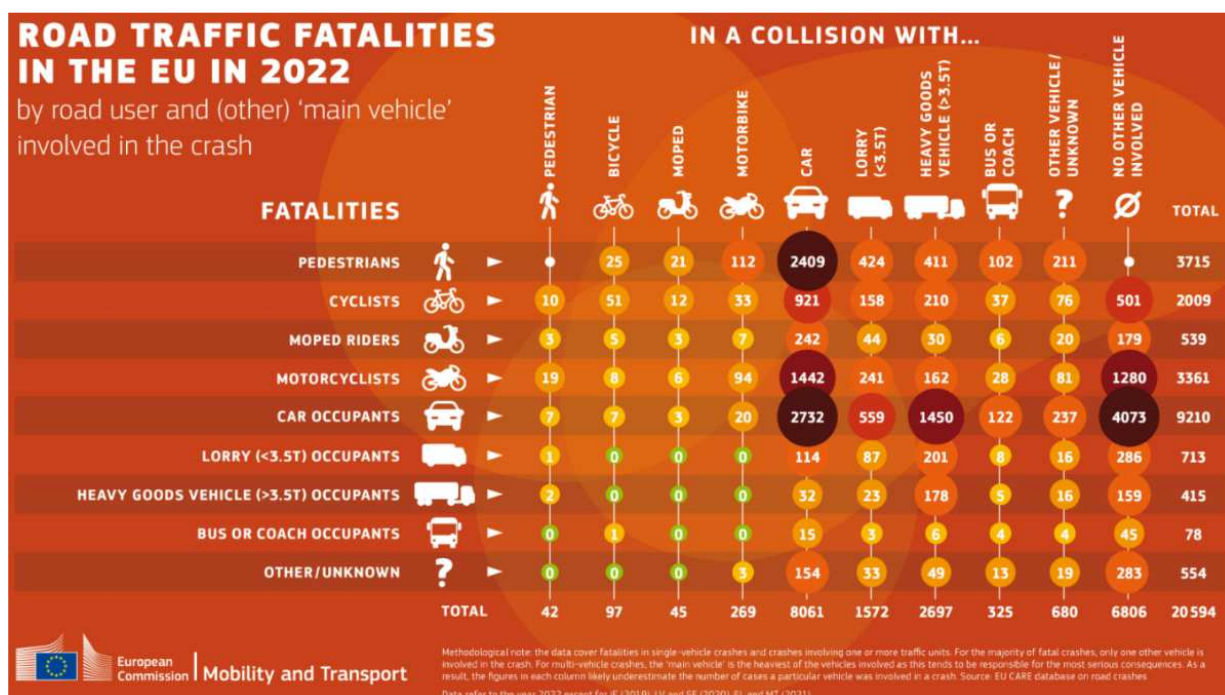


Figura 1. Víctimas mortales de accidentes de tráfico en la UE en 2022, según usuario y vehículo involucrado en el accidente. Fuente: CARE

Los ocupantes de automóviles (conductores y pasajeros) representaron el 45% de todas las víctimas mortales del 2022 en UE-27, mientras que los peatones supusieron el 18%, los usuarios de vehículos de dos ruedas con motor (motocicletas y ciclomotores) el 19% y los ciclistas el 10%. En la **Figura 1**, se puede observar el total de víctimas mortales en carreteras de la UE, en función del vehículo usado y el otro vehículo involucrado en el accidente.

España es el tercer país, por encima de la media en Europa, en tener una mayor tasa de mortalidad en motociclistas sobre el total de accidentes mortales en carreteras, tal y como indica la **Figura 2**.

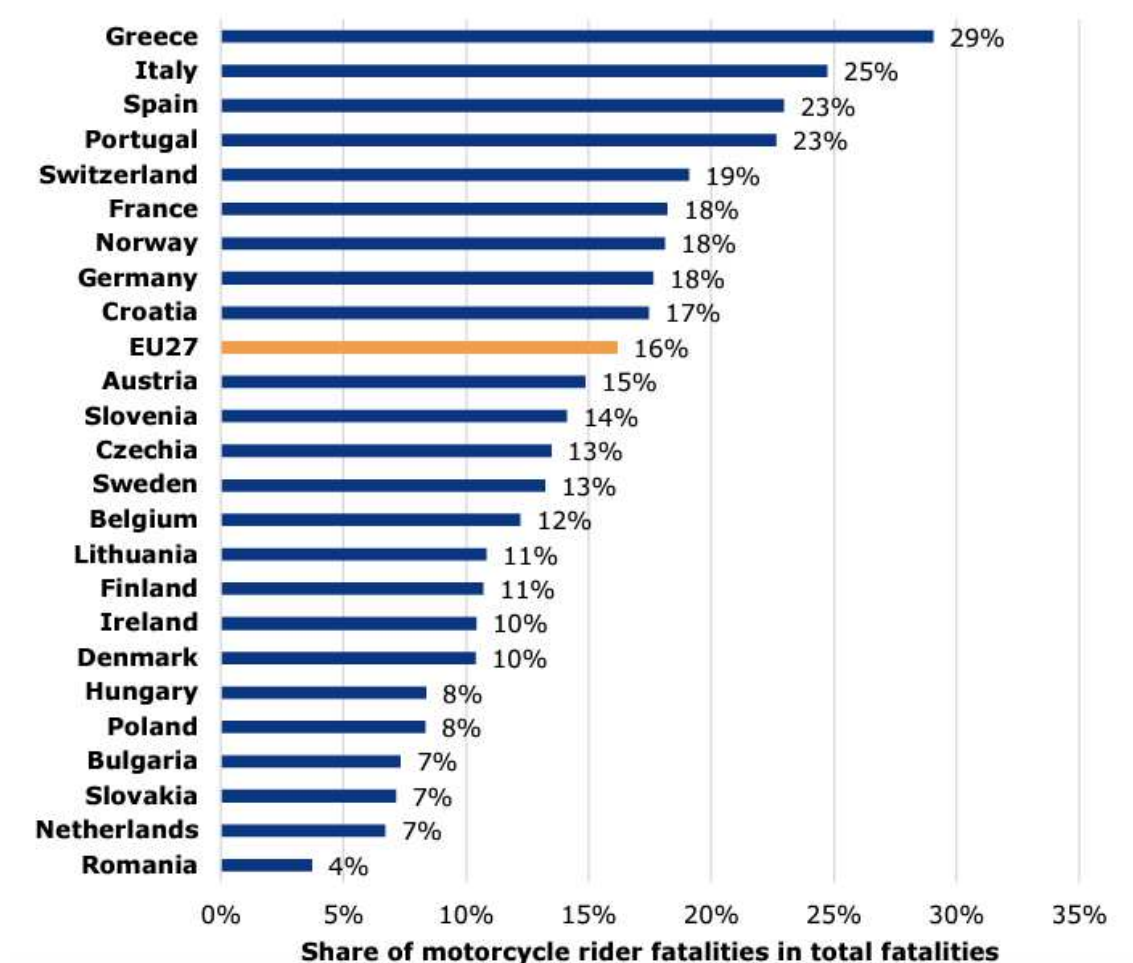


Figura 2. Proporción de motociclistas fallecidos respecto al número total de víctimas mortales, por países de la UE27 y la AELC (2022). Fuente: CARE

Al examinar la evolución del parque automovilístico en España, se detecta un crecimiento en la cantidad total de vehículos, con un incremento anual cercano al 1%. Sin embargo, esta tendencia al alza no se distribuye de manera uniforme entre los diferentes tipos de vehículos.

Desde 2013, se ha registrado una disminución en la proporción de camiones, furgonetas y ciclomotores con respecto al total del parque. En contraste, llama la atención el notable aumento de las motocicletas. Estas han pasado de representar el 9% del parque total de vehículos en 2013 a constituir el 11% en 2022, experimentando un crecimiento anual que oscila entre 50.000 y 150.000 unidades adicionales de motocicletas.

Ante este crecimiento en el registro, se suma el dato de que, en 2022, las motocicletas encabezaron las estadísticas de siniestralidad vial, registrando el mayor porcentaje de víctimas mortales y heridos que requirieron hospitalización. Esta tendencia ha mostrado un incremento significativo en los últimos diez años, elevándose del 24% en 2013 al 30% en 2022.

No obstante, al comparar con las cifras de 2019, se aprecia una leve mejoría, con una reducción del 1%, pasando de 3.146 casos (31% del total) a 3.022 (30% del total). Los datos de personas fallecidas según tipo de vehículo usado se ven representados en la **Error! Reference source not found..**

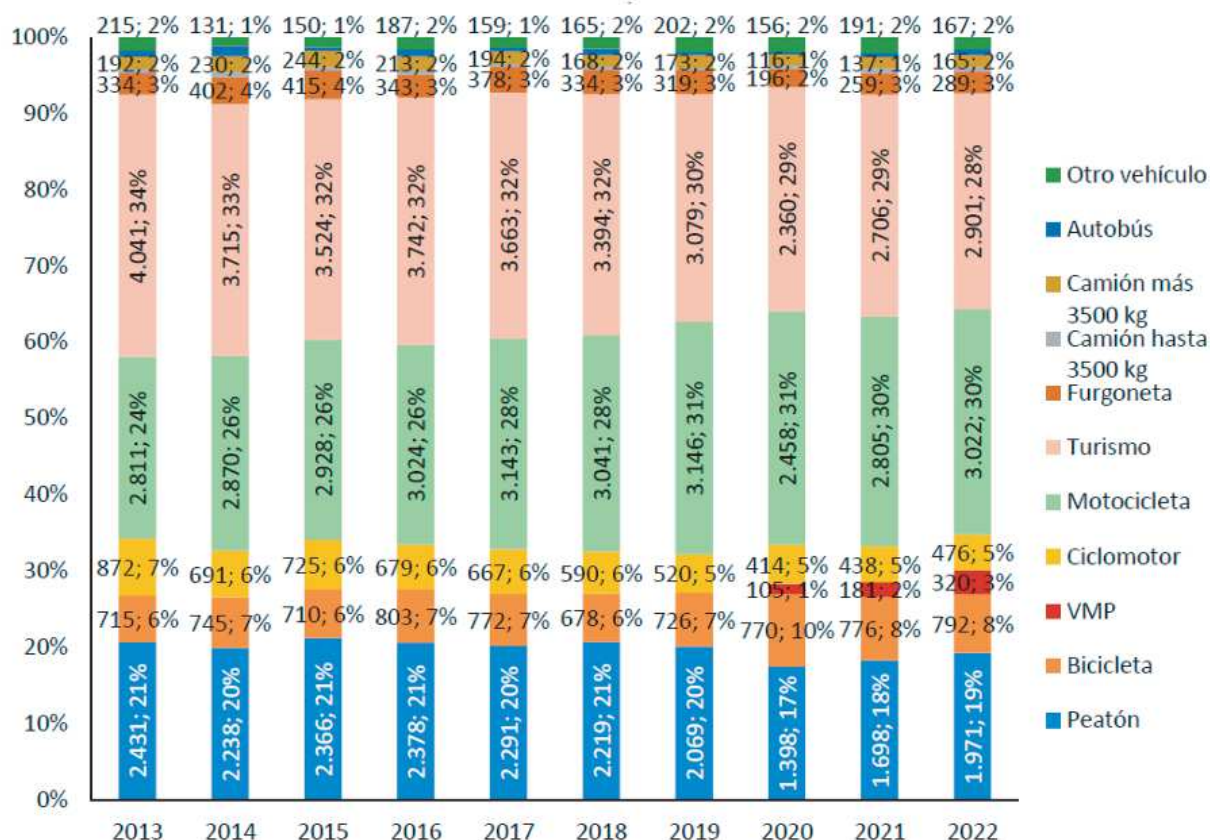


Figura 3. Evolución de la distribución de personas fallecidas y heridas hospitalizadas por medio de desplazamiento. España, 2013-2022. Fuente: DGT¹

¹ El medio de desplazamiento VMP se incluye desde el año 2020

Si nos fijamos en la evolución de las personas fallecidas según el medio de desplazamiento desde 2013 hasta 2022 (véase **Error! Reference source not found.**), se puede observar en los últimos años (2020-2022) que la tendencia no ha hecho más que aumentar para los motoristas, siendo el caso de las furgonetas y camiones hasta 3.500 kg el peor en evolución.

Cabe destacar que el descenso en fallecidos entre los años 2019 y 2020 coincide con la pandemia causada por la COVID-19, y por lo tanto la movilidad vehicular quedó restringida.

