

Análise exploratória dos usuários brasileiros de mCommerce para a criação de sistema uma recomendação através de similaridade de conteúdo

Mariana de Oliveira Goncalves Rodrigues¹; Prof. Ms. Fernando Freire Vasconcelos.

¹ Fließstraße, 14, Niederschöneweide; 12439, Berlim, Berlim, Alemanha

² Universidade de São Paulo. Orientador USP/Esalq /Doutorando FEA-USP. Universidade de São Paulo – Av. Prof. Luciano Gualberto, 908 – Butantã; 05508-010.

Análise exploratória dos usuários brasileiros de mCommerce para a criação de sistema uma recomendação através de similaridade de conteúdo

Resumo

O objetivo do presente estudo foi realizar uma análise dos usuários brasileiros de mCommerce para criar um sistema de recomendação de aplicativos, com o intuito de aumentar a visibilidade dos apps menos conhecidos entre usuários de Android. O estudo utilizou-se de dados demográficos, de investimento, avaliações e informações sobre quantidade de downloads totais tanto orgânicos quanto pagos do ranking geral dos aplicativos da categoria shopping do Brasil para que se pudesse realizar uma pesquisa demográfica a fim de entender melhor o comportamento dos usuários. Além disso a pesquisa também se aprofundou nas questões da correlação entre altos investimentos em publicidade e popularidade do aplicativo provando que os 10 aplicativos mais baixados eram também aqueles com mais de 50% dos investimentos totais em tráfego pago. Utilizando-se da API da maior plataforma de dados mobile do mundo, a data.ai, foi possível extrair esses importantes dados e que permitiram que todas as análises fossem feitas e utilizando técnicas de similaridade de conteúdo, como a matriz TD-IDF, o sistema foi capaz de recomendar aplicativos similares com sucesso. Concluiu-se que o sistema de recomendação desenvolvido foi uma ferramenta útil para aumentar a visibilidade de aplicativos menos conhecidos e permitiu aos usuários descobrir novas opções de apps.

Palavras-chave: Sistema de recomendação, matriz TF-IDF, similaridade de conteúdo, análise demográfica mCommerce, Python, recomendação de aplicativos, correlação de investimento

Exploratory analysis of Brazilian mCommerce users for the creation of a recommended system using content similarity

Resumo

The objective of this study was to analyze Brazilian mCommerce users to develop a mobile application recommendation system, with the goal of increasing the visibility of lesser-known apps among Android users. The study utilized demographic data, investment figures, reviews, and information on the total number of organic and paid downloads from the general ranking of shopping category apps in Brazil. This was to conduct a demographic survey to better understand user behavior. Moreover, the research delved into the correlation between high advertising investments and app popularity, proving that the 10 most downloaded apps also had over 50% of the total investments in paid traffic. Using the API from the world's largest mobile data platform, data.ai, it was possible to extract this critical data, allowing for comprehensive analysis. By utilizing content similarity techniques, such as the TF-IDF matrix, the system was able to successfully recommend similar apps. It was concluded that the developed recommendation system was a useful tool for increasing the visibility of lesser-known apps and allowed users to discover new app options.

Keywords: Recommendation system, TF-IDF matrix, content similarity, mCommerce demographic analysis, Python, app recommendation, investment correlation.

Introdução

É difícil imaginar o mundo moderno sem acessar ao menos uma vez o smartphone, seja para comunicação, pesquisa ou se informar sobre os acontecimentos recentes. Desde 2007 quando a Apple lançou seu primeiro dispositivo mobile, o iPhone, e o Google anunciou o seu sistema operacional, Android, o mercado não parou de crescer e abrir novas possibilidades.

Segundo pesquisa realizada pela Pew Research Center (2016), o número de usuários de smartphones disparou em 2013, alcançando mais de 25% de crescimento em países emergentes como o Brasil, Chile e Malásia; acelerando, também, o uso de internet, o que diminuiu o abismo entre países desenvolvidos e emergentes.

Entre 2013 e 2021 o mercado de smartphones no Brasil cresceu a ponto de colocar o país em quinto lugar em número de usuários no mundo. O ano de 2020 fechou com 234 milhões de acessos de telefonia móvel, o que representa um aumento de 7,39 milhões de acessos em relação ao mesmo período em 2019. A pandemia de COVID-19 foi um dos prováveis fatores para o aumento de acessos, uma vez que houve a transferência de atividades presenciais para o ambiente online neste período. (ANATEL,2020).

