Data in Brief 53 (2024) 110084

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Spatio-temporal dynamics of vehicles: Fusion of traffic data and context information

Daniel Bolaños-Martinez [a,*], Maria Bermudez-Edo [a], Jose Luis Garrido [a], Blanca L. Delgado-Márquez [b]

[a] Computer Science School and Research Centre for Information and Communication Technologies (CITIC-UGR), University of Granada, C/ Periodista Daniel Saucedo Aranda S/N, 18071 Granada, Spain
[b] Department of Business Management and European Institute of Sustainability Management, University of Granada, Campus of Cartuja, 18071 Granada, Spain

ARTICLE INFO

Article history:
Received 2 December 2023
Revised 11 January 2024
Accepted 16 January 2024
Available online 1 February 2024

Dataset link: Federation of Vehicular Data in Smart Villages with Socioeconomic Information (Original data)

Keywords:
Mobility analysis
Urban planning
Tourism data
LPR dataset
Vehicle tracking
Sierra Nevada

ABSTRACT

We present a dataset for vehicle tracking in a rural area. Specifically, in the Barranco de Poqueira region, which includes the municipalities of Pampaneira, Bubión, and Capileira in the Sierra Nevada National Park, Granada, Spain. Four Hikvision License Plate Recognition (LPR) cameras collect vehicle entries and exits to each village. Additional contextual data, including vacation calendars, vehicle origins, and socio-demographic information, enrich the dataset. The dataset comprises three files covering nine months from February to October 2022: one with raw data directly extracted from the cameras, another aggregated at the visit level and including context information, and a third aggregated by vehicles with context information. These datasets can be useful for mobility studies, urban planning, tourism, and socio-demographic analysis.

DOI of original article: 10.1016/j.inffus.2023.102164
* Corresponding author.
E-mail address: danibolanos@ugr.es (D. Bolaños-Martinez).
Social media: @d4nibomar (D. Bolaños-Martinez), @mariaberm (M. Bermudez-Edo), @bdelgadoUGR (B.L. Delgado-Márquez)

https://doi.org/10.1016/j.dib.2024.110084

Specifications Table

Subject	Transportation Management
Specific subject area	Spatio-temporal behavior of vehicles enriched with contextual information.
Data format	Raw and filtered data.
Type of data	Three tables (CSV format).
Data collection	The data was collected mainly through four Hikvision cameras with Automatic License Plate Recognition (ANPR) based on Deep Learning. The devices have 2MP resolution, 2.8-12 mm varifocal optics, and IR LEDs with a 50 m range. The files were constructed using additional information on holiday calendars (obtained from the holidays library in Python), vehicle origin (provided by the Spanish Directorate-General for Traffic (DGT)), distance in kilometers from the origin (geopy and pgeocode libraries in Python) and sociodemographic values (publicly available at the National Statistics Institute of Spain (Spanish: Instituto Nacional de Estadística, INE)).
Data source location	City/Town/Region: Pampaneira, Bubión, and Capileira. Alpujarra Region, Granada. Country: Spain Latitude and longitude of the LPR cameras: LPR Pampaneira 1 (36.939969, -3.363271) LPR Pampaneira 2 (36.938225, -3.360855) LPR Bubion (36.9457788, -3.3539488) LPR Capileira (36.960006, -3.358317)
Data accessibility	Repository name: Federation of Vehicular Data in Smart Villages with Socioeconomic Information Data identification number: 10245475 Direct URL to data: https://doi.org/10.5281/zenodo.10245475
Related research article	[1] Daniel Bolaños-Martinez, Maria Bermudez-Edo, Jose Luis Garrido, Clustering pipeline for vehicle behavior in smart villages, Information Fusion, 2023, 102164, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2023.102164.

1. Value of the Data

- Real-time traffic data aids in understanding and managing the impact of visitor influx on rural areas and national parks, facilitating sustainable tourism management strategies.
- Researchers in the fields of environmental science, ecology, and transportation can leverage this dataset to analyze the correlation between tourist activity and environmental impact, contributing to the development of effective conservation and traffic management policies [2].
- These data can be reused by other researchers to conduct comparative studies across different rural regions and national parks, enabling the identification of common patterns and the formulation of generalized sustainable management practices.
- The dataset provides an opportunity for interdisciplinary research collaboration, fostering partnerships between researchers, policymakers, and local communities to implement data-driven strategies for the preservation of natural landscapes while promoting responsible tourism.
- Additionally, the dataset proves valuable for generating predictive models and implementing machine learning techniques to optimize tourism management practices and anticipate visitor patterns [1], contributing to more efficient resource allocation and visitor experience enhancement.

