

Navegando el Ecosistema de Apps: EDA en Google Play



Tabla de contenido

1. Contexto	2
2. Hipótesis	3
3. Análisis exploratorio de datos.....	3
3.1 Análisis Univariante	3
Variables numéricas	3
Variables categóricas	8
3.2 Análisis Bivariante	10
3.3 Análisis Multivariante	15
4. Conclusiones.....	16
5. Recomendaciones	17

1. Contexto

Descripción del proyecto

El presente análisis exploratorio de datos (EDA), tiene como objetivo estudiar las aplicaciones de Google Play Store, identificando patrones, tendencias y factores que influyen en las calificaciones y el éxito de las aplicaciones, medido principalmente por el número de instalaciones y la valoración de los usuarios.

Objetivo

El principal objetivo es responder a las siguientes preguntas:

- ¿Cuáles son las categorías con apps mejor calificadas?
- ¿Existe alguna relación entre las calificaciones y el número de descargas/instalaciones?
- ¿Las apps gratuitas tienen mejores calificaciones que las de pago?
- ¿Cuál es la correlación entre el número de instalaciones y el precio?
- ¿Las apps que más duran en el mercado tienen mejores calificaciones?
- ¿Qué factores influyen más en las calificaciones?
- ¿Qué factores, en general, influyen más en las apps más exitosas?

Datasets utilizados

- google play store.csv: Contiene información sobre aplicaciones, como calificaciones, número de instalaciones, categoría de las apps, precio, tipo (gratuito/de pago), entre otros.
- google play store user reviews.csv: Contiene reseñas de usuarios, incluyendo el sentimiento (positivo, negativo, neutral), polaridad y subjetividad. Este dataset sin embargo no se ha empleado para el análisis, únicamente se ha llevado a cabo la limpieza y una primera observación. Se trata de una dataset con muchas más entradas que el primero, con lo cual sería muy complicado fusionarlos.

2. Hipótesis

Para guiar el análisis, se han seguido la hipótesis que se ven a continuación:

- Las apps de ciertas categorías están mejor valoradas.
- Existe una relación positiva entre las calificaciones ('Rating'), y el número de instalaciones ('Installs').
- Las apps gratuitas tienen en promedio mejores calificaciones que las de pago.
- Las apps con precios elevados tienen menos instalaciones.
- Las apps más antiguas tienden a tener mejores calificaciones.
- Las apps más exitosas (con mayor número de instalaciones), tienen mejores calificaciones y suelen ser gratuitas.

A continuación se muestra el análisis exploratorio de datos, mostrando únicamente aquellas gráficas más significativas para las conclusiones del final de este documento.

