



GOOGLE TRENDS E CIÊNCIA DE DADOS: EM PROL DE UM PROCESSO DECISÓRIO MAIS ASSERTIVO

Alex Ribeiro Maia Baroni¹

Resumo

Cabe como a matéria-prima de um Cientista de Dados, os próprios dados. Sem eles não é possível fazer análises, obter *insights* e tomar decisões. O acesso aos dados pode se dar sob diferentes formas, como bases primárias ou secundárias. Entre as bases secundárias menciona-se o Google Trends. Assim, objetiva-se com essa pesquisa, expor um estudo de caso no qual a análise para suporte a decisões está ancorada no uso do Google Trends. Do qual, se estabelece o problema de pesquisa: De que forma a Ciência de Dados pode utilizar-se do Google Trends em prol de um processo decisório mais assertivo? Como resultado alcançado, com o uso da linguagem R, foi possível expor a busca, análise, *insights* e decisões, por meio da base secundária do Google Trends. Assim, considera-se que essa plataforma pode ajudar com uma análise descritiva, preditiva ou até prescritiva, cooperando desta forma na tomada de decisão.

Palavras-chave: Google Trends, tomada de decisão, dados secundários.

Abstract

The raw material for a Data Scientist is the data itself. Without them, it is not possible to perform analyses, obtain insights and make decisions. Access to data can take place in different forms, such as primary or secondary databases. Among the secondary bases, Google Trends is mentioned. Thus, objectively with this research, we present a case study in which the analysis to support decisions is anchored in the use of Google Trends. From which, the research problem is established: How can Data Science use Google Trends in favor of a more assertive decision-making process? As a result, with R language, it was possible to export the search, analysis, insights and decisions, through the secondary Google Trends database. Therefore, consider that this platform can help with descriptive, predictive or even prescriptive analysis, thus cooperating in decision making.

Keywords: Google Trends, decision making, secondary data.

Introdução

O mundo empresarial se coloca em face a diversas tomadas de decisões todos os dias (Davenport, 2014; Kaplan; Norton, 1999; Sharda; Delen; Turban, 2019). Nas quais, decisões mais assertivas relacionam-se tanto a analisar dados do passado, como aqueles aspectos que podem ainda vir a ocorrer, ou seja, conjecturações (Kaplan; Norton, 1999). Os dados históricos de uma empresa podem advir de suas bases de dados, como um ERP, por exemplo, e servem para análise de aspectos passados (Kaplan; Norton, 1999). Enquanto isso, aspectos futuros demandam de investigação sobre esses dados passados, assim

¹ Universidade Federal Fluminense, UFF, baroni.alex@gmail.com



como outros que podem ainda não existir dentro das corporações, ou até mesmo já existem, no entanto a organização não tem acesso aos mesmos (Sharda; Delen; Turban, 2019). A partir dos dados, seguindo pela análise, obtendo os *insights*, é possível prover aspectos que suportem a tomada de decisão (Pantovic *et al.*, 2024). Cujas importância é alcançar valor e inovação empresarial (Davenport, 2014).

O desenvolvimento de uma pesquisa empírica, como a hipotético-dedutiva, está relacionada com a aquisição de dados para a verificação de hipóteses (Popper, 2013). Nesse cenário existem algumas possibilidades. Primeiro, o pesquisador pode ir de encontro aos indivíduos que farão parte de sua amostra que é representativa de uma determinada população, ou seja, quando ele produz dados primários (Gil, 2021). Mas também fazendo uso de dados secundários, por meio de bases de dados já coletadas por outrem (Gil, 2021).

Ao falar de buscar esses dados na internet é preciso mencionar o Google e sua relevância nesse cenário (Sharma *et al.*, 2018). Isso decorre, pois o Google é atualmente o buscador mais utilizado na internet (Sharma *et al.*, 2018). Quando um indivíduo efetiva uma busca nessa plataforma, os dados são registrados pelo Google em suas bases (Lolic; Matosec; Soric, 2024). O Google Trends vem sendo uma ferramenta que está auxiliando pesquisadores nas mais diferentes áreas, inclusive ajudou a produzir conhecimento durante o auge da pandemia de COVID-19 (Gadzalo *et al.*, 2020).

Objetivo

Apresenta-se como objetivo dessa pesquisa, analisar o Google Trends como uma ferramenta útil para a Ciência de Dados em prol de um processo decisório mais assertivo. Para tanto, as seguintes hipóteses de pesquisa foram estabelecidas:

H1 – Somente com o resultado das buscas pela plataforma Google Trends não é possível estabelecer relações estatísticas que denotem causalidade.

H2 – Através do Google Trends é possível encontrar dados que suportem uma tomada de decisão assertiva.

