

Case Study: Analisi strategica per un'azienda del benessere digitale

Enrico di Capua - Google Data Analytics Certificate

2025-06-18

Contents

1	Ask	3
1.1	Business Task	3
2	Data Source	4
2.1	Data Quality	6
2.2	Privacy, Security, and Accessibility	6
3	Prepare	7
3.1	Data Cleaning	7
4	Analyze	8
4.1	Insight 1: Attività fisica e sedentarietà	8
4.1.1	Coefficiente di correlazione Pearson	9
4.2	Insight 2: Sonno e attività fisica	9
4.3	Insight 3: Durata media del sonno	11
4.3.1	Durata media del sonno – utenti Fitbit (in ore)	11
4.3.2	Durata media del sonno – utenti non-Fitbit (in ore)	12
4.3.3	Durata media del sonno - Confronto utenti Fitbit e non-Fitbit	13
4.4	Insight 4: Qualità del sonno	14
4.4.1	Qualità del sonno - utenti Fitbit	14

4.4.2	Qualità del sonno - utenti non-Fitbit	15
4.4.3	Confronto della qualità del sonno	16
4.5	Insight 5: Confronto frequenza cardiaca	18
5	Act	20
5.1	Conclusione	22



1 Ask

Fondata nel 2013 da Urška Sršen e Sando Mur, Bellabeat è un'azienda high-tech all'avanguardia, specializzata nella produzione di dispositivi intelligenti per la salute e il benessere femminile. I prodotti offerti dalla compagnia sono in grado di monitorare attività fisica, sonno, stress e salute riproduttiva, incentivando un migliore e più corretto stile di vita. In questo contesto, l'analisi dei dati biometrici diventa un importante strumento per comprendere il comportamento delle utenti, anticipare le tendenze e rafforzare il posizionamento di Bellabeat come leader del benessere guidato dai dati, fornendo soluzioni sempre più innovative ed esteticamente apprezzabili.

1.1 Business Task

Come junior data analyst nel team marketing di Bellabeat, conduco attraverso RStudio, un'analisi sui dati ricavati dall'utilizzo quotidiano di dispositivi smart Fitbit, con l'obiettivo di ottenere insight da integrare nella futura campagna di marketing dell'azienda.

L'analisi si pone l'obiettivo di:

- Comprendere la relazione tra sonno, attività fisica e salute
- Dare consigli per promuovere abitudini salutari e strategie personalizzate
- Identificare target per campagne di marketing

Una volta formulati gli insight principali, questi verranno collegati a uno dei prodotti di punta di Bellabeat, selezionato in funzione dei comportamenti osservati. Lo scopo finale è proporre delle raccomandazioni concrete che guidino le scelte di comunicazione e promozione dei prodotti dell'azienda nel breve e lungo termine.

2 Data Source

Il dataset consigliato da Sršen è *FitBitFitnessTrackerData* (CC0: Dominio Pubblico, dataset reso disponibile da *Mobius* ed è stato generato tramite sondaggi distribuiti da Amazon Mechanical Turk tra il 12 Aprile 2016 e il 12 Maggio 2016. Attraverso i fitness tracker di trenta utenti Fitbit sono stati registrati parametri inerenti l'attività fisica come la frequenza cardiaca, il monitoraggio del sonno o il numero di passi giornalieri delle utenti, che utilizzerò per esplorare le abitudini delle stesse e ricavarne dei pattern. Nella trattazione, mi riferirò ai dati delle utenti provenienti da questo dataset con la dicitura "Fitbit".

```
# Importazione dei dataset activity, sleep e weight contenuti in Fitbase
path <- "Fitabase Data 4.12.16-5.12.16/"
activity <- read_csv(paste0(path, "dailyActivity_merged.csv"))
sleep <- read_csv(paste0(path, "sleepDay_merged.csv"))
heartrate <- read_csv(paste0(path, "heartrate_seconds_merged.csv"))
# Panoramica dei dati:
glimpse(activity)

## Rows: 940
## Columns: 15
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
## $ TotalSteps <dbl> 13162, 10735, 10460, 9762, 12669, 9705, 13019~
## $ TotalDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ TrackerDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ LightActiveDistance <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveMinutes <dbl> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ FairlyActiveMinutes <dbl> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ LightlyActiveMinutes <dbl> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ SedentaryMinutes <dbl> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ Calories <dbl> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203~
```

```
glimpse(sleep)
```

```
## Rows: 413
## Columns: 5
## $ Id          <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~
## $ SleepDay     <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "~
## $ TotalSleepRecords <dbl> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ TotalMinutesAsleep <dbl> 327, 384, 412, 340, 700, 304, 360, 325, 361, 430, 2~
## $ TotalTimeInBed   <dbl> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 3~
```

```
glimpse(heartrate)
```

```
## Rows: 2,483,658
## Columns: 3
## $ Id    <dbl> 2022484408, 2022484408, 2022484408, 2022484408, 2022484408, 2022~
## $ Time  <chr> "4/12/2016 7:21:00 AM", "4/12/2016 7:21:05 AM", "4/12/2016 7:21:~
## $ Value <dbl> 97, 102, 105, 103, 101, 95, 91, 93, 94, 93, 92, 89, 83, 61, 60, ~
```

Dal momento che i dati presenti nel database Fitbit Fitness Tracker presentano alcune limitazioni, ho ritenuto utile introdurre nell'analisi anche il dataset *SleepHealthandLifestyleDataset* (CC0: Dominio Pubblico, reso disponibile da *LaksikaTharmalingam*. Quest'ultimo comprende 400 righe e 13 colonne, coprendo un'ampia gamma di variabili self-reported, relative alla salute e alle abitudini quotidiane delle utenti, come durata e qualità del sonno o livello di attività fisica, e dati demografici come età e professione. Nella trattazione, mi riferirò ai dati delle utenti provenienti da questo dataset con la dicitura “non-Fitbit”.

```
sleep_health_global <- read_csv("Sleep_health_and_lifestyle_dataset.csv")
colnames(sleep_health_global) <- gsub(" ", ".", colnames(sleep_health_global))
# Panoramica dei dati:
glimpse(sleep_health_global)
```

```
## Rows: 374
## Columns: 13
## $ Person.ID    <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,~
## $ Gender       <chr> "Male", "Male", "Male", "Male", "Male", "Male"~
## $ Age          <dbl> 27, 28, 28, 28, 28, 28, 29, 29, 29, 29, 29, 29~
## $ Occupation   <chr> "Software Engineer", "Doctor", "Doctor", "Sale~
## $ Sleep.Duration <dbl> 6.1, 6.2, 6.2, 5.9, 5.9, 5.9, 6.3, 7.8, 7.8, 7~
```

```
## $ Quality.of.Sleep      <dbl> 6, 6, 6, 4, 4, 4, 6, 7, 7, 7, 6, 7, 6, 6, 6, 6~
## $ Physical.Activity.Level <dbl> 42, 60, 60, 30, 30, 30, 40, 75, 75, 75, 30, 75~
## $ Stress.Level          <dbl> 6, 8, 8, 8, 8, 8, 7, 6, 6, 6, 8, 6, 8, 8, 8, 8~
## $ BMI.Category          <chr> "Overweight", "Normal", "Normal", "Obese", "Ob~
## $ Blood.Pressure        <chr> "126/83", "125/80", "125/80", "140/90", "140/9~
## $ Heart.Rate            <dbl> 77, 75, 75, 85, 85, 85, 82, 70, 70, 70, 70, 70~
## $ Daily.Steps           <dbl> 4200, 10000, 10000, 3000, 3000, 3000, 3500, 80~
## $ Sleep.Disorder        <chr> "None", "None", "None", "Sleep Apnea", "Sleep ~
```

2.1 Data Quality

Entrambi i dataset soddisfano parzialmente i criteri ROCCC (Reliable, Original, Comprehensive, Current, Cited).