Segundo a pesquisa Impacto econômico e social do Android no Brasil (MOURA, Livia, CAMARGO, Gustavo, 2020) realizada pelo Google em parceria com a consultoria global Bain & Company, o Android está presente em mais de 90% dos smartphones no país, sendo responsável pela democratização do acesso à internet no Brasil, subindo de 41% para 70% entre 2010 e 2018. Nesse período, 24 milhões de brasileiros acessaram a internet pela primeira vez. Isso se deve, sobretudo, pela gratuidade deste sistema operacional e, também, pelo fato de ser aberto e sem custo, o que possibilita a sua utilização livre por muitos fabricantes de smartphones. Outro fator para a soberania do Android no Brasil é o preço elevado dos aparelhos de seu principal concorrente: Apple.

No estudo anual “State of Mobile 2022” da plataforma Data.ai, a principal ferramenta para analista de dados de aplicativos, aponta que, em 2021 foram lançados cerca de 2 milhões de novos aplicativos, sendo que, destes, 77% foram para o Google Play. Em 2020 foram realizados um total de 230 bilhões de downloads globalmente, fato que gerou uma receita de \$170 bilhões em transações feitas dentro de aplicativos. Embora seja um mercado global com muitos desenvolvedores, apenas 233 apps acumulam uma receita superior a \$100 milhões de dólares, sendo apenas 13 que ultrapassaram o valor de \$1 bilhão. O Brasil foi destacado no estudo ao alcançar o 4º país com o maior volume de downloads e o 14º em receita, além de ser o que passa o maior tempo no celular, ultrapassando 5 horas diárias.

Em relação a anunciantes, o estudo também divulgou algumas importantes cifras. Com tantas opções e um vasto mercado, a disputa pela atenção do usuário é grande, envolvendo investimentos bilionários em publicidade, os quais beiram \$295 bilhões de dólares em 2021. Há expectativas de ultrapassar \$350 bilhões de dólares globalmente em 2022.

A Appsflyer, conhecida plataforma de análise de dados e atribuição de aplicativos, também lançou seu estudo focado no eCommerce chamado “State of eCommerce (2021)”, o qual aponta que, somente em relação a Apps de e-Commerce, o gasto global com anúncios para aquisição de usuários foi de \$5.4 bilhões entre o último trimestre de 2020 e o primeiro trimestre de 2021. Atrás apenas dos Estados Unidos, o Brasil representa o segundo maior mercado investidor e o maior mercado em Android, com 19% do total de instalações de apps de compras do mundo.

Dados extraídos da plataforma Data.ai e apresentados no presente trabalho mostram que os grandes conglomerados de tecnologia estão entre os que mais investiram em publicidade, como, por exemplo, as chinesas *Shopee*, *Shein*, a argentina *MercadoLivre*; e a americana Amazon. Entre as empresas brasileiras está a MagazineLuiza e Lojas Americanas. Ao analisar o ranking dos aplicativos mais baixados pode-se, a princípio, fazer uma correlação entre os maiores anunciantes e os aplicativos com os maiores volumes de downloads.

A presente pesquisa propõe-se a analisar essa correlação e o perfil dos usuários dos m-Commerce brasileiros, e com isso gerar um sistema de recomendação utilizando-se similaridade de conteúdo para aplicativos que não se encontram entre os top 10 dos apps mais baixados entre os mCommerce no sistema Android. Utilizando-se, para tanto, dados demográficos e dimensionais com o objetivo de gerar maior visibilidade.

Material e Métodos

Esta pesquisa coletou dados disponibilizados pela plataforma Data.ai, uma das maiores plataformas utilizadas por analistas e profissionais de marketing voltados para o mercado de aplicativos, para o desenvolvimento da análise exploratória dos usuários de mCommerce no Brasil e para o sistema de recomendação. Após a coleta foi realizada um pré-processamento dos dados para que esses se encaixassem nos padrões necessários da análise, retirando valores nulos e incompletos. Por fim, foi gerado o sistema de recomendação com base em similaridade de conteúdos.

1.1 Coleta de Dados

A coleta de dados foi feita com auxílio da linguagem SQL, por meio da API (interface de programação de aplicações) da Data.ai, os quais estão armazenados em nuvem no programa “Snowflake” (empresa de computação em nuvem).

A pesquisa foi dividida em quatro amostras distintas, todas relacionadas a aplicativos da categoria de compras e disponíveis para o mercado brasileiro através da loja Google Play. Os dados coletados são de julho a setembro de 2022.

A primeira amostra, denominada Amostra Geral, contém dados relacionados ao ranking dos aplicativos, descrição, número de usuários ativos, número total de downloads e informações sobre o número de downloads gerados por cada mídia paga. A amostra contém um total de 347 observações e 44 variáveis.

A segunda amostra é relacionada aos dados demográficos, faixa etária e gênero, totalizando 2101 amostras através de 344 diferentes aplicativos.

