

Analisando aplicativos no Google Play Store Apps.

Introdução

Neste projeto de análise, utilizei um *dataset* referente aos aplicativos disponíveis na Google Play Store. Os dados foram obtidos de um repositório público do Kaggle, do autor Lavanya Gupta, intitulado "Google Play Store Apps" (Fevereiro de 2019). [Fonte: <https://www.kaggle.com/lava18/google-play-store-apps>]. A análise foi realizada em Python, utilizando as bibliotecas Pandas para manipulação dos dados e Matplotlib para visualizações.

Tratamento e Limpeza dos Dados

O dataset original é composto por 10841 linhas e 13 colunas, sendo elas: 'App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type', 'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver', 'Android Ver'.

Como parte da preparação inicial, colunas consideradas não essenciais para os objetivos analíticos deste estudo foram removidas para simplificar o conjunto de dados. As colunas excluídas foram: 'Genres', 'Last Updated', 'Current Ver' e 'Android Ver'.

Em seguida, iniciei o tratamento de valores ausentes (*missing values*) e inconsistências nas colunas selecionadas para análise:

1. **Coluna 'Rating' (Avaliação):** Esta coluna apresentava 1.474 valores nulos. Durante a inspeção, foi identificado um *outlier* significativo, com o valor 19 (em uma escala que deveria variar de 1 a 5). Considerando que este é um registro único e que distorce completamente métricas como a média, a decisão foi excluir a linha correspondente. Após sua remoção, a distribuição dos dados mostrou-se simétrica. Portanto, para preencher os valores nulos restantes, foi utilizado o valor da **média** aritmética (arredondada para duas casas decimais) das avaliações, uma abordagem adequada para distribuições simétricas.

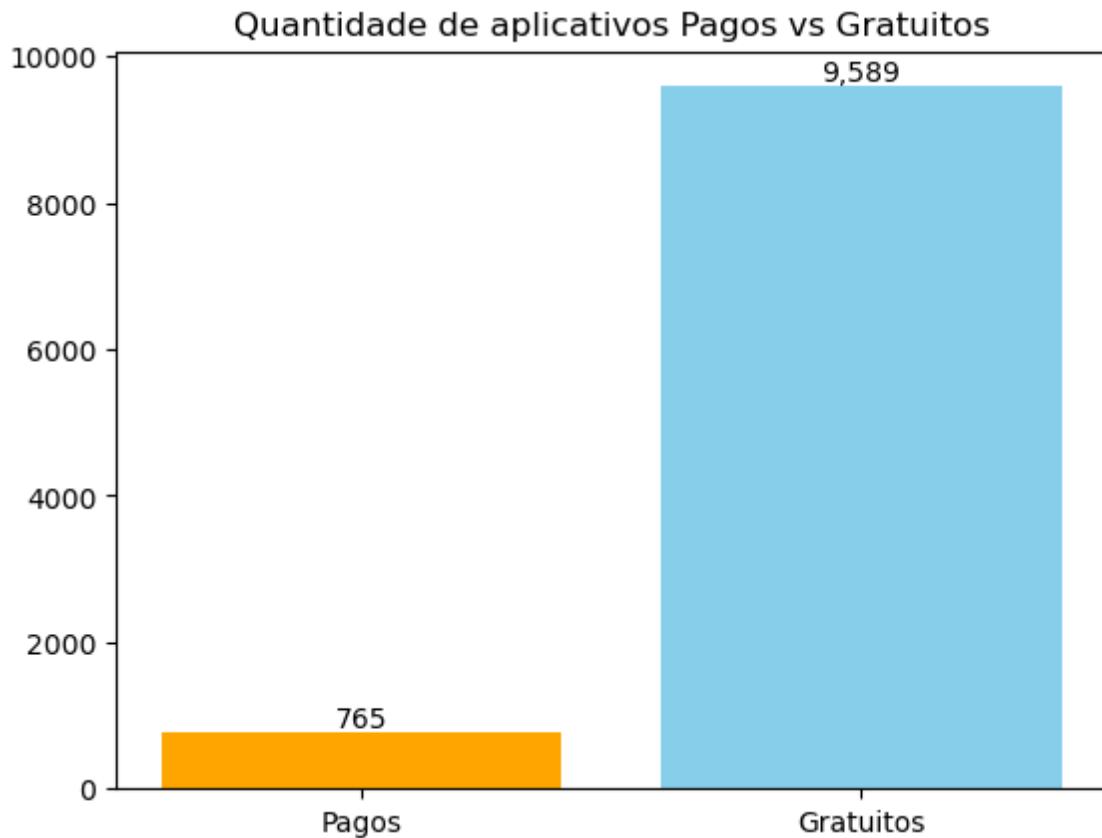
2. **Coluna 'Installs' (Instalações):** Os dados estavam no formato *string* (tipo *object*), com os números seguidos pelo símbolo "+" (ex.: "10.000+"). Seguindo a convenção padrão para este tipo de métrica, o símbolo "+" foi removido e os valores foram convertidos para o tipo numérico inteiro (*int*). Isso representa o **mínimo** de instalações reportado para cada aplicativo, garantindo consistência sem criar informações presunidas.
3. **Coluna 'Type' (Tipo):** Foi encontrado um único valor nulo (NaN). Ao investigar o registro, constatou-se que se tratava de um aplicativo com zero instalações e zero avaliações. Dada a impossibilidade de inferir de forma confiável se seria "Gratuito" ou "Pago" e sua aparente irrelevância para a análise (sem interação dos usuários), essa instância foi removida do *dataset*.
4. **Coluna 'Price' (Preço):** Os valores, precedidos pelo símbolo de dólar ("\$"), foram convertidos para o tipo numérico de ponto flutuante (*float*). O símbolo foi removido e as strings foram transformadas em números.
5. **Valores Duplicados:** A verificação revelou a existência de 483 registros duplicados (linhas idênticas). A presença de duplicatas poderia enviesar várias análises, como a contagem de aplicativos por categoria ou a identificação dos apps mais baixados. Todos os registros duplicados foram removidos, assegurando que cada linha no conjunto de dados final represente um **aplicativo único**.

