Visualització de dades

Part II: projecte de visualització

Marc Luengo Orús
Juny 2024

Índex

1. [10%] Justifiqueu breument la vostra selecció, sigui per motius personals o professionals.

Per a la realització d'aquest projecte que comprendrà la pràctica 1 i pràctica 2 de visualització de dades he escollit aquest conjunt de dades que es pot trobar al Zenodo, repositori d'accés obert de propòsit general desenvolupat sota el programa europeu OpenAIRE i operat per CERN .12. En els següents punts es donaran tots els detalls del conjunt de dades. (Punt 4)

(clicant damunt del conjunt de dades et porta a la pàgina del mateix per veure els detalls)

.

He estat cercant diferents plataformes de dades obertes com les dels governs de diferents països, de la Unió Europea, kaggle entre d'altres, Zenodo era una altra opció, aleshores vaig trobar aquest conjunt de dades anomenat: Wildlife–vehicle collisions (WVC) on interurban roads in Spain (2016-2021). Que vindria a representar les Col·lisions fauna-vehicle a les carreteres interurbanes a Espanya en l'interval de temps especificat, em va cridar l'atenció, ja que buscava un conjunt que afectés a Espanya o si més no a regions properes, que em pogués repercutir o que em despertes un interès personal per voler conèixer més i veure el comportament d'aquestes dades.

Abans d'escollir-lo, es va fer un primer anàlisi ràpid, sense codificar, vaig veure que era un dataset bastant complet i que complia amb els requisits demanats, comptava amb una gran varietat de variables i de diferents tipus, per tant, finalment es va decidir treballar sobre aquest conjunt de dades.

Dit això, es pot dir que l'elecció és purament personal, ja que és un conjunt de dades molt complet, s'ha combinat amb la recopilació d'altres fonts d'informació per la creació del mateix i perquè és un tema que em crida l'atenció i m'agrada; com és la fauna i la conducció.

2.[10%] La rellevància del conjunt de dades en el context. Són dades actuals? Tracten un tema important per algun col·lectiu concret? S'ha tingut en compte la perspectiva de gènere?

Com he comentat en el punt anterior, aquest és l'enllaç per accedir a la descàrrega del dataset emprat per a realitzar el projecte: conjunt de dades.

Col·lisions fauna-vehicle (WVC) a les carreteres interurbanes a Espanya (2016-2021)

Conté 1.000 registres de col·lisions vida salvatge-vehicle (WVC) a carreteres interurbanes d'Espanya entre 2016 i 2021. Per tant, les dades són bastant actuals, no són recents, però estan dintre d'un interval bastant actual, mateixa dècada, fet que va fer que es seguís treballant i usant aquest conjunt de dades pel projecte. No té incidència amb la perspectiva de gènere, ja que no afecta, perquè no està determinat el sexe dels individus que van patir l'accident, seria un punt a considerar per a propers estudis, si es trobés rellevant. Les col·lisions entre la fauna i el vehicle són importants en la gestió de la vida salvatge a causa dels seus impactes socioeconòmics creixents i l'efecte generalitzat en algunes espècies en perill d'extinció.

Em va semblar interessant, el fet que, la recopilació d'aquestes dades engloben sis anys, es una idea ocorrent i funcional realitzar la recopilació de dades i configuració del mateix dataset pels sis anys següents, que anirien del 2022 al 2027, es compten els darrers també, es podrien extreure diferents anàlisis que afectarien a diferents públics objectius: com podrien ser la Direcció General de Trànsit, El Comitè de Flora y Fauna Silvestres , entre d'altres investigadors, les espècies implicades i s'avaluaria la distribució geogràfica i el cost econòmic d'aquesta interacció home-animal a Espanya.

Com també, diferents estudis de millora que ha pogut implantar cada comunitat autònoma amb relació al millorament de les vies interurbanes.

Enfocant-se en les dades que engloba el dataset en qüestió (2016-2021) s'ha analitzat quins factors van afectar les col·lisions entre fauna i vehicle a les diferents carreteres de

l'estat Espanyol. A continuació, en el següent punt, es mostrarà tota la informació i detalladament la tipologia i variables del conjunt de dades.

3. [25%] La complexitat (mesura, variables disponibles, tipus de dades, etc.). Heu de tenir de l'ordre de milers de registres mínim. I ha de tenir un mínim de l'ordre de desenes de variables. Combina dades categòriques i quantitatives? Inclou altres tipus de dades? Evita els conjunts excessivament simples.

En aquest apartat del projecte, s'explicaran tots els detalls del conjunt de dades seleccionat per realitzar el procés d'anàlisi i de visualització de dades.

El projecte es diposita en aquest repositori GitHub personal amb el nom següent: Wildlife-vehicle-collisions-in-Spain.

(*Si cliqueu damunt del nom de repositori GitHub us conduirà a la pàgina on s'està treballant en el projecte i podreu visualitzar tots els passos realitzats fins a la data*.)

Com es pot veure en el repositori, el projecte es realitza amb Python, utilitzant les diferents llibreries (pandas, sys, matplotlib, seaborn, etc.) que ens ofereix el llenguatge per al tractament complet i general de les dades.

El repositori conté el codi principal en l'arxiu main.py, el read.me del projecte, el conjunt de dades en qüestió, un arxiu de text anomenat resultat.txt, on es visualitzaran les sortides del terminal, com també diferents imatges de gràfics que es van realitzant en el projecte, dades d'interès, com podrien ser: la distribució de les variables numèriques amb histogrames o gràfics de barres apilades per les variables categòriques.

Títol: Wildlife–vehicle collisions (WVC) on interurban roads in Spain (2016-2021)

A continuació es comentaran les diferents sortides, obtingudes del dataset:

```
dtypes: float64(13), int64(18), object(19)

Nombre de registres: 1000

Nombre de variables: 50
```

S'observa que el conjunt té 1000 registres amb un total de 50 variables, s'ha preparat un diccionari de les mateixes amb la descripció de cada una detallada.

(clicant en la paraula diccionari us condueix al mateix detallat de cada variable, per evitar mostrar-ho per duplicat).

Es podria pensar que el conjunt és escàs en termes de registres, ja que n'hi ha 1000, però he considerat continuar endavant, perquè contenia molta informació detallada d'aquests registres amb 50 variables, per tant, és un conjunt de dades molt complet, per tots els registres amb àmplia informació.

