

# Preferência e utilidade: O que os consumidores procuram em uma roupa da Zara?

Larissa Dolabella Gomide<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)  
– 31270-901 – Belo Horizonte – MG – Brazil

**Abstract.** *This study aims to explore how Game Theory can model consumer preferences and choices, influencing the sales performance of products. To achieve this, a framework is proposed that examines three possible methods to establish consumer preferences. This approach is exemplified through a practical case study utilizing a real dataset of products from Zara stores.*

**Resumo.** *Este estudo tem a finalidade de estudar como a Teoria dos Jogos pode modelar a preferência e as escolhas dos consumidores, de modo a resultar que um produto tenha mais vendas que o outro. Para isso é proposto um framework que estuda 3 possíveis métodos para estabelecer as preferências de consumo. Esta abordagem é exemplificada através de um estudo de caso prático que utiliza-se de uma base de dados real referente a produtos da loja Zara.*

## 1. Introdução

Com os recentes avanços da ciência da computação, em especial nas técnicas de inteligência artificial (IA), marcas de vários segmentos passaram a informatizar seus documentos e processos passando por uma transformação digital. Com o segmento de moda não foi diferente, segundo o relatório da [McKinsey and Company] 73% dos executivos da moda tem a IA generativa como prioridade para seus negócios em 2024. Segundo o mesmo relatório, a Shein vem modelando a demanda de tendências, para projetar e testar novos produtos, usando para isso dados de 'produtos virais' (que viralizaram nas redes sociais, por exemplo), tendências atuais e percepção do consumidor.

As indústrias de moda, ainda enfrentam o desafio de tentar prever tendências e as demandas de consumo anos ou meses antes das roupas chegarem às prateleiras. "O negócio médio da moda opera com prazos de entrega de 37 a 45 semanas, de acordo com o Boston Consulting Group. Isso significa que eles estão adivinhando todos os estilos, cores e tamanhos que os clientes comprarão com até 10 meses de antecedência."([Bain 2023], Tradução nossa). Nesse sentido, o uso de tecnologias de IA tem se mostrado promissoras para fazer essas previsões, conforme apresentado no artigo [Bain 2023], que discute o caráter probabilístico dos modelos e como eles podem ser usados de forma a agregar e trazer novos *insights* para as previsões humanas, evitando prejuízos e reduzindo riscos.

Tendo isso em vista, o presente trabalho consiste em responder a pergunta: quais critérios determinam que um produto tenha mais vendas que o outro? Para isso utilizaremos de uma série de conceitos relacionados a teoria dos jogos, e técnicas computacionais e estatísticas de modo a propor uma abordagem (*framework*) de como extrair e avaliar essas informações em bases de dados. Portanto, o objetivo principal deste trabalho é ajudar a prever a venda e/ou aceitação de uma peça de roupa antes dela entrar no mercado.

O trabalho está organizado da seguinte maneira. Na Seção 2.Trabalhos Relacionados, são apresentados artigos que fundamentam o trabalho, contextualizando a relação do problema com teoria dos jogos, além de apresentar um panorama das abordagens usadas para solucionar problemas semelhantes. Nas Seções 3.Dados e 4.Metodologia, descrevemos, respectivamente, o conjunto de dados utilizado e detalhamos o processo utilizado para a construção do *framework* proposto. Na Seção 5.Resultados, há um estudo de caso da contribuição do presente artigo, demonstrando uma aplicação da abordagem proposta e discutindo os resultados obtidos. Finalmente, na Seção 6.Conclusão, sintetiza-se as principais contribuições e resultados obtidos e propõe-se possíveis trabalhos futuros para aprimorar a abordagem.

## 2. Trabalhos Relacionados

### 2.1. Conceitos chave

Alguns conceitos são fundamentais para entender como consumidores tomam decisões e ajustam suas preferências diante de ofertas limitadas e preços variáveis. E portanto, devem estar claros para a melhor compreensão do presente estudo sobre previsão de demanda. São eles:

- **Agente Racional:** representação idealizada de uma entidade que realiza suas escolhas pautado em otimizar suas preferências.
- **Utilidade:** Dado um conjunto de preferências (por exemplo, prefiro chocolate a bala, matematicamente podemos escrever  $\text{chocolate} \succ \text{bala}$ ),  $U : X \rightarrow R$  representa a preferência se  $x, y \in X$ , então  $x \succ y \iff U(x) \geq U(y)$ . Deve ser:
  - Completa
  - Transitiva
- **Escolhas:** Modelam o comportamento do agente racional. Podem ser definidas como: "Dado um conjunto  $X$  de alternativas possíveis. Vemos um problema de escolha como um subconjunto não vazio de  $X$ , e nos referimos a uma escolha de  $A \subseteq X$  como especificando um dos membros de  $A$ ."([Rubinstein 2012]). Alguns dos critérios de escolhas incluem:
  - **Preferências lexicográficas:** Um ranking de critérios para efetuar escolhas. Quanto maior a prioridade do critério, mais ele se aproxima das primeiras posições do ranking.
  - **Satisfying Procedure:** melhor escolha dentre aquelas que apresentam critérios mínimos arbitrários.

### 2.2. Teoria dos Jogos e Preferências

O texto de [Baumol 2024] descreve como o conceito de 'utilidade' se relaciona com a dinâmica econômica. Ao longo do texto, o autor explica como a Terra em sua finitude tem recursos limitados, e portanto, todas as matérias primas e energia são recursos que são repartidos entre as pessoas e as indústrias. Ao escolher alocar um dado recurso, suponhamos por exemplo, o algodão para a produção de roupas esse mesmo recurso não pode ser usado como instrumento de higiene e cosmético.

Isso é a tragédia dos comuns, que segundo a Harvard Business School [Spiliakos 2019] refere-se a situação em que as pessoas ao ter acesso a um recurso e agindo conforme um agente racional podem depredar ou esgotar este recurso. E de acordo

com o raciocínio traçado por [Baumol 2024] uma forma de evitar que isso aconteça é associar um preço ao insumo, já que caso não houvesse essa limitação todos os recursos estariam sempre disponíveis e de graça. Como consequência surge a noção que quanto mais escasso um item mais caro ele deve ser. Mas ao mesmo tempo, ao introduzir a ideia de utilidade como o 'quão valioso ou preferível um item é para um consumidor' isso poderia gerar alguns paradoxos ao associar o valor de um produto com a utilidade percebida pelo consumidor. Nesse ponto o artigo deriva dois conceitos: a utilidade marginal e a utilidade ordinal. O primeiro mede a diferença na satisfação do cliente ao consumir uma unidade adicional de um bem ou serviço. Já o segundo foca na ordem das preferências ao invés da magnitude da utilidade. De maneira simplista poderíamos relacioná-los com a função de utilidade e escolhas lexicográficas, respectivamente. De todo modo, observa-se como a utilidade está relacionada à demanda de consumo, enquanto o recurso, por sua vez, determina a oferta.

