

MINERÍA DE DATOS

BlaBlaCar Pooling



María Garrido Arcos
Sergio Jiménez Fernández
Ismael Pérez Nieves
David Pinto Camacho

DESCRIPCIÓN DE LOS DATOS ORIGINALES

Hemos decidido usar el dataset relacionado con los [datos de BlaBlaCar](#). Este conjunto de datos contiene información sobre los trayectos ofertados:

Variable	Tipo	Descripción
Dia	Fecha	Fecha del trayecto (entre el 01/12/2017 y 30/11/2019)
Pais	Categórica	País donde se ha dado de alta el trayecto (ES o PT)
Origen	Categórica	Ciudad de origen del trayecto
Destino	Categórica	Ciudad de destino del trayecto
IMP_KM	Numérica	Importe medio por kilómetro
Asientos_ofertados	Numérica	Número de plazas ofertadas (sin incluir conductor)
Asientos_confirmados	Numérica	Número de plazas ocupadas (sin incluir conductor)
Viajes_ofertados	Numérica	Número de viajes ofertados
Viajes_confirmados	Numérica	Número de viajes realizados
Ofertantes	Numérica	Número de conductores que han ofrecido el trayecto
Ofertantes_nuevos	Numérica	Número de conductores que ofrecen un servicio por primera vez

Sería conveniente remarcar que la variable "País" no da la información suficiente para entender si la naturaleza del viaje es nacional o internacional. Por ello hemos decidido realizar un preprocesamiento que consiste en añadir dos columnas nuevas, llamadas "País Origen" y "País Destino", para así catalogar las ciudades de origen y destino como "ES" si son españolas, "PT" si son portuguesas o "FR" si son francesas. De esta forma, si queremos filtrar los viajes según su naturaleza podemos acceder a estas nuevas variables.

La tabla de variables final se quedaría como:

Variable	Tipo	Descripción
Dia	Fecha	Fecha del trayecto (entre el 01/12/2017 y 30/11/2019)
Pais	Categórica	País donde se ha dado de alta el trayecto (ES o PT)
Pais_origen	Categórica	País de origen del trayecto (ES, PT o FR)
Pais_destino	Categórica	País de destino del trayecto (ES, PT o FR)
Origen	Categórica	Ciudad de origen del trayecto
Destino	Categórica	Ciudad de destino del trayecto
IMP_KM	Numérica	Importe medio por kilómetro
Asientos_ofertados	Numérica	Número de plazas ofertadas (sin incluir conductor)
Asientos_confirmados	Numérica	Número de plazas ocupadas (sin incluir conductor)
Viajes_ofertados	Numérica	Número de viajes ofertados
Viajes_confirmados	Numérica	Número de viajes realizados
Ofertantes	Numérica	Número de conductores que han ofrecido el trayecto
Ofertantes_nuevos	Numérica	Número de conductores que ofrecen un servicio por primera vez

TRABAJOS SIMILARES

Nuestra propuesta plantea un enfoque similar al de [este estudio](#) de Segittur, en el que se analizan los diferentes flujos de turismo nacional (entre comunidades autónomas). Para nuestro caso, consideramos interesante un enfoque similar a dicho estudio aplicando los datos de BlaBlaCar.

Otro punto a tratar en nuestro estudio que difiere del anterior informe pero que consideramos relevante, y más actualmente con la crisis climática, sería realizar un análisis relativamente profundo de las emisiones y poder concluir en cómo están ayudando empresas de carpooling como BlablaCar a reducir las emisiones por movilizaciones. Ejemplos de estos análisis pueden ser este estudio de la [misma BlablaCar](#), o incluso [este otro](#) de Pagatelia.

PLANTEAMIENTO DE HIPÓTESIS

Como objetivo de este proceso de extracción de conocimiento a partir de los datos, trataremos de responder las siguientes hipótesis:

- Comprobar si la evolución del uso del servicio de carpooling (BlaBlaCar) ha afectado de alguna forma al uso de transportes interurbanos tradicionales públicos.
- Tratar de analizar el impacto medioambiental que ha supuesto el uso de carpooling comparado con el resto de vehículos para el transporte interurbano.

Responder a estas preguntas consideramos que puede tener cierta relevancia, pues el carpooling es un servicio que aún se encuentra en auge, siendo un tema de actualidad.

ENRIQUECIMIENTO DE DATOS

Para realizar la primera hipótesis que planteamos, necesitamos conjuntos de datos que nos permitan comparar los distintos tipos de transportes y su uso. Para ello, haremos uso de la [Base de datos del observatorio del transporte y la logística en España](#). En este conjunto podemos observar la movilidad en España en distintos medios de transporte. Dentro de este dataset podemos ver las siguientes columnas:

Variable	Tipo	Descripción
Modo	Categórica	Especifica el modo de transporte
Año	Numérica	Año en el que se han tomado los datos
Millones_viajeros_kilometro	Numérica	Millones de viajeros
Textbox7	Numérica	Total de millones de viajeros-kilómetro por modo

Para realizar la segunda hipótesis que planteamos, necesitamos conjuntos de datos que nos permitan comparar los impactos medioambientales de los distintos medios de transporte. Para ello, utilizaremos los datos también del observatorio de transporte y logística de España, pero en este caso, basados en los datos del [Sistema Español de Inventario y Proyecciones de Emisiones a la Atmósfera](#).

Variable	Tipo	Descripción
NombreGas	Categórica	Especifica el gas emitido
TipoTráfico	Categórica	Indica si es de cobertura nacional o internacional
Modo	Categórica	Especifica el modo de transporte
Año	Numérica	Año en el que se han tomado los datos
Valor	Numérica	Especifica las toneladas de gas emitido
Textbox4	Categórica	Indica el subtotal es nacional o internacional de los datos
Textbox6	Numérica	Indica el subtotal del TipoTráfico especificado
Textbox9	Numérica	Suma del subtotal nacional e internacional