

EADR: PEC

Enunciado

Vamos a trabajar con los ficheros “Estaciones.txt” y “Precios_2017_04_02.txt” (fuente), que contienen información sobre el precio de distintos carburantes el día 2 de Abril de 2017 en estaciones de toda España:

1. Lee fichero de precios. Asegúrate de que todas las variables de precios son numéricas.
2. Cambia los espacios en los nombres de las columnas por ‘_’.
3. Mira los tipos de datos del dataset de precios. Observa que la fecha no ha sido interpretada como tal. Conviértela a tipo fecha.
4. Calcula el porcentaje de NAs de cada columna para la del dataframe de precios.
5. Elimina las columnas que no tengan ningún valor distinto de NA.
6. Crea un dataframe con la media, máximo y mínimo de las columnas numéricas.
7. Habrás visto que el máximo del precio para gasolina 95 es un valor extremadamente alto. Para la variable precio de la gasolina 95, cambia todos los valores mayores a 3 (valores anómalos) por NA. Utiliza el data.frame resultado del ejercicio 5.
8. Selecciona las gasolineras que vendan gasolina 95 o 98 utilizando el dataframe resultado.
9. Selecciona las columnas que representan variables de precios usando el dataframe resultado del ejercicio 7.
10. Selecciona las 5 gasolineras con el precio de gasóleo B más barato.
11. Añade dos nuevas variables con los litros por euro de gasolina 95 y 98 usando el dataframe del ejercicio 7.
12. Crea un nuevo dataframe a partir del resultado del ejercicio 7 con las columnas `ID_estacion`, `fecha`, `tipo_gasolina` y `precio` y guarda el resultado en la variable `precios_long`. La variable `tipo_gasolina` tendrá los valores: “Precio gasolina 95”, “Precio gasóleo A”, “Precio gasóleo B”, “Precio bioetanol”, “Precio nuevo gasóleo A”, “Precio biodiesel”, “Precio gasolina 98”, “Precio gas natural comprimido”, “Precio gas natural licuado” y “Precio gases licuados del petróleo”.
13. En la variable `tipo_gasolina` del ejercicio anterior, elimina el texto “Precio”.
14. Haz un histograma de todas las variables de precio **usando facetas** de ggplot.
15. Lee el fichero de estaciones y júntalo con el dataframe resultado del ejercicio 7 utilizando como clave el `ID_ESTACION`. Considerar únicamente las gasolineras presentes en las dos tablas.
16. Calcula el precio medio de gasolina 95 por provincia.
17. Añade una variable de tipo lógico con nombre `ind_24H`, indicando si la gasolinera es 24H.
18. Convierte la variable rótulo en un factor que tenga 6 niveles: los 5 rótulos más frecuentes y un nivel “Otros” con el resto.
19. Realiza un gráfico de barras con el conteo de estaciones por cada tipo de rótulo.
20. Representa cada gasolinera de Repsol o Cepsa como un punto en función de su longitud y latitud. Cada punto debe estar coloreado de acuerdo al Rótulo.