

CLASIFICACIÓN DE CARACTERÍSTICAS DE DISPOSITIVOS MÓVILES

PRESENTACIÓN



ANÁLISIS DE TENDENCIAS EN CARACTERÍSTICAS DE DISPOSITIVOS MÓVILES

En este proyecto, el objetivo es analizar cómo han evolucionado las características de los teléfonos móviles a lo largo del tiempo, considerando factores como la capacidad de la batería, la resolución de pantalla, la memoria RAM y otras especificaciones técnicas.

El análisis se enfoca en identificar tendencias claves que podrían ayudar a anticipar las futuras necesidades de los consumidores, brindando información valiosa para el desarrollo de productos innovadores en la industria.

Estos datos serán explorados en las primeras etapas del análisis exploratorio de datos (EDA).

COLUMNAS DEL DATASET

1. ****id****: Identificador único para cada teléfono móvil.
2. ****battery_power****: La capacidad de la batería en mAh.
3. ****blue****: Indicador binario (1 o 0) de si el teléfono tiene Bluetooth.
4. ****clock_speed****: La velocidad del reloj del procesador en GHz.
5. ****dual_sim****: Indicador binario (1 o 0) de si el teléfono soporta dos SIMs.
6. ****fc****: Número de la cámara frotal en megapíxeles.
7. ****four_g****: Indicador binario (1 o 0) de si el teléfono soporta 4G.
8. ****int_memory****: Memoria interna en GB.
9. ****m_dep****: Profundidad del dispositivo en cm.
10. ****mobile_wt****: Peso del dispositivo en gramos.
11. ****n_cores****: Número de núcleos del procsador.
12. ****pc****: Número de la cámara principal en megapíxeles.
13. ****px_height****: Altura de la pantalla en píxeles.
14. ****px_width****: Ancho de la pantalla en píxeles.
15. ****ram****: Memoria RAM en MB.
16. ****sc_h****: Altura de la pantalla en cm.
17. ****sc_w****: Ancho de la pantalla en cm.
18. ****talk_time****: Tiempo de conversación en horas.
19. ****three_g****: Indicador binario (1 o 0) de si el teléfnso soporta 3G.
20. ****touch_screen****: Indicador binario (1 o 0) de si el teléfono tiene pantalla táctil.
21. ****wifi****: Indicador binario (1 o 0) de si el teléfnso tiene WiFi.

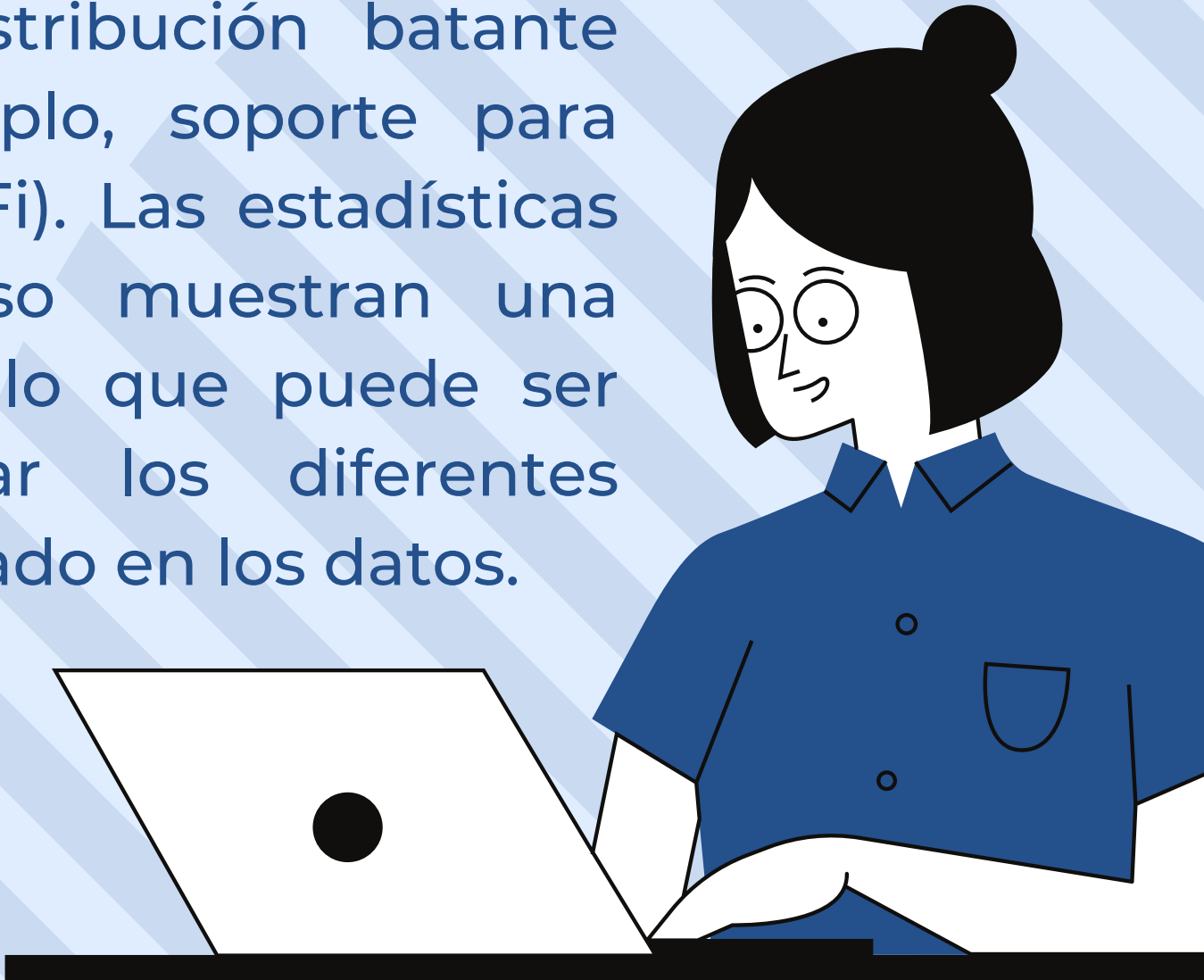


ANÁLISIS DE LA SALIDA DESCRIBE

	id	battery_power	blue	clock_speed	dual_sim	
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	500.500000	1248.510000	0.516000	1.540900	0.517000	
std	288.819436	432.458227	0.499994	0.829268	0.499961	
min	1.000000	500.000000	0.000000	0.500000	0.000000	
25%	250.750000	895.000000	0.000000	0.700000	0.000000	
50%	500.500000	1246.500000	1.000000	1.500000	1.000000	
75%	750.250000	1629.250000	1.000000	2.300000	1.000000	
max	1000.000000	1999.000000	1.000000	3.000000	1.000000	
	fc	four_g	int_memory	m_dep	mobile_wt	...
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	...
mean	4.593000	0.487000	33.652000	0.517500	139.51100	...
std	4.463325	0.500081	18.128694	0.280861	34.85155	...
min	0.000000	0.000000	2.000000	0.100000	80.00000	...
25%	1.000000	0.000000	18.000000	0.300000	109.75000	...
50%	3.000000	0.000000	34.500000	0.500000	139.00000	...
75%	7.000000	1.000000	49.000000	0.800000	170.00000	...
max	19.000000	1.000000	64.000000	1.000000	200.00000	...
	pc	px_height	px_width	ram	sc_h	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	10.054000	627.121000	1239.774000	2138.998000	11.995000	
std	6.095099	432.929699	439.670981	1088.092278	4.320607	
min	0.000000	0.000000	501.000000	263.000000	5.000000	
...						
75%	8.000000	16.000000	1.000000	1.000000	1.000000	
max	18.000000	20.000000	1.000000	1.000000	1.000000	

[8 rows x 21 columns]

El análisis muestra una variedad de características para los dispositivos móviles en el DataFrame. Los valores indican que hay una amplia gama de capacidades y especificaciones entre los dispositivos, con varias características que tienen una distribución bastante equitativa (por ejemplo, soporte para Bluetooth, 3G y Wi-Fi). Las estadísticas de memoria y peso muestran una amplia variabilidad, lo que puede ser útil para identificar los diferentes segmentos del mercado en los datos.