2. Background

Our dataset [3] offers information for rural mobility, particularly focusing on vehicle tracking in the Alpujarras region, which includes the municipalities of Pampaneira, Bubión, and Capileira, located in the Sierra Nevada National Park, Granada, Spain (see Fig. 1). It uses four LPR cameras

Fig. 1. Location of the study area in the Alpujarra region, Granada, Spain (Adapted from [4]).

to track vehicles entering and exiting the area, and the movements inside the area in each visit, providing a detailed view of their mobility. We have three cameras at the border of the area and we calculate the route of each visit as the cameras that detect the vehicle between the entry and exit cameras. If a vehicle exits through any exit camera and more than 30 minutes pass, the next entry into the zone will be considered a new visit in our dataset. Our dataset is unique as it contains anonymized plate numbers, from which we derived other variables such as time spent in the area or frequency of visits, which is not a common feature in the datasets.

We expanded the information from the cameras with different sets of contextual data, such as vacation calendars, vehicle origins, and socio-demographic data (see section "Experimental Design, Materials and Methods"). Information such as vehicle origin, is not displayed in any other public dataset. This information enriches the studies and analyses performed with our data.

3. Data Description

The dataset corresponds to the period between February and October 2022 (nine months). The information published in Zenodo is in CSV format. We have also divided the information into three separate files. The files, despite containing overlapping information, offer the information from three different perspectives: (i) centered on the vehicle (each row represents one vehicle, with all the visits to the area); (ii) centered on the visit (each row represents one visit of one vehicle); and (iii) centered on the raw data (each row represents one vehicle crossing one camera).

Table 1 describes the 43 variables in all files along with their name, data type, source data category (see section "Experimental Design, Materials, and Methods" for details), and a description of the variable.

Table 1
Variables description.

Variable	Type	Data source	Description
num_plate_ID	Integer (categorical)	LPR cameras, Vehicle information	Number that identifies each vehicle. It is the license plate number anonymized. Although it is a number it behaves as a categorical attribute.
camera_ID	Categorical	LPR cameras	Identifier of LPR device.
Date	Datetime (yyyy-mm-dd hh:mm:ss+TIME_ZONE)	LPR cameras,National calendar	Date and time of detection of the specific vehicle by a camera.
Direction	Categorical	LPR cameras	Binary value indicating whether the vehicle enters or leaves the village.
entry_cam	Categorical	LPR cameras	Vehicle entry camera_ID. It is the identification of the camera from which the vehicle enters the whole area in one visit.
exit_cam	Categorical	LPR cameras	Vehicle exit camera_ID. The camera from which the vehicle exits the area.
entry_date	Date (dd/mm/yyyy)	LPR cameras	Date of entry of the vehicle in the area in one visit.
exit_date	Date (dd/mm/yyyy)	LPR cameras	Date of exit of the vehicle of the area in one visit.
entry_time	Time (hh:mm:ss)	LPR cameras	Time of vehicle entry.
exit_time	Time (hh:mm:ss)	LPR cameras	Time of vehicle exit.
visit_time	Timedelta	LPR cameras	Time of stay of the vehicle in the area in one visit.
Route	Categorical	LPR cameras	List of all cameras_ID by which the vehicle has been recorded during one visit.
Distance	Float	LPR cameras	Distance traveled in kilometers by the vehicle within the area in one visit.
num_holiday	Integer	National calendar	Number of holiday days spent in the area.
num_workday	Integer	National calendar	Number of workday days spent in the area.
num_high_season	Integer	National calendar	Number of high season days spent in the area. Includes important holiday periods in Spain: Summer Holiday, Christmas and Holy Week.
num_low_season	Integer	National calendar	Number of low season days spent in the area.
Nights	Integer	LPR cameras	Number of overnights in the area.
visits_dif_weeks	Integer	LPR cameras	Number of different weeks with at least 1 visit.
visits_dif_months	Integer	LPR cameras	Number of different months with at least 1 visit.
entry_in_holiday	Integer	National calendar	Number of entries to the area on holiday by vehicle.
entry_in_high_season	Integer	National calendar	Number of entries to the area in high season by vehicle.
Country	Categorical (dichotomous)	Vehicle information	Indicates whether the vehicle comes from Spain or abroad (Other).
km_to_dest	Float	Geographic data	Distance in kilometers between the origin of the vehicle and the destination region (Pampaneira).
Population	Integer	Demographic and Economic data	Population size of the city/town of the provenance of the vehicle.
avg_gross_income	Float	Demographic and Economic data	Average gross income of the area of origin of the vehicle.
avg_disposable_income	Float	Demographic and Economic data	Average disposable income of the area of origin of the vehicle.

