IDIADA Automotive Technology SA
Passive Safety Department
L'Albornar – Apartado de correos 20
43710 Santa Oliva (Tarragona) Spain
Phone: +34 91 977 166 021
idiada@idiada.com
www.idiada.com

STUDY ON THE RELATIONSHIP BETWEEN MOTORCYCLE CHARACTERISTICS AND ROAD SAFETY

PROJECT CONDUCTED BY: Applus+ IDIADA

FOR: National Road Safety Observatory

General Directorate of Traffic (DGT)

Ministry of the Interior

Dossier: 3DGT6A000172

Produced by:

Diego Martin
ML Software Engineer
Digital Business

Approved by:

Genís Mensa
R&D Manager
Body and Passive Safety

Date delivered: 04/12/2024

This report contains 46 pages including this cover page and the annexes.

Published by: Directorate-General for Traffic
44, Josefa Valcárcel Street MADRID 28027

Author: National Road Safety Observatory

Catalogue of publications of the General State Administration: https://cpage.mpr.gob.es

O.P.I.N.: (on line ed.): 128-26-074-X

EXECUTIVE SUMMARY

According to the data published by the General Directorate of Traffic (DGT), in recent years there has been a notable reduction in the number of fatalities and serious injuries resulting from road traffic accidents. This reduction is particularly significant in view of the continuous growth of Spain's vehicle fleet. Even though its overall target has not been entirely met, compared to its fellow European Union countries Spain has achieved one of the highest proportionate reductions in the fatality rate.

However, the same level of progress has not been achieved in the case of motorcyclists. Although a considerable reduction in the number of fatal motorcycle accidents was achieved in the past, in recent years the trend has been reversed and fatalities are on the rise, in contrast to the general pattern observed for all other vehicles. For this reason, in 2009 the DGT commissioned IDIADA to conduct a pilot study focusing specifically on motorcyclists.

The aim of the study was to develop a methodology that would make it possible to assess motorcyclists' risk of suffering a fatal or serious accident, based on their level of exposure. The study analysed various motorcyclist profiles, which were defined on the basis of age and type of motorcycle ridden, and identified the profiles that were more likely to be exposed to the risk of a fatal accident.

This analysis applies the methodology that was developed in the aforementioned study. Like that study, the aim of this analysis is to identify the profiles that have a higher risk of suffering serious non-fatal accidents, based on their level of exposure.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION AND OBJECTIVES

1.1. BACKGROUND

According to data from the European Automobile Manufacturers' Association (ACEA), in 2022 the European Union (EU27) was home to a total of 437 million motorised vehicles. This total can be broken down into 252 million cars, 30 million light commercial vehicles, 6 million medium and heavy commercial vehicles, 720,000 buses, 15 million motorcycles and 11 million mopeds. The remainder are other types of vehicle (tractors, mobile cranes, fire engines, etc.).

In that same year (2022), more than 20,600 people were killed in road traffic accidents throughout the European Union. In view of these figures, the European Commission has set an ambitious target: to halve the number of road traffic fatalities by 2030, taking the 2019 data as a baseline. To date, significant progress has been made, with a reduction of 9% in road traffic fatalities in the EU27 between 2019 and 2022. This positive trend has been observed in almost all Member States. If we expand the time frame, since 2012 the fatality rate has fallen by 22%.

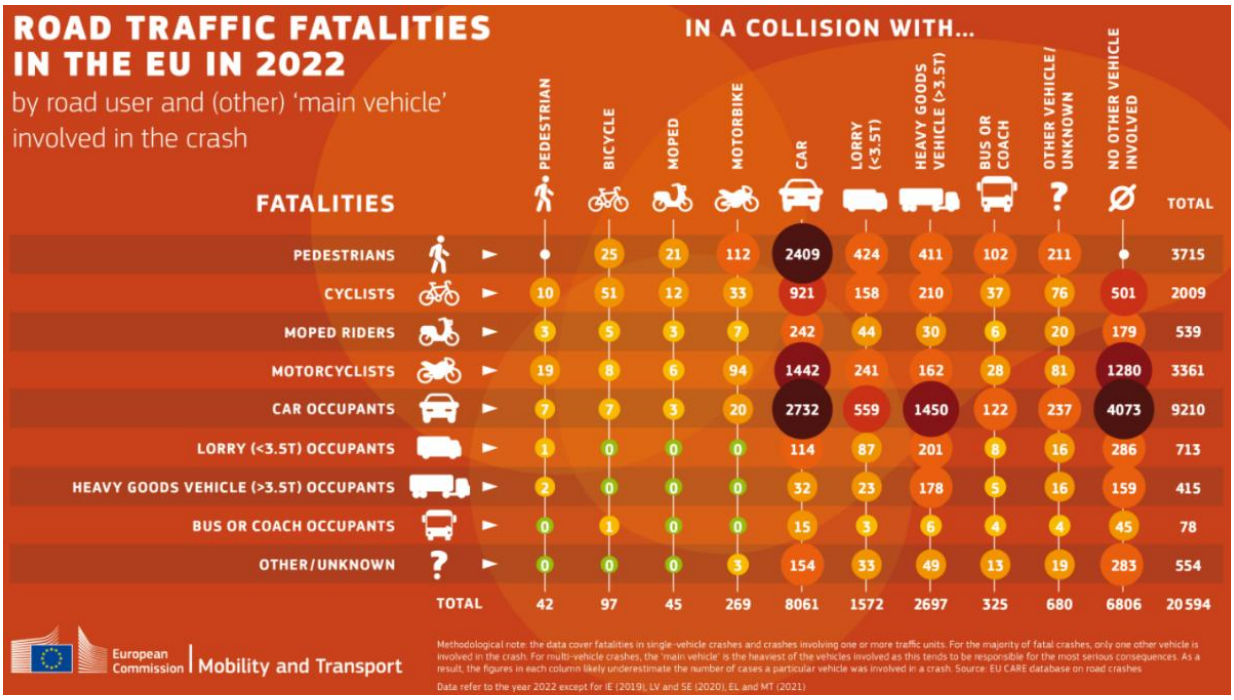

Figure 1. Fatalities resulting from road traffic accidents in the EU in 2022, by user and vehicle involved in the accident. Source: CARE.

In 2022, car occupants (drivers and passengers) accounted for 45% of all fatalities in the EU27. Pedestrians accounted for 18%, while riders of motorised two-wheeled vehicles (motorcycles and mopeds) accounted for 19% and cyclists for 10%. ***Figure 1*** shows the total number of road traffic accident fatalities in the EU, by type of transport used and type of transport involved in the accident.

In Spain, motorcyclists account for a higher proportion of road traffic fatalities than the European average; indeed, this proportion is the third-highest in Europe, according to ***Figure 2***.

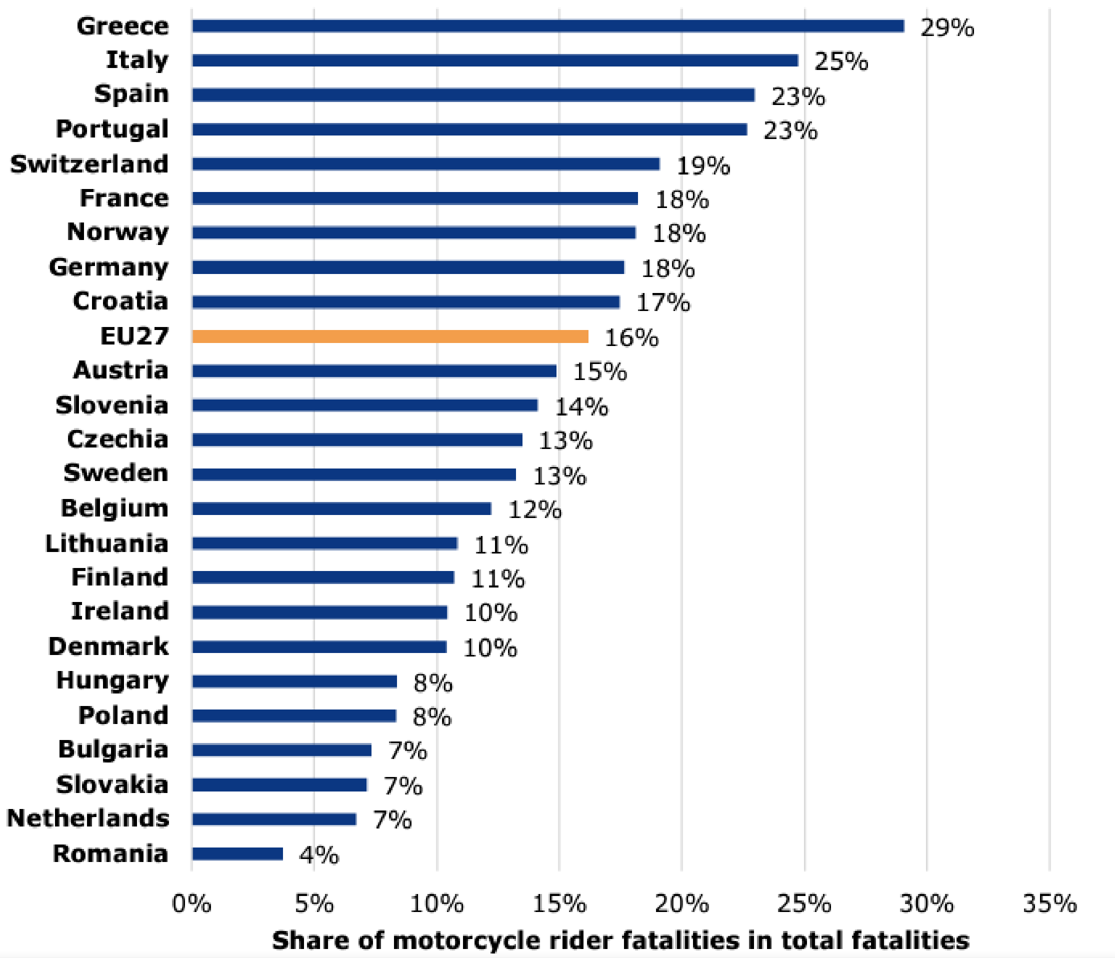

Figure 2. Motorcyclists as a proportion of total fatalities in the EU27 and EFTA countries (2022). Source: CARE.

If we examine the evolution of Spain's vehicle fleet, we see that the total number of vehicles has grown, with an annual increase of nearly 1%. However, this growth is not distributed evenly among the different types of vehicle.

Since 2013 there has been a reduction in the amount of lorries, vans and mopeds as a proportion of the fleet as a whole. In contrast, the country has witnessed a notable rise in the number of motorcycles: the category has grown from 9% of the total vehicle fleet in 2013 to 11% of the fleet in 2022, with between 50,000 and 150,000 motorcycles added to the fleet each year.

In parallel to this increase in numbers, the records show that in 2022, motorcycles topped the table of road traffic accident statistics, accounting for the highest percentage of fatalities and hospitalised injured casualties. Indeed, this trend has increased significantly over the last 10 years, rising from 24% in 2013 to 30% in 2022.

However, there has been a slight improvement compared to 2019, with cases falling from 3,146 (31% of the total) to 3,022 (30% of the total), i.e. a reduction of 1%. The data for fatalities by type of vehicle used are shown in Figure 3.

Figure 3. Evolution of the distribution of fatalities and hospitalised injured casualties, by mode of transport. Spain, 2013-2022. Source: DGT1

If we examine the evolution of fatalities by mode of transport between 2013 and 2022 (see Figure 4), we can see that in recent years (2020-2022) the trend has risen for motorcyclists; indeed, the only category with an even higher fatality rate is that of vans and lorries up to 3,500 kg.

It should also be noted that the drop in fatalities between 2019 and 2020 coincides with the COVID-19 pandemic, during which the use of vehicles was restricted.

[1] The "PMV" category has been included since 2020.

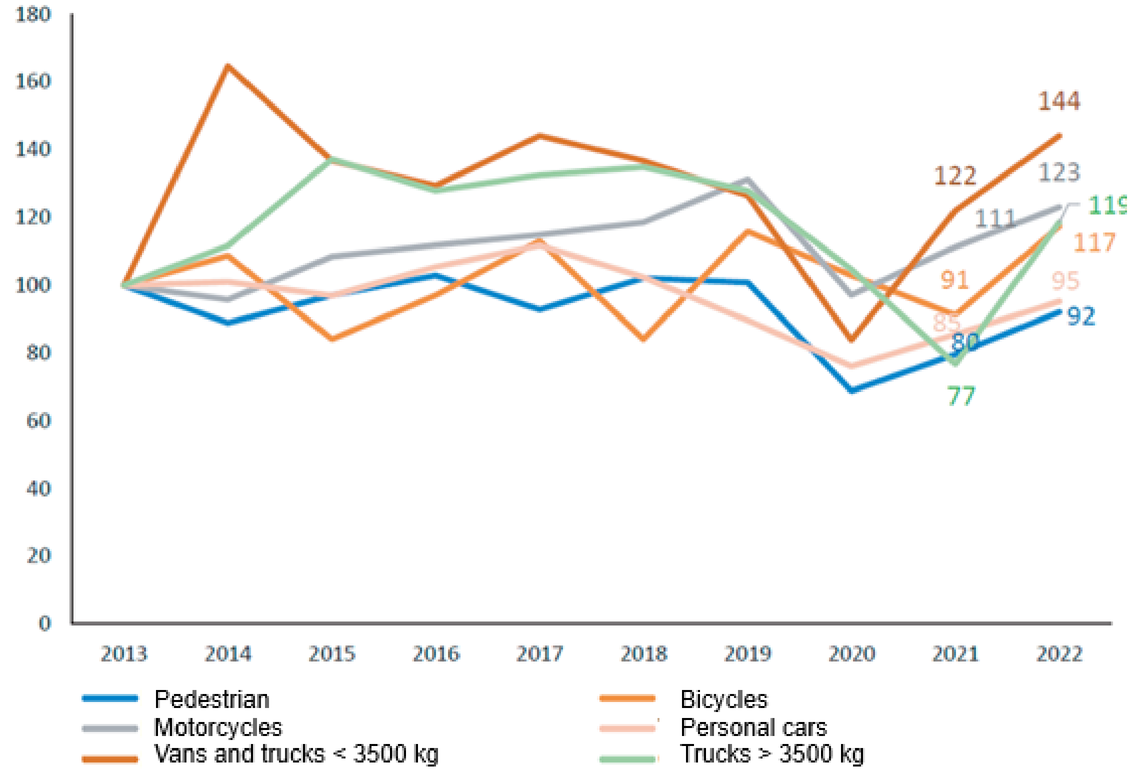

Figure 4. Evolution of fatalities by mode of transport. 2013 base = 100. Spain, 2013-2022.[2]

[2] The number of fatalities in Spain in 2013 has been set at a baseline of 100, in order to express the percentage changes between 2013 and 2022 by mode of transport.

