

A continuación tienes una **guía operativa** — en orden cronológico — para que apliques un Análisis Exploratorio de Datos (EDA) completo sobre `celulares_todos_base.csv`, junto con un banco de ideas de negocio y todas las líneas de análisis que podrías explotar más adelante.

1. Definir el propósito y las preguntas iniciales

Antes de tocar una sola línea de código, detalla **qué decisiones debes informar** con estos datos — por ejemplo:

Área del negocio	Pregunta orientativa
Pricing	¿Qué especificaciones técnicas explican la mayor parte del precio?
Promociones	¿Qué marcas responden mejor a los descuentos (variable <code>per_descuento</code>)?
Product-mix	¿Existen nichos de “gama media” con alto margen ocultos entre la RAM de 6–8 GB?
UX web	¿Qué combinaciones de atributos atraen visitas pero no compras?

La literatura de EDA recalca que **todo análisis debe nacer de preguntas de negocio** y refinarse iterativamente a medida que surgen hallazgos .

2. Reconocimiento rápido del dataset

Carga y estructura

```
celu <- read_csv("celulares_todos_base.csv")
glimpse(celu)      # tipos de dato
sapply(celu, n_distinct)
```

1. Valida dimensión, tipos y posibles claves (“`sitio + marca + modelo`” suele ser única).

Mapa de valores faltantes

```
naniar::vis_miss(celu)
colSums(is.na(celu))
```

2. Solo la columna **Procesador** muestra NAs según inspecciones previas; documenta su proporción y decide estrategia .

Duplicados e integridad

```
sum(duplicated(celu))  
validate::confront(celu, rule = Precio_comprador > 0)
```

3. Elimina o justifica cualquier infracción.
-

3. EDA univariante

Numéricas: media, mediana, IQR, asimetría; histogramas y boxplots para

Precio_comprador, **RAM (GB)**, **Almacenamiento**, etc.

Categorías: frecuencias y gráficos de barras para **marca**, **Sistema Operativo**, **NFC**.

Este paso detecta valores atípicos y distribuciones inusuales .

4. EDA bivalente

1. **Correlaciones numéricas** (**cor**, **ggcorrplot**) para identificar multicolinealidad.
2. **Numérica × Categórica**: boxplots de precio por marca, por presencia de NFC, por gama RAM.
3. **Categórica × Categórica**: chi-cuadrado entre **NFC** y **Sistema Operativo** para ver asociación.

Preguntas tácticas que podrías responder aquí:

- ¿El sobreprecio de los iPhone es lineal respecto al almacenamiento?
 - ¿Android 15 se vende sistemáticamente con NFC?
 - ¿El descuento medio varía por sitio?
-

5. EDA multivariante y segmentación

- **PCA**: reducir 10–12 numéricas y visualizar clusters naturales.
- **Clustering (k-means / hclust)**: segmentar teléfonos por *feature set* → sirve para campañas de marketing personalizadas.
- **Pairplots** (`GGally::ggpairs`) para ver relaciones no lineales complejas .

6. Manejo de outliers y transformación

- Detecta precios imposibles con regla $1.5 \times \text{IQR}$ o z-score ± 3 .
- Decide si se **winsoriza**, corrige o excluye: ¿es un error de carga o un “foldable” de nicho?
- Escala \log_{10} los precios antes de PCA para neutralizar colas largas .

7. Ingeniería de nuevas variables

Nueva feature	Cómo se calcula	Para qué sirve
<code>precio_por_GB</code>	<code>Precio_comprador / Almacenamiento</code>	Sensibilidad a memoria
<code>ram_band</code>	<code>Case_when(RAM ≤ 4, 4–8, ≥ 12)</code>	Simplificar análisis de gama
<code>flag_NFC</code>	<code>NFC == "Si"</code>	Modelo binario rápido
<code>descuento_pesos</code>	<code>Precio_anterior - Precio_comprador</code>	Evaluar elasticidad
<code>rel_camaras</code>	Principal / Frontal	Valor percibido de fotografía

Estas variables suelen mejorar modelos de predicción de precio o compra .

8. Automatizar y documentar

- Genera un **reporte HTML** con `DataExplorer::create_report(celu)`.

- Versiona los *scripts* y los datasets limpios; anota cada decisión (por qué imputaste, por qué eliminaste).
- Combina **exploración manual** y **checks automáticos** para no perder patrones sutiles ni dejar huecos sin revisar .

9. Ideas avanzadas / “todas las posibilidades”

Línea de trabajo	Qué podrías hacer	Valor para el negocio
Elasticidad de precio	Regresión robusta: <code>Precio_comprador ~ specs + descuento</code>	Ajustar promociones por rango de memoria
Market basket web	Si tienes logs de clics, usa asociación (<code>apriori</code>) para ver qué specs se consultan juntas	Diseñar packs y bundles
Forecast de stock	Enlazar ventas históricas y specs, usar modelos ARIMAX por ciclo release	Optimizar inventario
Análisis de sentimiento	Cruzar con reseñas y hacer NLP → <code>sentiment ~ specs</code>	Priorizar mejoras de producto
Simulación de escenarios	Tableros interactivos (<i>Shiny / Power BI</i>) para probar precios y descuentos	Soporte a la negociación con proveedores

Checklist final (condensado)

1. **Contexto** → Formular 3–5 preguntas de negocio clave.
2. **Calidad** → Tipos, NAs, duplicados, rangos.
3. **Univariante** → Estadísticos + gráficos.
4. **Bivariante** → Correlaciones, boxplots, tests.
5. **Multivariante** → PCA, clustering.
6. **Outliers** → Detectar y decidir acción.
7. **Nuevas features** → crear variables derivadas.

8. **Documentar y automatizar** → reportería reproducible.

Sigue cada paso con la filosofía **escéptica-visual-iterativa** del EDA ; regresa al paso previo cuando un hallazgo lo amerite, y mantén siempre el foco en las **decisiones de negocio** que necesitas respaldar.