Essa relação oferta-demanda, recursos e utilidade, foi estudada e mapeada para o denominado "Nash demand game"[Cabelka 2001] que por sua vez refere-se ao problema da barganha (Bargain Game) estudado por Nash em seus artigos de 1950 e 1953 (respectivamente, [Nash 1950] e [Nash 1953]). No presente contexto, que serão trabalhados dados referentes a loja Zara, podemos limitar o jogo de barganha ao problema do ultimato. Para isso vamos a uma breve contextualização.

A Zara é uma marca espanhola voltada para o mercado de massa e atua no segmento conhecido como *fast fashion*, e como tal apresenta uma premissa que não é válida para o segmento de luxo de moda: a loja quer maximizar os preços enquanto o cliente quer pagar o mínimo.

O jogo do ultimato, por sua vez, consiste em:

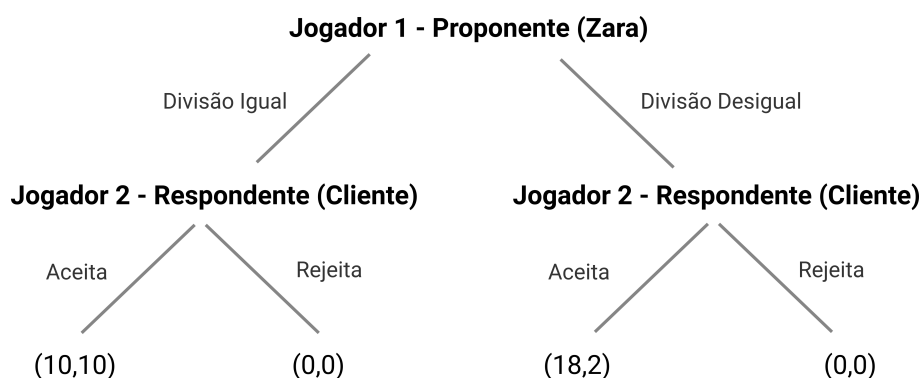
- Número de jogadores: 2 (Proponente e respondente)
- Estratégias:
  - **Proponente:** Escolhe uma divisão do recurso
  - **Respondente:** Escolhe aceitar ou rejeitar
- Se o respondente rejeitar a proposta, nenhum dos jogadores recebe nada

Logo, podemos simplificar a relação da Zara com seus clientes como um jogo do ultimato de dois jogadores de modo que:

- Número de jogadores: 2 (Proponente (Zara) e respondente (Cliente))
- Estratégias:
  - **Proponente:(Zara)** A loja faz uma oferta final ao cliente para um produto com um preço fixo.
  - **Respondente:(Cliente)** Escolhe aceitar ou rejeitar
- Caso o cliente rejeite, ele não efetua a compra. Portanto o cliente sai sem o produto e a loja sem a venda

Uma possível modelagem do problema seria a apresentada pela imagem abaixo:

Os *payoffs* ilustrados são uma possibilidade de saída. Ao observar a subárvore à esquerda, temos o cenário em que a Zara opta por uma divisão igual, neste caso é como se a loja considerasse seus custos de produção e seu lucro e fixasse um preço justo por suas peças. Caso o cliente opte por essa divisão, significa que o preço oferecido se aproxima



**Figura 1. Possíveis *payoffs* no Jogo do Ultimato**

do valor da utilidade marginal do cliente e portanto reconhece que os preços se adequam a sua percepção de custo benefício e de quanto ele está disposto a pagar por uma unidade adicional. Caso ele recuse é o cenário que apesar da empresa ter optado por um preço justo, esse preço provavelmente está além do que o consumidor está disposto a pagar. A empresa nesse caso, deve revisar sua cadeia produtiva e procurar minimizar seus custos de produção.

Na subárvore à direita, é apresentado o cenário que a Zara opta por uma maior margem de lucro. Caso o cliente aceite isso pode indicar um produto com características luxuosas, na qual a percepção de valor agregado do cliente é maior que o valor intrínseco do produto. Outra hipótese, é que o custo de produção está bem abaixo dos concorrentes, o que permite à empresa manter o valor do produto competitivo e aumentando sua margem de lucro. Caso o cliente rejeite, isso é sinal que o cliente percebeu que a relação custo-benefício está desproporcional. Logo, ao se tratar de um mercado de massa, o cliente irá atrás de propostas que se adequam a sua utilidade marginal e portanto tentará otimizar seus *payoffs* procurando por propostas mais justas.

Ao resolvermos o jogo apresentado utilizando o algoritmo de *Backward Induction* notamos que o ponto de equilíbrio (Equilíbrio de Nash) seria o Jogador 1 fazer uma divisão desigual e o Jogador 2 aceitar essa divisão, ou seja solução (18,2). O que reforça a ideia que os jogadores têm incentivos para diminuir custos e ainda manterem preços que se aproximam da utilidade marginal dos clientes e dos preços de mercado. No entanto, conforme apresentado no artigo de [Oosterbeek et al. 2004] há questões culturais e demográficas que influenciam a forma como as pessoas jogam e lidam com o Jogo do Ultimato. E conhecendo as peculiaridades de um mercado de massa, sabe-se que a tendência é que as empresas desse ramo tenham menores margens de lucro.

### 2.3. Trabalhos similares

Trabalhos recentes como "Demand Forecasting in the Fashion Industry: A Review"[Nenni et al. 2013] e "Demand forecasting for fashion products: A systematic review"[Swaminathan and Venkitasubramony 2024] apresentam os desafios de se prever demandas na indústria da moda. Dentre os principais pontos há o caráter sazonal dos produtos e seus ciclos de vida curtos. São apresentados um panorama de técnicas de predição que incluem desde modelos de Poisson até uso de técnicas de inteligência artificial. Em [Nenni et al. 2013] ainda é mencionado o artigo de [Bartezzaghi et al. 1999] que

se destaca pela proposta de um *framework* e sua respectiva simulação. Por fim o artigo [Capraro 2018] aborda a modelagem de preferências no contexto do Jogo do Ultimato - na sua versão original.

### 3. Base de Dados

A base de dados refere-se a loja de roupas Zara, e foi obtida através do Kaggle:<sup>1</sup>. Ela contém dados coletados no primeiro semestre de 2024, relativos a roupas vendidas pela marca, como número de produtos vendidos, preços, quais suas características, localização do produto entre outras características. Ao todo a base de dados contém 16 features e 252 produtos.