ANÁLISIS DE LA DISTRIBUCIÓN DE CARACTERÍSTICAS DE DISPOSITIVOS MÓVILES

Análisis de la Distribución de Características de Dispositivos Móviles

- Distribución de la batería: La capacidad de la batería muestra una distribución amplia, con concentración en ciertos rangos. Esto indica una variedad considerable en la duración de la batería entre los dispositivos.

- Distribución sesgada: Algunas características, como px_height y px_width, presentan una distribución sesgada hacia la derecha, lo que indica una mayor cantidad de dispositivos con resoluciones de pantalla más altas.

- Valores atípicos: Se observan valores atípicos en algunos histogramas, que podrían ser errores en los datos o representar dispositivos con características inusuales.

Análisis por Característica

- battery_power: La distribución de la capacidad de la batería sugiere una demanda significativa por dispositivos con baterías de mayor duración, aunque también hay muchos dispositivos con baterías más pequeñas.

- ram: La distribución de la RAM muestra un aumento en la cantidad de memoria a lo largo del tiempo, reflejando la tendencia hacia dispositivos más potentes.

- px_height y px_width: La distribución sesgada hacia pantallas más grandes y con mayor resolución refleja la evolución hacia mejores experiencias visuales.

- n_cores: La distribución de los núcleos del procesador indica que los dispositivos con un número moderado de núcleos son los más comunes.

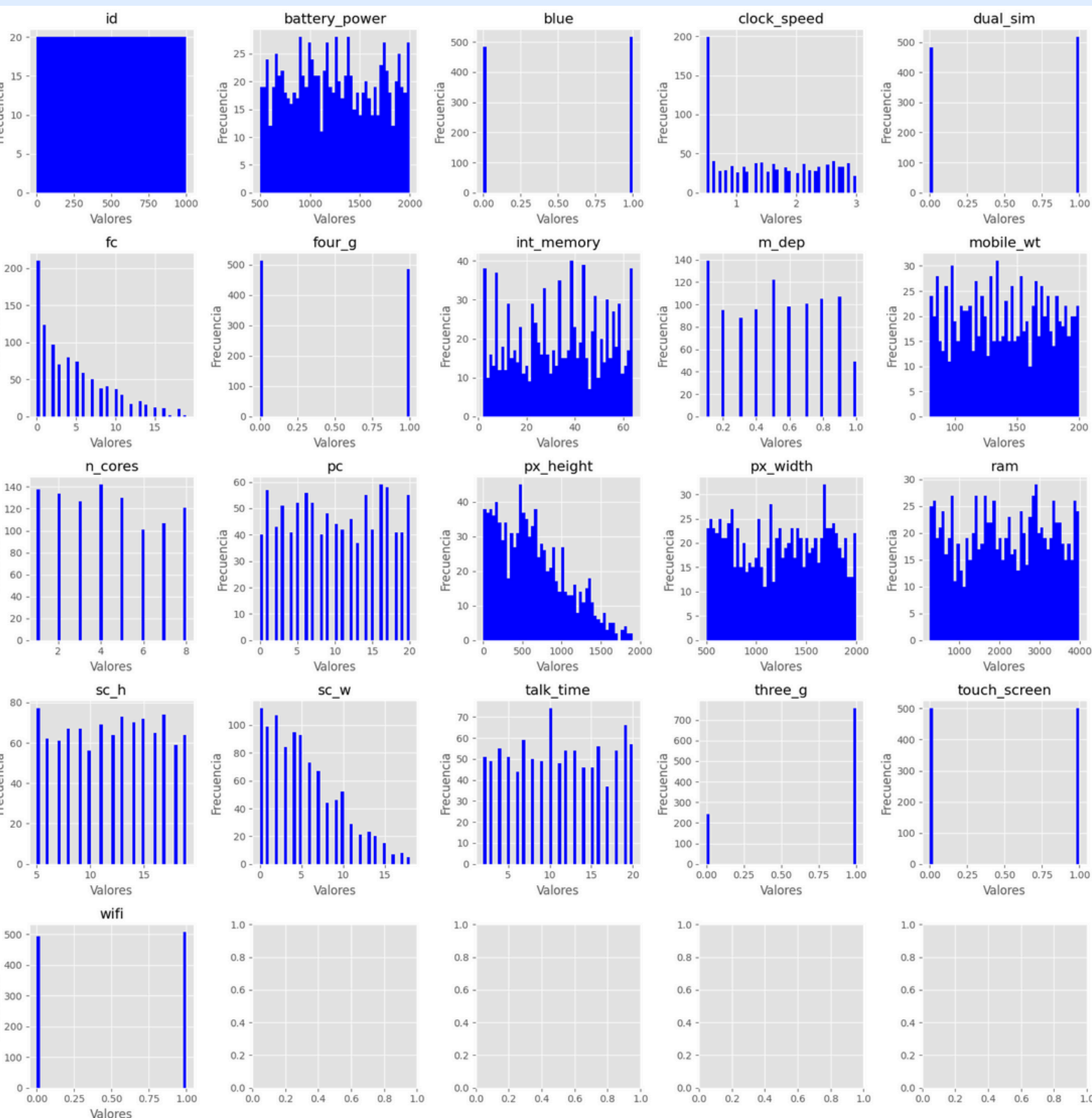
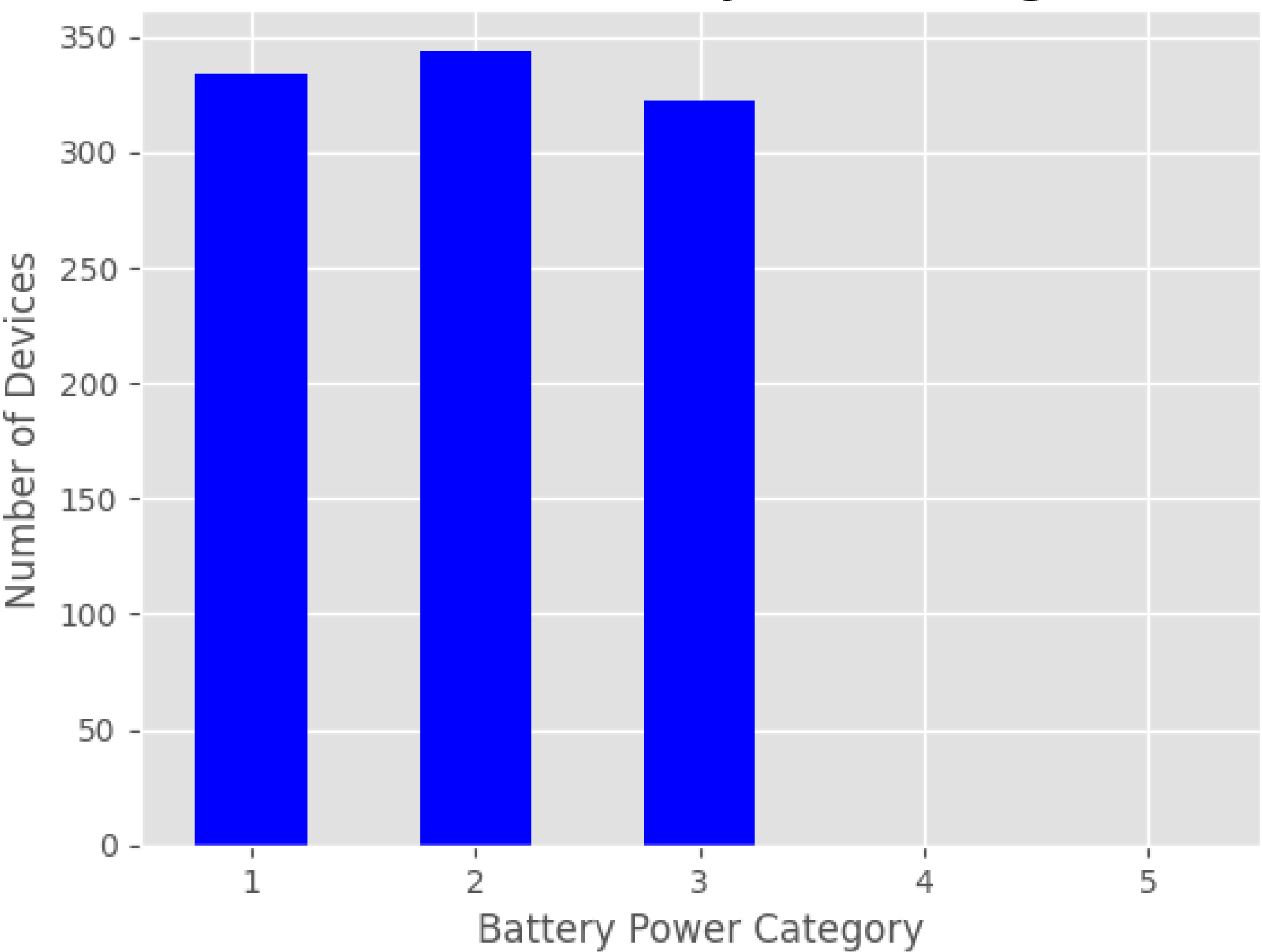


