

Estudo de Caso Bellabeat

Fase 1: Perguntar

Objetivo da análise:

Investigar como usuários de dispositivos Fitbit utilizam seus aparelhos e identificar padrões de comportamento relacionados à saúde física.

Pergunta de negócio:

Como os consumidores utilizam dispositivos inteligentes voltados ao monitoramento da saúde física e como esses hábitos podem orientar decisões estratégicas da Bellabeat?

Perguntas secundárias:

- Quais são as tendências no uso de dispositivos inteligentes?
- Como essas tendências podem se aplicar às clientes da Bellabeat?
- Como elas podem influenciar a estratégia de marketing da empresa?

Resumo da tarefa de negócios:

A Bellabeat deseja expandir sua presença no mercado de dispositivos inteligentes e de bem-estar. Esta análise busca compreender o comportamento dos usuários utilizando dispositivos inteligentes para o monitoramento da saúde física, de forma a gerar recomendações estratégicas alinhadas ao perfil do público feminino da marca.

Fase 2: Preparar

Fonte dos dados:

Os dados estão disponíveis em <https://www.kaggle.com/datasets/arashnic/fitbit/data>

Formato dos dados:

Os arquivos estão no formato longo, registrando atividades por usuário por dia.

Credibilidade e limitações:

- Amostra pequena (35 usuários).
- Ausência de dados demográficos (sexo, idade, altura).
- Não é possível saber se a base representa especificamente o público feminino — foco da Bellabeat.
- Possível viés por falta de diversidade de perfis.

Privacidade e licenciamento:

- Dados públicos.

- Usuários anonimizados.
- Uso permitido para análise.

Verificação de integridade:

- Remoção de duplicatas.
- Checagem de campos vazios.
- Consistência lógica entre distâncias.
- Padronização de datas e tipos numéricos.

Problemas encontrados:

- Inconsistências entre TotalDistance e distâncias parciais para alguns usuários (erro de GPS).
- No minutoSleep_merged: centenas de duplicatas (removidas).
- Usuários com registros incompletos de sono e peso.

Resumo da operação:

Os dados foram limpos, inconsistências pequenas foram justificadas e duplicatas removidas. Apesar das limitações, o dataset está apto para análises exploratórias.

Fase 3: Processar

Ferramentas utilizadas:

- Google Spreadsheets
- Python
- MySQL

Procedimentos de Limpeza:

- Padronização de datas
- Conversão de tipos numéricos
- Remoção de duplicatas
- Checagem de valores vazios
- Verificação de coerência entre valores relacionados (atividades x distância)

Critérios de validação:

- Id + Data únicos por registro
- Distâncias consistentes com atividades
- Outliers analisados individualmente
- Arquivos duplicados removidos antes da carga no banco de dados

Limpeza e Tratamento dos Dados

Etapa	Descrição	Ferramenta	Justificativa	Resultado
Padronização do Formato de Data	Ajuste do Formato de data para Ano-Mes-Dia HH:mm:ss	Python	Algumas datas estavam sendo interpretadas de uma forma e outras não, adoção de um método universal	4/12/2016 -> 2016-04-12 00:00:00
Remoção de Duplicatas	Filtragem e exclusão de linhas duplicadas	Spreadsheets - Remover Cópias	Evitar contagem repetida de registros	525 linhas removidas no arquivo minuteSleep_merged
Verificação de campos vazios	Filtro de células vazias em todas as colunas	Spreadsheets - Filtro por condição	Evitar distorções nas médias e totais	Nenhum campo vazio encontrado
Conferência de consistência	Comparação de TotalDistance com as distâncias parciais	Spreadsheets - Filtro por condição + fórmulas de comparação	Detectar erros de precisão do GPS	Diferenças pequenas aceitas (<0,1 mi); discrepâncias grandes (usuário 4020332650) sinalizadas

Organização para análise:

Após o processo de limpeza, o conjunto de dados está livre de duplicatas, com formatos padronizados e pequenas inconsistências tratadas ou justificadas. O dataset está pronto para análise exploratória. Os dados foram importados para um banco relacional no MySQL, com tabelas específicas para cada tipo de métrica (atividade diária, sono, calorias, intensidade, passos, batimentos e peso).