### 4. Metodologia

Como o objetivo do trabalho trata-se de entender sobre os fatores que influenciam as vendas da Zara, estamos tratando da parte da demanda de consumo, e para tratar dela focaremos nos conceitos de preferências e utilidades. Conforme tratado nos artigos apresentados na seção 2, existem vários fatores que determinam padrões de consumo como: demografia, cultura, alguns aspectos psicológicos como gatilhos de escassez e novidade, entre outros. Tendo isso em vista, além dos dados presentes na base original, optou-se por acrescentar dados que expressem aspectos visuais do produto. Já que, essa pesquisa se baseia nas seguintes hipóteses:

1. Preço: O cliente prefere pela peça de menor preço
2. Aspectos visuais: O cliente prefere peças bonitas e que valorizem sua imagem pessoal. No caso usaremos features genéricas de imagens através do modelo pré-treinado VGG 16, mas sugere-se que futuramente seja usado um modelo para extrair features mais concisas e do ponto de vista estético.
3. Localização na loja: É consenso na indústria e na academia que os padrões de consumo são influenciados pela forma como os produtos estão dispostos na loja. Os trabalhos de [Kempen et al. 2006] e do PROCON de Juiz de Fora [pjf 2024] reforçam essa ideia.

Tendo essas questões em vista, e assumindo que estamos tratando de um agente racional, podemos modelar suas demandas através de procedimentos de escolha através de Preferências Lexicográficas e do *Satisfying Procedure*, conforme supracitado. Entretanto, um agente racional não seria influenciado pela ordem que as opções são apresentadas ou por opções padronizadas, os chamados *default status* ou *status default* ([Rubinstein 2012]). No caso, tal influência é expressa pela disposição física dos elementos da loja. É como se o cliente ao ver a vitrine, ou uma roupa na entrada da loja fosse a escolha óbvia, ou um *status quo*. Portanto, devemos assumir os clientes como agentes econômicos que são influenciados pela forma como as opções são apresentadas.

Conforme mencionado no texto de [Rubinstein 2012] a opção padrão é vista de forma positiva, e tendo uma predileção ou tendência dos tomadores de decisão em permanecer nela. Uma forma de representar isso é através de funções de escolhas extendidas como por exemplo: a *Buridan's donkey* na qual se não houver uma alternativa que seja Pareto Ótima de Pareto dominar o *status default* o agente mantém sua escolha pelo *status*

---

<sup>1</sup><https://www.kaggle.com/datasets/xontoloyo/data-penjualan-zara/data>

*default*. Outra função de escolha estendida, seria a chamada *default bias*, na qual o agente escolhe a alternativa *default*, a menos que tenha uma alternativa que supere as vantagens da alternativa padrão.

Levando esses conceitos em consideração e os artigos apresentados na seção 2 apresenta-se a metodologia aplicada neste trabalho que inclui quatro etapas principais:

1. Preparação dos dados

- Análise de qualidade dos dados
- **Método:** Visualização dos dados e extração de características estatísticas, e curadoria através do Great Expectations.

2. Análise de Default Status

- **Dados Utilizados:** Localização dos produtos na loja e volume de vendas.
- **Método:** Análise de gráficos e cálculo de estatísticas, como a probabilidade condicional, ANOVA e *p-values*

3. Preferências Lexicográficas

- **Método:** Realizar pesquisas com consumidores ou aplicar o método Analytic Hierarchy Process (AHP) para determinar a ordem de prioridade dos atributos dos produtos.
- **Procedimento:** Extrair relações de preferência através de uma regressão linear. Os respectivos *p-values* são usados como métrica da importância relativa de um conjunto de critérios. No caso, filtramos pelos atributos que são significativos para a construção do modelo de regressão (*p-values* < 0,05)

4. *Satisfying Procedure*

- **Método:** Mineração de Padrões Frequentes.
- **Ferramenta:** Algoritmos de Apriori
- **Procedimento:** Extrair padrões frequentes das características dos produtos para identificar critérios mínimos comuns valorizados pelos consumidores. Correlacionando com o Sales Value

Em suma, a metodologia proposta é uma abordagem de como extrair de uma base de dados informações que possam ser analisadas da ótica dos procedimentos de escolha de um agente econômico. Os métodos utilizados foram selecionados de modo a dar embasamento estatístico/probabilístico a esta análise visando propiciar a predição de comportamentos de consumo e sua generalização para outros contextos. Na sessão de resultados apresentaremos detalhes da implementação da presente metodologia e das saídas obtidas.

## 5. Resultados

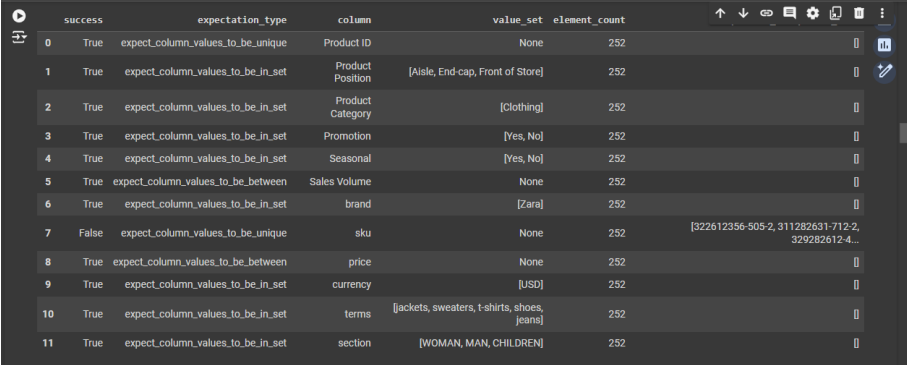
A presente sessão está organizada de acordo com a divisão apresentada na metodologia 4, e consiste em um estudo de caso efetuado na base de dados referente às vendas da Zara. O código elaborado ao longo do trabalho está disponível e pode ser acessado através do link do Google Colab<sup>2</sup>.

