

Caso de Estudo Bellabeat



28-07-2024

**Pedro Alexandre Pereira
Fernandes**

Índice

Introdução	3
Perguntar.....	4
Preparar	4
Processar	5
Analisar no SQL.....	9
Analisar no Tableau.....	17
Partilhar no Tableau	20
Atuar.....	27
Considerações Pessoais	30

Introdução

O estudo do caso insere-se na empresa Bellabeat. Vamos saber melhor do que se trata este negócio:

Urška Sršen e Sando Mur fundaram a Bellabeat, uma empresa de alta tecnologia que fabrica produtos inteligentes focados na saúde. A Urška aproveitou a sua experiência como artista para desenvolver uma tecnologia que informa e inspira mulheres em todo o mundo. A recolha de dados sobre a sua atividade, sono e saúde reprodutiva permitiu à Bellabeat capacitar as mulheres com conhecimento sobre a sua própria saúde e hábitos. Desde que foi fundada em 2013, a Bellabeat cresceu rapidamente e não levou muito tempo para se posicionar como uma empresa de bem-estar e tecnologia para as mulheres.

Urška Sršen: Cofundadora e CEO da Bellabeat

Sando Mur: Matemático e cofundador da Bellabeat; membro-chave da equipa executiva da Bellabeat

Equipa de análise de marketing da Bellabeat: Uma equipa de analistas de dados responsável por recolher, analisar e relatar dados que ajudam a orientar a estratégia de marketing da Bellabeat.

Falando agora dos produtos:

Aplicativo Bellabeat: O aplicativo Bellabeat fornece aos utilizadores, dados de saúde relacionados à sua atividade, sono, ciclo menstrual e hábitos de atenção plena. Estes dados podem ajudar os utilizadores a entender melhor os seus hábitos atuais e tomar decisões saudáveis. O aplicativo Bellabeat conecta-se à sua linha de produtos inteligentes de bem-estar.

Leaf: O rastreador de bem-estar clássico da Bellabeat pode ser usado como pulseira. O rastreador Leaf conecta-se ao aplicativo Bellabeat para rastrear a atividade e o sono.

Time: Este relógio de bem-estar combina a aparência atemporal de um relógio clássico com tecnologia inteligente para rastrear a atividade e o sono do cliente. O relógio Time conecta-se ao aplicativo Bellabeat para fornecer informações sobre o bem-estar diário.

Spring: Esta é uma garrafa de água que rastreia a ingestão diária de água por meio de tecnologia inteligente para garantir que o cliente está adequadamente hidratado ao longo do dia. A garrafa Spring conecta-se ao aplicativo Bellabeat para rastrear os níveis de hidratação.

Planos da Bellabeat: A Bellabeat também oferece aos utilizadores diferentes planos de assinatura. As assinaturas oferecem aos clientes acesso 24 horas por dia, 7 dias por semana, orientação totalmente personalizada sobre nutrição, atividade, sono e saúde, além de atenção plena com base no estilo de vida e objetivos.

Perguntar

O objetivo passa por analisar os dados do uso dos dispositivos inteligentes para obter informações sobre como os consumidores usam os dispositivos inteligentes que não são da Bellabeat. Em seguida, teremos de selecionar um produto da Bellabeat para aplicar esses *insights* na apresentação.

Dentro desta análise, espera-se que a Bellabeat consiga juntar mais subscritores aos seus serviços.

Perguntas para resolver:

1. Quais são algumas das tendências no uso de dispositivos inteligentes?
2. Como estas tendências podem ser aplicadas aos clientes da Bellabeat?
3. Como é que estas tendências podem ajudar a influenciar a estratégia de marketing da Bellabeat?

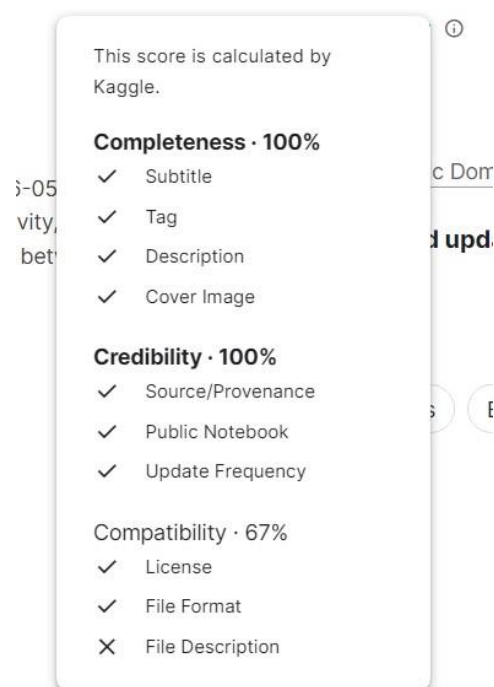
Preparar

Comecei por descarregar os dados do rastreador de condicionamento físico FitBit. Este conjunto de dados do Kaggle contém um rastreador de condicionamento físico pessoal de trinta usuários do Fitbit entre as datas de 12/03/2016 a 12/05/2016.

Trinta usuários elegíveis do Fitbit consentiram com o envio de dados pessoais do rastreador, incluindo os resultados a cada minuto de atividade física, frequência cardíaca e monitoramento do sono. São abrangidas informações sobre atividades diárias, passos e frequência cardíaca que podem ser usadas para explorar os hábitos dos usuários.

Os dados estão organizados numa pasta do ambiente de trabalho chamada “caso de estudo BellaBeat”

Visto que os dados estão na Kaggle, estes estão em conformidade com o licenciamento, privacidade e segurança.



Como podemos observar na imagem, este é um print screen que tirei do conjunto de dados que vamos abordar.

Processar

Os dados são um pouco limitados por apresentar uma amostra de 33 clientes e portanto, poderemos não ter resultados precisos.

Nem todos os usuários têm o conjunto de dados completo.

Comecei pelo processamento no Google Sheets.

Selecionei as tabelas que interessavam mais para a pesquisa e apliquei o seguinte:

- Remover os duplicados
- Limpeza dos espaços em branco
- Limpeza de colunas que não interessavam para o estudo
- Classificação dos dados para que apresentem uma maior consistência

Estes são os ficheiros onde foi feita a aplicação:

1. daily_activity_03
2. daily_activity_04
3. daily_calories_04
4. daily_sleep_04
5. hourly_calories_03
6. hourly_calories_04
7. hourly_steps_03
8. hourly_steps_04
9. min_sleep_03
10. min_sleep_04
11. sec_heartrate_03
12. sec_heartrate_04
13. weight_info_03
14. weight_info_04

De seguida, utilizei o MySQL para ter uma limpeza mais eficiente. Devido à grande quantidade de dados, penso que seria a ferramenta que mais se adequava para o processamento destes dados:

Comecei por criar novas tabelas, juntando cada tabela em uma só, para simplificar o processo.

```

1 CREATE TABLE daily_activity AS
2 SELECT Id, ActivityDate, TotalSteps, TotalDistance, TrackerDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDist
3 FROM bellabeat.dailyactivity_merged_03 - dailyactivity_merged_03.csv
4 UNION
5 SELECT Id, ActivityDate, TotalSteps, TotalDistance, TrackerDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDist
6 FROM bellabeat.dailyactivity_merged_04 - dailyactivity_merged_04.csv

```

De seguida, calculamos quantas entradas houve dos clientes no daily_activity e este foi o resultado:

Query 1		dailyactivity_merged_03 - dailya...	daily_activity	daily_activity	dailyactivity_merged_04 - dailya...	daily_activity
Result Grid		Filter Rows: <input type="text"/>				
		Export: Wrap Cell Content: <input type="text"/>				
	id	count				
	4020332650	63				
	1503960366	50				
	1624580081	50				
	4445114986	46				
	4702921684	46				
	6962181067	45				
	1844505072	43				
	1927972279	43				
	2022484408	43				
	2026352035	43				
	2320127002	43				
	2873212765	43				
	4319703577	43				
	4558609924	43				
	5553957443	43				
	7086361926	43				
	8378563200	43				
	8877689391	43				
	3977333714	42				
	8053475328	42				
	5577150313	41				
	8792009665	41				
	1644430081	40				
	4388161847	39				
	6290855005	39				
	8583815059	39				
	6117666160	38				
	7007744171	38				
	4057192912	36				
	6775888955	35				
	2347167796	33				
	8253242879	31				
	3372868164	30				
	6391747486	9				

Já tinha sido removido o user “2891001257” por insuficiência de dados e fiz o mesmo para o user “6391747486”. Dando ao todo 33 utilizadores. Foi aplicado a todas as tabelas este código:

```

3 DELETE FROM bellabeat.daily_activity
4 WHERE Id='2891001357'

