

# Tipologia i cicle de vida de les dades - Pràctica 1

Gerard Cegarra Dueñas

**FORMAR  
TRANS-  
FORMAR**

## Context

En aquest projecte s'han recollit dades sobre els accidents mortals de trànsit que s'han produït a les carreteres catalanes durant l'última dècada. La informació s'ha extret de l'Oficina de Premsa i Comunicació del Servei Català de Trànsit de la Generalitat de Catalunya. Les dades es poden trobar al web oficial de la Generalitat en forma de comunicats de premsa que es publiquen després de cada accident mortal amb la informació del fet.

El motiu pel qual el lloc web proporciona aquesta informació està explicat a la mateixa pàgina:

*«L'Oficina de Premsa i Comunicació del Servei Català de Trànsit és la unitat responsable de gestionar qualsevol petició realitzada pels mitjans de comunicació relacionada amb qualsevol dels aspectes de mobilitat, de seguretat viària i de gestió del trànsit a Catalunya.»*

(Extret de: [http://transit.gencat.cat/ca/el\\_servei/premsa\\_i\\_comunicacio/](http://transit.gencat.cat/ca/el_servei/premsa_i_comunicacio/))

## Títol del dataset

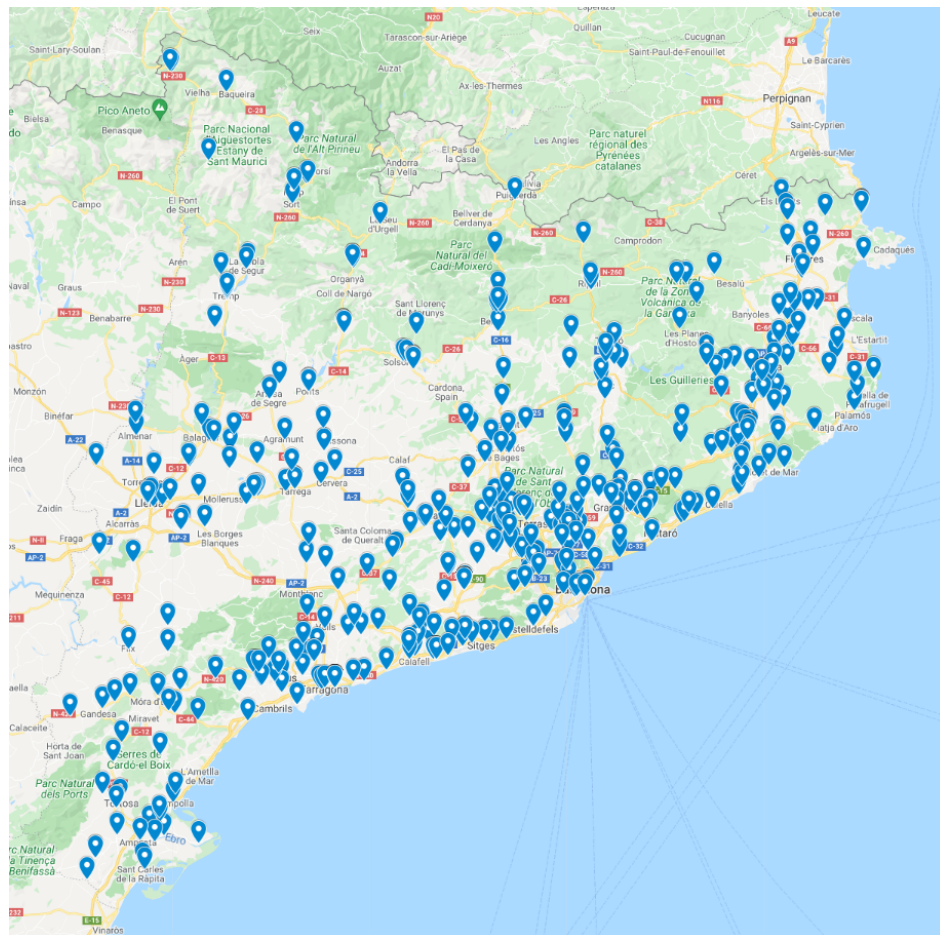
El títol escollit per a aquest dataset és: **Fatal traffic accidents in Catalonia.**

## Descripció del dataset

El dataset obtingut és un recull de 1024 accidents mortals de trànsit ocorreguts a Catalunya. Cada un dels accidents conté informació sobre la localització, la data i l'hora en què s'ha produït, així com el títol, la data i l'hora de publicació del comunicat de premsa. A més, també s'inclou una descripció detallada dels fets que explica com ha sigut l'accident, enumera les persones i els vehicles implicats, i llista els serveis de policia, bombers i emergències mèdiques que han sigut activats per a atendre l'accident. Finalment, per a molts dels accidents posteriors al 12 d'octubre de 2018, s'inclou també el nombre acumulat de víctimes mortals per accidents de trànsit a Catalunya durant l'any de l'accident.

