

Analizando dataset nycflights13:: flights

nycflights13 es un dataset que contiene información sobre vuelos de US. Si es necesario instala el paquete. Crear un nuevo proyecto de R con el nombre TrabajoPersonal. Recordar que debéis enviar tanto el fichero RMD y html comprimidos en un zip. Por favor, seguir el formato siguiente para entregar el trabajo: TrabajoPersonal_NombreApellido.zip.

1. Encuentra todos los vuelos que llegaron más de una hora tarde de lo previsto.
2. Encuentra todos los vuelos que volaron hacia San Francisco (aeropuertos SFO y OAK)
3. Encuentra todos los vuelos operados por United American (UA) o por American Airlines (AA)
4. Encuentra todos los vuelos que salieron los meses de primavera (Abril, Mayo y Junio)
5. Encuentra todos los vuelos que llegaron más de una hora tarde pero salieron con menos de una hora de retraso.
6. Encuentra todos los vuelos que salieron con más de una hora de retraso pero consiguieron llegar con menos de 30 minutos de retraso (el avión aceleró en el aire)
7. Encuentra todos los vuelos que salen entre medianoche y las 7 de la mañana (vuelos nocturnos).
8. ¿Cuántos vuelos tienen un valor desconocido de dep_time?
9. ¿Qué variables del dataset contienen valores desconocidos?
10. Ordena los vuelos de flights para encontrar los vuelos más retrasados en la salida. ¿Qué vuelos fueron los que salieron los primeros antes de lo previsto?
11. Ordena los vuelos de flights para encontrar los vuelos más rápidos. Usa el concepto de rapidez que consideres.
12. ¿Qué vuelos tienen los trayectos más largos?
13. ¿Qué vuelos tienen los trayectos más cortos?
14. El dataset de vuelos tiene dos variables, dep_time y sched_dep_time muy útiles pero difíciles de usar por cómo vienen dadas al no ser variables continuas. Fíjate que cuando pone 559, se refiere a que el vuelo salió a las 5:59... Convierte este dato en otro más útil que represente el número de minutos que pasan desde media noche.
15. Compara los valores de dep_time, sched_dep_time y dep_delay. ¿Cómo deberían relacionarse estos tres números? Compruébalo y haz las correcciones numéricas que necesitas.

16. Investiga si existe algún patrón del número de vuelos que se cancelan cada día.
17. Investiga si la proporción de vuelos cancelados está relacionada con el retraso promedio por día en los vuelos.
18. Investiga si la proporción de vuelos cancelados está relacionada con el retraso promedio por aeropuerto en los vuelos.
19. ¿Qué compañía aérea sufre los peores retrasos?
20. Queremos saber qué hora del día nos conviene volar si queremos evitar los retrasos en la salida.
21. Queremos saber qué día de la semana nos conviene volar si queremos evitar los retrasos en la salida.
22. Para cada destino, calcula el total de minutos de retraso acumulado.
23. Para cada uno de ellos, calcula la proporción del total de retraso para dicho destino.
24. Es hora de aplicar todo lo que hemos aprendido para visualizar mejor los tiempos de salida para vuelos cancelados vs los no cancelados. Recuerda bien qué tipo de dato tenemos en cada caso. ¿Qué deduces acerca de los retrasos según la hora del día a la que está programada el vuelo de salida?
25. Subir la carpeta a github y facilitar la url.
26. Al finalizar el documento agrega el comando `sessionInfo()`