1.2. OBJECTIVES

The main objective of this study is to analyse the risks associated with motorcycle use from a comprehensive and data-based perspective, in order to generate information that can be used to improve road safety. To this end, we will analyse three key databases corresponding to 2022.

The analysis of these databases will enable us to make detailed estimates of both the risk of an accident and its potential severity. The accident risk will be assessed from two complementary perspectives: the probability, for each vehicle in the fleet, of being involved in an accident; and the probability of such involvement per kilometre travelled. For its part, potential severity will be assessed by calculating the percentage of fatalities and serious injuries in relation to the total number of casualties, which will provide a more accurate view of the impact of the accidents in terms of severity of injury.

A key part of the study will be the classification of motorcycles based on a series of specific variables that enable us to identify differences in patterns of risk and severity of injury. These variables include:

- Capacity
- Power-to-weight ratio
- Vehicle age
- Motorcycle type

With regard to this latter variable, we will use a detailed classification designed to group motorcycles by type, such as sport, naked, scooter, touring, cruiser and trail/off-road/supermotard. To carry out this classification, we will identify the make and model of each motorcycle registered, with the aim of classifying at least 80% of the total fleet in line with these variables.

This approach will enable us not only to identify the most relevant risk factors, but to also understand how they vary based on the vehicles' specific characteristics. By linking technical and design-related variables to data on the fleet, kilometres travelled and accidents suffered, conclusions can be reached that may contribute to the development of effective strategies for reducing the number of accidents and mitigating their consequences.

Moreover, the analysis will provide evidence that can be used in drafting public policies designed to improve road safety for motorcyclists, who are one of the most vulnerable groups of road users. In summary, we intend for this study to serve as a key tool for understanding the relationship between motorcycle use and the accident rate, and for supporting data-based decision-making.

2. WORKING METHODOLOGY

2.1. DATA ANALYSIS METHODOLOGY

During the process of analysing the dataset for the motorcycle fleet, we performed an exhaustive evaluation of the quality of the data available. It is important to note that the database analysed was limited exclusively to static information regarding the vehicle fleet, and lacked dynamic data such as the number of kilometres travelled by the vehicles each year.

2.1.1. STAGE ONE: EVALUATING THE INTEGRITY OF THE DATA

During the quality analysis we identified various inconsistencies and anomalies in the records, principally in the following areas:

a) Name of makes and models:
- Typographical errors were found in the names of makes and models.
- These inconsistencies made it more difficult to categorise and statistically analyse the vehicle fleet correctly.
- The errors included variations in the spelling of the same make/model name, inconsistent use of lower/upper case, and typing errors.

b) Anomalous values for power-to-weight ratio (PWR):
- Records were found with PWRs outside of the logical range.
- There were cases where the PWR = 0, which is not technically feasible for a working vehicle.
- Records with infinite values (inf), indicating potential errors when inputting the data or in the calculation.

c) Capacity:
- Records were found citing capacities in excess of 3,000 cc.
- Such values are highly unlikely in the motorcycle segment.

These inconsistencies in the data present various challenges from an analytical perspective:
- They reduce the reliability of the statistical analyses.
- They require the data to undergo a cleansing and standardisation process.
- They may affect the accuracy of the conclusions drawn from the study.

For future records and updates to the database, we suggest the following:
- Implement validation systems for inputting data.
- Set acceptable ranges for numerical values (capacity, PWR).
- Create a standardised catalogue of makes and models.
- Develop protocols for the periodic verification of the quality of the data.

2.1.2. STAGE TWO: INTEGRATING EXTERNAL DATABASES

1. External data sources

The following databases were used as an external resource:

BIKEZ_KAGGLE: Database of the Bikez.com website, covering more than 38,684 models. This database holds technical information such as motorcycle make, model, power, tare weight, capacity, etc.

NHTSA: Database of the National Highway Traffic Safety Administration, covering more than 29,164 models. This database holds a variety of motorcycle-related information, such as make, model, chassis number, manufacturer, etc.

EURO NCAP: Database with information on 791 models, obtained from the website of the European New Car Assessment Programme (Euro NCAP).

2. Integration with the NHTSA

This process consisted of:

- Cross-referencing the chassis numbers of the motorcycle fleet with the NHTSA data.
- Validating makes and manufacturers by checking matches in the records.
- Identifying discrepancies between the two databases.

3. Results and findings

a) Positive results:

- We were able to correct and validate multiple records for makes.
- A data-verification protocol using external sources was established.
- Patterns of errors were identified in the original data.

b) Limitations encountered:

- We found that “chassis number” was not a reliable field in the original dataset.
- The success rate for cross-referencing the data was lower than expected (it was not possible to check the models).
- It became clear that a more robust approach to data cleansing was required.

The findings from this second stage led to the planning of a third phase of work, characterised by:

- The need to develop a more exhaustive approach.
- The need for more complex and robust methodologies.
- The need for more comprehensive solutions to cleanse the dataset.

2.1.3. STAGE THREE: APPLYING AI AND NLP ALGORITHMS

A process was designed to impute data and make corrections in the "makes" and "models" columns for the vehicles in a dataset, using semantic search techniques, string similarity algorithms and large language models (LLMs). The processing phases that were carried out are detailed below:

1. Loading and preprocessing data

The process begins by loading a base dataset containing vehicle information. The following preprocessing steps are then carried out:

- **Selecting relevant columns**: We work with specific columns, such as "`MODEL`", "`SHORT_DESCRIPTION`", "`MAKE_CODE`", "`DATE_FIRST_REG`", "`KW`" and "`OBV_TARE`".
- **Data transformations**:
  - The values are converted to upper case ("`MODEL`", "`SHORT_DESCRIPTION`") in order to ensure uniformity.
  - A new metric – power-to-weight ratio (PWR) – is calculated, for additional analyses.
  - Duplicates are eliminated from the index.
- **Renaming the columns**: The columns are adjusted so that the names are presented consistently.

This preprocessing ensures that the data are in a standardised format ahead of the following phases.

2. Imputation based on chassis number (VIN)

The aim during this phase is to obtain information on vehicle make directly from the chassis number/vehicle identification number (VIN), via integration with the National Highway Traffic Safety Administration (**NHTSA**) database. The steps are as follows:

1. **Extracting VIN data**:
   a. The "`NHTSA_VIN_EXTRACTOR`" library is used to obtain additional information, such as make, corresponding to the VIN of each record.
2. **Imputing makes**:
   a. Those records where a valid make has been obtained from the VIN are directly imputed to a new column ("`SEMANTIC_MAKE`") and the **origin of the information is labelled "`NHTSA VIN`".**
3. **Separating non-imputed data**:
   a. Those records whose VINs do not match the NHTSA database remain in a dataset named "`non_imputed`", for use in subsequent phases.

3. Semantic imputation with master databases

For those records that are not imputed during the previous phase, **various master databases** containing information on makes and models are used. The aim is to correct spelling mistakes, unify nomenclatures and make imputations based on semantic similarity.

- **Master databases used**: The following sources are included:
  - **BIKEZ_KAGGLE**: Master generated from the Bikez database (38,684 models).
  - **NHTSA**: Master generated from the NHTSA database (29,164 models).
  - **EURO NCAP**: Master generated from standardised data on vehicle safety tests (791 models).
  - **MANUAL**: A master database generated following manual research into known makes and models (317 models).

- **Imputation techniques**:
  - **String similarity**: The **Jaro-Winkler** algorithm is used to calculate metrics for the similarity between the values in the columns and the data in the master databases.
  - **Hierarchy of imputation**: The most reliable master databases (EURO NCAP and BIKEZ) are given priority ahead of the least specific (MANUAL and NHTSA).
  - **Similarity thresholds**: Two levels are set:
    - `threshold1` (0.90): high level of confidence in the imputation based on direct similarity.
    - `threshold2` (0.70): low level of confidence, where LLM support may be activated.

4. Additional correction using LLMs

In cases where semantic similarity does not exceed the higher threshold (`threshold1`) but is above the lower (`threshold2`), an LLM is used to make more precise imputations. This process involves:

1. **Building a prompt**: The LLM receives a detailed context, which includes the list of valid values in the master database and the problematic value in the dataset.
2. **Correction processing**: The model suggests the most likely value or issues a notification if it does not find a reasonable match.
3. **Updating the dataset**: The LLM results are incorporated into the imputed data and the origin of the correction is labelled as "LLM".

5. Replacing pseudonyms

Some makes and models have common variations or pseudonyms (popular names) in the dataset. A manual adjustment is made based on a predefined dictionary, where certain models are replaced with their standard equivalent. For example:

- **HONDA**: `WW125A` is replaced with `SH125AI` or `PCX 125`, as appropriate.
- **YAMAHA**: `YP125R` is replaced with `XMAX 125`.

6. Hierarchical imputation of models

Having imputed the makes, the next step is to impute the models. This process follows a hierarchical logic similar to that used for the makes, although now the master databases are filtered by imputed make, thereby reducing the number of comparisons required. This approach involves:

1. **Separation by make**: The records are grouped according to imputed make.
2. **Master database filtering**: The master databases are limited to models corresponding to the imputed make.
3. **Semantic imputation**: The same thresholds (`threshold1` and `threshold2`) and similarity algorithms are used to complete the "`MODEL`" `column`.

7. Results and final separation of the data cleansing process

At the end of the pipeline, the data are divided into three main datasets:

- **Imputed makes and models**: Records with successful imputations or corrections in the "`FINAL_MAKE`" `and` "`FINAL_MODEL`" `columns`. The source of each imputation (`NHTSA VIN`, `EURO NCAP`, etc.) is documented.
- **Imputed makes:** Records with successful imputations or corrections in the "`FINAL_MAKE`" column but not in the "FINAL_MODEL" column.
- **Not imputed**: Records where it was not possible to make an imputation due to a lack of matches in the master databases, or as a result of ambiguous values.

2.1.4. FINAL STAGE: CATEGORISING THE DATA

In this final stage a number of operations were carried out in order to consolidate, validate and categorise the vehicle data obtained from various sources. This process makes it possible to analyse, within a unified structure, the data on the vehicle fleet and vehicles involved in accidents. Each phase is described in detail below:

1. Loading and validation of initial data

1. Data sources

The datasets for vehicles (“`Fleet`”) and vehicles involved in accidents (“`Accidents`”) are loaded from CSV files generated during the previous stage. The records in each dataset are classified using a new column (“`VALIDATED`”) and assigned the following values:

- “`NULL`”: no validation.
- “`MAKE`”: partial validation (only makes).
- “`MAKE/MODEL`”: complete validation (makes and models).

2. Date formatting

As the datasets have different date formats, they are standardised in order to facilitate subsequent analysis:

- The dates for vehicles involved in accidents are converted from `MM/DD/YYYY` to `DD/MM/YYYY`.
- The dates for the fleet are already in the correct format (`DD/MM/YYYY`).

3. Identification of vehicles involved in accidents

- The following values are entered into the “`ACCIDENTS`” column:
    - “`NO`”: for records pertaining to the vehicle fleet.
    - “`YES`”: for records pertaining to vehicles involved in accidents.

4. Elimination of duplicate records

To prevent inconsistencies, duplicate records are eliminated when the datasets are merged. If a vehicle appears in both the “`Fleet`” and “`Accidents`” datasets, priority is given to the record pertaining to the latter.

5. Consolidation of make and model information

Consolidated versions of the “Makes” and “Models” columns are generated and given the names “`FINAL_MAKE`” and “`FINAL_MODEL`”:

- If “`SEMANTIC_MAKE`” or “`SEMANTIC_MODEL`” contain null values, the original values of the “`MAKE`” and “`MODEL`” columns are used.

Records where both the “`FINAL_MAKE`” and “`FINAL_MODEL`” columns are empty are eliminated from the dataset, as they cannot be categorised using external sources.

2. **Integration of external data**

   1. **Obtaining category information from the Bikez database**:

      a. The "FINAL_MAKE" + "FINAL MODEL" combinations are compared to the "MAKE" + "MODEL" combinations in the external dataset.
      b. The intersection of common values is calculated.

   2. **Extraction of ad hoc categories**:

      a. A dictionary is constructed in order to map makes and models to the corresponding categories, based on the external data.
      b. The missing categories are processed separately by means of a function that uses a set of categories extracted following a manual research process (updated version of the Bikez website, distributors, motorcyclists' associations, etc.), as a result of which a few more are categorised.

3. **Estimation of the missing categories using technical variables**

A classifier was developed to estimate the missing motorcycle categories, based on the "OBV_CAPACITY", "OBV_TARE" and "KW" variables. The process is detailed below:

1. **Filtering valid data**

We identified and eliminated records with missing values in the key variables for the analysis:

a. **OBV_CAPACITY**: Capacity of the engine in cm^3.
b. **OBV_TARE**: Net weight of the motorcycle.
c. **KW**: Power of the motorcycle in kilowatts.
d. **CATEGORY**: Motorcycle classification label.

2. **Division of the data into training and testing sets**

The data were divided into two sets:

e. **Training dataset** (67% of the data): this dataset was used to adjust the model.
f. **Validation dataset** (33% of the data; 1,310,503 motorcycles): this dataset was used to evaluate the model's performance.

3. **Standardisation of predictor variables**

To prevent absolute magnitudes from influencing the model disproportionately, the independent variables were standardised.

4. **Training the model: Random Forest**

We used a Random Forest classification model (`RandomForestClassifier`), which is known for its robustness and capacity to handle data with multiple characteristics without requiring extensive preprocessing.

5. Evaluating the model

The model was evaluated using the metrics included in the **classification report.** The metrics provided included precision, recall, F1-score and support by class.

Classification report:	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

As we can see, the general level of precision achieved is quite high. To support these results we defined a confidence metric: essentially, the probability that a motorcycle belongs to the assigned category. The distribution is shown in the following graph:

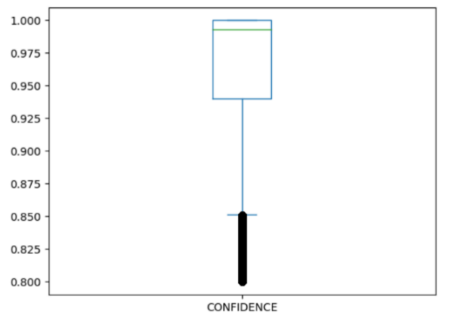

Figure 5. Distribution of confidence levels for the classification.

Some 75% of the estimates made have a confidence level in excess of 92%. In other words, there is a 92% probability that the motorcycle genuinely belongs to the category assigned. The remaining 25% all have a confidence level in excess of 80%.