---

<sup>2</sup>[https://colab.research.google.com/drive/14\\_EdZmPmD2Pj2dO-C\\_bwHvlyQQujXzyy?usp=sharing](https://colab.research.google.com/drive/14_EdZmPmD2Pj2dO-C_bwHvlyQQujXzyy?usp=sharing)

## 5.1. Preparação dos dados

Antes de iniciar as análises de demanda de produtos da Zara, através de conceitos de Teoria dos Jogos. É necessário garantir que os dados estão devidamente limpos e preparados para a tarefa, sendo uma amostra pertinente e coerente com a realidade. Para isso, optou-se pelo uso da biblioteca Great Expectations (GE) que é responsável por validar a base de dados e garantir que os valores presentes nela estejam dentro dos parâmetros. Conforme ilustrado na imagem abaixo 2,,a base de dados não precisou de grandes ajustes. A coluna SKU referente ao *Stock Keeping Unit*, foi a única que retornou problemas, já que, conforme a própria descrição da base de dados, deveria conter apenas elementos únicos (uma espécie de ID). Como ela não afeta diretamente as análises propostas, ela foi descartada.



	success	expectation_type	column	value_set	element_count
0	True	expect_column_values_to_be_unique	Product ID	None	252
1	True	expect_column_values_to_be_in_set	Product Position	[Aisle, End-cap, Front of Store]	252
2	True	expect_column_values_to_be_in_set	Product Category	[Clothing]	252
3	True	expect_column_values_to_be_in_set	Promotion	[Yes, No]	252
4	True	expect_column_values_to_be_in_set	Seasonal	[Yes, No]	252
5	True	expect_column_values_to_be_between	Sales Volume	None	252
6	True	expect_column_values_to_be_in_set	brand	[Zara]	252
7	False	expect_column_values_to_be_unique	sku	None	252
8	True	expect_column_values_to_be_between	price	None	252
9	True	expect_column_values_to_be_in_set	currency	[USD]	252
10	True	expect_column_values_to_be_in_set	terms	[jackets, sweaters, t-shirts, shoes, jeans]	252
11	True	expect_column_values_to_be_in_set	section	[WOMAN, MAN, CHILDREN]	252

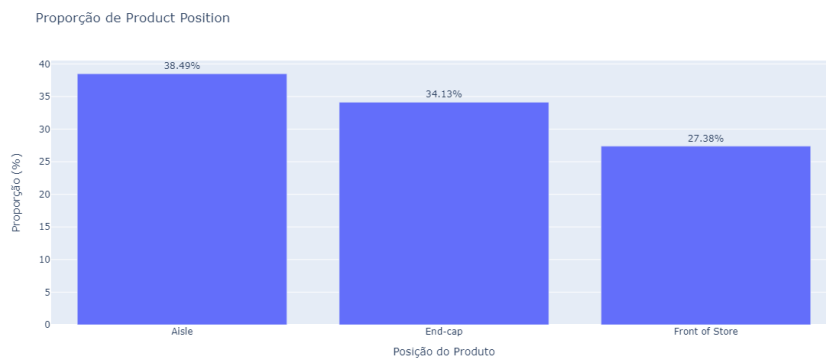
Figura 2. Resultados GE

### 5.1.1. Visualização dos dados

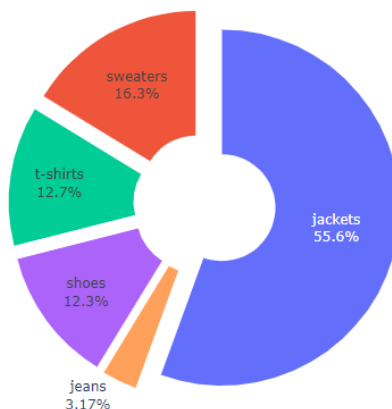
A fim de compreender melhor os dados, suas distribuições e proporções na composição da base de dados, foram elaboradas algumas visualizações. A seguir apresentamos três dos gráficos mais relevantes para o trabalho. Para mais detalhes veja a implementação completa no link do Google Colab.

Na figura 3, podemos observar quantos produtos estavam localizados em cada parte da loja. Nota-se que a maioria deles estava disposta no corredor (*Aisle*), seguidos da ponta de gôndola (*End-cap*) e da frente da loja (*Frente da loja*). O autor da base não deixa claro como esses dados foram coletados; no entanto, a formatação das colunas sugere que foi efetuada uma coleta de dados da loja virtual da Zara. Portanto, é provável que as categorias não estejam alinhadas com a realidade física. Caso essas informações se refiram à loja física, mas os dados de volume de vendas se refiram à loja online, não será possível ter uma análise precisa de como a disposição dos produtos impacta nas vendas e, conseqüentemente, no *Default Status*. De maneira análoga, é possível pensar que as categorias não capturam a lógica de disposição de produtos em um *e-commerce*, onde rótulos como *Homepage*, *Search* e afins parecem ser mais apropriados. Além disso, sites de *e-commerce* podem variar a distribuição e o posicionamento de um produto de acordo com o usuário, através de cookies de marketing e algoritmos de recomendação.

Já em 4 visualiza-se a proporção de palavras-chaves referente aos produtos. Nota-se pelo menos metade da base de dados se referem a jaquetas, estando a outra metade distribuída entre sweaters, camisetas, sapatos e jeans.



**Figura 3. Proporção de Product Position**



**Figura 4. Proporção de termos chaves dos produtos**

Por fim, na figura 5 é notável como que a maioria da base (86,51%) é constituída por produtos pertencentes a sessão masculina.

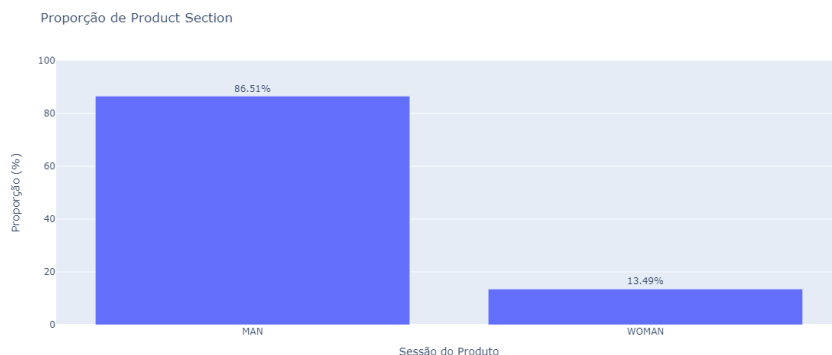
### 5.1.2. Preparação de dados

Para fazer análises que levassem em conta os atributos estéticos e visuais, foi necessário a elaboração de um *web crawler*, para coletar as imagens dos respectivos produtos na base. Para isso, desenvolveu-se o *script* disponível no GitHub <sup>3</sup>, e utilizou-se dos links presentes na tabela para fazer a extração das imagens correspondentes a cada instância. Cada link continha aproximadamente 6 fotos, além disso, algumas continham vídeos ou outras formas de interação e visualização do produto. Após a coleta, a base necessitou de uma limpeza mais fina executada manualmente, e por fim foi salva na pasta de Dados, disponível no repositório do GitHub supracitado.

