

# PROJETO DE MINERAÇÃO DE DADOS

UNIVERSIDADE ESTÁCIO DE SÁ

VICE-REITORIA DE PÓS-GRADUAÇÃO, PESQUISA E EXTENSÃO

Marcos Paulo Silvério Alves Nery de Souza

Rio de Janeiro

2025

# Introdução: O Desafio de Negócio

O supermercado SuperVida está lançando o "Programa Alimentação Saudável". A meta é incentivar hábitos alimentares mais nutritivos, combatendo a insegurança alimentar e a obesidade em comunidades locais. Para isso, o SuperVida vai firmar parceria com três ONGs focadas em diferentes públicos:

## ONG "Sabor da Horta"

Atende famílias de baixa renda no bairro Santa Rita. Eles se concentram em distribuir cestas básicas com produtos frescos (frutas, verduras, legumes) e educar sobre o uso de alimentos não-processados.

## ONG "Jovens em Movimento"

Atua no bairro Belo Horizonte, com foco em adolescentes e jovens adultos. Seu objetivo é combater o consumo excessivo de fast-food e bebidas açucaradas, promovendo escolhas saudáveis e a prática de esportes.

## ONG "Melhor Idade"

Presta serviço no bairro Vila Nova, atendendo idosos. O foco é a nutrição adequada para a saúde óssea e cardiovascular, com ênfase em produtos ricos em cálcio, fibras e ômega-3.

O SuperVida vai criar um programa de fidelidade especial para os clientes que se cadastrarem. Ao comprar produtos saudáveis (frutas, verduras, cereais integrais, etc.), esses clientes acumularão pontos que se converterão em doações para a ONG de sua escolha. Além disso, eles receberão descontos exclusivos nesses produtos.

# Resumo Executivo

A presente análise de mineração de dados teve como objetivo identificar o público ideal para o novo Programa de Alimentação Saudável da Rede SuperVida. Por meio de Análise Exploratória de Dados (AED) e do algoritmo de agrupamento K-Means, foi possível segmentar a base de clientes e identificar um cluster de alto potencial.

Este grupo se destaca por seu alto engajamento com produtos saudáveis, maior ticket médio e alta frequência de compra. As conclusões a seguir detalham o perfil desse público, fornecendo um caminho claro para o sucesso e a fidelização de clientes.

## Informações sobre a Base de Dados

Base de Dados: Origem e Tratamento. A base de dados utilizada neste projeto é fictícia, criada e modificada especificamente para fins de análise e segmentação de clientes. Apesar de não conter dados reais de consumidores, ela foi estruturada para simular um cenário de negócio real da SuperVida. O tratamento desses dados, incluindo a padronização e o ajuste de informações como bairros e CEPs para o Rio de Janeiro, foi uma etapa crucial para garantir a qualidade da análise.

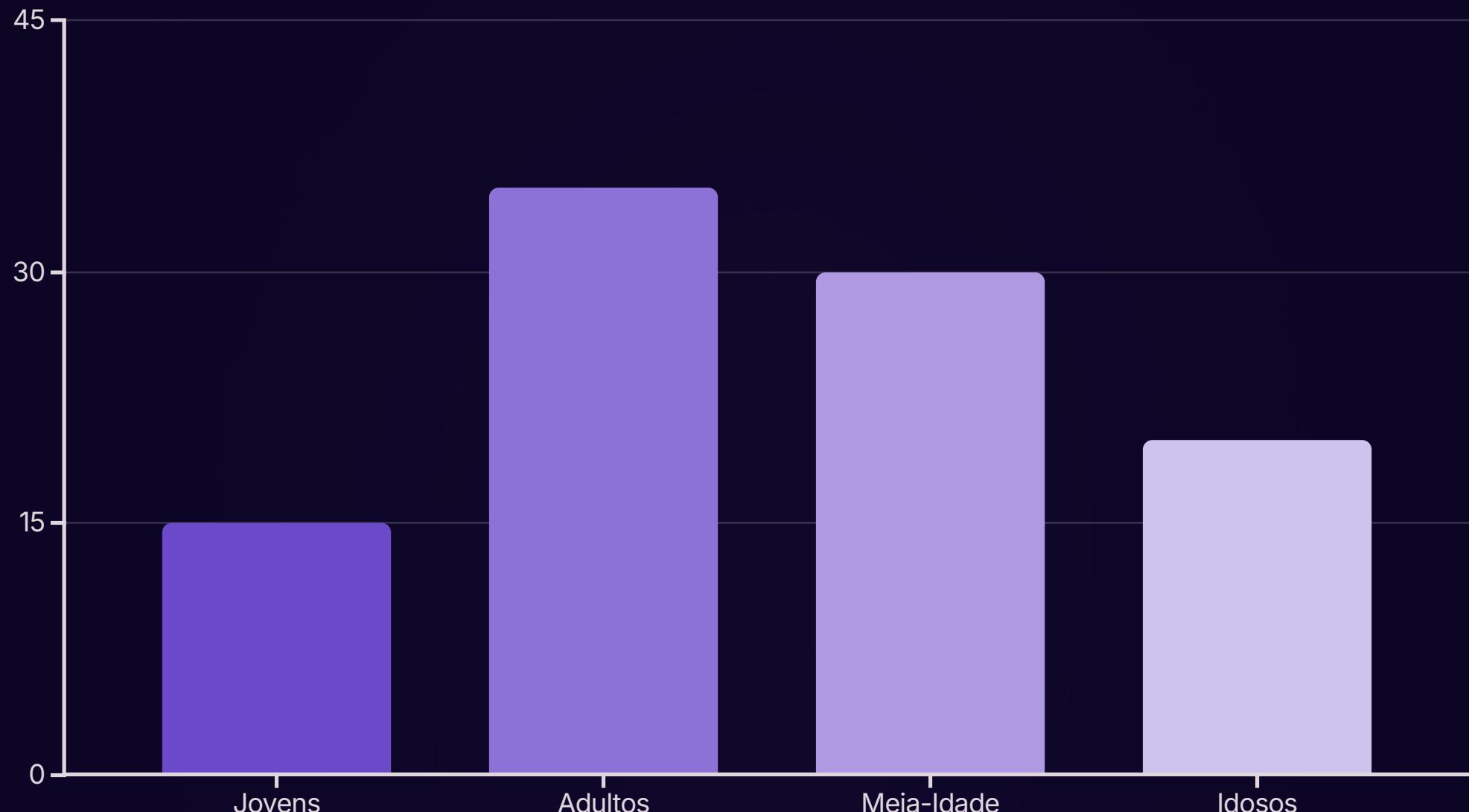
# Análise Demográfica: quem é o nosso cliente?