(*continued on next page*)

Table 1 (*continued*)

Variable	Type	Data source	Description
autonomous_community	Categorical	Geographic data	Spanish autonomous community of provenance of the vehicle.
Province	Categorical	Geographic data	Spanish province of provenance of the vehicle.
total_entries	Integer	LPR cameras	Total number of vehicle entries to the area.
avg_visit	Timedelta	LPR cameras	Average vehicle visit time in the area.
std_visit	Timedelta	LPR cameras	Standard deviation of the mean number of vehicle visits in the area.
avg_nights	Float	LPR cameras	Average number of vehicle overnights in the area.
std_nights	Float	LPR cameras	Standard deviation of the average number of vehicle nights in the area.
avg_holiday	Float	National calendar	Average number of holidays of the vehicle in the area.
std_holiday	Float	National calendar	Standard deviation of the average number of vehicle holidays in the area.
avg_workday	Float	National calendar	Average number of workdays of the vehicle in the area.
std_workday	Float	National calendar	Standard deviation of the average number of vehicle workdays in the area.
avg_high_season	Float	National calendar	Average number of days of high season for the vehicle in the area.
std_high_season	Float	National calendar	Standard deviation of the average number of days of high season for vehicles in the area.
avg_low_season	Float	National calendar	Average number of days of low season for the vehicle in the area.
std_low_season	Float	National calendar	Standard deviation of the average number of days of low season for vehicles in the area.
Postcode	Categorical	Vehicle information, Geographic, Demographic and Economic data	National postal code of the vehicle's provenance.

Table 2
RAW_SMART_POQUEIRA dataset summary statistics - datetime variables.

Variable	Min	Max	Median	Unique
Date	2022-02-04 12:19:55	2022-10-31 23:57:45	2022-07-30 10:58:23	993,240

Table 3
RAW_SMART_POQUEIRA dataset summary statistics - categorical variables.

Variable	Unique	Top counts
camera_ID	4	PAM_2: 302,744 CAP: 268,234 PAM_1: 267,510 BUB: 196,263
Direction	2	IN: 546,851 OUT: 487,900

The RAW_SMART_POQUEIRA.csv file contains 1,034,751 rows with 161,772 different registered vehicles. It contains information about 4 variables: num_plate_ID, camera_ID, date, and direction (see Table 1). These raw data have been extracted from the LPR cameras source, and a small preprocessing has been applied to the variables camera_ID and direction to facilitate their interpretation. Tables 2 and 3 show the descriptive statistics of the variables. In Tables 3, 5, and 7,

Table 4
VISITS_SMART_POQUEIRA dataset summary statistics - date and time variables.

Variable	Min	Max
entry_date	04/02/2022	31/10/2022
exit_date	06/02/2022	31/10/2022
entry_time	00:00:20	23:59:44
exit_time	00:00:16	23:59:57