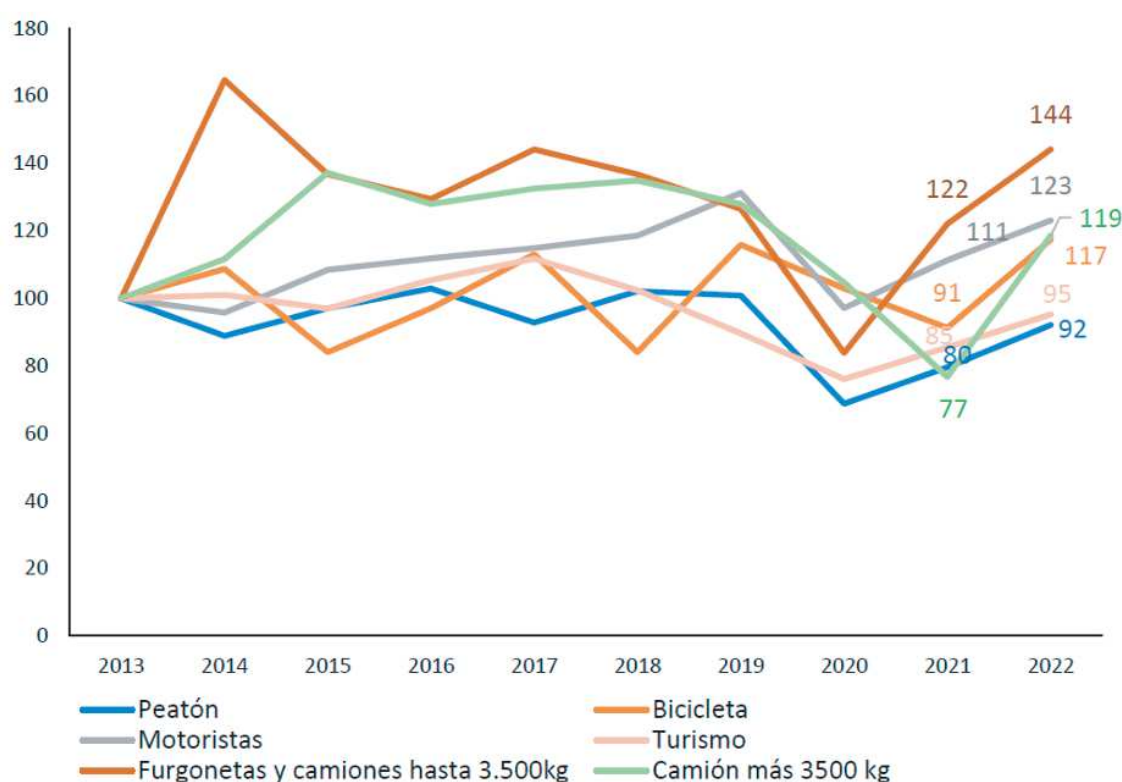


Figura 4. Evolución de las personas fallecidas por medio de desplazamiento. Base 2013 = 100. España, 2013-2022

² Las personas fallecidas en 2013 en España han sido estandarizadas a 100 para expresar los cambios en porcentaje en 2022 según el medio de desplazamiento con respecto al 2013.

1.2. OBJETIVOS

El objetivo principal de este trabajo es el de analizar los riesgos asociados al uso de motocicletas desde una perspectiva integral y basada en datos, con el propósito de generar información útil para la mejora de la seguridad vial. Para ello, el trabajo analizará tres bases de datos fundamentales correspondientes al año 2022.

El análisis de estas bases de datos permitirá realizar estimaciones detalladas tanto de riesgo como de lesividad. El riesgo se evaluará desde dos perspectivas complementarias: la probabilidad de implicación en un siniestro por cada vehículo perteneciente al parque y la probabilidad de implicación por kilómetro recorrido. Por otro lado, la lesividad se estudiará calculando el porcentaje de víctimas graves y mortales en relación con el total de víctimas, lo que proporcionará una visión más precisa del impacto de los accidentes en términos de gravedad de las lesiones.

Una parte clave del estudio será la clasificación de las motocicletas de acuerdo con una serie de variables específicas que permitan identificar diferencias en los patrones de riesgo y lesividad. Estas variables incluyen:

- la cilindrada
- la relación peso-potencia
- la antigüedad del vehículo
- el tipo de motocicleta.

En lo que respecta a esta última categoría, se empleará una clasificación detallada que busca agrupar las motocicletas en tipologías como sport, naked, scooter, touring, cruiser y trail/offroad/supermotard. Para realizar esta clasificación, se recurrirá a la identificación de la marca y el modelo de cada motocicleta registrada, con el objetivo de que al menos el 80% del parque total pueda ser clasificado según estas variables.

Este enfoque permitirá no solo identificar los factores de riesgo más relevantes, sino también comprender cómo varían en función de las características específicas de los vehículos. Al vincular los datos sobre el parque, los kilómetros recorridos y los siniestros con variables técnicas y de diseño, se podrán extraer conclusiones que contribuyan al diseño de estrategias efectivas para reducir los accidentes y sus consecuencias.

Además, el análisis aportará evidencia para la formulación de políticas públicas orientadas a mejorar la seguridad vial de los motociclistas, uno de los grupos más vulnerables en las vías. En resumen, este estudio se propone como una herramienta clave para comprender la relación entre el uso de motocicletas y la siniestralidad, y para apoyar la toma de decisiones fundamentadas en datos.

2. METODOLOGÍA DE TRABAJO

2.1. METODOLOGÍA DE ANÁLISIS DE DATOS

Durante el proceso de análisis del conjunto de datos (dataset) del parque de motocicletas, se llevó a cabo una evaluación exhaustiva de la calidad de los datos disponibles. Es importante señalar que la base de datos analizada se limitaba exclusivamente a información estática del parque de vehículos, careciendo de datos dinámicos como el kilometraje anual de los vehículos.

2.1.1. PRIMERA FASE: EVALUACIÓN DE LA INTEGRIDAD DE DATOS

En el curso del análisis de calidad, se identificaron diversas inconsistencias y anomalías en los registros, principalmente en las siguientes áreas:

- a) Nomenclatura de Marcas y Modelos:
 - Se detectaron errores tipográficos en la denominación de marcas y modelos
 - Estas inconsistencias dificultan la correcta categorización y análisis estadístico del parque de vehículos
 - Los errores incluían variaciones en la escritura de un mismo modelo/marca, uso inconsistente de mayúsculas/minúsculas y errores de tecleo
- b) Valores Anómalos en Relación Peso-Potencia (RPP):
 - Se identificaron registros con valores RPP fuera de rangos lógicos
 - Casos de $RPP = 0$, lo cual no es viable técnicamente para un vehículo operativo
 - Registros con valores infinitos (inf), indicando posibles errores en la entrada de datos o en el cálculo
- c) Cilindrada:
 - Se encontraron registros con valores de cilindrada superiores a 3000cc
 - Estos valores son altamente improbables en el segmento de motocicletas

Estas inconsistencias en los datos presentan varios desafíos para el análisis:

- Reducen la fiabilidad de los análisis estadísticos
- Requieren un proceso de limpieza y normalización de datos
- Pueden afectar a la precisión de las conclusiones derivadas del estudio

Para futuros registros y actualizaciones de la base de datos, se sugiere:

- Implementar sistemas de validación en la entrada de datos
- Establecer rangos aceptables para valores numéricos (cilindrada, RPP)
- Crear un catálogo estandarizado de marcas y modelos
- Desarrollar protocolos de verificación periódica de la calidad de los datos

2.1.2. SEGUNDA FASE: INTEGRACIÓN DE BASES DE DATOS EXTERNAS

2.2.2.1. FUENTES DE DATOS EXTERNAS

Como recurso externo se han obtenido una serie de bases de datos. Estas son las siguientes:

BIKEZ_KAGGLE: Base de datos de la web Bikez con más de 38684 modelos. En esta se guarda información técnica de las motocicletas, como pueden ser marca, modelo, potencia, tara, cilindrada, etc.

NHTSA: Base de datos de National Highway Traffic Safety Administration con más de 29164 modelos. En esta se guarda información diversa de las motocicletas, como puede ser marca, modelo, número de bastidor, fabricante, etc.

EURONCAP: Base de datos con 791 modelos obtenida de la web Euroncap (The European New Car Assessment Programme).

2.2.2.2. Proceso de Integración con NHTSA

El procedimiento consistió en:

- Cruce de código de bastidor del parque de motocicletas con datos de NHTSA
- Validación de marcas y fabricantes mediante correspondencia de registros
- Identificación de discrepancias entre ambas bases de datos

2.2.2.3. Resultados y Hallazgos

a) Aspectos Positivos:

- Se logró la corrección y validación de múltiples registros de marcas
- Se estableció un protocolo de verificación de datos mediante fuentes externas
- Se identificaron patrones de error en los datos originales

b) Limitaciones Encontradas:

- Se detectó que el código bastidor no era un campo fiable en el dataset original
- La tasa de éxito en el cruce de datos fue inferior a la esperada (no se pudieron comprobar los modelos).
- Se evidenció la necesidad de un enfoque más robusto para la limpieza de datos

Los hallazgos de esta segunda fase llevaron a la planificación de una tercera ola de trabajo, caracterizada por:

- La necesidad de desarrollar un enfoque más exhaustivo
- El requerimiento de metodologías más complejas y robustas
- La búsqueda de soluciones más integrales para la limpieza del dataset

2.2.3. TERCERA FASE: APLICACIÓN ALGORITMOS NLP E I.A

Se ha diseñado un proceso para realizar imputaciones y correcciones en las columnas de marcas y modelos de vehículos de un dataset, utilizando técnicas de búsqueda semántica, algoritmos de similitud de cadenas y modelos de lenguaje avanzado (LLM). A continuación, se detallan las etapas de procesamiento implementadas:

1. Carga y Preprocesamiento de Datos

El proceso comienza con la carga de un dataset base que contiene información vehicular. Se realizan los siguientes pasos de preprocesamiento:

- **Selección de columnas relevantes:** Se trabaja con columnas específicas, como MODELO, DESCRIPCION_CORTA, COD_MARCA, FEC_PRIM_MAT, KW, y TARA_OBV.
- **Transformaciones de datos:**
 - Conversión de valores a mayúsculas (MODELO, DESCRIPCION_CORTA) para garantizar la uniformidad.
 - Cálculo de una nueva métrica denominada RPP (Relación Potencia-Peso) para análisis adicionales.
 - Eliminación de duplicados en el índice.
- **Renombrado de columnas:** Las columnas son ajustadas para reflejar un esquema homogéneo de nombres.

Este preprocesamiento asegura que los datos estén en un formato estandarizado para las siguientes etapas.

2. Imputación Basada en VIN (Número de Bastidor)

En esta etapa, se intenta recuperar información de marcas directamente del número de bastidor (VIN) mediante una integración con la base de datos de la **NHTSA** (National Highway Traffic Safety Administration). Los pasos son los siguientes:

1. **Extracción de datos VIN:**
 - a. Utilizando la librería NHTSA_VIN_EXTRACTOR, se obtiene información adicional como la marca (Make) correspondiente al VIN de cada registro.
2. **Imputación de marcas:**
 - a. Los registros en los que se encuentra una marca válida a partir del VIN son imputados directamente en una nueva columna denominada MARCA_SEMANTIC, y el **origen de la información se etiqueta como NHTSA VIN.**
3. **Separación de datos no imputados:**
 - a. Los registros cuyo VIN no coincide con la base de datos de la NHTSA permanecen en un conjunto denominado no_imputados para etapas posteriores.

3. Imputación Semántica con Bases de Datos Maestras

Para los registros no imputados en la etapa anterior, se utilizan **varias bases de datos maestras** con información sobre marcas y modelos. El objetivo es corregir errores ortográficos, unificar nomenclaturas y realizar imputaciones basadas en similitud semántica.

- **Masters usados:** Se incluyen las siguientes fuentes:
 - **BIKEZ_KAGGLE:** Máster generado a partir de la base de datos de Bikez (38684 modelos).
 - **NHTSA:** Master generado a partir de la base de datos de National Highway Traffic Safety Administration (29164 modelos).
 - **EURONCAP:** Máster generado a partir de datos estandarizados de pruebas de seguridad vehicular (791 modelos).
 - **MANUAL:** Un máster generado tras una exploración manual con marcas y modelos conocidos (317 modelos).
- **Técnicas de imputación:**
 - **Similitud de cadenas:** Se utiliza el algoritmo **Jaro-Winkler** para calcular métricas de similitud entre los valores de las columnas y los datos de las bases maestras.
 - **Jerarquía de imputación:** Se asigna prioridad a las bases de datos maestras más confiables (EURONCAP y BIKEZ) sobre las menos específicas (MANUAL y NHTSA).
 - **Umbrales de similitud:** Se establecen dos niveles:
 - **threshold1 (0.90):** Alta confianza en la imputación basada en similitud directa.
 - **threshold2 (0.70):** Baja confianza, donde podría activarse el soporte de modelos LLM.

4. Corrección Adicional mediante Modelos LLM

En los casos donde la similitud semántica no supera el umbral alto (**threshold1**) pero excede el umbral bajo (**threshold2**), se emplea un modelo de lenguaje avanzado (LLM) para realizar imputaciones más precisas. Este proceso incluye:

1. **Construcción de un prompt:** El LLM recibe un contexto detallado, que incluye la lista de valores válidos de la base maestra y el valor problemático del dataset.
2. **Procesamiento de corrección:** El modelo sugiere el valor más probable o indica si no encuentra una correspondencia razonable.
3. **Actualización del dataset:** Los resultados del LLM se integran a los datos imputados, y el origen de la corrección se etiqueta como LLM.

5. Sustitución de Pseudónimos

Algunas marcas y modelos tienen variaciones comunes o pseudónimos (nombres populares) en el dataset. Se realiza un ajuste manual basado en un diccionario predefinido, donde se reemplazan ciertos modelos por su equivalente estándar. Por ejemplo:

- **HONDA:** WW125A se reemplaza por SH125AI o PCX 125, según corresponda.
- **YAMAHA:** YP125R se reemplaza por XMAX 125.

