

BOB ENTERPRISE

1. Selección y Justificación de la Base de Datos

a. Búsqueda de la Base de Datos:

Se decidió y optó por una Base de Datos de prueba.

El sitio de descarga de la base de datos fue: <https://www.kaggle.com>

El contexto de la base de datos seleccionada trata de la empresa de Bob, el cual, ha iniciado su propio negocio de telefonía móvil y quiere determinar a cuál giro de gama de venta de celular enfocarse, para competir con la grandes empresas como Samsung, Apple, entre otras.

b. Justificación de la Selección:

En esta problemática lo que se quiere descubrir es alguna relación entre las características de un teléfono móvil como lo son RAM, Batery, Resolución y Memoria interna, entre otras, Para poder determinar el Rango de Gama (Baja, Media, Alta) entran esas características, y poder apoyar a Bob a a elegir el giro al cual enfocarse en la venta de celulares.

2. Preparación y Limpieza de los Datos:

a. Descripción de la Base de Datos:

La base original

nombre BD original: celulares.csv

formato de BD: csv

tamaño: 120KB

contenido: tiene 2,000 registros y 21 campos.

Descripción de campos:

id	clave única
battery_power	Energía total que una batería puede almacenar en un tiempo medida en mAh
blue	tiene bluetooth o no
clock_speed	Velocidad a la que el microprocesador ejecuta instrucciones.
dual_sim	Tiene soporte dual sim o no
fc	Megapíxeles de la cámara frontal
four_g	Tiene 4G o no
int_memory	Memoria interna en Gigabytes
m_dep	Profundidad del móvil en cm
mobile_wt	Peso del teléfono móvil
n_cores	Número de núcleos de procesador
pc	Megapíxeles de la cámara principal
px_height	Altura de resolución de píxeles

px_width	Ancho de resolución de píxeles
ram	Memoria de acceso aleatorio en megabytes
sc_h	Altura de pantalla del móvil en cm
sc_w	Ancho de pantalla del móvil en cm
talk_time	mayor tiempo que durará una sola carga de batería cuando esté hablando
three_g	Tiene 3G o no
touch_screen	Tiene pantalla táctil o no
Wifi	Tiene wifi o no

El tipo de datos a utilizar fueron numéricos y categóricos.

Para determinar el tipo de datos de los campos se realizó un análisis en R,

```
> str(datos) #nos indica los tipos de datos de las columnas
spec_tbl_ [2,000 × 21] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ battery_power: num [1:2000] 842 1021 563 615 1821 ...
 $ blue         : num [1:2000] 0 1 1 1 1 0 0 0 1 1 ...
 $ clock_speed  : num [1:2000] 2.2 0.5 0.5 2.5 1.2 0.5 1.7 0.5 0.5 0.6 ...
 $ dual_sim     : num [1:2000] 0 1 1 0 0 1 0 1 0 1 ...
 $ fc           : num [1:2000] 1 0 2 0 13 3 4 0 0 2 ...
 $ four_g       : num [1:2000] 0 1 1 0 1 0 1 0 0 1 ...
 $ int_memory   : num [1:2000] 7 53 41 10 44 22 10 24 53 9 ...
 $ m_dep        : num [1:2000] 0.6 0.7 0.9 0.8 0.6 0.7 0.8 0.8 0.7 0.1 ...
 $ mobile_wt    : num [1:2000] 188 136 145 131 141 164 139 187 174 93 ...
 $ n_cores      : num [1:2000] 2 3 5 6 2 1 8 4 7 5 ...
 $ pc           : num [1:2000] 2 6 6 9 14 7 10 0 14 15 ...
 $ px_height    : num [1:2000] 20 905 1263 1216 1208 ...
 $ px_width     : num [1:2000] 756 1988 1716 1786 1212 ...
 $ ram          : num [1:2000] 2549 2631 2603 2769 1411 ...
 $ sc_h         : num [1:2000] 9 17 11 16 8 17 13 16 17 19 ...
 $ sc_w         : num [1:2000] 7 3 2 8 2 1 8 3 1 10 ...
 $ talk_time    : num [1:2000] 19 7 9 11 15 10 18 5 20 12 ...
 $ three_g      : num [1:2000] 0 1 1 1 1 1 1 1 1 1 ...
 $ touch_screen : num [1:2000] 0 1 1 0 1 0 0 1 0 0 ...
 $ wifi         : num [1:2000] 1 0 0 0 0 0 1 1 0 0 ...
 $ price_range  : num [1:2000] 1 2 2 2 1 1 3 0 0 0 ...
```