6. **Metrics for categorising motorcycle type**

Key performance indicators (KPIs) are calculated in order to evaluate the percentage of records categorised:

1. **Vehicle fleet**:
   Records categorised: **80%.**

2. **Vehicles involved in accidents**:
   Records categorised: **86%.**

4. Discretisation of variables present in the original data

A set of transformations was applied in order to categorise continuous variables and derive new metrics based on the data. These categorisations make it possible to analyse patterns and group key information, which in turn facilitates statistical analysis and interpretation of results.

1. **Categorisation of "OBV_CAPACITY"**

- The observed capacity ("`OBV_CAPACITY`") is grouped into predefined intervals using `pd.cut`, thereby generating a new column, "`CAT_CAPACITY`".
- The intervals and labels assigned are:
  - **0-125 cc**: low capacity.
  - **125-500 cc**: low-to-medium capacity.
  - **500-750 cc**: medium-to-high capacity.
  - **750+ cc**: very high capacity.

2. **Categorisation of "AGE"**

- The age ranges are grouped into interpretative categories using `pd.cut`, thereby creating the "`AGE`" column. The intervals and labels assigned for the variable "RIDER_AGE_CAT" are:
  - **Age 18-24**
  - **Age 25-34**
  - **Age 35-44**
  - **Age 44+**

3. **Categorisation of "DATE_FIRST_REG"**

- The vehicle's age is calculated on the basis of the difference between the current year ("current_year") and the year it was first registered ("DATE_FIRST_REG").
- Using this information, the "AGE" column is created, with the corresponding information stated in years.
- Subsequently, the data are grouped into categorical ranges in the "CAT_AGE" column, assigning labels such as:
    - **0-2 years old**: very new vehicle.
    - **3-5 years old**: new vehicle.
    - **6-10 years old**: vehicle of intermediate age.
    - **11-15 years old**: old vehicle.
    - **16-20 years old**: very old vehicle.
    - **20+ years old**: classic vehicle or extended use.

5. Solution diagram

The following diagram presents the flow of actions taken in order to obtain the dataset that is the product of this study:

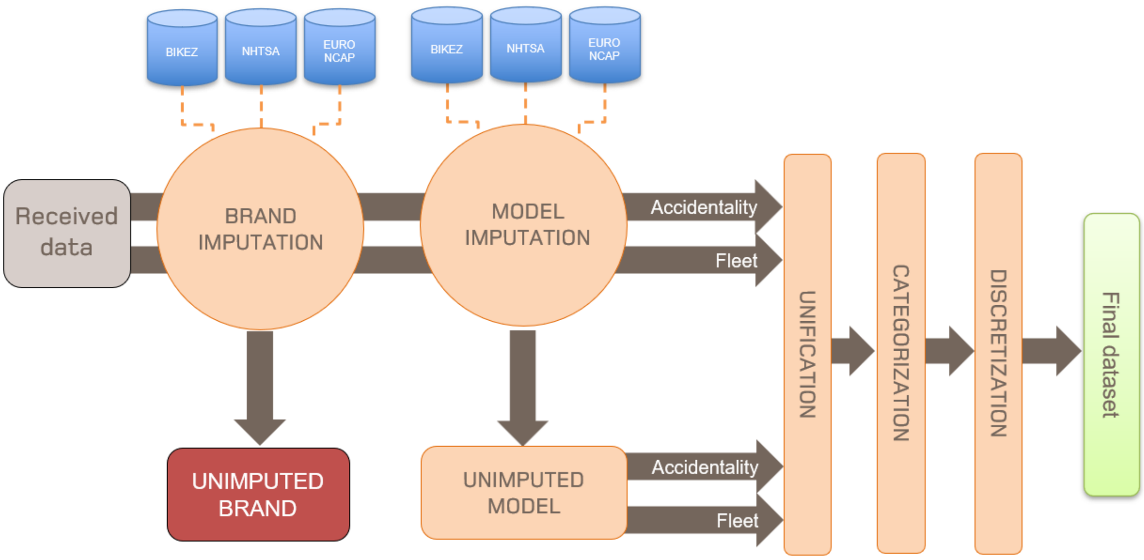

Figure 6. Diagram of the ETL process for the data on the vehicle fleet + vehicles involved in accidents.

3. DATA

3.1. ANALYSIS OF THE DATA RECEIVED

The study is based on three key datasets, each of which provides information essential to understanding the relationship between motorcycle characteristics and road safety. Each of these databases, which were provided by the DGT, are detailed below:

a) Motorcycle fleet:
   - Sample size: approximately 4,006,804 vehicles.
   - Description: this database represents the complete census of motorcycles registered in Spain in 2022.
   - Content: includes detailed information on each vehicle, such as make, model, year of manufacture, capacity, power and weight.
   - Relevance: these data are essential to determining the total size of the motorcycle fleet and will be used as a denominator for calculating the accident rates.
   - Potential limitations: this database may not account for motorcycles that are not in active use, despite being registered.

b) Kilometres travelled by motorcycles:
   - Volume of data: approximately 4,818,979 km recorded.
   - Source: observations made at vehicle roadworthiness inspection (ITV) stations.
   - Description: this database provides information on the distance travelled by the motorcycles, based on the odometer readings recorded at technical inspections.
   - Content: expected to include vehicle identification details, inspection date and odometer reading.
   - Relevance: these data are vital to estimating the motorcycles' exposure to risk. They will make it possible to calculate the accident rates per kilometre travelled, thereby providing a more accurate measurement of the real risk.
   - Potential limitations:
     - There may be a bias towards older motorcycles, as new vehicles do not require immediate inspection.
     - This database does not record the kilometres travelled by motorcycles that did not pass the ITV in the year studied.

c) Motorcycles involved in accidents with casualties:
   - Number of cases: approximately 30,875 vehicles.
   - Source: presumably police and emergency services records.
   - Description: this database contains detailed information on road traffic accidents with casualties involving motorcycles.

- Content: expected to include information such as date, time and location of the accident, the local conditions, the characteristics of the motorcycle involved, details of the rider and casualties, and information regarding the severity of the injuries.
- Relevance: this information is essential in order to evaluate the frequency and seriousness of accidents. It will make it possible to analyse accident patterns and risk factors associated with different motorcycle types and riding conditions.
- Potential limitations: the quality of the data may vary depending on the accuracy of the record-keeping at the place where the accident occurred.

The combination of these three databases will provide a comprehensive overview of road safety in relation to motorcycles. By correlating the information on the vehicle fleet with the data on kilometres travelled and accidents recorded, it will be possible to:

1. Calculate accident rates adjusted by exposure.
2. Identify risk factors associated with motorcycle characteristics.
3. Evaluate the influence of the rider's age and vehicle use on the likelihood of being involved in an accident.
4. Analyse safety trends in different segments of the motorcycle market.

3.2. SPAIN'S MOTORCYCLE FLEET (WITH ANNUAL_KM AS PER ITV)

3.2.1. DISTRIBUTION BY MOTORCYCLE TYPE

During a previous stage in the process, the motorcycles forming part of Spain's vehicle fleet in 2024 were classified by type, based on make and model and in accordance with the classification provided in Annex 1. The following figure shows the distribution of Spain's motorcycle fleet in 2024 by type.

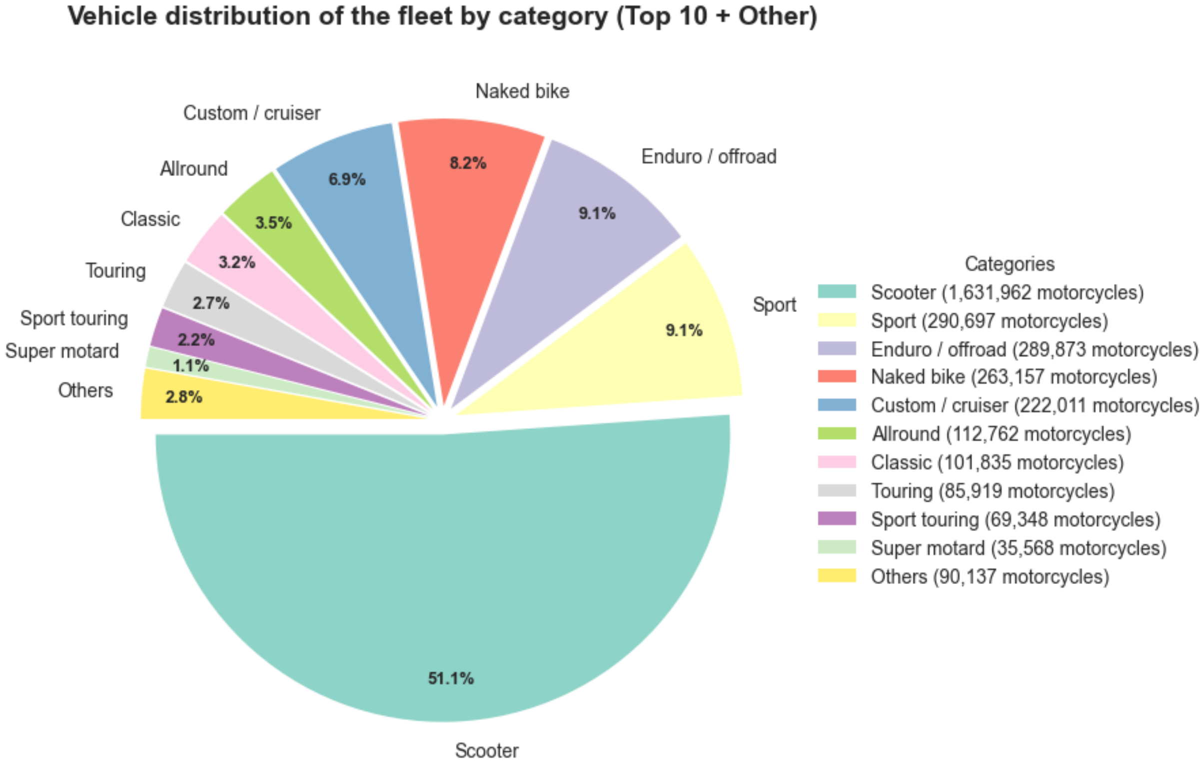

Figure 7. Distribution of Spain's motorcycle fleet in 2024 by type.

The graph shows that the predominant category is scooters, accounting for more than half of the fleet (51.1% of all motorcycles, equating to 1,631,961 vehicles).

In second and third place are the sports and enduro/off-road categories, both with 9.1% market share (equating to 290,697 sports motorcycles and 289,871 enduro/off-road vehicles). The naked-bike category is in fourth place with 8.2% (263,157 motorcycles).

The remaining categories are custom/cruiser (6.9%), all-round (3.5%), classic (3.2%), touring (2.7%), sport touring (2.2%) and supermotard (1.1%).

Some 2.8% of motorcycles do not fall into the main categories.

This distribution clearly illustrates consumer preferences in the motorcycle market, demonstrating a **strong inclination towards practical vehicles** for daily transport (scooters) **followed by more specialised options** for different riding styles and rider preferences.

3.2.2. DISTRIBUTION BY MOTORCYCLE AGE

The following figure shows the distribution of Spain's motorcycle fleet in 2024 by vehicle age.

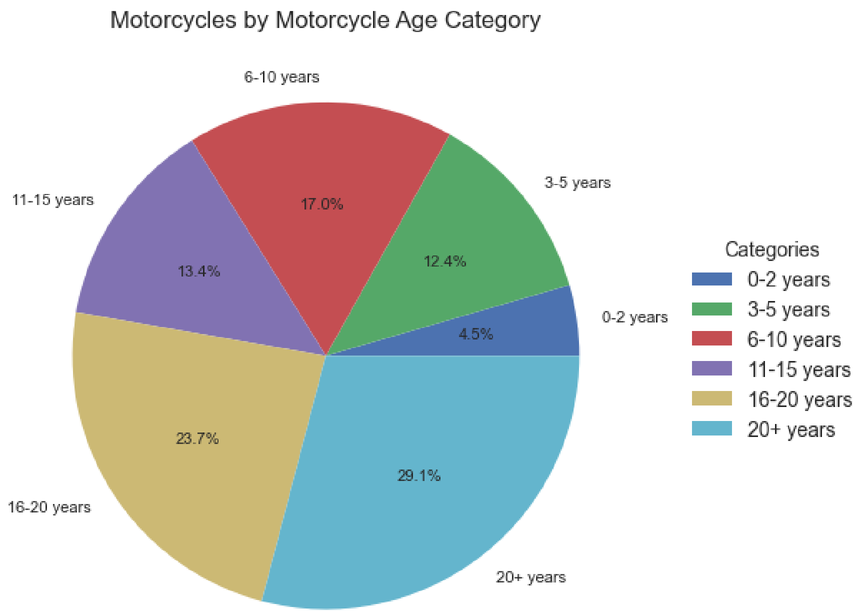

Figure 8. Distribution of Spain's motorcycle fleet in 2024 by age.

The graphs show the distribution of motorcycles within six age categories, ranging from new vehicles (0-2 years old) to the oldest vehicles (20+ years old).

Trend analysis:

- Newer vehicles (0-5 years old) account for 16.9% of the total. This suggests a moderate renewal rate for the motorcycle fleet.
- Vehicles of intermediate age (6-15 years old) account for 30.4%. This indicates that a sizeable proportion of the motorcycles on the road are quite old.
- Older vehicles (16+ years old) account for 52.8% of the total, meaning that a majority of the motorcycles on the road are older vehicles.

Implications:

- **Ageing fleet**: over half of the motorcycles (52.8%) are more than 16 years old, which may present challenges with regard to emissions and safety.
- **Slow renewal**: the relatively low percentage of new motorcycles (0-2 years old) suggests that the renewal rate is fairly slow.
- **Durability**: the significant presence of motorcycles aged 16+ indicates that many of the vehicles have a long useful life.

3.2.3. DISTRIBUTION BY MOTORCYCLE CAPACITY

The following figure shows the distribution of Spain's motorcycle fleet in 2024 by capacity.