Anotar que el conjunt en si, és una mostra de l'original, pel fet que si vols requerir el conjunt original al complet, t'has de posar en contacte amb la propietària i autora d'aquest conjunt de dades penjat al Zenodo, que més endavant anomenarem, com més detalls de drets de propietat, entre d'altres.
Informació variables: (50)

Nombre de variables numèriques: 31 (float64(13), int64(18))
Nom de les variables numèriques: ['id_num', 'ind_accda', 'ind_acciv', 'total_mu30df', 'total_hg30df', 'total_hl30df', 'mes_1f', 'anyo', 'ccaa_1f', 'provincia_1f', 'cod_municipio', 'km', 'sentido_1f', 'tipo_via_3f', 'titularidad_via_2f', 'tipo_animal_1f', 'tipo_animal_2f', 'longitud', 'latitud', 'dia_semana', 'luna', 'prec', 'tmin', 'tmax', 'sol', 'altitud', 'pendiente', 'taxonkey', 'tmed', 'imd_total', 'maxspeed']

Nombre de variables categòriques: 19 (object(19))
Nom de variables categòriques: ['nombre_ind_accd', 'nombre_ind_acciv', 'fecha_accidente', 'hora_accidente', 'nombre_mes', 'nombre_ccaa', 'nombre_provincia', 'nombre_municipio', 'carretera', 'nombre_sentido', 'nombre_tipo_via', 'nombre_titularidad_via', 'nombre_tipo_animal_1f', 'nombre_tipo_animal_2f', 'geom', 'nombre_dia_semana', 'parte_dia', 'uso_suelo', 'tipo_dia']

S'ha observat que no hi ha un valor molt alt de valors nulls en les 50 variables, sinó que només en 9 d'elles, es presenta el llistat a continuació amb el seu valor sobre el total dels registres que recordem eren 1000.

```
Variables amb valors Nulls:
nombre_municipio: 703 valors nulls
luna: 391 valors nulls
prec: 30 valors nulls
tmin: 30 valors nulls
tmax: 30 valors nulls
sol: 30 valors nulls
altitud: 13 valors nulls
taxonkey: 176 valors nulls
tmed: 30 valors nulls
```

En 6 d'elles el percentatge de null està per sota del 10%, per tant, no és tan significatiu, en canvi, en variables com el nom del municipi i la lluna sí que representen un valor bastant alt, que fa que la variable perdi valor i consistència a l'hora d'analitzar-la o representar-la envers les altres dades. Variables que o perquè es desconeixien, han ocorregut fora d'un municipi en concret o per l'hora de l'accident que fa que la lluna no entri en acció, són fets que expliquen aquests valors nulls tan alts en aquestes variables.
Tot seguit, en la imatge que veureu a continuació, s'ha volgut mostrar un anàlisi extens de les 31 variables numèriques, mostra els següents resums estadístics:
Comptatge: El nombre d'observacions per a la variable.
Mitjana: La mitjana aritmètica dels valors de la variable.
Desviació estàndard: La mesura de la dispersió dels valors de la variable al voltant de la mitjana.
Mínim: El valor mínim de la variable.
Primer quartil: El valor que divideix a la distribució de la variable en quatre parts iguals, de manera que el 25% de les observacions és menor que aquest valor.
Mediana: El valor que divideix a la distribució de la variable en dues parts iguals, de manera que el 50% de les observacions és menor que aquest valor i el 50% és major.
Tercer quartil: El valor que divideix a la distribució de la variable en quatre parts iguals, de manera que el 75% de les observacions és menor que aquest valor.

Màxim: El valor màxim de la variable.

```
         id_num    ind_accda    ind_acciv  total_mu30df  total_hg30df  total_hl30df       mes_1f         anyo      ccaa_1f  provincia_1f
count  1.000000e+03  1000.000000  1000.000000        1000.0        1000.0   1000.000000  1000.000000  1000.000000  1000.000000   1000.000000
mean   2.018829e+11     0.987000     0.013000           0.0           0.0      0.014000     7.086000  2018.563000     7.945000     26.539000
std    1.706543e+08     0.113331     0.113331           0.0           0.0      0.125777     3.437211     1.703212     4.140908     13.208594
min    2.016030e+11     0.000000     0.000000           0.0           0.0      0.000000     1.000000  2016.000000     1.000000      2.000000
25%    2.017310e+11     1.000000     0.000000           0.0           0.0      0.000000     4.000000  2017.000000     7.000000     15.000000
50%    2.019091e+11     1.000000     0.000000           0.0           0.0      0.000000     7.000000  2019.000000     8.000000     27.000000
75%    2.020311e+11     1.000000     0.000000           0.0           0.0      0.000000    10.000000  2020.000000    12.000000     36.000000
max    2.021502e+11     1.000000     1.000000           0.0           0.0      2.000000    12.000000  2021.000000    17.000000     50.000000
       cod_municipio           km   sentido_1f  tipo_via_3f  titularidad_via_2f  tipo_animal_1f  tipo_animal_2f     longitud      latitud   dia_semana
count    1000.000000  1000.000000  1000.000000  1000.000000         1000.000000     1000.000000     1000.000000  1000.000000  1000.000000  1000.000000
mean     7455.478000    93.091081     1.533000     1.860000            1.807000        9.008000        1.130000    -4.394195    41.315725     4.051000
std     13138.665648   145.593815     0.537773     0.347161            0.681333        5.993652        0.421043     2.422416     1.844537     2.058795
min         0.000000     0.100000     1.000000     1.000000            1.000000        0.000000        0.000000   -13.872740    28.456550     1.000000
25%         0.000000     7.297500     1.000000     2.000000            1.000000        5.000000        1.000000    -6.167665    40.418172     2.000000
50%         0.000000    21.550000     2.000000     2.000000            2.000000        8.000000        1.000000    -4.289750    41.929895     4.000000
75%     13048.250000   100.750000     2.000000     2.000000            2.000000        8.000000        1.000000    -2.708808    42.610008     6.000000
max     50298.000000   999.500000     4.000000     2.000000            5.000000       26.000000        2.000000     0.631860    43.731540     7.000000
             luna        prec        tmin        tmax         sol      altitud    pendiente      taxonkey        tmed      imd_total
count  609.000000  970.000000  970.000000  970.000000  970.000000   987.000000  1000.000000  8.240000e+02  970.000000    1000.000000
mean    50.973727    1.480206    7.781340   19.920825    5.059588   685.725898 -1992.963062  6.093773e+06   13.851082   10600.461000
std     35.944117    5.829136    6.151744    8.042053    4.628571   429.025626  4005.028829  1.765947e+06    6.689326   13181.117527
min      0.000000    0.000000  -11.800000    1.800000    0.000000  -999.000000 -9999.000000  2.433875e+06   -3.200000      56.000000
25%     14.000000    0.000000    3.100000   13.800000    0.000000   426.240000     0.610875  5.220126e+06    8.800000    2225.000000
50%     52.000000    0.000000    7.500000   19.400000    5.000000   733.993000     3.423250  5.220126e+06   13.500000    6689.500000
75%     88.000000    0.400000   12.200000   25.700000    9.275000   944.589000     9.980375  7.705930e+06   18.750000   14911.000000
max    100.000000  135.800000   26.000000   43.800000   14.100000  2341.796000    44.991500  7.705930e+06   34.100000  171452.000000
         maxspeed
count  1000.000000
mean     95.240000
std      18.062909
min      30.000000
25%      80.000000
50%      90.000000
75%     120.000000
max     120.000000
Empty DataFrame
Columns: []
Index: [count, mean, std, min, 25%, 50%, 75%, max]
Nombre de variables numèriques: 31
```