Essa base é anexada a tabela original, a medida que se faz o processamento das imagens e respectiva extração das características através do modelo VGG 16, pré-treinado

<sup>3</sup><https://github.com/LarissaDG/Mestrado>





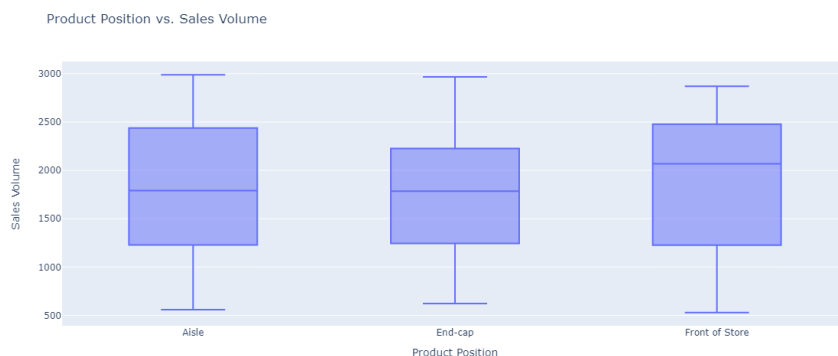
**Figura 5. Proporção de produtos por sessão (Masculina e Feminina)**

na base de dados denominada ImageNet. Como cada produto tem mais de uma imagem, novas linhas são criadas ao fazer um produto cartesiano das imagens (e suas features visuais) com as features originais da base do Kaggle. Ao final desse procedimento, é necessário aplicar um PCA (Principal Component Analysis), pois a extração de features do VGG 16 gera 25088 colunas, muitas delas com valores próximos ou iguais a zero.

Assim, ao final desse processo obtemos uma base de dados de 58 colunas e 1945 linhas. Além disso, alguns pré-processamentos foram realizados de modo a deixar os dados preparados para serem usados nos modelos de regressão de mineração de padrões frequentes.

## 5.2. Análise de *Default Status*

Para construção dessa etapa pretendeu-se responder à seguinte questão: como o volume de vendas é influenciado pela disposição dos produtos na loja. Para isso, em primeiro lugar, plotou-se os box plots apresentados na figura 6, a fim de observar a variação do volume de vendas por ponto de localização do produto. Nota-se que a ponta de gôndola apresentou menor variação no volume de vendas que os demais pontos. No entanto, a distribuição dos valores de volume de vendas é mais uniforme na categoria *Aisle*. Não há presença de *outliers* em nenhum dos pontos. E a frente da loja apresenta a maior mediana.



**Figura 6. Box Plot distribuição do volume de vendas por ponto da loja**

Em seguida, visualizou-se como é o volume de vendas total por pontos de vendas

por ponto, apresentado em 7. Nota-se que o corredor apresenta maior volume de vendas totalizando 177.396. Seguida da ponta de gôndola com 152.930 e a frente da loja com 129.247. A partir desses dados, calculou-se a probabilidade condicional de uma venda ocorrer por ponto de venda. Os resultados são apresentados na tabela 1

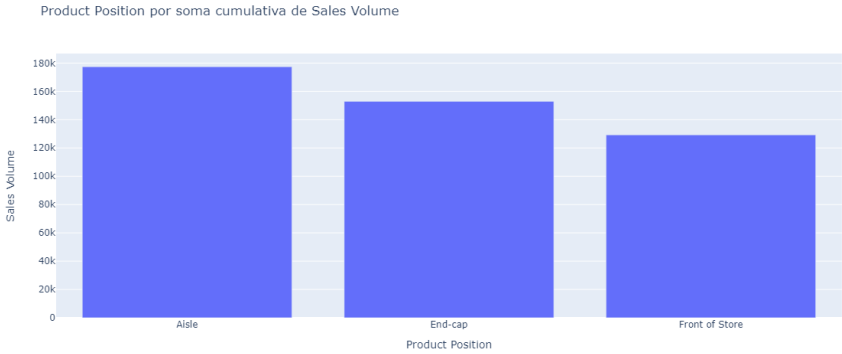


Figura 7. Venda acumulada de produtos por ponto

Posição do Produto	Volume de Vendas	Probabilidade (%)
Aisle/Corredor	177.396	38,60
Ponta de Gôndola/End-cap	152.930	33,28
Frente da loja/Front of store	129.247	28,12

Tabela 1. Tabela de probabilidades condicionais.

E por fim, para avaliar se há uma relação entre a disposição dos produtos nessas categorias, com a quantidade vendida, aplicou-se o ANOVA. Ele é responsável por fazer a análise da variância e gerar os *p-values*. Como resultado do ANOVA obtivemos os valores apresentados na tabela a seguir<sup>2</sup>. Nas linhas o valor de C(Product\_Position) refere-se a parte do modelo explicada pela variável dependente: volume de vendas. Já o Residual, refere-se às variações que não podem ser explicadas pelo modelo. Nas colunas temos os cálculos da soma dos quadrados (sum\_sq), df representa os graus de liberdade, F a estatística F que é usada para testar hipótese nula, e por fim PR(>F) representa o *p-value*. Conforme ilustrado, podemos concluir através do *p-value* que é maior do que 0,05 que não há relação entre a posição e o número de vendas. O que, por sua vez, não corresponde à nossa hipótese inicial.

Source	sum_sq	df	F	PR(>F)
C(Product_Position)	3.488437e+05	2.0	0.356473	0.700498
Residual	1.218356e+08	249.0	NaN	NaN

Tabela 2. Resultados da ANOVA para a relação entre posição do produto e volume de vendas.

### 5.3. Preferências Lexicográficas

Para a análise de preferências lexicográficas utilizou-se do algoritmo de *Analytic Hierarchy Process* (AHP) que tem por objetivo auxiliar na tomada de decisão e justificar a escolha. Ele consiste nos seguintes passos:

1. **Passo 1:** Definição de objetivos

- Selecionar o produto preferido para compra.

2. **Passo 2:** Estrutura da Hierarquia Critérios:

- Selecionar os critérios e os respectivos pesos para cada critério.

No caso, utilizamos de uma regressão linear para determinar os critérios e os seus respectivos pesos. Pois através dela podemos determinar, através dos respectivos *p-values* a importância relativa das variáveis para a construção do modelo. Nesse caso, filtramos pelas variáveis que apresentavam um *p-values*  $\leq 0.05$  e portanto são importantes para o modelo. No entanto, para a nossa modelagem de preferência queremos que o menor *p-values* apresente maior importância. Tendo isso em vista, e sabendo que os valores dessa variável estão entre 0 e 1, subtraímos 1 dos *p-values* para encontrar o valor relativo das variáveis.