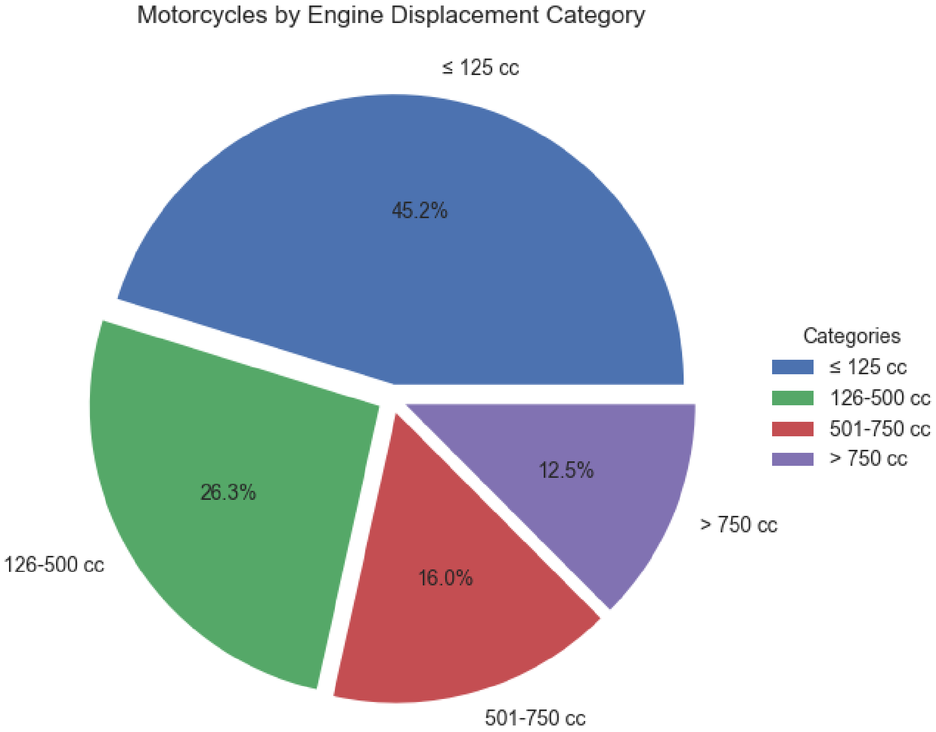

Figure 9. Distribution of Spain's motorcycle fleet in 2024 by capacity.

The distribution by capacity category is as follows:

- 125 cc or lower: 45.2% of vehicles (the largest category).
- 126-500 cc: 26.3% of vehicles.
- 501-750 cc: 16.0% of vehicles.
- Over 750 cc: 12.5% of vehicles.

The figures show that the **market is clearly dominated by motorcycles with a low capacity**. **Capacities of 125 cc or lower account for almost half of the fleet (45.2%)**, which indicates a strong preference for smaller vehicles, probably due to their economy, ease of handling in urban environments, and potential regulatory advantages.

We can see that as engine capacity increases, the number of motorcycles in each category decreases. The categories with the highest capacities (501-750 cc and

750+ cc) account for around 30% of the total. The intermediate segment (126-500 cc) accounts for a significant proportion (26.3%), evidently attracting users seeking a balance between power and practicality.

3.2.4. DISTRIBUTION BY AVERAGE NUMBER OF KILOMETRES TRAVELLED PER YEAR BY THE MOTORCYCLE

To round off the categorisation of the motorcycle fleet, the following figure shows the distribution of Spain's fleet in 2024 according to the average number of kilometres travelled per year.

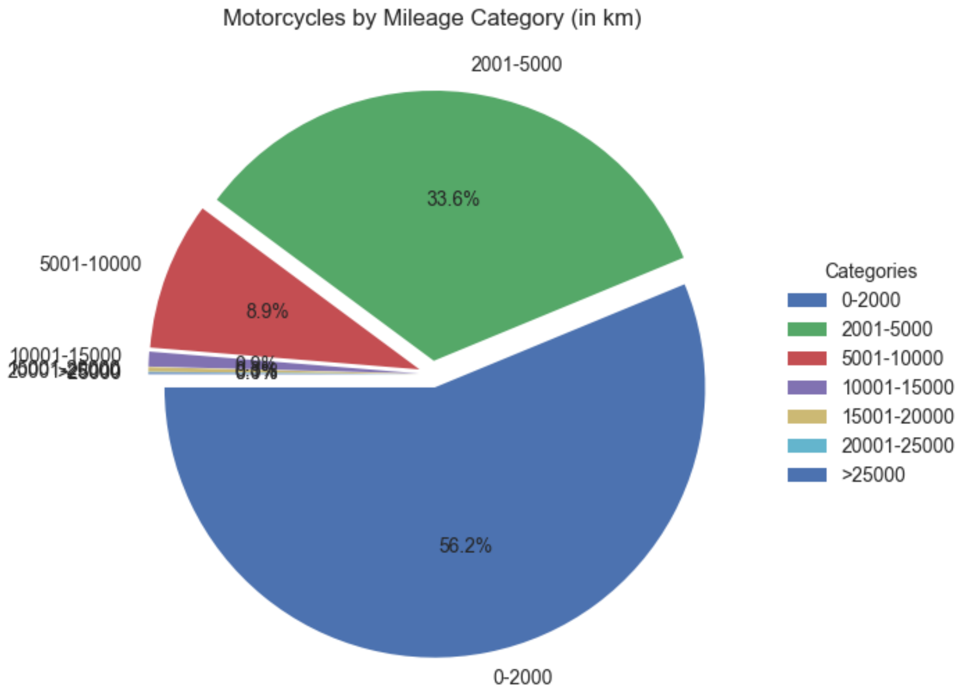

Figure 10. Distribution of Spain's motorcycle fleet in 2024 by average number of kilometres travelled per year by the motorcycle.

The distribution by category of annual kilometres travelled is as follows:

- Less than 2,000 km: 1,897,885 vehicles (56.2%; the largest category).
- 2,000-5,000 km: 1,133,804 vehicles (33.6%; the second-largest category).
- 5,000-10,000 km: 299,361 vehicles (8.9%).
- More than 10,000 km: 43,654 vehicles (1.3%).

The analysis reveals a **pattern of predominantly low use** for motorcycles, in terms of kilometres travelled. Some 89.8% of vehicles travel less than 5,000 km per year, which suggests they are used mainly for urban travel or short journeys. The distribution shows a marked downward trend as the number of kilometres travelled per year rises: just 10.2% of vehicles travel more than 5,000 km per year, and among them, only 1.3% travel more than 10,000 km.

3.3. MOTORCYCLE ACCIDENTS IN SPAIN

3.3.1. DISTRIBUTION BY MOTORCYCLE TYPE

The following figure shows the distribution of motorcycle accidents in Spain in 2024 by type of motorcycle involved in the accident.

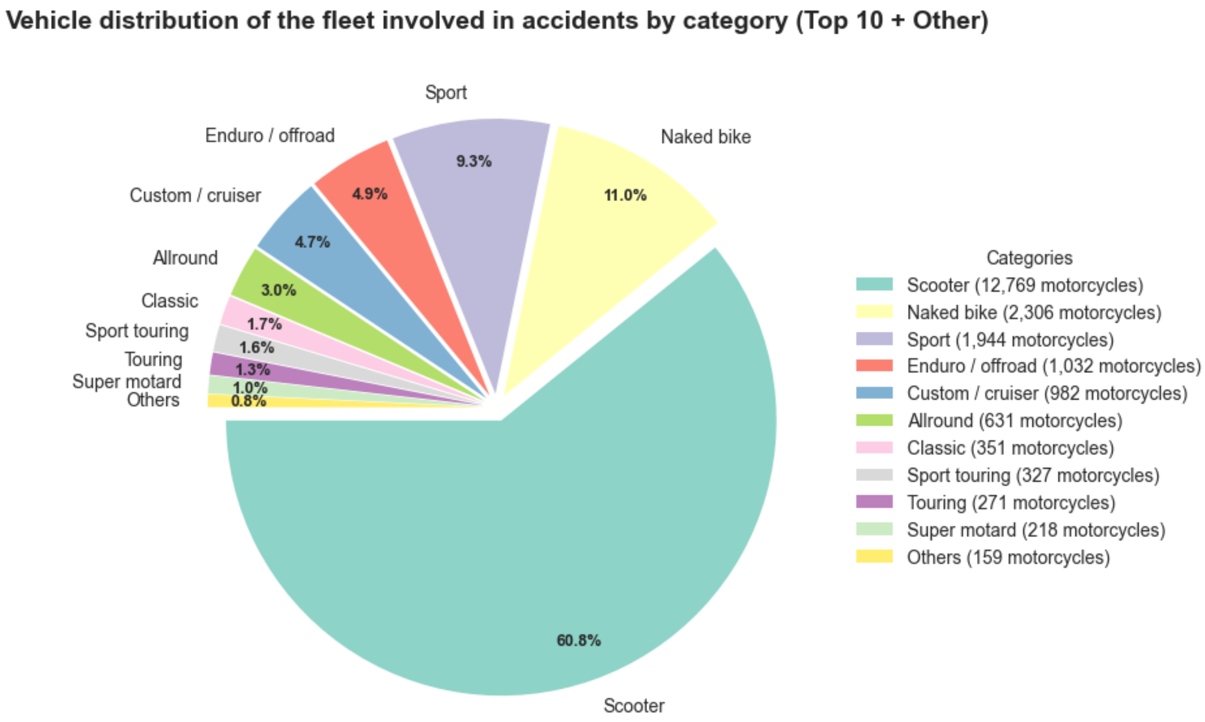

Figure 11. Distribution of motorcycle accidents in 2024 by motorcycle type.

The distribution by motorcycle type is as follows:

1. Scooters:
   - Proportion of total fleet: 51.1%.
   - Proportion of accidents: 60.8%.
   - Although they only represent 51.1% of the fleet, scooters are involved in a higher proportion of accidents (60.8%).
2. Naked bikes:
   - Proportion of total fleet: 8.2%.
   - Proportion of accidents: 11.0%.
   - Naked bikes are over-represented in accidents (+2.8 percentage points).

3. Sport:
   - Proportion of total fleet: 9.1%.
   - Proportion of accidents: 9.3%.
   - Slightly over-represented in accidents (+0.2 percentage points).

4. Enduro/off-road:
   - Proportion of total fleet: 9.1%.
   - Proportion of accidents: 4.9%.
   - Significantly under-represented in accidents (-4.2 percentage points).

Accidents are not distributed proportionately among the various categories, and significant trends can be observed:

1. **Scooters represent the majority of the fleet and account for an even higher proportion of accidents**, perhaps because they are predominantly used in urban environments and are more exposed to traffic.
2. **Naked bikes and sport motorcycles have accident rates proportionately higher than their share of the fleet**, suggesting a higher relative risk, probably owing to their power and associated riding style.
3. **Enduro/off-road motorcycles have a notably lower accident rate**, possibly because they are subject to more specialised use or are ridden in environments where there is less traffic.

In summary, although scooters dominate the figures for both the total fleet and involvement in accidents, the differences in the accident rate between the categories suggest that **the type of motorcycle and its typical use have a significant influence on the accident risk**.

3.3.2. DISTRIBUTION BY MOTORCYCLE AGE

The following figure shows the distribution of motorcycle accidents in Spain in 2024 by age of the motorcycle involved in the accident.

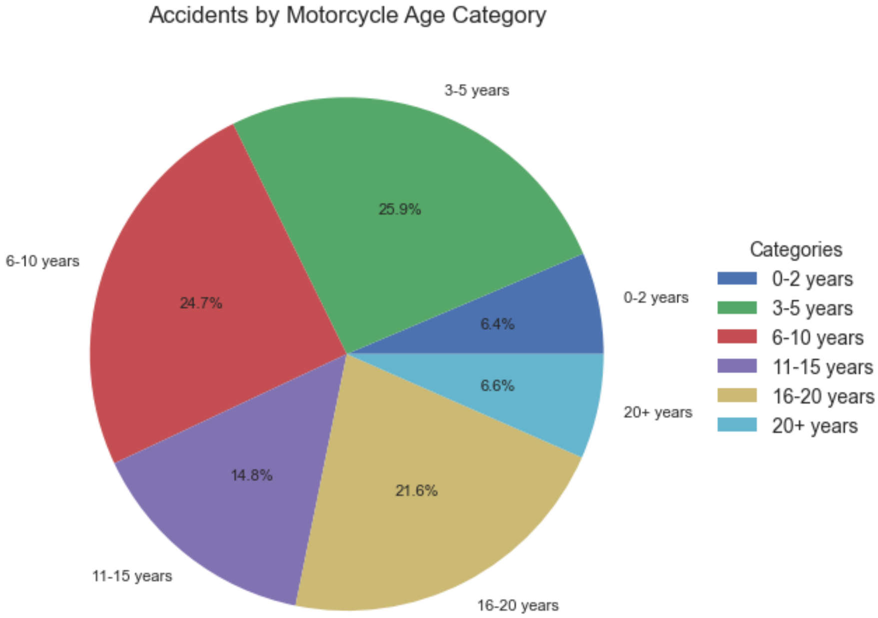

Figure 12. Distribution of motorcycle accidents in 2024 by motorcycle age.

If we compare the distribution of accidents by motorcycle age against the total proportion of the fleet, we find some notable discrepancies:

1. Newer vehicles (0-5 years old):
   - Proportion of total fleet: 16.9%.
   - Proportion of accidents: 32.3% (6.4% + 25.9%).
   - This category is significantly over-represented in accidents (+15.4 percentage points).
2. Vehicles of intermediate age (6-15 years old):
   - Proportion of total fleet: 30.4%.
   - Proportion of accidents: 39.5% (24.7% + 14.8%).

3. Older vehicles (>16 years old):
   - Proportion of total fleet: 52.8%.
   - Proportion of accidents: 28.2% (21.6% + 6.6%).
   - This category is extremely under-represented in accidents (-24.6 percentage points).

In summary, **new motorcycles (0-5 years old) have a disproportionately high accident rate, while the oldest motorcycles are involved in fewer accidents than their proportion of the vehicle fleet would suggest**.

3.3.3. DISTRIBUTION BY MOTORCYCLE CAPACITY

The following figure shows the distribution of motorcycle accidents in Spain in 2024 by capacity of the motorcycle involved in the accident.

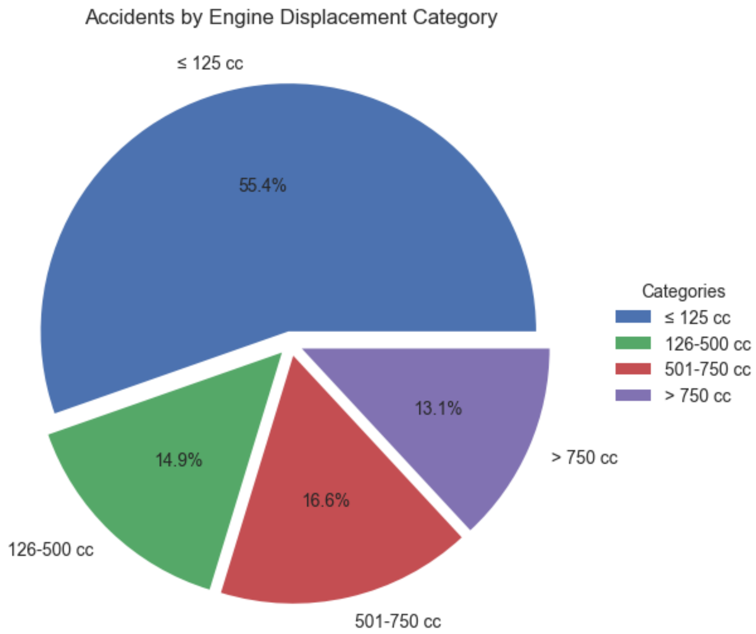

Figure 13. Distribution of motorcycle accidents in 2024 by motorcycle capacity.