Il dataset Fitabase è composto da dati reali raccolti tramite dispositivi e quindi può considerarsi affidabile (Reliable) e originale (Original), ma non aggiornato (Current), poiché si riferisce al periodo di tempo (aprile–maggio 2016). Inoltre è completo (Comprehensive) per quanto riguarda il tracking quotidiano (sonno, attività, calorie, passi).

Il dataset Sleep Health and Lifestyle è invece sintetico (generato artificialmente), quindi non è esattamente un dataset originale, ma è utile per scopi didattici.

2.2 Privacy, Security, and Accessibility

Entrambi i dataset non contengono dati personali identificabili (PII – Personally Identifiable Information), come nomi, indirizzi o email.

Dal punto di vista dell’accessibilità, entrambi i dataset sono pubblicamente disponibili, il che li rende utilizzabili da una vasta gamma di utenti e strumenti di analisi.

Nel complesso, i dati sono **appropriati** per una prima analisi, con un buon compromesso tra rappresentatività e scopo dell’analisi.

3 Prepare

3.1 Data Cleaning

Poiché Bellabeat si rivolge esclusivamente a un pubblico femminile, applico una pulizia del dataset Sleep Health & Lifestyle per escludere i soggetti di genere maschile rendendo l'analisi più mirata e rappresentativa del target aziendale. Inoltre, per valutare con maggiore chiarezza l'impatto dell'attività fisica sul benessere psico-fisico, ho escluso le utenti con livelli di attività fisica molto elevati (valori superiori a 50 su una scala da 1 a 100), potendo in questo modo confrontare i parametri delle utenti con attività elevata non.

```
sleep_health <- sleep_health_global %>%
  filter(Gender == "Female" & Physical.Activity.Level<50)
# Panoramica dei dati:
glimpse(sleep_health)

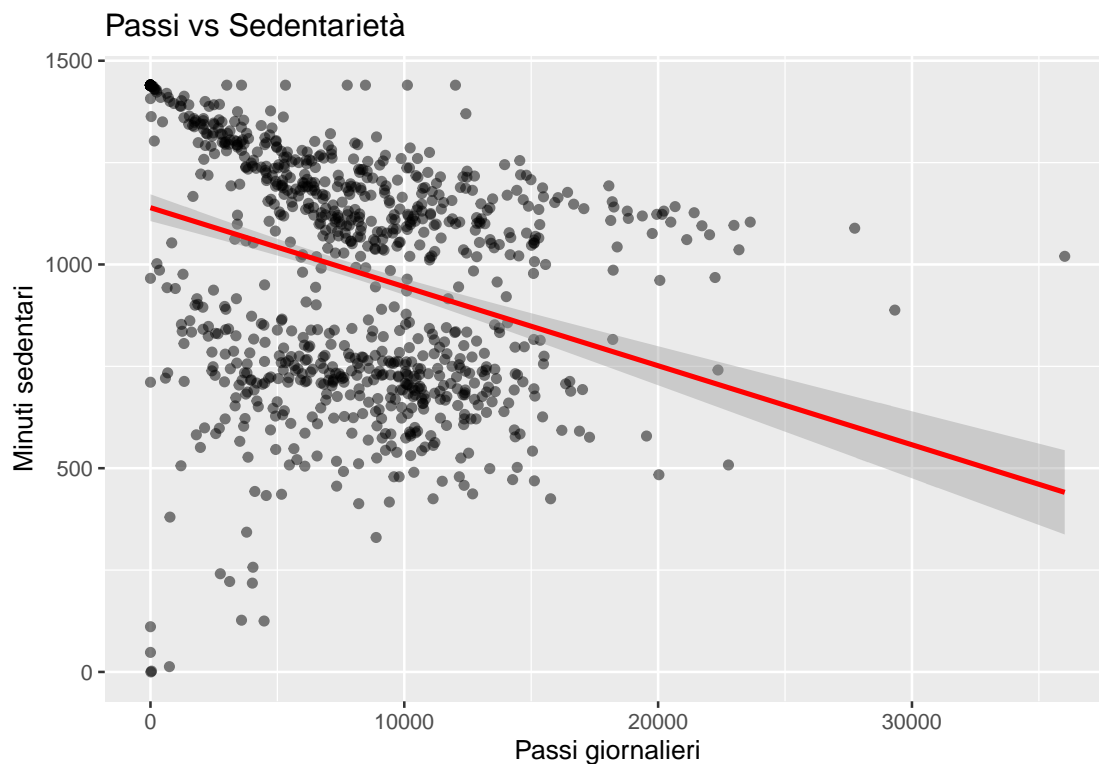
## Rows: 74
## Columns: 13
## $ Person.ID          <dbl> 17, 19, 31, 32, 81, 82, 107, 185, 186, 187, 18~
## $ Gender             <chr> "Female", "Female", "Female", "Female", "Femal~
## $ Age                <dbl> 29, 29, 30, 30, 34, 34, 37, 42, 42, 43, 43, 43~
## $ Occupation         <chr> "Nurse", "Nurse", "Nurse", "Nurse", "Scientist~
## $ Sleep.Duration     <dbl> 6.5, 6.5, 6.4, 6.4, 5.8, 5.8, 6.1, 6.8, 6.8, 6~
## $ Quality.of.Sleep   <dbl> 5, 5, 5, 5, 4, 4, 6, 6, 6, 7, 7, 7, 7, 7, 7~
## $ Physical.Activity.Level <dbl> 40, 40, 35, 35, 32, 32, 42, 45, 45, 45, 45~
## $ Stress.Level       <dbl> 7, 7, 7, 7, 8, 8, 6, 7, 7, 4, 4, 4, 4, 4, 4~
## $ BMI.Category       <chr> "Normal Weight", "Normal Weight", "Normal Weig~
## $ Blood.Pressure     <chr> "132/87", "132/87", "130/86", "130/86", "131/8~
## $ Heart.Rate         <dbl> 80, 80, 78, 78, 81, 81, 77, 78, 78, 65, 65, 65~
## $ Daily.Steps        <dbl> 4000, 4000, 4100, 4100, 5200, 5200, 4200, 5000~
## $ Sleep.Disorder     <chr> "Sleep Apnea", "Insomnia", "Sleep Apnea", "Ins~
```

4 Analyze

4.1 Insight 1: Attività fisica e sedentarietà

Per comprendere più a fondo i modelli comportamentali delle utenti, è fondamentale analizzare in che modo i livelli di attività fisica influenzino le abitudini sedentarie durante la giornata. L'obiettivo è individuare eventuali correlazioni tra il movimento quotidiano e i periodi di inattività. Il seguente grafico mostra la relazione tra il numero di passi giornalieri (indicatore del livello di attività fisica) e i minuti di sedentarietà, ovvero i periodi della giornata in cui non viene rilevato alcun movimento, per utenti Fitbit.

```
ggplot(activity, aes(x = TotalSteps, y = SedentaryMinutes)) +  
  geom_point(alpha = 0.5) +  
  geom_smooth(method = "lm", color = "red") +  
  labs(title = "Passi vs Sedentarietà", x = "Passi giornalieri", y = "Minuti sedentari")
```



I risultati mostrano una tendenza prevedibile: le persone che camminano di più tendono a essere meno sedentarie. Sebbene questo non sorprenda, tuttavia fornisce un valido riscontro sulla qualità dei dati a nostra disposizione, confermando la loro coerenza, affidabilità, e il fatto che le informazioni raccolte sono in linea con le aspettative.