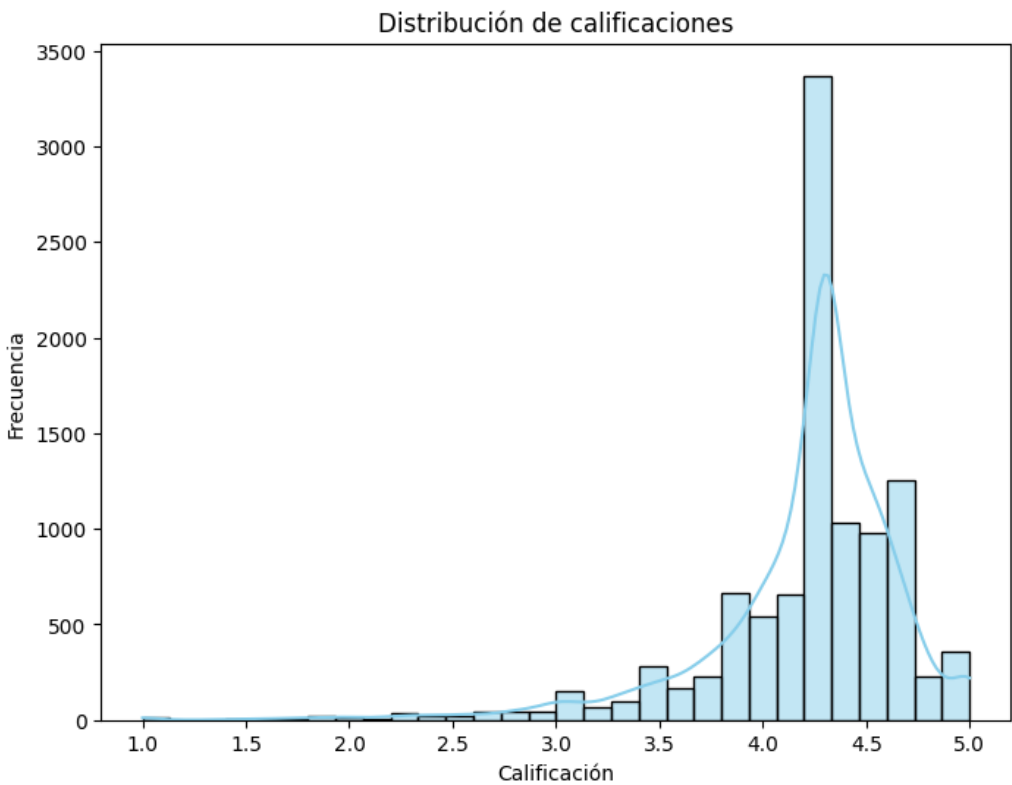
3. Análisis exploratorio de datos

3.1 Análisis Univariante

Variables numéricas

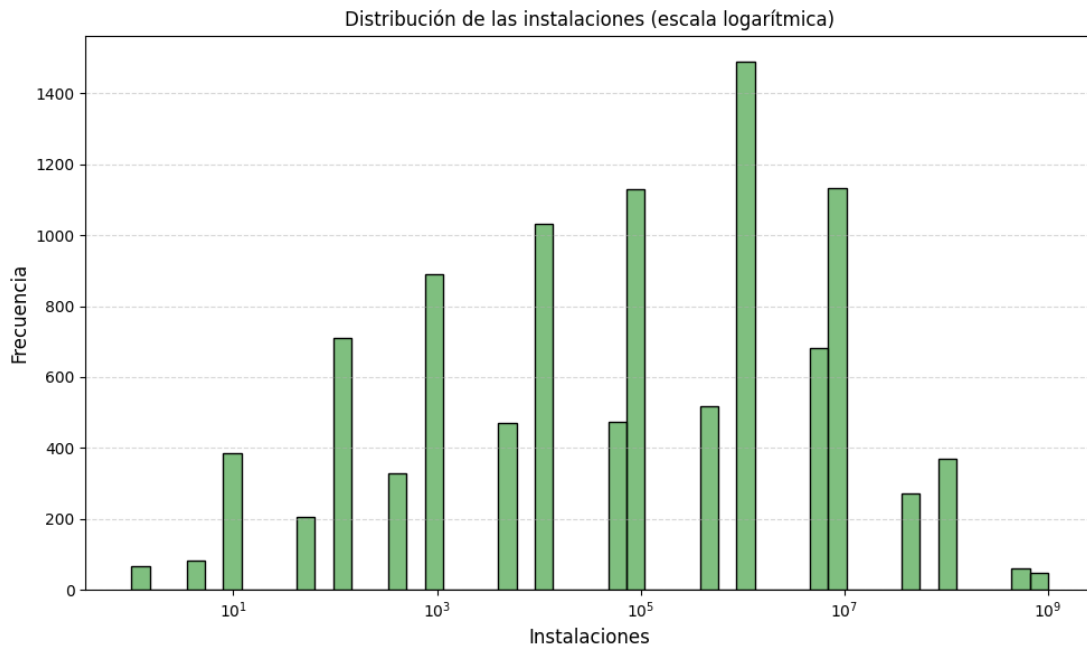
Rating (Calificaciones):

- Distribución: La mayoría de las apps tienen calificaciones entre 4.0 y 4.5
- Outliers: Se han identificado apps con calificaciones bajas (por debajo de 2.5).
- Gráficos utilizados: Histograma y boxplot.



Installs (Número de instalaciones):

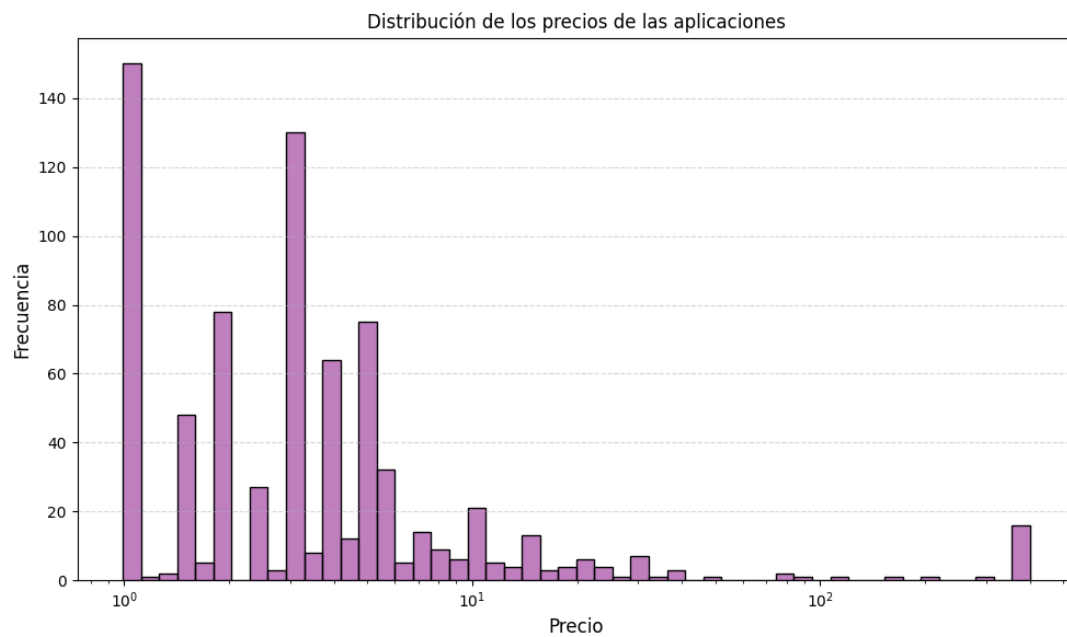
- La distribución está considerablemente sesgada hacia la derecha, con un pequeño número de apps muy populares.
- Se ha aplicado la transformación logarítmica para visualizar mejor los datos.
- Gráficos utilizados: Histograma con escala logarítmica.



Price (Precio):

- La mayoría de las apps son gratuitas (precio 0 y tipo 'free').
- Existe un pequeño número de apps con precios muy altos (outliers).

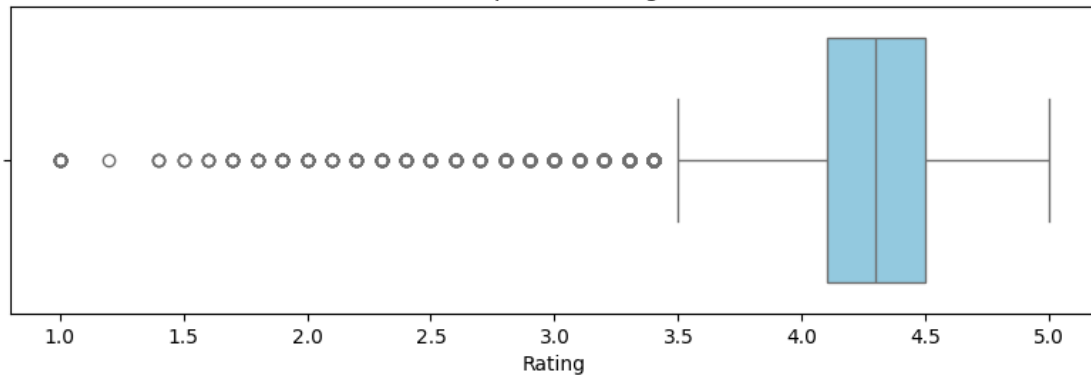
- Gráficos empleados: Histograma y boxplot.