## Perfil por Gênero

A análise demográfica mostra que a nossa base de clientes é bem equilibrada, com uma distribuição quase igual entre clientes masculinos e femininos. Há também uma pequena, mas significativa, parcela de clientes que se identificam como Outros. Esta diversidade de gênero na base reforça a nossa capacidade de atender a diferentes públicos, e a categoria "Outros" pode ser um campo fértil para futuras análises de nicho.

A base de clientes da SuperVida é notavelmente equilibrada em termos de gênero, com uma distribuição quase igual entre clientes masculinos e femininos. Além disso, a presença de uma pequena, mas importante, parcela de clientes no grupo 'Outros' reforça a diversidade da nossa base e pode ser um campo fértil para futuras ações de marketing direcionadas a nichos específicos.

# Perfil por Faixa Etária



A distribuição etária dos nossos clientes é bem variada, com uma concentração notável nas faixas de Adultos e Meia-Idade. Este resultado é ideal para o nosso programa de alimentação saudável, pois este público já tende a ter uma maior consciência sobre saúde e bem-estar.

A distribuição etária dos nossos clientes revela uma forte concentração nas faixas de Adultos e Meia-Idade. Este perfil é extremamente promissor, pois coincide com o público que tipicamente demonstra maior interesse e poder de compra para investir em saúde e bem-estar.

# Perfil por Bairro



A nossa base de clientes está predominantemente concentrada em três bairros estratégicos: Copacabana, Tijuca e Barra da Tijuca. Essa concentração geográfica é uma vantagem competitiva, pois permite que a nossa estratégia de marketing e logística seja otimizada para essas regiões, garantindo uma cobertura eficiente e de alto impacto.

Nossa base de clientes se concentra em três bairros estratégicos: Copacabana, Tijuca e Barra da Tijuca. Essa concentração geográfica é uma vantagem competitiva, permitindo otimizar a logística e as campanhas de marketing com foco em áreas de alta densidade de clientes.

A segmentação nos próximos passos nos ajudará a entender qual desses bairros tem o maior potencial para o programa "Alimentação Saudável" e direcionar esforços específicos para cada região.

# Análise de Padrões de Compra e Segmentação de Clientes

## Foco no Público de Alto Potencial

Nossa análise de segmentação identificou o grupo de clientes mais valioso para o sucesso do "Programa Alimentação Saudável". Conforme evidenciado no gráfico, o Cluster de Alto Potencial representa a oportunidade imediata e de maior retorno para a SuperVida.



### Identificação do Público-Alvo

O Cluster 1 se destaca por seu alto ticket médio e forte engajamento com a categoria de produtos saudáveis. Este grupo já possui o hábito de compra que buscamos, tornando-o o alvo ideal para ações de fidelização e aumento de receita.

### Ação Recomendada

A SuperVida deve concentrar a maior parte de seu investimento em marketing e comunicação neste público. A estratégia deve focar em aprofundar o relacionamento com esses clientes, reconhecendo seu valor para a marca.