Imatge 1: extreta del fitxer resultat.txt del GitHub del projecte

També s'ha fet un anàlisi exploratòria de les variables categòriques, per veure el seus components i freqüència de cada valor, a continuació es mostren els detalls de cada una:

```
Variable: nombre_ind_accd
Accidente de daños materiales exclusivamente    987
No es un accidente de daños exclusivamente       13
Name: nombre_ind_accd, dtype: int64

Variable: nombre_ind_acciv
No es un accidente con víctimas exclusivamente    987
Accidente con víctimas exclusivamente              13
Name: nombre_ind_acciv, dtype: int64
Variable: fecha_accidente
2021-11-10    4
2017-12-07    4
2017-11-27    4
2017-08-06    4
2016-09-25    4
              ..
Name: fecha_accidente, Length: 790, dtype: int64
Variable: hora_accidente
22:00    24
```

```
22:30    18
22:15    17
20:30    17
23:30    17
         ..
Name: hora_accidente, Length: 266, dtype: int64
Variable: nombre_mes
Noviembre     125
Octubre       116
Diciembre      91
Abril          87
Agosto         85
Mayo           84
Septiembre     78
Junio          74
Julio          69
Febrero        65
Marzo          63
Enero          63
Name: nombre_mes, dtype: int64
Variable: nombre_ccaa
Castilla y León                365
Galicia                        169
Aragón                          96
Castilla-La Mancha              93
Andalucía                       85
Asturias, Principado de         43
Extremadura                     34
Navarra, Comunidad Foral de     32
Comunitat Valenciana            30
Madrid, Comunidad de            16
Cantabria                       15
Rioja, La                       11
Murcia, Región de                9
Canarias                         2
Name: nombre_ccaa, dtype: int64
Variable: nombre_provincia
Burgos               98
León                 66
Lugo                 59
Soria                54
Ourense              48
Huesca               45
Asturias             43
Coruña, A            38
Palencia             35
```

```
Zamora                34
Navarra               32
Teruel                28
Cuenca                25
Ciudad Real           25
Pontevedra            24
Badajoz               23
Zaragoza              23
Segovia               21
Jaén                  21
Salamanca             20
Valladolid            19
Guadalajara           19
Ávila                 18
Madrid                16
Córdoba               15
Cantabria             15
Valencia/València     14
Granada               14
Albacete              13
Huelva                13
Rioja, La             11
Toledo                11
Cáceres               11
Alicante/Alacant      10
Murcia                 9
Sevilla                8
Málaga                 6
Castellón/Castelló     6
Almería                4
Cádiz                  4
Palmas, Las            2
Name: nombre_provincia, dtype: int64
Variable: nombre_municipio
Vilalba      7
Siero        6
Lugo         5
Córdoba      4
Ávila        4
            ..
Name: nombre_municipio, Length: 205, dtype: int64
Variable: carretera
N-234        22
N-122        19
N-525        17
N-120        17
```

```
A-6          12
             ..
Name: carretera, Length: 598, dtype: int64
Variable: nombre_sentido
Descendente     496
Ascendente      486
Ambos            17
Se desconoce      1
Name: nombre_sentido, dtype: int64
Variable: nombre_tipo_via
Resto vías interurbanas    860
Autopista y autovía        140
Name: nombre_tipo_via, dtype: int64
Variable: nombre_titularidad_via
Autonómica                     507
Estatal                        344
Provincial, Cabildo/Consell    148
Otra                             1
Name: nombre_titularidad_via, dtype: int64

Variable: nombre_tipo_animal_1f
Jabalí                     397
Corzo                      306
Canino                     128
Ciervo                      39
Animal no identificado      32
Zorro                       23
Tejón                       12
Otro animal                  9
Vacuno                       9
Felino                       9
Ovino                        8
Equino                       8
Ave                          7
Cabra montés                 6
Liebre                       3
Conejo                       2
Gamo                         1
Lobo                         1
Name: nombre_tipo_animal_1f, dtype: int64
Variable: nombre_tipo_animal_2f
Silvestre                  806
Doméstico                  162
Animal no identificado      32
Name: nombre_tipo_animal_2f, dtype: int64
Variable: geom
```

```
POINT (-3.56132 42.08382)    2
POINT (-7.84093 42.21992)    2
POINT (-3.6393 40.99974)     1
POINT (-5.94009 43.55046)    1
POINT (-1.92943 39.93018)    1
                            ..
Name: geom, Length: 998, dtype: int64
Variable: nombre_dia_semana
Domingo      161
Lunes        156
Sábado       144
Viernes      142
Jueves       139
Martes       136
Miércoles    122
Name: nombre_dia_semana, dtype: int64
Variable: parte_dia
Noche        609
Día          209
Anochecer    117
Amanecer      65
Name: parte_dia, dtype: int64
Variable: uso_suelo
Cultivos                                    703
Monte arbolado                               95
Monte desarbolado                            87
Artificial                                   43
Agua                                         24
Monte arbolado de plantación                 18
Monte arbolado adehesado                     12
Monte con arbolado ralo                       7
Monte con arbolado ralo de dehesa             4
Monte con arbolado disperso                   4
Monte con arbolado ralo de plantación         2
Monte con arbolado disperso de plantación     1
Name: uso_suelo, dtype: int64
Variable: tipo_dia
Diario    695
Finde     305
Name: tipo_dia, dtype: int64
```

Com es pot veure, s'ha estudiat i analitzat al detall el comportament tant, de les variables numèriques com categòriques, més endavant es mostraran gràfics que faran evidenciar el seu comportament, les seves característiques, visualment.

Aleshores, en aquest punt, que ja es van analitzar les dues tipologies de variables per separat, es va decidir, fer combinacions entre les dues variables, categòriques com numèriques per extreure resultats i fets d'estudi, anàlisi:

El nombre total de ferits sense hospitalització per accident per dia de la setmana:

```
nombre_dia_semana
Domingo      3
Jueves       0
Lunes        2
Martes       5
Miércoles    1
Sábado       2
Viernes      1
```

Total de ferits sense hospitalització per accident per tipus d'animal:

```
nombre_tipo_animal_1f
Animal no identificado    1
Ave                       0
Cabra montés              0
Canino                    2
Ciervo                    0
Conejo                    0
Corzo                     2
Equino                    1
Felino                    0
Gamo                      0
Jabalí                    5
Liebre                    0
Lobo                      0
Otro animal               0
Ovino                     1
Tejón                     1
Vacuno                    0
Zorro                     1
```

- Promig de la intensitat mitja diària de tràfic per tipus de carretera:

```
nombre_tipo_via
Autopista y autovía        20356.207143
Resto vías interurbanas     9012.316279
```

- Top 10 de combinacions de part del dia, mes i província amb més accidents:

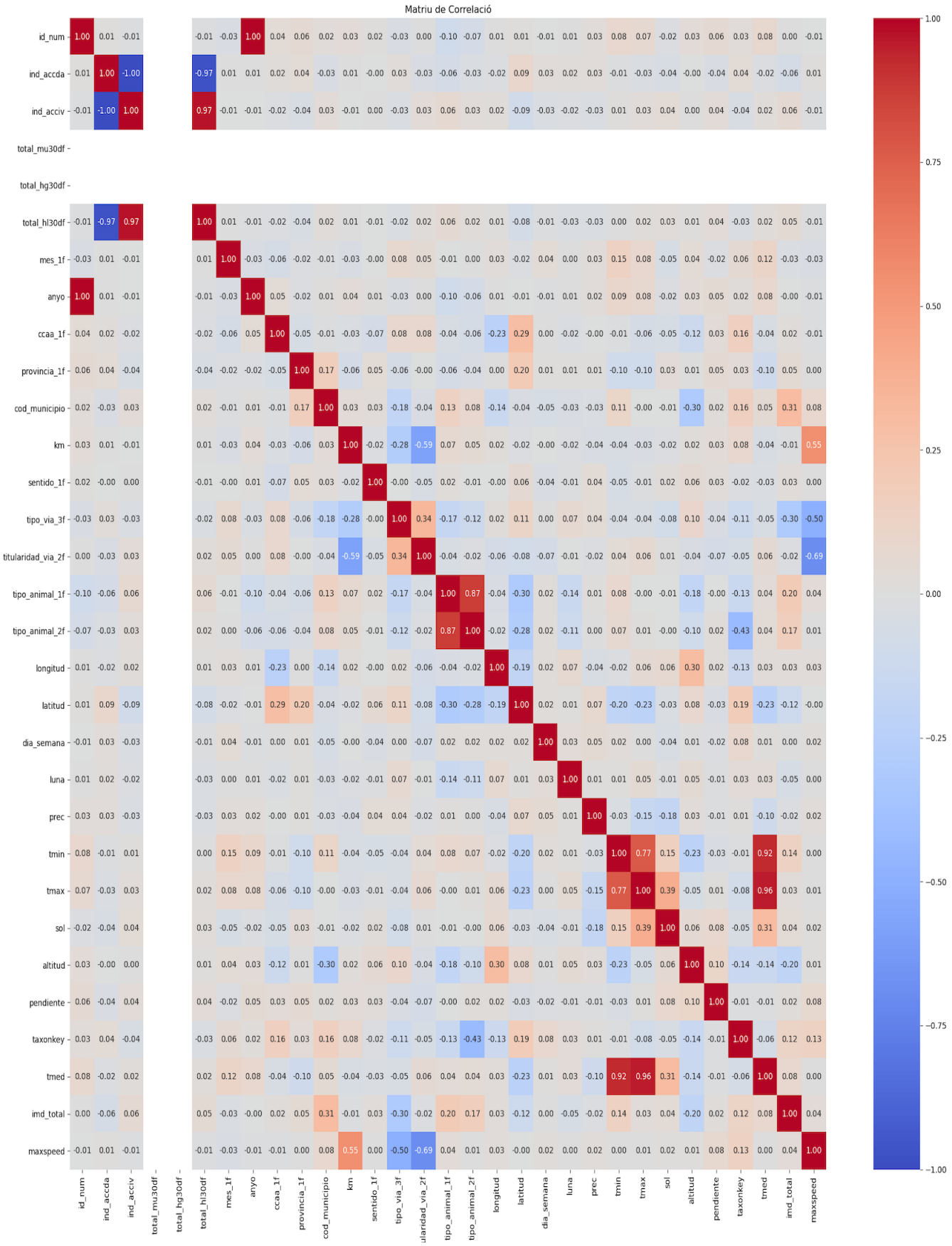
parte_dia	nombre_mes	nombre_provincia	
Noche	Septiembre	Asturias	8
	Diciembre	Ourense	7
	Noviembre	Burgos	7
		Huesca	7
	Agosto	Burgos	7
	Octubre	Asturias	7
	Diciembre	Huesca	6
	Noviembre	Ourense	6
		Lugo	6
	Marzo	Asturias	6

Continuant amb l'anàlisi en els següents punts es veuran les diverses visualitzacions que s'han representat del conjunt de dades.

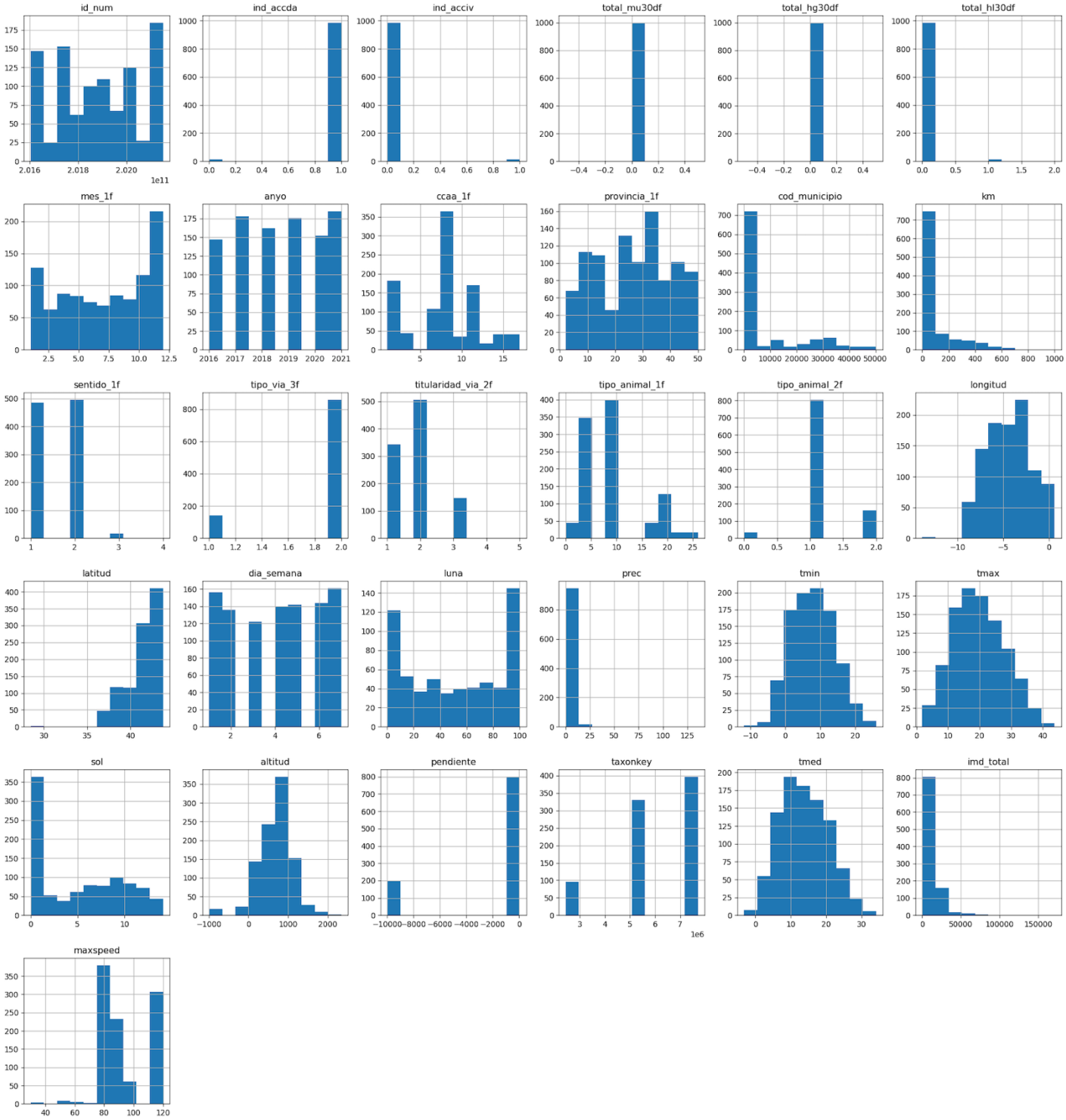
Primer, una anàlisi de la correlació, es veu la Matriu de correlació entre variables numèriques amb un mapa de calor.

