Strava Fitness Data Analytics: SQL Insights Report

By: Shruti Vinod Nair

This report provides a detailed analysis of user activity, sleep, weight, and heart rate data from the Strava Fitness App. We gathered data from multiple sources at daily, hourly, and minute levels. We conducted a thorough cleaning and validation process, along with exploratory analysis to uncover patterns in user behaviour, activity trends, and health insights.

The analysis includes checking data quality, detecting duplicates and outliers, segmenting users by how often they engage, and examining the relationships between steps, calories, sleep, BMI, and heart rate. We also looked at peak activity hours, differences by day of the week, user retention rates, and long-term trends like changes in weight.

Key findings show that users have high average daily activity and sleep efficiency. There is strong engagement and retention among users, along with clear times of peak physical activity. We also found important insights into how physical activity, sleep, and health metrics relate to each other. These results offer practical recommendations for improving user experience, engagement strategies, and health outcomes on the Strava platform.

DATASETS USED

For this analysis, we used several datasets from the Strava Fitness App. These included daily activity summaries, detailed intensity metrics, sleep logs, weight logs, and heart rate data. This varied data allowed us to explore trends among users as well as specific behavioural patterns.

DATA CLEANING AND VALIDATION

a. **Missing Values:** We checked each table for missing (NULL) values in all key columns. We found no missing values in numeric fields, so we did not need to impute or delete any data. This indicates high data quality, making it reliable for further analysis.

b. **Duplicates:** We looked for duplicate user-date pairs across all main tables. We only found duplicates in the sleep_day table, where three (Id, SleepDay) pairs were repeated. We resolved these duplicates by keeping only the first occurrence of each pair, which ensured accuracy in daily sleep aggregation.
Below is the screenshot attached of the query for resolving duplicate in this case.

```sql
-- 4) sleep_day
-- Duplicates in sleep logs
SELECT "Id", "SleepDay", COUNT(*)
FROM "sleep_day"
GROUP BY "Id", "SleepDay"
HAVING COUNT(*) > 1;
```

Data Output | Messages | Notifications

Showing rows: 1 to 3 | Page No: 1 of 1

	Id bigint	SleepDay timestamp without time zone	count bigint
1	4388161847	2016-05-05 00:00:00	2
2	8378563200	2016-04-25 00:00:00	2
3	4702921684	2016-05-07 00:00:00	2

DATA ANALYSIS AND INSIGHTS

a. **Top 10 Most Active Users:** The ten most active users recorded between 270,249 and 497,241 steps each during the analysis period. The top user logged 497,241 steps and burned 106,028 calories. This highlights a very engaged subgroup within the user base

```sql
--- Top 10 Most Active Users
SELECT "Id", SUM("TotalSteps") AS total_steps, SUM("Calories") AS total_calories
FROM "daily_activity"
GROUP BY "Id"
ORDER BY total_steps DESC
LIMIT 10;
```

Data Output | Messages | Notifications

Showing rows: 1 to 10

	Id bigint	total_steps bigint	total_calories bigint
1	8877689391	497241	106028
2	8053475328	457662	91320
3	1503960366	375619	56309
4	2022484408	352490	77809
5	4388161847	335232	95910
6	3977333714	329537	45410
7	6962181067	303639	61443
8	7007744171	294409	66144
9	7086361926	290525	79557
10	8378563200	270249	106534

b. **Average Daily Metrics:** On average, users walked about 7,638 steps per day, covered 5.49 kilo-meters, and burned around 2,304 calories daily. This suggests a moderately active group of users.

```
441  --- Average Daily Metrics
442  SELECT
443      AVG("TotalSteps") AS avg_steps,
444      AVG("TotalDistance") AS avg_distance,
445      AVG("Calories") AS avg_calories
446  FROM "daily_activity";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	avg_steps numeric	avg_distance double precision	avg_calories numeric
1	7637.9106382978723404	5.489702121915416	2303.6095744680851064

c. **Activity Frequency Segmentation:** We segmented users by the number of activity days. Out of 33 users, most were highly engaged, with 24 classified as "High" frequency (active on 31 days). A few users fell into "Medium" or "Low" engagement segments. This shows strong user retention and frequent app usage.