GRÁFICO DE CAPACIDADES DE BATERÍA

Distribution of Battery Power Categories



Eje X:

El eje X representa las categorías de capacidad de batería que has creado:

- 1. Batería de hasta 1000 mAh (miliamperios hora)
- 2. Batería de 1000 mAh a 1500 mAh
- 3. Batería de 1500 mAh a 2000 mAh
- 4. Batería de 2000 mAh a 2500 mAh
- 5. Batería de más de 2500 mAh

Eje Y:

El eje Y representa el número de dispositivos en cada categoría.

Observaciones:

- 1. Distribución: La distribución de las capacidades de batería parece estar sesgada hacia la derecha. Esto significa que la mayoría de los dispositivos tienen una capacidad de batería relativamente baja (categorías 1 y 2), mientras que hay menos dispositivos con baterías de mayor capacidad (categorías 4 y 5).
- 2. Categoría más frecuente: La categoría 2 (de 1000 mAh a 1500 mAh) parece ser la más frecuente, lo que indica que una porción considerable de los dispositivos tienen una capacidad de batería moderada.
- 3. Categorías menos frecuentes: Las categorías 1 (menos de 1000 mAh) y 5 (más de 2500 mAh) son las menos frecuentes.

Interpretaciones:

En el conjunto de datos, parece predominar el uso de baterías de capacidad moderada. Existen menos dispositivos en los extremos del espectro, tanto con baterías muy pequeñas como con baterías de gran capacidad.

ANÁLISIS DE PROPORCIONES DE CATEGORÍAS DE POTENCIA DE BATERÍA

La tabla muestra la comparación de proporciones de categorías de potencia de batería entre tres conjuntos de datos:

- 1. **Overall %:** Proporción de cada categoría en el conjunto de datos completo (test_set).
- 2. **Stratified %:** Proporción de cada categoría en el conjunto de datos de prueba estratificado (strat_test_set).
- 3. **Random %:** Proporción de cada categoría en el conjunto de datos de entrenamiento aleatorio (train_set).
- 4. **Strat. Error %:** Error porcentual entre las proporciones estratificadas y las proporciones generales.
- 5. **Rand. Error %:** Error porcentual entre las proporciones aleatorias y las proporciones generales.

Detalles

Battery Power Category	Overall %	Stratified %	Random %	Strat. Error %	Rand. Error %
1	0.34	0.00	0.33	-98.51	-0.37
2	0.34	0.33	0.34	-4.35	-0.36
3	0.32	0.34	0.32	7.81	0.78
4	0.00	0.32	0.00	inf	NaN
5	0.00	0.00	0.00	NaN	NaN

Observaciones

- 1. **Categorías con Proporciones en Overall %:**
 - Las categorías 1, 2, y 3 tienen proporciones en el conjunto de datos completo (Overall %) que oscilan entre 0.32 y 0.34.
 - Las categorías 4 y 5 no están presentes en el conjunto de datos completo.
- 2. **Desviaciones en Stratified %:**
 - La categoría 1 muestra una proporción de 0.00 en el conjunto estratificado, lo que es inusual dado que debería reflejar una proporción proporcional similar a la del conjunto completo. Esto puede indicar un error en la partición estratificada.
 - Las categorías 4 y 5 en el conjunto estratificado tienen proporciones inusuales (0.32 para la categoría 4 y 0.00 para la categoría 5), lo que también puede indicar una distribución no representativa.
- 3. **Errores de Estratificación y Aleatorización:**
 - Categoría 1:** La proporción en el conjunto estratificado es 0.00, resultando en un error porcentual muy alto (-98.51%). Esto sugiere que la categoría 1 está ausente en la muestra estratificada.
 - Categoría 4:** La proporción en el conjunto estratificado es 0.32, lo que lleva a un error porcentual infinito, ya que no hay representación en el conjunto general (Overall % es 0.00).
 - Categoría 5:** La proporción es 0.00 en ambos conjuntos, lo que resulta en un error porcentual no disponible (NaN).

Conclusión

El análisis indica que algunas categorías están mal representadas en el conjunto estratificado en comparación con el conjunto de datos completo. Las categorías 4 y 5 tienen problemas de representación que podrían sugerir errores en el proceso de estratificación o en los datos originales.

GRÁFICO DE DISPERSIÓN

Este gráfico de dispersión nos permite visualizar la relación entre dos variables del conjunto de datos: la potencia de la batería (en el eje x) y el tiempo de conversación (en el eje y) de diferentes modelos de teléfonos móviles.

¿Qué podemos interpretar de este gráfico?

- No hay una relación lineal clara: Los puntos en el gráfico están dispersos y no forman una línea recta ascendente o descendente. Esto sugiere que no hay una relación directa y proporcional entre la potencia de la batería y el tiempo de conversación. Es decir, un teléfono con mayor potencia de batería no necesariamente tiene un tiempo de conversación más largo.

- Hay un rango de valores: Los valores de la potencia de la batería se encuentran entre aproximadamente 500 y 2000, mientras que los valores del tiempo de conversación están entre 0 y 20.

- Hay agrupamientos: Se pueden apreciar algunos grupos de puntos en ciertas zonas del gráfico, lo que podría indicar que hay modelos de teléfonos con características similares en cuanto a batería y tiempo de conversación.

- No hay valores atípicos evidentes: A simple vista, no se observan puntos muy alejados del resto, lo que sugiere que los datos son relativamente consistentes.

Conclusión

Con base a este gráfico, podemos concluir que, en este conjunto de datos, no existe una relación lineal clara entre la potencia de la batería y el tiempo de conversación de los teléfonos móviles. Otros factores, como la eficiencia energética del procesador, la capacidad de la pantalla y las aplicaciones en uso, podrían influir más en el tiempo de conversación.