A terceira amostra incluiu um total de 19.092 observações com avaliações e classificações dos usuários variando de um a cinco, realizadas entre 16 e 22 de julho de 2022. Essas avaliações foram feitas para os aplicativos que ocupavam as três primeiras posições na categoria de e-commerce no mesmo período, que são, respectivamente, Shopee, Shein e Mercado Livre.

A quarta amostra inclui dados sobre os anúncios feitos por cada aplicativo entre julho e setembro de 2022, totalizando 389 observações, com informações sobre quantidade de anúncios feitos pelos aplicativos e quantidade de plataformas utilizadas para anúncios.

1.2 Pré-processamento

Foi utilizado a ferramenta “Jupyter Notebooks” em conjunto com a linguagem Python e bibliotecas como Pandas e Numpy para fazer uma triagem inicial dos dados, excluindo-se, assim, aqueles que não seriam necessários para nenhuma análise.

De maneira geral, textos em sua forma natural não são bem formatados nem padronizados. O pré-processamento envolveu usar diversas técnicas para converter o texto em uma sequência de componentes linguísticos bem estruturados para que outros sistemas e aplicativos de NLP possam interpretar corretamente (SARKAR, 2016). Durante o processo foram retirados comentários com links em HTML, avaliações repetidas, stop words (palavras de baixa relevância semântica) e, a fim de concentrar a amostra em adjetivos, foram removidos os verbos, utilizando o pacote do Python chamado Spac.

Para analisar os dados de investimento em publicidade, utilizou-se também os dados de downloads pagos coletados da Amostra Geral. Ao combinar as duas amostras, foram encontrados alguns valores nulos, os quais foram descartados, resultando em 317 observações. Além disso, os valores dos três meses foram somados para cada observação, resultando em um total de 118 observações:

Tabela 1. Total investimento em publicidade estimado dos aplicativos no top 10

Posição no Ranking	Nome do Aplicativo	Estimativa do Investimento em Aquisição de Usuários	Número de criativos	Plataformas anunciadas	Downloads pagos	Quantidade estimada de anúncios
Primeiro	Shein	\$30.598.014	129.344	992	13.659.828	44.665.746
Segundo	Shopee	\$28.303.396	47.926	12360	12.635.445	197.838.154
Terceiro	MercadoLivre	\$8.630.543	12.216	846	3.852.921	3.257.938
Quarto	Magalu	\$9.728.987	44.637	5129	4.343.298	87.635.892
Quinto	Americanas	\$10.255.647	35.894	624	4.578.414	7.843.272
Sexto	Amazon	\$9.066.509	23.725	443	4.047.549	3.195.293
Sétimo	Casas Bahia	\$7.977.110	22.471	216	3.561.210	1.853.635
Oitavo	Aliexpress	\$7.640.223	63.842	1101	3.410.814	24.836.256
Nono	OLX	\$6.407.909	1.833	45	2.860.674	29.796
Decimo	O Boticario	\$3.173.063	3.074	20	1.416.546	14.768

Fonte: dados originais da pesquisa

De acordo com o Relatório de Apps Mobile de Compras (2022) disponibilizado pela Liftoff e Singular, cerca de 85% dos gastos com publicidade são destinados a novos usuários que estão baixando o aplicativo pela primeira vez. O custo médio de aquisição de usuário, conhecido como modelo CPI (custo por instalação), é de cerca de \$2.24. Com os dados

coletados, foi possível estimar o valor de investimento, multiplicando a quantidade de downloads pagos pelo valor médio de CPI.

1.3 Definição do modelo para o sistema de recomendação

Uma das principais técnicas utilizadas é através da similaridade de conteúdo baseada em TF-IDF, por ser de simples implementação, eficiente e largamente utilizada quando existe uma limitação maior no volume de dados (Muthurasu et al, 2019). Diversos estudos foram conduzidos utilizando essa mesma técnica, principalmente para sistemas de recomendações para filmes e series em “streaming”. Um deles realizados por Chiny e colaboradores (2021) que utilizou dados dos títulos e descrições e, diferentemente do streaming estudado, foram considerados também dados demográficos para que se pudesse gerar uma diferencial na análise, algo que também foi considerado para o presente estudo.

Segundo o “ASO Guide 2020”, uma pesquisa realizada pela TheTool e PickASO, duas empresas especialistas em gerenciar e criar ferramentas para profissionais da área de marketing digital, uma das técnicas mais utilizadas para gerar visibilidade aos aplicativos é o ASO – “App Store Optimization” ou Otimização da busca na loja de aplicativos, em tradução livre. De acordo com a referida pesquisa, há dois fatores que influenciam o ranking: “on metadata”, que são fatores que podem ser modificados pelos desenvolvedores como, por exemplo, o título do aplicativo, descrição, gênero e subgênero e palavras-chaves; o segundo fator é o “off metadata”, que são os fatores em que não se tem controle e que dependem quase exclusivamente de investimentos, como, por exemplo, número de usuários ativos e total de downloads.