Table 5
VISITS_SMART_POQUEIRA dataset summary statistics - categorical variables

Variable	Unique	Top counts
entry_cam	4	PAM_1 59,136 PAM_2 39,455 BUB 16,330 CAP 14,446
exit_cam	3	PAM_1 58,282 PAM_2 47,260 BUB 23,825
Route	12,181	['PAM_1', 'PAM_2'] 18,021 ['PAM_2', 'PAM_1'] 11,280 ['PAM_1', 'PAM_1'] 9,376 ['PAM_2', 'PAM_2'] 8,940 ['PAM_2', 'PAM_1', 'PAM_2'] 4,466
Country	2	España 122,816 Other 6,514
autonomous_community	19	Andalucía 90,970 Comunidad de Madrid 13,255 Catalunya 5,218 Comunitat Valenciana 3,426 Región de Murcia 2,476
Province	52	Granada 69,674 Comunidad de Madrid 13,255 Málaga 8,070 Barcelona 4,365 Almería 4,286

we have added a column named 'Top Counts'. This column shows all the distinct values present in the file for each categorical variable, along with their respective frequencies. In cases where there is a multitude of unique values, we present the top 5 values ordered by their occurrence frequency.

The file VISITS_SMART_POQUEIRA.csv contains 129,367 rows with 50,901 different registered vehicles. It contains information about 26 variables: num_plate_ID, entry_cam, entry_date, entry_time, exit_cam, exit_date, exit_time, visit_time, route, distance, num_holiday, num_workday, num_high_season, num_low_season, nights, visits_dif_weeks, visits_dif_months, entry_in_holiday, entry_in_high_season, country, km_to_dest, population, avg_gross_income, avg_disposable_income, autonomous_community, and province (see Table 1). This file defines a complete entry and exit to the area by a vehicle in each row. Hence, multiple rows exist for the same vehicle if it has made multiple visits to the area. Tables 4, 5, and 6 show the descriptive statistics for all variables in the file.

The file VEHICLES_SMART_POQUEIRA.csv contains 50,901 rows with 50,901 different registered vehicles. It contains information about 33 variables: num_plate_ID, visit_time, distance, num_holiday, num_workday, num_high_season, num_low_season, entry_in_high_season, entry_in_holiday, nights, visits_dif_weeks, visits_dif_months, total_entries, avg_visit, std_visit, avg_nights, std_nights, avg_holiday, std_holiday, avg_workday, std_workday, avg_high_season, std_high_season, avg_low_season, std_low_season, route, country, km_to_dest, population, avg_gross_income, avg_disposable_income, autonomous_community, and province (see Table 1).

Table 6
VISITS_SMART_POQUEIRA dataset summary statistics - integer, numeric, and timedelta variables.

Variable	Mean	Std	Min	Max	P25	P50	P75
visit_time	8 days 11:17:32	26 days 15:53:49	0 days 00:00:00	267 days 19:55:45	0 days 00:46:21	0 days 05:45:24	2 days 20:16:31
distance	6.267	10.557	0.0	489.5	1.5	3.0	8.5
num_holiday	2.881	8.068	0.0	83.0	0.0	1.0	2.0
num_workday	6.570	18.640	0.0	186.0	1.0	1.0	3.0
num_high_season	1.610	5.316	0.0	39.0	0.0	0.0	1.0
num_low_season	7.841	22.640	0.0	230.0	1.0	1.0	3.0
nights	8.448	26.667	0.0	268.0	0.0	0.0	3.0
visits_dif_weeks	3.463	4.186	1.0	31.0	1.0	1.0	4.0
visits_dif_months	1.772	1.309	1.0	9.0	1.0	1.0	2.0
entry_in_holiday	0.300	0.458	0.0	1.0	0.0	0.0	1.0
entry_in_high_season	0.226	0.418	0.0	1.0	0.0	0.0	0.0
km_to_dest	153.874	221.194	0.113	1700.62	5.705	33.580	242.423
population	148,399.925	472,646.506	157.0	3,305,408.0	10,726.0	43,013.0	58,545.0
avg_gross_income	22,440.267	7,473.958	12,638.0	79,327.0	16,084.0	19,976.0	26,947.0
avg_disposable_income	18,855.488	5,315.840	11,218.0	57,956.0	14,342.0	17,097.0	22,202.0

Table 7
VEHICLES_SMART_POQUEIRA dataset summary statistics - categorical variables.