3. **Passo 3:** Cria matriz comparações Pairwise

- Realiza comparações pairwise para determinar a importância relativa dos critérios. Ao final obtemos uma matriz como a ilustrada abaixo:

Matriz de comparação pairwise:

	price	section	pca_9	pca_21	pca_22	pca_29	pca_33	pca_47
price	1.000	1.000	1.012146	1.016260	1.009082	1.010101	1.011122	1.025641
section	1.000	1.000	1.012146	1.016260	1.009082	1.010101	1.011122	1.025641
pca_9	0.988	0.988	1.000000	1.004065	0.996973	0.997980	0.998989	1.013333
pca_21	0.984	0.984	0.995951	1.000000	0.992936	0.993939	0.994944	1.009231
pca_22	0.991	0.991	1.003036	1.007114	1.000000	1.001010	1.002022	1.016410
pca_29	0.990	0.990	1.002024	1.006098	0.998991	1.000000	1.001011	1.015385
pca_33	0.989	0.989	1.001012	1.005081	0.997982	0.998990	1.000000	1.014359
pca_47	0.975	0.975	0.986842	0.990854	0.983855	0.984848	0.985844	1.000000

Figura 8. Matrix Pairwise

4. **Passo 4:** Cálculo dos Pesos

- Normalizar a matriz Pairwise, dividindo cada valor de uma coluna pela soma dos valores dessa coluna.
- Verificação de Consistência através do cálculo da Razão de Consistência definido por:

$$CR = \frac{CI}{RI}$$

Em que *CR* é a Razão de Consistência, *CI* é o Índice de Consistência e *RI* o Índice de Consistência Aleatório. No caso, obtemos  $CR = 0$  indicando que não há inconsistência aleatória esperada na comparação dos critérios, ou seja, a matriz de comparação pairwise é perfeitamente consistente, sem erros aleatórios esperados.

- Tirar média aritmética dos valores de cada linha para obter a matriz de pesos

5. **Passo 5:** Avaliação das Alternativas

- Avalie cada produto em relação a cada critério. Isso pode envolver normalizar os valores dos critérios, como o preço, ou usar métodos qualitativos para critérios categóricos.

#### 6. Passo 6: Cálculo da Pontuação Final

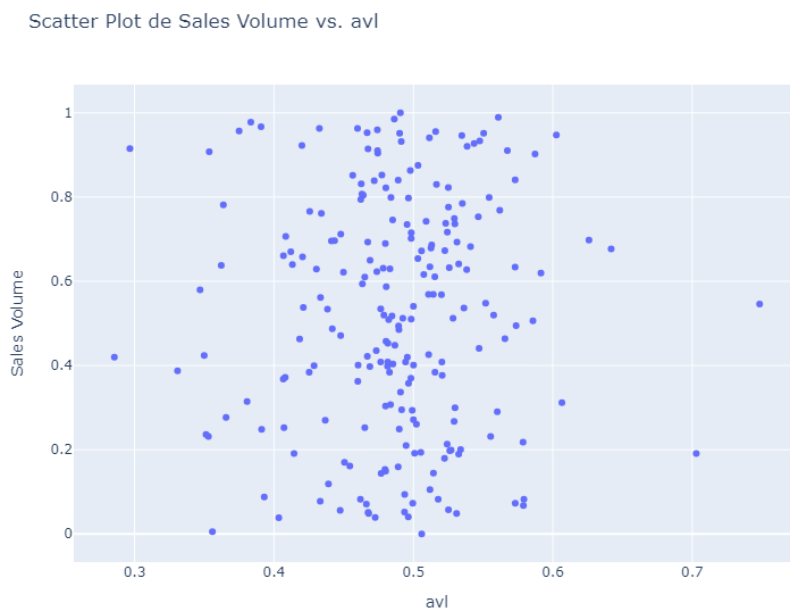
- Multiplique as avaliações normalizadas das alternativas pelos pesos dos critérios e some os resultados para obter a pontuação final de cada produto. (Média ponderada). Ordenando do que tem maior pontuação para o que tem menor.

No caso, os 5 produtos mais bem classificados, da direita para a esquerda, foram:



**Figura 9. Top 5 produtos mais bem classificados segundo o AHP**

Apesar do método ter se mostrado uma forma interessante de simular a tomada de decisão através das preferências lexicográficas, não é possível traçar uma correlação entre a pontuação avaliada pelo AHP e o volume de vendas. O gráfico da figura 10 ilustra isso:



**Figura 10. Gráfico de pontuação da avaliação (avl) por volume de vendas (Sales Volume)**

#### 5.4. Satisfying Procedure

Para ilustrar o *Satisfying Procedure* optamos por utilizar uma mineração de padrões frequentes a fim de determinar quais os requisitos/características costumam aparecer juntos na base de dados. Pois se tais critérios costumam aparecer juntos podem significar que são relevantes na tomada de decisão, algo como, por exemplo, o produto deve ser uma jaqueta, estar em promoção e ser masculina. No caso, utilizou-se de um *support* de 0.2, ou seja, um conjunto para ser considerado frequente deve aparecer em pelo menos 20% da base de dados. Na primeira execução do algoritmo obtivemos os resultados apresentados na figura 11. Observe que, conforme mostrado na parte de visualização na etapa de preparação dos dados, 'section=MAN' representa 80% da base de dados. Logo, a mineração de padrões frequentes retornou consequentes apenas de 'section=MAN' indicando, por exemplo, que dado que apareceu o termo 'Product Position=Aisle' tem grandes chances de aparecer 'section=MAN' com uma probabilidade de quase 90% (vide coluna 'confidence'), o que é óbvio. Portanto essa abordagem não trouxe resultados interessantes.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
6	(Product Position_Aisle)	(section_MAN)	0.384921	0.865079	0.345238	0.896907	1.036792
11	(Promotion_No)	(section_MAN)	0.523810	0.865079	0.456349	0.871212	1.007089
16	(Seasonal_No)	(section_MAN)	0.492063	0.865079	0.444444	0.903226	1.044096
20	(terms_jackets)	(section_MAN)	0.555556	0.865079	0.555556	1.000000	1.155963
22	(terms_jackets, Product Position_Aisle)	(section_MAN)	0.214286	0.865079	0.214286	1.000000	1.155963
30	(Seasonal_No, Promotion_No)	(section_MAN)	0.257937	0.865079	0.234127	0.907692	1.049259
37	(terms_jackets, Promotion_No)	(section_MAN)	0.285714	0.865079	0.285714	1.000000	1.155963
40	(Seasonal_No, Promotion_Yes)	(section_MAN)	0.234127	0.865079	0.210317	0.898305	1.038408
44	(terms_jackets, Promotion_Yes)	(section_MAN)	0.269841	0.865079	0.269841	1.000000	1.155963
50	(terms_jackets, Seasonal_No)	(section_MAN)	0.269841	0.865079	0.269841	1.000000	1.155963
54	(terms_jackets, Seasonal_Yes)	(section_MAN)	0.285714	0.865079	0.285714	1.000000	1.155963

Figura 11. Tabela resultados Apriori todos os dados

Visando contornar tal problema, os dados são filtrados pela coluna sessão de modo a criar dois pandas DataFrames. O primeiro se refere aos dados que apresentam 'section=MAN' e o segundo com 'section=WOMAN', e removi a coluna 'section' da análise de padrões frequentes. Dessa forma é possível extrair relações mais interessantes já que elimina o desbalanceamento provocado por essa coluna e cria dois perfis de consumidores, cujas regras de associação podem ser modeladas através do algoritmo Apriori.