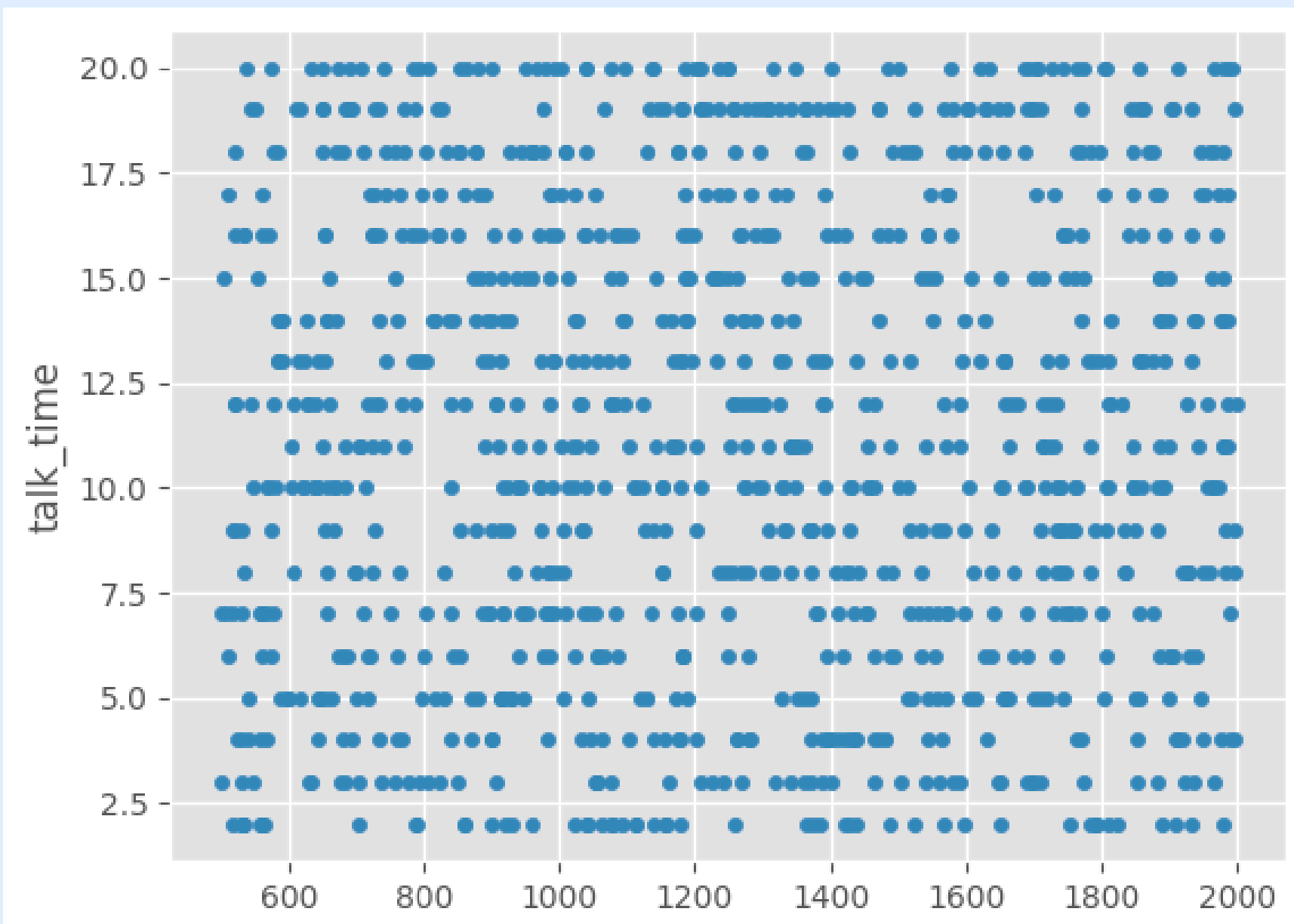
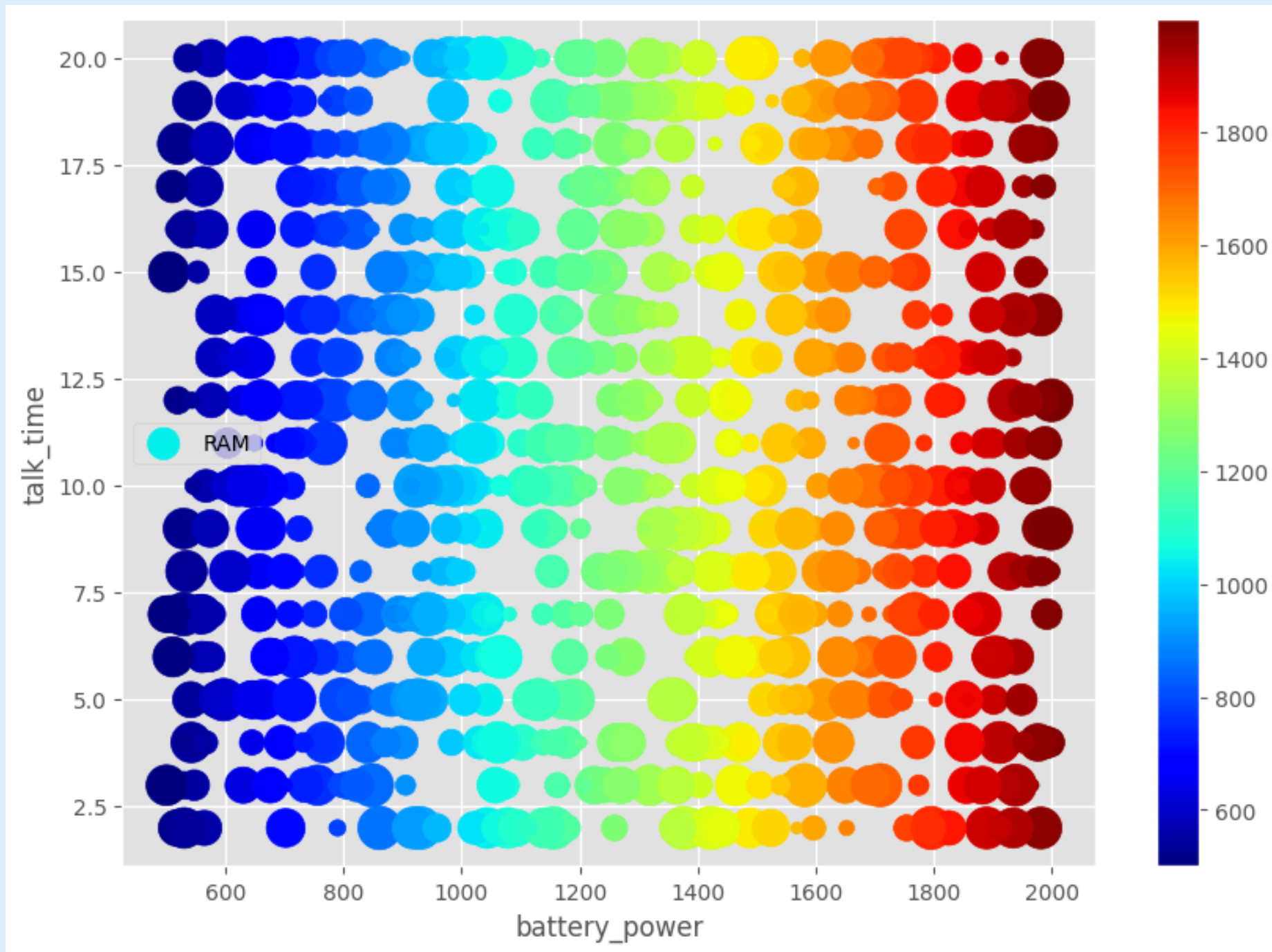


GRÁFICO DE DISPERSIÓN DE BURBUJAS



Este gráfico de dispersión de burbujas nos muestra la relación entre tres variables de un conjunto de datos sobre teléfonos móviles:

- Eje X (battery_power): Representa la potencia de la batería de cada teléfono.
- Eje Y (talk_time): Representa el tiempo de conversación de cada teléfono.
- Tamaño de las burbujas (RAM): El tamaño de cada burbuja representa la cantidad de RAM (memoria RAM) de cada teléfono. El color de las burbujas varía según la potencia de la batería, utilizando una escala de colores que va del azul (menor potencia) al rojo (mayor potencia).