Variable	Unique	Top counts
route	11,998	['PAM_1', 'PAM_1'] 5,727 ['PAM_1', 'PAM_2'] 5,144 ['PAM_2', 'PAM_1'] 1,395 ['PAM_1', 'PAM_2', 'BUB', 'BUB', 'CAP', 'CAP', 'BUB', 'BUB', 'PAM_2', 'PAM_1'] 1,129 ['BUB', 'PAM_1'] 991
country	2	España 46,815 Other 4,051
autonomous_community	19	Andalucía 27,665 Comunidad de Madrid 7,595 Catalunya 3,074 Comunitat Valenciana 2,385 Región de Murcia 1,684
province	52	Granada 13,882 Comunidad de Madrid 7,595 Málaga 5,104 Almería 2,835 Barcelona 2,539

This file groups the information from the VISITS_SMART_POQUEIRA.csv file at the vehicle level, so there is one row per vehicle, which defines its own behavior based on all its accumulated visits in the area. Tables 7 and 8 show the descriptive statistics for all variables in the file.

4. Experimental Design, Materials, and Methods

4.1. LPR cameras data

Our primary information source was the vehicle tracking system, specifically the LPR cameras. We strategically positioned these four cameras to cover the entrances and exits of each village in the target area, as shown in Fig. 2. The locations were (i) entrance to Pampaneira from

Table 8
VEHICLES_SMART_POQUEIRA dataset summary statistics - Integer, numeric and timedelta variables.

Variable	Mean	Std	Min	Max	P25	P50	P75
visit_time	21 days 12:40:33	50 days 14:17:32	0 days 00:00:07	268 days 11:31:40	0 days 01:51:19	0 days 05:23:06	2 days 20:12:33
Distance	15.928	58.339	0.0	2,213.0	1.5	7.0	13.0
num_holiday	7.323	15.949	0.0	131.0	0.0	1.0	2.0
num_workday	16.698	37.786	0.0	367.0	0.0	1.0	4.0
num_high_season	4.092	9.423	0.0	75.0	0.0	0.0	2.0
num_low_season	19.929	45.404	0.0	378.0	1.0	1.0	4.0
entry_in_holiday	0.763	1.823	0.0	91.0	0.0	0.0	1.0
entry_in_high_season	0.574	1.534	0.0	57.0	0.0	0.0	1.0
Nights	21.472	50.599	0.0	269.0	0.0	0.0	3.0
visits_dif_weeks	1.511	2.012	1.0	31.0	1.0	1.0	1.0
visits_dif_months	1.210	0.741	1.0	9.0	1.0	1.0	1.0
total_entries	2.541	7.247	1.0	307.0	1.0	1.0	2.0
avg_visit	10 days 14:15:42	31 days 18:28:39	0 days 00:00:07	267 days 19:55:45	0 days 01:38:45	0 days 04:49:18	1 days 22:14:27
std_visit	4 days 07:04:15	16 days 03:53:45	0 days 00:00:00	182 days 11:47:14	0 days 00:00:00	0 days 00:00:00	0 days 00:34:03
avg_nights	10.557	31.773	0.0	268.0	0.0	0.0	2.0
std_nights	4.296	16.165	0.0	182.433	0.0	0.0	0.0
avg_holiday	3.639	9.584	0.0	83.0	0.0	1.0	2.0
std_holiday	1.335	4.860	0.0	55.154	0.0	0.0	0.0
avg_workday	7.918	22.218	0.0	186.0	0.0	1.0	2.0
std_workday	3.032	11.302	0.0	127.279	0.0	0.0	0.0
avg_high_season	2.088	5.985	0.0	39.0	0.0	0.0	1.0
std_high_season	0.863	3.288	0.0	27.577	0.0	0.0	0.0
avg_low_season	9.469	26.631	0.0	230.0	1.0	1.0	2.5
std_low_season	3.692	13.725	0.0	154.856	0.0	0.0	0.0
km_to_dest	238.491	233.913	0.113	1700.622	36.516	161.880	390.694
population	191,723.207	556,626.029	157.0	3,305,408.0	7,557.0	25,611.0	111,932.0
avg_gross_income	24,703.433	7,675.179	12,638.0	79,327.0	18,779.0	23,569.0	28,738.0
avg_disposable_income	20,464.639	5,463.695	11,218.0	57,956.0	15,932.0	19,775.0	23,520.0