Material e Método

Quanto à natureza, foi desenvolvida uma pesquisa aplicada, pois ela visa produzir conhecimento científico de forma prática para um determinado tema (Gil, 2021). É caracterizada como explicativa quanto ao seu objetivo (Gil, 2021). Mostra-se explicativa, pois há intenção de “esclarecer quais fatores contribuem, de alguma forma, para a



ocorrência de determinado fenômeno” (Vergara, 2016, p. 49). O método epistemológico utilizado para a construção do conhecimento foi o hipotético-dedutivo (Popper, 2013). Assim, foram conjecturadas hipóteses e depois testadas. Por fim, o procedimento adotado é o estudo de caso (Malhotra, 2006), no qual analisou-se um fictício empreendedor brasileiro que deseja abrir uma filial de sua fábrica de sorvetes nos Estados Unidos. Os dados referentes ao Google Trends (trends.google.com.br) foram coletados por meio desta própria plataforma usando-se filtros. Para os procedimentos estatísticos, foram desenvolvidos modelos de análise de correlação (Malhotra, 2006), utilizando o R por meio do RStudio para os cálculos. O R Core usado está na versão 4.1.1 (R Core Team, 2021) e o ambiente de desenvolvimento integrado RStudio na versão 2021.09.0+351 (RStudio Team, 2021).

Em função dos recursos estatísticos utilizados serem nativos do R, não houve a necessidade de uso de nenhuma biblioteca adicional. A correlação estatística no R é dada por - *cor.test*, a qual foi realizada entre os vetores *summer* e *ice cream*, ou seja, desta forma - *cor.test(summer, ice_cream)*. Enquanto isso, o teste de normalidade foi feito por meio do Shapiro-Wilk - *shapiro.test(summer)* e *shapiro.test(ice_cream)*.

Resultados e Discussão

Em vias de analisar o estudo de caso apresentado, a coleta dos dados focou em análises no Google Trends, nos últimos 5 anos (2020 até 2024) pelos termos “*summer*” e “*ice cream*” nos Estados Unidos, em todas as categorias. O resultado merece destaque, pois visualmente as linhas abaixo parecem seguir um padrão, ou seja, parecem estar correlacionadas.

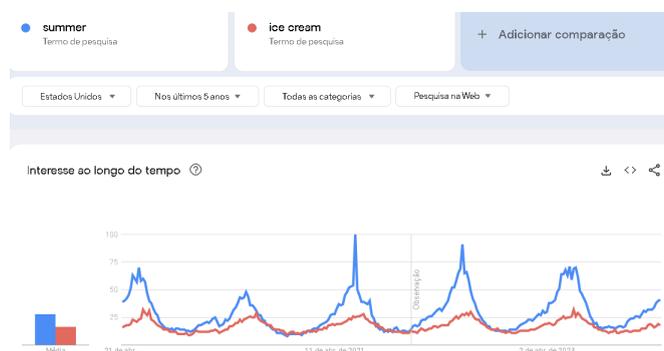


Figura 1 – Buscas no Google Trends pelas palavras *summer* e *ice cream*.

Fonte: Google Trends, 2024.



O termo correlação significa uma relação nos dois sentidos, ou seja, descreve a relação linear entre duas variáveis, não fazendo julgamento se há causa ou consequência estabelecida, definição contrária a uma regressão (Malhotra, 2006).

A análise mostrou que esses dados obtidos pelo Google Trends não obedecem a uma normalidade, ou seja, não são paramétricos. Para esse teste, foi verificado por meio do Shapiro-Wilk o valor de p , que no caso foi menor do que 0,05. Portanto, foi utilizado para análise o Rho de Spearman, que apresentou o valor de p igual à 0,001, o qual é considerado estatisticamente significativo por ser menor do que 0,05, dentro do intervalo de confiança utilizado de 95%. Considera-se então estatisticamente significativa a correlação, a qual apresenta-se no valor de 0,835. Esse valor pode variar entre -1 e $+1$, e no caso como o valor é positivo tem-se uma relação direta entre as variáveis. O valor de 0,835 se mostra como de intensidade alta (Hair Jr et al., 2005).

Segundo a análise desenvolvida, quanto maiores as buscas por *summer*, maiores também são as buscas por *ice cream*. Destaca-se assim, que a sazonalidade apresentada pelo Google Trends expõe uma série temporal, ou seja, a dinâmica das buscas se repete ao longo de determinado período de tempo, caracterizando-se como um processo estocástico (Nomelini, 2017). Além disso, foi possível observar que a relação entre as duas curvas obedece uma tendência, ciclo e sazonalidade, componentes esses presentes em uma série temporal (Morettin; Tolo, 2018). Assim, a tendência das curvas mostra o seu padrão entre as subidas e descidas. Enquanto o ciclo, permite observar que há uma variação ao longo dos períodos do ano (inverno e verão). A sazonalidade expõe a repetição das curvas em semelhantes períodos. Diante do exposto, uma das grandes vantagens das séries temporais está na possibilidade de realizar a previsão de situações, ou seja, uma análise preditiva. Assim, usando desta análise, foi reconhecido um padrão no resultado apresentado pelo Google Trends, permitindo que seja possível antever uma situação futura.