Com a conclusão dessas etapas de limpeza, padronização e consolidação, o *dataset* está preparado para as análises subsequentes.

Como os aplicativos estão distribuídos

- **Gratuitos vs Pagos**

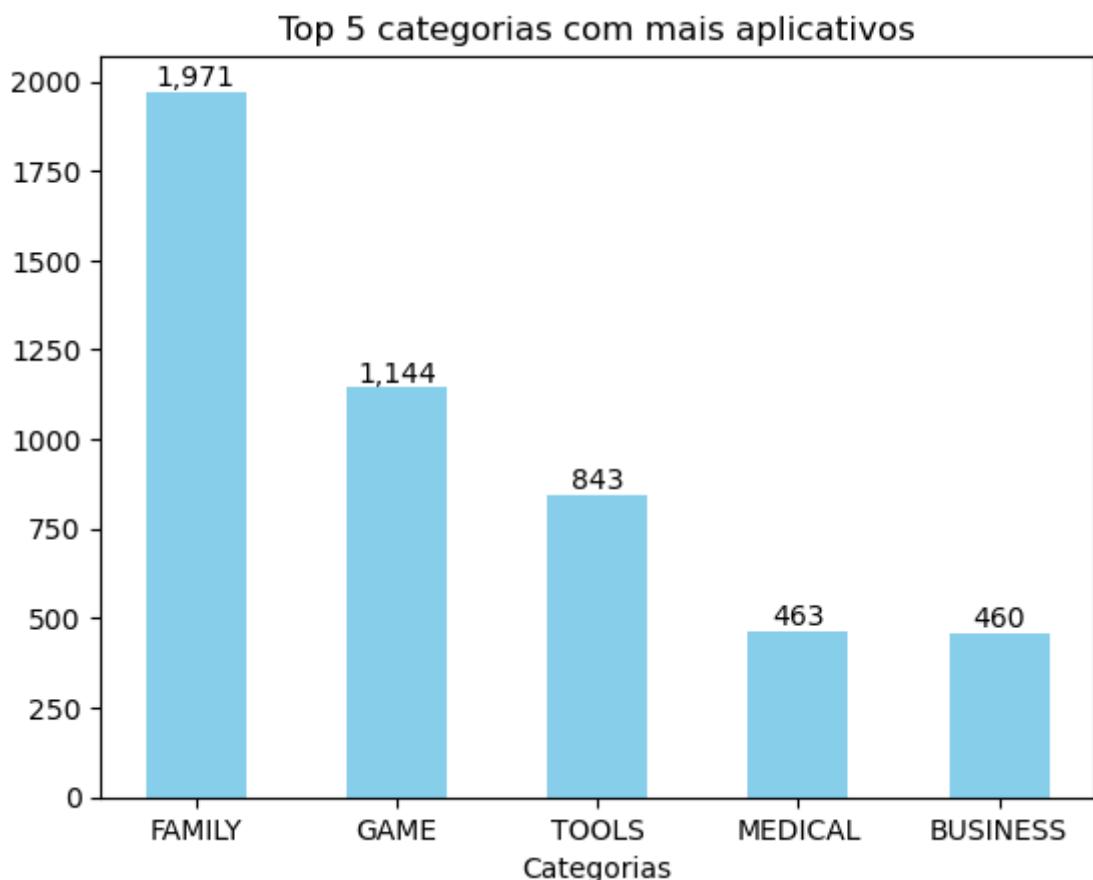
Após a etapa de limpeza, a base de dados consolidada contém 10.354 aplicativos únicos, distribuídos em 33 categorias distintas, dos quais 9589 são gratuitos, representando assim 92,6% dos aplicativos.