Observe, em primeiro lugar, os resultados das regras referente aos dados de 'section=MAN'. Observe que agora que a mineração retornou regras de associação com valores de *lift* próximos de 1, indicando não haver relações fortes e surpreendentes entre os itens no antecedente com relação ao consequente. Note ainda, que em nenhuma das regras a 'confidence' obteve valores maiores que 70%.

Repare, agora nos resultados das regras referente aos dados de 'section=WOMAN'. Observe que agora que a mineração retornou várias regras de associação com valores de *lift* iguais a 1, indicando não haver relações fortes e surpreendentes entre os itens no antecedente com relação ao consequente. Bem como valores de 'confidence' iguais a 1, também. Indicando, por sua vez, que apesar de sempre que ocorre o antecedente aparecer o consequente, essas características são independentes. Além disso, há

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(Product Position_Aisle)	(Promotion_No)	0.399083	0.527523	0.211009	0.528736	1.002299
1	(Promotion_No)	(Product Position_Aisle)	0.527523	0.399083	0.211009	0.400000	1.002299
2	(Seasonal_Yes)	(Promotion_No)	0.486239	0.527523	0.256881	0.528302	1.001477
3	(Promotion_No)	(Seasonal_Yes)	0.527523	0.486239	0.256881	0.486957	1.001477
4	(Seasonal_No)	(Promotion_Yes)	0.513761	0.472477	0.243119	0.473214	1.001560
5	(Promotion_Yes)	(Seasonal_No)	0.472477	0.513761	0.243119	0.514563	1.001560
6	(terms_jackets)	(Promotion_Yes)	0.642202	0.472477	0.311927	0.485714	1.028017
7	(Promotion_Yes)	(terms_jackets)	0.472477	0.642202	0.311927	0.660194	1.028017
8	(terms_jackets)	(Seasonal_Yes)	0.642202	0.486239	0.330275	0.514286	1.057682
9	(Seasonal_Yes)	(terms_jackets)	0.486239	0.642202	0.330275	0.679245	1.057682

**Figura 12. Tabela resultados Apriori em que 'section=MAN'**

relações em que o *lift* é um pouco maior que 1 e 'confidence' igual a 90%. Na imagem 13 são apresentados três casos assim:

- Product\_Position\_Aisle → Seasonal\_Yes
- {Product\_Position\_Aisle, terms\_sweaters} → Seasonal\_Yes
- Seasonal\_Yes → {Product\_Position\_Aisle, terms\_sweaters}

Observe que as duas últimas são simétricas, o que por sua vez indica uma associação bidirecional, além de uma provável correlação entre os termos. Logo, quando um produto está no corredor (*Aisle*) e é um sweater há uma probabilidade de 90% dele ser classificado como sazonal. De maneira similar dado que o produto é sazonal há 90% de chance dele estar no corredor e ser um sweater.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(Product Position_Aisle)	(Seasonal_Yes)	0.294118	0.647059	0.264706	0.9	1.390909
2	(Product Position_Aisle)	(terms_sweaters)	0.294118	1.000000	0.294118	1.0	1.000000
7	(Product Position_End-cap)	(terms_sweaters)	0.382353	1.000000	0.382353	1.0	1.000000
8	(Product Position_Front of Store)	(terms_sweaters)	0.323529	1.000000	0.323529	1.0	1.000000
13	(Promotion_No)	(terms_sweaters)	0.500000	1.000000	0.500000	1.0	1.000000
17	(Promotion_Yes)	(terms_sweaters)	0.500000	1.000000	0.500000	1.0	1.000000
18	(Seasonal_No)	(terms_sweaters)	0.352941	1.000000	0.352941	1.0	1.000000
21	(Seasonal_Yes)	(terms_sweaters)	0.647059	1.000000	0.647059	1.0	1.000000
22	(Product Position_Aisle, terms_sweaters)	(Seasonal_Yes)	0.294118	0.647059	0.264706	0.9	1.390909
23	(Product Position_Aisle, Seasonal_Yes)	(terms_sweaters)	0.264706	1.000000	0.264706	1.0	1.000000
25	(Product Position_Aisle)	(terms_sweaters, Seasonal_Yes)	0.294118	0.647059	0.264706	0.9	1.390909

**Figura 13. Gráfico de pontuação da avaliação (avl) por volume de vendas (Sales Volume)**

## 6. Conclusão

Apesar dos dados e análises efetuadas, não foi possível detectar relações que evidenciam fortes relações entre antecedente e consequente, em especial se tratando de 'section=MAN'. No caso de 'section=WOMAN', poderíamos dizer que os critérios de ser

sazonal, suéter e corredor apresentam uma relação interessante. No que se refere ao volume de vendas, aplicou-se o seguinte procedimento:

1. Calcula o impacto médio no volume de vendas para cada regra
2. Codifica (one-hot-encoding) as características para análise
3. Utiliza a correlação de Pearson para identificar a relação entre as características codificadas e o impacto no volume de vendas.

Neste caso, observando a correlação calculada, as variáveis mais interessantes são apresentadas na tabela 3. Note como que a correlação de um produto estar em promoção apresenta uma correlação positiva tanto para homens quanto para mulheres, mas apresentando uma correlação maior com o número de vendas dentro do público feminino, agora se o produto não está em promoção ele não impacta as vendas de forma negativa, como impacta para o público feminino.

Já a disposição dos produtos em diferentes pontos tem relações de Pearson distintas para homens e mulheres quando se trata da frente da loja (*front-of-store*), havendo um efeito negativo no volume de vendas para homens. Já as pontas de gôndolas têm relações próximas entre homens e mulheres. E o corredor (*aisle*) tende a reduzir o volume de vendas para ambos os sexos, com um impacto mais forte para mulheres.