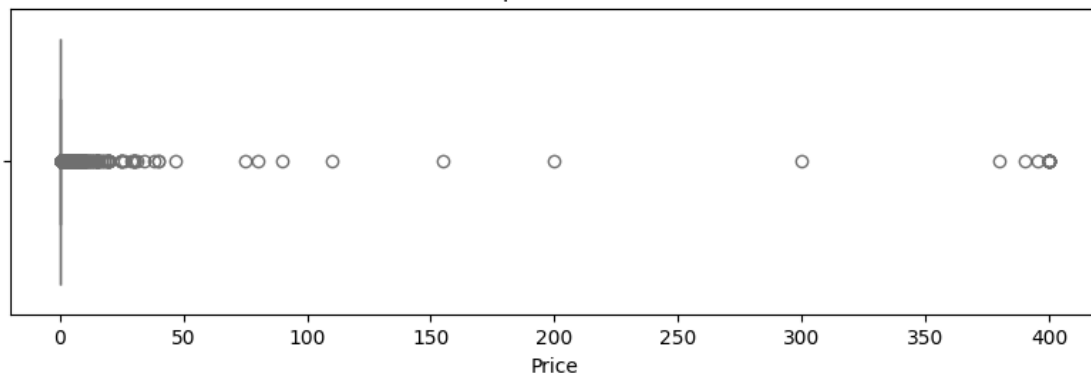
App Age (Antigüedad de la app en años).

- La mayoría de las apps tienen menos de 8 años de antigüedad
- Gráficos utilizados: Histograma

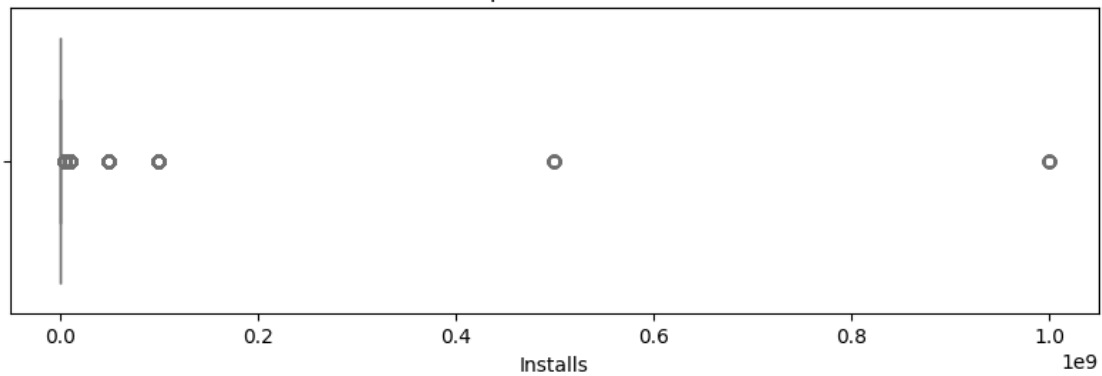
Boxplot de Rating



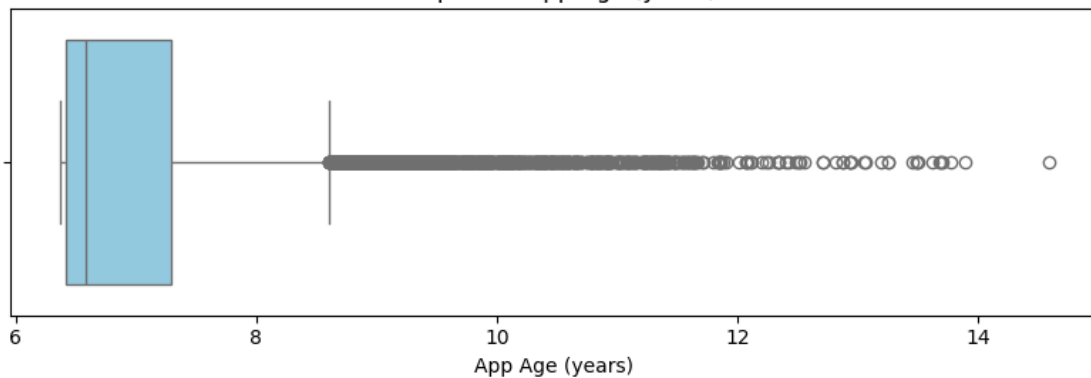
Boxplot de Price



Boxplot de Installs



Boxplot de App Age (years)



Variables categóricas

Categoría (Categoría de la app):

- Las categorías más frecuentes son "FAMILY" y "GAME".
- Gráfico utilizado: Gráfico de barras.

Type (Tipo: Free or Paid):

- La mayoría de las apps son 'Free', es decir, gratuitas.
- Gráficos utilizados: Gráfico de barras.