### Como Alcançar esse Público

Sugerimos a criação de ofertas e programas de fidelidade exclusivos, além do envio de comunicações personalizadas sobre lançamentos de produtos e eventos. Essas ações irão fortalecer o vínculo com a marca e incentivar compras ainda mais frequentes.

# Recomendações Estratégicas para o Programa "Alimentação Saudável"

Com base no perfil do Cluster 1, recomendamos as seguintes estratégias para maximizar o retorno sobre o investimento e fortalecer o relacionamento com esses clientes:

1

## Comunicação Personalizada e Segmentada

Utilize canais de comunicação direta, como WhatsApp, SMS e e-mail, para enviar mensagens personalizadas. O conteúdo deve ser focado em lançamentos exclusivos de produtos saudáveis, dicas de nutrição e notícias sobre a causa das doações para as ONGs.

2

## Programa de Fidelidade Exclusivo

Para este cluster, o programa de pontos pode oferecer benefícios diferenciados, como descontos progressivos, acesso antecipado a eventos de lançamento de produtos e parcerias com nutricionistas ou academias.

3

## Ofertas de Valor (Cross-Selling e Up-Selling)

Crie kits de produtos saudáveis ou promoções de "compre e ganhe" que incentivam o aumento do ticket médio. A estratégia deve ser focada em oferecer mais valor, e não apenas em reduzir o preço.

4

## Coleta de Feedback Contínua

Mantenha um canal de comunicação aberto para ouvir as necessidades e sugestões deste grupo. Eles são os seus clientes mais valiosos e o feedback deles é a chave para o desenvolvimento de novos produtos e serviços.

# Visão para o Futuro: Oportunidades de Expansão

O modelo de segmentação de clientes desenvolvido neste projeto é uma ferramenta poderosa para o crescimento contínuo da SuperVida. Além de otimizar o programa de alimentação saudável, a mesma metodologia pode ser replicada para identificar novas oportunidades de mercado.

## Expansão para Outros Bairros

O mesmo processo de análise exploratória e de clusterização pode ser aplicado a dados de clientes de novos bairros. Isso permitiria à SuperVida avaliar o potencial de clientes com alto engajamento em saúde antes de investir em novas unidades ou campanhas de marketing localizadas.

## Otimização de Campanhas para Outros Clusters

O Cluster 2, com sua alta afinidade por produtos saudáveis, representa uma oportunidade de crescimento. A SuperVida pode desenvolver campanhas específicas para este grupo, focando em elevar o ticket médio e a frequência de compra.

## Tomada de Decisão Baseada em Dados

A implementação desta metodologia garante que as futuras decisões de negócio, desde o lançamento de produtos até a expansão física, sejam embasadas em dados sólidos, reduzindo o risco e aumentando a probabilidade de sucesso.

# ANEXOS

## Informativo sobre os Scripts

Anexo: Scripts e Metodologia – O projeto foi desenvolvido com uma metodologia clara e documentada, que pode ser replicada para futuras análises. Foram criados três scripts principais, cada um com uma função específica:

### `preparar_dados.py`

Responsável pela limpeza, padronização e enriquecimento da base de dados.

### `analise_exploratoria.py`

Utilizado para a geração dos gráficos e visualizações iniciais do perfil do cliente.

### `segmentacao_clientes.py`

Contém o modelo de machine learning (K-Means) para a segmentação de clientes.

## Checklist do projeto de mineração de dados

### Parcerias Comunitárias

Doações e educação alimentar local

### Programa de Fidelidade

Pontos que geram doações e descontos



### Segmentação de Clientes

Identificar cluster de alto potencial

### Comunicação Personalizada

Mensagens e ofertas por canal

### Etapa 1: Preparação do Ambiente e Geração de Dados Fictícios

Gerar e salvar o arquivo `clientes_supervida.csv`.

1

### Etapa 3: Análise Exploratória de Dados (AED)

Criar o script `analise_exploratoria.py` para gerar gráficos de perfil demográfico (gênero, faixa etária e bairro) e padrões de compra (ticket médio e produtos saudáveis).

2

### Etapa 2: Pré-processamento e Limpeza dos Dados

Criar o script `preparar_dados.py` para limpar os dados, padronizar as colunas e gerar o arquivo `clientes_supervida_ajustado_final.csv`.

3

### Etapa 4: Descoberta de Perfis e Segmentação de Clientes

Criar o script `segmentacao_clientes.py` para aplicar o algoritmo K-Means e identificar o cluster de alto potencial.

4

### Etapa 5: Relatório Final e Propostas de Estratégia

Consolidar os resultados da análise em um relatório executivo.

5

### Etapa 6: Visão para o Futuro

Propor um plano de expansão e escalabilidade do modelo de análise para futuros projetos.

6