Variável	Homem (Correlação)	Mulher (Correlação)
Promotion_Yes	0.236689	0.721496
Promotion_No	0.086459	-0.310754
Product Position_Aisle	-0.330559	-0.580341
Product Position_End-cap	0.577659	0.443557
Product Position_Front of Store	-0.742527	0.221447

**Tabela 3. Correlação entre a posição do produto e o volume de vendas para homens e mulheres.**

Nesta seção aplicamos os conceitos de Teoria dos Jogos em uma base de dados real, a fim de mostrar um estudo de caso utilizando o *framework* elaborado. Na aplicação envolvendo o *Default Status* e as Preferências Lexicográficas, não foi possível evidenciar correlações nos dados em relação ao volume de vendas. Já na etapa de demonstração do *Satisfying Procedure* não foi possível gerar relações de critérios expressivas, em especial para os homens. Já para as mulheres podemos dizer que entre os principais critérios está a sazonalidade e o fato de ser um sweater. Ao olhar as métricas de Pearson das associações com relação ao volume de vendas fica evidente a preferência das mulheres por promoções, a rejeição dos homens por produtos na frente da loja e a similaridade do comportamento com relação às peças em pontas de gôndolas.

## 6.1. Conclusão

Por fim, através do presente trabalho, foi possível concluir e estabelecer uma metodologia para se trabalhar com conceitos de Teoria dos Jogos, no contexto de análise de demandas de consumo e estabelecer relações de preferência. Destaca-se em especial o resultado promissor alcançado no uso do AHP como simulação das preferências lexicográficas. Além disso, conforme esperado nas premissas e demonstrado na parte de preferências

lexicográficas o preço e atributos visuais são relevantes para estabelecer critérios ao escolher produtos. Já a premissa da influência do local onde o produto é disponibilizado, no geral, não confirmou a hipótese no que concerne ao volume de vendas.

Como trabalhos futuros, propõe-se fazer análises comparativas com outros nichos de mercado de moda. Por exemplo, a Zara, a qual a base de dados se refere é uma marca *fast fashion* voltada para o mercado de massa, poderia ser comparada com dados de uma marca como a Louis Vitton do segmento de luxo. Além disso, pretende-se utilizar modelos capazes de extrair features estéticas, ou características como cor, forma e simetria para analisar como a beleza influencia as preferências dos consumidores e seus comportamentos de consumo.

Do ponto de vista de modelagem, propõe-se categorizar o preço dentro de faixas para incorporá-lo à análise de padrões frequentes. Além disso, é possível estudar outras correlações entre as variáveis de modo a estabelecer um parâmetro mais acurado para o *Default Status*. E, uma vez feito isso, estudar funções de escolha estendidas como as mencionadas *Buridan's donkey* e *default bias*. De modo a compreender os limites que fazem com que o agente econômico mude de ideia sobre suas preferências.

## Referências

- [pjf 2024] (2024). Produtos de supermercados e decisão de compra. [https://www.pjf.mg.gov.br/administracao\\_indireta/procon/pesquisas/arquivos/produtos\\_supermercados\\_e\\_decisao\\_de\\_compra.pdf](https://www.pjf.mg.gov.br/administracao_indireta/procon/pesquisas/arquivos/produtos_supermercados_e_decisao_de_compra.pdf). Acessado em: 23 de junho de 2024.
- [Ahlstrom 2023] Ahlstrom, L. (2023). Ultimatum game. <https://inomics.com/terms/ultimatum-game-1538668>. Economics Terms A-Z. Acessado em: 23 de junho de 2024.
- [Bain 2023] Bain, M. (2023). Can ai predict what shoppers will buy? *Business of Fashion*. Acessado em: 23 de junho de 2024.
- [Bartezzaghi et al. 1999] Bartezzaghi, E., Verganti, R., and Zotteri, G. (1999). A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*, 59(1-3):499–510.
- [Baumol 2024] Baumol, W. J. (2024). Utility and value. <https://www.britannica.com/money/utility-economics>. Acessado em: 23 de junho de 2024.
- [Bocconi ] Bocconi, U. C. L. Curso coursera - bocconi. <https://www.coursera.org/learn/mafash>. Acessado em: 23 de junho de 2024.
- [Brasil 2018] Brasil, K. A. (2018). Tragédia dos comuns — microeconomia — khan academy. <https://www.youtube.com/watch?v=bT5E62UKhhc>. Acessado em: 23 de junho de 2024.
- [Cabelka 2001] Cabelka (2001). The nash demand game revisited: A logit-equilibrium approach. *CERGE-EI Working Paper Series*. First version: August 2000. This version: March 2001. A joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences.
- [Capraro 2018] Capraro, V. (2018). Social versus moral preferences in the ultimatum game: A theoretical model and an experiment. *arXiv preprint arXiv:1804.01044*.



- [Kaggle ] Kaggle. Zara sales data. <https://www.kaggle.com/datasets/xontoloyo/data-penjualan-zara/data>. Acessado em: 23 de junho de 2024.
- [Kempen et al. 2006] Kempen, E., Merwe, D. V. d., and Sonnenberg, N. (2006). The effect of fashion store layout and misual merchandising on female consumer walking patterns: a systems perspective. *ACR Asia-Pacific Advances*.
- [McKinsey and Company ] McKinsey and Company. State of fashion. [https://www.mckinsey.com/industries/retail/our-insights/state-of-fashion#/.](https://www.mckinsey.com/industries/retail/our-insights/state-of-fashion#/) Acessado em: 23 de junho de 2024.
- [Nash 1950] Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49. Acessado em: 23 de junho de 2024.
- [Nash 1953] Nash, J. (1953). The bargaining problem. *Econometrica*, 18(2):155–162. Acessado em: 23 de junho de 2024.
- [Nenni et al. 2013] Nenni, M. E., Giustiniano, L., and Pirolo, L. (2013). Demand forecasting in the fashion industry: a review. *International Journal of Engineering Business Management*, 5:37.
- [Oosterbeek et al. 2004] Oosterbeek, H., Sloof, R., and Van De Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental economics*, 7:171–188.
- [Rubinstein 2012] Rubinstein, A. (2012). *Lecture Notes in Microeconomic Theory: The Economic Agent*. Princeton University Press, second edition.
- [Spiliakos 2019] Spiliakos, A. (2019). The tragedy of the commons: Impact on sustainability issues. <https://online.hbs.edu/blog/post/tragedy-of-the-commons-impact-on-sustainability-issues>. Acessado em: 23 de junho de 2024.
- [Swaminathan and Venkitasubramony 2024] Swaminathan, K. and Venkitasubramony, R. (2024). Demand forecasting for fashion products: A systematic review. *International Journal of Forecasting*, 40(1):247–267.