6. Imputación Jerárquica de Modelos

Con las marcas ya imputadas, se procede a realizar la imputación de modelos. El proceso sigue una lógica jerárquica similar a la etapa de marcas, pero ahora los maestros se filtran por la marca imputada, reduciendo la cantidad de comparaciones necesarias. Este enfoque incluye:

1. **Separación por marca:** Los registros se agrupan según su marca imputada.
2. **Filtrado de maestros:** Las bases maestras se limitan a modelos correspondientes a la marca imputada.
3. **Imputación semántica:** Se utilizan los mismos umbrales (threshold1 y threshold2) y algoritmos de similitud para completar la columna MODELO.

7. Resultados y Separación Final del proceso de saneamiento de datos

Al final del pipeline, los datos se dividen en tres datasets principales:

- **Marcas y modelos imputados:** Registros con correcciones o imputaciones exitosas en las columnas MARCA_FINAL y MODELO_FINAL. Se documenta la fuente de cada imputación (NHTSA VIN, EURONCAP, etc.).
- **Marcas imputadas:** Registros con correcciones o imputaciones exitosas en la columna MARCA_FINAL pero no en la columna MODELO_FINAL.
- **No imputados:** Registros donde no fue posible realizar imputaciones debido a falta de coincidencias en los maestros o valores ambiguos.

2.2.4. FASE FINAL: CATEGORIZACIÓN DE LOS DATOS

En esta última fase se han realizado un conjunto de operaciones para consolidar, validar y categorizar los datos de vehículos provenientes de diversas fuentes. Este proceso permite analizar los datos de parque vehicular y vehículos accidentados en una estructura unificada. A continuación, se describe en detalle cada etapa:

1. Carga y Validación de Datos Iniciales

1. Fuentes de datos

Los datasets de vehículos (parque) y vehículos accidentados (accidentados) se cargan desde archivos CSV fruto de la fase anterior. Los registros de cada dataset se clasifican mediante una nueva columna VALIDADO con valores:

- "NADA": Sin validación.
- "MARCA": Validación parcial (solo marcas).
- "MARCA/MODELO": Validación completa (marcas y modelos).

2. Formateo de Fechas

Los datasets presentan formatos de fecha distintos. Se realizan ajustes uniformes para facilitar el análisis posterior:

- Las fechas en accidentados se convierten de MM/DD/YYYY a DD/MM/YYYY.
- Las fechas en parque ya están en el formato correcto (DD/MM/YYYY).

3. Identificación de Vehículos Accidentados

- Se introduce la columna ACCIDENTADO con valores:
 - "NO": Para registros del parque vehicular.
 - "SI": Para registros de vehículos accidentados.

4. Eliminación de Registros Duplicados

Para evitar inconsistencias, se eliminan registros duplicados al unir los datasets. Si un vehículo aparece tanto en parque como en accidentados, se prioriza el registro de accidentados.

5. Consolidación de Información de Marcas y Modelos

Se genera una versión consolidada de las columnas de marcas y modelos bajo los nombres MARCA_FINAL y MODELO_FINAL:

- Si MARCA_SEMANTIC o MODELO_SEMANTIC contienen valores nulos, se utilizan los valores originales de las columnas MARCA y MODELO.

Los registros con ambas columnas MARCA_FINAL y MODELO_FINAL vacías se eliminan del conjunto de datos (ya que no se pueden categorizar usando fuentes externas).

2. Integración de Datos Externos

1. Obtención de la categoría a partir de la base de datos de bikez:

- Se comparan las combinaciones MARCA_FINAL + MODELO_FINAL con MARCA + MODELO del dataset externo.
- Se calcula la intersección de valores comunes.

2. Extracción de categorías ad-hoc:

- Se construye un diccionario para mapear marcas y modelos a las categorías correspondientes, basado en los datos externos.
- Las categorías faltantes se procesan a parte mediante una función que usa un conjunto de categorías extraídas tras un proceso manual de exploración (web bikez más actualizada, distribuidores, asociaciones de motoristas, etc.). Se consiguen categorizar algunas más.

3. Estimación de las categorías faltantes a través de variables técnicas.

Se ha desarrollado un clasificador para la estimación de las categorías de motocicletas faltantes a partir de las variables CILINDRADA_OBV, TARA_OBV y KW. A continuación se detalla el proceso:

1. Filtrado de Datos válidos

Se identificaron y eliminaron los registros con valores faltantes en las variables clave para el análisis:

- CILINDRADA_OBV**: Capacidad del motor en cm³.
- TARA_OBV**: Peso neto de la motocicleta.
- KW**: Potencia del motor en kilovatios.
- CATEGORIA**: Etiqueta de clasificación de la motocicleta.

2. División de Datos en Conjuntos de Entrenamiento y Prueba

Los datos fueron divididos en dos conjuntos:

- Conjunto de Entrenamiento** (67% de los datos): Utilizado para ajustar el modelo.
- Conjunto de Validación** (33% de los datos, 1.310.503 motocicletas): Empleado para evaluar el desempeño del modelo.

3. Estandarización de las Variables Predictoras

Para evitar que magnitudes absolutas influyan de manera desproporcionada en el modelo, las variables independientes fueron estandarizadas.

4. Entrenamiento del Modelo: Random Forest

Se utilizó un modelo de clasificación Random Forest (RandomForestClassifier), conocido por su robustez y capacidad de manejar datos con múltiples características sin requerir un extenso preprocesamiento.

5. Evaluación del Modelo

La evaluación del modelo se realizó mediante las métricas incluidas en el **reporte de clasificación**. Proporciona métricas como precisión, recall, F1-score y soporte por clase.

Reporte de clasificación:

	precision	recall	f1-score	support
ATV	0.96	0.97	0.96	3424
Allround	0.86	0.87	0.87	31058
Classic	0.89	0.95	0.92	17740
Cross / motocross	0.84	0.81	0.82	6343
Custom / cruiser	0.92	0.87	0.90	45062
Enduro / offroad	0.93	0.92	0.93	67457
Minibike, cross	0.99	0.89	0.93	597
Minibike, sport	0.60	0.00	0.01	730
Naked bike	0.96	0.88	0.92	60160
Prototype / concept model	0.99	0.97	0.98	698
Scooter	0.96	0.99	0.97	383358
Speedway	0.67	0.67	0.67	3
Sport	0.95	0.93	0.94	61830
Sport touring	0.95	0.89	0.92	16003
Super motard	0.94	0.85	0.89	9495
Touring	0.93	0.92	0.93	18923
Trial	0.95	0.92	0.93	1787
accuracy			0.95	724668
macro avg	0.90	0.84	0.85	724668
weighted avg	0.94	0.95	0.94	724668

Como podemos apreciar, la precisión general obtenida es bastante alta. Para apoyar estos resultados se ha definido una métrica de confianza, que en esencia es la probabilidad de pertenencia de una motocicleta a la categoría asignada. En la siguiente gráfica podemos observar la distribución:

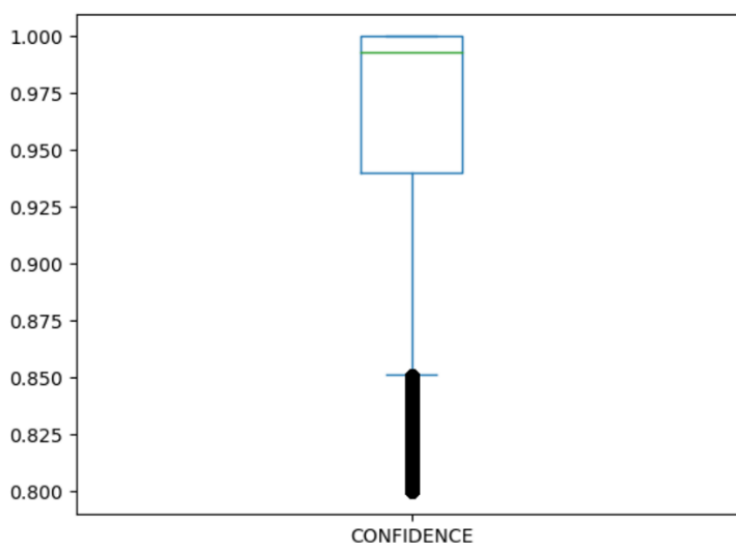


Figura 5: Distribución niveles de confianza clasificación

Un 75% de las estimaciones realizadas se hallan por encima del 92% de confianza. Es decir, 92% de probabilidades de pertenecer realmente a la categoría asignada. El 25% restantes todas se hallan por encima del 80%.

6. Métricas de Categorización del tipo de motocicleta

Se calculan indicadores clave de desempeño (KPIs) para evaluar el porcentaje de registros categorizados:

1. **Parque vehicular:**
Registros categorizados: **80%**.
2. **Vehículos accidentados:**
Registros categorizados: **86%**.

4. Discretización de Variables presentes de los datos originales

Se han aplicado un conjunto de transformaciones para categorizar variables continuas y derivar nuevas métricas basadas en los datos. Estas categorizaciones permiten analizar patrones y agrupar información clave, lo que facilita el análisis estadístico y la interpretación de los resultados.

1. Categorización de CILINDRADA_OBV

- Se agrupa la cilindrada observada (CILINDRADA_OBV) en intervalos predefinidos mediante `pd.cut`, generando la nueva columna CILINDRADA_CAT.
- Los intervalos y etiquetas asignados son:
 - **0-125cc**: Cilindrada pequeña.
 - **125-500cc**: Cilindrada media-baja.
 - **500-750cc**: Cilindrada media-alta.
 - **750cc+**: Cilindrada muy alta.

2. Categorización EDAD

- Se agrupa el rango de edades en categorías interpretativas mediante `pd.cut`, creando la columna EDAD. Los intervalos y etiquetas asignados para la variable CAT_EDAD_CONDUCTOR son:
 - **18-24 años**
 - **25-34 años**
 - **35-44 años**
 - **44+ años**

3. Categorización de FEC_PRIM_MAT

- La antigüedad del vehículo se calcula como la diferencia entre el año actual (`current_year`) y el año de la primera matriculación (FEC_PRIM_MAT).
- Con esta información, se crea la columna ANTIGÜEDAD en años.
- Posteriormente, se agrupa en rangos categóricos con la columna ANTIGÜEDAD_CAT, asignando etiquetas como:
 - **0-2 años**: Vehículo muy reciente.
 - **3-5 años**: Vehículo reciente.
 - **6-10 años**: Vehículo de mediana antigüedad.
 - **11-15 años**: Vehículo antiguo.
 - **16-20 años**: Vehículo muy antiguo.
 - **20+ años**: Vehículo clásico o en uso prolongado.

5. Diagrama de la solución

En el siguiente diagrama se puede ver el flujo de acciones hasta obtener el dataset producto de este estudio:

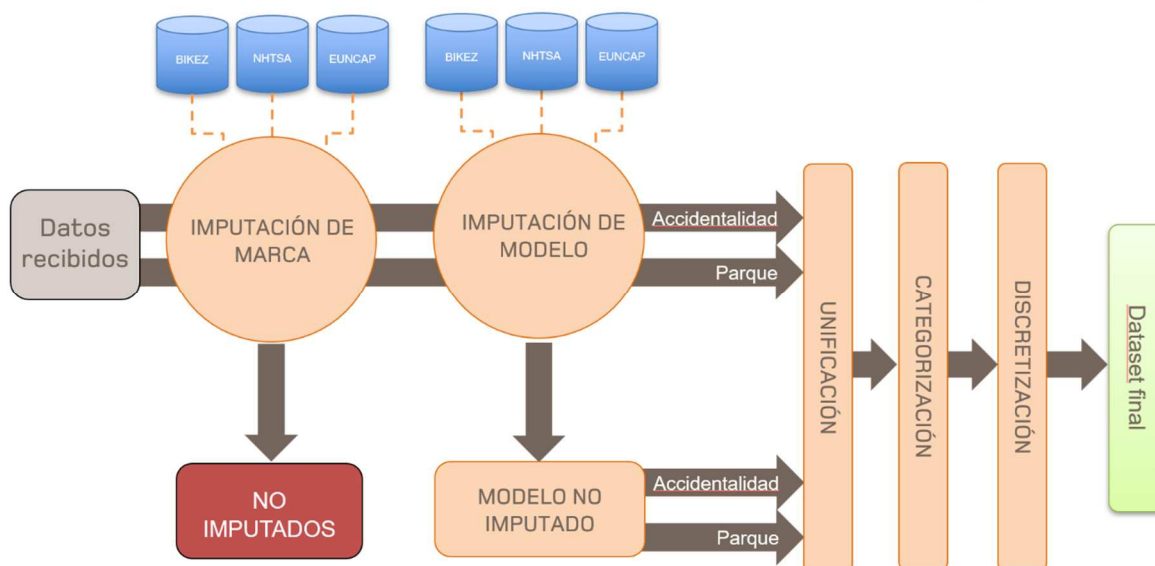


Figura 6: Diagrama proceso ETL datos parque+accidentadas

3. DATOS

3.1. ANÁLISIS DE DATOS RECIBIDOS

El estudio se fundamenta en tres conjuntos de datos esenciales, cada uno de los cuales proporciona información crucial para comprender la relación entre las características de las motocicletas y el riesgo vial. A continuación, se detalla cada una de estas bases de datos proporcionadas por la Dirección General de Tráfico:

a) Parque de motocicletas:

- Tamaño de la muestra: Aproximadamente 4.006.804 vehículos
- Descripción: Esta base de datos representa el censo completo de motocicletas registradas en España durante el año 2022.
- Contenido: Incluye información detallada sobre cada vehículo, como marca, modelo, año de fabricación, cilindrada, potencia, y peso.
- Relevancia: Estos datos son fundamentales para establecer la población total de motocicletas y servirán como denominador en los cálculos de tasas de siniestralidad.
- Limitaciones potenciales: Puede no reflejar las motocicletas que, aunque registradas, no están en uso activo.

b) Kilómetros recorridos por motocicletas:

- Volumen de datos: Aproximadamente 4.818.979 km registrados
- Fuente: Observaciones realizadas en las estaciones de inspección técnica de vehículos (ITV)
- Descripción: Esta base de datos proporciona información sobre la distancia recorrida por las motocicletas, basada en las lecturas de odómetros registradas durante las inspecciones técnicas.
- Contenido: Se espera que incluya la identificación del vehículo, la fecha de inspección, y la lectura del odómetro.
- Relevancia: Estos datos son cruciales para estimar la exposición al riesgo de las motocicletas. Permitirán calcular tasas de siniestralidad por kilómetro recorrido, ofreciendo una medida más precisa del riesgo real.
- Limitaciones potenciales:
 - Puede haber un sesgo hacia motocicletas más antiguas, ya que las nuevas no requieren inspección inmediata.
 - No captura los kilómetros recorridos por motocicletas que no han pasado la ITV en el año de estudio.

c) Motocicletas implicadas en siniestros con víctimas:

- Número de casos: Aproximadamente 30.875 vehículos
- Fuente: Presumiblemente, registros policiales y de emergencias.
- Descripción: Esta base de datos contiene información detallada sobre los siniestros viales que involucran motocicletas y que han resultado en víctimas.
- Contenido: Se espera que incluya datos como la fecha y hora del siniestro, ubicación, condiciones del entorno, características de la motocicleta implicada, información sobre el conductor y las víctimas, y la gravedad de las lesiones.
- Relevancia: Esta información es esencial para evaluar la frecuencia y gravedad de los siniestros. Permitirá analizar patrones de accidentalidad y factores de riesgo asociados a diferentes tipos de motocicletas y condiciones de conducción.
- Limitaciones potenciales: La calidad de los datos puede variar dependiendo de la precisión del registro en el lugar del siniestro.

La combinación de estas tres bases de datos proporcionará una visión integral del panorama de seguridad vial relacionado con las motocicletas. Al correlacionar la información del parque de vehículos con los kilómetros recorridos y los siniestros ocurridos, será posible:

1. Calcular tasas de siniestralidad ajustadas por exposición.
2. Identificar factores de riesgo asociados a características de las motocicletas.
3. Evaluar la influencia de edad del conductor y el uso del vehículo en la probabilidad de verse involucrado en un siniestro.
4. Analizar tendencias de seguridad en diferentes segmentos del mercado de motocicletas.

3.2. PARQUE NACIONAL DE MOTOCICLETAS (CON KM_ANUALES DE ITV)

3.2.1. DISTRIBUCIÓN SEGÚN EL TIPO DE MOTOCICLETA

En una fase anterior, las motocicletas del parque nacional en el año 2024 fueron clasificadas por tipo de motocicleta a partir de la marca y el modelo, de acuerdo con la clasificación expuesta en el Anexo 1. En la siguiente figura se muestra la distribución del parque nacional de motocicletas en 2024, según su tipo.

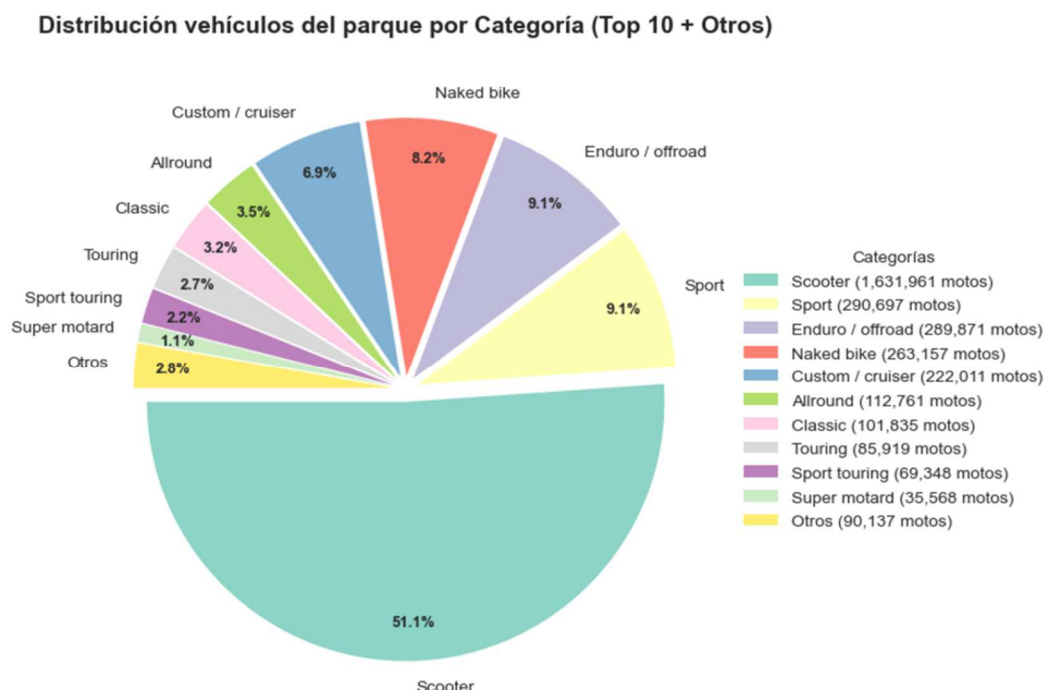


Figura 7: Distribución del parque nacional de motocicletas durante 2024 según el tipo

Se puede observar que la categoría dominante son las scooters (ciclomotores), representando a más de la mitad del parque, con el 51,1% del total de vehículos. Esto equivale a 1.631.961 motos.

Como categorías secundarias están las deportivas y las enduro/offroad, ambas con un 9,1% del mercado. Esto representa 290.697 motos para las deportivas y 289.871 para las enduro/offroad. Las motocicletas naked bike ocupan el cuarto lugar con un 8,2% (263.157 motos).

El resto de categorías son: custom/cruiser (6,9%), allround (3,5%), classic (3,2%), touring (2,7%), sport touring: (2,2%), y super motard (1,1%).

Un 2,8% de las motos no se encuentran en las categorías principales.

Esta distribución proporciona una visión clara de las preferencias de los consumidores en este mercado de motocicletas, mostrando una **fuerte inclinación hacia vehículos prácticos** para el transporte diario (scooters), **seguidos por opciones más especializadas** para diferentes estilos de conducción y preferencias de los motociclistas.

3.2.2. DISTRIBUCIÓN SEGÚN LA ANTIGÜEDAD DE LA MOTOCICLETA

En la siguiente figura se muestra la distribución del parque nacional de motocicletas en 2024, según la antigüedad del vehículo.

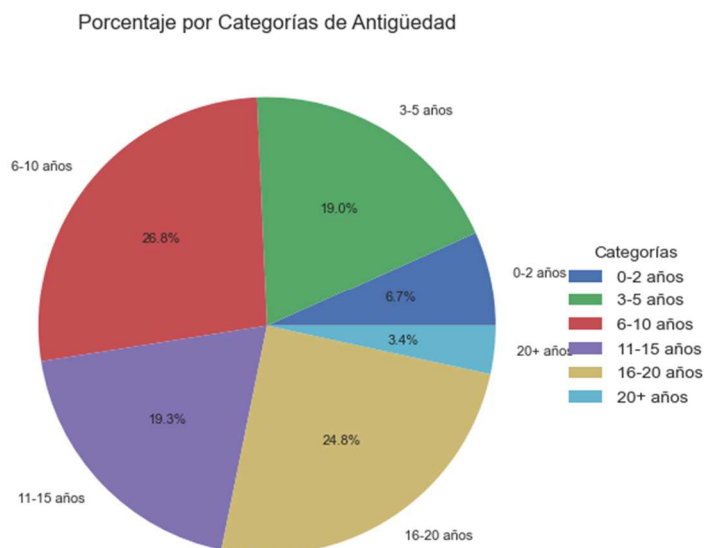


Figura 8: Distribución del parque nacional de motocicletas durante 2024 según la antigüedad

Las gráficas muestran la distribución de motocicletas en seis categorías de antigüedad, desde vehículos nuevos (0-2 años) hasta los más antiguos (20+ años).

Análisis de tendencias:

- Vehículos más nuevos (0-5 años): Representan el 25,7% del total. Esto sugiere una tasa moderada de renovación del parque de motocicletas.
- Vehículos de edad media (6-15 años): Constituyen la mayor parte del parque, sumando el 46,1%. Esto indica que una gran proporción de las motocicletas en circulación tiene una antigüedad considerable.
- Vehículos más antiguos (16+ años): Representan el 28,2% del total, lo que es una proporción significativa de vehículos antiguos aún en circulación.

Implicaciones:

- **Envejecimiento del parque:** Más de la mitad de las motocicletas (52,9%) tienen más de 10 años, lo que podría implicar desafíos en términos de emisiones y seguridad.
- **Renovación lenta:** El porcentaje relativamente bajo de motocicletas nuevas (0-2 años) sugiere una tasa de renovación del parque relativamente lenta.
- **Durabilidad:** La presencia significativa de motocicletas de más de 16 años indica que muchos vehículos tienen una larga vida útil.

3.2.3. DISTRIBUCIÓN SEGÚN LA CILINDRADA DE LA MOTOCICLETA

En la siguiente figura se muestra la distribución del parque nacional de motocicletas en 2024, según su cilindrada.

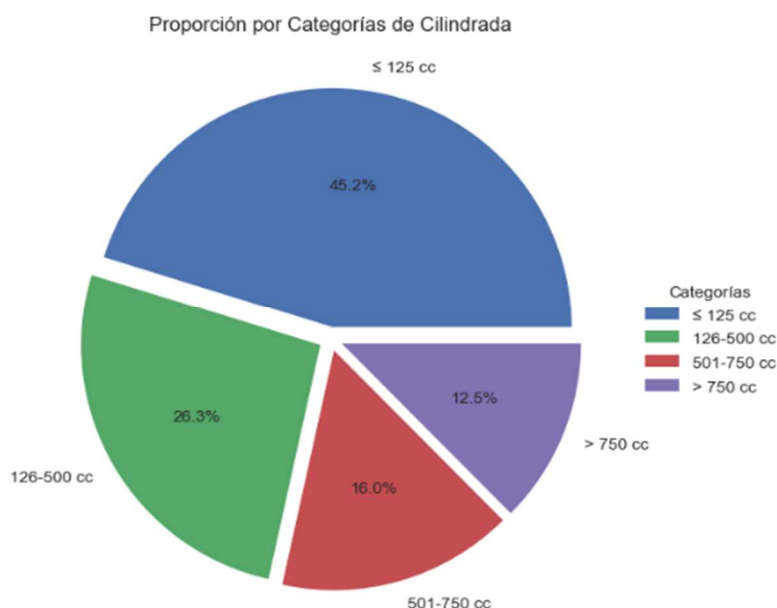


Figura 9: Distribución del parque nacional de motocicletas durante 2024 según la cilindrada

La distribución por categorías de cilindrada es la siguiente:

- 125 cc o menos: 1.215.239 vehículos (49,2%, la categoría más numerosa)
- 126-500 cc: 512.489 vehículos (20,8%)
- 501-750 cc: 376.760 vehículos (15,3%)
- Más de 750 cc: 363.873 vehículos (14,7%)

Se observa un claro **predominio de motocicletas de baja cilindrada en el mercado. Las de 125 cc o menos representan casi la mitad del parque (49,2%)**, lo que indica una fuerte preferencia por vehículos más pequeños, probablemente debido a su economía, facilidad de manejo en entornos urbanos y posibles ventajas regulatorias.

Se observa una tendencia decreciente en el número de motocicletas a medida que aumenta la cilindrada. Las categorías de mayor cilindrada (501-750 cc y más de 750 cc) suman alrededor del 30% del total. El segmento medio (126-500 cc) representa un 20,8% significativo, atrayendo a usuarios que buscan equilibrio entre potencia y practicidad.

3.2.4. DISTRIBUCIÓN SEGÚN EL KILOMETRAJE MEDIO ANUAL RECORRIDO POR LA MOTOCICLETA

Para finalizar la categorización del parque de motocicletas, en la siguiente figura se muestra la distribución del parque nacional de motocicletas en 2024, según su kilometraje anual medio.