```
448  --- Activity Frequency Segmentation
449  SELECT "Id",
450         COUNT(DISTINCT "ActivityDate") AS active_days,
451         CASE
452             WHEN COUNT(DISTINCT "ActivityDate") < 10 THEN 'Low'
453             WHEN COUNT(DISTINCT "ActivityDate") < 30 THEN 'Medium'
454             ELSE 'High'
455         END AS frequency_segment
456  FROM "daily_activity"
457  GROUP BY "Id";
```

Data Output | Messages | Notifications

Showing rows: 1 to 33 | Page No: 1 of 1

	Id bigint	active_days bigint	frequency_segment text
1	1503960366	31	High
2	1624580081	31	High
3	1644430081	30	High
4	1844505072	31	High
5	1927972279	31	High
6	2022484408	31	High
7	2026352035	31	High
8	2320127002	31	High
9	2347167796	18	Medium

Total rows: 33 | Query complete 00:00:00.067 | CRLF | Ln 454, Col 23

d. Intensity And Sleep Analysis:

- *Minutes per Intensity Type:* Users spent an average of 21.2 minutes per day in very active physical activity, 13.6 minutes fairly active, 192.8 minutes lightly active, and 991.2 minutes in sedentary behavior.

```
461  -- a) Minutes per intensity type
462  SELECT
463      AVG("VeryActiveMinutes") AS avg_very_active,
464      AVG("FairlyActiveMinutes") AS avg_fairly_active,
465      AVG("LightlyActiveMinutes") AS avg_lightly_active,
466      AVG("SedentaryMinutes") AS avg_sedentary
467  FROM "daily_intensities";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 Page No: 1 of 1

	avg_very_active numeric	avg_fairly_active numeric	avg_lightly_active numeric	avg_sedentary numeric
1	21.1648936170212766	13.5648936170212766	192.8127659574468085	991.2106382978723404

- *Distances per Intensity Type:* Regarding distances, users covered an average of 1.50 km during very active minutes, 0.57 km in moderate activity, 3.34 km during light activity, and they were sedentary for nearly the rest of the day.

```
469  -- b) Distances per intensity type
470  SELECT
471      AVG("VeryActiveDistance") AS avg_very_active_distance,
472      AVG("ModeratelyActiveDistance") AS avg_moderate_distance,
473      AVG("LightActiveDistance") AS avg_light_distance,
474      AVG("SedentaryActiveDistance") AS avg_sedentary_distance
475  FROM "daily_intensities";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 Page No: 1 of 1

	avg_very_active_distance double precision	avg_moderate_distance double precision	avg_light_distance double precision	avg_sedentary_distance double precision
1	1.502680850999945	0.5675425513706943	3.3408191485885292	0.001606382956688709

- *Sleep Summary:* The average user spent 419.2 minutes (about 7 hours) sleeping each day and 458.5 minutes in bed. This results in an overall sleep efficiency of 91.4% (minutes asleep divided by minutes in bed). This indicates a healthy sleep pattern among users.

```sql
-- c) -- Sleep summary
SELECT
    AVG("TotalMinutesAsleep") AS avg_sleep,
    AVG("TotalTimeInBed") AS avg_time_in_bed
FROM "sleep_day";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	avg_sleep numeric	avg_time_in_bed numeric
1	419.1731707317073171	458.4829268292682927

e. **Weight and BMI:** The average user weight was 72.0 kg, with a mean BMI of 25.2, slightly above the "normal" BMI range. This suggests that some users may be overweight.
When we examined weight changes over time, we saw that most users had minor fluctuations, with the largest observed change being 1.8 kg.

```sql
--- Weight and BMI
SELECT
    AVG("WeightKg") AS avg_weight_kg,
    AVG("BMI") AS avg_bmi
FROM "weight_log";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	avg_weight_kg double precision	avg_bmi double precision
1	72.03582137378294	25.185223792915917

```sql
--- Weight Change Over Time (per user)
SELECT "Id", MIN("Date") AS first_entry, MAX("Date") AS last_entry,
       MIN("WeightKg") AS min_weight, MAX("WeightKg") AS max_weight,
       (MAX("WeightKg") - MIN("WeightKg")) AS weight_change
FROM "weight_log"
GROUP BY "Id"
ORDER BY weight_change DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 8 | Page No:

	Id bigint	first_entry timestamp without time zone	last_entry timestamp without time zone	min_weight double precision	max_weight double precision	weight_change double precision
1	8877689391	2016-04-12 06:47:11	2016-05-12 06:42:53	84	85.8000030517578	1.8000030517577983
2	6962181067	2016-04-12 23:59:59	2016-05-12 23:59:59	61	62.5	1.5
3	4558609924	2016-04-18 23:59:59	2016-05-09 23:59:59	69.0999984741211	70.3000030517578	1.2000045776367045
4	2873212765	2016-04-21 23:59:59	2016-05-12 23:59:59	56.7000007629395	57.2999992370605	0.5999984741209943
5	4319703577	2016-04-17 23:59:59	2016-05-04 23:59:59	72.3000030517578	72.4000015258789	0.09999847412110796
6	5577150313	2016-04-17 09:17:55	2016-04-17 09:17:55	90.6999969482422	90.6999969482422	0
7	1503960366	2016-05-02 23:59:59	2016-05-03 23:59:59	52.5999984741211	52.5999984741211	0
8	1927972279	2016-04-13 01:08:52	2016-04-13 01:08:52	133.5	133.5	0

f. **Heart Rate Analysis:** The average heart rate for users was 77.3 bpm. The lowest recorded value was 36 bpm, while the highest was 203 bpm.

```
489  --- Heart Rate
490  SELECT
491      AVG("Value") AS avg_hr,
492      MIN("Value") AS min_hr,
493      MAX("Value") AS max_hr
494  FROM "heartrate_seconds";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	avg_hr numeric	min_hr integer	max_hr integer
1	77.3284236396476488	36	203

Looking at the distribution of heart rate values helps us understand the typical cardiovascular responses among users during activity. By analyzing percentiles (like the 25th, median, and 75th), we can see where most users' heart rates fall and compare these to healthy norms. It is important to flag any unusually low or high readings, which could indicate data recording errors or unusual physiological events.

The analysis shows that users' heart rates are generally within a healthy range during their activities. The median heart rate is 73 bpm, with most users' heart rates between 63 bpm (25th percentile) and 88 bpm (75th percentile). The average heart rate is 77 bpm, with minimum and maximum recorded values of 36 bpm and 203 bpm, respectively. While most values are standard for adults during exercise and rest, a few outliers on the low and high ends likely reflect either short intense efforts or measurement errors. Overall, the heart rate data supports the presence of active and healthy users.

```
598  --- Heart Rate: Distribution and Outliers
599  SELECT
600    PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY "Value") AS hr_25th,
601    PERCENTILE_CONT(0.50) WITHIN GROUP (ORDER BY "Value") AS hr_median,
602    PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY "Value") AS hr_75th
603  FROM "heartrate_seconds";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	hr_25th double precision	hr_median double precision	hr_75th double precision
1	63	73	88

g. **Steps by Month (Seasonality):** Analysis of monthly step counts revealed a noticeable seasonal trend, with total user steps dropping significantly from April (4,777,221 steps) to May (2,406,915 steps)—a reduction of nearly 50%. This suggests a substantial decrease in overall physical activity as the months progress, which may be attributed to seasonal factors, motivation loss, or external events.

```
496  --- Steps by Month (Trend/Seasonality)
497  SELECT DATE_TRUNC('month', "ActivityDate") AS month, SUM("TotalSteps") AS total_steps
498  FROM "daily_activity"
499  GROUP BY month
500  ORDER BY month;
501
```

Data Output | Messages | Notifications

Showing rows: 1 to 2 | Page No: 1 of 1

	month timestamp with time zone	total_steps bigint
1	2016-04-01 00:00:00+05:30	4772721
2	2016-05-01 00:00:00+05:30	2406915

Further, user-level analysis showed that several individuals experienced a month-on-month decline in their step totals, confirming that this drop is not limited to the overall group but affects multiple users individually. These findings highlight the importance of proactive engagement strategies or app interventions during periods where activity naturally declines, to help sustain healthy user behaviour and improve retention.

```
502  --- Users With Declining Monthly Steps
503  WITH monthly_steps AS (
504      SELECT "Id", DATE_TRUNC('month', "ActivityDate") AS month, SUM("TotalSteps") AS steps
505      FROM "daily_activity"
506      GROUP BY "Id", month
507  )
508  SELECT ms1."Id", ms1.month, ms1.steps, ms2.steps AS prev_steps
509  FROM monthly_steps ms1
510  JOIN monthly_steps ms2
511      ON ms1."Id" = ms2."Id" AND ms1.month = ms2.month + interval '1 month'
512  WHERE ms1.steps < ms2.steps
513  ORDER BY ms1."Id", ms1.month;
```