S'ha visualitzat la distribució de les variables numèriques amb un gràfic histogrames per cada una d'elles.

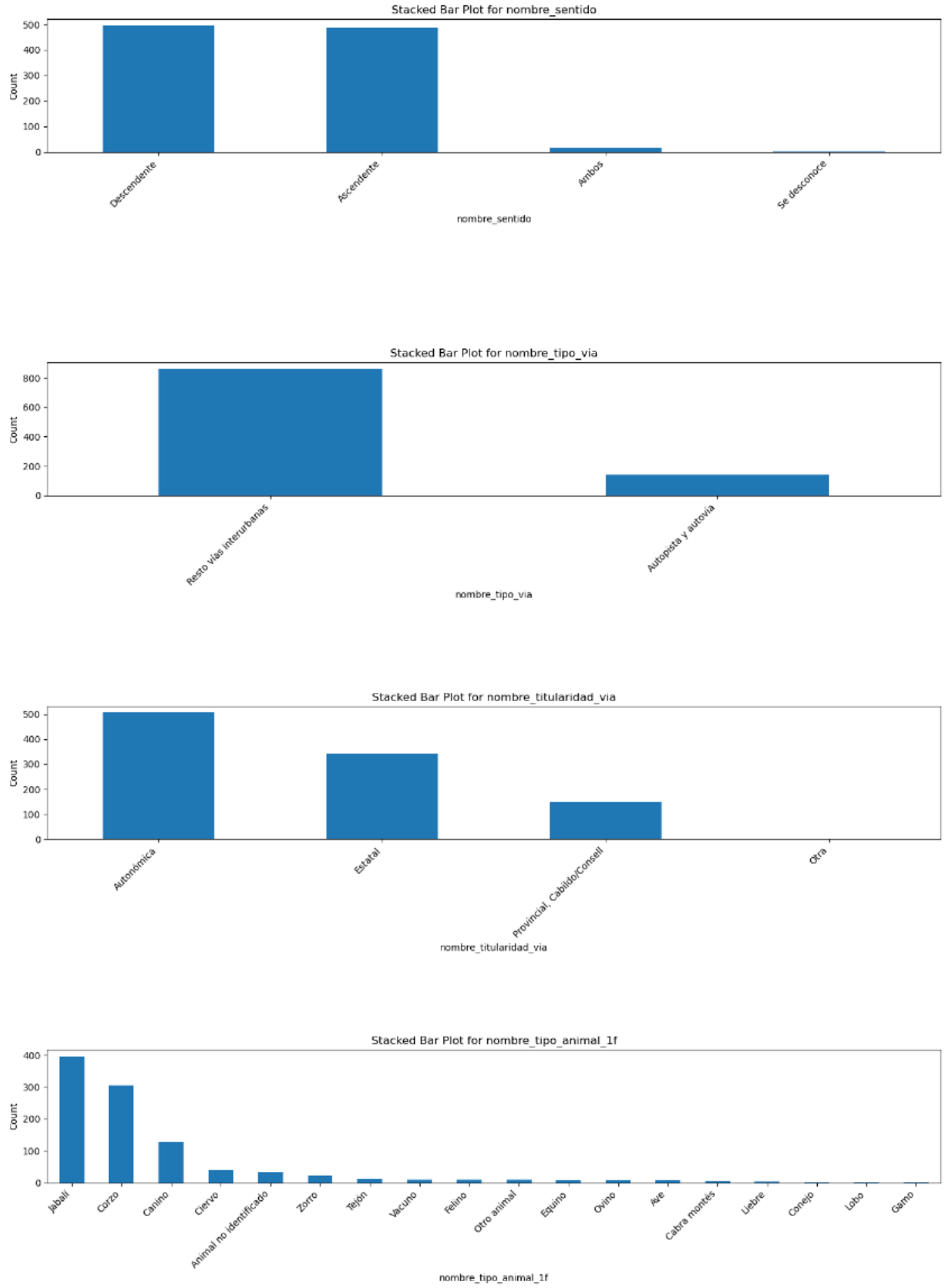
En relació a les variables categòriques s'ha realitzat un gràfic de barres apilades per cada una de elles, a la imatge es mostra només exemple de 4 d'elles, en el projecte, és pot veure en detall.



