# 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 ou manipulação de dados  
  
Verificação dos dados em busca de erros Ao analisar os dados de uso de dispositivos inteligentes, é importante realizar uma verificação inicial para identificar possíveis erros ou inconsistências nos dados. Esses erros podem incluir valores ausentes, dados duplicados, formatos incorretos, valores discrepantes ou outliers. Para essa tarefa, utilizei a ferramenta de planilhas Microsoft Excel.

Escolha das ferramentas Para realizar a análise e limpeza dos dados, escolhi utilizar o Microsoft Excel devido à sua familiaridade e ampla capacidade de manipulação de dados. O Excel oferece recursos robustos para filtrar, classificar e formatar dados, além de fornecer funções para transformações e cálculos básicos.

Transformação dos dados Após identificar os erros nos dados, executei as seguintes etapas de transformação para garantir que eles estejam prontos para análise:

1. Remoção de valores duplicados: Utilizei a funcionalidade de remoção de duplicatas do Excel para eliminar entradas repetidas nos dados.
2. Formatação de dados: Verifiquei se as colunas estavam formatadas corretamente. Por exemplo, garanti que as colunas de data e hora estivessem no formato apropriado, para facilitar análises futuras.
3. Tratamento de outliers: Analisei os dados em busca de valores discrepantes que pudessem afetar a análise. Em casos de outliers extremos, decidi remover esses registros do conjunto de dados, pois poderiam distorcer as tendências.

Documentação do processo de limpeza É essencial documentar todas as etapas e transformações realizadas nos dados para que seja possível revisar, reproduzir e compartilhar os resultados. Criei um documento no formato de relatório, descrevendo detalhadamente cada passo do processo de limpeza, desde a verificação inicial dos erros até a transformação final dos dados. Incluí capturas de tela relevantes, fórmulas utilizadas e explicações claras sobre as decisões tomadas durante o processo.

Após a limpeza dos dados, realizei as seguintes verificações para garantir sua integridade:

1. Contagem de registros: Comparei o número total de registros após a limpeza com o número inicial para identificar possíveis perdas significativas de dados.
2. Verificação de valores únicos: Verifiquei se todos os valores em atributos-chave eram únicos para garantir a consistência dos dados.
3. Testes de consistência: Realizei testes para verificar a consistência entre dados relacionados, como datas correspondentes a valores corretos ou valores numéricos seguindo regras esperadas.
4. Validação de formatos de dados: Garanti que os formatos dos dados estavam corretos, como datas no formato apropriado e campos de texto sem caracteres inválidos.

Essas verificações foram documentadas para revisão e compartilhamento dos resultados.