4.1.1 Coefficiente di correlazione Pearson

Il coefficiente di correlazione Pearson misura la forza e la direzione della relazione lineare tra due variabili. Nel mio caso le variabili sono TotalSteps, che indica i passi totali fatti e SedentaryMinutes, che indica i minuti sedentari. Inoltre ho escluso le righe che contenevano valori mancanti NA.

```
cor(activity$TotalSteps, activity$SedentaryMinutes, use = "complete.obs")
```

```
## [1] -0.3274835
```

Insight: correlazione moderatamente negativa ($r = -0.33$) → chi cammina di più tende ad essere meno sedentaria. Tuttavia esistono eccezioni, ad esempio intense camminate seguite da lunghi periodi di inattività oppure intensa attività fisica non attribuibile a una camminata.

Strategia data-driven consigliata: suggerire brevi sessioni di movimento nei giorni caratterizzati da alta sedentarietà tramite notifiche personalizzate. L'attività fisica può essere incentivata attraverso un programma a premi, che ricompensi gli utenti al raggiungimento di obiettivi minimi giornalieri (gamification). Questo approccio potrebbe contribuire a ridurre i livelli medi di sedentarietà.

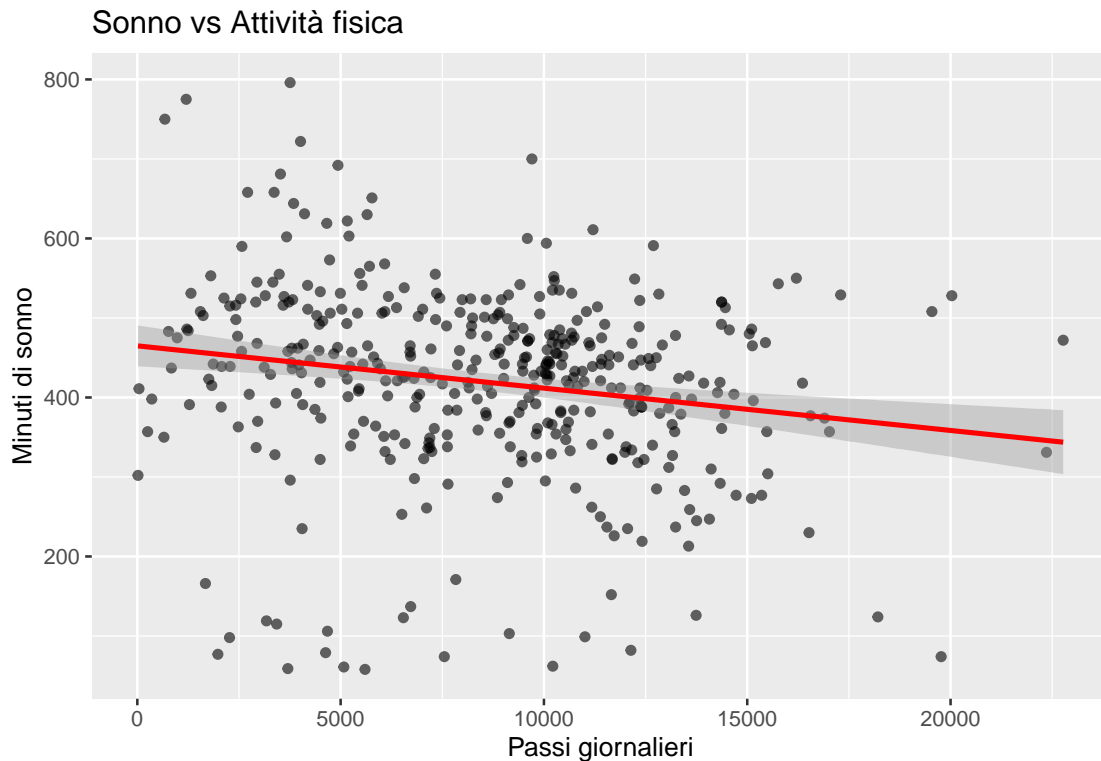
4.2 Insight 2: Sonno e attività fisica

Voglio scoprire se esista una correlazione tra il numero di passi giornalieri e la quantità di sonno nella notte **dello stesso giorno** per utenti allenate, ovvero Fitbit. Ciò che voglio indagare è se, in alcune situazioni, dormire più del solito possa essere legato a un bisogno di recupero dopo giornate intense oppure a uno stile di vita più sedentario. A tal fine è stato necessario uniformare i formati delle date nei due dataset. In particolare, nel dataset activity la colonna ActivityDate è stata convertita da stringa a formato date. Nel dataset sleep, invece, la colonna SleepDay è stata trasformata in date partendo da un formato con data e ora. Un comando inner_join permette di unire i due dataset, in maniera che le due date coincidano per ogni utente.

```
activity <- activity %>% mutate(ActivityDate = mdy(ActivityDate))
sleep <- sleep %>% mutate(SleepDay = as_date(mdy_hms(SleepDay)))
merged <- inner_join(activity, sleep, by = c("Id", "ActivityDate" = "SleepDay"))
```

Il grafico risultante mostra la relazione tra passi giornalieri e minuti di sonno registrati:

```
ggplot(merged, aes(x = TotalSteps, y = TotalMinutesAsleep)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "red") +
  labs(title = "Sonno vs Attività fisica", x = "Passi giornalieri", y = "Minuti di sonno")
```



```
cor( merged$TotalSteps, merged$TotalMinutesAsleep, use = "complete.obs")
```

```
## [1] -0.1868665
```

Insight: c'è una correlazione leggermente negativa ($r = -0.19$) e cioè chi si muove di meno tende a dormire di più. Questo riflette il trend comportamentale delle utenti nei periodi di bassa attività fisica, e cioè probabilmente i periodi di “recupero”.

Strategia data-driven consigliata: classificare le giornate come “recupero” vs “attive”. Nelle giornate “recupero” fornire consigli mirati tipo: *“Hai dormito molto? Oggi ti suggeriamo yoga o stretching leggero.”* Questo potrebbe aiutare a svolgere attività fisica monitorata e controllata, seppur meno intensa, durante i periodi di recupero.

