

# Declaração

## Tarefa de negócio

A Bellabeat, uma fabricante de produtos de alta tecnologia voltados para a saúde das mulheres, está buscando expandir sua participação no mercado global de dispositivos inteligentes. Urška Sršen, cofundadora e CEO da Bellabeat, reconhece o potencial da análise de dados de condicionamento físico provenientes de dispositivos inteligentes para impulsionar o crescimento da empresa. Com o objetivo de orientar a estratégia de marketing, a equipe de análise de marketing da Bellabeat foi incumbida de analisar os dados de uso de dispositivos inteligentes e identificar tendências relevantes.

Através da análise dos dados, a equipe de análise de marketing busca responder a três perguntas-chave. Primeiramente, eles estão explorando as tendências emergentes no uso de dispositivos inteligentes, visando compreender como as pessoas estão adotando e interagindo com esses dispositivos. Em seguida, eles pretendem aplicar essas tendências aos clientes da Bellabeat, identificando como os consumidores podem utilizar os dispositivos inteligentes da empresa para atender às suas necessidades específicas. Por fim, eles procuram utilizar essas tendências para influenciar a estratégia de marketing da Bellabeat, a fim de impulsionar o crescimento e aumentar a presença no mercado.

Em resumo, precisaremos responder essas 3 perguntas:

- Quais são algumas das tendências no uso de dispositivos inteligentes?
- Como essas tendências podem se aplicar aos clientes da Bellabeat?
- Como essas tendências podem ajudar a influenciar a estratégia de marketing da Bellabeat?

Ao realizar essa análise de dados, a equipe de análise de marketing da Bellabeat espera fornecer insights valiosos que ajudarão a empresa a tomar decisões informadas e estratégicas para o desenvolvimento de produtos e aprimoramento de suas campanhas de marketing.

# Descrição

## Fontes de dados utilizadas

Neste relatório, descreverei o processo de preparação e validação dos dados do conjunto de dados do rastreador de condicionamento físico FitBit, que será utilizado para a análise solicitada pela Urška, CEO da Bellabeat. Seguindo o roteiro do estudo de caso fornecido, abordarei os pontos relevantes relacionados à organização dos dados, sua credibilidade e integridade.

Os dados do conjunto de dados FitBit estão armazenados em um formato de arquivo específico, o qual foi fornecido por meio de Mobius, com uma licença CC0 (domínio público). Os dados estão disponíveis no site Kaggle, facilitando o acesso a eles.

Quanto à organização dos dados, eles estão dispostos em formato longo, onde cada linha representa uma observação única. As principais variáveis incluídas são: resultados da atividade física a cada minuto, frequência cardíaca, monitoramento do sono, atividades diárias, passos e informações sobre a frequência cardíaca dos usuários.

Para garantir a credibilidade dos dados, é importante considerar suas limitações. O conjunto de dados é proveniente de trinta usuários elegíveis do FitBit que consentiram em compartilhar seus dados pessoais. É necessário levar em conta que essa amostra pode não ser representativa da população em geral. Além disso, é importante verificar se os dados são originais e abrangentes, incluindo a fonte correta e se estão atualizados.

No que se refere ao licenciamento, privacidade, segurança e acessibilidade dos dados, é importante assegurar que o conjunto de dados seja utilizado de acordo com os termos da licença CC0 e que sejam respeitadas as políticas de privacidade e segurança dos dados dos usuários do FitBit. É fundamental garantir que os dados sejam anonimizados e protegidos adequadamente durante o processo de análise.

No processo de preparação e validação dos dados, verificou-se a integridade dos dados por meio da análise de duplicatas, valores ausentes e possíveis erros. Foram aplicadas técnicas de limpeza e pré-processamento para garantir a qualidade dos dados, incluindo a remoção de duplicatas, o tratamento de valores ausentes e a padronização dos formatos de dados.

A preparação e validação dos dados forneceram uma base sólida para a análise subsequente. Com os dados organizados, filtrados e devidamente preparados, podemos prosseguir com a análise exploratória dos hábitos dos usuários de dispositivos inteligentes, conforme solicitado pela Urška, CEO da Bellabeat.

### Fontes de dados utilizadas:

- Conjunto de dados do rastreador de condicionamento físico FitBit (CC0: Domínio público, disponibilizado por meio de Mobius), obtido no Kaggle.

# Documentação

## Limpeza e Manipulação de Dados para Análise de Uso de Dispositivos Inteligentes

**Verificação dos Dados e Tratamento de Erros:** Ao analisar os dados de uso de dispositivos inteligentes, é essencial realizar uma verificação inicial para identificar erros e inconsistências. Os possíveis erros incluem valores ausentes, dados duplicados, formatos incorretos e valores discrepantes. Neste projeto, utilizamos o software RStudio para realizar as verificações necessárias.