Imatge 2: Matriu de correlació variables numèriques



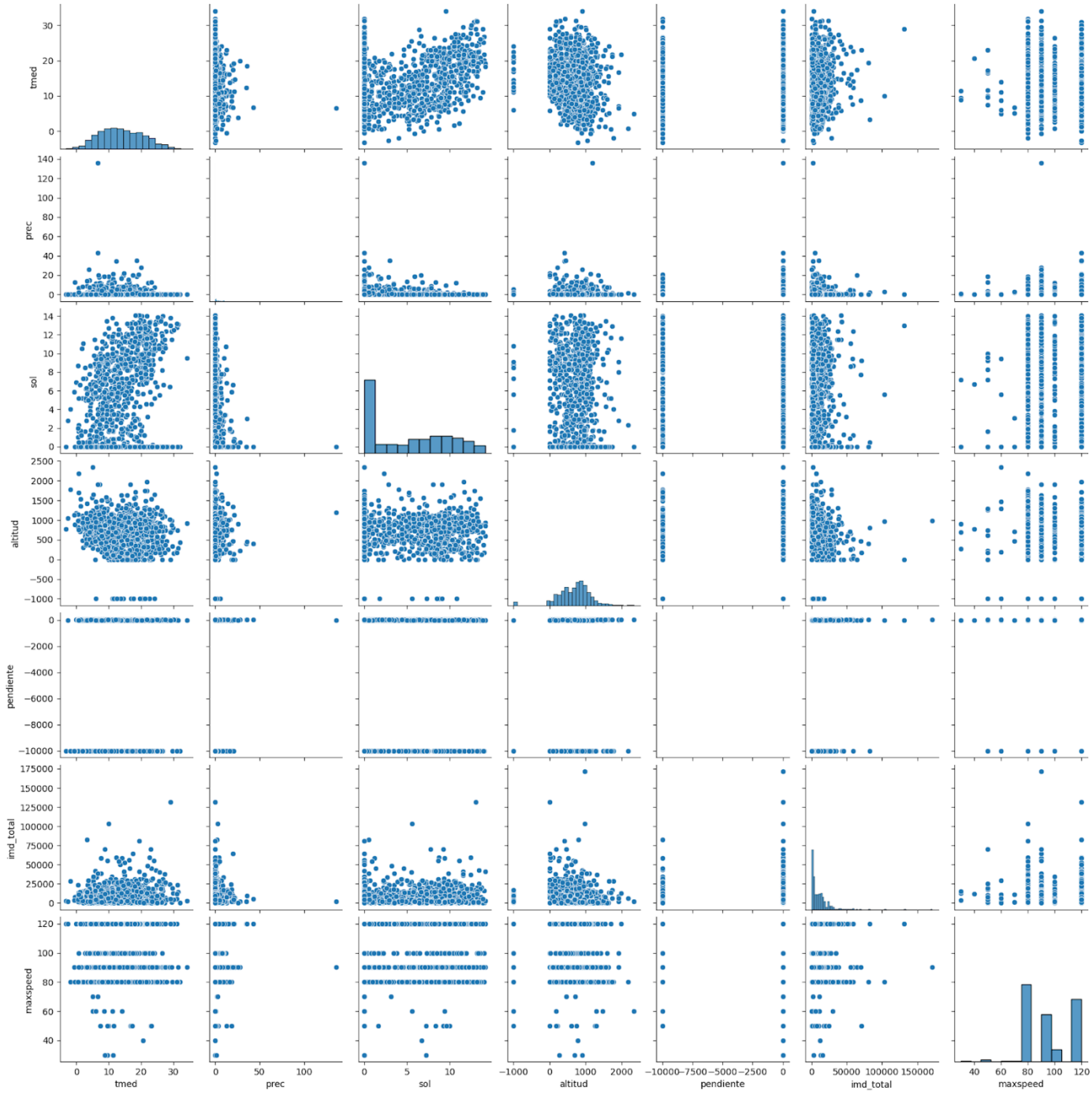
Imatge 3: Distribució de les variables numèriques amb histogrames



Imatge 4: Gràfic de barres apilades de 4 variables categòriques

Cal dir, que es poden fer i es faran més representacions gràfiques per mostrar el detall, comportament i relacions entre les variables del conjunt de dades, a continuació es mostra un gràfic matriu de dispersió.

S'ha seleccionat les variables numèriques representatives en termes de condicions climàtiques que ens ofereix el dataset, és a dir['tmed', 'prec', 'sol', 'altitud', 'pendiente', 'imd_total', 'maxspeed'].



Imatge 5: scatter_matrix de variables numèriques representatives

4. [25%] L'originalitat. Es valora no repetir els conjunts de dades clàssiques o molt treballades Links to an external site. Ni temes ja molt tractats (p. ex. Covid-19, trànsit, criminalitat...) Podeu combinar o millorar el conjunt de dades. En el primer cas, enriquir el conjunt de dades amb altres de diferents per donar un enfocament nou. En el segon cas, generant noves mètriques o indicadors amb les variables existents mitjançant transformacions. Hi ha altres visualitzacions basades en aquest conjunt de dades? És una evolució o una actualització d'un conjunt anterior? Heu enriquit un conjunt de dades ja existent?

Com s'ha pogut observar el conjunt de dades extret del Zenodo: **Wildlife–vehicle collisions (WVC) on interurban roads in Spain (2016-2021)** és propietat de l'usuari: **Alba Gómez Varela** *(clicant en damunt el seu nom es poden veure la resta de datasets que ha pujat a la plataforma Zenodo)* publicat Publicat el 10 de gener de 2023 | Versió 1.0 esta indexat al OpenAIRE amb els drets: Creative Commons Reconeixement 4.0 Internacional (La llicència d'atribució de Creative Commons permet la redistribució i la reutilització d'una obra amb llicència amb la condició que el creador sigui acreditat adequadament.)

- Aquesta es la citació del conjunt de dades: **Gómez Varela, A. (2023). Wildlife–vehicle collisions (WVC) on interurban roads in Spain (2016-2021) (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7523379**

Sobre l'elecció del conjunt he parlat en els punts anterior, però comentat que aquest dataset, és molt complet també, ja que la font de dades de cada registre WVC és la Direcció General de Trànsit (DGT), però el conjunt de dades s'ha millorat gràcies a la integració d'altres fonts: OpenStreetMap (OSM), Global Biodiversity Information Facility (GBIF), l'Institut Geogràfic Nacional d'Espanya (IGN), Agència Estatal de Meteorologia (AEMET).

Per tant, el conjunt escollit per realitzar el projecte ha estat modificat, per l'autora i s'ha enriquit d'altres fonts, clarament clarificar que ha estat fet per l'autora i no s'ha creat el dataset fet per mi, que era una idea si no es trobaven d'altres conjunts de dades amb suficient informació o detall per treballar-hi, aquest, considero que sí que en té.

Un cop es té clara la procedència del conjunt i els components d'aquest, després de fer les primeres anàlisis, el que he pogut millor seria l'eliminació de columnes com "nombre_municipio" i "lluna", ja que, no aporten valor al conjunt al tenir un nombre molt alt de valors nuls. Un cop definits els següents passos i objectius de visualització de dades, també es pot reduir el conjunt amb altres variables que no ens aportaran informació que desencadeni en fets a estudiar, com els punt de longitud, latitud de l'accident, si no es vol seguir per aquest camí, ja que tens altres variables com la carretera que et dirien el punt de l'accident.

S'ha realitzat una recerca extensa per veure si hi havia conjunt de dades semblants, però no se n'ha trobat, ja que es bastant únic al estar construït per diferents punts de recol·lecció de les dades.

Però es mostren diferents webs consultades on hi ha informació rellevant enfocada en les Col·lisions fauna-vehicle (WVC) a Espanya:

- https://www.researchgate.net/publication/273758330_Wildlife-vehicle_collisions_in_Spain
- https://docta.ucm.es/entities/publication/e6b3bac4-9725-47e0-9819-196d05d06bc7
- https://www.dgt.es/export/sites/web-DGT/.galleries/downloads/dgt-en-cifras/publicaciones/Principales_Cifras_Siniestralidad/Main-figures-on-Road-Safety-Data-2021.pdf

Aleshores, per enriquir aquest projecte, personalment, es crearan noves mètriques, indicadors i s'enfocaran en certes variables del conjunt de dades que facin extreure dades i visualitzacions de dades que siguin explicatives i es puguin treure conclusions definitòries.

5. [30%] Les qüestions que respondreu amb la visualització de dades, tenen en compte els punts anteriors? Han estat plantejades en altres visualitzacions o en altres projectes? Són adequades per al conjunt de dades escollit? En aquest punt, elaboreu un diccionari de les variables, el seu significat i si és un fet a estudiar o una dimensió que el mesura, us pot ajudar.