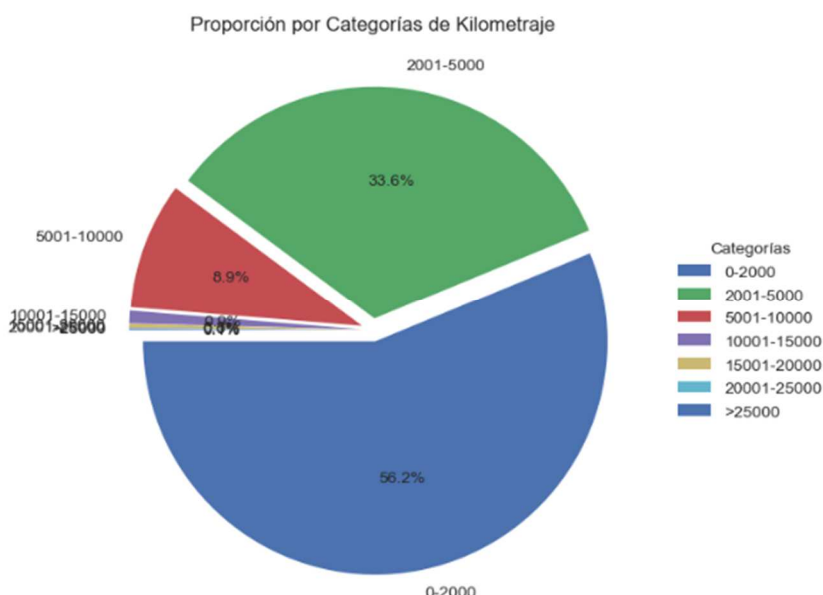


Figura 10: Distribución del parque nacional de motocicletas durante 2024 según el kilometraje medio anual recorrido por la motocicleta

La distribución por categorías de kilometraje anual es la siguiente:

- Menos de 2.000 km: 1.897.885 vehículos (56,2%, la categoría más numerosa)
- 2.000-5.000 km: 1.133.804 vehículos (33,6%, la categoría más numerosa)
- 5.000-10.000 km: 299.361 vehículos (8,9%)
- Más de 10.000 km: 43.654 vehículos (1,3%)

El análisis revela un **patrón de uso predominante de bajo kilometraje** en motocicletas. El 89,8% recorre menos de 5.000 km al año, sugiriendo un uso principalmente urbano o para trayectos cortos. La distribución muestra una tendencia fuertemente decreciente a medida que aumenta el kilometraje anual, con solo el 10,2% recorriendo más de 5.000 km al año, de los cuales un 1,3% recorre más de 10.000 km.

3.3. ACCIDENTES DE MOTOCICLETA EN ESPAÑA

3.3.1. DISTRIBUCIÓN SEGÚN EL TIPO DE MOTOCICLETA

En la siguiente figura podemos observar la distribución de los accidentes de motocicleta en España durante 2024 según el tipo de motocicleta implicado en el accidente.

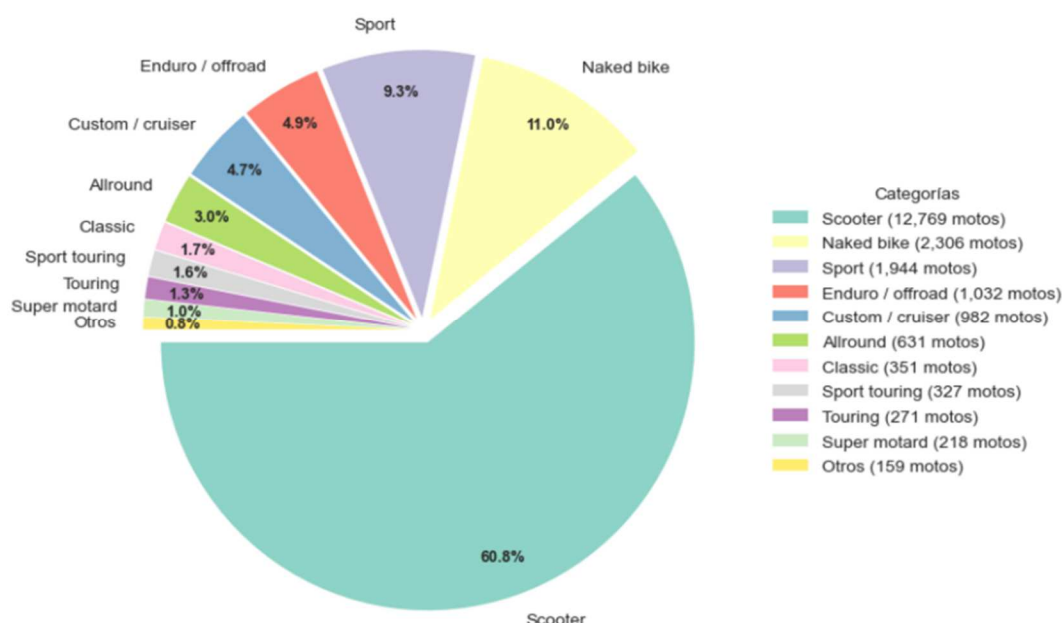


Figura 11: Distribución de los accidentes de motocicleta durante 2024 según el tipo de motocicleta

La distribución por tipo de motocicleta es la siguiente:

1. Scooters:
 - Total del parque: 51,1%
 - Accidentadas: 60,8%
 - A pesar de representar el 51,1% del parque, las scooters tienen una mayor proporción de accidentes (60,8%)
2. Naked bike:
 - Total del parque: 8,2%
 - Accidentadas: 11,0%
 - Las naked bikes están sobrerrepresentadas en accidentes (+2,8 puntos)
3. Sport (deportivas):
 - Total del parque: 9,1%
 - Accidentadas: 9,3%
 - Ligera sobrerrepresentación en accidentes (+0,2 puntos)

4. Enduro/offroad:

- Total del parque: 9,1%
- Accidentadas: 4,9%
- Significativamente subrepresentadas en accidentes (-4,2 puntos)

El análisis revela que **aproximadamente el 0,66% del parque total de motocicletas ha sufrido accidentes** (20.990 motocicletas de 3.193.265). La distribución de accidentes no es proporcional entre categorías, mostrando tendencias significativas:

1. **Las scooters, siendo mayoritarias, tienen una proporción aún mayor de accidentes**, posiblemente debido a su uso predominante en entornos urbanos con mayor exposición al tráfico.
2. **Las naked bikes y deportivas muestran tasas de accidentalidad superiores a su representación en el parque**, sugiriendo un mayor riesgo relativo, probablemente debido a su potencia y tipo de conducción.
3. **Las enduro/offroad tienen una tasa de accidentalidad notablemente inferior**, posiblemente por su uso más especializado o en entornos menos congestionados.

En resumen, aunque las scooters dominan tanto el parque total como los accidentes, las diferencias en las tasas de accidentalidad entre categorías sugieren que **el tipo de moto y su uso típico influyen significativamente en el riesgo de accidente**.

3.3.2. DISTRIBUCIÓN SEGÚN LA ANTIGÜEDAD DE LA MOTOCICLETA

En la siguiente figura se muestra la distribución de los accidentes de motocicleta en España durante 2024 según la antigüedad de la motocicleta implicada en el accidente.

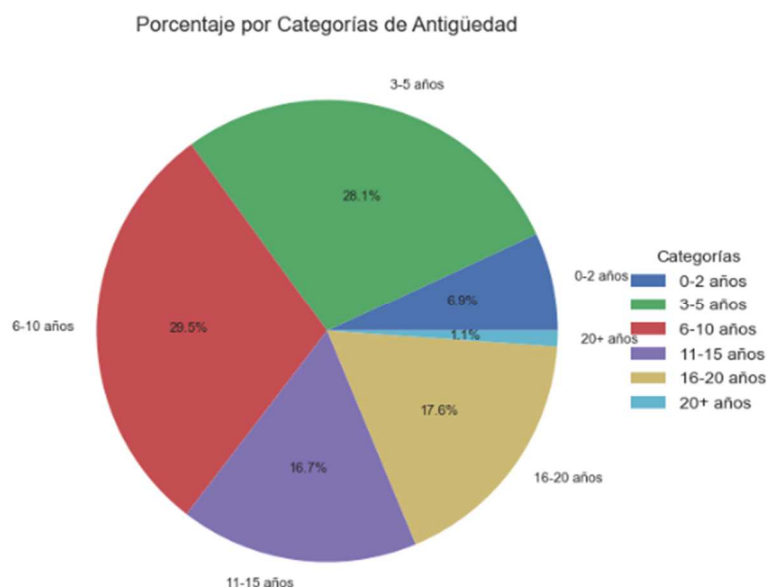


Figura 12: Distribución de los accidentes de motocicleta durante 2024 según la antigüedad de la motocicleta

Comparando la distribución por antigüedad de las motos accidentadas con el total del parque, se observan diferencias notables:

1. Motos de edad media (3-10 años):
 - Parque total: 45,8% (19,0% + 26,8%)
 - Accidentadas: 57,6% (28,1% + 29,5%)
 - Están significativamente sobrerrepresentadas en accidentes (+11,8 puntos)
2. Motos nuevas (0-2 años):
 - Parque total: 6,7%
 - Accidentadas: 6,9%
 - Mantienen una proporción casi idéntica (+0,2 puntos)
3. Motos antiguas (>10 años):
 - Parque total: 47,5% (19,3% + 24,8% + 3,4%)
 - Accidentadas: 35,4% (16,7% + 17,6% + 1,1%)
 - Están claramente subrepresentadas en accidentes (-12,1 puntos)

En resumen, **las motos de edad media (3-10 años) tienen una tasa de accidentalidad desproporcionadamente alta, mientras que las más antiguas se accidentan menos de lo que su presencia en el parque sugeriría**. Las motos nuevas mantienen una tasa de accidentalidad proporcional a su presencia en el parque.

3.3.3. DISTRIBUCIÓN SEGÚN LA CILINDRADA DE LA MOTOCICLETA

En la siguiente figura se muestra la distribución de los accidentes de motocicleta en España durante 2024 según la cilindrada de la motocicleta implicada en el accidente.

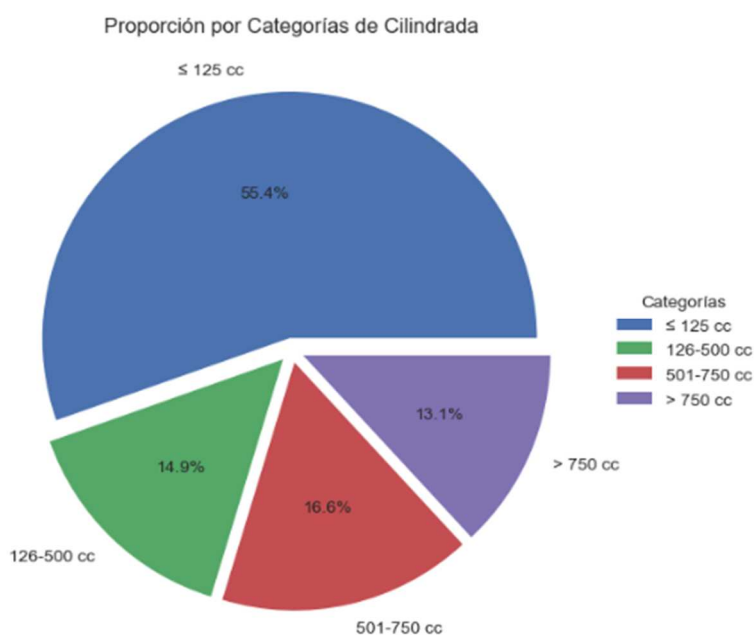


Figura 13: Distribución de los accidentes durante 2024 según la cilindrada de la motocicleta

Comparando la distribución por cilindrada de las motos accidentadas con el total del parque, observamos diferencias significativas:

1. Motos de 125cc o menos:
 - Parque total: 49,2%
 - Accidentadas: 60,1%
 - Están significativamente sobrerrepresentadas en accidentes (+10,9 puntos)
2. Motos intermedias:
 - 126-500cc: baja del 20,8% al 15,1% (-5,7 puntos)
 - 501-750cc: similar, del 15,3% al 16,5% (+1 punto)
3. Motos de alta cilindrada (más de 750cc):
 - Parque total: 14,7%
 - Accidentadas: 8,4%
 - Están significativamente subrepresentadas (-6,3 puntos)

En resumen, **las motos de menor cilindrada (125cc o menos) tienen una tasa de accidentalidad desproporcionadamente alta**, mientras que las motos de alta cilindrada se accidentan menos de lo esperado según su presencia en el parque.

3.3.4. DISTRIBUCIÓN SEGÚN EL KILOMETRAJE MEDIO ANUAL RECORRIDO POR LA MOTOCICLETA

En la siguiente figura se muestra la distribución de los accidentes de motocicleta en España durante 2024 según el kilometraje medio anual de la motocicleta implicada en el accidente.