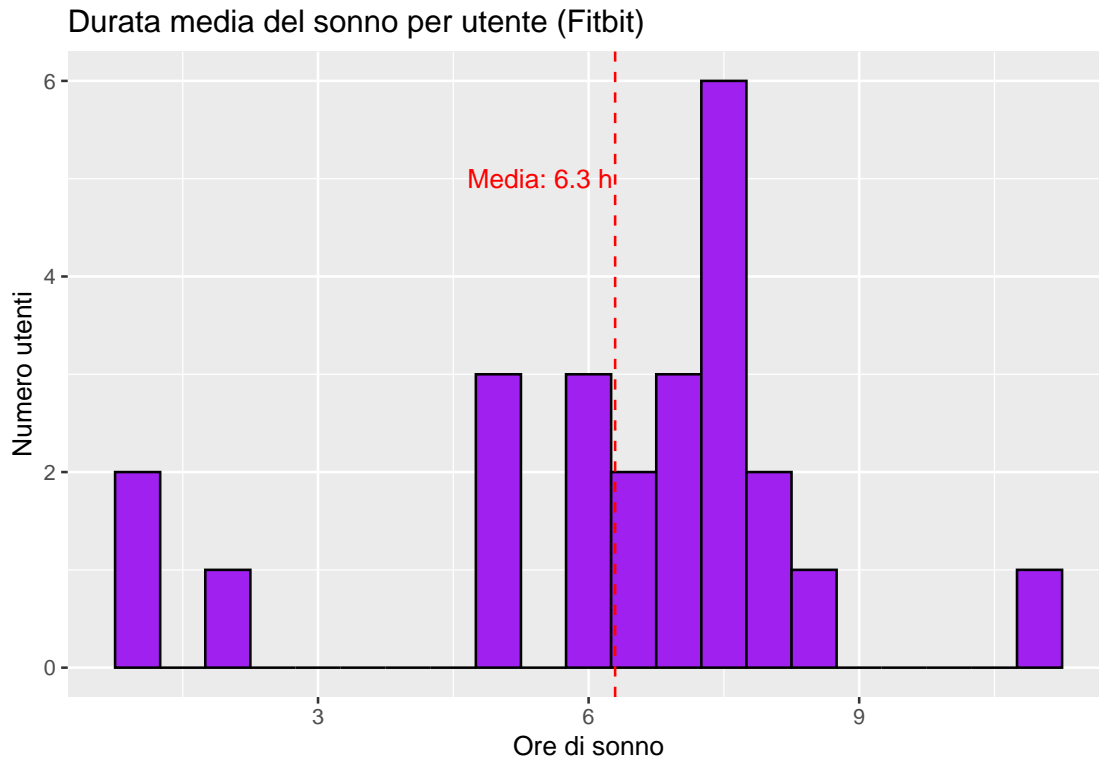
4.3 Insight 3: Durata media del sonno

È interessante confrontare le ore di sonno medie delle utenti Fitbit e quelle delle utenti non-Fitbit, ovvero quelle del dataset Sleep health and lifestyle che praticano poca attività fisica.

4.3.1 Durata media del sonno – utenti Fitbit (in ore)

L'istogramma sottostante rappresenta la distribuzione delle ore medie di sonno per le utenti Fitbit. Ogni barra indica il numero delle utenti a cui è associata una certa durata di sonno medio, con intervallo di 0,5 ore. La linea rossa tratteggiata indica la media complessiva delle ore di sonno e il valore esatto è riportato accanto alla linea stessa.

```
# Calcolo della media delle ore di sonno per ogni utente Fitbit
sleep_avg <- sleep %>%
  group_by(Id) %>%
  summarise(AvgSleep = mean(TotalMinutesAsleep, na.rm = TRUE) / 60)
# Calcolo del valore medio delle ore di sonno tra tutti gli utenti Fitbit
sleep_avg_fitbit <- mean(sleep_avg$AvgSleep)
# Plot del risultato
ggplot(sleep_avg, aes(x = AvgSleep)) +
  geom_histogram(binwidth = 0.5, fill = "purple", color = "black") +
  labs(title = "Durata media del sonno per utente (Fitbit)",
       x = "Ore di sonno", y = "Numero utenti") +
  geom_vline(xintercept = sleep_avg_fitbit, linetype = "dashed", color = "red") +
  annotate("text", x = sleep_avg_fitbit - 0.83, y = 5,
         label = paste0("Media: ", round(sleep_avg_fitbit, 1), " h"),
         color = "red")
```



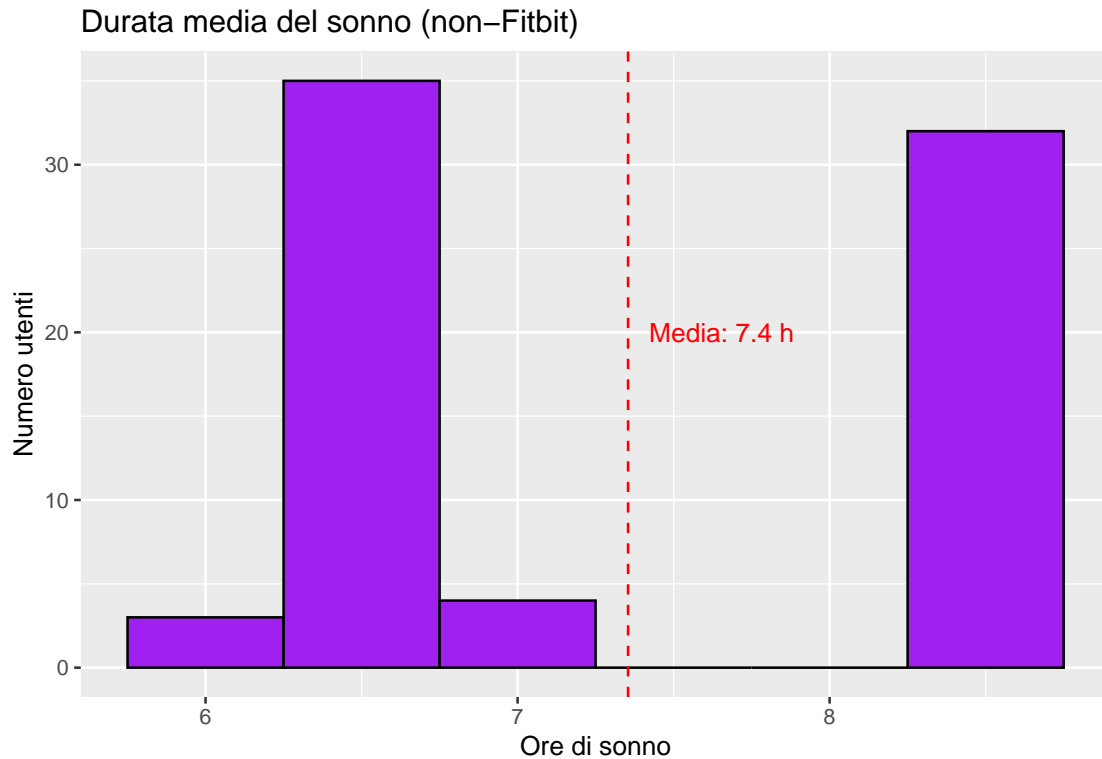
4.3.2 Durata media del sonno – utenti non-Fitbit (in ore)

Effettuo le stesse operazioni per le utenti non-Fitbit:

```
# Rimozione dei valori N/A
sleep_health_clean <- sleep_health %>%
  filter(!is.na(Sleep.Duration))

# Calcolo della media delle ore di sonno per utenti non-Fitbit
sleep_avg_non_fitbit <- mean(sleep_health_clean$Sleep.Duration)

# Plot del risultato
ggplot(sleep_health_clean, aes(x = Sleep.Duration)) +
  geom_histogram(binwidth = 0.5, fill = "purple", color = "black") +
  labs(title = "Durata media del sonno (non-Fitbit)",
       x = "Ore di sonno", y = "Numero utenti") +
  geom_vline(xintercept = sleep_avg_non_fitbit, linetype = "dashed", color = "red") +
  annotate("text", x = sleep_avg_non_fitbit + 0.3, y = 20,
         label = paste0("Media: ", round(sleep_avg_non_fitbit, 1), " h"),
         color = "red")
```