Veient les diferents qüestions que es respondran a continuació a la part II, s'adeqüen al conjunt de dades escollit. S'estudiaran mètriques i anàlisis tant de clima, nombre de víctimes relacionades en l'accident, geografia, velocitat de la via, entre d'altes, com exemples que s'han mostrat en el punt 3. En l'anterior punt es referenciava que no s'havien trobat més conjunt de dades semblants com tampoc visualitzacions enfocades en aquest tòpic que combina la fauna i les col·lisions dels vehicles a les carreteres d'Espanya.

Com s'ha observat, en l'anàlisi extens de les variables categòriques, referent a les Comunitat Autònomes (no estan referenciades totes) es mostran aquestes visualitzacions també:

```
Top 10 províncies amb més accidents i el seu animal predominant:
Província: Burgos, Animal predominant: Corzo
Província: León, Animal predominant: Corzo
Província: Lugo, Animal predominant: Jabalí
Província: Soria, Animal predominant: Corzo
Província: Ourense, Animal predominant: Jabalí
Província: Huesca, Animal predominant: Jabalí
Província: Asturias, Animal predominant: Jabalí
Província: Coruña, A, Animal predominant: Jabalí
Província: Palencia, Animal predominant: Corzo
Província: Zamora, Animal predominant: Jabalí

Top 10 províncies de Galícia amb més accidents i el seu animal
predominant:
Provincia: Lugo, Animal predominant: Jabalí
Provincia: Ourense, Animal predominant: Jabalí
Provincia: Coruña, A, Animal predominant: Jabalí
Provincia: Pontevedra, Animal predominant: Jabalí
```

6. Objectius Part II: projecte de visualització

En aquest apartat, he volgut representar 6 tipologies de gràfics interactius per explicar diferents objectius extrets gràcies al conjunt de dades escollit, a part dels gràfics estàtics analítics que s'han pogut veure en la primera part del projecte.

1. Distribució d'accidents per dia de la setmana

- Objectiu: Analitzar en quins dies de la setmana es produeixen més accidents.
- Visualització: Gràfic de barres interactiu mostrant la quantitat d'accidents per dia de la setmana.

2. Relació entre el tipus de via i la quantitat d'accidents

- Objectiu: Verificar si uns certs tipus de vies tenen més accidents que uns altres.
- Visualització: Gràfic de barres interactiu mostrant la quantitat d'accidents per tipus de via.

3. Distribució d'accidents per províncies i tipus d'animals

- Objectiu: Identificar les províncies amb més accidents i el tipus d'animal predominant.
- Visualització: Mapa interactiu que mostri la ubicació dels accidents i el tipus d'animal involucrat.

4. Anàlisi de correlació entre condicions climàtiques i accidents

- Objectiu: Entendre com les variables climàtiques afecten la quantitat d'accidents.
- Visualització: Gràfic de dispersió i mapa de calor interactiva per a visualitzar les correlacions.

5. Freqüència d'accidents per mes i part del dia

- Objectiu: Determinar si hi ha una estacionalitat en els accidents i en quina part del dia són més freqüents.
- Visualització: Gràfic de línies interactiu que mostri la freqüència d'accidents per mes i part del dia.

En la següent pàgina es mostra el diccionari creat per al conjunt de dades seleccionat on es veuran tots els detalls de cada una, conté descripció, si és un fet a estudiar o dimensió com el total de valors not-null i identificador de tipus de variable.

diccionari

Aquest es el diccionari de les 50 variables que conté el conjunt de dades:

RangeIndex: 1000 entries, 0 to 999
Data columns (total 50 columns):
dtypes: float64(13), int64(18), object(19)

- 0 **id_num**: Identificador únic d'un accident. Aquest és un fet a estudiar, ja que identifica de manera única cada accident en el conjunt de dades.
  (1000 non-null int64)
- 1 **ind_accda**: Variable binària per a danys a la propietat implicats o no (codificat). Aquesta és una dimensió que mesura si hi ha danys a la propietat involucrats en l'accident.
  (1000 non-null int64)
- 2 **nombre_ind_accd**: Declaració per danys a la propietat implicats o no (descodificat). Aquesta és una dimensió que mesura la declaració sobre si hi ha danys a la propietat implicats en l'accident.
  (1000 non-null object)
- 3 **ind_acciv**: Variable binària per a danys personals implicats o no (codificat). Aquesta és una dimensió que mesura si hi ha danys personals implicats en l'accident.
  (1000 non-null int64)
- 4 **nombre_ind_acciv**: Declaració per danys personals implicats o no (descodificat). Aquesta és una dimensió que mesura la declaració sobre si hi ha danys personals implicats en l'accident
  (1000 non-null object)

- 5 **total_mu30df**: Nombre total de morts per l'accident. Aquest és un fet a estudiar, ja que representa una mètrica important en termes d'impacte mortal de l'accident.
  (1000 non-null int64)
- 6 **total_hg30df**: Nombre total de ferits amb hospitalització per l'accident. Aquest és un fet a estudiar, ja que representa una mètrica important en termes de ferits greus que requereixen hospitalització a causa de l'accident.
  (1000 non-null int64)
- 7 **total_hl30df**: Nombre total de ferits sense hospitalització per l'accident. Aquest és un fet a estudiar, ja que representa una mètrica important en termes de ferits que no requereixen hospitalització a causa de l'accident.
  (1000 non-null int64)
- 8 **fecha_accidente**: Data informada de la col·lisió, seguint la norma ISO 8601. Aquesta és una dimensió que mesura la data en què va succeir l'accident.
  (1000 non-null object)
- 9 **hora_accidente**: Hora informada de la col·lisió en notació de 24 hores. Aquesta és una dimensió que mesura l'hora en què va succeir l'accident.
  (1000 non-null object)
- 10 **mes_1f**: Mes com a nombre enter de la data de l'esdeveniment (codificat). Aquesta és una dimensió que mesura el mes en què va succeir l'accident.
  (1000 non-null int64)
- 11 **nombre_mes**: Nom del mes de la data de l'esdeveniment (descodificat). Aquesta és una dimensió que mesura el nom del mes en què va succeir l'accident.
  (1000 non-null object)
- 12 **anyo**: Any de quatre dígits de la data de l'esdeveniment. Aquesta és una dimensió que mesura l'any en què va succeir l'accident.
  (1000 non-null int64)
- 13 **ccaa_1f**: Codi autonòmic de l'INE on es registra l'accident (codificat). Aquesta és una dimensió que mesura la comunitat autònoma on es va registrar l'accident.
  (1000 non-null int64)