- **Distribuição por Categoria**

A distribuição é bastante heterogênea, com concentração significativa em poucas categorias. O destaque é a categoria "FAMILY", que representa sozinha 18.8% do total, com 1.942 aplicativos. Em segundo lugar aparece a categoria "GAME", com 1.121 aplicativos (10.8% do total).

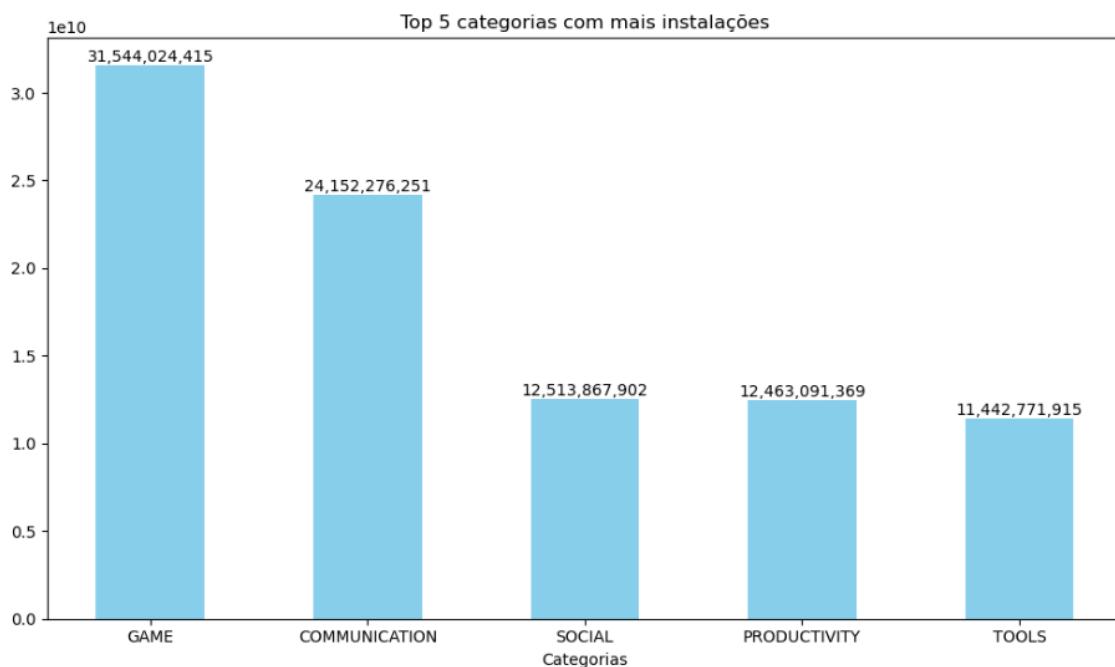
Na outra extremidade, as categorias com menor representatividade são "BEAUTY", com apenas 53 aplicativos (0.5%), e "COMICS", com 60 aplicativos (0.6%).



- **Número de Aplicativos vs. Volume de Instalações**

Surge, então, uma questão: a categoria com mais aplicativos também é a que concentra o maior número de instalações?

Ao cruzar os dados, observa-se uma inversão significativa. A categoria GAME, que era a segunda em quantidade de aplicativos, assume a liderança absoluta no ranking de instalações, acumulando impressionantes mais de 31 bilhões de downloads. Por outro lado, a categoria FAMILY, apesar de possuir o maior catálogo de apps, cai para a sexta posição no volume total de instalações.

