# Análisis descriptivo Bike Sharing

## Objetivo

Realizar un análisis descriptivo de Bike Sharing para entender la variable respuesta cnt y su relación con predictores, preparar datos y separar conjuntos de modelización y validación.

## Datos

Conjuntos: day.csv (agregado diario) y hour.csv (por hora). La variable respuesta es cnt (total de alquileres), donde cnt = registered + casual.

## Metodología

Lectura y tipado de fecha (dteday); perfilado (dimensiones, tipos, faltantes, estadísticos); tratamiento defensivo de faltantes; correlaciones con cnt; split temporal 80/20 por dteday; visualizaciones (distribución de cnt y heatmaps).

## Resultados clave

Relación con cnt (day): registered (0.946), casual (0.673), atemp (0.631), temp (0.627), instant (0.629), yr (0.567), weathersit (-0.297), windspeed (-0.235), hum (-0.101).

Relación con cnt (hour): registered (0.972), casual (0.695), temp (0.405), atemp (0.401), hr (0.394), hum (-0.323), weathersit (-0.142).

### Distribución de cnt

Figura 1. Distribución de cnt (day)

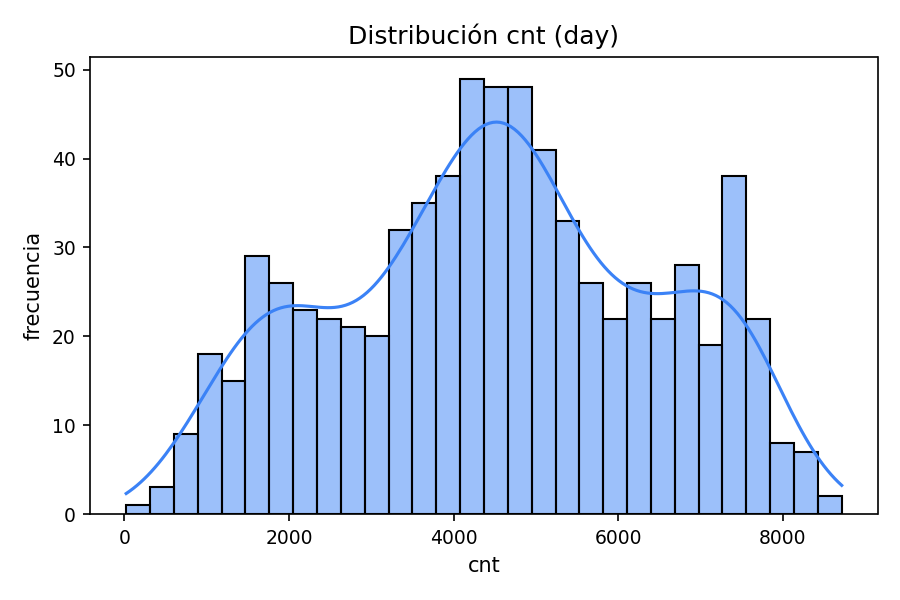
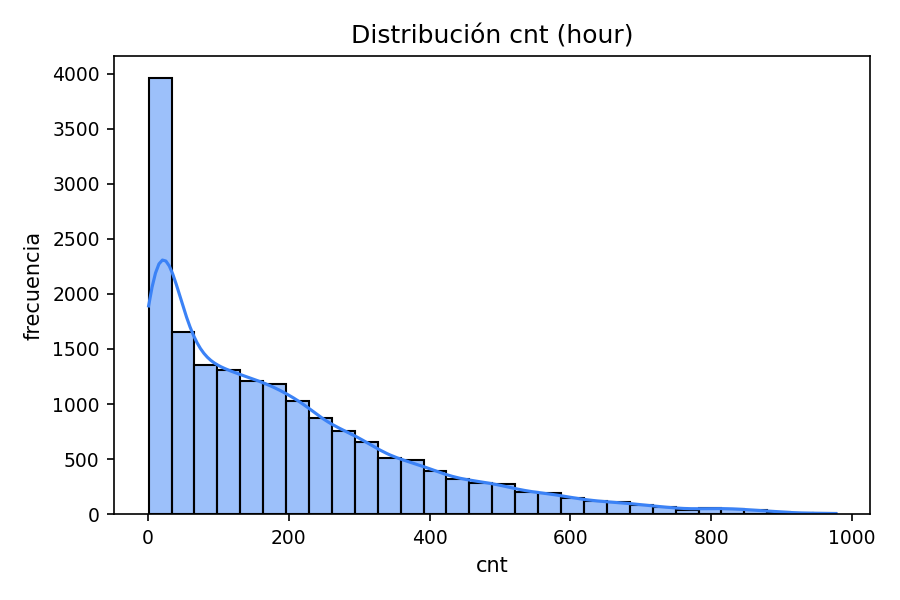


Figura 2. Distribución de cnt (hour)



## Comentarios sobre predictores y cnt

registered y casual explican gran parte de la variación de cnt pero no deben usarse como predictores (fuga de información). Temperatura real/aparente y variables temporales muestran relación positiva; peor clima, humedad y viento altos se asocian con menor demanda.