```

Fui ver a consistência dos passos feitos por hora:

```
SELECT
  id,
  COUNT(Id) as count
FROM
  bellabeat.hourlysteps
GROUP BY
  Id
ORDER BY
  count DESC;
```



id	count
1624580081	1480
1927972279	1480
2022484408	1480
2026352035	1480
4558609924	1480
8053475328	1479
8378563200	1479
2320127002	1479
4445114986	1479
8877689391	1479
7086361926	1477
2873212765	1476
4020332650	1476
6962181067	1476
1844505072	1475
5553957443	1474
4702921684	1473
1503960366	1461
3977333714	1440
5577150313	1438
8792009665	1416
1644430081	1408
4319703577	1395
8583815059	1384
6117666160	1355
7007744171	1345
6775888955	1292
6290855005	1186
3372868164	1183
8253242879	1171
2347167796	1158
4057192912	832
4388161847	735

As conclusões que podemos chegar é que:

- 18 users** falharam menos de 1 dia
- 8 users** falharam menos de 7 dias
- 5 users** falharam menos de 14 dias
- 2 users** falharam mais que 14 dias

Verifiquei também o número de registo de sono, este foi o resultado:

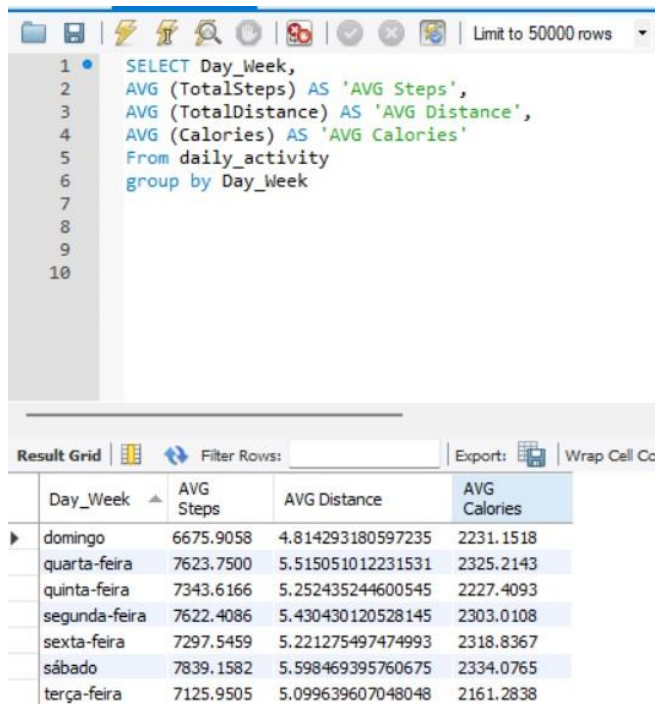
id	count
2026352035	28645
5553957443	25004
8378563200	23876
6962181067	23566
5577150313	22209
4319703577	22009
2347167796	21444
4702921684	21170
6117666160	19577
7086361926	17424
4445114986	17004
3977333714	16939
1503960366	16791
1927972279	14582
8792009665	9648
4020332650	9117
6775888955	6647
4388161847	5821
1844505072	4676
1644430081	3316
4558609924	1152
8053475328	1040
7007744171	711
2022484408	171
2320127002	47

Eliminei o users “2320127002”, “2022484408” e “7007744171” por insuficiência de registo de sono. Depois da limpeza, temos 22 utilizadores disponíveis dos 33.

Analisar no SQL

Calculei a média dos passos, distância e calorias pelos dias da semana

eis o resultado:



The screenshot shows a SQL IDE interface. At the top, there's a toolbar with various icons and a dropdown menu set to 'Limit to 50000 rows'. Below the toolbar, a SQL query is entered in a text area. The query is as follows:

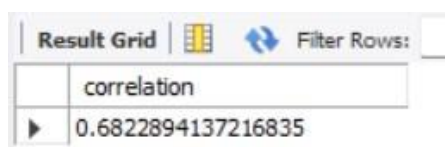
```
1 SELECT Day_Week,
2     AVG (TotalSteps) AS 'AVG Steps',
3     AVG (TotalDistance) AS 'AVG Distance',
4     AVG (Calories) AS 'AVG Calories'
5 From daily_activity
6 group by Day_Week
```

Below the query editor, there's a 'Result Grid' section. It includes a 'Filter Rows:' input field, an 'Export:' button, and a 'Wrap Cell Co' option. The main part of the result grid is a table with the following data:

Day_Week	AVG Steps	AVG Distance	AVG Calories
domingo	6675.9058	4.814293180597235	2231.1518
quarta-feira	7623.7500	5.515051012231531	2325.2143
quinta-feira	7343.6166	5.252435244600545	2227.4093
segunda-feira	7622.4086	5.430430120528145	2303.0108
sexta-feira	7297.5459	5.221275497474993	2318.8367
sábado	7839.1582	5.598469395760675	2334.0765
terça-feira	7125.9505	5.099639607048048	2161.2838

De seguida, quis saber o coeficiente de correlação entre a média de calorias e os passos, fazendo a sua fórmula:

```
SELECT
    (n * sum_xy - sum_x * sum_y) /
    SQRT((n * sum_x2 - sum_x * sum_x) * (n * sum_y2 - sum_y * sum_y)) AS correlation
FROM (
    SELECT
        COUNT(*) AS n,
        SUM(`AVG Calories`) AS sum_x,
        SUM(`AVG Steps`) AS sum_y,
        SUM(`AVG Calories` * `AVG Steps`) AS sum_xy,
        SUM(`AVG Calories` * `AVG Calories`) AS sum_x2,
        SUM(`AVG Steps` * `AVG Steps`) AS sum_y2
    FROM
        `bellabeat`.`avg_daily_activity`
) AS stats;
```



The screenshot shows the 'Result Grid' section of the SQL IDE. It includes a 'Filter Rows:' input field. The main part of the result grid is a table with the following data:

correlation
0.6822894137216835

Já o coeficiente de correlação entre a média das calorias e a distância (aplicando a mesma fórmula):

Result Grid	Filter Rows:
correlation	
0.7047047320051782	

A correlação entre a média da distância e os passos:

Result Grid	Filter Rows:
correlation	
0.994628668154024	

Conclusão: Podemos concluir que existe uma correlação natural entre a média das calorias e dos passos por dia da semana.

As médias das calorias e da distância tiveram uma correlação ligeiramente maior que 0,70. Os passos estarão focados na distância devido aos dados disponíveis e devem representar tendências com precisão, já que a média dos passos e da distância tiveram uma correlação significativamente forte de 0,99.

Foi calculado a média de calorias pelos dias da semana na tabela “hourlycalories”.

código:

```
1 SELECT
2     AVG(Calories) as avg_calories,
3     CASE
4         WHEN DAYOFWEEK(date_time) = 1 THEN 'Sunday'
5         WHEN DAYOFWEEK(date_time) = 2 THEN 'Monday'
6         WHEN DAYOFWEEK(date_time) = 3 THEN 'Tuesday'
7         WHEN DAYOFWEEK(date_time) = 4 THEN 'Wednesday'
8         WHEN DAYOFWEEK(date_time) = 5 THEN 'Thursday'
9         WHEN DAYOFWEEK(date_time) = 6 THEN 'Friday'
10        WHEN DAYOFWEEK(date_time) = 7 THEN 'Saturday'
11    END AS day_of_week
12 FROM
13     bellabeat.hourlycalories
14 GROUP BY
15     day_of_week
16 ORDER BY
17     avg_calories desc
18 -- FIELD(day_of_week, 'Sunday', 'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday')
```

O resultado foi:

	avg_calories	day_of_week
▶	97.3707	Saturday
	96.9916	Tuesday
	96.6123	Wednesday
	96.3748	Thursday
	96.1156	Friday
	95.0837	Monday
	94.0369	Sunday

Concluímos que em média, **sábado** foi o dia da semana que mais se queimou calorias e **domingo** o menos.