Para refinar o sistema de recomendação e mostrar aplicativos relevantes a seus usuários, o Google utiliza dados internos do próprio aparelho Android e de ações feitas dentro do aplicativo Google Play, como compra de algum aplicativo ou serviço, histórico de busca e aplicativos instalados. (Support Google, 2023). Para o desenvolvimento do sistema de recomendação, foram considerados os dois fatores relevantes para ASO: o título, o subgênero do aplicativo e a principal audiência, sendo esta feminina ou masculina (on-metadata); e a quantidade de usuários ativos (off-metadata). Neste último, os valores foram divididos em segmentos de acordo com o seu percentil; a técnica foi utilizada para que se pudesse converter os dados quantitativos em qualitativos e assim usar no algoritmo de recomendação.

Empregou-se uma metodologia fundamentada na similaridade de conteúdo, recorrendo à matriz TF-IDF, que significa *Term Frequency – Inverse Document Frequency* ou Frequência do Termo – Inverso da Frequência no Documento, que consiste em um instrumento estatístico eficaz para determinar a relevância de termos em um texto.

$$idf_t = \log \frac{N}{df_t}$$

onde t é um termo específico do documento, N é o número total de documentos e df_t é o número de documentos que contém o termo. O IDF mede a raridade de um termo em uma coleção de documentos. Já para o cálculo da frequência de termo:

$$TF = \frac{\text{numero de vezes em que o termo aparece no documento}}{\text{numero total de termos no documento}}$$

A matriz TF-IDF visa converter cada texto, também chamado de documento, em um vetor. Cada componente desse vetor está relacionado a um termo específico e seu respectivo valor é determinado por um peso, que representa a relevância desse termo no contexto do documento.

O peso é uma combinação dos termos de frequência com a frequência inversa do documento para, assim, produzirmos um peso diferente para cada documento, podendo ser traduzida na equação abaixo (SCHÜTZE; MANNING; RAGHAVAN, 2008):

$$tf - idf_{t,d} = tf_{t,d} \cdot idf_t$$

onde $tf - idf$ é o valor do $tf-idf$ para cada termo (t) e a quantidade de vezes em que aparece no documento igual a frequência do termo no documento multiplicado pelo IDF do termo específico. A multiplicação do TF pelo IDF resulta em um valor numérico a qual indica a importância e relevância do termo no documento.

Para a obtenção da matriz TD-IDF foi utilizada a biblioteca de python *sklearn* chamada *TfidfVectorizer*.

Resultados e Discussão

2.1 Estatísticas descritivas

Com um total 240 observações após o pré-processamento dos dados, foram extraídos alguns dados gerais dos aplicativos da categoria Shopping. Entre eles tem-se o total de downloads, downloads pagos, downloads orgânicos (que não tiveram a sua origem através de uma propaganda) usuários ativos, sessões por usuário que podem ser interpretados como uma métrica de retenção, uma vez que indicam quantas vezes o usuário abriu o aplicativo.

Pelo desvio padrão é possível observar que há uma grande variabilidade entre os dados e a presença de discrepâncias na amostra em praticamente todas as métricas, exceto em sessão por usuário, a qual os dados possuem uma distribuição mais próxima em relação a média, ou seja, a retenção do usuário é consistente entre todos os aplicativos, independente de estarem no top 10 ou não. Contudo a diferença maior fica por conta da quantidade de downloads, principalmente aqueles gerados de maneira paga.

Tabela 2. Principais métricas dos dados gerais dos aplicativos de Shopping

Métricas	Total de downloads	Downloads pagos	Downloads orgânicos	Usuários ativos	Sessão por Usuário
Média	306.532	142.207	164.325	1.295.025	11,12
Desvio Padrão	908.860	457.762	460.495	5.079.521	10,39
Primeiro Quartil	18.193	3.813	8.080	13.382	5,88
Segundo Quartil	46.864	18.110	26.342	40.682	8,02
Terceiro Quartil	235.796	88.679	134.853	217.497	11,56
Máximo	8.674.918	4.553.276	4.121.642	44.589.200	100,74

Fonte: dados originais da pesquisa

O número de downloads totais dos aplicativos no top 10 é de 37.462.837, o que corresponde a 51% do total, sendo os outros 49% divididos entre 230 aplicativos. Entre os downloads pagos, esse número sobe para 53% do total. O gráfico abaixo detalha a quantidade de downloads totais dividido pelos aplicativos no top 10:

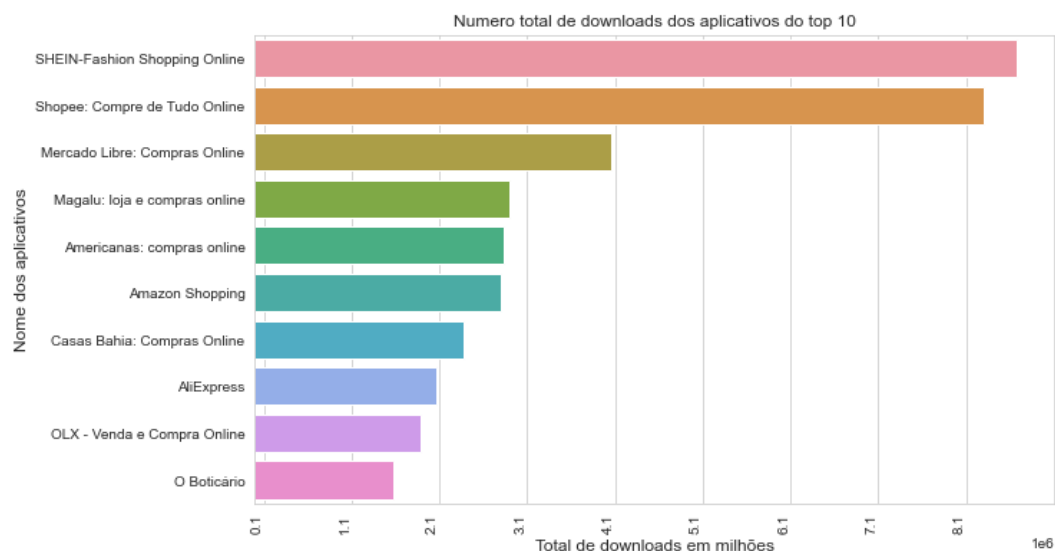


Figura 1. Total de downloads do top 10 dos aplicativos mais baixados da categoria Shopping do Google play

Fonte: Dados originais da pesquisa

2.2 Análise Exploratória dos dados demográficos e avaliações dos usuários

Foi realizada uma análise para identificar as principais características dos usuários dos aplicativos de mCommerce do Brasil a fim de traçar um perfil de usuário.

Os dados demográficos coletados mostram que 54,5% dos usuários são mulheres e 66,1% do total estão entre os 25 e 44 anos.

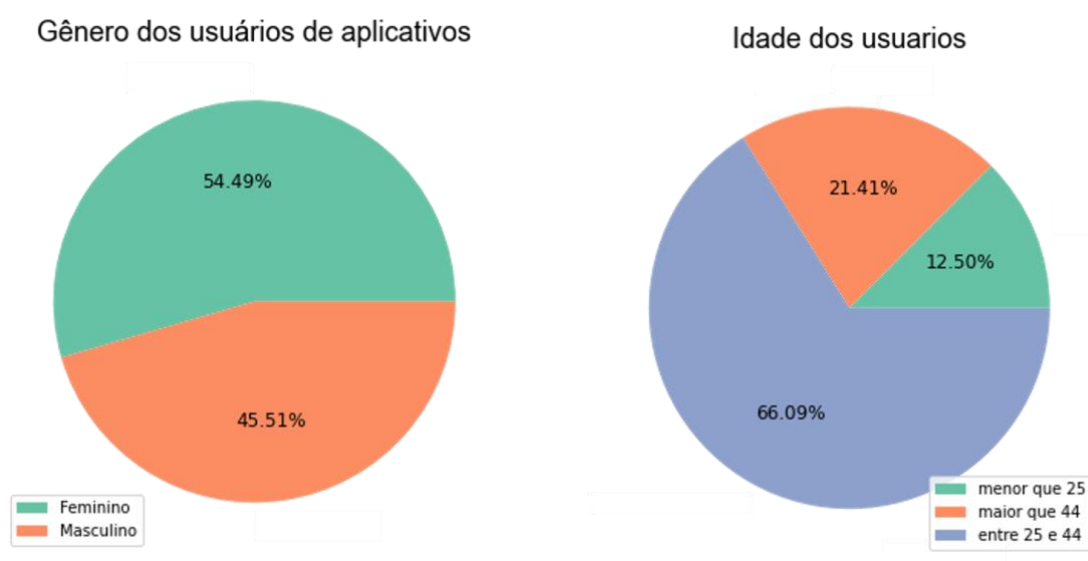


Figura 2. Gênero e faixa-etária dos usuários de aplicativos de Shopping no Brasil

Fonte: Dados originais da pesquisa

Os dados mostram que as mulheres têm uma tendência maior em utilizar aplicativos do subgênero E-Commerce (Retailer), que, geralmente, são aqueles que não possuem lojas físicas, apenas online. Já para os homens, os aplicativos mais usados também são E-Commerce (Retailer) porém, com uma distribuição maior entre os outros subgêneros, como mostra a tabela abaixo:

Tabela 3. Frequência de subgêneros do mCommerce entre homens e mulheres

Subgênero	Mulheres	M. Frequência	Homens	H. Frequência
E-Commerce (Retailer)	53	50.00%	25	38.46%
E-Commerce (B2C)	24	22.64%	14	21.54%
Other Shopping	10	9.43%	5	7.69%
E-Commerce (C2C)	9	8.49%	7	10.77%
Coupons and Rewards	7	6.60%	11	16.92%
Resell	2	1.89%	1	1.54%
Pharmacy and Drugstore	1	0.94%	1	1.54%
BNPL	0	0.00%	1	1.54%

Fonte: dados originais da pesquisa

A amostra de avaliações e classificações dos usuários forneceu dados para melhor compreensão do nível de satisfação dos usuários e o que eles mais gostam ou não nos aplicativos posicionados entre os três mais populares, quais sejam, na ordem: Shein (com.zzkko), Shopee (com.shopee.br) e Mercado Livre (com.mercadolivre).



Figura 3. Nota das avaliações dos usuários para os três aplicativos mais baixados da categoria mCommerce

Fonte: Dados originais da pesquisa

Cerca de 75% da amostra é composta pela nota máxima 5, o que indica grande satisfação dos usuários com os serviços oferecidos pelos aplicativos. Também é notável a disparidade no volume de avaliações deixadas pelos usuários, o qual Shopee, apesar de estar em segundo lugar no ranking, possui quase o dobro de avaliações positivas em relação ao primeiro colocado, Shein. Mercado Livre também se destaca como o aplicativo com a maior proporção de avaliações negativas em relação aos demais.

A fim de entender o que os usuários gostam ou não dos aplicativos, foi feita uma análise dos comentários da página do Google Play dos três aplicativos. A tabela abaixo mostra a contagem das 10 palavras mais utilizadas por usuários que deixaram a maior e pior nota.

Nas avaliações negativas é possível ver que existem insatisfações referentes a reembolso, entrega e pedido. “Vendedor” também é uma palavra frequente o que pode justificar o fato de o aplicativo “Mercado Livre” possuir uma proporção maior de notas negativas, haja vista ser um *market place*, ou seja, existem diversos vendedores que utilizam a plataforma para oferecer seus produtos. Em relação às avaliações positivas, nota-se o uso de palavras que reforçam o sentimento de satisfação, como “bom”, “ótimo” ou “melhor”; destaque para o termo “preço” ter aparecido diversas vezes de maneira positiva, indicando ser um dos pontos de maior interesse dos usuários.

Algumas palavras se repetem tanto nas avaliações positivas como negativas como “compra”, “produto” e “aplicativo”.

Tabela 4 e 5. Frequência de subgêneros do mCommerce entre homens e mulheres

Palavras Positivas	Contagem	Palavras Negativas	Contagem
Bom	3154	Compra	679
Muito	2041	Produto	474
Ótimo	2016	Nao	359
Compra	1968	Aplicativo	359
Produto	1654	Nada	334
Aplicativo	1124	Dia	317
Tudo	1051	Reembolso	292
Melhor	954	Agora	270
Qualidade	903	Entrega	241
Preço	877	Pedido	229

Fonte: dados originais da pesquisa

2.3 Correlação entre investimentos em publicidade e ranking

Para que se pudesse quantificar a correlação existente entre alto investimento e posicionamento em um ranking, foi utilizada o coeficiente de Spearman, que é uma medida de associação entre variáveis ordinais cujo resultado varia entre -1 a 1, sendo -1 uma associação linear negativa perfeita entre as variáveis e 1 uma associação positiva perfeita entre as variáveis de -1 a 1 (Fávero e Belfiore, 2017).

A correlação entre o posicionamento no ranking e o investimento estimado foi calculada utilizando-se o coeficiente de Spearman. O resultado foi de -0.93, ou seja, quanto menor o investimento, mais distante do topo do ranking o app está. O resultado corrobora com a premissa da pesquisa que se propõe em recomendar apps que possuem um maior investimento com base no perfil do usuário.