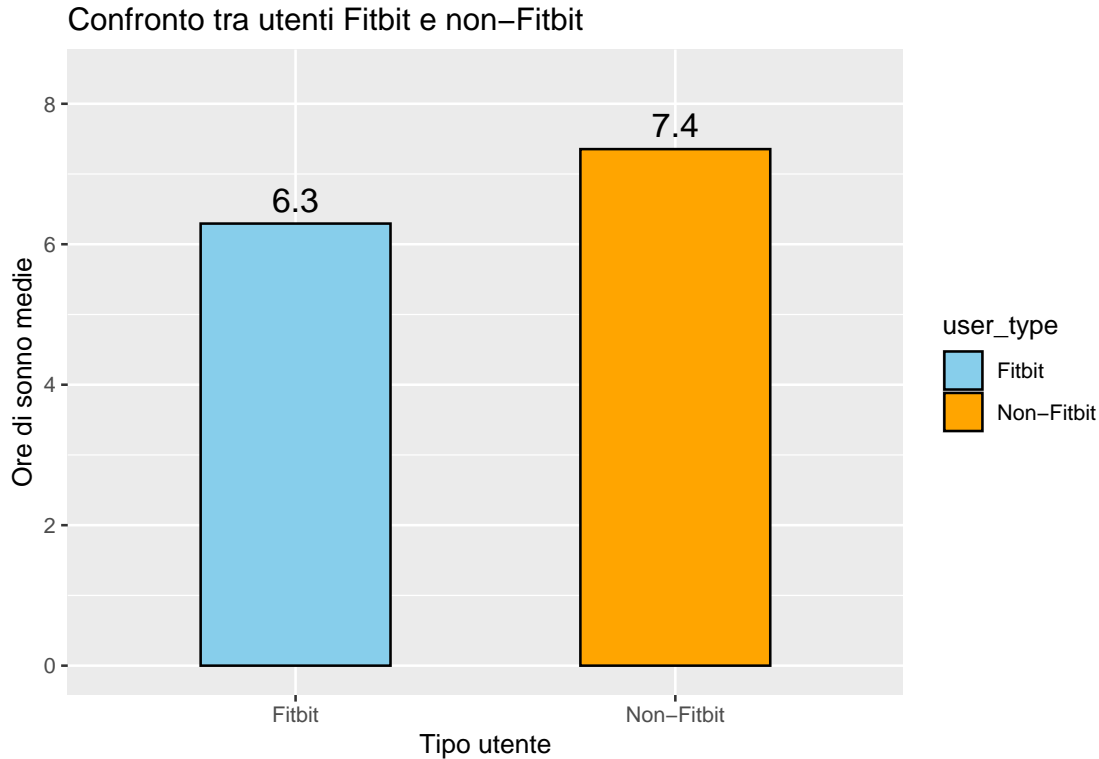


4.3.3 Durata media del sonno - Confronto utenti Fitbit e non-Fitbit

Posso infine confrontare i risultati per ricavare le informazioni cercate:

```
# Dataframe riassuntivo per il confronto:
sleep_comparison <- tibble(
  user_type = c("Fitbit", "Non-Fitbit"),
  avg_hr_sleep = c(sleep_avg_fitbit, sleep_avg_non_fitbit)
)

#Grafico comparativo:
ggplot(sleep_comparison, aes(x = user_type, y = avg_hr_sleep, fill = user_type)) +
  geom_col(width = 0.5, color = "black") +
  labs(title = "Confronto tra utenti Fitbit e non-Fitbit",
       x = "Tipo utente", y = "Ore di sonno medie") +
  scale_fill_manual(values = c("Fitbit" = "skyblue", "Non-Fitbit" = "orange")) +
  geom_text(aes(label = round(avg_hr_sleep, 1)), vjust = -0.5, size = 5) +
  ylim(0, max(sleep_comparison$avg_hr_sleep) + 1)
```



Il confronto tra le ore di sonno medie delle utenti **Fitbit** e **non-Fitbit** offre uno spunto interessante: contrariamente a quanto ci si potrebbe aspettare, i dati mostrano che le utenti con un basso livello di attività fisica dormono mediamente di più rispetto a quelle più attive. Questo risultato, seppur sorprendente, potrebbe essere influenzato da diversi fattori, ad esempio la qualità del sonno potrebbe essere inferiore nei soggetti meno attivi. È infatti importante sottolineare che una maggiore quantità di sonno non implica necessariamente una migliore qualità. Pertanto, risulta fondamentale andare a valutare la differenza della qualità del sonno tra le due tipologie di utenti.

4.4 Insight 4: Qualità del sonno

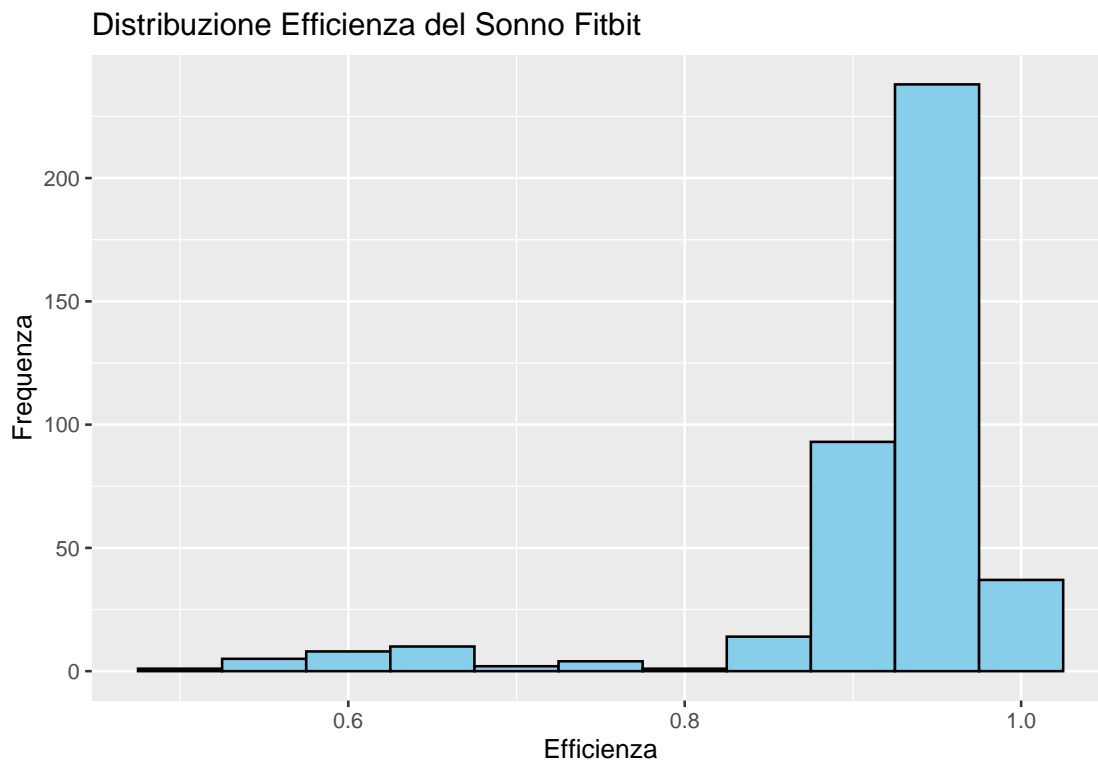
4.4.1 Qualità del sonno - utenti Fitbit

Non essendo presente nel dataset Fitbit una variabile che quantifichi direttamente la qualità del sonno, ho deciso di stimarla utilizzando il parametro *Sleep Efficiency*, definito come il rapporto tra *TotalMinutesAsleep* e *TotalTimeInBed*. Questa scelta si basa sull'assunto che una buona qualità del sonno si ottenga quando la maggior parte del tempo trascorso a letto viene effettivamente dedicato al sonno.