Data Output | Messages | Notifications

Showing rows: 1 to 29 | Page No: 1

	Id bigint	month timestamp with time zone	steps bigint	prev_steps bigint
1	1503960366	2016-05-01 00:00:00+05:30	136812	238807
2	1624580081	2016-05-01 00:00:00+05:30	68001	110060
3	1644430081	2016-05-01 00:00:00+05:30	76434	142055
4	1844505072	2016-05-01 00:00:00+05:30	10993	68989
5	1927972279	2016-05-01 00:00:00+05:30	13207	15193
6	2022484408	2016-05-01 00:00:00+05:30	125556	226934
7	2026352035	2016-05-01 00:00:00+05:30	81679	90894
8	2320127002	2016-05-01 00:00:00+05:30	54072	92151
9	2873212765	2016-05-01 00:00:00+05:30	89106	145123

Total rows: 29 | Query complete 00:00:00.073

h. **Peak Activity Hours (from hourly tables) -** What are the hours when users are most active? Analysis of hourly activity patterns shows that users take the highest average number of steps during the evening hours. The most active hour is **18:00 (6 PM)**, when the average step count peaks. This suggests that users are most likely to engage in walking or exercise routines after work or in the early evening, making this a key window for app engagement and notifications.

```
521  --- Peak Activity Hours (from hourly tables)
522  SELECT EXTRACT(hour FROM "ActivityHour") AS hour, AVG("StepTotal") AS avg_steps
523  FROM "hourly_steps"
524  GROUP BY hour
525  ORDER BY avg_steps DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 24 | Page No: 1 of 1

	hour numeric	avg_steps numeric
1	18	599.1699779249448124
2	19	583.3907284768211921
3	17	550.2328918322295806
4	12	548.6420824295010846
5	14	540.5135722041259501
6	13	537.6981541802388708
7	16	496.8456449834619625
8	10	481.6652314316469322
9	11	456.8867313915857605
10	9	433.3018259935553169
11	8	427.5445757250268528
12	15	406.3191256830601093
13	20	353.9050772626931567

Total rows: 24 | Query complete 00:00:00.061 | CRLF | Ln 532, Col 1

- *Peak Activity Hours – Calories:* The hourly breakdown of calories burned indicates that energy expenditure is also highest in the evening, with the greatest average calories burned at **18:00 (6 PM)**. This alignment with peak step counts further reinforces the evening as the primary period of intense physical activity for most users.

```
527  --- Peak Hours for Calories
528  SELECT EXTRACT(hour FROM "ActivityHour") AS hour, AVG("Calories") AS avg_calorie
529  FROM "hourly_calories"
530  GROUP BY hour
531  ORDER BY avg_calories DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 24 | Page No: 1 of 1

	hour numeric	avg_calories double precision
1	18	123.49227373068433
2	17	122.75275938189846
3	19	121.48454746136865
4	12	117.19739696312364
5	14	115.7328990228013

Total rows: 24 | Query complete 00:00:00.054 | CRLF | Ln 532, Col 1

- *Peak Activity Hours – Intensities:* When examining intensity metrics, users exhibit the highest total intensity values during the **18:00 to 19:00 (6–7 PM)** window. This demonstrates that not only are users more active in terms of steps and calories in the evening, but they are also engaging in higher-intensity activities during these hours.

```sql
--- Peak Hours for Intensities
SELECT EXTRACT(hour FROM "ActivityHour") AS hour, AVG("TotalIntensity") AS avg_i
FROM "hourly_intensities"
GROUP BY hour
ORDER BY avg_intensity DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 24 | Page No: 1 of 1

	hour numeric	avg_intensity numeric
1	18	21.9216335540838852
2	17	21.6556291390728477
3	19	21.3852097130242826
4	12	19.8470715835140998
5	14	18.8686210640608035
6	13	18.7752442996742671
7	16	17.7166482910694598
8	10	17.6437029063509150
9	11	16.9212513484358145
10	15	15.5846994535519126
11	9	15.3877551020408163

Total rows: 24 | Query complete 00:00:00.056 | CRLF | Ln 537, Col 29

In summary, the early evening (especially around 6 PM) consistently emerges as the peak period for user activity—whether measured by steps, calories, or intensity. This time frame offers the greatest opportunity for targeted app interactions, challenges, and motivational prompts to maximize engagement.

i. **Daily Patterns: Steps by Day of Week:** Are people more active on certain days?
Average steps varied by day, with the highest counts on Saturdays (8,152 steps) and Mondays (8,126 steps), and the lowest on Sundays (6,933 steps). This suggests more activity at the start and end of the week.

(0 = Sunday, 1 = Monday, ... 6 = Saturday in PostgreSQL)