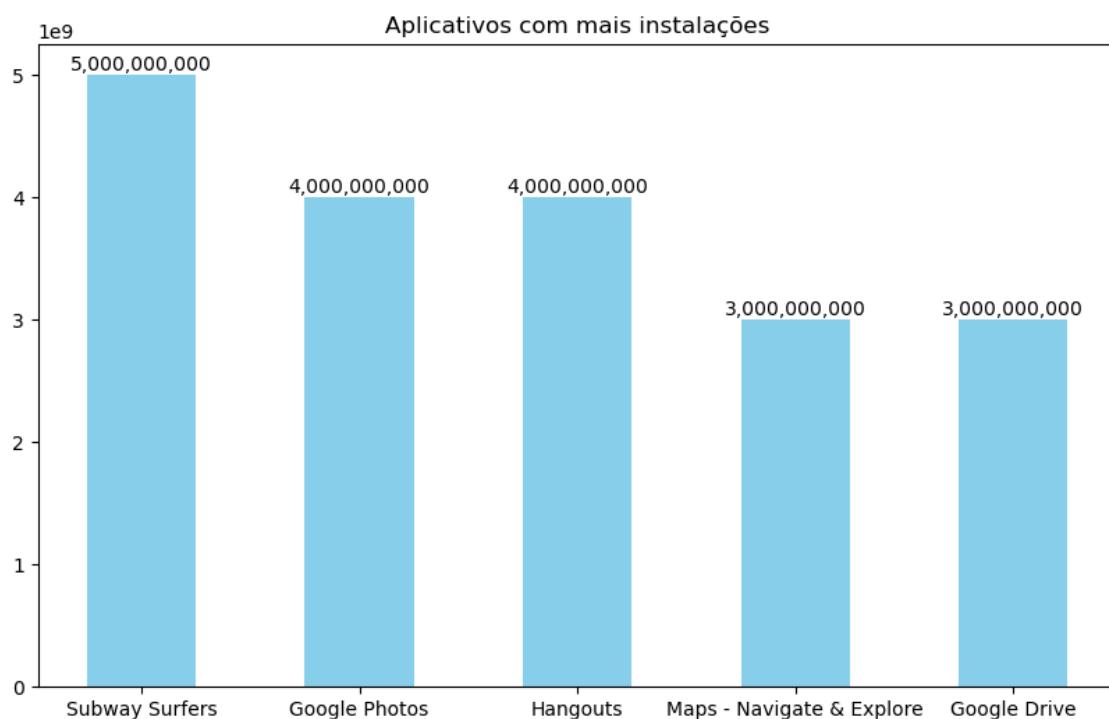


Essa mudança no ranking evidencia que sucesso em volume (número de apps) não se traduz automaticamente em sucesso de mercado (número de instalações). Enquanto a categoria Family parece ter um mercado mais fragmentado com muitos apps de nicho, a categoria game demonstra ter apps com apelo massivo, capazes de gerar bilhões de downloads individualmente, concentrando o interesse dos usuários.

- **Aplicativo mais baixado e outros dados relevantes**

Confirmando a dominância da categoria GAME, o aplicativo individual mais baixado de todo o conjunto analisado é o Subway Surfers, com aproximadamente 5 bilhões de instalações. Embora seja possível identificar aplicativos com mais de 1 bilhão de instalações, é importante destacar que os dados da Google Play Store são apresentados em faixas mínimas de download. Dessa forma, não é possível ordenar com precisão os aplicativos dentro da mesma faixa, apenas classificá-los como pertencentes a um mesmo nível de popularidade.

Como curiosidade e para ilustrar a amplitude dos dados, observa-se que, em contraste extremo, apenas 14 aplicativos (0,14% do total) registraram zero downloads.



Outros pontos interessantes de observar é que o top 5 de aplicativos mais baixados é composto em seu total por aplicativos gratuitos e 4 deles são da própria Google, o que pode demonstrar forte domínio do ecossistema Google talvez pelo fato desses aplicativos consequentemente demonstrarem maior compatibilidade e desempenho com o sistema operacional.