Calculou-se também a média dos passos por semana

código:

```
1 SELECT
2     AVG(StepTotal) as avg_Steps,
3     CASE
4         WHEN DAYOFWEEK(`Date`) = 1 THEN 'Sunday'
5         WHEN DAYOFWEEK(`Date`) = 2 THEN 'Monday'
6         WHEN DAYOFWEEK(`Date`) = 3 THEN 'Tuesday'
7         WHEN DAYOFWEEK(`Date`) = 4 THEN 'Wednesday'
8         WHEN DAYOFWEEK(`Date`) = 5 THEN 'Thursday'
9         WHEN DAYOFWEEK(`Date`) = 6 THEN 'Friday'
10        WHEN DAYOFWEEK(`Date`) = 7 THEN 'Saturday'
11    END AS day_of_week
12 FROM
13     bellabeat.hourlysteps
14 GROUP BY
15     day_of_week
16 ORDER BY
17     avg_Steps desc
18 -- FIELD(day_of_week, 'Sunday', 'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday')
```

Eis o resultado:

	avg_Steps	day_of_week
▶	319.9526	Saturday
	314.1951	Tuesday
	313.2068	Thursday
	311.3550	Wednesday
	303.2663	Friday
	301.9186	Monday
	285.9645	Sunday

Sábado foi o dia com mais média de passos e **domingo** o menos, coincidindo com os resultados da média de calorias.

Calculei a média das calorias e passos nas várias fases do dia:

```

1  SELECT
2      AVG(cal.Calories) as avg_calories,
3      AVG(step.StepTotal) as avg_steps,
4      CASE
5          WHEN HOUR(cal.date_time) BETWEEN 6 AND 11 THEN 'Morning'
6          WHEN HOUR(cal.date_time) BETWEEN 12 AND 17 THEN 'Afternoon'
7          WHEN HOUR(cal.date_time) BETWEEN 18 and 21 THEN 'Evening'
8          ELSE 'Night'
9      END AS time_of_day
10 FROM
11     bellabeat.hourlycalories AS cal
12 JOIN bellabeat.hourlysteps AS step ON cal.id=step.id
13 GROUP BY
14     time_of_day
15 ORDER BY
16     avg
17     -- FIELD(time_of_day, 'Morning', 'Afternoon', 'Evening', 'Night');

```

avg_calories	avg_steps	time_of_day
99.9293	309.2183	Morning
113.3529	309.2506	Afternoon
109.5216	309.3714	Evening
73.5460	309.2796	Night

Concluimos que se queima mais calorias à **tarde** (afternoon) e menos à **noite** (night). Os passos são aproximadamente os mesmo ao longo do dia.

Para se ir um pouco mais a fundo, fui explorar onde se queima em média, mais calorias ao longo de cada hora do dia

código:

```

1  SELECT
2      CASE
3          WHEN HOUR(date_time)=0 THEN '0'
4          WHEN HOUR(date_time)=1 THEN '1'
5          WHEN HOUR(date_time)=2 THEN '2'
6          WHEN HOUR(date_time)=3 THEN '3'
7          WHEN HOUR(date_time)=4 THEN '4'
8          WHEN HOUR(date_time)=5 THEN '5'
9          WHEN HOUR(date_time)=6 THEN '6'
10         WHEN HOUR(date_time)=7 THEN '7'
11         WHEN HOUR(date_time)=8 THEN '8'
12         WHEN HOUR(date_time)=9 THEN '9'
13         WHEN HOUR(date_time)=10 THEN '10'
14         WHEN HOUR(date_time)=11 THEN '11'
15         WHEN HOUR(date_time)=12 THEN '12'
16         WHEN HOUR(date_time)=13 THEN '13'
17         WHEN HOUR(date_time)=14 THEN '14'
18         WHEN HOUR(date_time)=15 THEN '15'
19         WHEN HOUR(date_time)=16 THEN '16'
20         WHEN HOUR(date_time)=17 THEN '17'
21         WHEN HOUR(date_time)=18 THEN '18'
22         WHEN HOUR(date_time)=19 THEN '19'
23         WHEN HOUR(date_time)=20 THEN '20'
24         WHEN HOUR(date_time)=21 THEN '21'
25         WHEN HOUR(date_time)=22 THEN '22'
26         WHEN HOUR(date_time)=23 THEN '23'
27     END AS time_of_day, AVG(Calories) as avg_calories
28 FROM
29     bellabeat.hourlycalories
30 GROUP BY
31     time_of_day
32 ORDER BY
33     FIELD(time_of_day, '0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19', '20', '21', '22', '23')

```

O resultado foi:

	time_of_day	avg_calories
0	72.1024	
1	69.8083	
2	68.5341	
3	67.4280	
4	67.8849	
5	79.6688	
6	83.9469	
7	92.5124	
8	100.9642	
9	106.7407	
10	108.6405	
11	109.3636	
12	115.7206	
13	112.3298	
14	114.0540	
15	106.5866	
16	112.2535	
17	117.8846	
18	119.7718	
19	119.4110	
20	103.7464	
21	94.4588	
22	86.5815	
23	77.2882	

Reparamos que a hora onde se queima mais calorias é às **18 horas** e a que menos queima é às **3 da manhã**.

Faremos igual para a média dos passos:

```
1  SELECT
2
3  CASE
4      WHEN HOUR(`Date`) = 0 THEN '0'
5      WHEN HOUR(`Date`) = 1 THEN '1'
6      WHEN HOUR(`Date`) = 2 THEN '2'
7      WHEN HOUR(`Date`) = 3 THEN '3'
8      WHEN HOUR(`Date`) = 4 THEN '4'
9      WHEN HOUR(`Date`) = 5 THEN '5'
10     WHEN HOUR(`Date`) = 6 THEN '6'
11     WHEN HOUR(`Date`) = 7 THEN '7'
12     WHEN HOUR(`Date`) = 8 THEN '8'
13     WHEN HOUR(`Date`) = 9 THEN '9'
14     WHEN HOUR(`Date`) = 10 THEN '10'
15     WHEN HOUR(`Date`) = 11 THEN '11'
16     WHEN HOUR(`Date`) = 12 THEN '12'
17     WHEN HOUR(`Date`) = 13 THEN '13'
18     WHEN HOUR(`Date`) = 14 THEN '14'
19     WHEN HOUR(`Date`) = 15 THEN '15'
20     WHEN HOUR(`Date`) = 16 THEN '16'
21     WHEN HOUR(`Date`) = 17 THEN '17'
22     WHEN HOUR(`Date`) = 18 THEN '18'
23     WHEN HOUR(`Date`) = 19 THEN '19'
24     WHEN HOUR(`Date`) = 20 THEN '20'
25     WHEN HOUR(`Date`) = 21 THEN '21'
26     WHEN HOUR(`Date`) = 22 THEN '22'
27     WHEN HOUR(`Date`) = 23 THEN '23'
28  END AS time_of_day, AVG(StepTotal) as avg_steps
29  FROM
30      bellabeat.hourlysteps
31
32  GROUP BY
33      time_of_day
34  ORDER BY
35      avg_steps DESC
--      FIELD(time_of_day, '0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14',
```

Obtivemos o seguinte resultado:

time_of_day	avg_steps
19	563.1170
18	558.4618
12	540.9011
14	513.5209
17	507.0876
13	503.5844
16	477.9594
11	460.1958
10	459.5293
9	436.7859
15	403.9734
8	401.1794
20	383.2472
21	287.6108
7	286.9669
22	206.9887
6	150.5034
23	113.7504
0	44.0294
5	35.4590
1	22.1570
2	13.9102
4	11.2781
3	6.9548

Onde há mais média de passos é às **19 horas** e menos às **3 da manhã**.

Agora vamos analisar o registo do sono para ver se existe algum padrão com os passos e as calorias.

Nota: Cada registo é um registo de 1 minuto de sono monitorado pelo rastreador de fitness, podemos assim, contar cada registo por utilizador como um minuto de sono. Irei calcular a média dos minutos a dormir por dia da semana código:

```

1  SELECT
2  AVG(total_sleep_min) AS avg_total_sleep_min,
3  day_of_week
4  FROM (
5  SELECT
6  id,
7  COUNT(id) AS total_sleep_min,
8  CASE
9  WHEN DAYOFWEEK(`date`) = 1 THEN 'Sunday'
10 WHEN DAYOFWEEK(`date`) = 2 THEN 'Monday'
11 WHEN DAYOFWEEK(`date`) = 3 THEN 'Tuesday'
12 WHEN DAYOFWEEK(`date`) = 4 THEN 'Wednesday'
13 WHEN DAYOFWEEK(`date`) = 5 THEN 'Thursday'
14 WHEN DAYOFWEEK(`date`) = 6 THEN 'Friday'
15 WHEN DAYOFWEEK(`date`) = 7 THEN 'Saturday'
16 END AS day_of_week
17 FROM
18 `bellabeat`.`minutes_sleep`
19 GROUP BY
20 day_of_week,
21 id
22 ) AS total_sleep
23 GROUP BY
24 day_of_week
25 ORDER BY
26 avg_total_sleep_min DESC
27 -- FIELD(day_of_week, 'Sunday', 'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Satu

```


Obtivemos:

	avg_total_sleep_min	day_of_week
▶	2439.3636	Sunday
	2414.0909	Saturday
	2205.5000	Wednesday
	2100.3333	Thursday
	2090.0000	Monday
	2073.9545	Tuesday
	1847.5455	Friday

Em média, onde se dormiu mais foi no **domingo** e menos foi na **sexta**.