## Representació gràfica

El següent mapa s'ha construït mitjançant un mapa personalitzat de Google Maps utilitzant les dades de localització presents al dataset. La localització inclou, en la majoria dels casos, el nom de la carretera i el punt quilomètric en el qual s'ha produït l'accident. Aquestes dades s'han filtrat, netejat i formatat per a crear un dataset secundari que es pot importar directament a Google Maps, el qual crea de forma automàtica el mapa que es mostra a la visualització. El codi desenvolupat per a crear el dataset secundari de localitzacions, així com el mateix dataset, es poden trobar a la carpeta "map" del codi del projecte.



## Consideracions

Els casos en què la localització no està present, o no segueix el format de nom de carretera i punt quilomètric, s'han descartat de la visualització. També és possible que

Google Maps hagi errat alguna de les localitzacions, ja que la combinació de nom de carretera i punt quilomètric pot donar resultats erronis, com per exemple, en el cas de carreteres nacionals, on es pot confondre el punt quilomètric amb el número d'un bloc d'edificis que es troba a la mateixa carretera.

## Contingut


El dataset obtingut en aquest projecte conté informació de 1024 accidents mortals de trànsit ocorreguts entre el 13 de juny de 2014 i el 30 d'octubre de 2021. Per cada accident, es recullen els següents camps d'informació:

- **titol\_accident:** títol del comunicat de premsa publicat en relació a l'accident
- **data\_publicacio:** data de publicació del comunicat de premsa publicat en relació a l'accident
- **hora\_publicacio:** hora de publicació del comunicat de premsa publicat en relació a l'accident
- **total\_victimes:** total acumulat de víctimes en el mateix any de l'accident
- **data\_accident:** data en què s'ha produït l'accident
- **hora\_avis:** hora en què s'ha rebut l'avis d'accident
- **via:** Carretera i punt quilomètric en què s'ha produït l'accident
- **descripcio:** descripció dels fets inclosa en el comunicat de premsa publicat en relació a l'accident
- **enllac:** enllaç al comunicat de premsa publicat al web de la Generalitat en relació a l'accident

## Com s'ha recollit

Els comunicats de premsa dels quals s'han extret les dades es troben organitzats al web en tres nivells, els quals cadascun correspon a un tipus de pàgina diferent. El primer nivell és un índex anual amb un llistat de tots els anys dels quals es disposa de dades. Per cada any, es troba un segon índex amb paginació, on cada pàgina conté una taula amb part dels accidents de l'any. Finalment, per cada accident, es troba una pàgina amb tota la informació de l'accident presentada de forma semiestructurada.

De la primera pàgina, la més senzilla, s'extreuen els enllaços corresponents a cada índex anual. Per a aquesta extracció, només cal trobar tots els elements de tipus enllaç de la secció de la pàgina on està l'índex.


[Contacte](#) | [CA](#) | [Q](#)

Servei Català de Trànsit

[Inici](#) | [El Servei](#) | [Informació viària](#) | [Gestions](#) | [Seguretat viària](#) | [Formació i educació](#)

[Inici](#) > [El Servei](#) > [Premsa i comunicació](#) > [Comunicats d'accidents mortals](#)

Comunicats d'accidents mortals

[Twitter](#) [Facebook](#) [Telegram](#) [WhatsApp](#) [YouTube](#)

**El conductor d'una motocicleta ha mort en un accident a la C-26 a Algerri (Noguera)**  
02/11/2021

**El conductor d'un turisme ha mort en un accident a la C-153 a la Vall d'en Bas (Garrotxa)**  
01/11/2021

**Un motorista ha mort en un accident en un camí paral·lel a l'AP-7 a Ulldecona (Montsià)**  
01/11/2021

**Un motorista ha mort en un accident a la C-65 a Llagostera (Gironès)**  
30/10/2021

**Una ciclista ha mort en un accident a la GIV-6703 a Quart (Gironès)**  
30/10/2021

Accidents mortals 2021	Accidents mortals 2020	Accidents mortals 2019
Accidents mortals 2018	Accidents mortals 2017	Accidents mortals 2016
Accidents mortals 2015	Accidents mortals 2014	Accidents mortals 2013
Accidents mortals 2012	Accidents mortals 2011	

Un cop dins de l'índex anual, s'extreuen els enllaços corresponents a cadascun dels accidents obtenint tots els elements de tipus enllaç dins la taula d'accidents. Un cop trobats tots els enllaços visibles, es passa a la següent pàgina. Es repeteix aquest procediment fins que el botó de passar a la següent pàgina estigui desactivat, la qual cosa indica que s'ha arribat al final de la paginació.