Fase 4: Analisar

Descobertas gerais:

- Usuários apresentam perfis variados: alguns altamente ativos, outros pouco ativos.
- Atividade física média: ~7.247 passos/dia e grande quantidade de tempo sedentário.
- Sono médio: ~7,7 horas, mas com grande variação individual.
- Baixa adesão ao registro de sono e peso reduz a profundidade das análises.

Padrões Identificados:

- Sono adequado relaciona-se a melhor equilíbrio energético e atividade leve.
- Padrões semanais estáveis: pouca diferença entre dias úteis e fins de semana.
- Atividades intensas são raras, mas influenciam significativamente o gasto calórico.
- Há usuários com registros muito irregulares de utilização do dispositivo.

Insights chave:

- A combinação de sono e atividade física revela fortes correlações com o gasto calórico.
- A inconsistência de uso entre usuários sugere oportunidade de aumentar engajamento.
- A maior parte dos comportamentos indica rotina sedentária — possível alvo de melhorias nos produtos Bellabeat.

Fase 5: Compartilhar

Panorama geral de atividade física

O primeiro conjunto de gráficos mostra que, em média, as usuárias apresentam:

- 2.264 kcal gastas por dia,
- 5,19 km percorridos,
- 7.247 passos realizados por dia

Esses valores sugerem um nível de atividade moderado, porém ainda abaixo das recomendações internacionais de saúde (10k passos/dia geralmente sugeridos como referência). A distância percorrida acompanha diretamente o volume de passos, reforçando um padrão de movimentação consistente, mas não elevado.

Distribuição do tempo em atividades por intensidade + sono

A decomposição do tempo diário mostra que:

- 43,4% do tempo total inativo é sono (≈431 min)
- O restante se distribui entre inatividade acordada e atividade leve.

As usuárias passam mais de 16 horas/dia sedentárias, entre sono e períodos acordadas, enquanto o tempo em:

- atividade intensa é muito baixo (≈ 20 min/dia),
- moderada ainda menor (≈ 14 min/dia),
- e leve, apesar de mais presente, também não compensa o volume de inatividade.

Essa distribuição revela um comportamento predominantemente sedentário, o que é coerente com o número moderado de passos.

Comparativo de passos - dias úteis x fim de semana

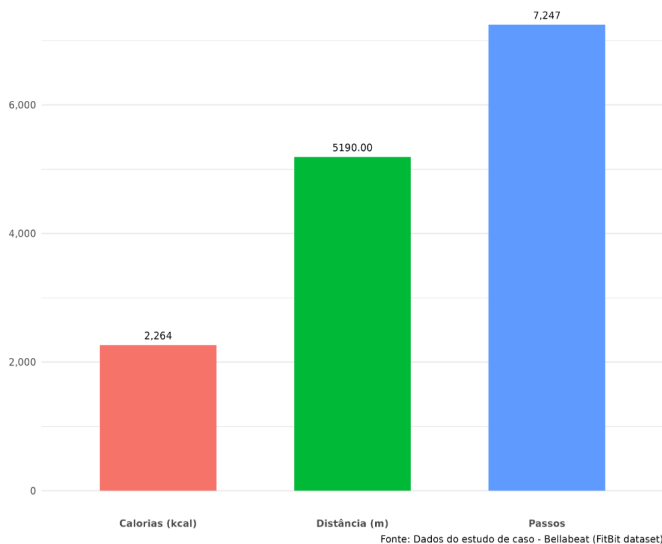
O gráfico comparativo mostra:

- Semana: 7.271 passos em média
- Fim de semana: 7.188 passos

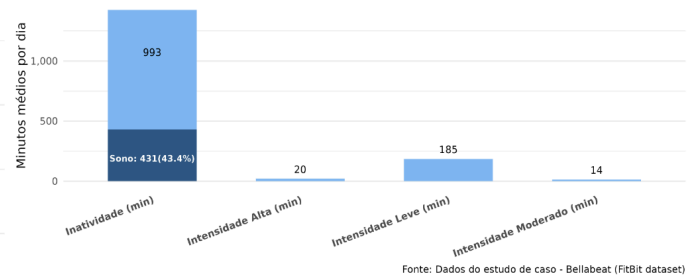
A diferença é pequena, indicando padrão estável de comportamento, sugerindo que as usuárias não variam muito sua rotina de movimentação entre dias úteis e finais de semana.

Resumo Visual - Estudo de Caso Bellabeat.
Panorama de sono, atividade e comportamento das usuárias

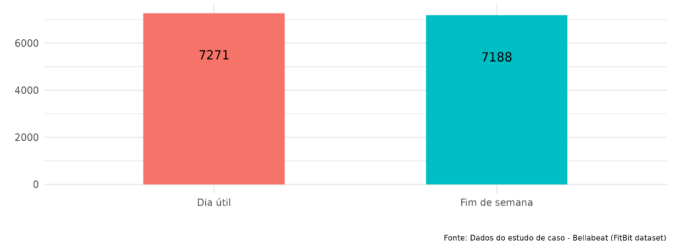
Panorama Geral dos Usuários - Média Calorias, Distância e Passos
Indicadores agregados de atividade física e gasto calórico



Panorama Geral das Usuárias - Tempo Médio em Atividade e Sono
O sono representa em média 43.4% do tempo total inativo



Comparativo de Passos Médio
Dia útil = Seg até Sex / Fim de Semana = Sáb e Dom



Insights gerais

Existe uma forte predominância de comportamento sedentário com quase metade do tempo inativo diário sendo passado dormindo, mas a outra metade ocorre acordado. Isso reforça a oportunidade de o dispositivo incentivar períodos de repouso mais ativos, com lembretes de movimentação e metas progressivas.

Além disso, os minutos diários em intensidades mais altas(intensa e moderada) são muito reduzidos. Para o contexto Bellabeat, isso abre espaço para gamificação, desafios diários e recomendações de treinos curtos de alta eficiência.

A estabilidade na quantidade de passos sugere que os usuários seguem rotinas previsíveis e podem responder bem a metas semanais ou ajustes de hábitos baseados em consistência, ao invés de estratégias focadas apenas em picos de atividade.

Os indicadores de distância, calorias e passos caminham juntos, mas todos ficam

em faixas moderadas, com nenhum destaque. Isso pode indicar que os usuários caminham bastante, mas sem intensidade, não incorporam exercícios estruturados a sua rotina e têm longos períodos de inatividade.

História dos dados:

Os dados mostram que grande parte das usuárias apresenta uma rotina com baixa atividade física, longos períodos sedentários e registros incompletos de sono. Mesmo com uma média de sono adequada, a distribuição de atividades ao longo do dia é irregular e insuficiente.

Visualizações Produzidas:

- Relação entre Intensidade da Atividade Física, Gasto Calórico e Batimentos Cardíacos
- Impacto da Duração do Sono sobre o Desempenho Físico Diário
- Relação entre Peso e Componentes de Atividade Física

Que podem ser encontradas também em pdf junto a este repositório.

Fase 6: Agir

Conclusão Geral:

A análise indica que o comportamento típico das usuárias é sedentário, apesar de dormirem em média 7,7 horas por noite. A atividade física é distribuída em pequenos períodos, e a ausência de pausas regulares limita benefícios metabólicos.

Principais Recomendações:

- Criar notificações inteligentes de movimento.
- Oferecer metas individualizadas de atividade e sono.
- Reforçar conteúdos educativos sobre hábitos saudáveis.
- Direcionar campanhas para mulheres com rotinas intensas.
- Incentivar registros consistentes de peso e sono.