Agora vamos calcular por cada hora do dia

código:

```
1 SELECT
2     AVG(total_sleep_min),
3     time_of_day
4 FROM (
5     SELECT
6         id,
7         COUNT(id) AS total_sleep_min,
8         CASE
9             WHEN HOUR(date_time)=0 THEN '0'
10            WHEN HOUR(date_time)=1 THEN '1'
11            WHEN HOUR(date_time)=2 THEN '2'
12            WHEN HOUR(date_time)=3 THEN '3'
13            WHEN HOUR(date_time)=4 THEN '4'
14            WHEN HOUR(date_time)=5 THEN '5'
15            WHEN HOUR(date_time)=6 THEN '6'
16            WHEN HOUR(date_time)=7 THEN '7'
17            WHEN HOUR(date_time)=8 THEN '8'
18            WHEN HOUR(date_time)=9 THEN '9'
19            WHEN HOUR(date_time)=10 THEN '10'
20            WHEN HOUR(date_time)=11 THEN '11'
21            WHEN HOUR(date_time)=12 THEN '12'
22            WHEN HOUR(date_time)=13 THEN '13'
23            WHEN HOUR(date_time)=14 THEN '14'
24            WHEN HOUR(date_time)=15 THEN '15'
25            WHEN HOUR(date_time)=16 THEN '16'
26            WHEN HOUR(date_time)=17 THEN '17'
27            WHEN HOUR(date_time)=18 THEN '18'
28            WHEN HOUR(date_time)=19 THEN '19'
29            WHEN HOUR(date_time)=20 THEN '20'
30            WHEN HOUR(date_time)=21 THEN '21'
31            WHEN HOUR(date_time)=22 THEN '22'
32            WHEN HOUR(date_time)=23 THEN '23'
33        END AS time_of_day
34    FROM
35        `bellabeat`.`minutesleep`
36    GROUP BY
37        time_of_day,
38        id
39 ) AS hourly_sleep
40 GROUP BY
41     time_of_day
42 ORDER BY
43     FIELD(time_of_day, '0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15')
```

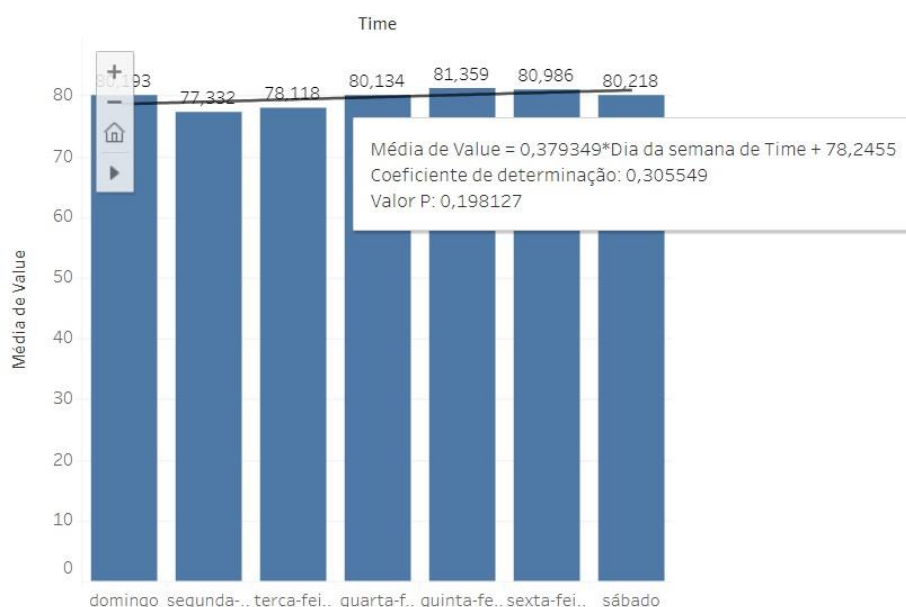
Result Grid			Filter Rows:
	AVG(total_sleep_min)	time_of_day	
▶	47.9231	18	
	48.7500	14	
	50.4000	17	
	54.9286	19	
	56.5000	13	
	56.6111	15	
	56.8889	16	
	61.8667	12	
	84.8889	20	
	112.5882	11	
	187.8571	10	
	260.8571	21	
	308.1818	9	
	498.8182	8	
	595.8182	22	
	826.1429	7	
	1020.0909	23	
	1182.6818	6	
	1265.6364	0	
	1413.8636	1	
	1521.6818	5	
	1562.2273	2	
	1643.8636	4	
	1681.9545	3	

Podemos concluir que a hora em que se dorme mais é às **3 da manhã** e a que menos se dorme é às **18 horas**.

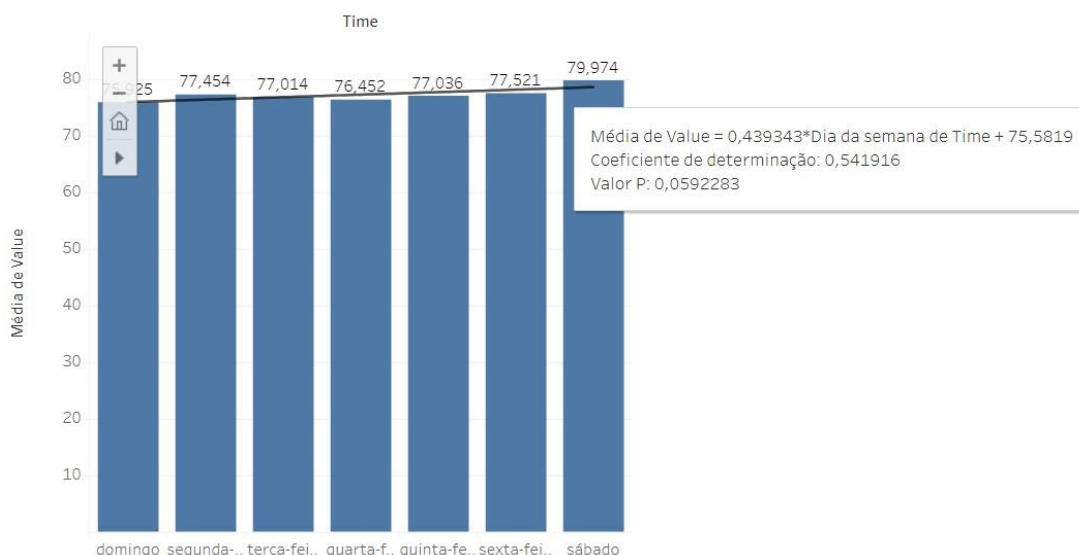
Estes resultados demonstram que existe uma correlação entre as **horas dormidas** e as **calorias queimadas**, enquanto os **passos** estão bastante próximos dessa correlação (diferenciando-se apenas em que a hora com maior média de passos é às 19 horas).

Analisar no Tableau

Sendo que não consegui carregar os dados do *sec_hearttrate* no SQL por se tratar de um ficheiro demasiado pesado, tratei desses dados no Tableau, analisando a média pelos dias da semana e saber a correlação entre março/abril e abril/maio.