Ainda assim, é imperativo destacar que a análise da correlação entre as buscas por sorvete e verão (*summer* e *ice cream*), não permite afirmar, somente com base nos dados fornecidos pelo Google Trends, que há uma causalidade. Para ser possível inferir que há uma causalidade, há necessidade de analisar outros aspectos além daqueles fornecidos pelo Google Trends. Portanto, corroborando a hipótese H_1 , já que somente por meio do Google Trends não é possível assumir que há relação de causalidade.

Segundo Hunt (2010), há quatro condições que se mostram necessárias para assumir que há uma causalidade envolvida no estudo, são elas: variação concomitante, sequência temporal, uma relação não espúria e evidências teóricas por meio da literatura. Assim,



aprofundando nesse sentido em respeito aos achados, parece haver uma causalidade. Isso se dá, pois durante o verão com o aumento das temperaturas, há maior calor e por conta disso há uma busca por bebidas e comidas refrescantes (variação concomitante). Além disso, há uma temporalidade, ou seja, exclusão de uma hipótese alternativa. Ademais, a literatura denota que há relação entre os constructos analisados (Perrone, 2016; Zimmerman, 2022). Assim, se corrobora a hipótese H2, ao afirmar que utilizando o Google Trends é possível encontrar dados que suportem uma tomada de decisão mais assertiva.

Conclusão

O artigo expôs a importância dos dados secundários para a área da Ciência de Dados. Ainda que com cenários hipotéticos, foi possível demonstrar como trabalhar com os dados do Google Trends e análises estatísticas.

Embora existam limitações na ferramenta Google Trends, ela mostra-se de grande utilidade ao Cientista de Dados. Saliendo que essas limitações podem fazer com que outros dados também sejam necessários para responder as questões que estão sendo analisadas. Mas nem por isso, invalidando o resultado alcançado com o Google Trends, o qual coopera para a tomada de decisão.

Referências

- DAVENPORT, T. **Big Data no trabalho – Derrubando mitos e descobrindo oportunidades**. Editora Elsevier, 2014.
- GADZALO, Iaroslav et al. Assessment of global food demand in unexpected situations. **Innovative Marketing**, v. 16, n. 4, p. 91-103, 2020. DOI: 10.21511/im.16(4).2020.08.
- GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 7ª Edição. São Paulo: Editora Atlas, 2021.
- HAIR JR., J. F.; BABIN, Barry; MONEY, Arthur H.; SAMUEL, Phillip. **Fundamentos de Métodos de Pesquisa em Administração**. Porto Alegre: Bookman, 2005.
- KAPLAN, Robert e NORTON, David. **A estratégia em ação: Balanced scorecard**. 5ª edição. Rio de Janeiro: Editora Campus, 1999.
- LOLIC, I.; MATOSEC, M.; SORIC, P. DIY Google Trends indicators in social sciences: A methodological note. **Technology in Society**, v. 77, 2024, DOI: 10.1016/j.techsoc.2024.102477.
- MALHOTRA, N. K. **Pesquisa de marketing. Uma orientação aplicada**. Porto Alegre: Editora Bookman, 2006.
- MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de séries temporais: modelos lineares univariados**. 3ª Edição, v 1. São Paulo: Editora Blucher, 2018.
- NOMELINI, Quintiliano Siqueira Schroden; FERREIRA, Eric Batista; NOGUEIRA, Denismar Alves; GOLYNSKI, Anselmo Afonso; GOLYNSKI, Adelmo; VILLA, Thacyo Euqueres de. Uso de Modelagem Univariada e Multivariada com Séries Temporais como Ferramenta de Gestão do Agronegócio na Cultura de Soja do Brasil. **Revista Espacios**, 38-08, 2017.



PANTOVIC, V.; VIDOJEVIC, D.; VUJICIC, S.; SOFIJANIC, S.; JOVANOVIC-MILENKOVIC, M. Data-Driven Decision Making for Sustainable IT Project Management Excellence. **Sustainability**, Open Access, Volume 16, Issue 7, 2024, Article number 3014. DOI: 10.3390/su16073014.

PERRONE, H. Consumers' quality choices during demand peaks. **International Journal of Industrial Organization**, v. 44, p. 154-162, 2016. DOI: 10.1016/j.ijindorg.2015.11.002.

POPPER, Karl. **A lógica da pesquisa científica**. 2ª Edição. São Paulo: Editora Cultrix, 2013.

RSTUDIO Team (2021). **RStudio: Integrated Development Environment for R**. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

R Core Team (2021). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

SHARDA, D.; Delen, D; Turban, E. **Business Intelligence e Análise de Dados para gestão do negócio**: 4ª Edição. Porto Alegre: Editora Bookman, 2019.

SHARMA, D.; GUPTA, A.; MATEEN, A.; PRATAP, S. Making sense of the changing face of Google's search engine results page: an advertiser's perspective. **Journal of Information, Communication and Ethics in Society**, v. 16, n. 1, p. 90–107, 2018. DOI: 10.1108/jices-06-2017-0035.

VERGARA, Sylvia Constant. **Projetos e Relatórios de Pesquisa em Administração**. 16ª Edição. Editora Atlas, 2016.