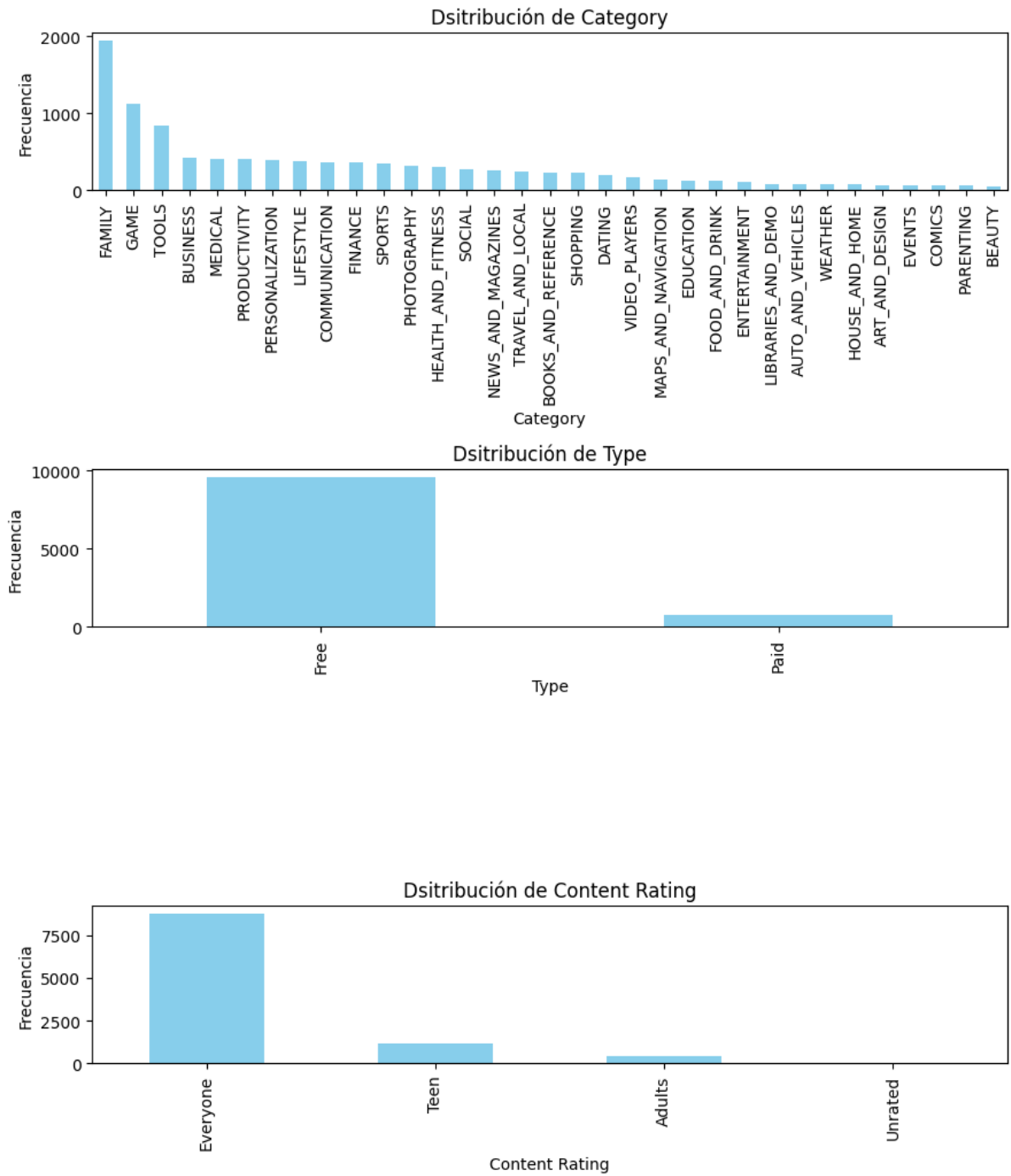
Content Rating (Clasificación por edad):

- La mayoría de las apps son gratuitas (precio 0 y tipo 'free').
- Existe un pequeño número de apps con precios muy altos (outliers).
- Gráficos empleados: Histograma y boxplot.

App Age (Antigüedad de la app en años)

- La mayoría de las apps están calificadas como 'Everyone', es decir, para todos los públicos.
- Gráficos utilizados: Gráfico de barras.

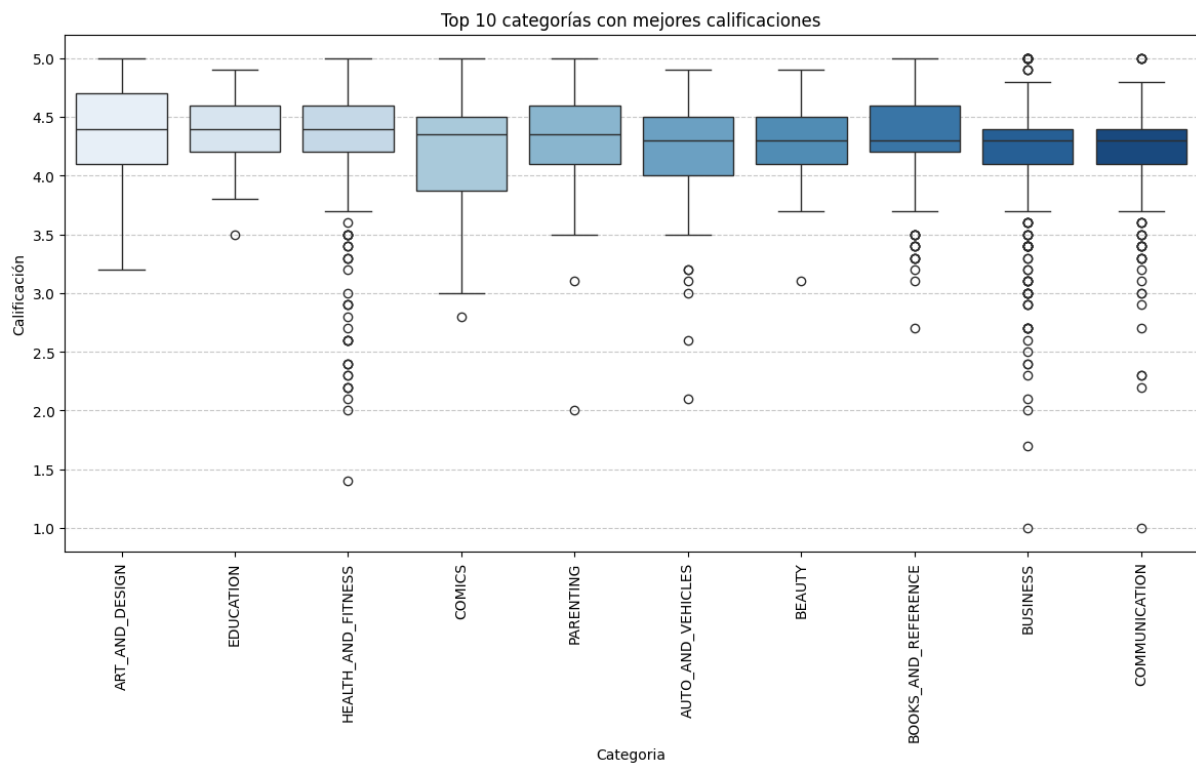
Frecuencia total de las variables categóricas clave