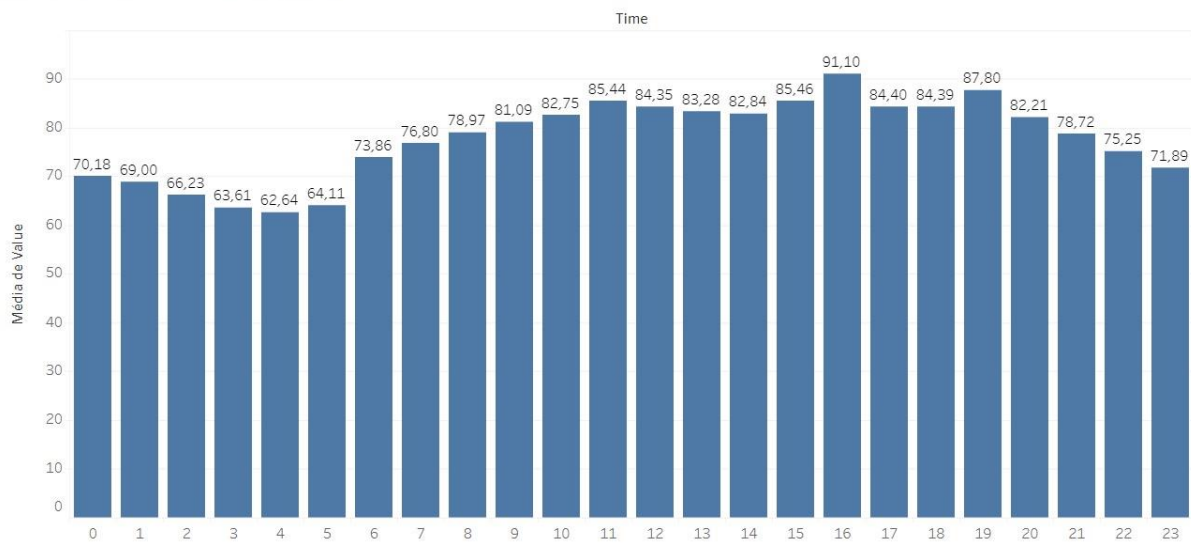


Podemos observar que o coeficiente de correlação entre março e abril é de aproximadamente 0,31 o que corresponde a uma **fraca correlação**.

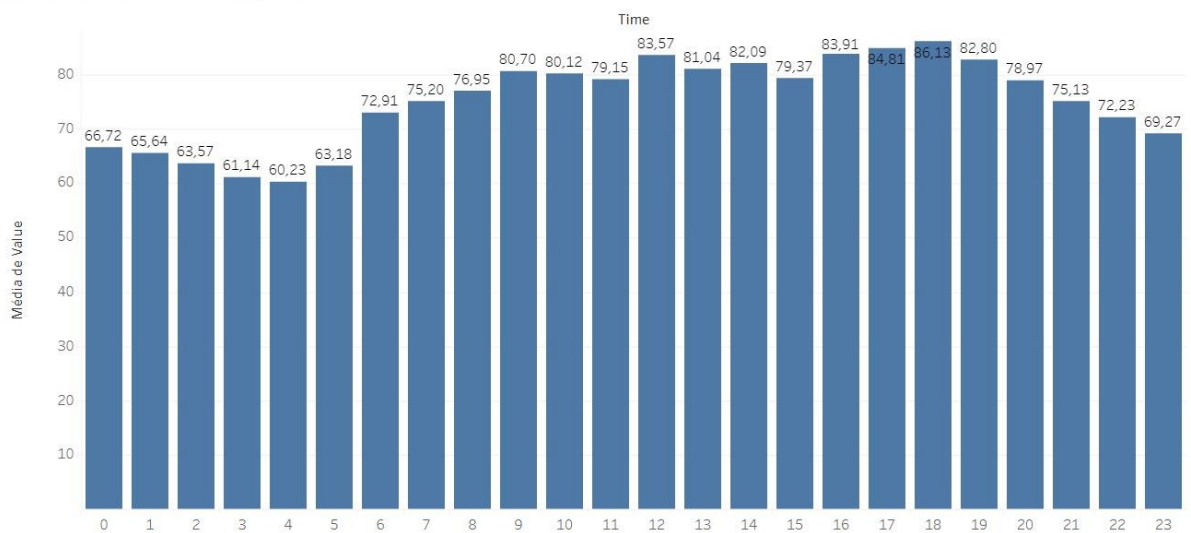


Já entre abril e maio, observamos um modesto coeficiente de correlação de 0,54. Fui calcular também a média da frequência cardíaca em cada hora do dia:

Média por hora em março/abril



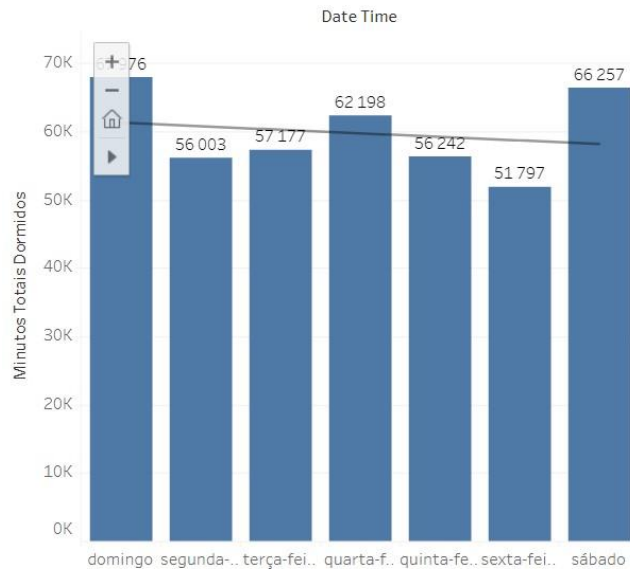
Média por hora em abril/maio



Conseguimos perceber que as frequências cardíacas mais baixas nos utilizadores são às **4 da manhã** em ambos os meses e a maior em março/abril é às **16 horas** e em abril/maio foi às **18 horas**.

Fomos também perceber se havia uma tendência do total de minutos dormidos e total de minutos de atividade com os dias da semana e este foi o resultado:

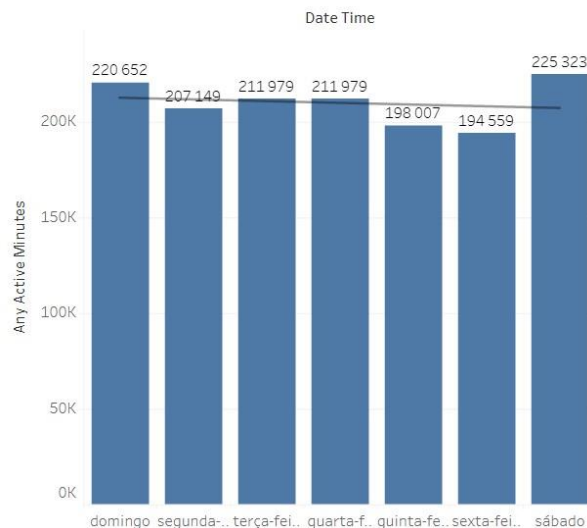
Comparação do total de minutos dormidos com o total de minutos de atividade



Apresentando um coeficiente de correlação de 0,04 aproximadamente, conseguimos perceber que **não existe essa tendência**.

Fizemos o mesmo para o total de minutos de atividade:

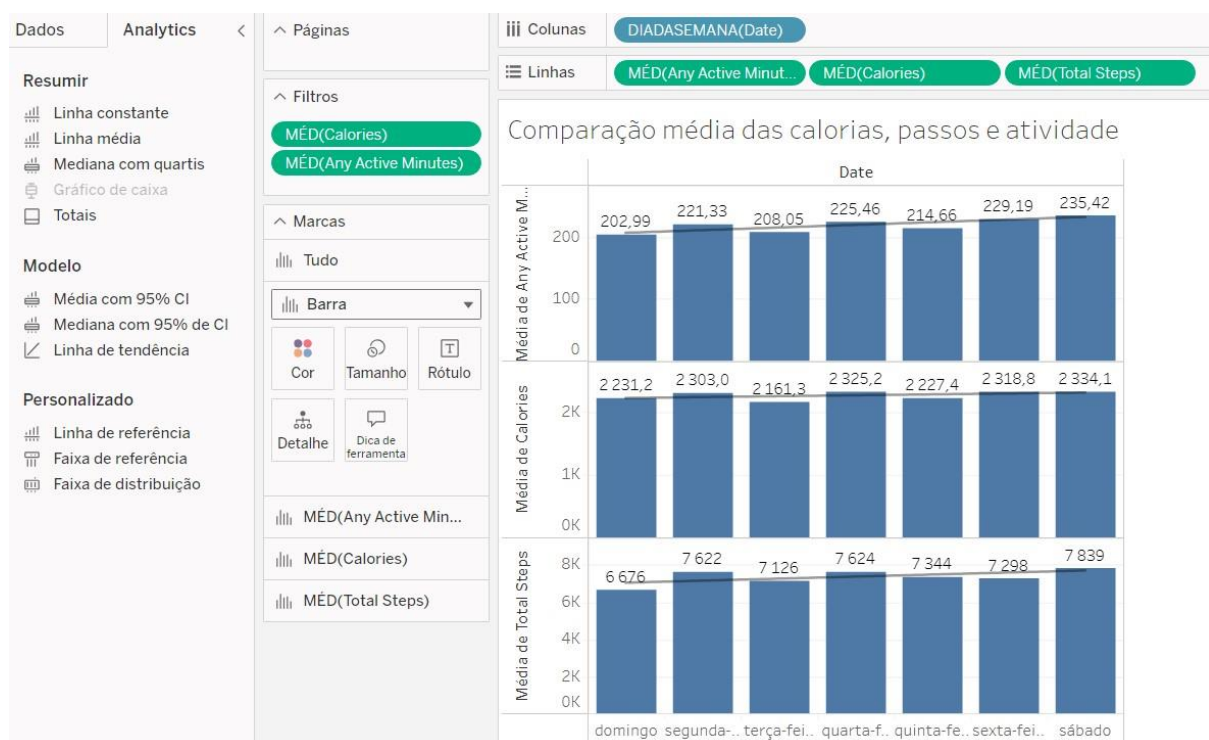
Comparação do total de minutos dormidos com o total de minutos de atividade



Ficámos a perceber que com um coeficiente de correlação de 0,03, também **não existe tendência com os dias da semana**.

