



innovation technologies

Aerolíneas



WORKING
DATA

Francisco Pérez Hernández
Francisco Javier Vélchez Torralba
David Jiménez Paredes

Índice

1 Introducción

2 Preprocesamiento

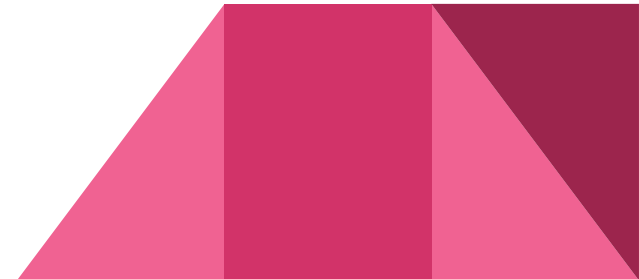
3 Análisis del dataset

3.1 Análisis por compañía

3.2 Análisis de retardo

3.3 Análisis visual

4 Conclusiones



Introducción

Partimos de 3 ficheros:

- Vuelos: Fecha, Aeropuertos de Origen y Destino, Retrasos, etc.
- Operadoras: Código de operadora y Operadora.
- Fichero que Indica la latitud y la longitud de cada uno de los aeropuertos.

Usando las herramientas:

- RStudio
- CartoDB



CARTO

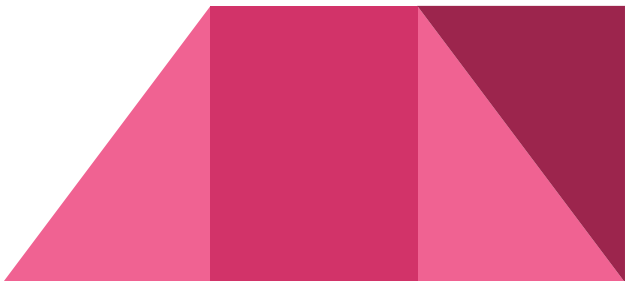


Preprocesamiento

División del DataSet:

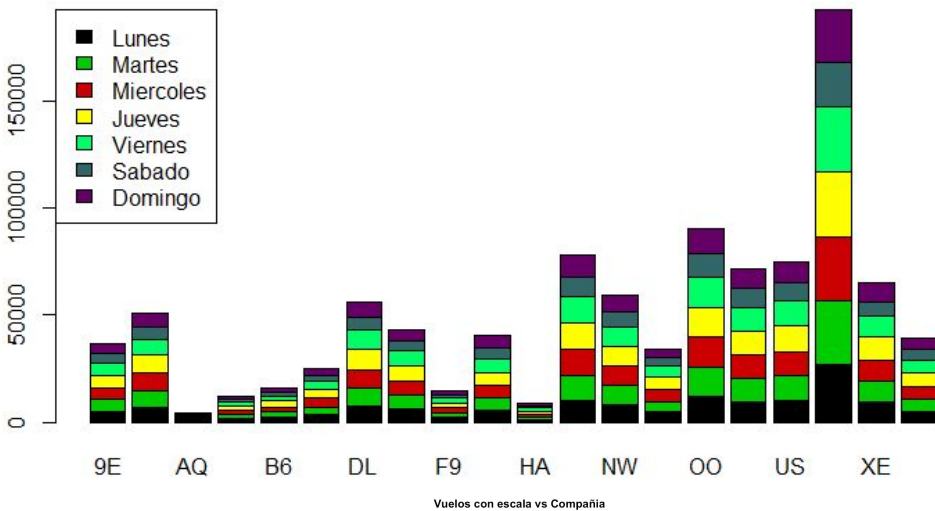
- Variable “Canceled” y “CancellationCode” para DataSet con vuelos cancelados (3,2%).
- Variable “Diverted” para vuelos (0,25%), consecuencia del cambio de aeropuerto:
 - Con escalas, de un punto A -> B -> C (71,4%)
 - Vuelta al origen, A -> B -> A (28,6%)
- Vuelos directos (96,55%).