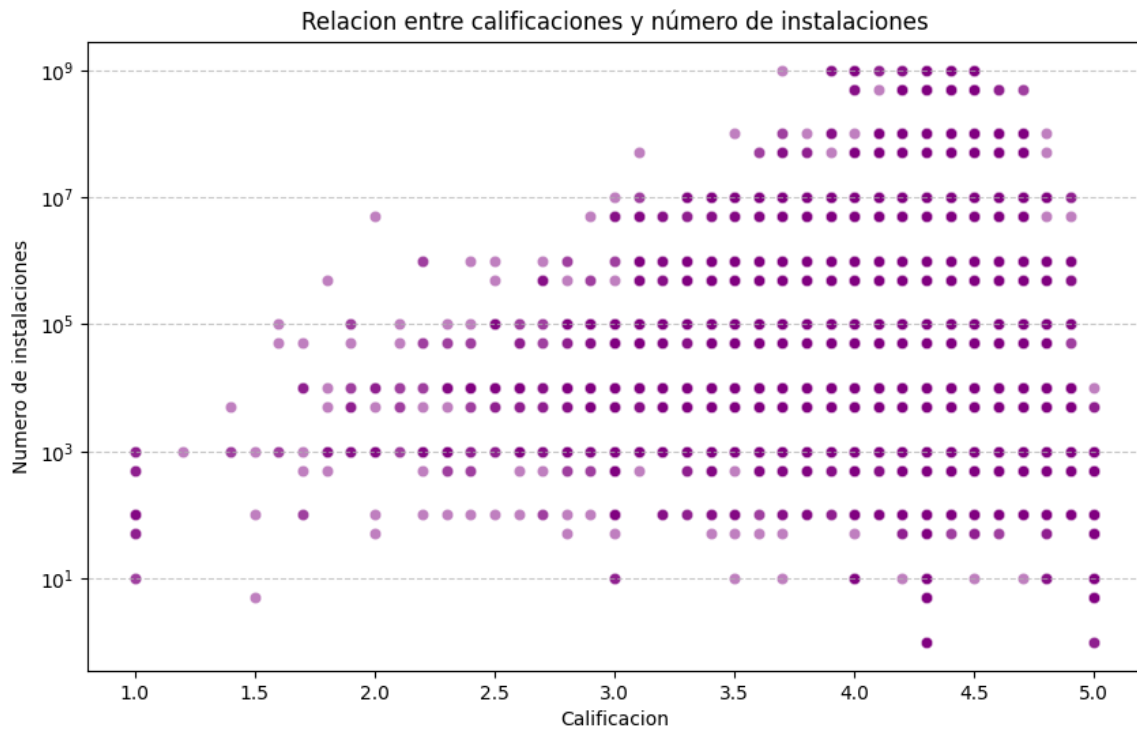
3.2 Análisis Bivariante

Respuesta a algunas de las preguntas del EDA:

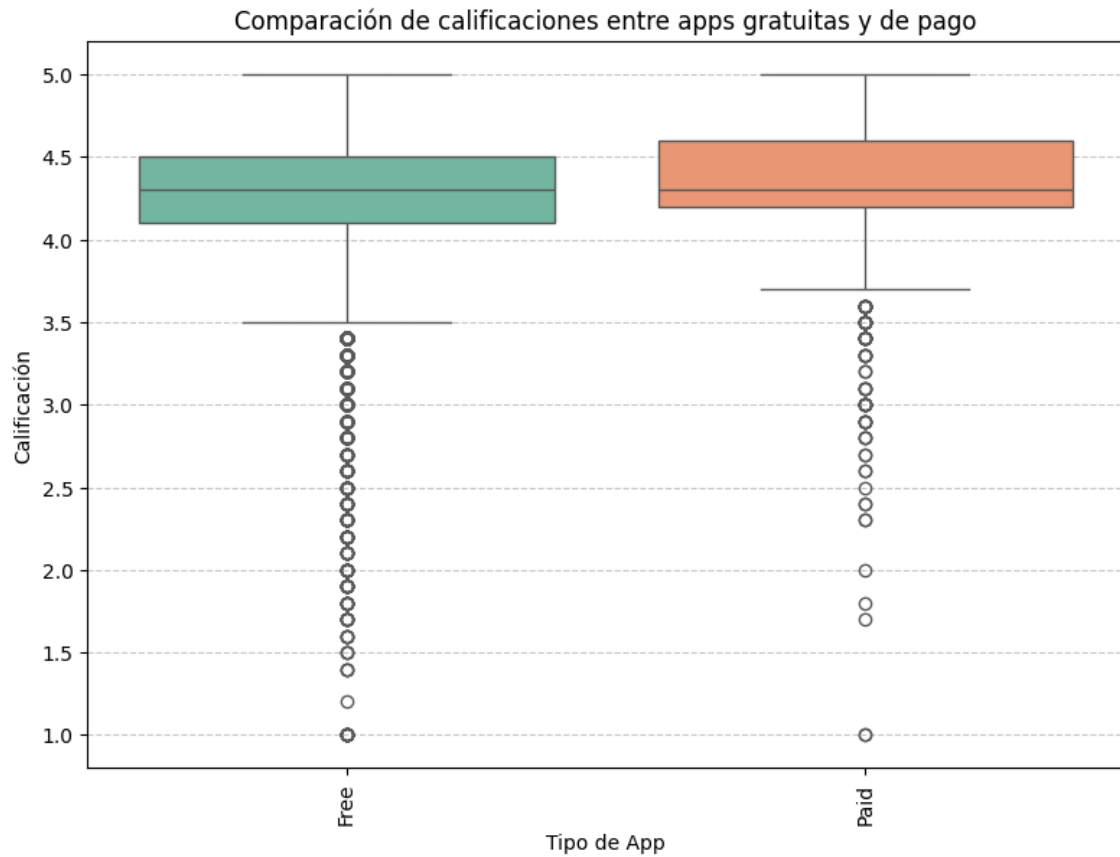
- ¿Cuáles son las categorías con apps mejor calificadas?
 - Gráfico utilizado: Boxplot de Rating y Category
 - Conclusión: Las categorías con calificaciones más altas en promedio son: ART AND DESIGN, EDUCATION, HEALTH_AND_FITNESS.
 - Categorías como Comics y Parenting, aunque con menos aplicaciones, muestran valoraciones muy altas.
 - Esto podría sugerir que, en esos nichos, las apps tienden a tener buenas reseñas de su audiencia específica.