If we compare the distribution of accidents by motorcycle capacity against the total proportion of the fleet, we find some significant discrepancies:

1. Motorcycles of 125 cc or lower:
   - Proportion of total fleet: 45.2%.
   - Proportion of accidents: 55.4%.
   - This category is significantly over-represented in accidents (+10.2 percentage points).
2. Intermediate-capacity motorcycles:
   - 126-500 cc: a decrease in proportion from 26.3% to 14.9% (-11.3 percentage points).
   - 501-750 cc: similar proportions, rising slightly from 16% to 16.6% (+0.6 percentage points).
3. High-capacity motorcycles (more than 750 cc):
   - Proportion of total fleet: 12.5%.
   - Proportion of accidents: 13.1%.

In summary, **motorcycles with lower capacities (125 cc or lower) have a disproportionately high accident rate**.

3.3.4. DISTRIBUTION BY AVERAGE NUMBER OF KILOMETRES TRAVELLED PER YEAR BY THE MOTORCYCLE

The following figure shows the distribution of motorcycle accidents in Spain in 2024 by average number of kilometres travelled per year by the motorcycle involved in the accident.

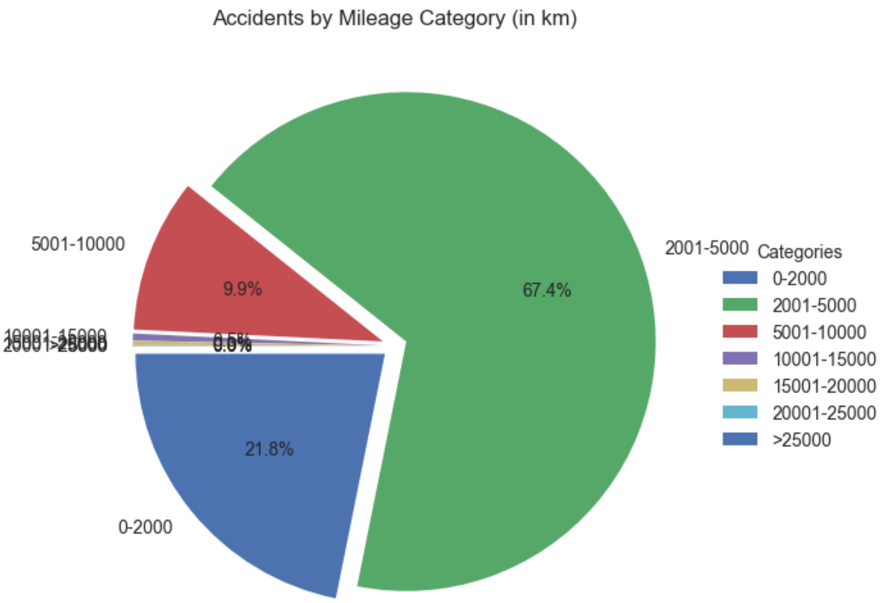

Figure 14. Distribution of motorcycle accidents in 2024 by average number of kilometres travelled per year by the motorcycle.

If we compare the two graphs against Spain's total fleet, we can see the following discrepancies in the distribution by kilometres travelled:

1. Motorcycles travelling less than 2,000 km:
   - Proportion of total fleet: 56.2%.
   - Proportion of accidents: 21.9%.
   - Clearly under-represented in accidents (-34.3 percentage points).
2. Motorcycles travelling 2,000-5,000 km:
   - Proportion of total fleet: 33.6%.
   - Proportion of accidents: 67.4%.
   - Clearly over-represented in accidents (+33.8 percentage points).
3. Motorcycles travelling 5,000-10,000 km:
   - Proportion of total fleet: 8.9%.
   - Proportion of accidents: 9.9%.
   - Slightly over-represented in accidents (+1.0 percentage points).

4. Motorcycles travelling more than 10,000 km:
   - Proportion of total fleet: 1.3%.
   - Proportion of accidents: 0.9%.
   - Under-represented in accidents (-0.4 percentage points).

In summary, **motorcycles that travel fewer kilometres** are enormously under-represented within the group of motorcycles involved in accidents. In contrast, motorcycles travelling an intermediate number of kilometres are clearly over-represented, although this trend is subsequently reversed when we consider that the motorcycles travelling the furthest (more than 10,000 km) are once again under-represented.

4. RESULTS

In this section we analyse the risk of being involved in a serious or fatal accident based on the profile of the motorcyclist and/or motorcycle type.

To assess this risk, and with specific regard to serious accidents, we used three different indicators for two distinct variables:

- **Serious accident total** (the number of accidents with serious injuries in each group).
- **Serious accident rate** (the absolute number of accidents with serious injuries, divided by the total number of motorcycles in the fleet for the group in question).
- **Serious accident IREx** (Exposure-Based Risk Index) (the absolute number of accidents with serious injuries, divided by the total annual kilometres travelled by the group as a whole).

For fatal accidents:

- **Fatal accident total** (the number of fatal accidents in each group).
- **Fatal accident rate** (the absolute number of fatal accidents, divided by the total number of motorcycles in the fleet for the group in question).
- **Fatal accident IREx** (the absolute number of fatal accidents, divided by the total annual kilometres travelled by the group as a whole).

4.1. EXPOSURE-BASED RISK

The **IREx** is an index which was designed and presented in a previous study. It improves the representativeness of the accident or fatality rate by taking into account the kilometres travelled by each rider profile. The following equations show how the index – which is expressed in parts per 100 million (10^8) – is calculated. It should be noted that the index is used to make two separate calculations, one for serious injuries, and the other for fatalities:

IREx = (no. of fatalities/fleet's annual kilometres) * 100,000,000 = (fatality rate/motorcycle's annual kilometres) * 100,000,000

Equation 1: IREx for fatalities.

IREx = (no. of serious injuries/fleet's annual kilometres) * 100,000,000 = (serious accident rate/motorcycle's annual kilometres) * 100,000,000

Equation 2: IREx for serious injuries.

4.1.1. SERIOUS INJURY RATE VS ANNUAL KILOMETRES TRAVELLED

The serious injury rate is obtained by dividing the total number of serious injuries by the total number of motorcycles in the fleet for the group in question. This is a more representative value than the absolute number of serious injuries, as it takes into account the total number of motorcycles in the fleet for each profile, thereby providing a relative view of the number of serious injuries.

To visualise the added value provided by the IREx, the data can also be presented in the form of a scatter graph, with serious accident rate (or fatality rate) on the x-axis and kilometres travelled on the y-axis.

The following figure shows the serious accident rate based on annual kilometres travelled in 2022. We decided to show only those groups for which there is a significant amount of data (i.e. those with a minimum of three accidents, at least one of which is serious), so that the graph did not include incidental situations that might dilute the relevant information.

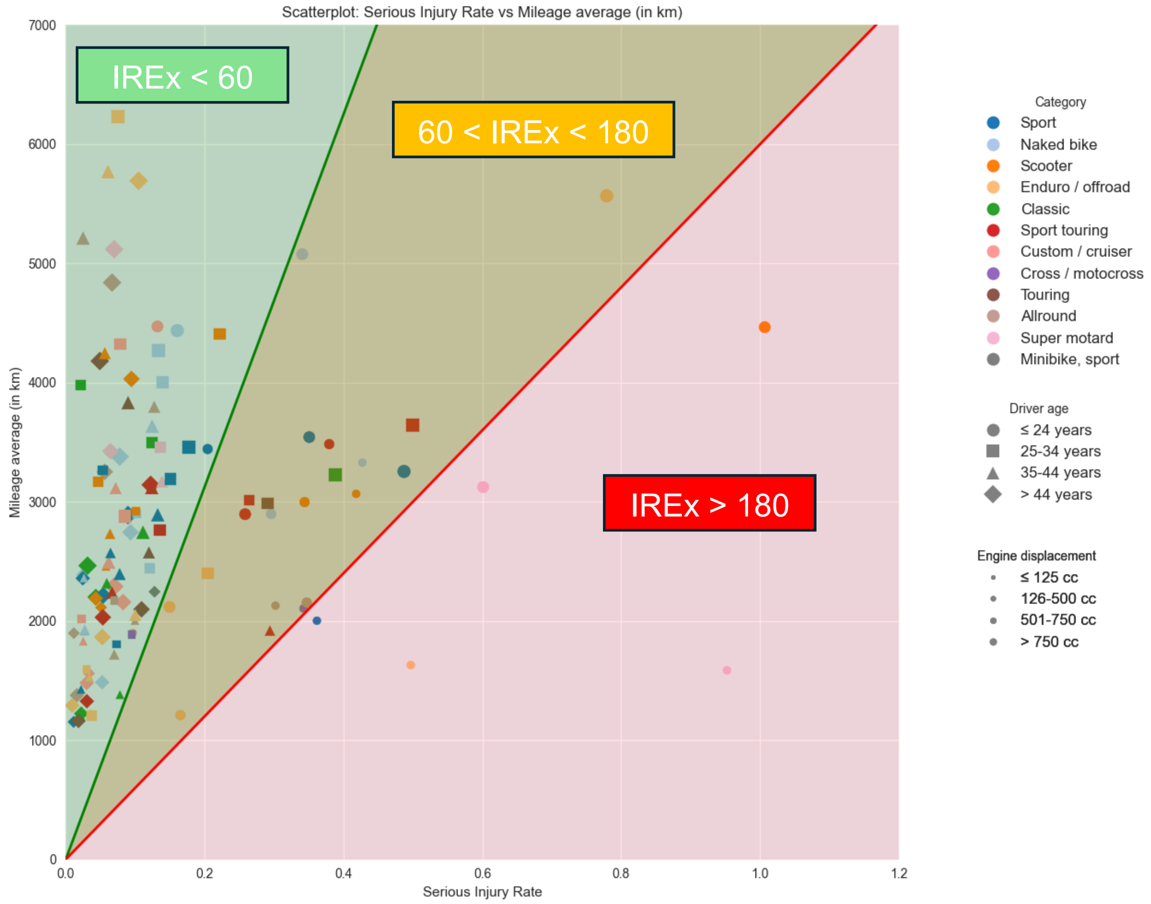

Figure 15. Serious accident rate by kilometres travelled. Each point represents the intersection of category (colour), rider age (shape) and capacity (size).

4.1.2. FATALITY RATE VS ANNUAL KILOMETRES TRAVELLED

The fatality rate is obtained by dividing the total number of fatalities by the total number of motorcycles in the fleet for the group in question. This is a more representative value than the absolute number of fatalities, as it takes into account the total number of motorcycles in the fleet for each profile, thereby providing a relative view of the number of fatalities.

The same type of graph was used to visualise the fatality rate based on kilometres travelled. The following figure shows the fatality rate based on annual kilometres travelled in 2022.

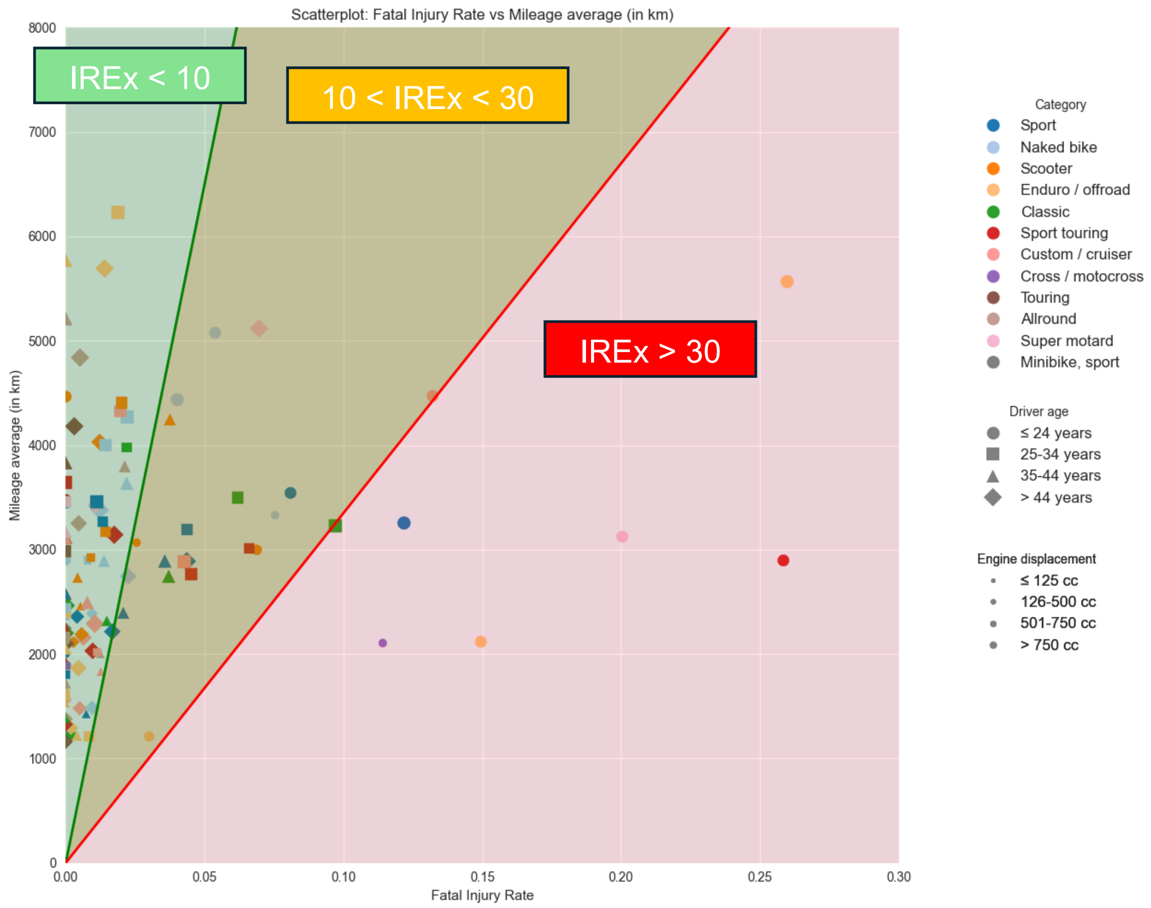

Figure 16. Fatality rate by kilometres travelled.