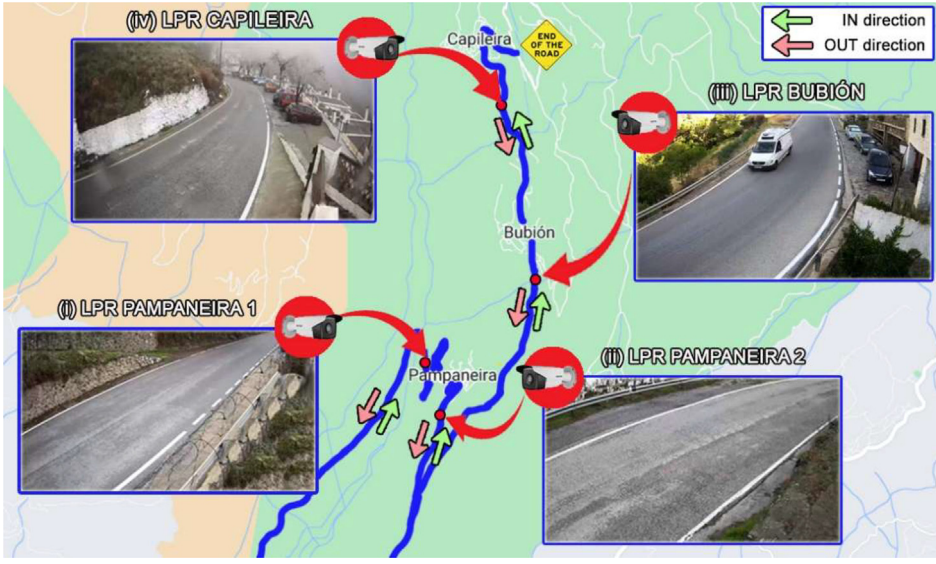

Fig. 2. Setup of the 4 LPR that obtain the data from the license plates of the vehicles.

the western part of the Alpujarra, (ii) entrance to Pampaneira from the eastern part of the Alpujarra, (iii) entrance to Bubión via a single road, and (iv) entrance to Capileira via a single road. By leveraging the road layout, we could efficiently monitor vehicle entrance/exit to the three villages using just four LPRs, which reduced system costs and complexity. The data collected by the cameras were stored in a cloud platform. This data comprises the variables: num_plate_ID, camera_ID, date, direction. Additionally we add the calculated variables: entry_cam, exit_cam, entry_date, exit_date, entry_time, exit_time, visit_time, route, distance, nights, visits_dif_weeks, visits_dif_months, total_entries, avg_visit,m std_visit, avg_nights and std_nights.

In the realm of IoT, sensor data production can sometimes be inaccurate, leading to missing records. In our case, we introduced two data cleaning steps for our LPR cameras dataset (see Fig. 3).

First, we aimed to reduce errors in incomplete or incorrectly detected license plates by the LPRs. Approximately 2% of the total records had missing license plate values. For instance, if one record had the correct license plate, “1111 ZZZ,” and another record showed “1#11 ZZZ,” missing the second digit, we inferred that both records belonged to the same plate, assigning the correct value, “1111 ZZZ,” to both records. We assigned the same plate number to records where at least four characters out of seven matched in the same position.

Second, we reduced the percentage of vehicles that went undetected by any LPR device. These errors occurred when a camera failed to detect a passing vehicle (up to 10% of cases, according to camera specifications). In our setup, if a vehicle moved from camera 1 to camera 3, and camera 2, situated between them, didn’t detect the car, we inferred that the car had passed through camera 2, and it calculated the vehicle’s stay time based on the newly recorded values.

We obtained the remaining variables from various sources, as shown in Fig. 3.

4.2. Vehicle information data

The Spanish Directorate-General for Traffic (DGT) provided additional vehicle data [5]. This data comprises the variables: num_plate_ID, country, and postcode. The dataset links each vehicle to a fiscal address (postcode) utilized for road tax payment, thereby allowing us to determine