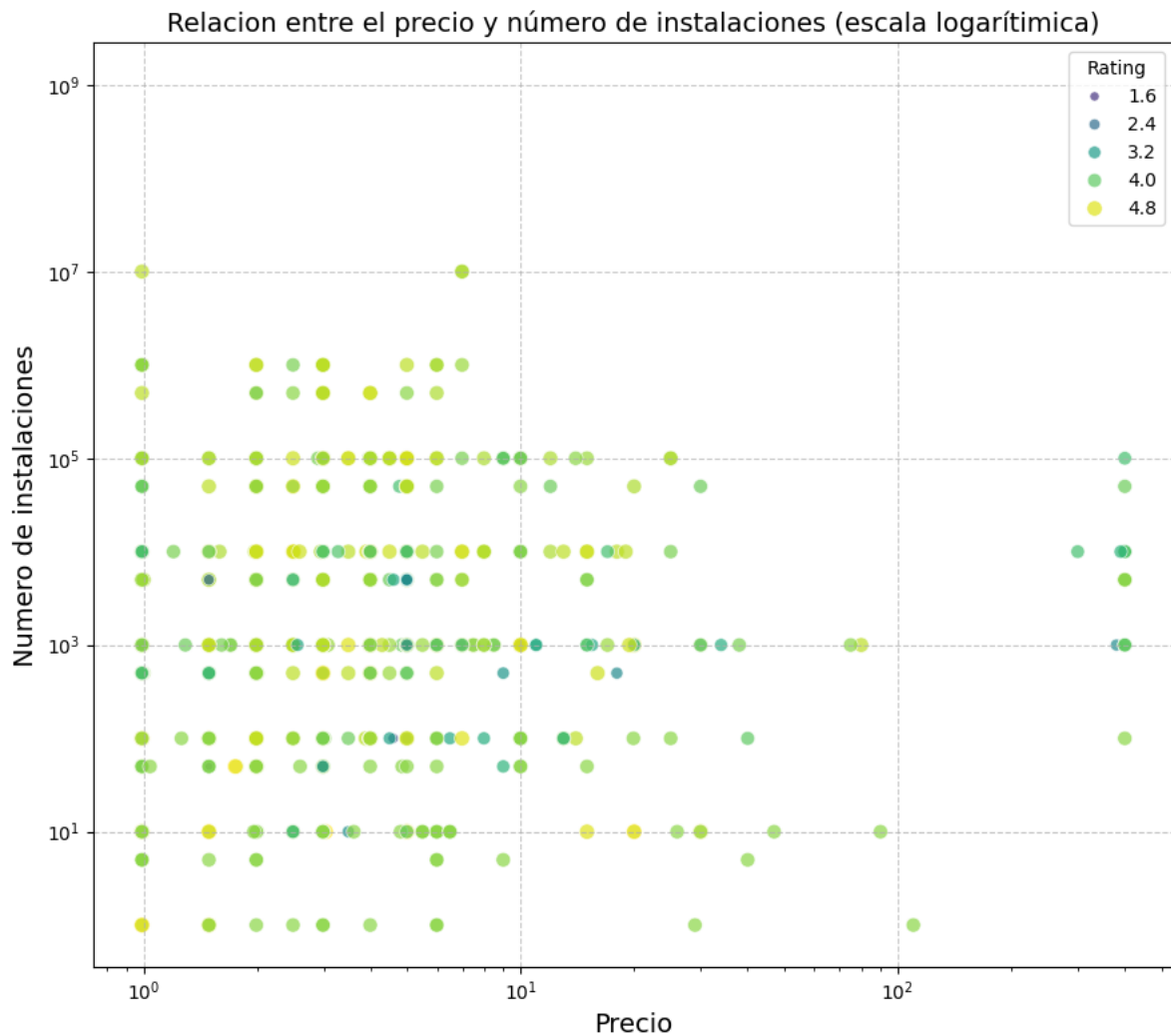
- ¿Existe una relación entre las calificaciones y el número de descargas?
 - Gráfico utilizado: Scatterplot de Rating VS Installs.
 - Existe una correlación positiva débil entre las calificaciones y el número de instalaciones, pero se trata de una correlación de alrededor de 0.044, es decir, prácticamente despreciable. Concluiría, por tanto, que no hay apenas relación.