Accidents mortals 2021

[Twitter](#) [Facebook](#) [Telegram](#) [WhatsApp](#) [YouTube](#)

<b>El conductor d'una motocicleta ha mort en un accident a la C-26 a Algerri (Noguera)</b> 02/11/2021	<b>El conductor d'un turisme ha mort en un accident a la C-153 a la Vall d'en Bas (Garrotxa)</b> 01/11/2021
<b>Un motorista ha mort en un accident en un camí paral·lel a l'AP-7 a Ulldecona (Montsià)</b> 01/11/2021	<b>Un motorista ha mort en un accident a la C-65 a Llagostera (Gironès)</b> 30/10/2021
<b>Una ciclista ha mort en un accident a la GIV-6703 a Quart (Gironès)</b> 30/10/2021	<b>Un ciclista ha mort en un accident a l'N-340 a Cambrils (Baix Camp)</b> 25/10/2021
<b>Un motorista ha mort en una col·lisió frontal a la B-112 a Collbató (Baix Llobregat)</b> 24/10/2021	<b>Un motorista ha mort en un accident amb un camió a l'A-2 a Castellbisbal (Vallès Occidental)</b> 18/10/2021
<b>El conductor d'un turisme ha mort en un accident a la C-66 a Torrent (Baix Empordà)</b> 17/10/2021	<b>El conductor d'un totterreny ha mort en un accident a l'A-2 a Bell-lloc d'Urgell (Pla d'Urgell)</b> 17/10/2021
<b>Un motorista ha mort en una sortida de via en una pista forestal a Muntanyola (Osona)</b> 12/10/2021	<b>El conductor d'un turisme ha mort en una sortida de via a la C-59 a Caldes de Montbui (Vallès Oriental)</b> 12/10/2021
<b>La conductora d'un turisme ha mort en una col·lisió frontal a la C-37 a Manlleu (Osona)</b> 10/10/2021	<b>Atropellament mortal d'un vianant a la C-32 a Gavà (Baix Llobregat)</b> 09/10/2021
<b>La conductora d'un turisme ha mort en un accident a l'N-260 a Avinyonet de Puigventós (Alt Empordà)</b> 04/10/2021	

[<<](#) [<](#) anterior [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) següent [>](#) [>>](#)

Total de pàgines: 8

Finalment, de les pàgines que contenen la informació individual de cada accident, s'extreuen totes les dades rellevants. En aquest tipus de pàgina, la informació no està tan

estructurada com en les anteriors. Moltes de les dades a obtenir es troben dins el mateix element HTML, separades per salts de línia i, en alguns dels casos, precedides per etiquetes de text pla. S'ha fet ús d'expressions regulars per a poder fer una separació efectiva de cada dada.

**gencat** Contacte CA Q

Servei Català de Trànsit

Inici El Servei Informació viària Gestions Seguretat viària Formació i educació

Inici > El conductor d'una motocicleta ha mort en un accident a la C-26 a Algerri (Noguera)

## El conductor d'una motocicleta ha mort en un accident a la C-26 a Algerri (Noguera)

02/11/2021 20:13

Amb aquesta víctima, són 116 les persones que han mort en accident de trànsit enguany a la xarxa viària interurbana de Catalunya (dades provisionals)

DIA: Dimarts, 2 de novembre de 2021  
HORA D'AVÍS: 09.46 h  
VIA: C-26, al punt quilomètric 6 a Algerri (Noguera)

El Servei Català de Trànsit informa que aquest matí s'ha registrat un sinistre viari mortal a la C-26, al punt quilomètric 6, a l'altura d'Algerri (Noguera). Els Mossos d'Esquadra han rebut l'avís a les 09.46 h.

Per causes que encara s'estan investigant, hi ha hagut una col·lisió entre un camió i una motocicleta. Com a conseqüència de l'accident, el conductor de la motocicleta ha mort. Es tracta de L.B.C., de 56 anys i veí de Barcelona.

Arran de la incidència, s'han activat sis patrulles dels Mossos d'Esquadra, dues dotacions dels Bombers de la Generalitat, l'helicòpter medicalitzat i dues ambulàncies del Sistema d'Emergències Mèdiques (SEM).