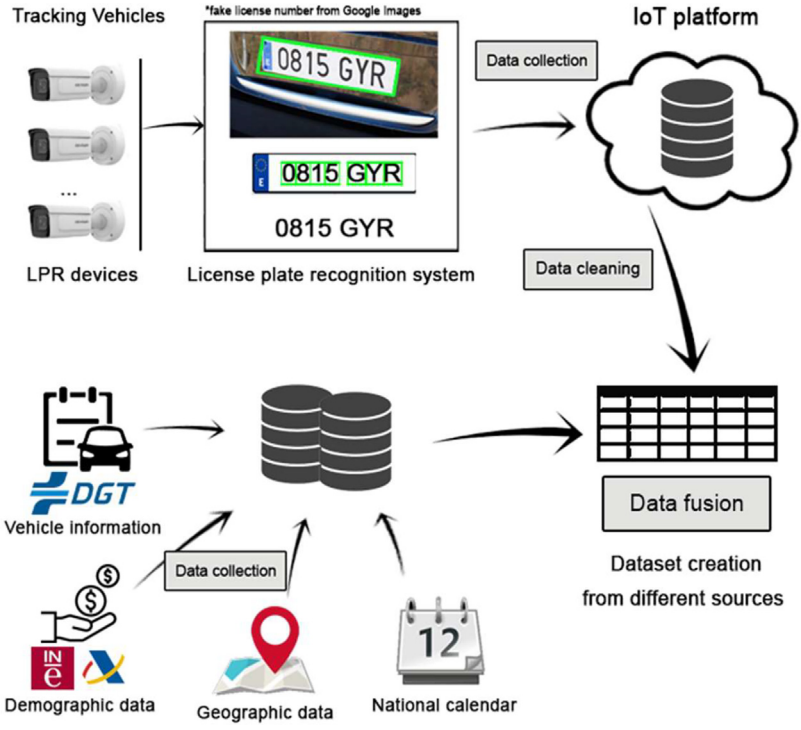

Fig. 3. Methodology for collecting, cleaning and merging data from various sources.

the vehicle's provenance. To the best of our knowledge, this information is not available in other vehicular datasets.

4.3. Demographic and economic data

We took the demographic and economic data from the National Statistics Institute (INE) website [6]. This dataset is available for regions with more than 1,000 inhabitants and updated until 2020. The variables are: postcode, population size, average gross income, and average disposable income per person. The link of this information with vehicular data is not available in other vehicular datasets.

4.4. National calendar data

We used a holiday library [7] to obtain holiday data. This library allowed us to create custom calendars for local holidays, long weekends, and bank holidays, tailored to Spain. We defined holiday periods based on the country's three most important national holidays: Summer, Christmas, and Holy Week [8]. This data comprises the variables: num_plate_ID, country, autonomous_community, postcode, and province. This data comprises the variables: date, num_holiday, num_workday, num_high_season, num_low_season, entry_in_holiday, entry_in_high_season, avg_holiday, std_holiday, avg_workday, std_workday, avg_high_season, std_high_season, avg_low_season, and std_low_season.

4.5. Geographic data

We collected geographic information about the vehicles' origins using postcodes and two libraries: pgeocode [9] and geopy [10]. These libraries allowed us to query GPS coordinates, region names, municipality names, and validate the vehicle's location at different levels, such as municipality, county, or suburb. Additionally, data from the INE helped verify the province and autonomous community codes related to the postcode. This data was essential for calculating the distance from the vehicle's origin and identifying the autonomous community and province of provenance. The variables derived from this geographic data are: postcode, km_to_dest, autonomous_community, and province. The link of this information with vehicular data is not available in other vehicular datasets.

In the final stage, we combined all the databases, cross-referencing information from license plates and postcodes. This resulted in the construction of the VISITS_SMART_POQUEIRA file. From this file, we aggregated information by num_plate_ID and calculated averages for specific variables, leading to the creation of the VEHICLES_SMART_POQUEIRA file.

Limitations

The proposed dataset has certain limitations. We lack information on the provenance and distance in kilometers for foreign vehicles (country = 'Other'). The Demographic and Economic data from the INE provides only overall information on foreigners in Spain, and we applied this single value to all foreign vehicles, which may not reflect reality. However, this discrepancy affects a relatively small percentage of the vehicles (7.96% of the total) and can be disregarded if desired. In addition, some vehicles in the RAW_SMART_POQUEIRA file do not appear in the other two files due to camera errors. This error, common in sensor data, can lead to occasional failures in detecting vehicles entering or leaving the area, resulting in an incomplete calculation of visits (entry and exit) in the VISITS_SMART_POQUEIRA and VEHICLES_SMART_POQUEIRA files. Lastly, the sample size of the dataset is limited to a 9-month period, although we intend to periodically update the repository in Zenodo with new data.