```sql
--- Daily Patterns: Steps by Day of Week
SELECT EXTRACT(dow FROM "ActivityDate") AS day_of_week, AVG("TotalSteps") AS avg_steps
FROM "daily_activity"
GROUP BY day_of_week
ORDER BY day_of_week;
```

Data Output | Messages | Notifications

Showing rows: 1 to 7 | Page No: 1 of 1

	day_of_week numeric	avg_steps numeric
1	0	6933.2314049586776860
2	1	7780.8666666666666667
3	2	8125.0065789473684211
4	3	7559.3733333333333333
5	4	7405.8367346938775510
6	5	7448.2301587301587302
7	6	8152.9758064516129032

j. **Sleep Patterns by Day of Week:** Users slept the most on Sundays (average 452.7 minutes), while sleep duration dipped to its lowest on Wednesdays (about 404.5 minutes). This indicates that users tend to recover sleep on weekends.

```
545  --- Sleep Patterns by Day of Week
546  SELECT EXTRACT(dow FROM "SleepDay") AS day_of_week, AVG("TotalMinutesAsleep") AS avg_s
547  FROM "sleep_day"
548  GROUP BY day_of_week
549  ORDER BY day_of_week;
```

Data Output | Messages | Notifications

Showing rows: 1 to 7 | Page No: 1 of 1

	day_of_week numeric	avg_sleep numeric
1	0	452.7454545454545455
2	1	419.5000000000000000
3	2	404.5384615384615385
4	3	434.6818181818181818
5	4	401.2968750000000000
6	5	405.4210526315789474
7	6	419.0701754385964912

k. **Correlation Analysis**

- *Steps and Calories:* A correlation of 0.59 was found, suggesting a moderate-to-strong connection between step count and calories burned.

```
551  --- Correlation Between Steps and Calories
552  SELECT CORR("TotalSteps", "Calories") AS corr_steps_cal
553  FROM "daily_activity";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	corr_steps_cal double precision
1	0.5915680862453353

- *Steps (Activity) and Sleep:* Do users who walk more sleep better? The correlation between steps and total minutes asleep was -0.19, indicating no significant or possibly a slightly negative relationship between daily activity and sleep duration.

```sql
--- Correlation Between Activity and Sleep
SELECT CORR(a."TotalSteps", s."TotalMinutesAsleep") AS corr_steps_sleep
FROM "daily_activity" a
JOIN "sleep_day" s
  ON a."Id" = s."Id" AND a."ActivityDate" = DATE(s."SleepDay");
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	corr_steps_sleep double precision
1	-0.19034391666763165

- *BMI and Activity:* Are heavier users less active? There was a weak negative correlation (-0.16) between BMI and steps, suggesting that heavier users tend to walk less, though the link is not strong.

```sql
--- BMI vs Activity: Are heavier users less active?
SELECT CORR(w."BMI", a."TotalSteps") AS corr_bmi_steps
FROM "weight_log" w
JOIN "daily_activity" a
  ON w."Id" = a."Id" AND w."Date"::date = a."ActivityDate";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	corr_bmi_steps double precision
1	-0.16314953799528079