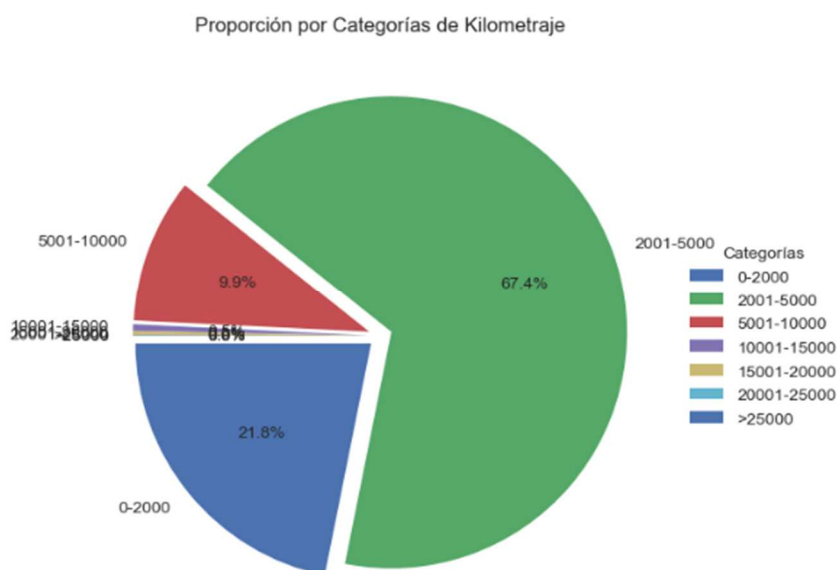


Figura 14: Distribución de los accidentes de motocicleta durante 2024 según el kilometraje medio anual recorrido por la motocicleta

Comparando los dos gráficos con el total del parque nacional, observamos las siguientes diferencias en la distribución por kilometraje:

1. Motos de menos de 2.000 km:
 - Parque total: 56,2%
 - Accidentadas: 67,4%
 - Ligera sobrerrepresentación en accidentes (+11,2 puntos)
2. Motos de 2.000-5.000 km:
 - Parque total: 33,6%
 - Accidentadas: 21,8%
 - Subrepresentadas en accidentes (-11,8 puntos)
3. Motos de 5.000-10.000 km:
 - Parque total: 8,9%
 - Accidentadas: 9,9%
 - Ligeramente sobrerrepresentadas en accidentes (+1,0 puntos)
4. Motos de alto kilometraje (más de 10.000 km):
 - Parque total: 1,3%
 - Accidentadas: 0,9%
 - Subrepresentadas en accidentes (-0,4 puntos)

En resumen, **las motos que menos kilómetros recorren al año tienen una tasa de accidentalidad ligeramente superior a su presencia en el parque**, mientras que las motos que más kilómetros recorren tienen proporcionalmente menos accidentes.

4. RESULTADOS

En este apartado analizamos el riesgo de sufrir un accidente grave o mortal en función del perfil del motorista y/o tipo de motocicleta.

Para valorar este riesgo, se han empleado tres indicadores diferentes de víctimas:

- **Total de accidentes graves o de muertes** (número de heridos graves o de víctimas mortales de cada colectivo)
- **Tasa de heridos graves o de muertes** (número absoluto dividido por el total de motocicletas del parque para dicho colectivo).
- **IREx** (Índice de Riesgo en función de la Exposición) **de heridos graves o de muertes** (número absoluto dividido por el kilometraje anual total recorrido por el todo el colectivo)

4.1. RIESGO EN FUNCIÓN DE SU EXPOSICIÓN

4.1.1. TASA DE HERIDOS DE GRAVEDAD VS KILOMETRAJE ANUAL

La tasa de heridos de gravedad se obtiene dividiendo el número total de víctimas heridas de gravedad entre el número total de motocicletas en el parque dado un colectivo del parque. Es un índice más representativo que el valor de víctimas heridas de gravedad absoluto, dado que considera el número total de motocicletas que hay en el parque para cada perfil, relativizando así los casos de heridos graves. Este índice se representa en porcentaje y es ampliamente conocido y utilizado.

Para visualizar el valor añadido del IREx, también podemos representar un gráfico de dispersión, con la tasa de heridos de gravedad (o de víctimas mortales) en el eje de las abscisas y el kilometraje en el eje de las ordenadas.

La siguiente figura muestra la tasa de heridos de gravedad en función del kilometraje anual en 2022.

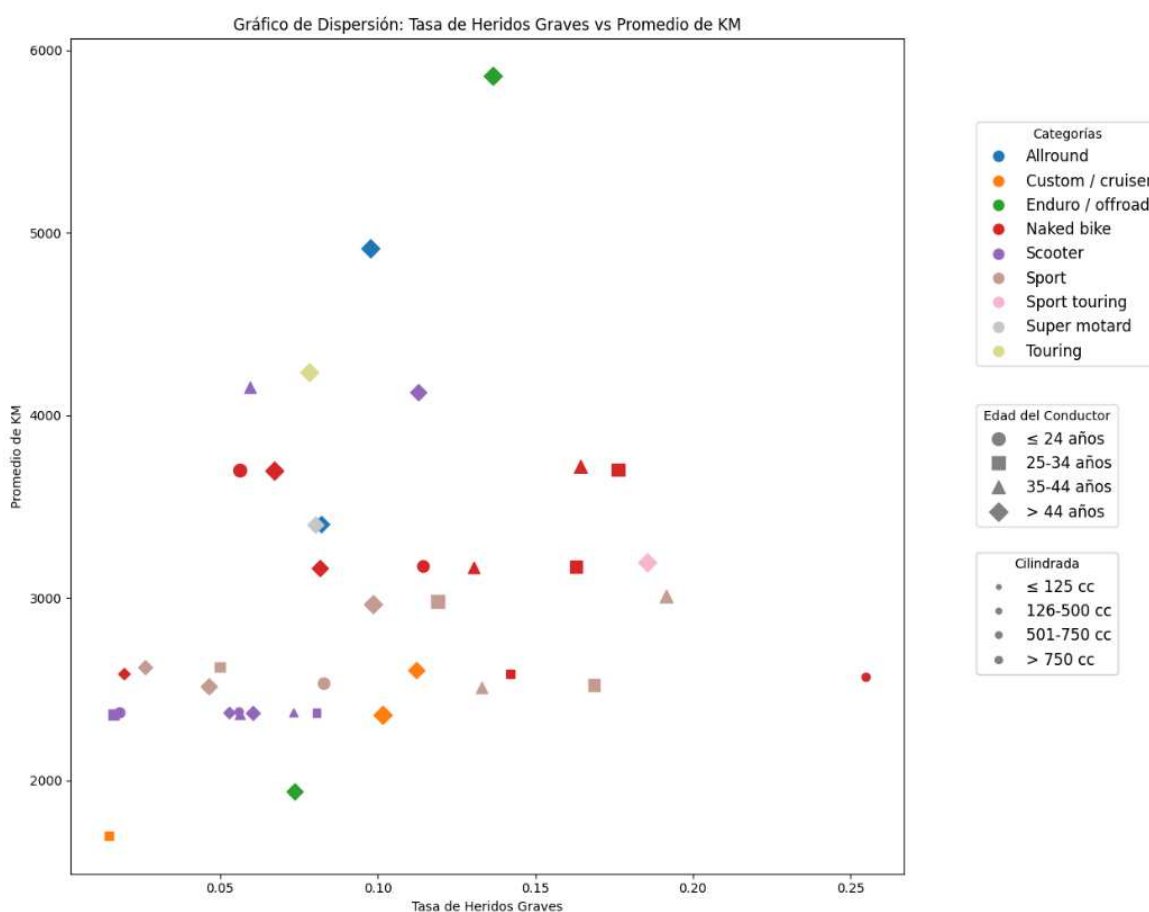


Figura 15: Tasa de heridos graves según kilometraje. Cada punto representa la intersección de categoría (color), edad del conductor (forma) y cilindrada (tamaño).

Sólo se muestran las combinaciones para las cuales hay como mínimo 62 accidentes, y al menos un accidente mortal. De manera que se evita prestar atención a colectivos que presentan unas cantidades de accidentes insuficientemente grandes como para ser considerados de riesgo.

4.1.2. TASA DE MORTALIDAD VS KILOMETRAJE ANUAL

La tasa de mortalidad se obtiene dividiendo el número total de víctimas mortales entre el número total de motocicletas en el parque dado un colectivo del parque. Es un índice más representativo que el valor de víctimas mortales absoluto, dado que considera el número total de motocicletas que hay en el parque para cada perfil, relativizando así los fallecimientos. Este índice se representa en porcentaje y es ampliamente conocido y utilizado.

Para visualizar la tasa de mortalidad en función del kilometraje, se ha optado por el mismo tipo de gráfico. La siguiente figura muestra la tasa de mortalidad en función del kilometraje anual en 2022.

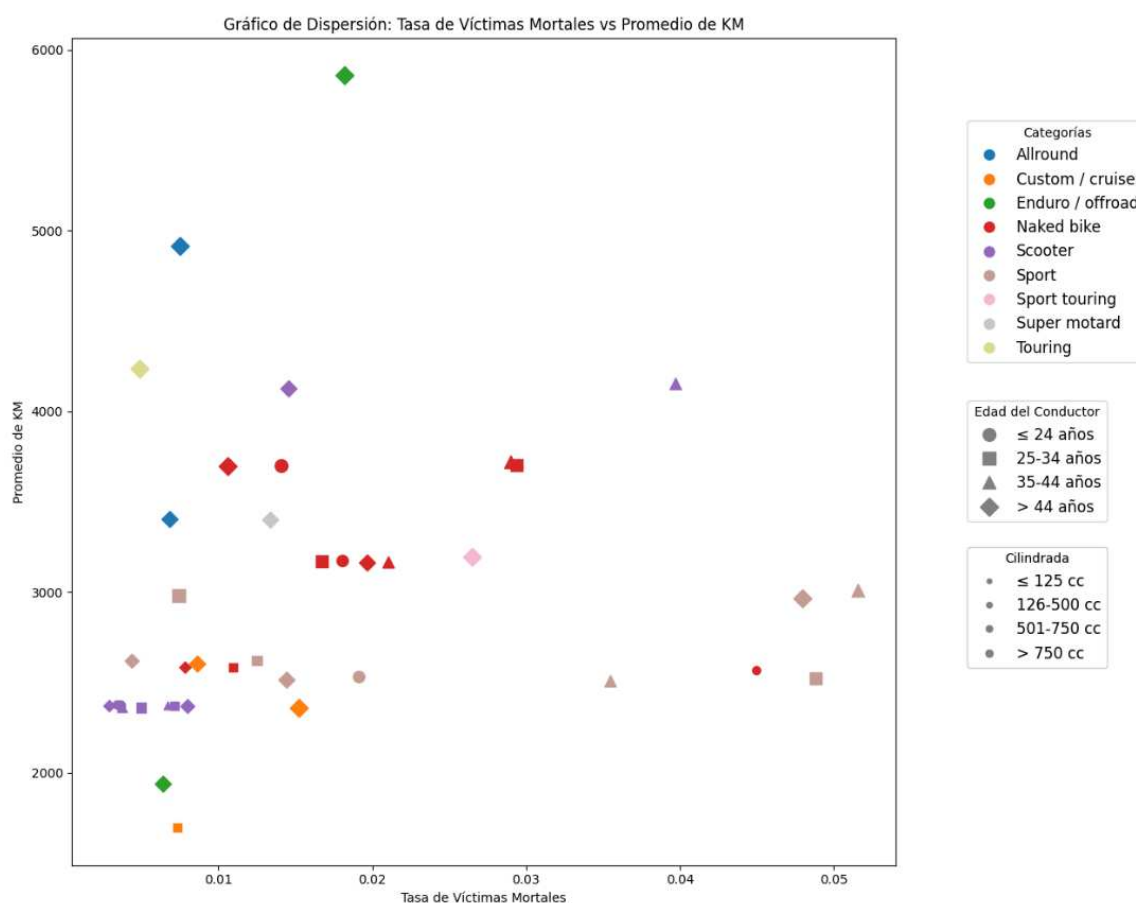


Figura 16: Tasa de víctimas mortales según kilometraje

Sólo se muestran las combinaciones para las cuales hay como mínimo 62 accidentes, y al menos un accidente mortal. De manera que se evita prestar atención a colectivos que presentan unas cantidades de accidentes insuficientemente grandes como para ser considerados de riesgo.