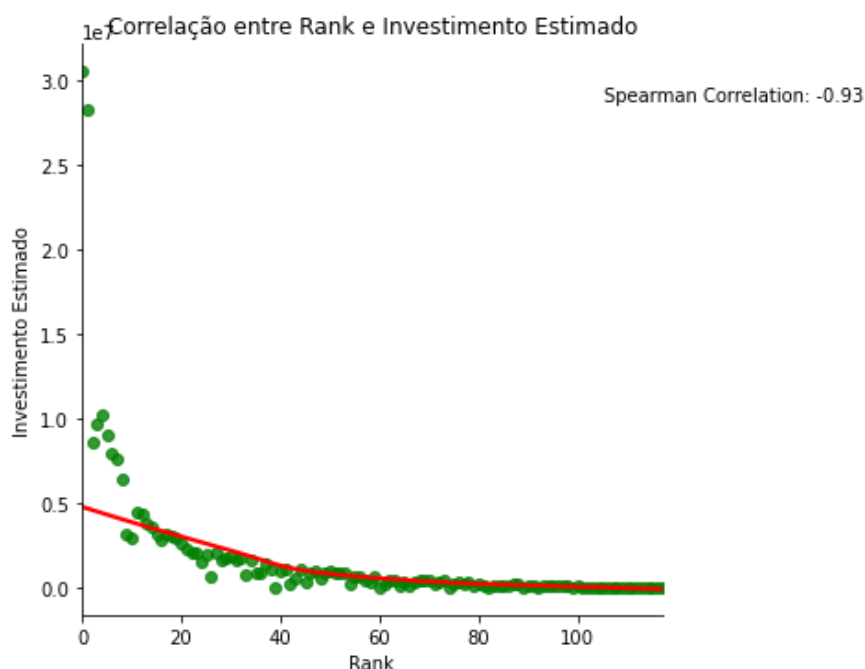


Figura 4. Gráfico da correlação entre posicionamento no ranking e investimento

Fonte: Dados originais da pesquisa

2.4 Modelo de recomendação

Para a criação do modelo de recomendação de aplicativos, foram usados os nomes dos aplicativos disponíveis no Google Play e as informações dos subgêneros dos aplicativos, que são otimizados para obter um melhor ranqueamento no sistema de busca da plataforma, utilizando-se métodos "ondata" de ASO, conforme mencionado anteriormente. Além disso, dados demográficos relacionados ao gênero predominante dos usuários foram coletados para

umentar a precisão dos resultados. Informações sobre o número de usuários ativos também foram empregados, utilizando-se o seu percentil para que pudesse, também, isolar os maiores aplicativos dos menores. A fim de evitar a duplicação, foi adicionada uma lógica de exclusão para impedir que um aplicativo recomendasse a si mesmo, uma vez que o maior índice de similaridade seria entre ele próprio e outro aplicativo.

Criou-se uma coluna no dataframe que consiste na combinação dos valores mencionados anteriormente. A partir desta coluna, foi possível realizar a análise de similaridade entre os aplicativos e criar a matriz TF-IDF. A matriz resultante tem dimensões de 240x240 e apresenta os respectivos pesos entre os aplicativos. Para melhor compreensão, segue abaixo uma amostra da matriz como exemplo:

Tabela 6. Amostra da matriz TD-IDF com o índice de afinidade entre os aplicativos

Nome do Aplicativo	Shein – Fashion Shopping Online	Shopee: Compre de tudo online	MercadoLibre: Compras online	Magalu: loja e compras online	Americanas: compras online
Shein Fashion Shopping Online	1.00	0.18	0.16	0.21	0.21
Shopee: Compre de tudo online	0.18	1.00	0.24	0.17	0.17
MercadoLibre: Compras online	0.16	0.24	1.00	0.31	0.32
Magalu: loja e compras online	0.21	0.17	0.32	1.00	1.00
Americanas: compras online	0.23	0.19	0.35	0.35	0.35

Fonte: dados originais da pesquisa

Gerou-se uma função para que se obtenha os 5 aplicativos mais similares considerando o índice de afinidade e o resultado; ao buscarmos pelo aplicativo “Shein Fashion Shopping Online” o resultado foi o seguinte:

Tabela 7. Amostra da matriz TD-IDF com o índice de afinidade entre os aplicativos

Posição no rank	AppName	Percentil	Afinidade
225	NewChic – Fashion Online	1	0.40
89	Hibobi – Kids fashion online	3	0.37
156	Light in the box – online shopping	2	0.34
88	Banggood – Online Shopping	4	0.32
222	Fashion Nova	1	0.31

Fonte: dados originais da pesquisa

O resultado da função deu como resposta os aplicativos que mais se assemelham a “Shein Fashion Shopping Online” e que estão fora do top 10 do mais baixados e seu respectivo peso ou, aqui definido como afinidade, que varia de 1 a 0, sendo 1 quando o texto é completamente igual a palavra inicial. Com esse resultado é possível oferecer ao usuário novas alternativas de locais para fazer suas compras e assim expandir mais o leque de aplicativos baixados.

Conclusão

As compras online, em particular via dispositivos móveis, têm visto um aumento significativo nos últimos anos. O comportamento dos usuários em plataformas de mCommerce é fundamental para empresas e desenvolvedores otimizarem suas ofertas e interfaces.