I. **User Retention Rate:** What % of users are still active in week 2, 3, etc.? The retention analysis showed that all 33 users were present in week 0, with 32 retained in week 1 and 29 active through week 4. This indicates a good retention rate, with only slight user drop-off after the first few weeks.

```sql
--- User Retention Rate
WITH first_date AS (
  SELECT "Id", MIN("ActivityDate") AS first_activity
  FROM "daily_activity"
  GROUP BY "Id"
), activity_weeks AS (
  SELECT d."Id",
         FLOOR(EXTRACT(epoch FROM (DATE_TRUNC('week', d."ActivityDate") - DATE_TRUNC('
  FROM "daily_activity" d
  JOIN first_date f ON d."Id" = f."Id"
)
SELECT week_since_start, COUNT(DISTINCT "Id") AS retained_users
FROM activity_weeks
GROUP BY week_since_start
ORDER BY week_since_start;
```

Data Output | Messages | Notifications

Showing rows: 1 to 5 | Page No: 1 of 1

	week_since_start numeric	retained_users bigint
1	0	33
2	1	32
3	2	32
4	3	29
5	4	27

Total rows: 5 | Query complete 00:00:00.058 | CRLF | Ln 582, Col 1

m. **Sleep Efficiency: Minutes Asleep vs Time in Bed:** The analysis reveals that users maintain a high average sleep efficiency of 91.4%, calculated as the ratio of minutes asleep to total time spent in bed. This indicates that most users are able to fall asleep and remain asleep for the majority of their time in bed, reflecting generally healthy sleep patterns across the user base.

```
583  --- Sleep Efficiency: Minutes Asleep vs Time in Bed
584  SELECT
585    AVG("TotalMinutesAsleep") AS avg_asleep,
586    AVG("TotalTimeInBed") AS avg_inbed,
587    (AVG("TotalMinutesAsleep")::float / NULLIF(AVG("TotalTimeInBed"),0)) AS avg_efficien
588  FROM "sleep_day";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	avg_asleep numeric	avg_inbed numeric	avg_efficiency double precision
1	419.1731707317073171	458.4829268292682927	0.9142612433369862

n. **Active vs. Inactive Users:** Most users in the dataset remained engaged over a long span which is one whole month (31 days), with several users active for 25 or more days in period. The engagement analysis shows that a majority of users are still considered "Active" by the end of the dataset, with only a few dropping off and being labelled as "Inactive." This indicates strong user engagement and a high likelihood of sustained app usage over time.

```
605  --- Active vs. Inactive Users
606  WITH first_last AS (
607    SELECT "Id", MIN("ActivityDate") AS first_date, MAX("ActivityDate") AS last_date,
608           COUNT(DISTINCT "ActivityDate") AS active_days
609    FROM "daily_activity"
610    GROUP BY "Id"
611  )
612  SELECT *,
613    (last_date - first_date) AS engagement_span,
614    CASE WHEN last_date >= (SELECT MAX("ActivityDate") FROM "daily_activity") - INTERVAL
615         ELSE 'Inactive'
616    END AS status
617  FROM first_last
618  ORDER BY engagement_span DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 33 | Page No: 1 of 1

	Id bigint	first_date date	last_date date	active_days bigint	engagement_span integer	status text
25	8792009665	2016-04-12	2016-05-10	29	28	Active
26	6290855005	2016-04-12	2016-05-10	29	28	Active
27	6117666160	2016-04-12	2016-05-09	28	27	Active
28	7007744171	2016-04-12	2016-05-07	26	25	Active
29	6775888955	2016-04-12	2016-05-07	26	25	Active
30	3372868164	2016-04-12	2016-05-01	20	19	Inactive
31	8253242879	2016-04-12	2016-04-30	19	18	Inactive

Total rows: 33 | Query complete 00:00:00.078 | CRLF | Ln 618, Col 31

o. **Days Until Dropoff (Churn Analysis):** When analysing the number of days users stayed active before dropping off, most users maintained high activity for the majority of the observation period. Many users were active for 26– 31 days, suggesting consistent engagement and only gradual attrition rather than abrupt churn. Only a small minority became inactive early (e.g., 4 or 18–20 days).

```
620  --- Days Until Dropoff (Churn Analysis)
621  WITH activity_days AS (
622    SELECT "Id", "ActivityDate",
623           ROW_NUMBER() OVER (PARTITION BY "Id" ORDER BY "ActivityDate") AS day_num
624    FROM "daily_activity"
625  )
626  SELECT "Id", MAX(day_num) AS days_active
627  FROM activity_days
628  GROUP BY "Id"
629  ORDER BY days_active DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 33 | Page No: 1 of 1

	Id bigint	days_active bigint
25	8792009665	29
26	6290855005	29
27	6117666160	28
28	7007744171	26
29	6775888955	26
30	3372868164	20
31	8253242879	19
32	2347167796	18
33	4057192912	4

Total rows: 33 | Query complete 00:00:00.076 | CRLF | Ln 620, Col 1

p. **Percentage of Users Meeting Recommended Activity Levels:** An impressive 75.76% of users consistently met or exceeded the World Health Organization's recommended threshold of 10,000 steps per day. This demonstrates a highly active user population and suggests that Strava's platform is effectively encouraging users to achieve globally recognized physical activity standards.