También se hizo una revisión de las características de los campos, utilizando un resumen.

```
> summary(datos)
battery_power      blue      clock_speed      dual_sim      fc
Min.   : 501.0    Min.   :0.000    Min.   :0.500    Min.   :0.0000    Min.   : 0.000
1st Qu.: 851.8    1st Qu.:0.000    1st Qu.:0.700    1st Qu.:0.0000    1st Qu.: 1.000
Median :1226.0    Median :0.000    Median :1.500    Median :1.0000    Median : 3.000
Mean   :1238.5    Mean   :0.495    Mean   :1.522    Mean   :0.5095    Mean   : 4.309
3rd Qu.:1615.2    3rd Qu.:1.000    3rd Qu.:2.200    3rd Qu.:1.0000    3rd Qu.: 7.000
Max.   :1998.0    Max.   :1.000    Max.   :3.000    Max.   :1.0000    Max.   :19.000

four_g      int_memory      m_dep      mobile_wt      n_cores
Min.   :0.0000    Min.   : 2.00    Min.   :0.1000    Min.   : 80.0    Min.   :1.000
1st Qu.:0.0000    1st Qu.:16.00    1st Qu.:0.2000    1st Qu.:109.0    1st Qu.:3.000
Median :1.0000    Median :32.00    Median :0.5000    Median :141.0    Median :4.000
Mean   :0.5215    Mean   :32.05    Mean   :0.5018    Mean   :140.2    Mean   :4.521
3rd Qu.:1.0000    3rd Qu.:48.00    3rd Qu.:0.8000    3rd Qu.:170.0    3rd Qu.:7.000
Max.   :1.0000    Max.   :64.00    Max.   :1.0000    Max.   :200.0    Max.   :8.000

pc      px_height      px_width      ram      sc_h
Min.   : 0.000    Min.   : 0.0    Min.   : 500.0    Min.   : 256    Min.   : 5.00
1st Qu.: 5.000    1st Qu.: 282.8    1st Qu.: 874.8    1st Qu.:1208    1st Qu.: 9.00
Median :10.000    Median : 564.0    Median :1247.0    Median :2146    Median :12.00
Mean   : 9.916    Mean   : 645.1    Mean   :1251.5    Mean   :2124    Mean   :12.31
3rd Qu.:15.000    3rd Qu.: 947.2    3rd Qu.:1633.0    3rd Qu.:3064    3rd Qu.:16.00
Max.   :20.000    Max.   :1960.0    Max.   :1998.0    Max.   :3998    Max.   :19.00

sc_w      talk_time      three_g      touch_screen      wifi
Min.   : 0.000    Min.   : 2.00    Min.   :0.0000    Min.   :0.000    Min.   :0.000
1st Qu.: 2.000    1st Qu.: 6.00    1st Qu.:1.0000    1st Qu.:0.000    1st Qu.:0.000
Median : 5.000    Median :11.00    Median :1.0000    Median :1.000    Median :1.000
Mean   : 5.767    Mean   :11.01    Mean   :0.7615    Mean   :0.503    Mean   :0.507
3rd Qu.: 9.000    3rd Qu.:16.00    3rd Qu.:1.0000    3rd Qu.:1.000    3rd Qu.:1.000
Max.   :18.000    Max.   :20.00    Max.   :1.0000    Max.   :1.000    Max.   :1.000

price_range
Min.   :0.00
1st Qu.:0.75
Median :1.50
Mean   :1.50
3rd Qu.:2.25
Max.   :3.00
```

b. Limpieza de Datos:

Eliminamos categorías que se consideraron irrelevantes para el análisis.

Analizando y limpiando la base de datos quedaron 2,000 registros y 8 campos, los cuales son los que se están utilizando en el visualizador final.

3. Análisis Exploratorio de Datos (EDA)

Análisis Descriptivo:

- Estadísticos descriptivos de la base de datos por rango de gama
- Se utilizaron las herramientas: Excel, R e IA.

Para definir los rangos de precios utilizados en este análisis, haciendo uso de la herramienta de excel, comenzamos con un proceso manual de clasificación basado en los valores mínimos y máximos de las características técnicas de los teléfonos móviles. Por ejemplo, en el caso de la RAM, si el valor más bajo era 256 MB y el más alto 3998 MB, dividimos estos valores en tres rangos. El rango “Baja” comprendió los valores entre 256 y 1503 MB, el rango “Media” incluyó valores entre 1504 y 2750 MB, y el rango “Alta” abarcó desde 2751 hasta 3998 MB. Este mismo procedimiento se aplicó a otras características clave como la memoria interna y la resolución de pantalla.