Il seguente istogramma mostra la distribuzione dell'efficienza del sonno per utenti Fitbit, ovvero il numero di volte in cui il parametro *SleepEfficiency* viene calcolato pari ad un dato valore tra 0 e 1.

```
sleep <- sleep %>% mutate(SleepEfficiency = TotalMinutesAsleep / TotalTimeInBed)

ggplot(sleep, aes(x = SleepEfficiency)) +
  geom_histogram(binwidth = 0.05, fill = "skyblue", color = "black") +
  labs(title = "Distribuzione Efficienza del Sonno Fitbit", x = "Efficienza", y = "Frequenza")
```



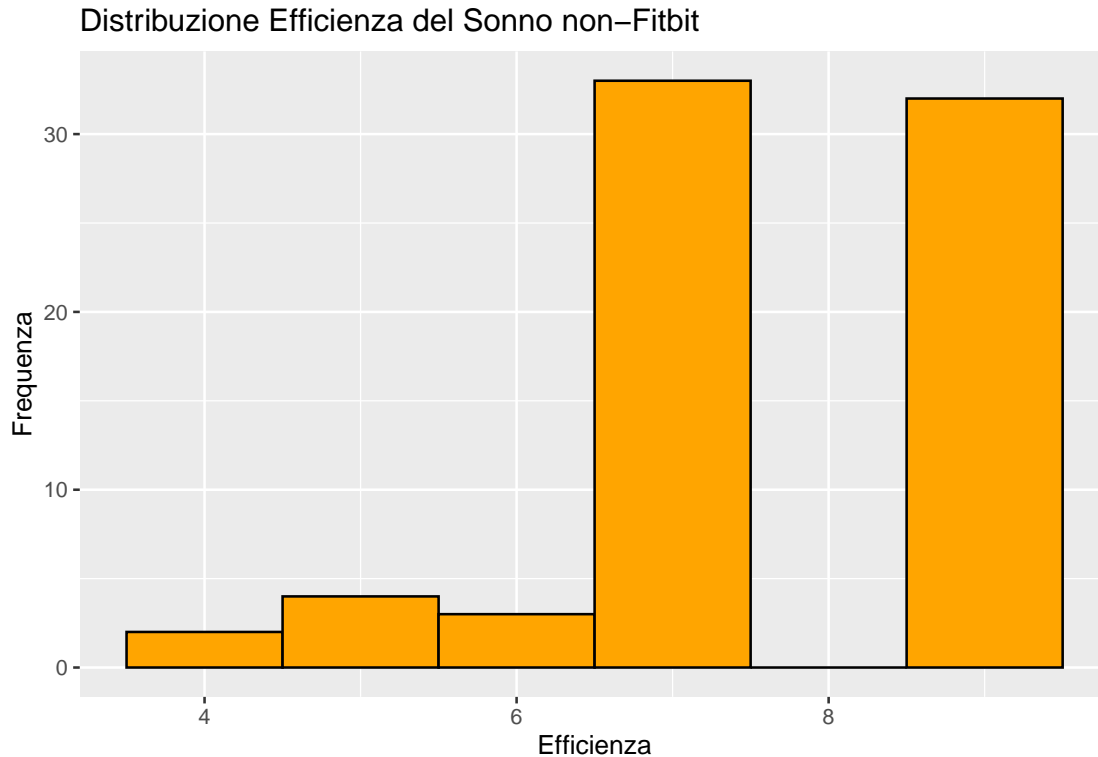
Insight: le utenti Fitbit presentano un'efficienza del sonno mediamente alta (>85%), il che suggerisce che generalmente il tempo trascorso a letto viene sfruttato in modo ottimale per il riposo. Tuttavia, una parte della popolazione mostra una bassa efficienza del sonno, con valori inferiori al 70%, suggerendo possibili casi di insonnia e disturbi del sonno.

Strategia data-driven consigliata: focalizzarsi sulla promozione e la sensibilizzazione di un corretto ciclo del sonno attraverso notifiche personalizzate per utenti con un livello di attività fisica medio-alto.

4.4.2 Qualità del sonno - utenti non-Fitbit

Per le utenti non-Fitbit, l'efficienza del sonno è un parametro già presente nella variabile *Quality.of.Sleep* del rispettivo dataset. L'istogramma mostra che le utenti non-Fitbit presentano un'efficienza del sonno mediamente più bassa (<85%) rispetto ai valori di salute solitamente consigliati.

```
ggplot(sleep_health_clean, aes(x = Quality.of.Sleep)) +
  geom_histogram(binwidth = 1, fill = "orange", color = "black") +
  labs(title = "Distribuzione Efficienza del Sonno non-Fitbit", x = "Efficienza", y = "Frequenza")
```



4.4.3 Confronto della qualità del sonno

Confrontando i risultati appena trovati, si può ottenere il seguente plot comparativo:

```
# Calcolo Sleep Efficiency per ogni utente Fitbit
sleep_efficiency_per_user <- sleep %>%
  group_by(Id) %>%
  summarise(avg_efficiency = mean(SleepEfficiency, na.rm = TRUE))

# Media della qualità del sonno per utenti Fitbit (0-1)
quality_fitbit <- mean(sleep_efficiency_per_user$avg_efficiency, na.rm = TRUE)

# Media della qualità del sonno per utenti non-Fitbit (normalizzata su scala 0-1)
quality_non_fitbit <- mean(sleep_health_clean$Quality.of.Sleep / 10, na.rm = TRUE)
```