Partilhar no Tableau

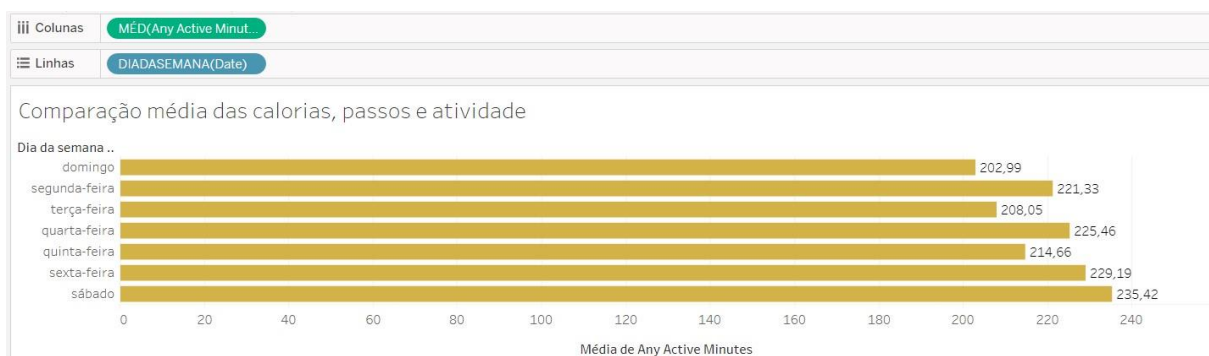
O primeiro painel mostra se há a tendência entre os minutos de atividade e as calorias e passos pelos dias da semana utilizando apenas os dados da Daily_Activity:



Concluimos com a linha de tendência nos gráficos que demonstram uma **correlação fraca na média de calorias de 0,23** aproximadamente e de **0,37 para a média de passos**. Já com uma correlação mais forte temos a média dos minutos de atividade onde apresenta **0,63** aproximadamente e portanto conseguimos perceber que existe uma pequena tendência dos minutos ativos com os dias da semana.

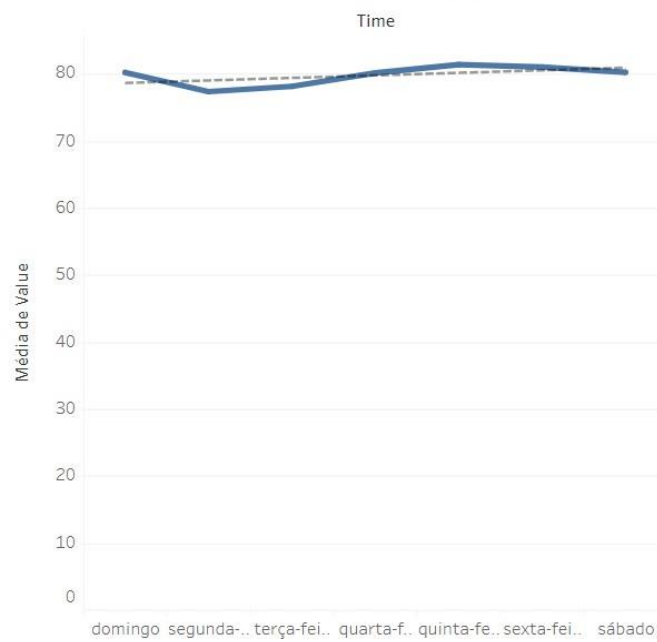
De seguida, fui comparar as médias das calorias, passos e minutos de atividade por cada dia da semana, utilizando nas calorias a base de dados “hourly_calories”, nos passos a “hourly_steps” e nos minutos de atividade a “daily_activity”.

Eis o resultado:

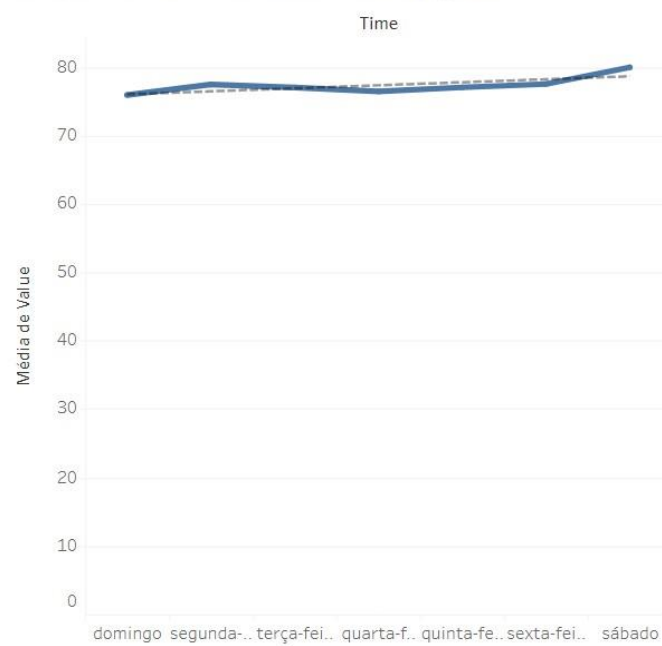


Temos também as frequências médias cardíacas de março/abril e abril/maio:

Média por dia da semana em março/abril

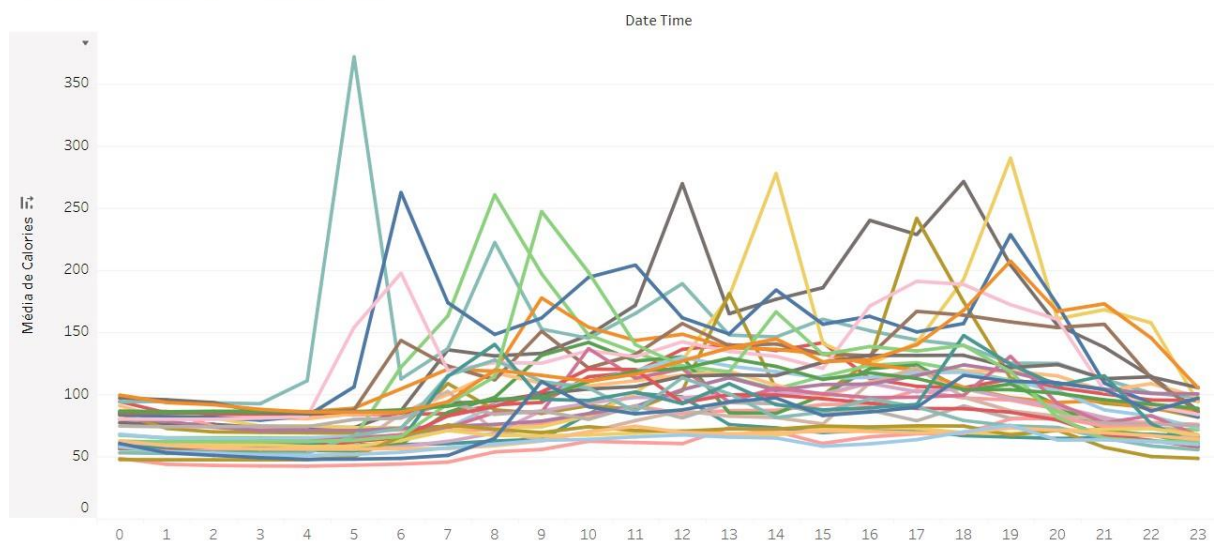


Média por dia da semana em abril/maio

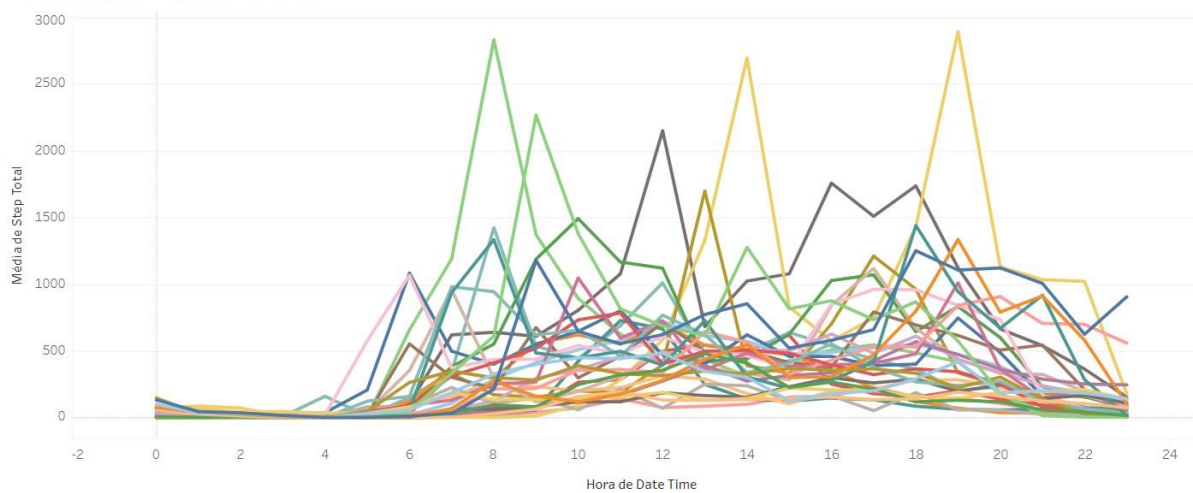


Podemos observar que os utilizadores registam os minutos mais ativos aos **sábados** e os minutos menos ativos aos **domingos**. As frequências cardíacas médias por dia da semana também aumentam ao longo da semana, embora os dados sejam limitados:

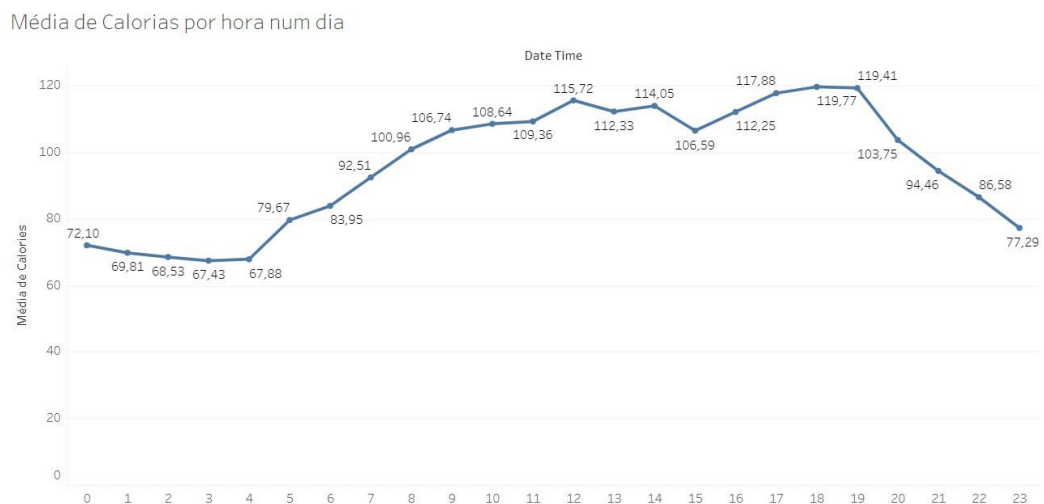
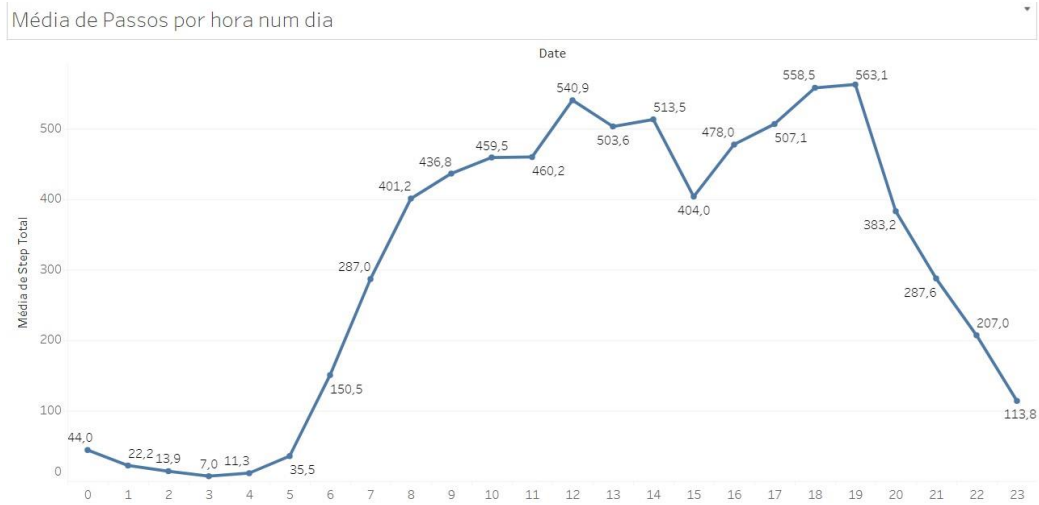
Média de calorias por hora do dia



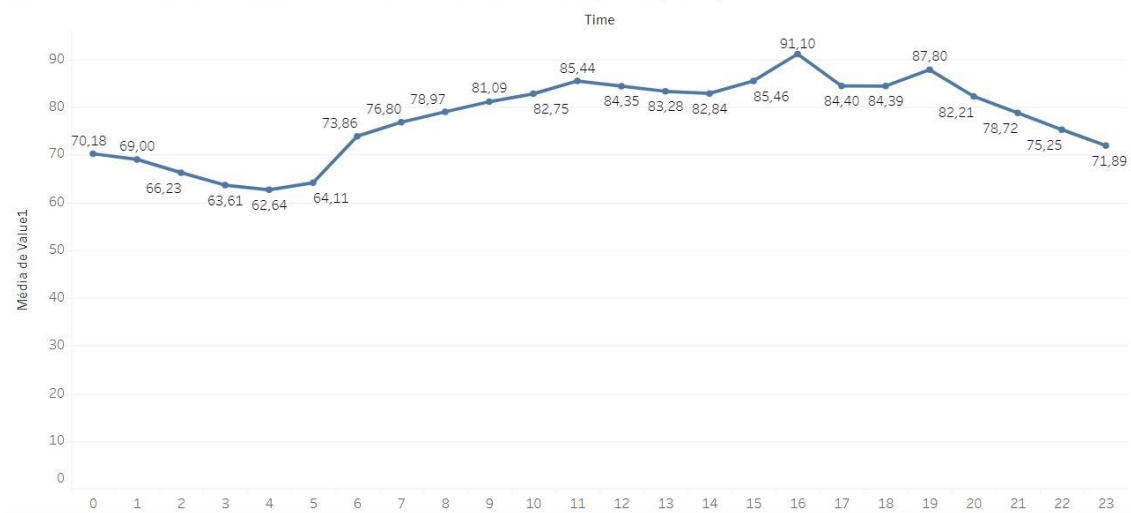
Média de passos por hora do dia



Estes painéis demonstram principalmente tendências na atividade dos utilizadores ao longo do dia. Os utilizadores são **menos ativos** e têm as frequências cardíacas **mais baixas** quando têm **o maior total de registos/minutos de sono registados**.