La C-26 ha estat tallada durant dues hores a l'altura d'Algerri.

El Servei Català de Trànsit recorda que el SIAVT (Servei d'Informació i Atenció a les Víctimes de Trànsit) es troba a disposició dels familiars i afectats per sinistres de trànsit els 365 dies de l'any per oferir-los un servei d'orientació i informació sobre els tràmits després del sinistre, conèixer els recursos existents, rebre informació sobre els drets que els reconeix la legislació vigent i el suport psicològic davant d'un sinistre de trànsit. S'hi pot contactar mitjançant el telèfon gratuït 900 100 268 i a través del web [victimestransit.gencat.cat](http://victimestransit.gencat.cat).

## Agraïments

El propietari de les dades extretes en aquest projecte és la Generalitat de Catalunya. L'organisme encarregat de la publicació i manteniment de les dades és l'Oficina de Premsa i Comunicació del Servei Català de Trànsit de la Generalitat de Catalunya.

Quant a l'àmbit legal i ètic del projecte, les dades recopilades en el dataset han estat extretes a partir de comunicats de premsa, d'una pàgina pública i que pertany a l'administració pública. Per tant, les dades obtingudes són de naturalesa pública i la seva recopilació no comporta cap implicació ètica ni legal.

Hi ha molta recerca al voltant de l'anàlisi de les causes i predicció dels accidents de trànsit. A treballs com a [1] s'utilitzen mètodes de mineria de text per a extreure les causes principals d'accident a partir de les descripcions d'accidents passats i poder treure conclusions que ajudin a la modificació de les carreteres per a fer-les més segures. A [2], s'usen tècniques similars per trobar els principals factors de distracció que provoquen els accidents de trànsit. D'altra banda, a [3], s'usen mètodes d'anàlisi espacial per a relacionar la densitat d'accidents a les carreteres amb la congestió del trànsit.

## Inspiració

La reducció d'accidents mortals a les carreteres és un objectiu important de salut pública per a l'administració, ja que les fatalitats a la carretera són una de les primeres causes de mort en certs sectors de la població. Les dades obtingudes en aquest projecte presenten informació de text i de localització dels accidents de trànsit mortals a les carreteres catalanes, que es poden utilitzar per a dur a terme anàlisis textuais i espacials com els citats a l'apartat anterior, ajudant així a entendre els factors de risc més importants per a facilitar la prevenció de la mortalitat al volant.

## Llicència

He escollit la llicència **CC BY-SA 4.0** per a distribuir el dataset del projecte. Aquesta llicència permet la redistribució del material així com la seva modificació i ús, ja sigui per a fins comercials o no, sempre i quan es faci atribució a l'autor original i els resultats obtinguts es distribueixin sota la mateixa llicència.

Quant a la permissió d'ús amb fins comercials, he escollit aquesta opció ja que qualsevol treball realitzat amb les dades obtingudes tindrà, molt probablement, un objectiu de salut pública, independentment del benefici que un tercer en pugui obtenir. En la mateixa línia, el manteniment de la llicència també és un punt clau en la meva decisió, ja que garanteix la continuïtat del treball futur generat a partir el dataset, el qual contribuirà en l'àmbit de la salut pública, aportant així un benefici a tot el conjunt de la societat.

## Codi

Enllaç al repositori: <https://github.com/gerardcd/TrafficAccidents>

## Dataset

Enllaç del DOI: <https://doi.org/10.5281/zenodo.5652001>

## Bibliografia

[1] Nayak R., Piyatrapoomi N., Weligamage J. (2010) Application of text mining in analysing road crashes for road asset management. In: Kiritsis D., Emmanouilidis C., Koronios A., Mathew J. (eds) Engineering Asset Lifecycle Management. Springer, London.  
[https://doi.org/10.1007/978-0-85729-320-6\\_7](https://doi.org/10.1007/978-0-85729-320-6_7)

[2] Tseng, W., Nguyen, H., Liebowitz, J. and Agresti, W. (2005), "Distractions and motor vehicle accidents: Data mining application on fatality analysis reporting system (FARS) data files", Industrial Management & Data Systems, Vol. 105 No. 9, pp. 1188-1205.  
<https://doi.org/10.1108/02635570510633257>

[3] Wang, C., Quddus, M. A., & Ison, S. G. (2009). Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England. Accident Analysis & Prevention, 41(4), 798-808.  
<https://doi.org/10.1016/j.jtte.2017.05.005>

## Contribucions

Contribucions	Signatura
Investigació prèvia	GCD
Redacció de les respostes	GCD
Desenvolupament del codi	GCD