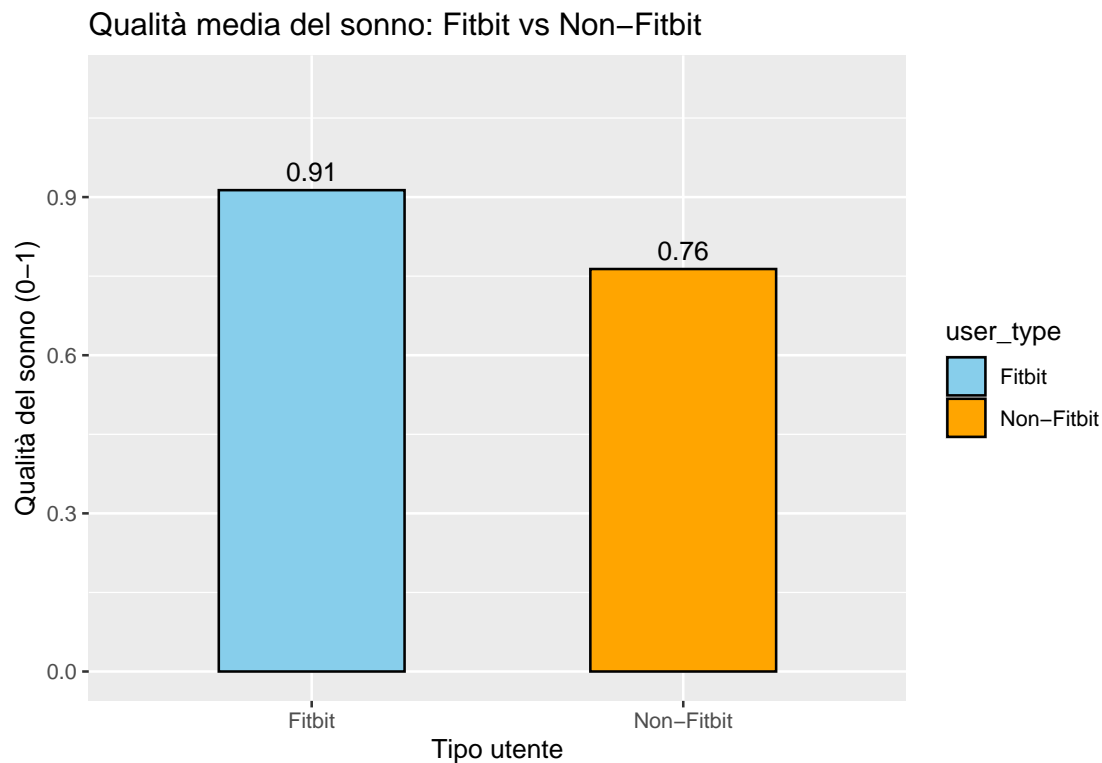


```

# Tabella comparativa
quality_comp <- tibble(
  user_type = c("Fitbit", "Non-Fitbit"),
  QualitySleep = c(quality_fitbit, quality_non_fitbit)
)

# Grafico
ggplot(quality_comp, aes(x = user_type, y = QualitySleep, fill = user_type)) +
  geom_col(width = 0.5, color = "black") +
  labs(title = "Qualità media del sonno: Fitbit vs Non-Fitbit",
       x = "Tipo utente", y = "Qualità del sonno (0-1)") +
  scale_fill_manual(values = c("Fitbit" = "skyblue", "Non-Fitbit" = "orange")) +
  geom_text(aes(label = round(QualitySleep, 2)), vjust = -0.5) +
  ylim(0, max(quality_comp$QualitySleep) + 0.2)

```



Nel confronto, emergono differenze interessanti. Nonostante le utenti **Fitbit** dormano mediamente meno rispetto alle utenti **non-Fitbit** (6.3 ore contro 7.4 ore, con una differenza di circa il **15%**), la qualità del sonno risulta superiore. Infatti per le utenti **Fitbit** risulta pari a 0,91, mentre quella delle utenti **non-Fitbit** si ferma a 0,76. Ciò significa che, pur dormendo meno, le utenti Fitbit sfruttano meglio il tempo a disposizione

a letto, con una differenza del **19,7%** in più nell'efficienza del sonno. Questo potrebbe suggerire che le utenti Fitbit, pur avendo una durata di sonno inferiore, ottimizzano meglio il loro riposo grazie a una qualità superiore, forse supportata dalla funzione di monitoraggio attivo del sonno, che le aiuta a migliorare la gestione del riposo notturno.

4.5 Insight 5: Confronto frequenza cardiaca

Un ulteriore livello di analisi può essere ricavato dai dati relativi alla frequenza cardiaca. La domanda che mi pongo è: *le utenti Fitbit presentano una frequenza cardiaca media inferiore, che potenzialmente riflette un livello di forma più elevato, oppure mostrano valori più elevati, forse a causa di un monitoraggio più frequente durante l'attività fisica?* Questo confronto apre un'interessante possibile linea di ricerca su come l'utilizzo dei dispositivi di monitoraggio accoppiato ai livelli di attività, influenzino le letture biometriche.

```
# Media complessiva della frequenza cardiaca per utenti Fitbit
```

```
heart_rate_avg_fitbit <- heartrate %>%  
  group_by(Id) %>%  
  summarise(AvgHeartRate = mean(Value, na.rm = TRUE)) %>%  
  summarise(overall_avg = mean(AvgHeartRate)) %>%  
  pull(overall_avg)
```

```
# Calcolo della media per gli utenti non-Fitbit
```

```
heart_rate_avg_nonfitbit <- mean(sleep_health$Heart.Rate, na.rm = TRUE)
```

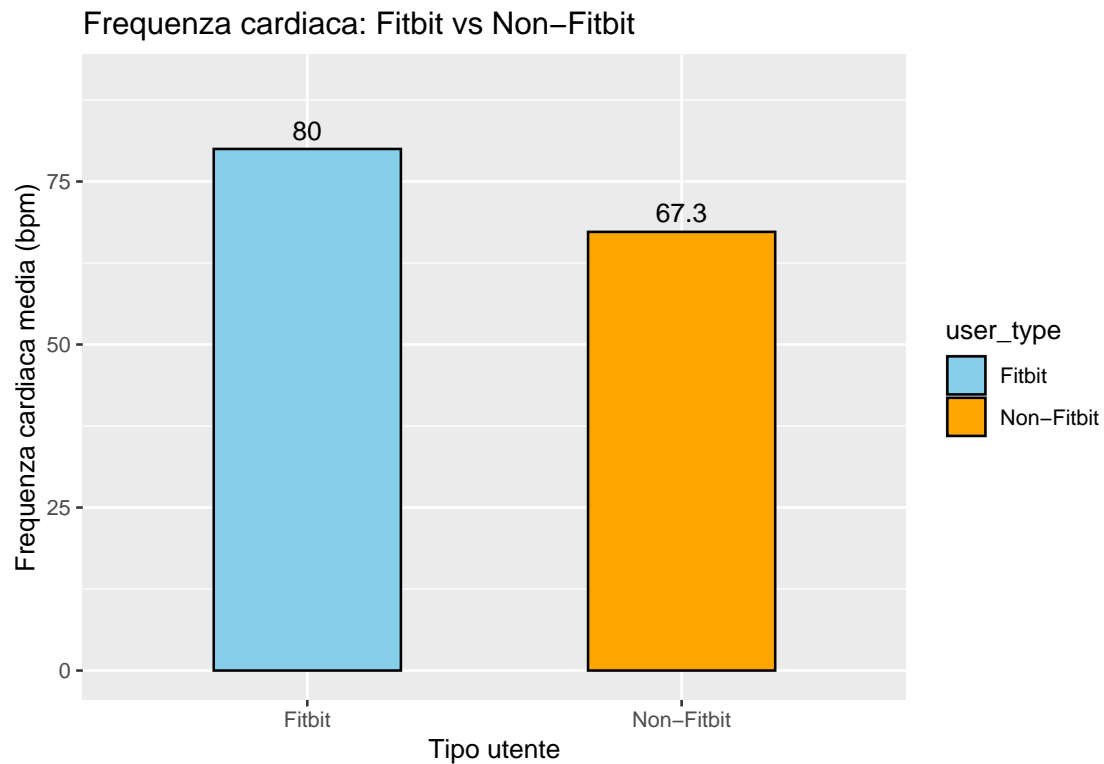
```
# Tabella comparativa
```

```
heart_rate_comp <- tibble(  
  user_type = c("Fitbit", "Non-Fitbit"),  
  heart_rate_c = c(heart_rate_avg_fitbit, heart_rate_avg_nonfitbit)  
)
```

```
# Grafico a barre
```

```
ggplot(heart_rate_comp, aes(x = user_type, y = heart_rate_c, fill = user_type)) +  
  geom_col(width = 0.5, color = "black") +  
  labs(title = "Frequenza cardiaca: Fitbit vs Non-Fitbit",  
        x = "Tipo utente", y = "Frequenza cardiaca media (bpm)") +  
  scale_fill_manual(values = c("Fitbit" = "skyblue", "Non-Fitbit" = "orange")) +
```

```
geom_text(aes(label = round(heart_rate_c, 1)), vjust = -0.5) +
ylim(0, max(heart_rate_comp$heart_rate_c) + 10)
```



Le utenti **Fitbit** presentano una frequenza cardiaca media decisamente più alta (80 bpm) rispetto alle utenti **non-Fitbit** (67,3 bpm). Questo scarto potrebbe essere spiegato da diversi fattori, ad esempio le utenti fitbit:

- sono mediamente più attive, e quindi mostrano una frequenza cardiaca più elevata dovuta a un metabolismo più stimolato
- registrano dati anche durante l'attività fisica, mentre i dati self-reported del secondo gruppo sono riferiti soltanto a stati di riposo.