Además, para facilitar la graficación en Tableau, en lugar de etiquetar los rangos como “Baja”, “Media” o “Alta”, se les asignaron valores numéricos (1, 2 y 3). Esto permitió una visualización más clara y sencilla de los datos.

Dado que los datos contenían una gran cantidad de valores, también utilizamos técnicas de inteligencia artificial (IA) para refinar aún más esta clasificación. Esto nos permitió determinar con mayor precisión qué valores correspondían a cada rango, mejorando la exactitud del análisis y asegurando que los rangos reflejaran de manera fiel las características técnicas de los dispositivos.

Formulas en excel para Clasificacion Ram:

=SI(O2>2259,3,SI(O2>=387,2,1))

Formulas en excel para Clasificacion Battery:

=SI(B2>1499,3,SI(B2>=1000,2,1))

Formulas en excel para Clasificacion Resolucion:

=SI((M2*N2)>2000000,3,SI((M2*N2)>=1000000,2,1))

Formulas en excel Clasificacion Memoria Interna:

=SI(H2>44,3,SI(H2>=23,2,1))

Formulas en excel Puntuación

=SUMA(W2:Z2)

c) Distribuciones de variables/campos

Los campos fueron renombrados y recalculados, haciendo uso de la herramienta de excel.

id	clave única
Clasificacion Ram	Valor numérico (1,2,3)
Clasificacion Battery	Valor numérico (1,2,3)
Clasificacion Resolucion	Valor numérico (1,2,3)
Clasificacion Memoria Interna	Valor numérico (1,2,3)
Puntuacion	Permite almacenar la puntuación, la cuál se obtiene con la suma de Clasificacion Ram, Clasificacion Battery, Clasificacion Resolucion y Clasificacion Memoria Interna.
Rango de Gama	valores (Baja, Media, Alta)

4. Visualización de Datos

a) Herramienta de Visualización:

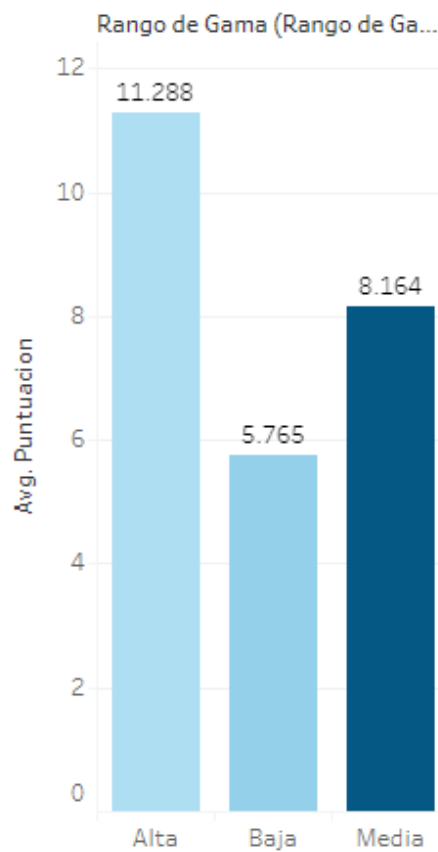
Se utilizó la herramienta de Tableau

b) Principales Visualizaciones:

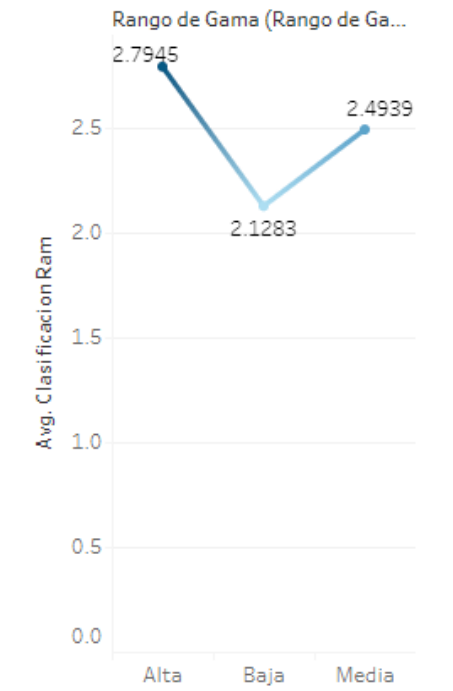
Relación entre Clasificaciones y Rango de Precios:

	Rango de Gama (Rango de Ga...		
	Alta	Baja	Media
Avg. Puntuacion	11	6	8
Clasificacion Battery	206	474	3,287
Clasificacion Memoria Int..	209	475	3,204
Clasificacion Ram	204	796	3,873
Clasificacion Resolucion	205	411	2,315