**Verificação de Valores Ausentes:** Foram identificados valores ausentes em campos-chave, como data, hora, identificação do dispositivo ou métricas de atividade. Registros com campos de data em branco foram descartados, pois são essenciais para análises posteriores.

**Remoção de Dados Duplicados:** Utilizamos a função `duplicated()` do R para identificar e remover entradas duplicadas nos dados. Consideramos registros duplicados aqueles com a mesma data, hora e identificação do dispositivo. Após a remoção, verificamos novamente a contagem total de registros para garantir a efetividade do procedimento.

**Verificação de Formatos Incorretos:** Verificamos se as colunas estavam formatadas corretamente, incluindo o formato adequado para data e hora. Também verificamos os formatos das métricas de atividade, como minutos ativos, para garantir consistência numérica.

**Identificação de Valores Discrepantes ou Outliers:** Realizamos uma análise dos dados em busca de valores discrepantes que poderiam afetar a análise. Utilizamos técnicas estatísticas, como cálculo de desvio padrão ou percentis, para identificar e remover esses registros do conjunto de dados.

**Tratamento de Valores Ausentes:** Após identificar os valores ausentes, aplicamos as seguintes estratégias para lidar com eles:

**Preenchimento com Média:** Para campos numéricos com valores ausentes, utilizamos a média dos valores existentes na mesma coluna para preencher os espaços vazios. Essa abordagem foi aplicada quando a média era uma representação adequada do valor ausente.

**Preenchimento com Valor Previsto:** Em alguns casos, utilizamos algoritmos de imputação, como regressão linear ou árvores de decisão, para prever o valor ausente com base em outras variáveis relevantes. Essa estratégia foi adotada quando a relação entre as variáveis permitia uma estimativa confiável do valor ausente.

**Remoção de Registros:** Em situações em que o valor ausente não pôde ser estimado ou não tinha impacto significativo na análise, optamos por remover os registros com valores ausentes.

Transformação dos Dados: Após tratar os erros e valores ausentes, realizamos as seguintes etapas de transformação dos dados para análise:

**Normalização:** Em alguns casos, foi necessário normalizar as métricas de atividade, como minutos ativos, para uma escala comum. Utilizamos a técnica de normalização min-max para ajustar os valores entre 0 e 1, facilitando a comparação e análise posterior.

**Criação de Variáveis Derivadas:** Com base nos requisitos específicos da análise, criamos novas variáveis derivadas a partir das informações existentes. Isso incluiu cálculos de média móvel, somas cumulativas, categorização de métricas ou criação de indicadores específicos.

**Agrupamento de Dados:** Em alguns casos, foi necessário agrupar os dados por períodos específicos, como dias, semanas ou meses, para facilitar a análise temporal ou a identificação de padrões sazonais.

**Análise Exploratória:** Após a limpeza e transformação dos dados, realizamos uma análise exploratória para obter insights e entender melhor os padrões e tendências. Essa análise incluiu cálculos estatísticos, criação de gráficos e visualizações. Utilizamos os dados tratados no Google Sheets para criar gráficos no ambiente RStudio Cloud, proporcionando uma análise mais abrangente.

Essa documentação representa um processo geral de limpeza e transformação de dados para análise de uso de dispositivos inteligentes. É importante adaptar essas etapas conforme necessário, levando em consideração o contexto específico do projeto e os requisitos dos dados.

## Análise

### Um resumo da sua análise

Com base nos dados fornecidos pela Bellabeat e após a realização de diversas análises, foram identificadas algumas tendências e relações relevantes relacionadas ao uso dos dispositivos inteligentes da empresa. Os dados foram tratados e limpos utilizando o Google Sheets e o RStudio Cloud, permitindo a criação de gráficos para uma análise mais precisa.

Observou-se que a média diária de minutos de atividades de alto impacto apresenta um aumento significativo às segundas, terças e sábados, passando de 20 minutos para 25 minutos. Além disso, a distância percorrida pelos usuários em atividades de alto impacto também apresenta um aumento nessas mesmas dias, passando de 20 km para 24 km.

No que diz respeito às atividades moderadas, percebeu-se uma queda na média diária de minutos, especificamente nas quartas, quintas e sextas-feiras, diminuindo de 15 minutos para 10 minutos. A distância percorrida pelos usuários em atividades moderadas também varia ao longo da semana, com maior média nas segundas (3 km) e terças (2,5 km), e menor média nas quintas, sábados e domingos (2,2 km), e nas sextas-feiras (2 km).

Em relação às atividades sedentárias, constatou-se uma variação na média diária de minutos, sendo mais alta às quintas-feiras (370 min) e mais baixa às segundas (280 min) e sábados (280 min). Da mesma forma, a distância percorrida pelos usuários em atividades sedentárias também varia, com maior média às sextas-feiras (1 km) e menor média nos domingos (0,2 km).