Questa differenza mette in evidenza una possibile *criticità del database*, che va tenuta in considerazione nel valutare la rappresentatività dei dati. Per una comparazione più coerente in sviluppi futuri, potrebbe essere utile distinguere tra frequenza cardiaca a riposo, durante l'attività fisica e durante il sonno.

5 Act

Nel corso dell'analisi ho confrontato i comportamenti e i parametri di salute tra utenti Fitbit e non-Fitbit, ottenendo insight concreti e direttamente sfruttabili.

Le utenti Fitbit registrano in media un numero decisamente maggiore di passi giornalieri, un'attività fisica più intensa e una durata del sonno leggermente inferiore, delineando un profilo più attivo ma potenzialmente anche più esposto a stress. Anche la frequenza cardiaca media più elevata per queste ultime suggerisce uno stile di vita più dinamico, con picchi fisiologici legati sia all'attività fisica, che a un possibile maggior carico psico-fisico quotidiano. In questo scenario, Bellabeat ha l'opportunità di **differenziarsi** sul mercato con una proposta che va oltre il semplice monitoraggio delle performance fisiche. Una campagna marketing efficace dovrebbe rivolgersi sia alle utenti attive, sia a quelle meno attive.

Per le prime, Bellabeat può sottolineare il valore aggiunto dei propri dispositivi nel monitorare non solo il movimento e le attività fisiche, ma anche la qualità del sonno, i livelli di stress e l'equilibrio psico-fisico quotidiano. Confrontare questi dati con quelli delle utenti attive ma che utilizzano dispositivi non-Bellabeat per il tracciamento, potrebbe evidenziare significative differenze nei livelli di benessere percepito, offrendo spunti per **strategie promozionali** orientate al miglioramento della qualità della vita nel suo complesso e non solo sul piano fisico.

Anche per le utenti meno attive, l'accento dovrebbe essere posto sui benefici meno evidenti nel breve periodo, ma fondamentali per il benessere psico-fisico: miglioramento del sonno, riduzione dello stress, promozione di uno stile di vita più sano e consapevole. Bellabeat può essere il ponte ideale tra salute fisica e benessere quotidiano.



"Non tutte vogliono diventare atlete, ma tutte vogliono uno stile di vita più sano."

Il dispositivo *BellabeatLeaf* si allinea perfettamente con la tipologia di analisi effettuata. Oltre a monitorare attività fisica, sonno, ciclo mestruale e livelli di stress, il Leaf è ideato come un supporto quotidiano al benessere psicofisico, piuttosto che come un semplice fitness tracker. I risultati dell'analisi suggeriscono nuove linee di sviluppo per renderlo ancora più competitivo e coinvolgente, soprattutto per chi cerca equilibrio più che performance.

1. Si rivolge ad un target ben definito e coerente con i dati:

Bellabeat Leaf è uno dei pochi dispositivi progettati specificamente per un pubblico femminile attento non solo all'attività fisica, ma anche al benessere fisico e mentale. Questo si allinea perfettamente con gli insight emersi dalla mia analisi: le utenti con attività fisica moderata o bassa ma con un forte bisogno di equilibrio e benessere, rappresentano un segmento ad *alto potenziale*. Altri dispositivi sul mercato sono più orientati alla performance o alla tecnologia, mentre il Leaf può essere considerato come uno strumento di benessere, più che di prestazione.

2. Approccio centrato sulla salute mentale: A differenza di altri fitness tracker, Bellabeat integra fin dall'inizio funzioni legate a meditazione, ciclo mestruale, stress e mindfulness. Puntare a quella parte di utenza le cui esigenze consistono nel desiderio di ridurre lo stress, dormire meglio, sentirsi in equilibrio, si sposa perfettamente con i bisogni evidenziati dai dati.

3. Design discreto ed elegante: Il Bellabeat Leaf si distingue per la sua estetica, che richiama quella di un gioiello, e che lo rende perfettamente integrabile nell'abbigliamento quotidiano, senza l'aspetto "tech" degli altri fitness tracker. Questa caratteristica favorisce un utilizzo costante e quotidiano anche in contesti lavorativi o formali, migliorando la frequenza e la qualità del monitoraggio. Una maggiore continuità nei dati raccolti permette analisi più affidabili, suggerimenti personalizzati più efficaci e una maggiore aderenza agli obiettivi di benessere.

Alla luce di ciò, suggerisco di rivolgere la campagna pubblicitaria a:

- utenti di altri fitness tracker, evidenziando il valore aggiunto nell'integrare all'attività fisica, anche un piano del sonno personalizzato, distinguendosi nettamente dai competitor - donne attive ma non necessariamente sportive, in cerca di equilibrio tra lavoro, famiglia e salute, che troverebbero nel Leaf il giusto compromesso tra estetica e funzionalità - persone soggette a stress o disturbi del sonno, che trarrebbero beneficio da un monitoraggio attento e da suggerimenti personalizzati del piano del sonno

Temi su cui far leva:

- riduzione dello stress quotidiano
- prevenzione e consapevolezza del proprio corpo
- benessere olistico, inteso come equilibrio tra mente, corpo ed emozioni, senza pressioni da "prestazione"

Funzionalità aggiuntive che consiglio di implementare:

- avvisi intelligenti per suggerire pause rigenerative (stretching, esercizi di respirazione, yoga) quando si



Figure 1: Bellabeat Leaf Urban e Bellabeat Leaf Chakra

rilevano segnali di stress o inattività prolungata

- raccomandazioni smart basate sui dati registrati, come l'orario ideale per andare a dormire in base allo stato fisico e di stress rilevato nei giorni precedenti
- gamification (obiettivi, premi, sfide) e community, per rafforzare l'engagement quotidiano
- partnership con app di meditazione o nutrizionisti digitali, creando un ecosistema wellness integrato a 360° - raccolta dati direttamente da utenti Bellabeat per potenziare le future analisi - distinzione tra sonno leggero/profondo, per migliorare le raccomandazioni sui cicli di sonno
- integrazione di indicatori psicologici dello stato di stress riscontrato

5.1 Conclusione

Il Bellabeat Leaf è già oggi un dispositivo unico nel panorama del wearable tech, ma con alcuni interventi mirati può diventare un **punto di riferimento** per il benessere femminile. In un mercato saturo di dispositivi che semplicemente *tracciano*, Bellabeat ha l'opportunità concreta di emergere come l'**unico** ad offrire un'esperienza di benessere completa, personalizzata e realmente utile nella vita quotidiana.