### Correlaciones (heatmaps)

Figura 3. Heatmap de correlaciones (day)

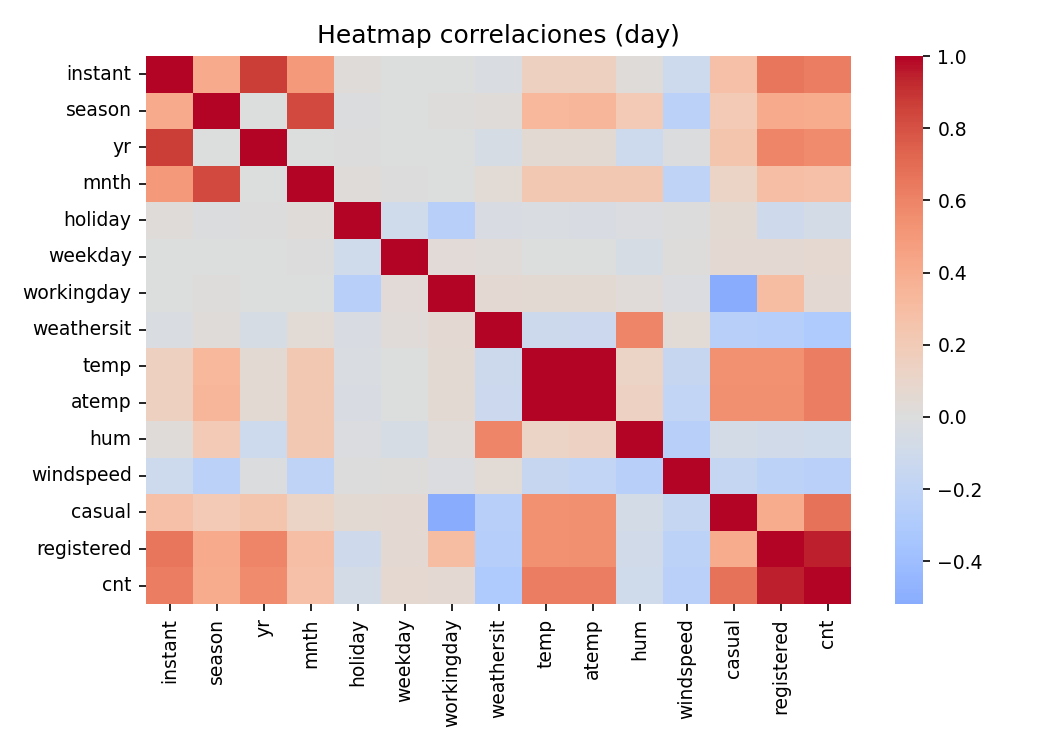
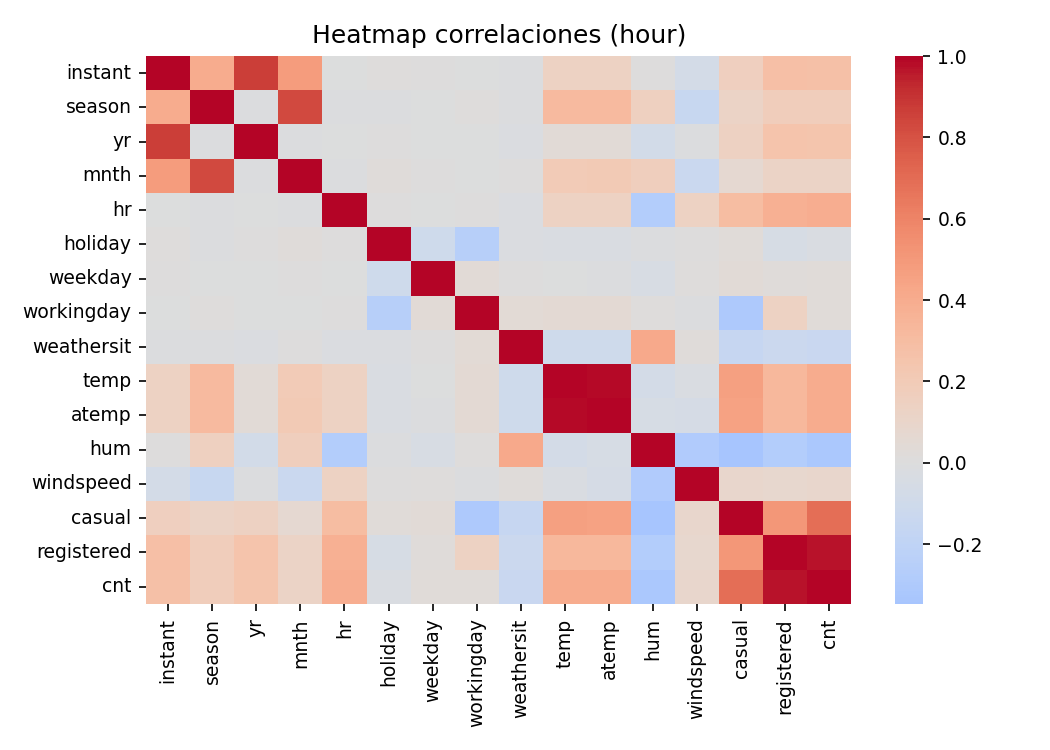


Figura 4. Heatmap de correlaciones (hour)



## Preparación de datos y validación

No se observaron faltantes relevantes; se dejó imputación defensiva. Se aplicó división temporal 80/20 por dteday para modelización y validación.

## Conclusiones

cnt crece con temperatura y con el paso del tiempo (yr), y disminuye con malas condiciones meteorológicas, humedad y viento. Las horas del día influyen marcadamente (picos en horas punta en hour.csv). Para modelar cnt, excluir registered y casual; centrarse en clima, tiempo y calendario.

## Limitaciones y próximos pasos

La correlación no implica causalidad; conviene explorar modelos con validación temporal (p.ej., regresión regularizada, árboles/boosting) y variables derivadas (fines de semana/feriados, no linealidades en clima).