4.2. ANALYSIS BY PROFILE

4.2.1. MOTORCYCLIST PROFILES WITH A HIGHER SERIOUS INJURY IREx IN SPAIN

To visualise the results more easily, and in accordance with the risk groups defined previously, the following figure shows the groups that have the highest observed risk of serious injury, based on their exposure:

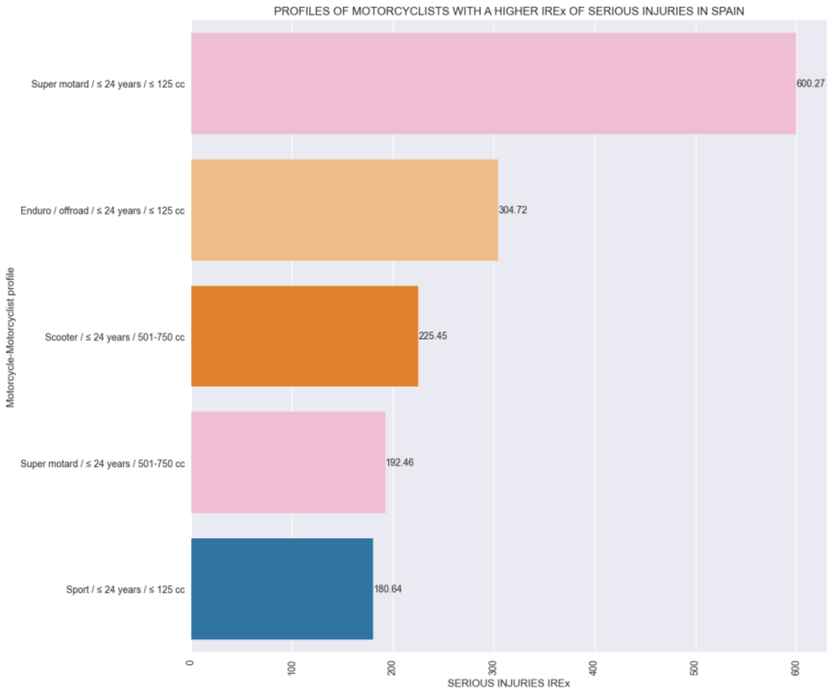

Figure 17. Motorcyclist profiles with a higher serious injury IREx in Spain.

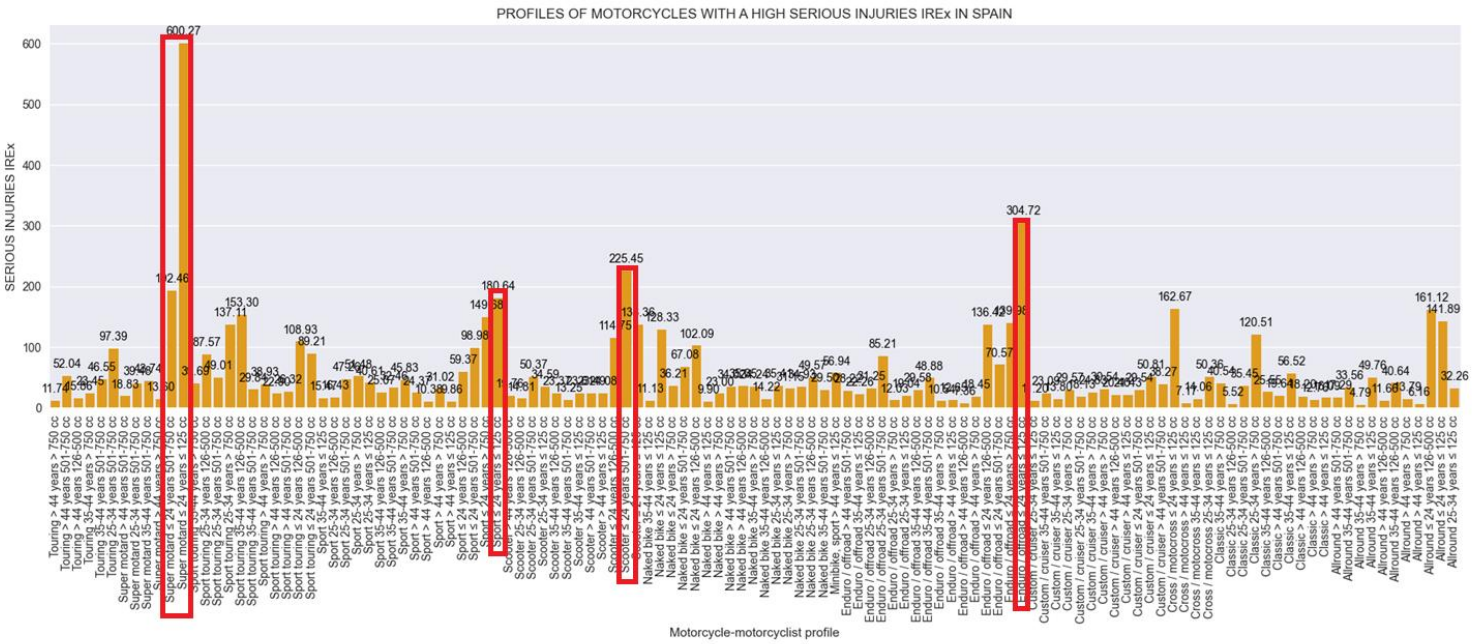

Figure 18. Serious injury IREx for all motorcyclist profiles.

4.2.2. MOTORCYCLIST PROFILES WITH A HIGHER FATALITY IREx IN SPAIN

To visualise the results more easily, and in accordance with the risk groups defined previously, the following figure shows the groups that have the highest observed risk of a fatal accident, based on their exposure:

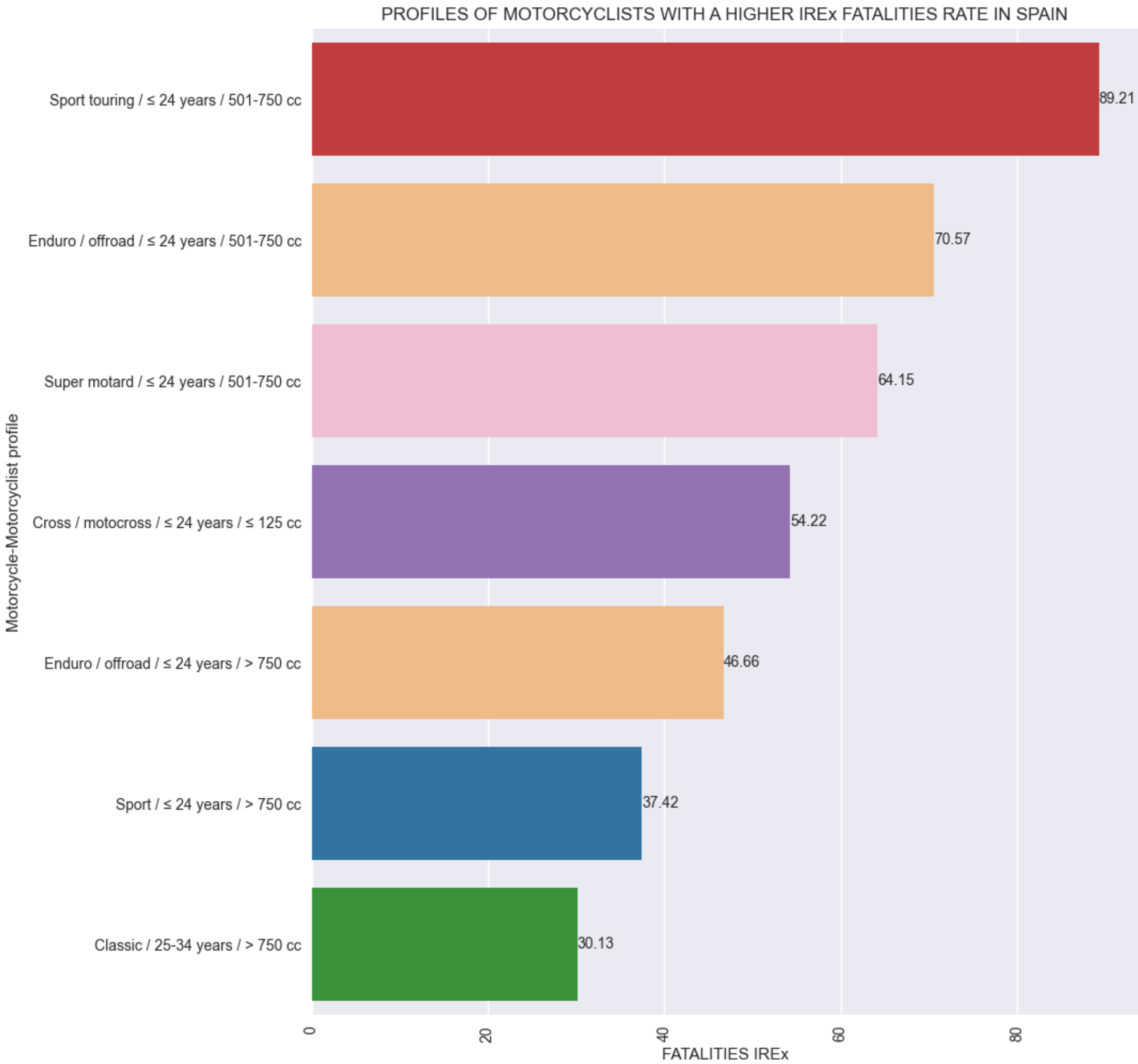

Figure 19. Motorcyclist profiles with a higher fatality IREx in Spain.

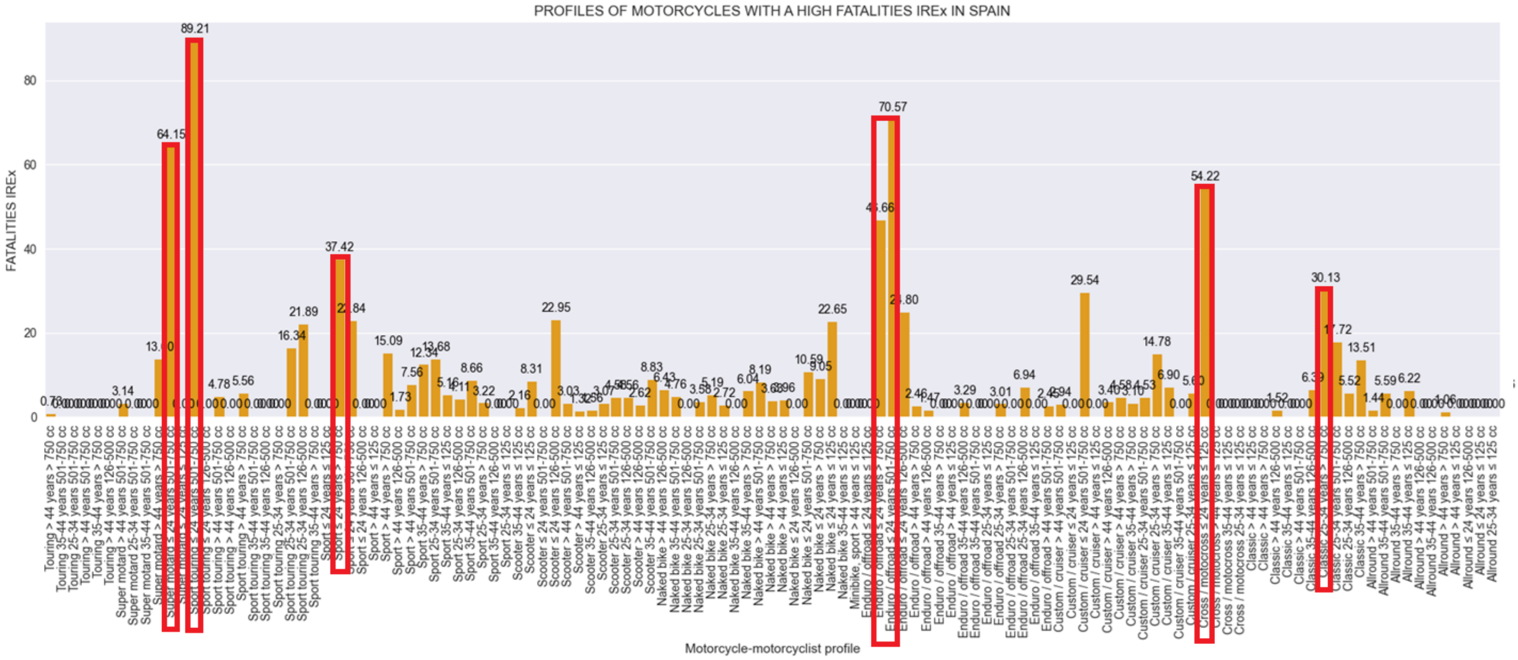

Figure 20. Fatality IREx for all motorcyclist profiles.

4.2.3. IREx CORRESPONDING TO SPORT MOTORCYCLES

The following graph presents a visual illustration of the influence of capacity within the age groups that form part of the category with the highest observed risk:

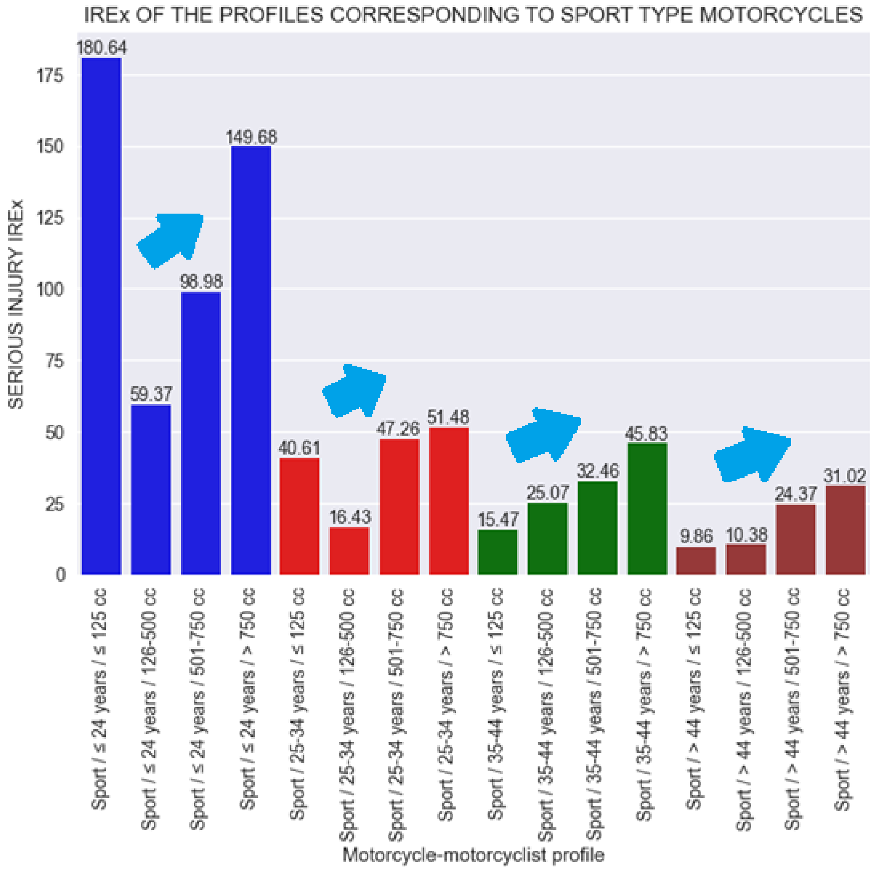

Figure 21. Serious injury IREx for the profiles corresponding to sport motorcycles.