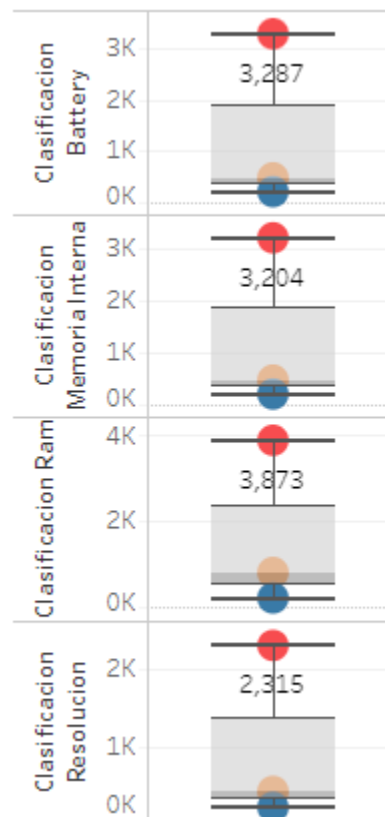
Comparación de Gamas por Puntuación:



Tendencias de Puntuación en Cada Gama:



Distribución de Clasificaciones por Gama:



5. Interpretación y Conclusiones

1. **Tabla: Relación entre Clasificaciones y Rango de Precios:**

La primera tabla proporciona una comparación de cuatro atributos clave (Batería, Memoria Interna, RAM y Resolución) a lo largo de los tres rangos de precios. Muestra la clasificación de cada atributo para teléfonos de gama alta, media y baja. La tabla revela que los teléfonos en el rango “Media” generalmente tienen mejores puntajes promedio en todos los atributos, lo que indica un rendimiento equilibrado. Por ejemplo, la RAM en la gama media tiene un puntaje promedio de 3.873, notablemente más alto que en la gama alta con 204 y en la baja con 796. De manera similar, la batería y la memoria también muestran un mejor desempeño en la gama media.

2. **Gráfico de Barras: Comparación de Gamas por Puntaje Promedio:**

El gráfico de barras ilustra el puntaje promedio general (“Prom. Puntuación”) para cada rango de precios. Los teléfonos en la gama “Alta” tienen el puntaje promedio más alto, con un valor de 11.288, seguidos de los teléfonos de gama “Media” con 8.164. La gama “Baja” se queda atrás con un promedio de 5.765. Esto indica que, si bien los teléfonos de gama alta tienen el mejor rendimiento, los de gama media aún ofrecen resultados competitivos, lo que los convierte en una opción fuerte para los productos de Bob.

3. **Gráfico de Líneas: Tendencias en el Rendimiento de la RAM por Gama:**

El gráfico de líneas se centra en el rendimiento de la RAM a lo largo de los tres rangos de precios. Muestra una caída en el rendimiento para los teléfonos de la gama “Baja”, con un puntaje promedio de 2.1283, mientras que las gamas “Alta” y “Media” tienen puntajes más altos, alcanzando la gama “Media” un puntaje de 2.4939. Esto sugiere que el rendimiento de la RAM en los teléfonos de gama media es comparable al de los teléfonos de gama alta, reforzando la competitividad de este segmento.

4. **Diagrama de Cajas: Distribución de Clasificaciones por Gama:**

La última visualización es un diagrama de cajas que muestra la distribución de las clasificaciones para cada atributo (Batería, Memoria Interna, RAM y Resolución) en los tres rangos de precios. Las medianas y los patrones de distribución indican que los teléfonos de gama “Media” ofrecen un rendimiento equilibrado en varios aspectos. Por ejemplo, las clasificaciones de RAM y memoria interna para los teléfonos de gama media son similares a las de los teléfonos de gama alta, e incluso en algunos casos las superan, especialmente en comparación con los teléfonos de gama baja.

Conclusión

Con base en este análisis, el producto de Bob encajaría mejor en la **gama media de precios**. Los datos muestran que los teléfonos en la gama media ofrecen un equilibrio sólido en el rendimiento de atributos clave como la RAM, la capacidad de la batería y la memoria interna. Aunque los teléfonos de gama alta logran los mejores puntajes en general, los de gama media proporcionan un rendimiento fuerte a un precio más competitivo, lo que los convierte en una opción viable en el mercado. Al posicionar su producto en la gama media, Bob puede competir de manera efectiva sin sacrificar demasiado en términos de calidad o rendimiento.