Imputación de los valores perdidos en las variables CarrierDelay, WeatherDelay, NASDelay, SecurityDelay y LateAircrafDelay creando una nueva variable OtherDelay, basándose en la variable ArrDelay.

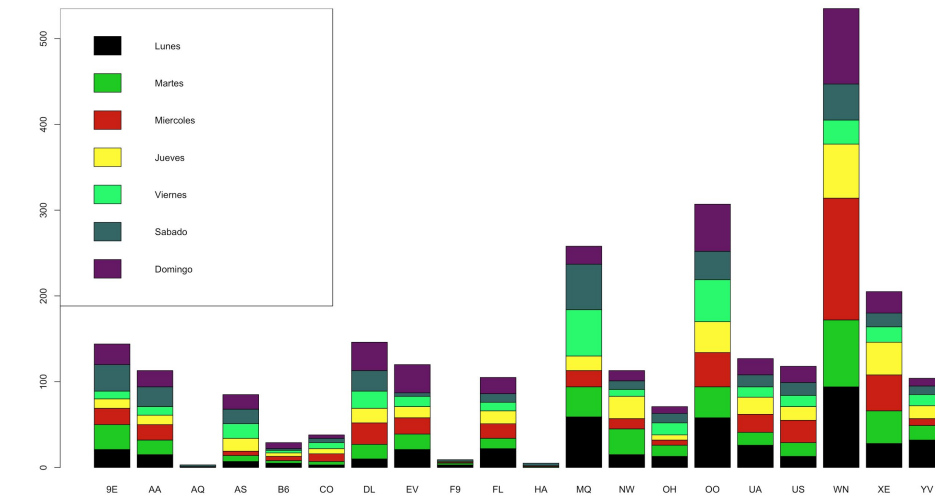


