

GRUPO DE USUARIOS DE R DE MADRID

Fecha: Jueves 7 de julio 2022

Lugar: Facultad de Estudios Estadísticos (aula B-03) + Zoom

Hora: 7:00pm - 8:30pm

Acceso: Libre (plazas limitadas por protocolo Covid)

Cómo llegar: Buses desde Moncloa, 133, 83, 162 (cruzar la A6 por túnel) - F y G (dejan en Historia, cruzar A6 por debajo). Hay parking gratuito pero con número limitado de plazas.

Más detalles en: <http://madrid.r-es.org>

Presentaciones:

Pedro Concejero

"Accidentes de Tráfico. Datos de la DGT y ¿Big Data?"



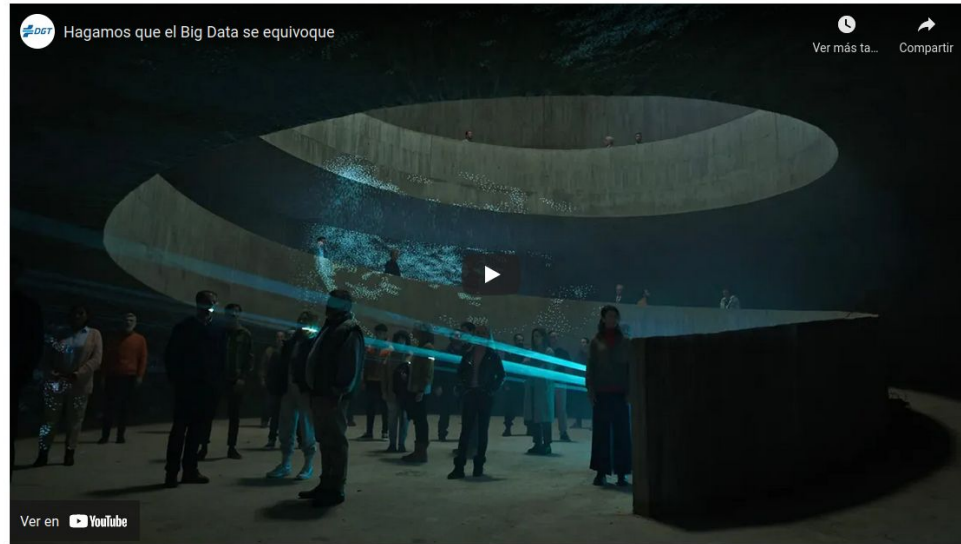
Nuestros patrocinadores:



Las últimas campañas DGT

<https://www.dgt.es/comunicacion/campanas/Los-Treinta-y-Seis/>

La campaña identifica, gracias al Big Data, el perfil de las 36 personas con más probabilidad de morir durante esta Semana Santa en las carreteras y los hace protagonistas de la campaña.



El Big Data, los algoritmos y las estadísticas lo pronostican todo. ¿Por qué no anticipar cuántas personas morirán esta Semana Santa en accidentes de tráfico y así poder evitarlo? Con la ayuda de Nicolás Rodríguez, experto en Analítica e Inteligencia de Datos y Predicción de Riesgos del comportamiento, se ha diseñado la nueva campaña de publicidad "hacer que el Big Data se equivoque".

Las últimas campañas DGT: un poco rimbombante

<https://www.youtube.com/watch?v=Miu6o11Hck8>

Los 36. Cómo los encontramos.

10.822 visualizaciones • 6 abr 2022



DGTes

16.400 suscriptores

Gracias al Big Data, podemos pronosticar el número y el perfil demográfico de las personas con más probabilidades de morir esta Semana Santa en las carreteras españolas. Podemos predecir también el tipo de vehículo, la hora, el lugar y la clase de accidente. Ahora que tenemos todos

MOSTRAR MÁS

¿Qué datos hay en dgt.es?

Microdatos (¡¡¡ojo, de 2020!! pero publicados en 2022)

<https://www.dgt.es/menusecundario/dgt-en-cifras/dgt-en-cifras-resultados/dgt-en-cifras-detalle/?id=00808>

Resto (recientes) son agregados y ... pdf

<https://www.dgt.es/menusecundario/dgt-en-cifras/dgt-en-cifras-resultados/dgt-en-cifras-detalle/?id=00001>

<https://www.dgt.es/menusecundario/dgt-en-cifras/dgt-en-cifras-resultados/?tema=accidentes-de-trafico&pag=1&order=DESC>

Pero ... ay esa calidad del dato

Múltiples problemas:

- No hay día del mes, solo día de la semana -> no podemos hacer serie temporal detallada
- Los missing son códigos numéricos (999, 9999...)
- No hay etiquetas para los factores (para muuuchos factores)

¿Dónde pueden estar los significados de los factores? Hay que ir a un Real Decreto:

Orden INT/2223/2014, de 27 de octubre, por la que se regula la comunicación de la información al Registro Nacional de Víctimas de Accidentes de Tráfico. BOE 29 / 11 / 2014

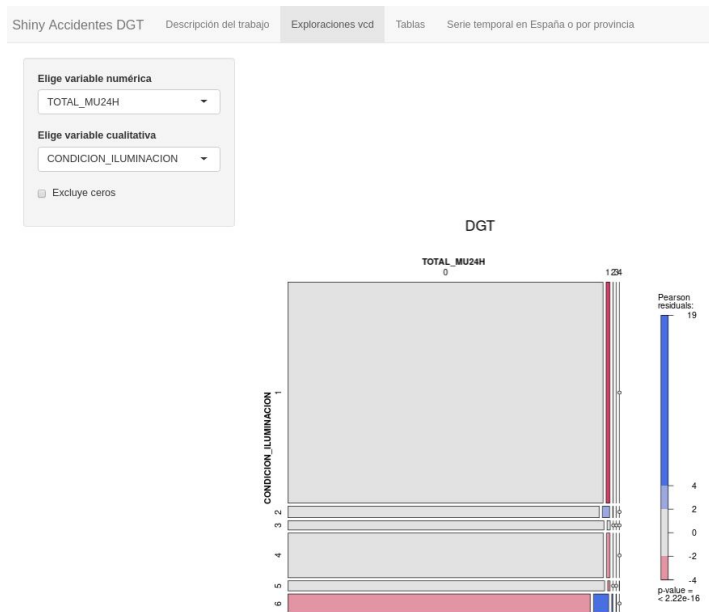
<https://www.boe.es/buscar/doc.php?id=BOE-A-2014-12411>

A ver qué podemos hacer con estos datos

Todo en:

https://github.com/pedroconcejero/accidentes_DGT

(Cómo no) Un shiny para explorar:



(Todavía) Working project

glm : regresión de Poisson

zip: zero-inflated Poisson Regression

<https://stats.oarc.ucla.edu/r/dae/zip/>

GRACIAS!!!!
(Y FELICES VACACIONES)