As we can see, within each age group the indicator increases as capacity grows, which suggests that capacity is a decisive risk factor.

4.2.4. SERIOUS ACCIDENTS BY ROAD TYPE

The following graph shows the different groups distributed by road type, and their respective serious accident rate for the road type in question.

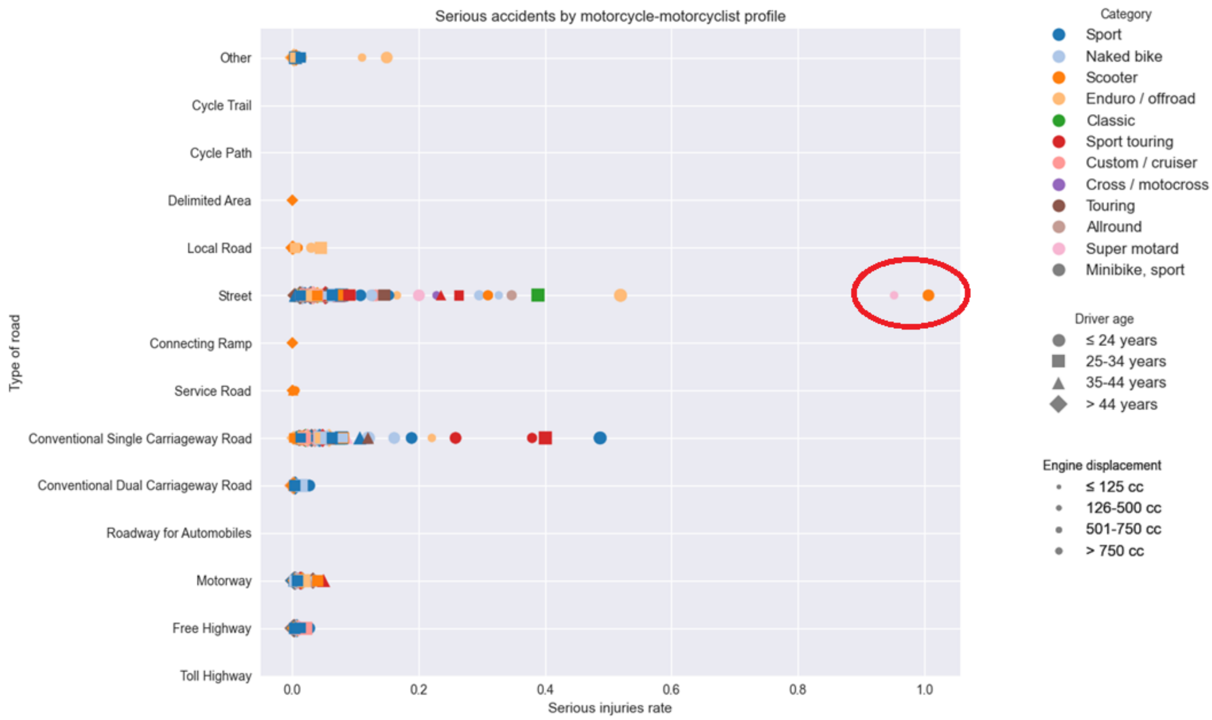

Figure 22. Serious accidents by road type.

4.2.5. FATAL ACCIDENTS BY ROAD TYPE

The following graph shows the different groups distributed by road type, and their respective fatality rate for the road type in question.

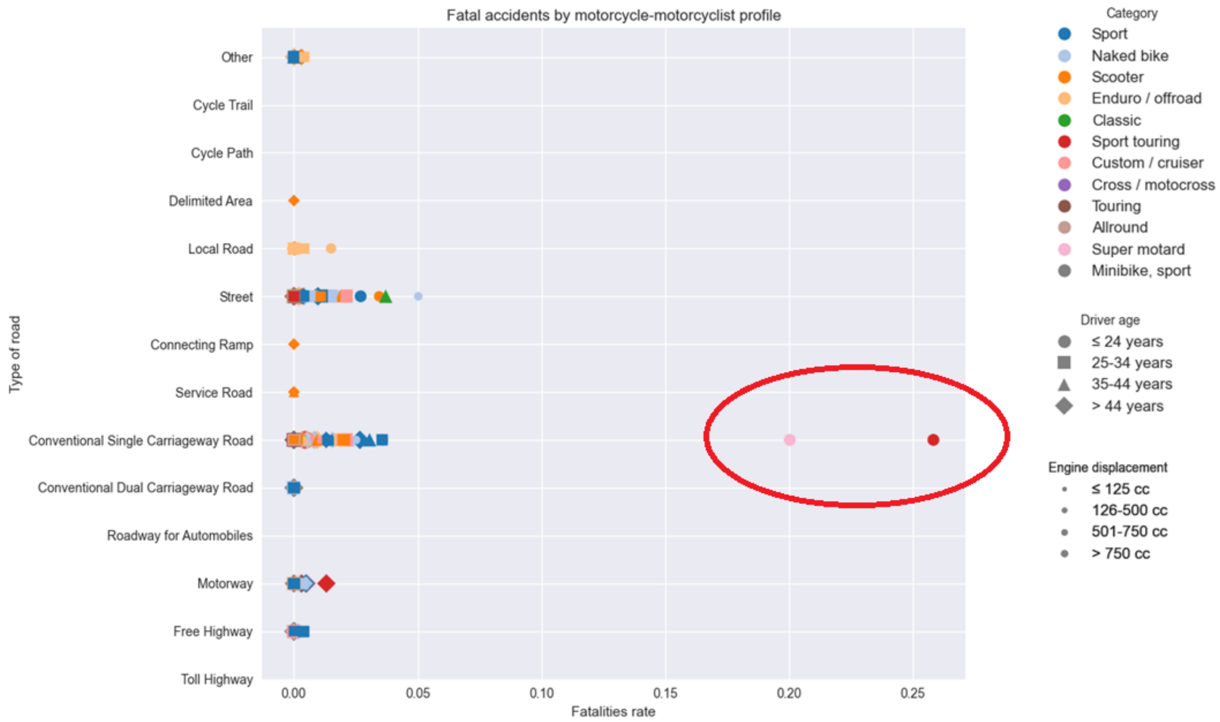

Figure 23. Fatal accidents by road type.

4.3. ANALYSIS OF RESULTS

4.3.1. GRAPH FOR POINT 4.1.1

Analysis by rider age vs risk rate

Young riders (≤ 24 years old)

- **Risk pattern**: they are **more scattered** in the graph, appearing in areas of both low and high risk.
- **Key observation**: some of the points with **higher serious injury rates** correspond to this age group.
- **Relation to kilometres travelled**: this group features in both low and high numbers of kilometres travelled, which suggests that inexperience may be a relevant factor regardless of distance travelled.

Riders of intermediate age (25-34 years old)

- These riders are **moderately concentrated** in the green-beige area (low to moderate risk).

- They have a **significant presence** in the intermediate range of kilometres travelled (2,000-4,000 km).
- Some of the riders of certain motorcycle types (such as enduro/off-road) in this group appear in areas of higher risk.

Experienced riders (35-44 years old)

- These riders are located **predominantly** in the lower-risk area (green).
- **Broad vertical distribution**: regardless of whether they cover a few or many kilometres, their risk rate generally remains low.
- Exception: some triangles appear in the beige area, with moderate serious injury rates.

Older riders (> 44 years old)

- These riders account for the **most well-defined concentration** in the low-risk area (green).
- They have a **smaller presence** in the high- and moderate-risk areas.
- They also account for various points in the **low kilometres travelled and low risk** area, which suggests a more cautious riding approach.

Analysis of capacity vs risk rate

Small motorcycles (≤ 125 cc)

- **Higher concentration** in the lower left area: low kilometres travelled and low serious injury rate.
- **Less of a presence** in high-risk areas.
- Potential correlation with urban use and more moderate speeds.

Medium-sized motorcycles (126-500 cc)

- **Broader distribution** in terms of risk.
- Some points have moderate to high-risk rates.
- Greater variation, possibly related to the more diverse use of these motorcycles.

Large motorcycles (501-750 cc and > 750 cc)

- **Significant presence** in areas with medium to high kilometres travelled.
- Those with the highest capacity (> 750 cc) have **points at the extremes** of the graph:
  - Some are in the very low-risk area (potentially associated with experienced riders).
  - Others are in the high-risk area (potentially associated with increased power and speed).

Key conclusions

1. **Age factor**: the data suggest that **rider experience** has an inverse correlation to the risk of serious injury, with riders over 44 years old presenting the lowest risk profile in general.
2. **Capacity factor**: the relationship between capacity and risk is **non-linear**. Instead, it appears to be moderated by other factors (possibly rider experience and type of use).
3. **Age-capacity interaction**: the combination of **young riders with high-capacity motorcycles** appears to have a stronger presence in high-risk areas.

4. **Risk mitigation**: the data suggest that experience (age) may be a more decisive factor than motorcycle type in reducing the risk of a serious accident.

4.3.2. GRAPH FOR POINT 4.1.2

Analysis by rider age vs fatality risk rate

Young riders (≤ 24 years old)

- **Risk pattern**: some of the points are located in the high-risk area (IREx > 30), which indicates a higher probability of fatal accidents.
- These riders are distributed across the three risk areas, which confirms their greater variation with regard to safety.
- Some young riders of scooters and sport motorcycles **appear in the areas with a higher fatality risk**.

Riders of intermediate age (25-34 years old)

- These riders have a smaller presence in the high-risk area (red) compared to younger riders.
- There is a notable concentration in the green and beige areas (low and medium risk).
- The presence of various points in the yellow area suggests a moderate risk for this group.

Experienced riders (35-44 years old)

- Higher concentration in the green area (low fatality risk).
- Some points in the yellow area, but almost none in the red area.
- The distribution suggests that experience contributes significantly to lowering the risk of a fatal accident.

Older riders (> 44 years old)

- Clear predominance in the low-risk area (green).
- Very small presence in the areas with a high fatality risk.
- The data confirm that **this age group presents the lowest risk profile** for a fatal accident.

Analysis of capacity vs fatality risk rate

Small motorcycles (≤ 125 cc)

- Larger presence in the low-risk area (IREx < 10).
- Some exceptions appear in higher-risk areas, especially scooters.
- The risk appears to be influenced more by type of use than by the motorcycle's capacity.

Medium-sized motorcycles (126-500 cc)

- More balanced distribution between areas of low and medium risk.
- Some models in particular (sport, cross/motocross) appear in the high-risk area.

- This category shows greater variation with regard to fatality risk.

Large motorcycles (501-750 cc and > 750 cc)

- High-capacity motorcycles have a significant presence in all risk areas.
- There are a number of large points in the red area, which suggests that **more powerful motorcycles may pose a higher fatality risk**.
- Notably, some large motorcycles ridden by experienced users remain in the green area, which indicates that experience may mitigate risk even for high-powered vehicles.

Key conclusions

1. **Age factor**: the data confirms an even more marked inverse correlation between age/experience and the risk of a fatal accident, compared to the risk of serious injury.
2. **Capacity factor**: higher-capacity motorcycles have a stronger presence in areas with a higher IREx, which suggests that **power increases the risk of fatal accidents**.
3. Categories with a higher fatality risk:
   - The enduro/off-road and sport categories have points in the area of high fatality risk.
   - The custom/cruiser categories also have points with a high IREx.
4. **Safety implications: young riders with powerful motorcycles consistently feature** as the combination **with the highest risk of a fatal accident**.

4.3.3. GRAPH FOR POINT 4.2.1

Age factor

- **All of the high-risk profiles** correspond to **young riders (≤ 24 years old)**.
- The data decisively confirms the previous analysis: that youth/inexperience is the **most consistent risk factor**.
- No motorcyclist profile over the age of 24 features in this ranking.

Motorcycle type factor

- The **supermotard** category appears twice in the ranking (first and fourth position).
- **Enduro/off-road** motorcycles also present a high risk.
- Surprisingly, **high-capacity scooters** emerge as a significant risk factor.

Capacity factor

- **Counter-intuitive finding**: three of the five highest-risk profiles correspond to **low-capacity motorcycles (≤ 125 cc)**.
- This partially contradicts the common perception that higher capacity = higher risk.
- For the supermotard category, the IREx is **three times higher** for low capacities than it is for medium to high capacities.

Conclusions

1. **Profile with the highest risk**: a young rider with a low-capacity supermotard vehicle presents an exceptionally high risk, possibly due to:
   - The riding style encouraged by these motorcycles (similar to that of the supermoto category), which engenders a false sense of security owing to their low capacity.
2. **Age as a determining factor**: youth is a common factor across all high-risk profiles, regardless of motorcycle type or capacity.
3. **Implications for safety policy**: these data indicate a need to:
   - Develop specific programmes for young riders.
   - Place greater focus on certain types of motorcycle, regardless of their power.
   - Revise the perception that small motorcycles are inherently safer.

4.3.4. GRAPH FOR POINT 4.2.2

Age factor vs fatality rate

- **Six out of seven** profiles correspond to **young riders (≤ 24 years old)**.
- **The data confirm that youth is a determining factor of risk.**
- The only profile with riders aged 25-34 has the lowest IREx in the ranking.

Motorcycle type factor vs fatality rate

- **Sport touring motorcycles** appear to be the most dangerous in terms of fatality rate.
- The data confirm that **enduro/off-road** motorcycles are high risk (they feature twice in the ranking).
- The **supermotard** category remains dangerous in terms of both serious injuries and the fatality rate.

Capacity factor vs fatality rate

- **Significant contrast** with the previous graph: for the fatality rate, five out of seven profiles correspond to **medium-to-high-capacity motorcycles**.
- Only one profile corresponds to a low-capacity motorcycle (≤ 125 cc).
- This suggests that a **higher capacity increases the risk of a fatal accident**, although not necessarily the risk of serious injury.

Conclusions

1. **Profile with the highest fatality risk**: **young riders of medium-capacity sport touring motorcycles have the highest fatality risk**, which may be explained by the following:
   - The higher speeds that these motorcycles can reach.
   - The use of such vehicles on interurban roads, where accidents can be more serious.
   - Lack of experience in handling powerful motorcycles designed for long journeys.