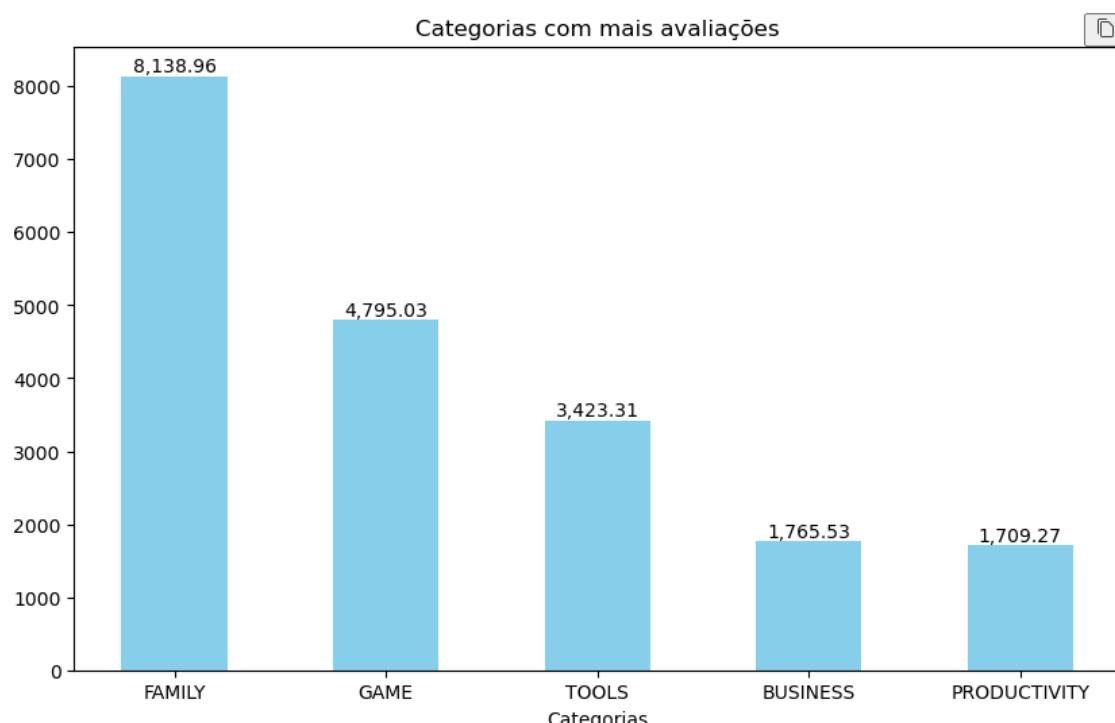
Analisando as avaliações

- **Avaliações dos aplicativos**

No total, **271 aplicativos alcançaram a nota máxima (5 estrelas)** nas avaliações dos usuários. Quando consideramos uma faixa de classificação mais ampla, entre **4.5 e 5 estrelas**, temos **2.815 aplicativos** que podem ser considerados como bem avaliados. Esse grupo representa **25,97% do total de aplicativos** analisados e é responsável por mais de **45 bilhões de instalações** na Google Play Store, correspondendo a aproximadamente **30,80% de todos os downloads**.

- **Categorias com mais avaliações**

A categoria **FAMILY** apresentou o maior nível de engajamento em avaliações por parte dos usuários. Esse comportamento pode estar relacionado à natureza desses aplicativos e ao perfil das pessoas que consomem esse tipo de aplicativo. Como são direcionados à família, os usuários podem sentir-se mais responsáveis e dedicando um momento para avaliar esses aplicativos da mesma forma que dão valor as avaliações também feitas por outras pessoas anteriormente.