- 14 **nombre_ccaa**: Nom de la comunitat autònoma on es registra l'accident (descodificat). Aquesta és una dimensió que mesura el nom de la comunitat autònoma on es va registrar l'accident.
  (1000 non-null object)
- 15 **provincia_1f**: Codi de província de l'INE on es registra l'accident (codificat). Aquesta és una dimensió que mesura la província on es va registrar l'accident.
  (1000 non-null int64)
- 16 **nombre_provincia**: Nom de la província on es registra l'accident (descodificat). Aquesta és una dimensió que mesura el nom de la província on es va registrar l'accident.
  (1000 non-null object)
- 17 **cod_municipio**: Codi del municipi de l'INE on es registra l'accident (codificat). Aquesta és una dimensió que mesura el codi del municipi on es va registrar l'accident.
  (1000 non-null int64)
- 18 **nombre_municipio**: Nom del municipi on es registra l'accident (descodificat). Aquesta és una dimensió que mesura el nom del municipi on es va registrar l'accident.
  (297 non-null object)
- 19 **carretera**: Via atesa pel sistema de numeració de carreteres nacionals d'Espanya on es situa l'accident. Aquesta és una dimensió que mesura la via on va succeir l'accident.
  (1000 non-null object)
- 20 **km**: Punt quilomètric de la via on es troba l'accident. Aquesta és una dimensió que mesura el punt quilomètric de la via on va succeir l'accident.
  (1000 non-null float64)
- 21 **sentido_1f**: Sentit de la circulació del vehicle informat com a nombre sencer quan va succeir l'accident (codificat). Aquesta és una dimensió que mesura el sentit de la circulació del vehicle en el moment de l'accident.
  (1000 non-null int64)

- 22 **nombre_sentido**: Direcció de circulació del vehicle informada quan va succeir l'accident (descodificat). Aquesta és una dimensió que mesura la direcció de circulació del vehicle en el moment de l'accident.
  (1000 non-null object)
- 23 **tipo_via_3f**: Tipus de via com a nombre sencer atenent a la classificació de via del projecte (codificat). Aquesta és una dimensió que mesura el tipus de via on va succeir l'accident.
  (1000 non-null int64)
- 24 **nombre_tipo_via**: Tipus de descripció de la via atenent a la classificació de la via del projecte (descodificat). Aquesta és una dimensió que mesura la descripció del tipus de via on va succeir l'accident.
  (1000 non-null object)
- 25 **titularidad_via_2f**: Tipus de propietat de la via com a nombre sencer (codificat). Aquesta és una dimensió que mesura el tipus de propietat de la via on va succeir l'accident.
  (1000 non-null int64)
- 26 **nombre_titularidad_via**: Descripció del tipus de propietat de la via (descodificat). Aquesta és una dimensió que mesura la descripció del tipus de propietat de la via on va succeir l'accident.
  (1000 non-null object)
- 27 **tipo_animal_1f**: Espècie animal implicada en l'accident com a nombre enter (codificat). Aquesta és una dimensió que mesura l'espècie animal implicada en l'accident.
  (1000 non-null int64)
- 28 **nombre_tipo_animal_1f**: Nom de l'espècie animal implicada en l'accident (descodificat). Aquesta és una dimensió que mesura el nom de l'espècie animal implicada en l'accident.
  (1000 non-null object)
- 29 **tipo_animal_2f**: Tipus d'animal reportat com a nombre sencer (codificat). Aquesta és una dimensió que mesura el tipus d'animal implicat en l'accident.
  (1000 non-null int64)

- 30 **nombre_tipo_animal_2f**: Descripció del tipus de cria d'animal (descodificat). Aquesta és una dimensió que mesura la descripció del tipus de cria d'animal implicada en l'accident.
  (1000 non-null object)
- 31 **longitud**: Longitud de les coordenades del lloc de l'accident en graus decimals. Aquesta és una dimensió que mesura la longitud del lloc de l'accident.
  (1000 non-null float64)
- 32 **latitud**: Latitud de les coordenades del lloc de l'accident en graus decimals. Aquesta és una dimensió que mesura la latitud del lloc de l'accident.
  (1000 non-null float64)
- 33 **geom**: Geometria des de la posició de latitud i longitud. Desenvolupat per a aquest projecte. Aquesta és una dimensió que mesura la geometria del lloc de l'accident.
  (1000 non-null object)
- 34 **dia_semana**: Dia sencer de la setmana en què va succeir l'accident (codificat). Aquesta és una dimensió que mesura el dia de la setmana en què va succeir l'accident.
  (1000 non-null int64)
- 35 **nombre_dia_semana**: Nom del dia en què va succeir l'accident (descodificat). Aquesta és una dimensió que mesura el nom del dia en què va succeir l'accident.
  (1000 non-null object)
- 36 **parte_dia**: Nom de la part del dia en què es registra l'accident incloent el dia, la nit i les transicions. Aquesta és una dimensió que mesura la part del dia en què va succeir l'accident.
  (1000 non-null object)
- 37 **lluna**: Porció de la superfície lunar il·luminada representada com a valor enter de 0 a 100. Aquesta és una dimensió que mesura la porció de la superfície lunar il·luminada en el moment de l'accident.
  (609 non-null float64)

- 38 **prec**: Mesura de pluja diària del dia de l'esdeveniment basada en dies pluviomètrics. Aquesta és una dimensió que mesura la quantitat de pluja en el dia de l'accident.
  (970 non-null float64)
- 39 **tmin**: Temperatura mínima en graus Celsius del dia de l'esdeveniment. Aquesta és una dimensió que mesura la temperatura mínima en el dia de l'accident.
  (970 non-null float64)
- 40 **tmax**: Temperatura màxima en graus Celsius del dia de l'esdeveniment. Aquesta és una dimensió que mesura la temperatura màxima en el dia de l'accident.
  (970 non-null float64)
- 41 **sol**: Hores de sol acumulades del dia de l'esdeveniment. Aquesta és una dimensió que mesura la quantitat d'hores de sol en el dia de l'accident.
  (970 non-null float64)
- 42 **uso_suelo**: Ús principal del sòl de l'àrea de l'accident. Aquesta és una dimensió que descriu l'ús principal del sòl a l'àrea on va succeir l'accident.
  (1000 non-null object)
- 43 **altitud**: Altitud en metres sobre el nivell del mar. Aquesta és una dimensió que mesura l'altitud del lloc de l'accident.
  (987 non-null float64)
- 44 **pendiente**: Valor mitjà del pendent d'una zona d'amortiment de 30 metres al voltant del lloc de l'accident. Aquesta és una dimensió que mesura el pendent del terreny al lloc de l'accident.
  (1000 non-null float64)
- 45 **taxonkey**: Clau de taxó de la columna vertebral de GBIF. Aquesta és una dimensió que identifica el taxó relacionat amb l'accident.
  (824 non-null float64)
- 46 **tipo_dia**: Nom de la categoria del tipus de dia per separar el dia laborable del cap de setmana (descodificat). Aquesta és una dimensió que mesura el tipus de dia en què va succeir l'accident.
  1000 non-null object

- 47 **tmed**: Temperatura mitjana en graus Celsius del dia de l'esdeveniment. Aquesta és una dimensió que mesura la temperatura mitjana en el dia de l'accident. (970 non-null float64)
- 48 **imd_total**: Intensitat mitjana de trànsit diari de l'any de l'accident. Aquesta és una dimensió que mesura la intensitat mitjana de trànsit al lloc de l'accident. (1000 non-null float64)
- 49 **maxspeed**: Velocitat màxima del tram de carretera on es va informar de la col·lisió. Aquesta és una dimensió que mesura la velocitat màxima permesa al tram de carretera on va ocórrer l'accident.
  (1000 non-null int64)