Ethics Statement

Prior to camera installation and license plate detection, necessary agreements were established with municipal authorities and the camera installation company to ensure compliance with national laws. The License Plate Recognition (LPR) cameras transmitted license plate data to a secure server at our provider's facilities. We exclusively utilized anonymized data, replacing each license plate with a unique integer value. All other datasets were publicly available, with the exception of the one obtained from the DGT. DGT shared sensitive data with license plates and their associated postal codes solely for research purposes. The postal code information has been removed from the published dataset, leaving only general information such as distance in kilometers to the origin area and province/autonomous community. These variables are non-identifying and safeguard individual privacy while enabling meaningful analysis.

Data Availability

Federation of Vehicular Data in Smart Villages with Socioeconomic Information (Original data) (Zenodo).

CRediT Author Statement

Daniel Bolaños-Martinez: Methodology, Validation, Investigation, Resources, Software, Writing – review & editing; **Maria Bermudez-Edo:** Conceptualization, Investigation, Resources, Writing – review & editing, Supervision; **Jose Luis Garrido:** Conceptualization, Investigation, Resources, Writing – review & editing, Supervision; **Blanca L. Delgado-Márquez:** Investigation, Writing – review & editing, Project administration.

Acknowledgements

This publication is part of the R&D&i Project Ref. PID2019-109644RB-I00 funded by Ministerio de Ciencia e Innovación/ Agencia Estatal de Investigación/ 10.13039/501100011033, and the R&D&i Project Ref. C-SEJ-128-UGR23 funded by Junta de Andalucía and "ERDF A way of making Europe", and also by the project "Thematic Center on Mountain Ecosystem & Remote sensing, Deep learning-AI e-Services University of Granada-Sierra Nevada" (LifeWatch-2019-10-UGR-01), which has been co-funded by the Ministry of Science and Innovation through the FEDER funds from the Spanish Pluriregional Operational Program 2014-2020 (POPE), LifeWatch-ERIC action line. The project has also been co-financed by the Provincial Council of Granada.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Daniel Bolaños-Martinez, Maria Bermudez-Edo, Jose Luis Garrido, Clustering pipeline for vehicle behavior in smart villages, Inf. Fusion (2023) 102164 ISSN 1566-2535, doi:10.1016/j.inffus.2023.102164.
[2] European CommissionEU Guide on Data for Tourism Destinations July, Smart Tourism Destinations, 2022 https://smarttourismdestinations.eu/wp-content/uploads/2022/07/Smart-Tourism-Destinations_EU-guide_v1_EN.pdf.
[3] D. Bolaños-Martinez, M. Bermudez-Edo, J.L. Garrido, B.L. Delgado Márquez, Federation of vehicular data in smart villages with socioeconomic information (1.0) [Data set], Zenodo (2023), doi:10.5281/zenodo.10245475.
[4] Category:Alpujarra granadina. Wikimedia.org. https://commons.wikimedia.org/wiki/Category:Alpujarra_Granadina, (n.d.), (accessed 10 November 2023)
[5] Informe de un vehículo. DGT. Gob.es. https://sede.dgt.gob.es/en/index.html, (n.d.), (accessed 31 October 2022).
[6] Instituto Nacional de Estadística. INE. https://www.ine.es/en/index.htm, (n.d.), (accessed 31 October 2022).
[7] Python-holidays — holidays documentation. Readthedocs.Io. https://python-holidays.readthedocs.io/en/latest/, (n.d.), (accessed 31 October 2022).
[8] Las vacaciones de los españoles - Datos estadísticos. Statista. https://es.statista.com/temas/3585/vacaciones-en-espana/, 2023, (accessed 20 July 2023).
[9] Pgeocode — pgeocode 0.3.0 documentation. Readthedocs.Io. https://pgeocode.readthedocs.io/en/latest/, (n.d.), (accessed 31 October 2022).
[10] Welcome to GeoPy's documentation! — GeoPy 2.4.0 documentation. Readthedocs.Io. https://geopy.readthedocs.io/en/latest/, (n.d.), (accessed 31 October 2022).