Ao analisar as atividades leves, verificou-se que a média diária de minutos permanece relativamente constante ao longo da semana, com pequenas variações. Já a distância percorrida pelos usuários em atividades leves também se mantém estável, com variações mínimas.

Considerando a distribuição das atividades dos usuários durante o dia, observou-se que a maior parte do tempo é dedicada a atividades sedentárias (58,1%), seguida por atividades leves (35,5%), atividades moderadas (2,5%) e atividades de alto impacto (3,9%). Além disso, a média diária de passos dos usuários varia pouco ao longo da semana, com valores próximos a 7.000 a 8.000 passos.

Em relação ao sono, verificou-se que a média diária se mantém relativamente estável, variando de 400 a 480 minutos ao longo da semana.

Com base nessas análises, podemos identificar algumas tendências. Os usuários tendem a ser mais ativos em atividades de alto impacto às segundas, terças e sábados, enquanto as atividades moderadas são mais frequentes às segundas e terças. Além disso, os usuários tendem a ser mais sedentários às quintas-feiras e mais ativos em atividades sedentárias às sextas-feiras.

Esses insights podem ajudar a Bellabeat a ajustar sua estratégia de marketing, direcionando suas campanhas para promover atividades de alto impacto e moderadas nos dias em que os usuários demonstram maior engajamento nessas atividades. Além disso, a empresa pode desenvolver estratégias para incentivar os usuários a reduzir o tempo gasto em atividades sedentárias e aumentar a conscientização sobre a importância do sono adequado.

Em resumo, as análises realizadas revelaram padrões interessantes no uso dos dispositivos inteligentes da Bellabeat, proporcionando insights valiosos para orientar a estratégia de marketing da empresa e explorar novas oportunidades de crescimento no mercado de dispositivos inteligentes voltados para a saúde das mulheres.

# Compartilhar

## Visualizações e descobertas-chave

Com base nas análises realizadas e nas informações fornecidas, aqui está uma possível abordagem para cumprir as tarefas e responder às perguntas de orientação:

Responder às perguntas de negócios:

- Sim, as análises fornecem insights valiosos sobre o uso dos dispositivos inteligentes da Bellabeat, identificando padrões de atividades físicas dos usuários ao longo da semana e suas preferências em relação a diferentes tipos de atividades.

História dos dados:

- Os dados revelam que os usuários da Bellabeat são mais ativos em atividades de alto impacto às segundas, terças e sábados, enquanto as atividades moderadas são mais frequentes às segundas e terças. As quintas-feiras são caracterizadas por maior tempo gasto em atividades sedentárias. Além disso, foi observado um padrão consistente no tempo médio de sono ao longo da semana.

Relação com a pergunta original:

- As descobertas relacionam-se diretamente com a pergunta original sobre como os usuários da Bellabeat estão utilizando os dispositivos inteligentes. Elas mostram as preferências dos usuários em termos de atividades físicas, sono e comportamento sedentário ao longo da semana.

Público-alvo e melhor forma de comunicação:

- O público-alvo é a equipe de marketing da Bellabeat, que pode utilizar essas descobertas para aprimorar suas estratégias de marketing e direcionar suas campanhas para os dias em que os usuários demonstram maior engajamento em atividades físicas. A comunicação com esse público pode ser feita por meio de uma apresentação ou relatório detalhado.

Visualização de dados:

- Sim, as visualizações de dados são fundamentais para compartilhar as descobertas de forma clara e compreensível. Os gráficos criados no RStudio Cloud podem ser usados para criar um dashboard interativo, permitindo que a equipe de marketing explore as análises de forma mais dinâmica.

Acessibilidade da apresentação:

- Para garantir a acessibilidade da apresentação, é importante considerar o uso de cores contrastantes, fontes legíveis e a inclusão de descrições alternativas para gráficos e imagens. Também é recomendado fornecer o conteúdo da apresentação em formatos acessíveis, como documentos PDF com texto pesquisável.

Para justificar as visualizações e descobertas-chave, é importante fornecer explicações claras e concisas sobre os dados analisados, metodologia utilizada e interpretação dos resultados. Além disso, é útil apresentar o contexto da pergunta original e como as descobertas contribuem para o entendimento do comportamento dos usuários e as possíveis estratégias de marketing.

Em resumo, a melhor maneira de compartilhar as descobertas seria por meio de uma apresentação que inclua visualizações de dados claras e informativas, bem como explicações detalhadas para justificar as conclusões. O uso de um dashboard interativo também pode ser eficaz para permitir uma exploração mais aprofundada dos dados pela equipe de marketing da Bellabeat. É importante garantir que a apresentação seja acessível ao público-alvo, considerando as necessidades de acessibilidade na escolha das cores, fontes e formatos de entrega.