Os resultados do estudo apresentado evidenciam o perfil do usuário de aplicativos de compras em dispositivos Android, bem como a disparidade nos investimentos em publicidade de uma pequena parcela dos mais de 200 aplicativos disponíveis, o que reforça o monopólio das grandes empresas em relação ao ranking de usuários.

As técnicas de ASO são amplamente utilizadas no mercado de *mobile marketing*, com mudanças frequentes nos títulos e descrições dos aplicativos para facilitar sua localização pelos usuários. A técnica TF-IDF permitiu a avaliação simplificada de nomes e outras variáveis para gerar recomendações aos usuários, possibilitando um leque maior de opções ao buscar por um aplicativo específico.

Sugere-se a implementação de novas variáveis, como descrições e avaliações de usuários, aumentando-se a quantidade de texto e dando mais contexto sobre o que é cada aplicativo, o algoritmo aplicado se tornara mais eficiente considerando, não somente o nome, dados demográficos e número de downloads mas também serviços prestados pelos aplicativos conseguindo assim aproximar ainda mais os resultados do aplicativo inicial.

Referências

Agência Nacional de Telecomunicações [ANATEL]. 2020. Relatório de acompanhamento do setor de telecomunicações. Disponível em:

<https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?eEP-wqk1skrd8hSlk5Z3rN4EVg9uLJqrLYJw_9INcO4NT86aq4DZSJMWWh9gBoilhtRgvXnEhjT6dqYhPLelC2xMriZOLrD6LEYnf1psEzILJAq9-LHeI_G9fbuXR7UR>. Acesso em: 21 out 2022

Appsflyer. 2021. The state of eCommerce apps in Brazil. Disponível em:

<<https://www.appsflyer.com/pt/infograms/state-of-ecommerce-apps-brazil-2021/>>. Acesso em: 23 out 2022

Chiny, M; Chihab, M, Bencharef, O; Chinab, Y. Netflix recommendation system based on TF-IDF and cosine similarity algorithms. Disponível em:

<<https://pdfs.semanticscholar.org/22fd/585a75a52264bf2f3ecaabd7a53d5a2ef465.pdf>>.

Acesso em 22 Mar 2023

Data.ai. 2022. State of Mobile 2022. Disponível em: <<https://www.data.ai/en/go/state-of-mobile-2022>>. Acesso em: 28 set 2022

Fávero, L.P; Belfiore, P. 2017. Análise de dados. Estatística e modelagem multivariada com Excel, SPSS e Stata. 1ed. Elsevier Editora Ltda. Rio de Janeiro, RJ, Brasil.

Google Support. 2023. User policies for apps and digital content. Disponível em:

<<https://support.google.com/googleplay/answer/11416267?hl=en&co=GENIE.Platform%3DAndroid#zippy=%2Cdata-collection%2Cdata-types>>. Acesso em 01 de Abr 2023

Liftoff. 2022. Relatório de apps de compra. Próximas tendências em e-commerce e delivery.

Disponível em: <https://info.liftoff.io/pt-br/2022-mobile-shopping-apps-report?utm_source=pr&utm_medium=article&utm_campaign=DL-2022-ShoppingApps-PR&utm_content=Portuguese>. Acesso em: 10 jan 2023

Moura, Livia; Camargo, Gustavo. 2020. Impacto econômico e social do Android no Brasil. Disponível em:

<<https://www.bain.com/contentassets/a9200a057a0241b8963c05a9b09e33fe/impactos-do-android-no-brasil.pdf>>. Acesso em: 4 out 2022

Muthurasu, N; Rengaraj, N; Mohan, K.C. 2019. Movie recommendation system using Term Frequency-Inverse Document Frequency and cosine similarity method. Disponível em: <<https://www.ijrte.org/wp-content/uploads/papers/v7i6s3/F1018376S19.pdf>>. Acesso em 28 Abr 2023

Pew Research Center. 2016. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. Disponível em: <https://www.diapoimansi.gr/PDF/pew_research%201.pdf>. Acesso em: 20 out 2022

PickAso e TheTool.2020.ASO Guide 2020. Disponível em <<https://thetool.io/wp-content/uploads/2020/04/aso-app-store-optimization-guide-2020-pickaso-thetool.pdf>>. Acesso em: 11 mar 2023

Sarkar, D. 2016. Text Analytics with Python a Practical Real-World Approach to Gaining Actionable Insights from Your Data. 1ed. Apress, Bangalore, Karnataka, Índia.

Schütze, H; Manning, C.D; Raghavan, P. 2008. Introduction to information retrieval. Online edition. Cambridge University Press. Disponível em <https://nlp.stanford.edu/IR-book/pdf/06vect.pdf>. Acesso em: 16 mar 2023