4.2. INDICE DE RIESGO EN FUNCIÓN DE SU EXPOSICIÓN (IREx)

El IREx es un índice desarrollado y presentado en un estudio precedente. Mejora la representatividad de la tasa de mortalidad teniendo en cuenta el kilometraje recorrido por cada perfil de conductor. La siguiente ecuación muestra cómo se calcula dicho índice, que se representa en partes por cien millones (10-8):

$$\begin{aligned} \text{IREx} &= (\text{Nº Víctimas mortales} / \text{Kilometraje anual parque}) * 100.000.000 = \\ &= (\text{Tasa mortalidad} / \text{Kilometraje anual motocicleta}) * 100.000.000 \end{aligned}$$

Ecuación 1: Índice de Riesgo en función de la Exposición

4.2.1. PERFILES DE MOTORISTAS CON UN IREx DE HERIDOS GRAVES MÁS ALTO EN ESPAÑA

Para visualizar más fácilmente los resultados y en concordancia con los grupos de riesgo definidos anteriormente, en la siguiente figura se muestran los colectivos con un riesgo de heridas graves en función de su exposición más alto observados:

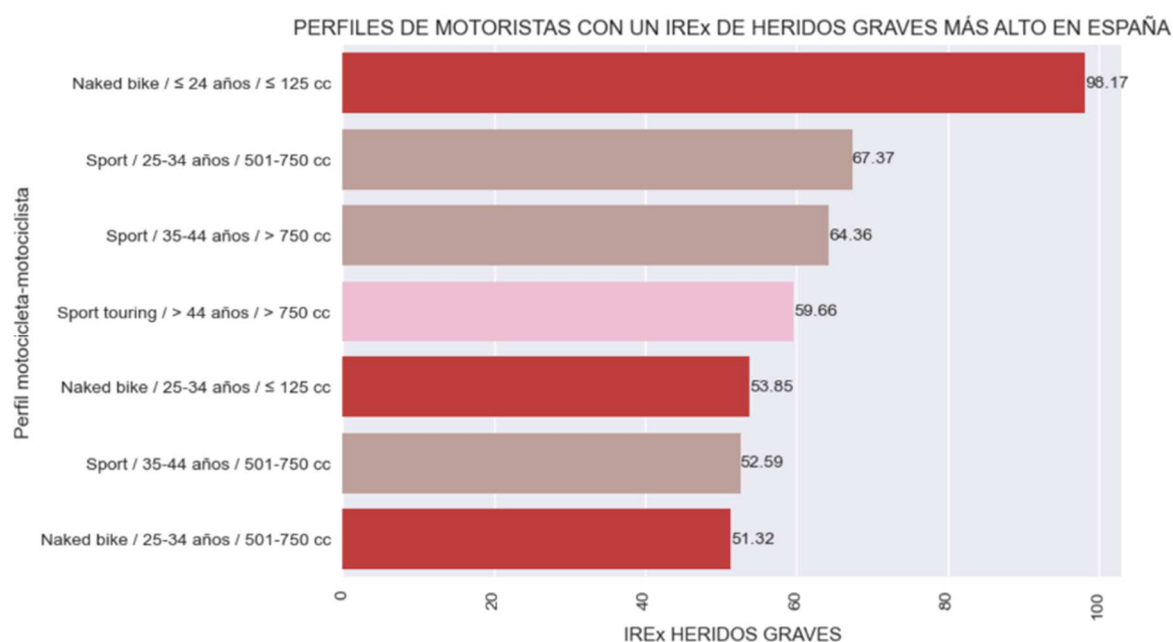


Figura 17: Perfiles de motoristas con un IREx de heridos graves más alto en España

4.2.2. PERFILES DE MOTORISTAS CON UN IREx DE MORTALIDAD MÁS ALTO EN ESPAÑA

Para visualizar más fácilmente los resultados y en concordancia con los grupos de riesgo definidos anteriormente, en la siguiente figura se muestran los colectivos con un riesgo de defunción en función de su exposición más alto observados:

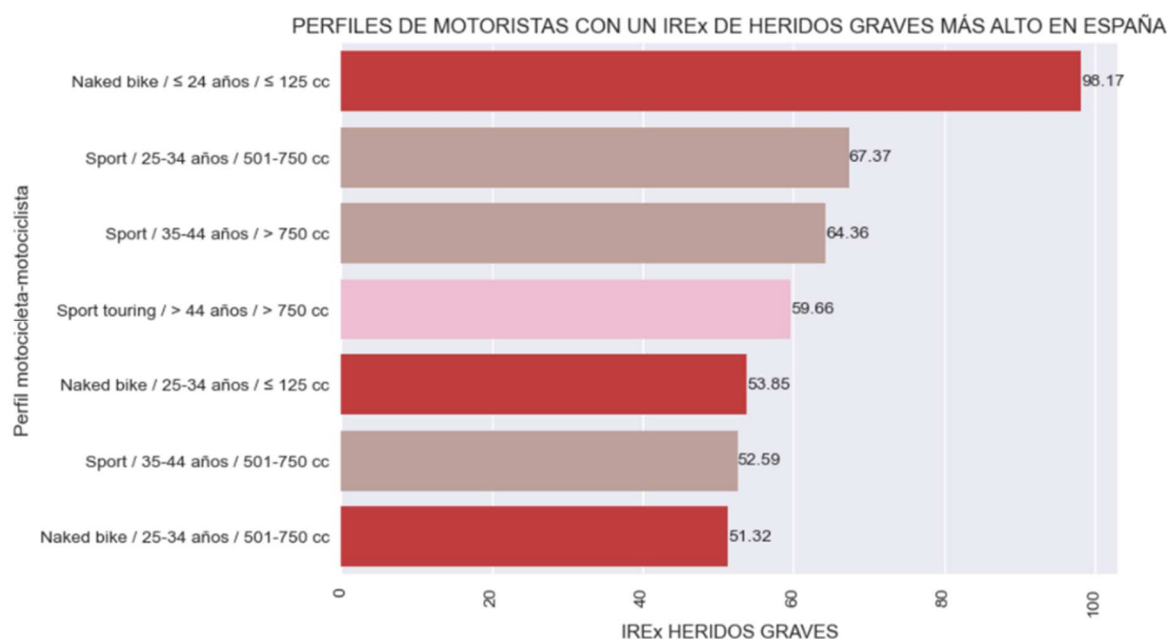


Figura 18: Perfiles de motoristas con un IREx de mortalidad más alto en España

4.2.3. IREx CORRESPONDIENTES A MOTOCICLETAS TIPO SPORT

Con el objetivo de visualizar la influencia de la cilindrada dentro de los grupos de edad dentro de la categoría con un mayor riesgo observado presentamos la siguiente gráfica:

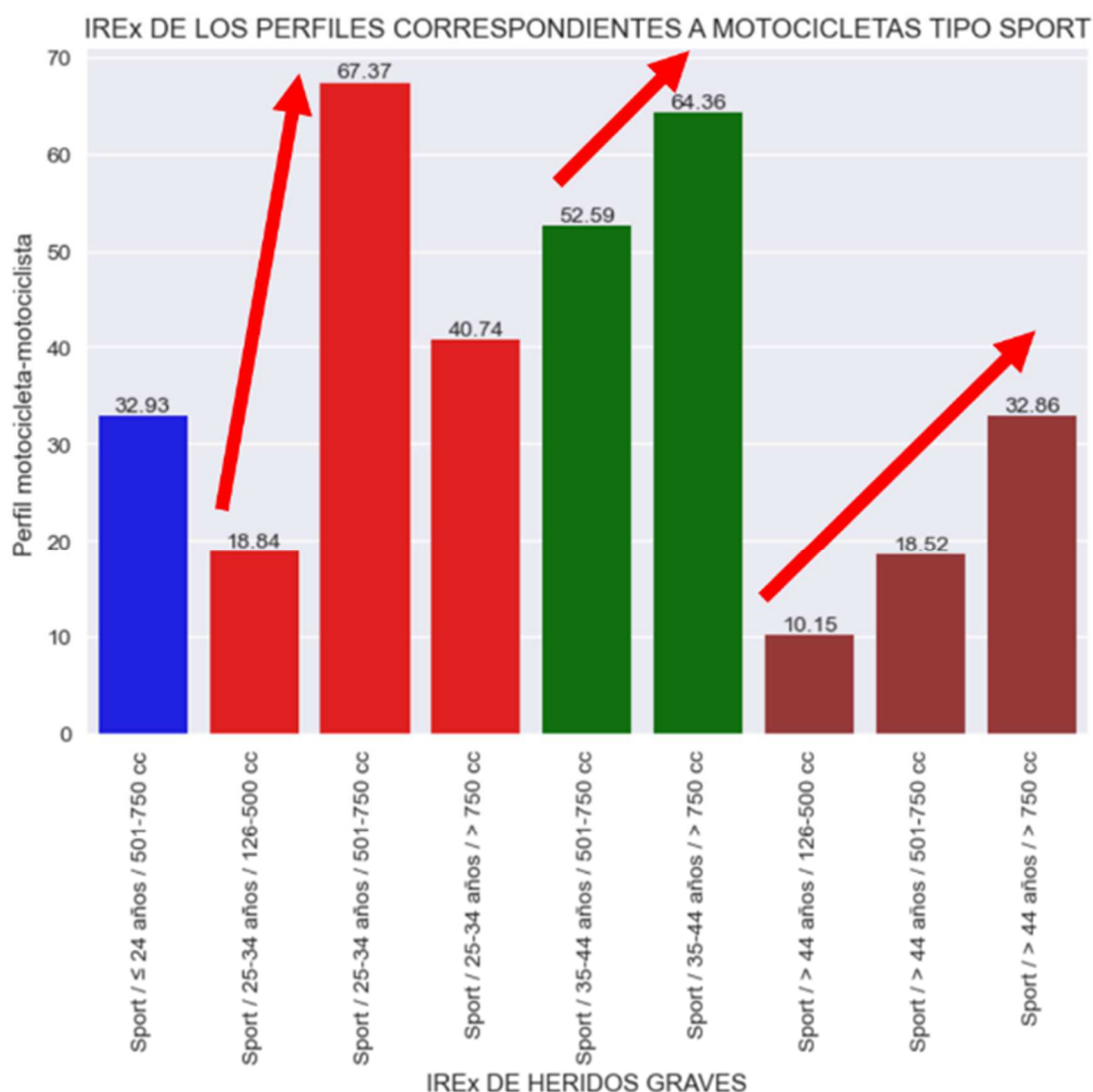


Figura 19: IREx heridos graves de los perfiles correspondientes a motocicletas tipo Sport

4.2.4. IREx CORRESPONDIENTES A MOTOCICLETAS TIPO NAKED

Con el objetivo de visualizar la influencia de la cilindrada dentro de los grupos de edad dentro de la segunda categoría con un mayor riesgo observado presentamos la siguiente gráfica:

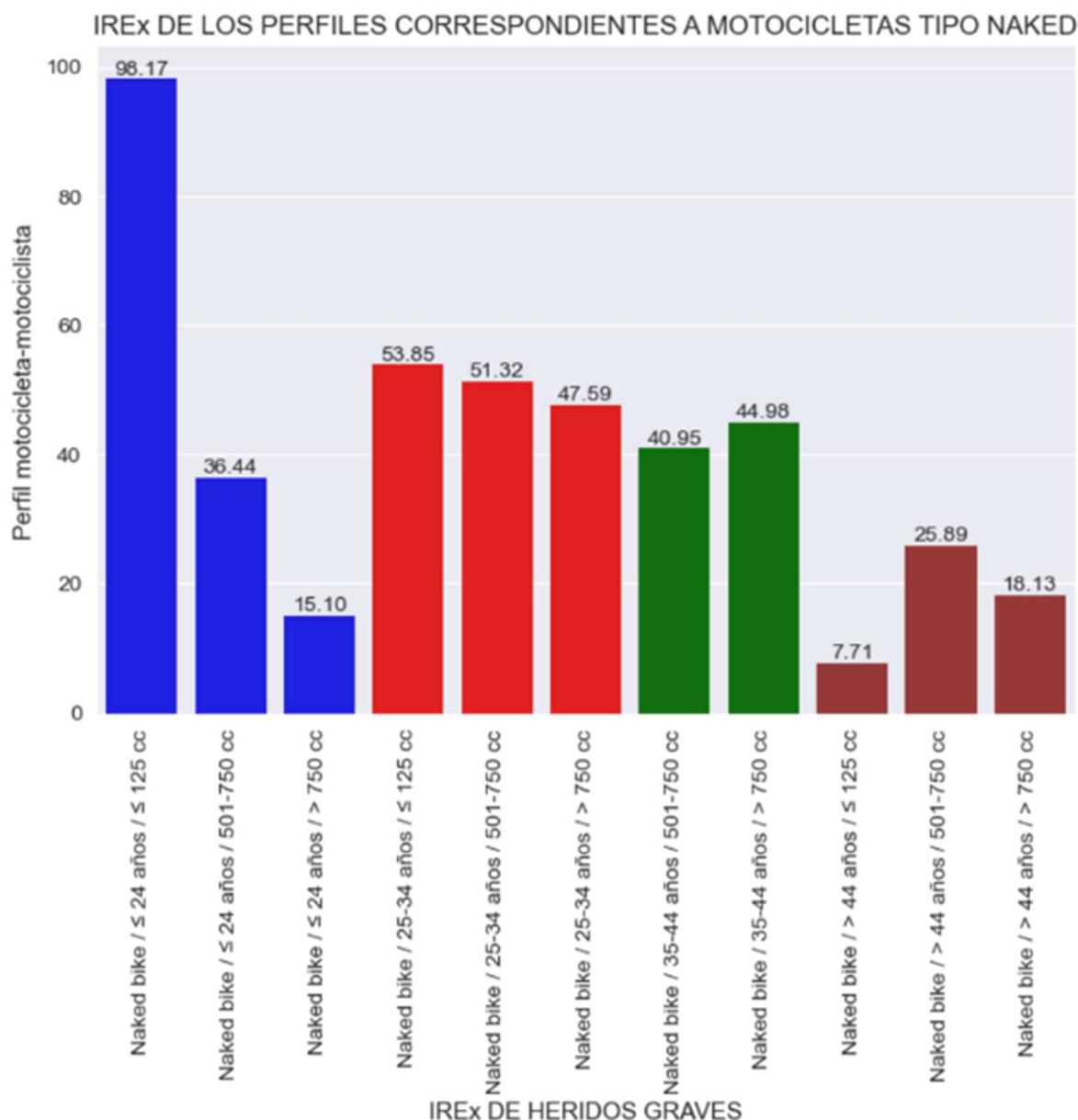


Figura 20: IREx heridos graves de los perfiles correspondientes a motocicletas tipo Naked

5. CONCLUSIONES

El estudio sobre la relación entre las características de una motocicleta y el riesgo vial arroja luz sobre varios aspectos importantes del parque de motocicletas en España y su impacto en la seguridad vial. En cuanto a la distribución del parque, se observa un claro predominio de las scooters, que representan más de la mitad del total de motocicletas. Las categorías deportivas, enduro/offroad y naked les siguen en importancia, aunque con porcentajes significativamente menores.