Este gráfico de dispersión de burbujas es una herramienta muy útil para explorar datos multivariados y visualizar relaciones entre diferentes variables. Al combinar información sobre la potencia de la batería, el tiempo de conversación y la RAM, podemos obtener una mejor comprensión de las características de los teléfonos móviles.

MATRIZ DE CORRELACIÓN

Interpretación

- Relaciones Esperadas: Las correlaciones fuertes entre características como la altura y el ancho de la pantalla, o entre la capacidad de procesamiento y la memoria interna, son lógicas y coinciden con lo que esperaríamos.

- Relaciones Interesantes: Las correlaciones entre características como las cámaras frontales y la conectividad 4G, o entre la RAM y el tamaño de la pantalla, pueden ser útiles para entender las tendencias en el diseño de los teléfonos móviles.

- Ausencia de Correlación: La ausencia de correlación entre muchas variables indica que no hay una relación lineal directa entre ellas. Esto podría deberse a que otros factores no considerados en el análisis influyen en estas variables, o simplemente a que no existe una relación lineal.

¿Qué Podemos Hacer con Esta Información?

- Selección de Características: Al identificar las variables altamente correlacionadas, podemos seleccionar un subconjunto de variables para construir modelos de machine learning, evitando problemas de multicolinealidad.

- Generación de Hipótesis: Las correlaciones pueden sugerir relaciones causales que pueden ser investigadas a fondo. Por ejemplo, ¿por qué los teléfonos con muchas cámaras frontales tienden a tener conectividad 4G?

- Visualizaciones: Podemos crear visualizaciones más detalladas para explorar las relaciones entre las variables, como gráficos de dispersión o diagramas de caja.

En Resumen

La matriz de correlación es una herramienta útil para explorar las relaciones entre las variables de un conjunto de datos. Al analizarla, podemos obtener información valiosa sobre las características de los teléfonos móviles y generar hipótesis para futuras investigaciones.

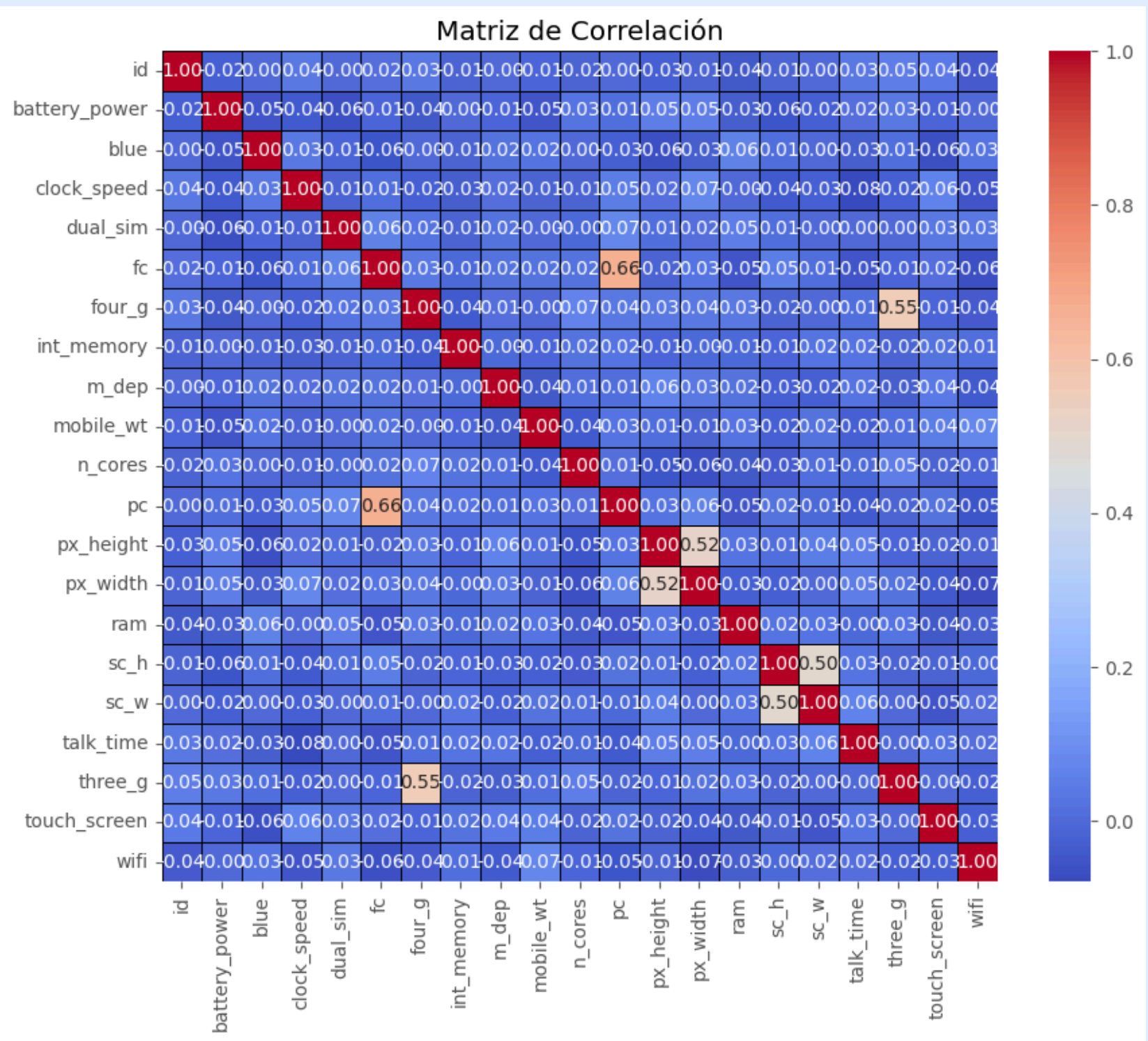


GRÁFICO DE DISPERSIÓN DE MATRIZ

Relaciones entre características:

- Batería y RAM: No hay una relación clara entre la batería y la RAM. Los puntos están dispersos por todo el gráfico, lo que sugiere que tener una batería grande no necesariamente significa tener mucha RAM, y viceversa.

- Batería y tiempo de conversación: Tampoco hay una relación lineal clara. Algunos teléfonos con baterías grandes tienen un tiempo de conversación largo, pero otros no.

- Peso y otras características: El peso del teléfono parece estar relacionado con la batería. Los teléfonos más pesados tienden a tener baterías más grandes. Sin embargo, no hay una relación clara con la RAM o el tiempo de conversación.