Análisis de los Datos - Análisis por compañía

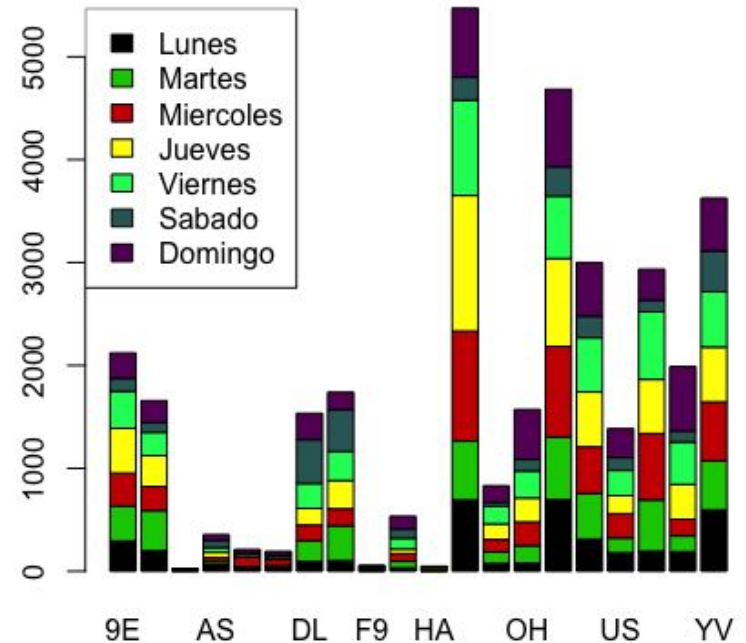
Vuelos Directos vs Compañía



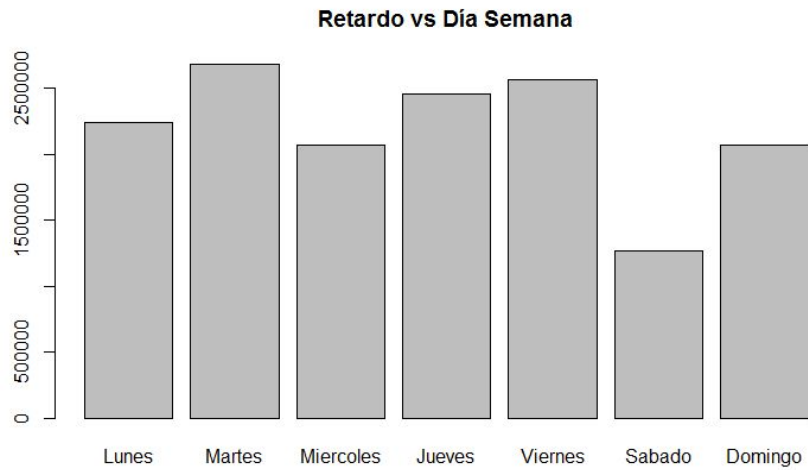
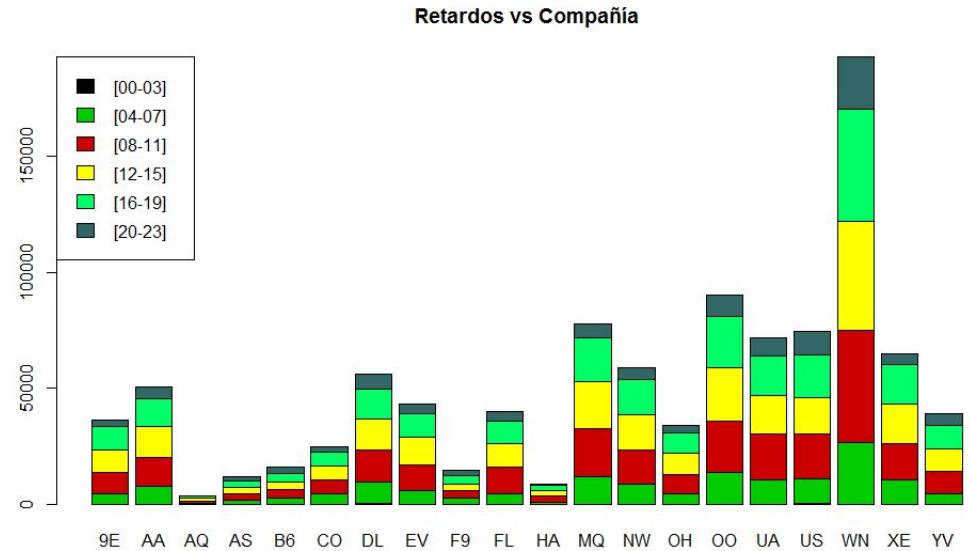
Vuelos con escala vs Compañía



