PREPARACIÓN Y LIMPIEZA BASE DE DATOS

VARIABLES: *priceDescription_rent, metro_count, tram_count, train_count, supermarket_count, pharmacy_count, hospital_count, university_count y college_count.*

1. Tabla resumen-> todas las variables son numéricas.

2. Resumen estadístico:
   - Borrar la variable tram_count (es siempre 0).
     fotocasa$"tram_count" = NULL

   - Borrar fila anómala (sin barrio).
     fotocasa= fotocasa[fotocasa$"metro_count"!= 64,]

3. Valores inconsistentes:
   - Cambiar valores en filas con priceDescription_rent=0 (fotocasa no consideraba que 3 pisos fueran suficientes para calcular la media, calculado a mano).
     -Poble Nou: 1320€/80m2
     fotocasa$priceDescription_rent[fotocasa$priceDescription_rent == 0 & fotocasa$neighborhood == "Poble Nou"]=1320
     -Massarrojos: 832€/80m2
     variables$priceDescription_rent[variables$priceDescription_rent == 0 & variables$neighborhood == "Massarrojos"]=832

4. Valores faltantes: no hay valores faltantes en ninguna de las variables

5. Agrupar datos:
   - metro_count+train_count=public_transport_count
     variables$public_transport_count=variables$metro_count+variables$train_count

6. Distribuciones:
   - PriceDescription_rent es relativamente simétrica, con varios picos (posibles subgrupos).
   - Variables conteo servicios tienen asimetría positiva, a lo mejor es necesario aplicar alguna transformación en posteriores análisis (en ese caso, el df ya está creado pero no te lo paso porque de momento no sirve).