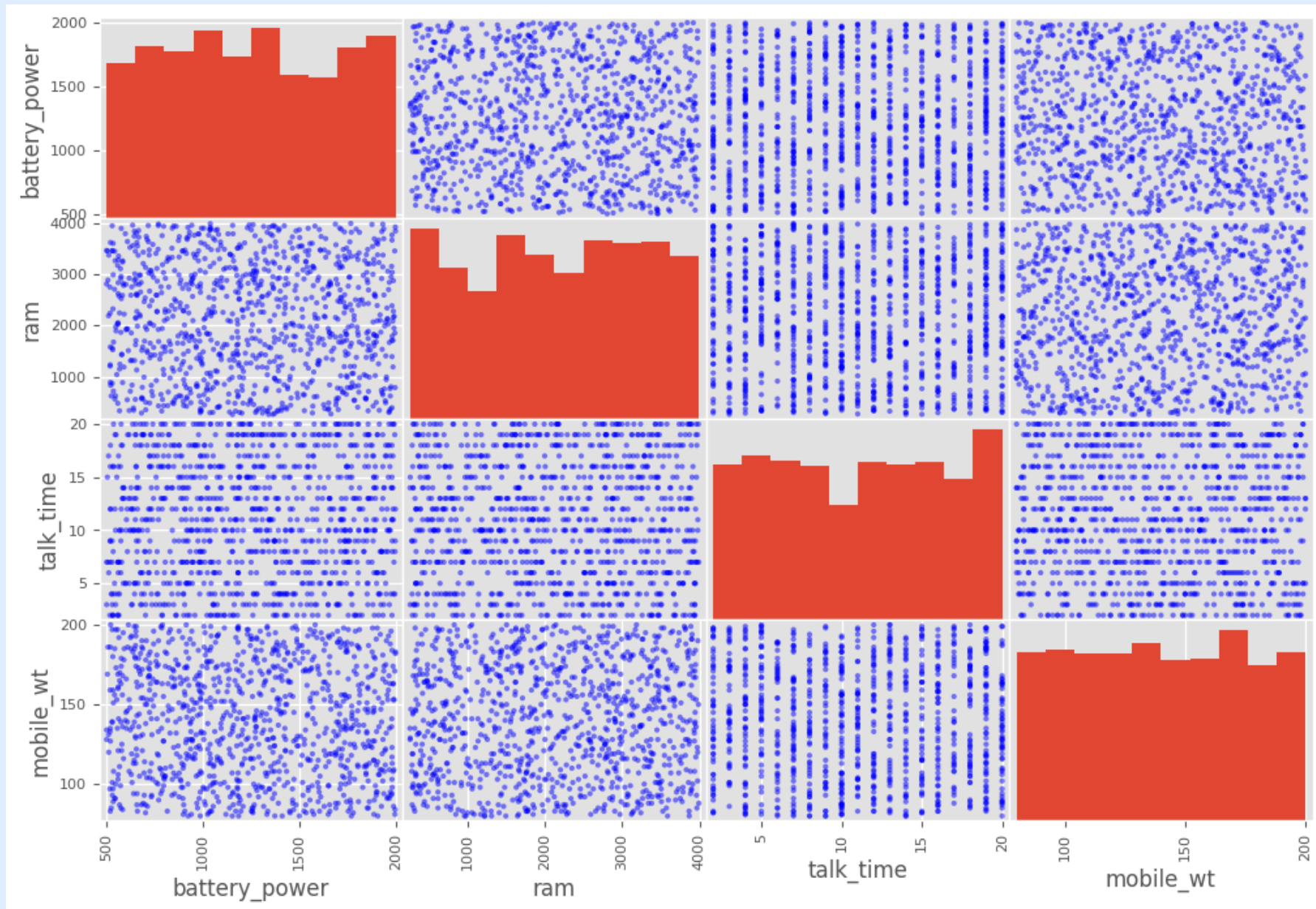
Distribución de los datos:

Los histogramas muestran que la mayoría de los teléfonos tienen una batería de tamaño medio, una cantidad moderada de RAM y un peso promedio.

Este gráfico de dispersión de matriz nos proporciona una visión general de las relaciones entre las diferentes características de los teléfonos móviles en tu conjunto de datos. Sin embargo, es importante recordar que:

- Correlación no implica causalidad: El hecho de que dos variables estén relacionadas no significa necesariamente que una cause la otra. Podría haber otras variables que influyan en ambas.

- Otras variables: Este análisis solo considera cuatro características. Podría haber otras variables importantes que no se han incluido y que podrían afectar las relaciones observadas.



DISTRIBUCIÓN DE PUNTOS: POTENCIA DE BATERÍA VS. TIEMPO DE CONVERSACIÓN

Análisis:
Tendencia General:

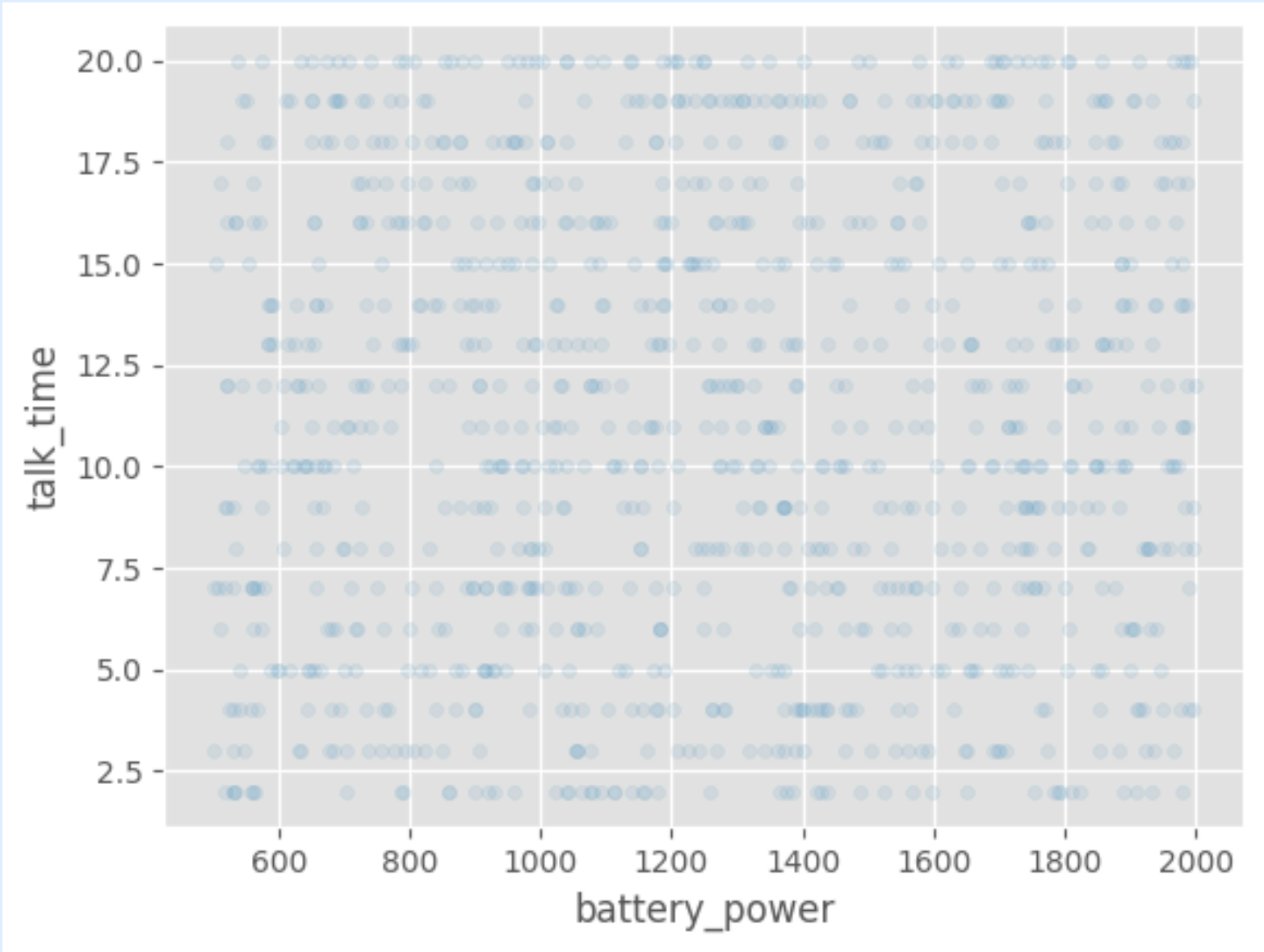
Se observa una tendencia general a que los dispositivos con mayor potencia de batería tengan también un mayor tiempo de conversación. Sin embargo, esta relación no es necesariamente lineal y puede haver mucha variabilidad.