```
631  --- Percentage of Users Hitting WHO/CDC Recommended Activity (10,000+ steps/day)
632  SELECT 100.0 * COUNT(DISTINCT "Id")
633      FILTER (WHERE "TotalSteps" >= 10000) / COUNT(DISTINCT "Id") AS percent_10k
634  FROM "daily_activity";
```

Data Output | Messages | Notifications

Showing rows: 1 to 1 | Page No: 1 of 1

	percent_10k numeric
1	75.7575757575757576

q. **Average Sleep Duration by Activity Level:** The relationship between activity and sleep duration is somewhat counterintuitive in this dataset. Highly active users averaged the least

sleep (284 minutes per night), while moderately active and low active users slept more, averaging 416 and 379 minutes respectively. This suggests that higher physical activity in this cohort does not directly correlate with longer sleep durations and may indicate that the most active users are trading off some sleep for activity or have different lifestyle patterns.

```sql
--- Average Sleep Duration by Activity Level
WITH user_activity AS (
  SELECT "Id", AVG("TotalSteps") AS avg_steps
  FROM "daily_activity"
  GROUP BY "Id"
), user_sleep AS (
  SELECT "Id", AVG("TotalMinutesAsleep") AS avg_sleep
  FROM "sleep_day"
  GROUP BY "Id"
)
SELECT
  CASE WHEN a.avg_steps >= 10000 THEN 'Highly Active'
       WHEN a.avg_steps >= 5000 THEN 'Moderately Active'
       ELSE 'Low Active'
  END AS activity_group,
  AVG(s.avg_sleep) AS avg_sleep_mins
FROM user_activity a
JOIN user_sleep s ON a."Id" = s."Id"
GROUP BY activity_group;
```

Data Output | Messages | Notifications

Showing rows: 1 to 3 | Page No: 1 of 1

	activity_group text	avg_sleep_mins numeric
1	Highly Active	283.9193540372670808
2	Low Active	378.5552721088435374
3	Moderately Active	415.7691316722902744

Total rows: 3 | Query complete 00:00:00.059 | CRLF | Ln 637, Col 1

CONCLUSION & KEY TAKEAWAYS

This analysis of Strava user data revealed several important insights about user behavior, engagement, and health trends:

- **High Data Quality:** The datasets required minimal cleaning, with almost no missing values or duplicates after initial processing.
- **Strong User Engagement:** Most users were very active, with many logging activity nearly every day during the observation period. User retention remained strong over multiple weeks.
- **Activity Patterns:** Users averaged around 7,600 steps and 2,300 calories burned each day. Most activity and calorie burn peaked in the evening hours, especially around 6 PM. Step counts were highest on Saturdays and Mondays, but declined notably in May, suggesting potential seasonality or a drop in motivation.
- **Health Benchmarks:** Over 75% of users met the daily recommendation of 10,000 steps. Average sleep efficiency was high at 91%, indicating generally healthy habits among the user base. Most users' heart rates stayed within safe ranges. BMI analysis showed a slight tendency toward being overweight.

- **Correlations:** Steps and calories showed a moderate positive correlation. Activity and sleep duration, as well as BMI and steps, had only weak negative associations.
- **Sleep & Activity:** Contrary to expectations, the most active users tended to sleep less. This suggests a need to further investigate user routines and balance.
- **User Retention & Churn:** Most users remained active for the majority of the period, with very few dropping off early. Weight changes over time were minor for most users, indicating stability in health metrics.

Overall, the Strava Fitness App user base in this sample showed high engagement, healthy behaviors, and strong retention. There are clear opportunities for further engagement during periods of declining activity or seasonal drops. These insights can guide future feature development, personalized interventions, and strategies to improve user experience and outcomes.