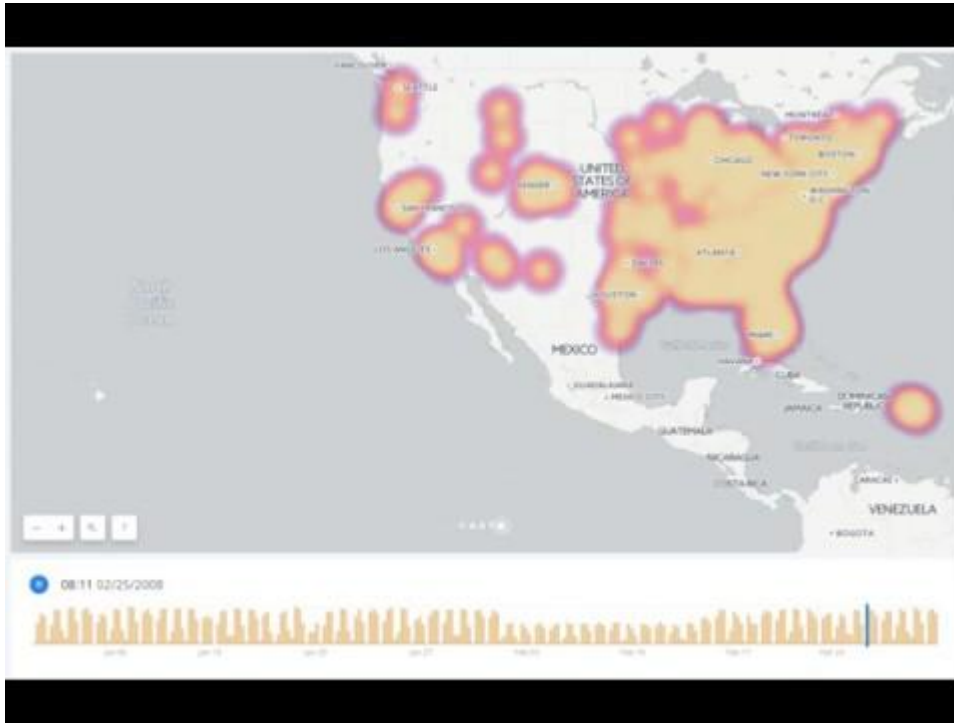
Vuelos Cancelados vs Compañía



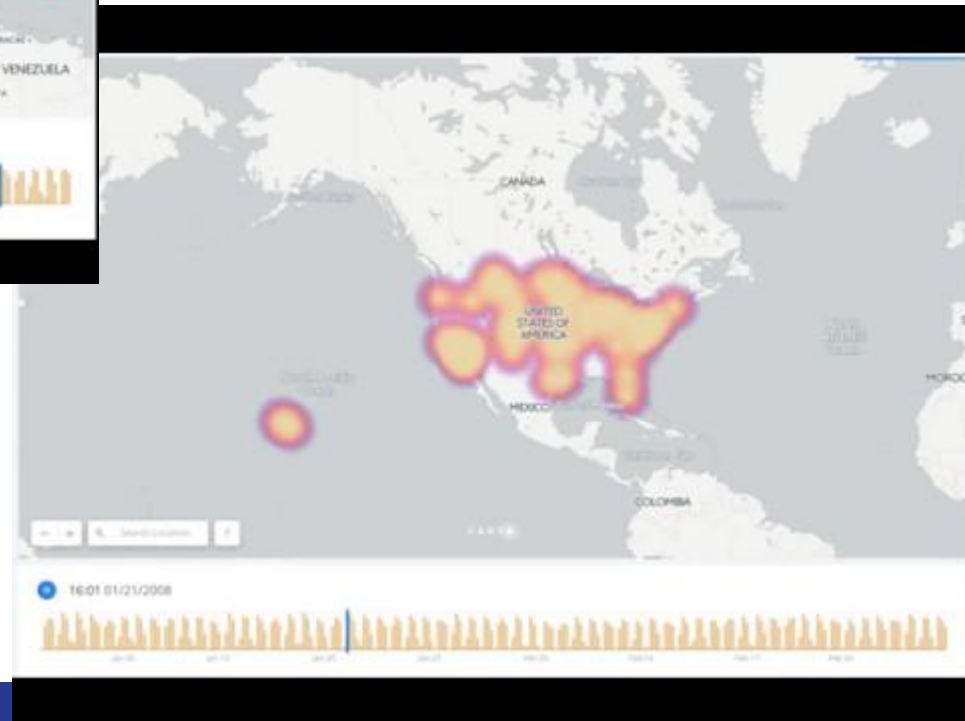
Análisis de los Datos - Análisis de retardo



Análisis de los Datos - Análisis visual



DENVER



Conclusiones

Para el dataset sobre aerolíneas:

- Se ha preprocesado el dataset, imputando valores perdidos y sacado conclusiones sobre las variables.
- Se ha realizado un análisis por compañías viendo conclusiones como que la tercera con más vuelos, es la que más vuelos cancela, al contrario que la compañía que más vuelos realiza, la cual no tiene tantas cancelaciones.
- Se han analizado los retardos, viendo como el mejor día para viajar sería el sábado, y viendo las franjas horarias por las que hay menos retardos.
- Se ha visualizado, mediante un mapa de calor, las zonas donde más tráfico hay, viendo como por ejemplo, hay vuelos fuera de la zona central de Estados Unidos, como se pensaba de inicio.

El repositorio del código es: <https://github.com/PacoPollos/uCodeByadidas2017>