Densidad de Puntos:
Áreas con una alta densidad de puntos indican que hay un gran número de dispositivos con características similares en cuanto a batería y tiempo de conversación.

Outliers:
Los puntos que se encuentran alejados de la nube princital de datos podrían representar dispositivos con características inusuales, como una batería muy potente pero un tiempo de conversación bajo, o viceversa.

Correlación:
La distribución de los puntos puede sugerir si existe una correlación positiva o negativa entre la potencia de la batería y el tiempo de conversación. Una alineación general de los puntos en una dirección indica una correlación, mientras que una dispersión aleatoria sugiere una correlación débil o inexistente.

Limitaciones de la Visualización:
La transparencia de los puntos (alpha=0.1) puede dificultar la identificación de patrones claros en áreas con una alta densidad de datos, ya que los puntos se superponen.



LÍNEA DE TENDENCIA

Análisis de la Gráfica:

Línea de tendencia:

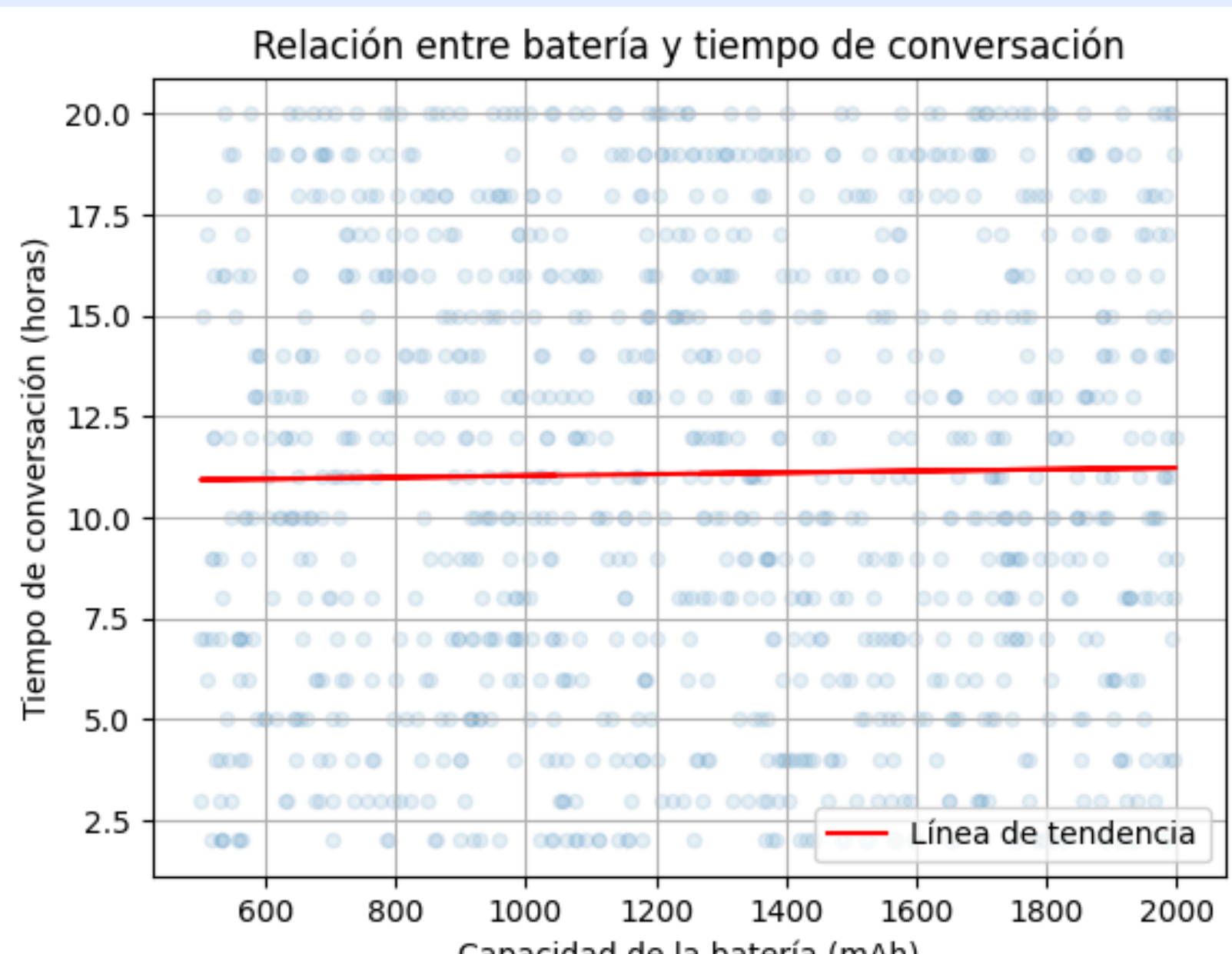
La línea de tendencia roja muestra una ligera tendencia positiva, lo que sugiere que a medida que aumenta la capacidad de la batería, también tiende a aumentar ligeramente el tiempo de conversación. Sin embargo, esta tendencia es débil y no es muy pronunciada.

Valores atípicos:

Hay algunos puntos que se encuentran bastante alejados de la tendencia general. Estos podrían corresponder a dispositivos con características inusuales o errores en los datos.

Conclusiones:

- La capacidad de la batería no es el único factor determinante del tiempo de conversación: Otros factores como la eficiencia del procesador, el tamaño de la pantalla, las aplicaciones en ejecución y las configuraciones del dispositivo también influyen significativamente en el tiempo de conversación.
- La relación entre la capacidad de la batería y el tiempo de conversación no es lineal**: La línea de tendencia muestra una relación aproximadamente lineal, pero la dispersión de los datos sugiere que la relación real puede ser más compleja.
- Hay una gran variabilidad en el tiempo de conversación para una misma capacidad de batería**: Esto se debe a la influencia de otros factores que no están siendo considerados en este análisis.



DISTRIBUCIÓN DE LA CANTIDAD DE RAM

Análisis de los histogramas

1. Histograma de RAM (MB):

- Forma: La distribución de la RAM parece estar sesgada hacia la derecha. Esto significa que la mayoría de los dispositivos tienen una cantidad relativamente pequeña de RAM, mientras que hay pocos dispositivos con mucha RAM.
- Concentración: La mayor parte de los dispositivos se concentran en un rango específico de valores de RAM, con una disminución gradual en la frecuencia a medida que aumenta la cantidad de RAM.

2. Histograma de Log de RAM (MB):

- Forma: Al aplicar el logaritmo, la distribución se vuelve más simétrica y se aproxima a una distribución normal. Esto facilita la interpretación de los datos.
- Patrones: Podemos observar con más claridad los diferentes grupos de dispositivos según su cantidad de RAM. Por ejemplo, se pueden identificar un grupo de dispositivos con poca RAM, otro con una cantidad intermedia y un grupo más pequeño con una cantidad muy alta de RAM.