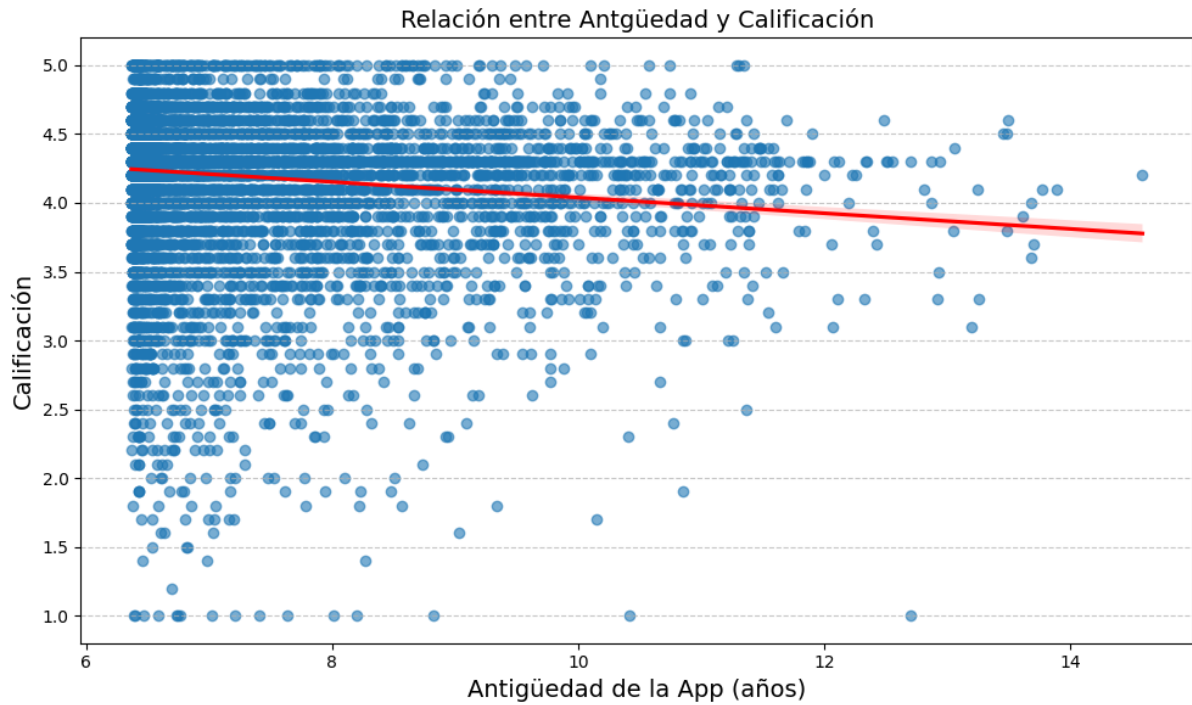
- ¿Las apps gratuitas tienen mejores calificaciones que las de pago?
 - Gráfico utilizado: Boxplot de Rating por Type
 - La diferencia entre ambas medianas es prácticamente imperceptible, con lo cual, no hay una influencia notable del tipo de aplicación, gratuita o de pago en la valoración de los usuarios.



- ¿Cuál es la relación entre precio e instalaciones?
 - Gráfico utilizado: Scatterplot con escala logarítmica
 - Las apps gratuitas muestran más instalaciones en promedio que las apps de pago.



- ¿Las apps más antiguas tienen mejores calificaciones?
 - Gráfico utilizado: Scatterplot de App Age VS Rating
 - No se ha observado una relación significativa entre la antigüedad y las calificaciones. Existe de hecho una correlación débil