Un hallazgo preocupante es el envejecimiento del parque de motocicletas, con más de la mitad de los vehículos superando los 10 años de antigüedad. Esto sugiere una tasa de renovación relativamente lenta, lo que podría tener implicaciones en términos de seguridad y emisiones. En cuanto a la cilindrada, se evidencia una marcada preferencia por las motocicletas de baja potencia, con casi la mitad del parque compuesto por vehículos de 125cc o menos. Esta tendencia decrece a medida que aumenta la cilindrada. Los patrones de uso de las motocicletas revelan que la gran mayoría se utiliza para trayectos cortos o en entornos urbanos, con casi el 90% recorriendo menos de 5.000 km al año. Esto sugiere que las motocicletas se emplean principalmente como medio de transporte para desplazamientos cotidianos más que para viajes largos o de ocio.

En términos de accidentalidad, se observan disparidades significativas entre la representación en el parque y la implicación en accidentes. Las scooters, por ejemplo, están sobrerrepresentadas en los siniestros en comparación con su presencia en el parque. De manera similar, las motocicletas naked y deportivas muestran tasas de accidentalidad superiores a lo que su proporción en el parque sugeriría. Por el contrario, las motocicletas enduro/offroad presentan una tasa de accidentalidad notablemente inferior. El análisis del riesgo por edad y cilindrada revela que los motoristas jóvenes con motocicletas de baja cilindrada presentan el mayor índice de riesgo de lesiones graves. Asimismo, las motocicletas deportivas y naked de cilindrada media-alta conducidas por personas de mediana edad también muestran un alto riesgo de accidentes graves y mortales.

Estas conclusiones subrayan la complejidad de los factores que influyen en la seguridad vial de los motoristas. Los resultados de este estudio pueden ser fundamentales para el diseño de estrategias de seguridad vial más efectivas y orientadas, centrándose en los grupos y tipos de vehículos que presentan mayor riesgo.

Nota importante: Algunos cálculos se han realizado a partir de unos datos en la distribución de edad de los titulares del parque que ha sido generada matemáticamente de manera aleatoria. Tan pronto la nueva extracción de datos esté disponible, los cálculos se realizarán sobre los datos extraídos y serán presentados en una nueva versión del documento. La generación de datos aleatorios puede afectar a varios de los factores que se presentan en los gráficos de este reporte.

ANEXO 1: CLASIFICACIÓN DE LOS TIPOS DE MOTOCICLETAS

La clasificación de los diferentes tipos de motocicletas implicadas en el siguiente estudio piloto, se ha llevado a cabo según el siguiente esquema:

Sport:	Naked:
	
Scooter:	Touring:
	
Custom/cruiser:	Sport touring:
	

Enduro/offroad:	Super motard:
	
Classic	Allround:
	

Figura 21: Clasificación llevada de los tipos de motocicletas

ANEXO 2: EXPOSICIÓN AL RIESGO DE LOS MOTORISTAS EN ESPAÑA

Este anexo contiene las tablas resumen a partir de las cuales se han realizado la mayoría de los gráficos para la interpretación de resultados.

Los diferentes perfiles están ordenados según los siguientes parámetros:

- Tipo de motocicleta
- Edad del conductor
- Cilindrada de la motocicleta

Y la información básica que se puede encontrar en las tablas para cada uno de ellos es la siguiente

- Número de víctimas absoluto
- Porcentaje de muertos y heridos graves respecto el total de motocicletas en el parque nacional
- Índice de Riesgo en función de su Exposición

En la Tabla 1 queda definido el código de colores que facilita la interpretación de los resultados obtenidos en la Tabla 2, y los clasifica de acuerdo con tres grados de riesgo diferentes (bajo, medio o alto).

CÓDIGO DE COLORES	GRADO DE RIESGO	MEDIA KILOMETRAJE	TOTAL VÍCTIMAS GRAVES	TOTAL VÍCTIMAS MORTALES	TASA DE HERIDOS DE GRAVEDAD [%]	TASA DE MORTALIDAD [%]	IREx HERIDOS GRAVES [x10e8]	IREx MUERTOS [x10e8]
VERDE	BAJO	> 10.000	< 20	< 4	<0,6%	<0,1%	<60	<10
NARANJA	MEDIO	5.000 - 10.000	20 - 60	4 - 9	0,6%-1,5%	0,1%-0,25%	60-180	10-30
ROJO	ALTO	< 5.000	> 60	> 9	>1,5%	>0,25%	>180	>30

Tabla 1: Código de valores para la interpretación de la tabla 2

Tipo moto	Edad [años]	Cilindrada [cc]	Media Km/año	H. Grave	Muertos	Tasa de heridos de gravedad	Tasa de mortalidad	IREx heridos graves	IREx fallecidos
Allround	≤ 24 años	≤ 125 cc	1966	3	0	0,07	0,0	35,19	0,0
		126-500 cc	1441	1	0	0,03	0,0	21,85	0,0
		501-750 cc	3382	1	0	0,03	0,0	7,98	0,0
		> 750 cc	4908	0	0	0,0	0,0	0,0	0,0
	25-34 años	≤ 125 cc	2030	5	0	0,08	0,0	39,39	0,0

Tipo moto	Edad [años]	Cilindrada [cc]	Media Km/año	H. Grave	Muertos	Tasa de heridos de gravedad	Tasa de mortalidad	IREx heridos graves	IREx fallecidos
		126-500 cc	1477	0	0	0,0	0,0	0,0	0,0
		501-750 cc	3478	0	0	0,0	0,0	0,0	0,0
		> 750 cc	4961	0	0	0,0	0,0	0,0	0,0
	35-44 años	≤ 125 cc	1986	8	1	0,13	0,02	64,18	8,02
		126-500 cc	1478	2	0	0,05	0,0	30,73	0,0
		501-750 cc	3352	6	1	0,11	0,02	34,04	5,67
		> 750 cc	4918	1	0	0,02	0,0	4,22	0,0
	> 44 años	≤ 125 cc	1991	2	0	0,01	0,0	5,71	0,0
		126-500 cc	1491	3	0	0,02	0,0	16,44	0,0
		501-750 cc	3449	12	1	0,08	0,01	23,81	1,98
		> 750 cc	4965	13	1	0,1	0,01	19,82	1,52
Custom / cruiser	≤ 24 años	≤ 125 cc	1673	1	0	0,01	0,0	6,17	0,0
		126-500 cc	1554	0	0	0,0	0,0	0,0	0,0
		501-750 cc	2631	1	1	0,02	0,02	6,63	6,63
		> 750 cc	2354	0	0	0,0	0,0	0,0	0,0
	25-34 años	≤ 125 cc	1673	2	1	0,01	0,01	8,66	4,33
		126-500 cc	1495	0	0	0,0	0,0	0,0	0,0
		501-750 cc	2568	4	1	0,05	0,01	18,97	4,74
		> 750 cc	2359	4	2	0,03	0,01	12,15	6,08
	35-44 años	≤ 125 cc	1669	4	2	0,03	0,01	17,53	8,76
		126-500 cc	1558	0	1	0,0	0,02	0,0	13,93
		501-750 cc	2590	5	0	0,06	0,0	23,59	0,0
		> 750 cc	2373	8	1	0,06	0,01	24,11	3,01
	> 44 años	≤ 125 cc	1677	16	0	0,04	0,0	25,07	0,0
		126-500 cc	1535	6	1	0,05	0,01	30,52	5,09
		501-750 cc	2617	26	2	0,11	0,01	43,31	3,33
		> 750 cc	2373	40	6	0,1	0,02	42,5	6,37
Enduro / offroad	≤ 24 años	≤ 125 cc	1477	9	0	0,13	0,0	90,23	0,0
		126-500 cc	1265	11	2	0,08	0,01	61,25	11,14
		501-750 cc	1972	1	1	0,01	0,01	6,49	6,49
		> 750 cc	5828	3	1	0,04	0,01	6,2	2,07

Tipo moto	Edad [años]	Cilindrada [cc]	Media Km/año	H. Grave	Muertos	Tasa de heridos de gravedad	Tasa de mortalidad	IREx heridos graves	IREx fallecidos
	25-34 años	≤ 125 cc	1456	2	0	0,02	0,0	14,57	0,0
		126-500 cc	1262	9	2	0,04	0,01	35,11	7,8
		501-750 cc	1918	9	0	0,08	0,0	42,24	0,0
		> 750 cc	5886	4	1	0,03	0,01	5,8	1,45
	35-44 años	≤ 125 cc	1435	3	0	0,03	0,0	21,54	0,0
		126-500 cc	1290	9	1	0,04	0,0	34,21	3,8
		501-750 cc	1945	10	0	0,09	0,0	46,04	0,0
		> 750 cc	5782	7	0	0,06	0,0	10,3	0,0
	> 44 años	≤ 125 cc	1432	6	0	0,02	0,0	15,48	0,0
		126-500 cc	1245	5	1	0,01	0,0	7,07	1,41
		501-750 cc	1930	23	2	0,07	0,01	38,07	3,31
		> 750 cc	5889	45	6	0,14	0,02	23,19	3,09
Naked bike	≤ 24 años	≤ 125 cc	2605	17	3	0,26	0,05	100,43	17,72
		126-500 cc	1628	2	0	0,05	0,0	33,73	0,0
		501-750 cc	3177	19	3	0,11	0,02	35,84	5,66
		> 750 cc	3695	4	1	0,06	0,01	15,2	3,8
	25-34 años	≤ 125 cc	2575	13	1	0,14	0,01	54,57	4,2
		126-500 cc	1633	3	0	0,06	0,0	35,59	0,0
		501-750 cc	3159	39	4	0,16	0,02	51,7	5,3
		> 750 cc	3704	18	3	0,18	0,03	47,46	7,91
	35-44 años	≤ 125 cc	2571	3	0	0,03	0,0	12,62	0,0
		126-500 cc	1605	1	0	0,02	0,0	11,84	0,0
		501-750 cc	3210	31	5	0,13	0,02	41,2	6,65
		> 750 cc	3687	17	3	0,16	0,03	44,74	7,89
	> 44 años	≤ 125 cc	2570	5	2	0,02	0,01	7,67	3,07
		126-500 cc	1667	11	2	0,08	0,01	45,01	8,18
		501-750 cc	3149	54	13	0,08	0,02	25,87	6,23
		> 750 cc	3708	19	3	0,07	0,01	18,16	2,87
Scooter	≤ 24 años	≤ 125 cc	2378	82	5	0,06	0,0	23,6	1,44
		126-500 cc	2360	10	2	0,02	0,0	7,56	1,51
		501-750 cc	4157	6	0	0,09	0,0	21,42	0,0

Tipo moto	Edad [años]	Cilindrada [cc]	Media Km/año	H. Grave	Muertos	Tasa de heridos de gravedad	Tasa de mortalidad	IREx heridos graves	IREx fallecidos
	25-34 años	> 750 cc	2430	0	0	0,0	0,0	0,0	0,0
		≤ 125 cc	2374	169	15	0,08	0,01	33,89	3,01
		126-500 cc	2354	13	4	0,02	0,0	6,87	2,11
		501-750 cc	4113	11	1	0,11	0,01	26,98	2,45
		> 750 cc	2315	0	0	0,0	0,0	0,0	0,0
	35-44 años	≤ 125 cc	2372	153	14	0,07	0,01	30,91	2,83
		126-500 cc	2397	45	3	0,06	0,0	23,43	1,56
		501-750 cc	4152	6	4	0,06	0,04	14,74	9,83
		> 750 cc	2293	0	0	0,0	0,0	0,0	0,0
	> 44 años	≤ 125 cc	2369	309	17	0,05	0,0	22,34	1,23
		126-500 cc	2362	136	18	0,06	0,01	25,66	3,4
		501-750 cc	4124	31	4	0,11	0,01	27,14	3,5
		> 750 cc	2318	0	0	0,0	0,0	0,0	0,0
Sport	≤ 24 años	≤ 125 cc	1328	7	0	0,11	0,0	81,37	0,0
		126-500 cc	2541	4	0	0,07	0,0	28,13	0,0
		501-750 cc	2520	13	3	0,08	0,02	33,27	7,68
		> 750 cc	2963	4	1	0,04	0,01	14,51	3,63
	25-34 años	≤ 125 cc	1366	6	0	0,06	0,0	47,59	0,0
		126-500 cc	2608	4	1	0,05	0,01	18,74	4,69
		501-750 cc	2532	38	11	0,17	0,05	67,0	19,39
		> 750 cc	3020	16	1	0,12	0,01	39,38	2,46
	35-44 años	≤ 125 cc	1362	3	1	0,03	0,01	23,82	7,94
		126-500 cc	2605	6	0	0,07	0,0	28,06	0,0
		501-750 cc	2507	30	8	0,13	0,04	53,22	14,19
		> 750 cc	2960	26	7	0,19	0,05	65,15	17,54
	> 44 años	≤ 125 cc	1355	3	0	0,01	0,0	8,54	0,0
		126-500 cc	2633	6	1	0,03	0,0	10,02	1,67
		501-750 cc	2513	29	9	0,05	0,01	18,42	5,72
		> 750 cc	2967	37	18	0,1	0,05	33,21	16,15

Tabla 2: Tabla resumen de los motoristas heridos graves y muertos en España durante 2022