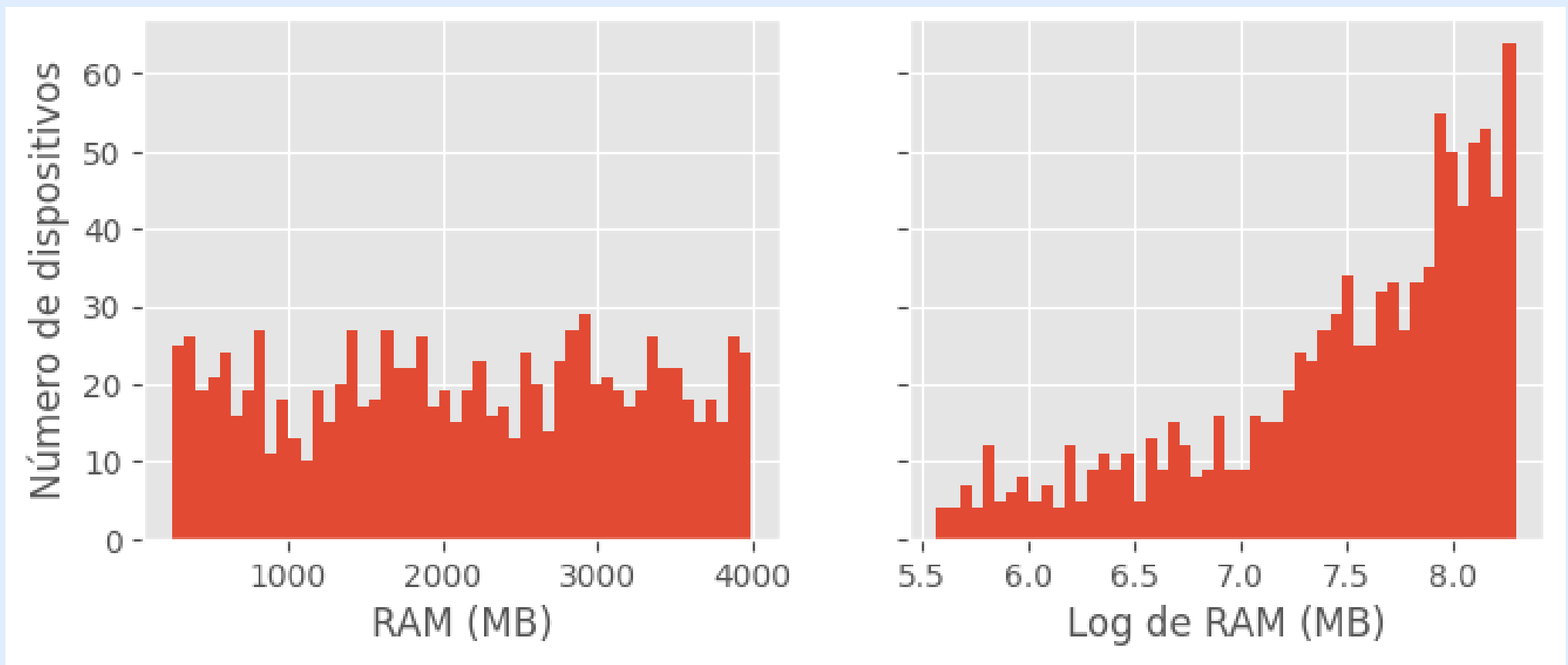


GRÁFICO COMBINADO: HISTOGRAMA DE RAM Y SIMILITUD RBF

Interpretación del Gráfico

1. Histograma: La distribución de la RAM muestra que la mayoría de los dispositivos tienen una cantidad de RAM entre 1000 y 3000 MB.

2. Similitud RBF(Función base radial):

- Gamma = 0.1: La línea azul sólida muestra una similitud alta solo para los valores de RAM muy cercanos a 2000 MB. Fuera de este rango, la similitud cae rápidamente a cero.

- Gamma = 0.03: La línea azul punteada muestra una similitud más amplia. Incluso los valores de RAM que están bastante lejos de 2000 MB tienen una similitud no nula, aunque menor.

¿Qué nos dice esto?

- Influencia de gamma:

- Un gamma alto mide la similitud de manera muy local, identificando valores que están muy cerca del valor de referencia.

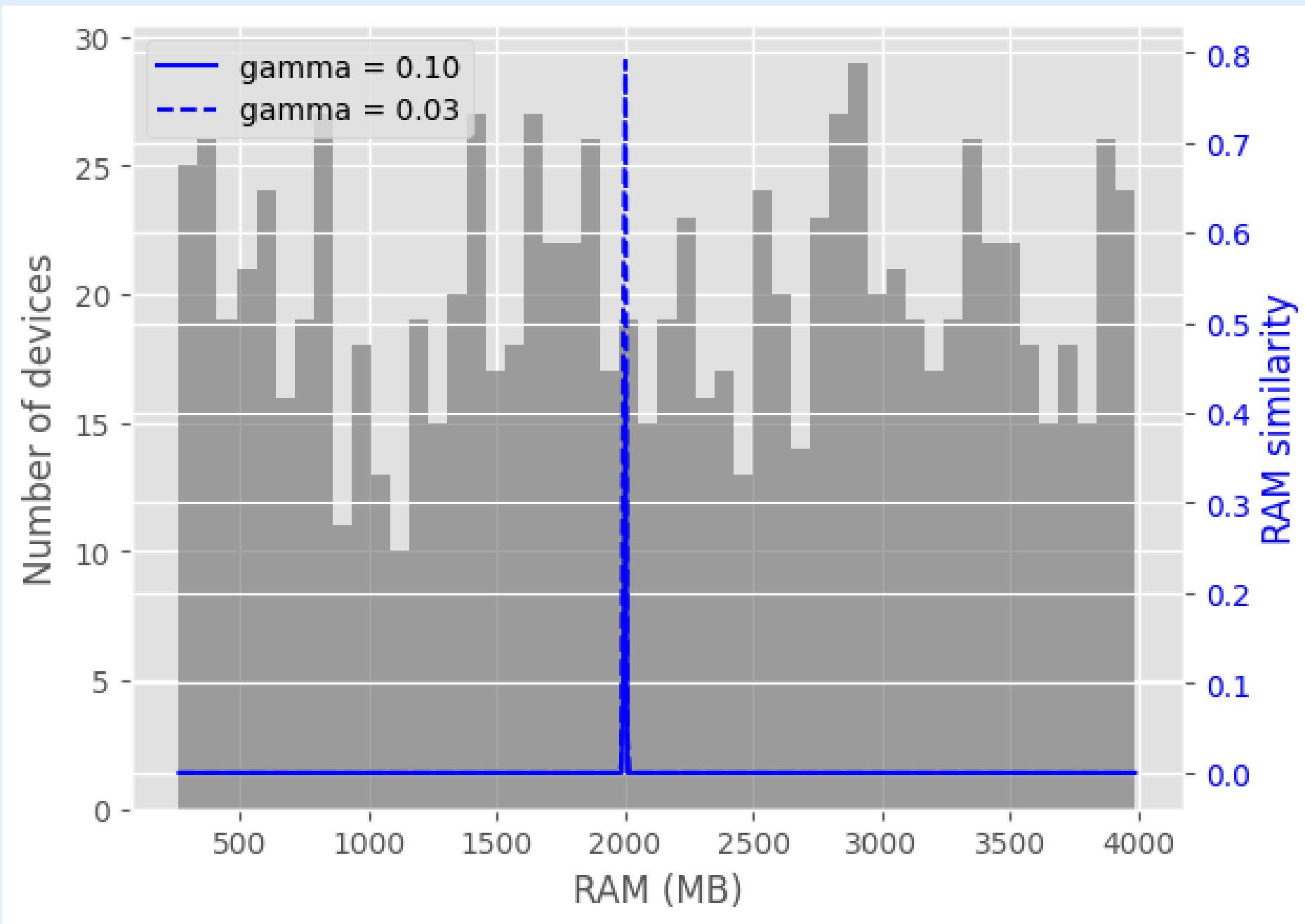
- Un gamma bajo mide la similitud en un rango más amplio, proporcionando una visión más general.

- Distribución de la RAM: La combinación del histograma y las curvas de similitud nos permite ver cómo se distribuye la RAM en relación con el valor de referencia de 2000 MB. Según el valor de gamma, podemos identificar qué rango de valores de RAM es más similar a 2000 MB.

Aplicaciones:

- Este análisis es útil para tareas como clustering, clasificación o detección de anomalías.

- Por ejemplo, podemos identificar grupos de dispositivos con características similares o detectar dispositivos con una cantidad de RAM inusualmente alta o baja.



GRACIAS