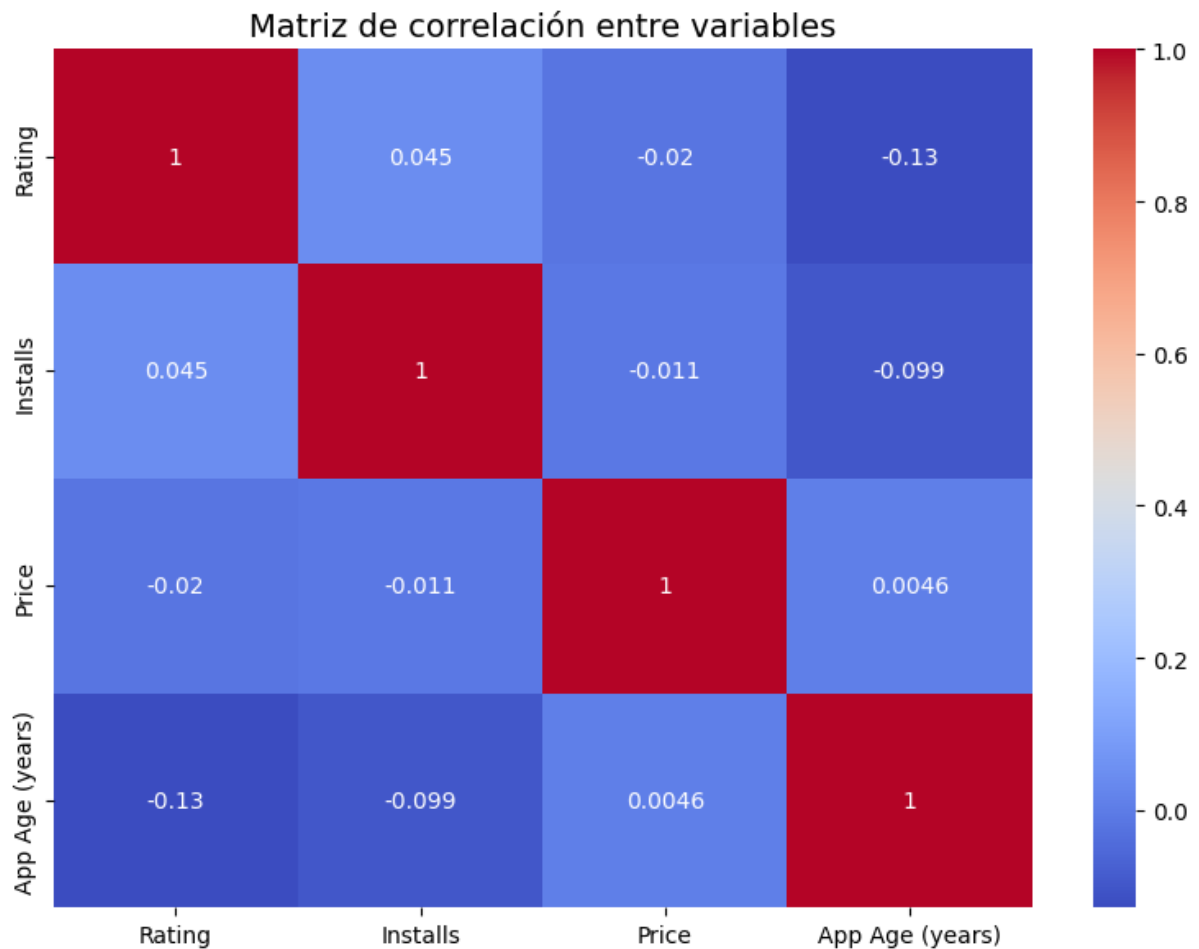


negativa.

3.3 Análisis Multivariante

3.3.1 Factores que influyen en las calificaciones

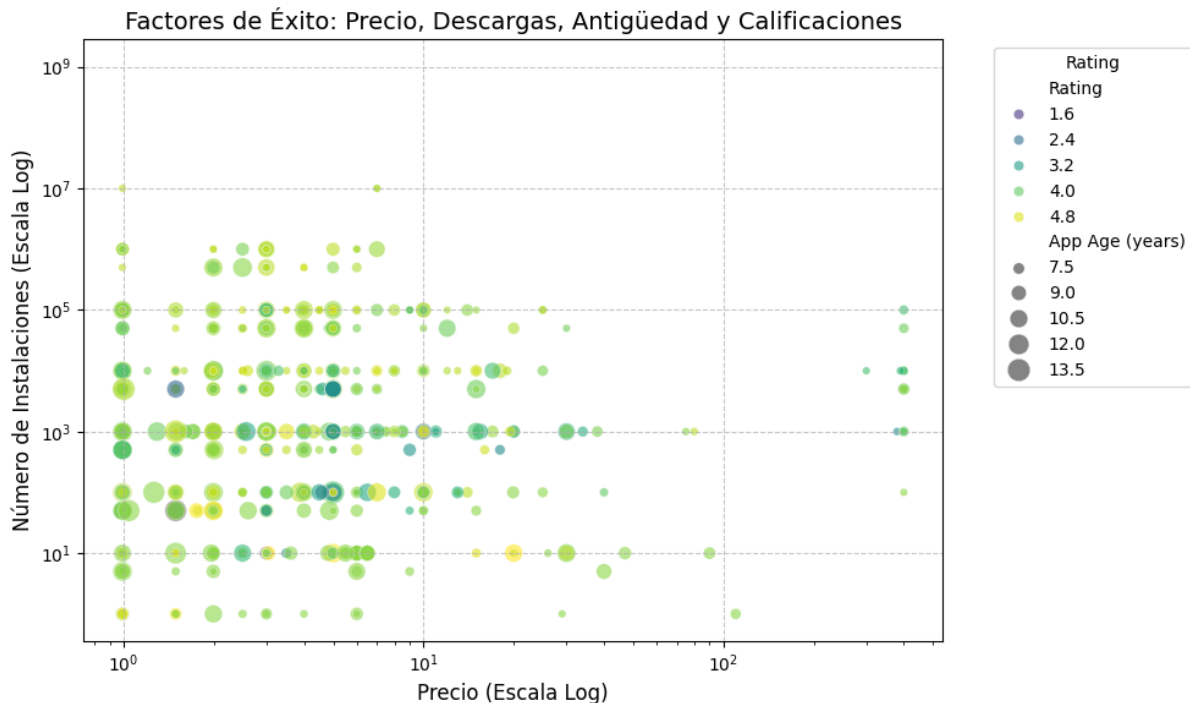
- Gráfico utilizado: Matriz de correlaciones
- De dicha matriz, podríamos destacar que el número de descargas (Installs), tiene varias relaciones de las que podríamos sacar conclusiones.



- Dada su relación positiva débil con Rating, se observa que las apps con más descargas tienen mejor valoración, pero no la consideraría una conclusión dada la bajísima relación (0.045).
- Por otro lado, con Price se da una relación negativa, aunque también muy débil, lo cual sugiere que, a mayor precio, menores descargas, en coherencia con el gráfico anterior (precio vs. instalaciones).
- Por último, la antigüedad de la app presenta una influencia mínima en las calificaciones.

3.3.2 Factores que influyen en las apps más exitosas

- Gráfico utilizado: Gráfico de burbujas
- Las apps gratuitas, con buenas calificaciones, tienden a tener más instalaciones



4. Conclusiones

4.1 Las categorías con más aplicaciones (Family o Game), no necesariamente lideran en calificaciones promedio. Esto destaca la importancia de enfocarse en nichos de alta satisfacción para diferenciarse.

4.2 Existe una relación positiva entre calificaciones y descargas, pero es débil.

4.3 Las apps gratuitas reciben mejores calificaciones en promedio.

4.4 El precio está negativamente correlacionado con el número de instalaciones.

4.5 La antigüedad de la app no influye significativamente en las calificaciones.

5. Recomendaciones

5.1 Promover apps gratuitas o modelos 'freemium' donde las funciones básicas sean gratuitas y las avanzadas tengan un coste asociado para atraer una mayor base de usuarios iniciales.

5.2 Priorizar el desarrollo de apps en categorías con mejores calificaciones.

5.3 Optimizar la estrategia de precios: Realizar un análisis detallado de la disposición a pagar de los usuarios, utilizando datos históricos de precios y ventas. Implementar además estrategias como precios dinámicos, en los que los costes de las apps puedan variar según la demanda, las temporadas o haciendo promociones específicas.