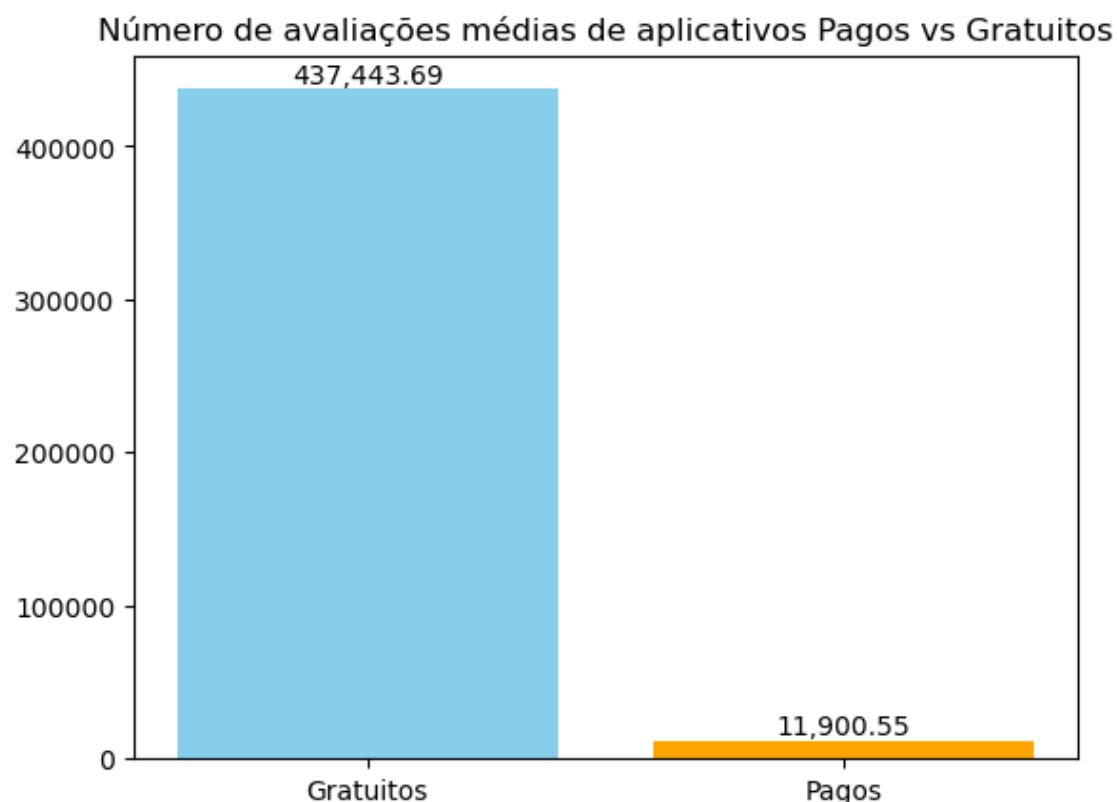
- **Aplicativos mais baixados e suas avaliações**

Aplicativo	Instalações Estimadas	Nº de Avaliações	% de Usuários que Avaliaram
Subway Surfers	5.000.000.000	138.606.606	2,77%
Google Photos	4.000.000.000	43.423.827	1,09%
Hangouts	4.000.000.000	13.676.659	0,34%
Maps	3.000.000.000	27.702.141	0,92%
Google Drive	3.000.000.000	8.191.323	0,27%

Observa-se que, mesmo entre os aplicativos mais populares, a taxa de usuários que efetivamente realizam avaliações é inferior a 3%. Isso indica que o número de reviews representa apenas uma pequena fração da base total de usuários, sendo mais um indicador de engajamento do que de alcance.

- Apps pagos têm, em média, avaliações mais altas?

Os dados indicam que aplicativos pagos apresentam, em média, avaliações ligeiramente superiores quando comparados aos aplicativos gratuitos. Enquanto os apps pagos registram uma nota média de 4,25, com aproximadamente 11.900 avaliações, os aplicativos gratuitos apresentam uma nota média de 4,18, concentrando um volume significativamente maior de avaliações, cerca de 437.444 reviews.



O volume expressivamente maior de avaliações em aplicativos gratuitos reflete diretamente sua base de usuários muito mais ampla. Entretanto, a nota média inferior observada nesse grupo pode estar associada a diversos fatores que impactam negativamente a experiência do usuário, como a presença de anúncios invasivos, compras dentro do aplicativo, funcionalidades limitadas e instabilidades ou bugs, que tendem a gerar frustração e, consequentemente, avaliações mais baixas.

Além disso, o comportamento do usuário difere conforme o modelo de monetização. Em aplicativos gratuitos, os usuários podem se sentir mais propensos a avaliar de forma impulsiva diante de qualquer insatisfação, uma vez que não houve um custo financeiro direto. Por outro lado, usuários de aplicativos pagos tendem a apresentar um engajamento mais qualificado, com maior propensão a deixar avaliações, ainda que em menor volume. Nesse contexto, cada review em aplicativos pagos possui um peso relativo maior, considerando a base de usuários mais restrita.

Conclusão

Esta análise revelou padrões importantes sobre a dinâmica da Play Store, mas abre caminho para investigações mais específicas. Futuros estudos poderiam explorar: A correlação entre preço e avaliação nos aplicativos pagos; Comparações diretas entre categorias específicas (ex: Educação vs. Entretenimento); Análise segmentada por faixas de preço; O impacto de outros fatores como tamanho do app ou atualizações.

Os dados confirmam que diferentes modelos de negócio atraem públicos distintos: aplicativos pagos tendem a nichos onde a qualidade percebida justifica o investimento, enquanto os gratuitos dominam estratégias de alcance massivo. Essa segmentação natural do mercado oferece oportunidades distintas para desenvolvedores, dependendo de seus objetivos e público-alvo.