2. **Capacity factor in relation to fatality rate**: unlike the risk of serious injury, **higher capacities do appear to be directly correlated with a higher fatality risk**.
3. **Consistency of the age factor**: the age factor remains decisive for both serious injuries and fatalities.
4. **Implications for safety policy**:
    - There is a need for specific restrictions for young riders of powerful motorcycles.
    - Special attention must be paid to sport touring, enduro and supermotard motorcycles.
    - Specific training programmes are required for handling medium- to high-capacity motorcycles.

4.3.5. GRAPH FOR POINT 4.2.3

Two trends can be observed:

- A decrease in risk as rider age increases.
- An increase in risk as capacity increases, within the same age group.

This reaffirms our previous observation, i.e. that age and capacity are the two most important risk factors.

4.3.6. GRAPH FOR POINT 4.2.4

This graph shows how the groups consisting of **young riders of medium- to high-capacity scooters and low-capacity supermotard motorcycles have a much higher level of risk than all other groups on the same type of road**.

The supermotard category is a specific exception to the **general pattern whereby risk increases in line with the vehicle's capacity**.

4.3.7. GRAPH FOR POINT 4.2.5

This graph shows how the groups consisting of **young riders of medium- to high-capacity sport touring and medium- to high-capacity supermotard motorcycles have a much higher level of risk than all other groups on the same type of road**.

5. CONCLUSIONS

This study on the relationship between motorcycle characteristics and road safety sheds light on several important aspects of Spain's motorcycle fleet and their impact on road safety. With regard to the distribution of the fleet, we can see that scooters clearly predominate, accounting for more than half of all motorcycles. The sport, enduro/off-road and naked categories are next in the ranking, albeit with significantly smaller percentages.

One concerning discovery is that the motorcycle fleet is ageing: over half of all vehicles are more than 10 years old. This suggests a relatively low renewal rate, which may in turn have implications with regard to safety and emissions. In terms of capacity, the data reveal a clear preference for low-powered motorcycles, as nearly half the fleet is comprised of vehicles with a capacity of 125 cc or lower. The data also show that the number of vehicles decreases as engine capacity increases. The figures for patterns of use reveal that the great majority of motorcycles are used for short journeys or in urban environments, with almost 90% of vehicles travelling less than 5,000 km per year. This suggests that motorcycles are chiefly used as a mode of transport for daily journeys, rather than for long journeys or leisure.

With regard to the quality of the data, we have managed to create a dataset of vehicles involved in accidents in which **86%** of the data are fully classified. The equivalent figure for the vehicles in the fleet is **80%**.

In terms of the accident rate, we can make the following affirmations:

Regarding rider age:

1. **Young riders (≤ 24 years old) are the highest-risk group**, featuring in all of the profiles with a high accident rate.
2. **The rider's level of experience drastically reduces the accident risk**, with the safest profile corresponding to riders aged over 44.
3. **None of the profiles corresponding to riders aged over 24 feature in the high-risk rankings**, which confirms that age/experience is the most important protective factor.
4. **Experience offsets the risk posed by powerful motorcycles**: experienced riders still have low risk profiles even with high-capacity motorcycles.

Regarding capacity:

1. **In terms of serious injury, low-capacity (≤ 125 cc) motorcycles may be more dangerous** than medium- to high-capacity vehicles, contrary to popular belief.
2. **With regard to fatal accidents, medium- to high-capacity motorcycles do represent a higher risk**, accounting for five out of seven profiles with a high fatality rate.
3. **The influence of capacity varies according to type of injury**: it has little bearing on serious injuries, but plays a very significant role in fatal accidents.
4. **The risk of serious injury posed by low-capacity supermotard motorcycles is three times higher** than that of medium- to high-capacity models.

Regarding motorcycle type:

1. **Supermotard is the most dangerous category of motorcycle in terms of serious injury**, occupying first and fourth place in the risk ranking.
2. **Sport touring motorcycles are the most lethal**, topping the ranking for fatalities.
3. **Enduro/off-road motorcycles pose a consistently high risk** in terms of both serious injury and the fatality rate.
4. **High-capacity scooters are surprisingly dangerous**, which contradicts the perception that they are safer or easier to handle.

Regarding combinations of factors:

1. **The highest-risk profile for serious injury is young riders with low-capacity supermotard motorcycles**.
2. **The highest-risk profile in terms of fatality rate is young riders with medium-capacity sport touring motorcycles**.
3. **The combination of youth and power is especially lethal** and represents the most significant risk factor for fatal accidents.
4. **Certain groups pose a disproportionately high risk on certain types of roads**, serving to multiply the danger in specific scenarios.

Safety implications:

1. **Policies based solely on capacity restrictions are insufficient**; motorcycle type should also be taken into consideration.
2. **Specific training for young riders is the most pressing measure** given their predominance across all of the high-risk profiles.

3. **The perception that small motorcycles are safer is wrong** and potentially dangerous, especially for certain motorcycle types.

4. **Motorcycle type should be taken into consideration for training programmes and licences**, with particular attention paid to the supermotard, enduro and sport touring categories.

These conclusions underline the complexity of the factors that influence road safety for motorcyclists. The results of this study may be essential to designing more effective and targeted road safety strategies, with a focus on the groups and types of vehicles that pose the highest risk.

ANNEX 1: CLASSIFICATION OF MOTORCYCLE TYPES

The different types of motorcycle involved in the following pilot study were classified in accordance with the following table:

Sport:	**Naked:**
	
Scooter:	**Touring:**
	

Custom/cruiser:	**Sport touring:**
	
Enduro/off-road:	**Supermotard:**
	
Classic:	**All-round:**
	

Figure 24. Classification of motorcycle types.

ANNEX 2: RISK EXPOSURE FOR MOTORCYCLISTS IN SPAIN

This annex contains the summary tables used to create the majority of the graphs for interpreting the results.

The different profiles are ordered in line with the following parameters:

- Motorcycle type
- Rider age
- Motorcycle capacity

The basic information for each parameter found in the table is as follows:

- Absolute number of victims
- Percentage of fatalities and serious injuries in relation to the total number of motorcycles in Spain's fleet
- Exposure-Based Risk Index

Table 1 defines the colour-based code used to facilitate interpretation of the results obtained in Table 2, and classifies said results based on three different levels of risk (low, medium or high).

COLOUR CODE FOR THE NATIONWIDE SUMMARY TABLE								
COLOUR CODE	LEVEL OF RISK	AVERAGE KM (KM/YEAR)	TOTAL SERIOUS INJURIES	TOTAL FATALITIES	SERIOUS INJURY RATE (%)	FATALITY RATE (%)	SERIOUS INJURY IREx (x10e8)	FATALITY IREx (x10e8)
GREEN	LOW	> 10,000	< 20	< 4	< 0.6%	< 0.1%	< 60	< 10
ORANGE	MEDIUM	5,000 – 10,000	20 – 60	4 – 9	0.6% – 1.5%	0.1% – 0.25%	60 – 180	10 – 30
RED	HIGH	< 5,000	> 60	> 9	> 1.5%	> 0.25%	> 180	> 30

Table 1. Code of values for interpreting Table 2.

Motorcycle type	Age (years)	Capacity (cc)	Average km/year	Serious inj.	Fatalities	Serious injury rate	Fatality rate	Serious injury IREx	Fatality IREx
Enduro/off-road	Age ≤ 24	≤ 125 cc	1,631	9	0	0.5%.	0.0%.	304.72	0.0
		126-500 cc	1,211	11	2	0.17%.	0.03%.	136.42	24.8
		501-750 cc	2,118	1	1	0.15%.	0.15%.	70.57	70.57
		> 750 cc	5,567	3	1	0.78%.	0.26%.	139.98	46.66
	Age 25-34	≤ 125 cc	1,590	2	0	0.03%.	0.0%.	19.04	0.0
		126-500 cc	1,209	9	2	0.04%.	0.01%.	31.25	6.94
		501-750 cc	2,401	9	0	0.2%.	0.0%.	85.21	0.0
		> 750 cc	6,231	4	1	0.07%.	0.02%.	12.03	3.01
	Age 35-44	≤ 125 cc	1,528	3	0	0.03%.	0.0%.	22.26	0.0
		126-500 cc	1,219	9	1	0.04%.	0.0%.	29.58	3.29
		501-750 cc	2,050	10	0	0.1%.	0.0%.	48.88	0.0
		> 750 cc	5,769	7	0	0.06%.	0.0%.	10.54	0.0
	Age > 44	≤ 125 cc	1,381	6	0	0.02%.	0.0%.	12.65	0.0
		126-500 cc	1,293	5	1	0.01%.	0.0%.	7.36	1.47
		501-750 cc	1,865	23	2	0.05%.	0.0%.	28.22	2.45
		> 750 cc	5,693	45	6	0.11%.	0.01%.	18.45	2.46
Naked bike	Age ≤ 24	≤ 125 cc	3,329	17	3	0.43%.	0.08%.	128.33	22.65
		126-500 cc	2,898	2	0	0.3%.	0.0%.	102.09	0.0
		501-750 cc	5,078	19	3	0.34%.	0.05%.	67.08	10.59
		> 750 cc	4,437	4	1	0.16%.	0.04%.	36.21	9.05
	Age 25-34	≤ 125 cc	2,897	13	1	0.1%.	0.01%.	35.41	2.72
		126-500 cc	2,442	3	0	0.12%.	0.0%.	49.57	0.0
		501-750 cc	4,001	39	4	0.14%.	0.01%.	34.93	3.58
		> 750 cc	4,270	18	3	0.13%.	0.02%.	31.15	5.19
	Age 35-44	≤ 125 cc	2,360	3	0	0.03%.	0.0%.	11.13	0.0

Motorcycle type	Age (years)	Capacity (cc)	Average km/year	Serious inj.	Fatalities	Serious injury rate	Fatality rate	Serious injury IREx	Fatality IREx
		126-500 cc	1,925	1	0	0.03%.	0.0%.	14.22	0.0
		501-750 cc	2,891	31	5	0.09%.	0.01%.	29.5	4.76
		> 750 cc	3,634	17	3	0.12%.	0.02%.	34.24	6.04
	Age > 44	≤ 125 cc	2,388	5	2	0.02%.	0.01%.	9.9	3.96
		126-500 cc	1,486	11	2	0.05%.	0.01%.	35.35	6.43
		501-750 cc	2,745	54	13	0.09%.	0.02%.	34.02	8.19
		> 750 cc	3,379	19	3	0.08%.	0.01%.	23.0	3.63
Scooter	Age ≤ 24	≤ 125 cc	3,067	82	5	0.42%.	0.03%.	136.36	8.31
		126-500 cc	2,998	10	2	0.34%.	0.07%.	114.75	22.95
		501-750 cc	4,465	6	0	1.01%.	0.0%.	225.45	0.0
	Age 25-34	≤ 125 cc	2,923	169	15	0.1%.	0.01%.	34.59	3.07
		126-500 cc	3,169	13	4	0.05%.	0.01%.	14.81	4.56
		501-750 cc	4,407	11	1	0.22%.	0.02%.	50.37	4.58
	Age 35-44	≤ 125 cc	2,458	153	14	0.06%.	0.01%.	23.61	2.16
		126-500 cc	2,731	45	3	0.06%.	0.0%.	23.37	1.56
		501-750 cc	4,246	6	4	0.06%.	0.04%.	13.25	8.83
	Age > 44	≤ 125 cc	2,114	309	17	0.05%.	0.0%.	24.08	1.32
		126-500 cc	2,190	136	18	0.04%.	0.01%.	19.76	2.62
		501-750 cc	4,031	31	4	0.09%.	0.01%.	23.49	3.03
Sport	Age ≤ 24	≤ 125 cc	2,003	7	0	0.36%.	0.0%.	180.64	0.0
		126-500 cc	3,443	4	0	0.2%.	0.0%.	59.37	0.0
		501-750 cc	3,543	13	3	0.35%.	0.08%.	98.98	22.84
		> 750 cc	3,255	4	1	0.49%.	0.12%.	149.68	37.42
	Age 25-34	≤ 125 cc	1,806	6	0	0.07%.	0.0%.	40.61	0.0
		126-500 cc	3,265	4	1	0.05%.	0.01%.	16.43	4.11
		501-750 cc	3,192	38	11	0.15%.	0.04%.	47.26	13.68
		> 750 cc	3,456	16	1	0.18%.	0.01%.	51.48	3.22
	Age 35-44	≤ 125 cc	1,426	3	1	0.02%.	0.01%.	15.47	5.16
		126-500 cc	2,571	6	0	0.06%.	0.0%.	25.07	0.0
		501-750 cc	2,393	30	8	0.08%.	0.02%.	32.46	8.66
		> 750 cc	2,891	26	7	0.13%.	0.04%.	45.83	12.34
	Age > 44	≤ 125 cc	1,154	3	0	0.01%.	0.0%.	9.86	0.0
		126-500 cc	2,358	6	1	0.02%.	0.0%.	10.38	1.73
		501-750 cc	2,215	29	9	0.05%.	0.02%.	24.37	7.56
		> 750 cc	2,889	37	18	0.09%.	0.04%.	31.02	15.09
Sport touring	Age ≤ 24	126-500 cc	3,484	2	0	0.38%.	0.0%.	108.93	0.0
		501-750 cc	2,896	1	1	0.26%.	0.26%.	89.21	89.21
	Age 25-34	126-500 cc	3,015	4	1	0.26%.	0.07%.	87.57	21.89
		501-750 cc	2,764	3	1	0.14%.	0.05%.	49.01	16.34
		> 750 cc	3,647	5	0	0.5%.	0.0%.	137.11	0.0
	Age 35-44	126-500 cc	1,919	5	0	0.29%.	0.0%.	153.3	0.0
		501-750 cc	2,238	3	0	0.07%.	0.0%.	29.84	0.0
		> 750 cc	3,119	5	0	0.12%.	0.0%.	39.69	0.0
	Age > 44	126-500 cc	1,327	2	0	0.03%.	0.0%.	22.9	0.0
		501-750 cc	2,031	11	2	0.05%.	0.01%.	26.32	4.78
		> 750 cc	3,142	28	4	0.12%.	0.02%.	38.93	5.56
Supermotard	Age ≤ 24	≤ 125 cc	1,587	1	0	0.95%.	0.0%.	600.27	0.0
		501-750 cc	3,124	3	1	0.6%.	0.2%.	192.46	64.15
	Age 25-34	501-750 cc	3,460	4	0	0.14%.	0.0%.	39.48	0.0
	Age 35-44	501-750 cc	3,169	8	0	0.14%.	0.0%.	43.74	0.0
	Age > 44	501-750 cc	3,424	12	2	0.06%.	0.01%.	18.83	3.14
		> 750 cc	5,119	1	1	0.07%.	0.07%.	13.6	13.6

Table 2. Summary of serious injuries and fatalities among motorcyclists in Spain during 2022