Média de batimentos cardíacos por hora ao longo do dia (março/abril)

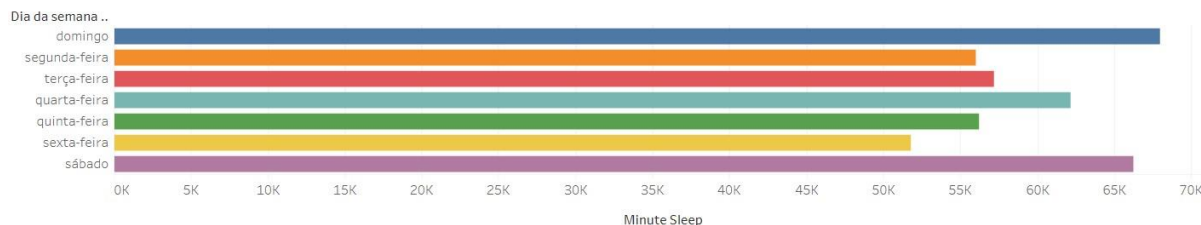


Média de batimentos cardíacos por hora ao longo do dia (abril/maio)

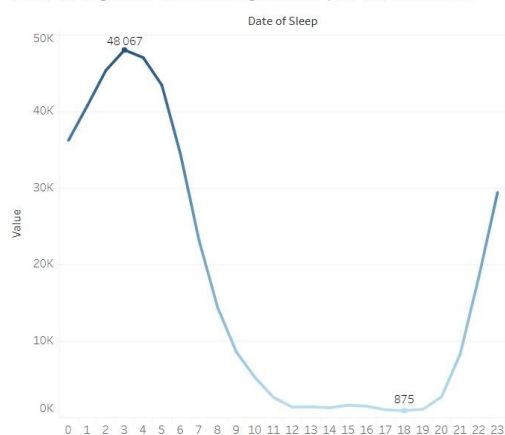


O terceiro painel compara as tendências gerais do sono. Os utilizadores registaram mais minutos de sono no **domingo** e menos minutos de sono na **sexta-feira** como podemos observar:

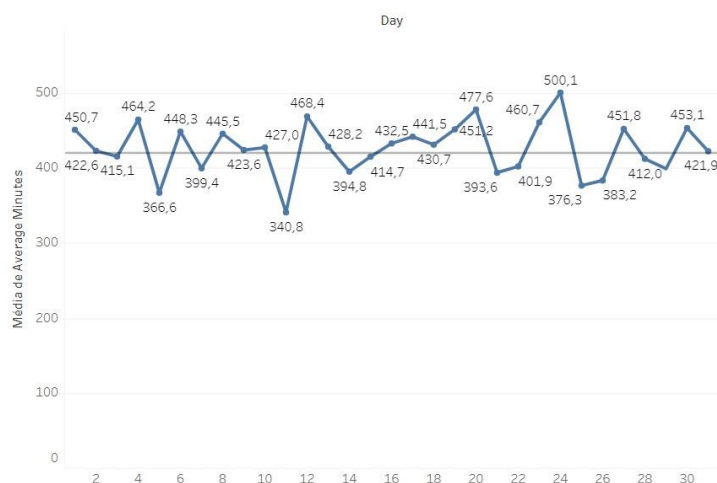
Total de minutos dormidos por cada dia da semana



Total de registos de sono registados por hora num dia



Média do total de minutos dormidos em cada dia



As horas recomendadas de sono são 7 horas e percebemos pelo gráfico que em 31 dias, apenas 19 dias foi atingido as horas recomendadas pelos utilizadores.

Atuar

Nesta fase será respondido às perguntas anteriores, feitas na fase Perguntar:

Quais são algumas das tendências no uso de dispositivos inteligentes?

Os utilizadores que usaram um rastreador de condicionamento físico mostram uma correlação moderada entre as médias das calorias e passos, com uma correlação modesta entre calorias médias e passos médios por dia da semana. A menor média de calorias e passos ocorreu no domingo e a maior no sábado, o que também se correlacionou com a média de minutos ativos dos utilizadores. Isso pode ser reflexo dos horários de trabalho das pessoas, pois é comum que as pessoas descansem aos domingos e tenham mais tempo livre nas noites de sexta e sábado para sair ou praticar exercícios. Curiosamente, a frequência cardíaca média dos utilizadores também mostrou uma correlação modesta com os dias da semana, mas não há dados suficientes para fazer qualquer análise.

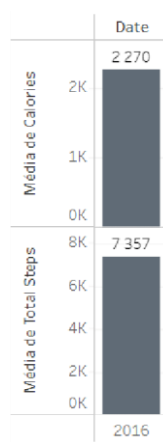
A frequência média cardíaca dos utilizadores aumenta a partir das 4 horas da manhã e aumenta até ao meio-dia, caindo ligeiramente antes de atingir o pico às 16h, diminuindo e aumentando ligeiramente às 19h antes de diminuir continuamente até às 4 horas da manhã.

Complementando com o total de registos de sono observados por hora ao longo do dia, demonstra que a maioria dos utilizadores tende a dormir das 19 horas às 4 horas da manhã sensivelmente. Embora o gráfico mostre que o registo mais baixo de minutos de sono é às 3 da manhã, a frequência cardíaca mais baixa é às 4 da manhã e portanto, o utilizador normalmente dorme até às 3 da manhã ou 4. Isto pode demonstrar que alguns podem fazer uma sesta entre o meio-dia e 14h e entre 16h e 18h ou que os utilizadores exercitam ou são mais ativos durante o meio-dia e à noite, mas não há dados suficientes para tirar quaisquer conclusões.

Os utilizadores que usaram um rastreador de fitness também tendem a monitorar o sono. 22 usuários de um total de 33 usuários usaram os registos de sono.

Segundo os dados, os users dormem mais aos domingos e menos às sextas. Novamente, isso pode ser um reflexo do trabalho e das agendas pessoais dos utilizadores. Os utilizadores atingiram a duração diária recomendada do sono apenas em 19 dos 31 dias.

Em média são queimadas 2.270 calorias e caminham 7.357 passos por dia.



2. Como é que estas tendências podem ser aplicadas aos clientes da Bellabeat?

Segundo os dados, as mulheres que usam rastreadores de exercício físico têm maior probabilidade de queimar mais calorias com mais passos.

Devido a muitos clientes serem menos ativos aos domingos e mais ativos aos sábados, sugere-se em adaptar recomendações em torno dos horários de trabalho e descanso dos utilizadores.

Observar o aumento da frequência cardíaca pode orientar recomendações de saúde personalizadas, especialmente em relação à idade ou a alguma doença associada.

2 em cada 3 utilizadores de rastreadores de fitness monitoram o seu sono, portanto, os clientes atuais do Bellabeat também são extremamente propensos a monitorar o sono.

A média por utilizador do monitor de condicionamento físico não atingiu a quantidade recomendada de sono por dia em 38,71% do tempo, refletindo que muitos utilizadores não dormem o suficiente (apenas foi atingido em 19 dos 31 dias observados).

3. Como é que estas tendências podem ajudar a influenciar a estratégia de marketing da BellaBeat?

Iremos considerar como se estas tendências pudessem influenciar a subscrição da BellaBeat com recomendações personalizadas

Para as mulheres que têm metas para a perda de peso, a Bellabeat poderia destacar que o monitoramento de passos poderia ajudar ao gasto calórico e a incentivar a movimentar nos horários em que as pessoas fazem menos atividades sem estar durante as horas de sono.

A compreensão dos horários de trabalho e descanso do utilizador, pode fornecer recomendações personalizadas de atividade ou sono, como atividades de baixo impacto nos dias úteis e treinos mais intensos em finais de semana, que é onde os utilizadores têm mais tempo livre normalmente. A Bellabeat também poderá considerar a atividade registada recentemente do utilizador e recomendar exercícios como alongamentos após uma sessão de caminhada registada ou um breve exercício cardiovascular após um longo trajeto.

Mais recomendações também poderiam ser feitas aos clientes com base nas tendências da frequência cardíaca, como um lembrete de relaxamento baseado em picos irregulares fora da programação dos utilizadores.

Complementando à frequência cardíaca, as recomendações seriam reforçadas com mais informações sobre os minutos ativos ao longo do dia.

A Bellabeat poderia considerar a implementação de um sistema de recompensas para frequência cardíaca regular e/ou monitoramento de minutos ativos para aumentar o envolvimento e produzir mais dados para entender melhor as tendências dos seus clientes.

Com uma alta percentagem de utilizadores monitorando o sono com rastreadores de fitness, a Bellabeat beneficiava-se muito com o fortalecimento dos aspetos relacionados ao sono no seu aplicativo.

A Bellabeat poderia implementar o uso de sons/notas de áudio para dormir, numa subscrição premium da BellaBeat, ajudando assim os seus clientes a dormir melhor.

A gravação do áudio e a poluição luminosa dos utilizadores ao seu redor enquanto dormem iria fornecer aos seus clientes, informações sobre o seu ambiente de sono e melhorar o tempo e a qualidade do sono.

Considerações Pessoais

Para ter uma análise mais precisa, recomenda-se uma visão completa e mais ampla dos aspectos interessantes.

Um conjunto de dados com informações sobre períodos mais longos de muito mais utilizadores poderia permitir a análise do impacto das estações e feriados na atividade das pessoas.

Alguns dados como sexo ou idade, seriam úteis para fazer recomendações mais específicas para os clientes do Bellabeat.

Também seria interessante complementar o conjunto de dados atual com ingestão de água ao longo do tempo, período menstrual registado e rastreador de stress diário e minutos de atenção plena.