

Case Study



Google data analytics Case Study-Sheet as a Tool-

In this document, I applied my skills in **data cleaning, analysis, and visualization using Google Sheets**. The objective of this analysis was to clean the Bellabeat user data, calculate key metrics, and visualize insights that can inform business decisions.

Key Skills Demonstrated:

- **Data Cleaning:** Removing duplicates, handling missing data, and ensuring data accuracy.
- **Descriptive Statistics:** Summarizing user activity, sleep quality, and other health metrics.
- **Pivot Tables & Visualization:** Creating pivot tables and charts to visualize data trends.
- **Comparison with External Benchmarks:** Comparing Bellabeat's data with World Health Organization (WHO) guidelines to provide actionable insights for business strategy.

TABLE

of contents

01. Introduction

02. Company overview

03. Scope of the analysis

04. The analysis

05 .Conclusion and Recommendation

Introduction



Company Overview

Bellabeat is an innovative company specializing in smart devices, focusing on wellness and fitness technology. Known for its success in the niche market of health-tracking devices, Bellabeat has established itself as a leader in integrating technology with personal wellness

Current Market Position

While Bellabeat has achieved significant success as a small company, it is well-positioned to expand its influence and capture a larger share of the global smart device market. The company's innovative approach and existing customer base provide a strong foundation for scaling operations

Objective of the Analysis

Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, envisions that a detailed analysis of smart device fitness data can unlock new growth opportunities. This analysis aims to explore fitness data trends, identify potential areas for product improvement, and uncover insights that could drive strategic growth.

The business task

Analyzing data fitness App to unlock new growth opportunities for the company

Scope of the Analysis

1 Data aggregation

2 Data analyzing

3 Identify key trends and relationships

4 Identify opportunities for company growth

My data source

- Google data analytics capstone course



dailyActivity_merged.csv



heartrate_seconds_merged.csv



hourlyCalories_merged.csv



hourlyIntensities_merged.csv



hourlySteps_merged.csv



minuteCaloriesNarrow_merged.csv



minuteIntensitiesNarrow_merged.csv



minuteMETsNarrow_merged.csv



minuteSleep_merged.csv



minuteStepsNarrow_merged.csv



weightLogInfo_merged.csv

My data source

- Google data analytics capstone course

⏪ ⏴ ⏵ ⏩ 🔍 Untitled query × 📊 B > + ▾

📊 Bellaapp 🔍 ▾ + 👤 📄 ⋮ ↻

< SCHEMA DETAILS PREVIEW TABLE EX >

Primary key(s)

Tags

Storage info ⓘ

Number of rows	457
Total logical bytes	53.55 KB
Active logical bytes	53.55 KB
Long term logical bytes	0 B
Total physical bytes	15.84 KB
Active physical bytes	15.84 KB
Long term physical bytes	0 B
Time travel physical bytes	0 B

🔍 *Untitled query × + ▾ ⋮

🔍 Untitled query ⏮ RUN ⋮ ✓ Query completed.

1 SELECT COUNT(*) AS Total_rows

2 FROM `new-project-exercice.Bellasta.Bellaapp`;

Query results ⬇️ SAVE RESULTS ▾ 📊 ▴

< JOB INFORMATION RESULTS CHART >

Row	Total_rows ▾	
1	457	

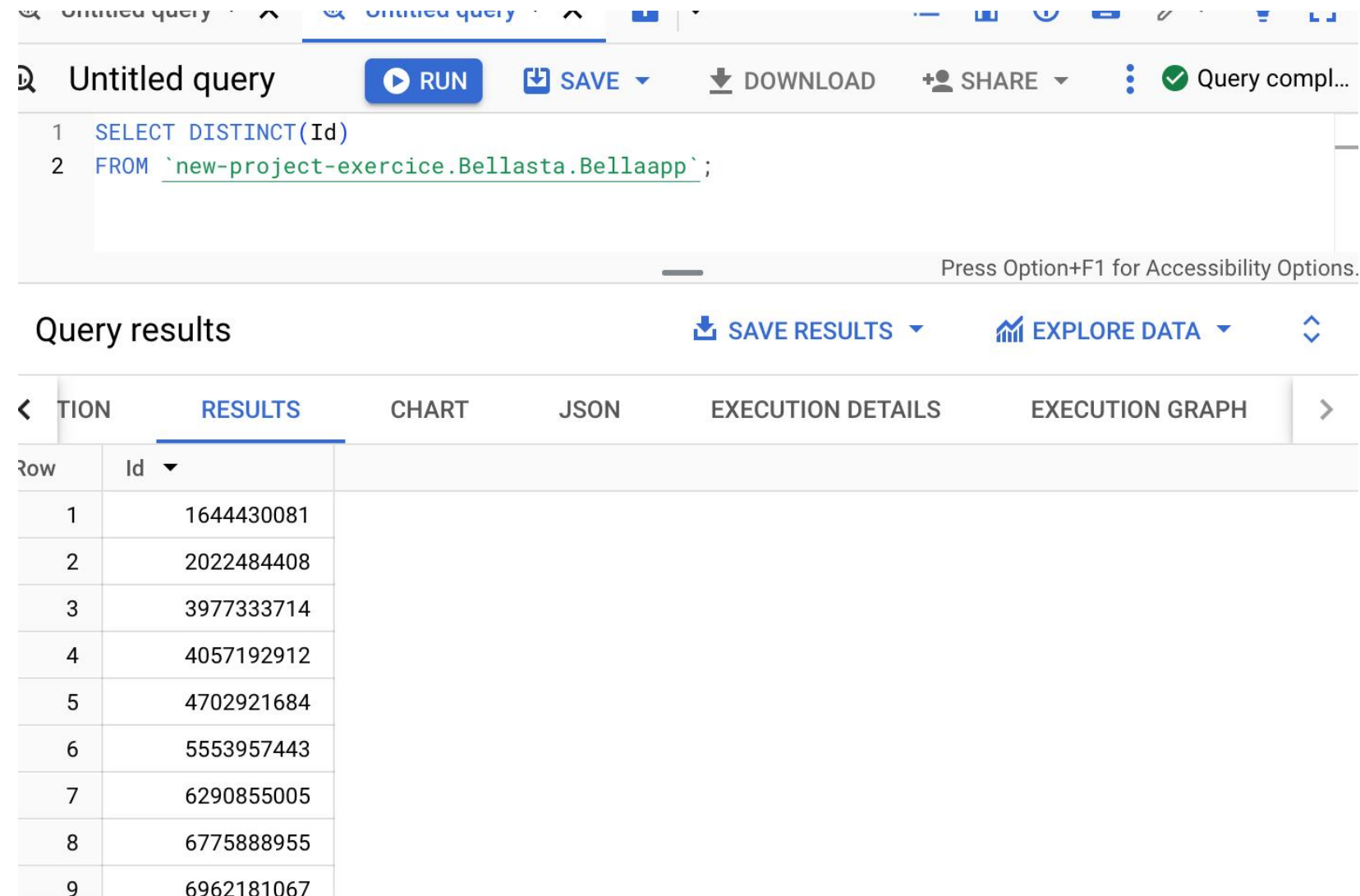
1. Count Rows

To get an overview of how large your dataset is, you can count the number of rows in your table:

Purpose: Helps you understand the volume of data you're dealing with and can also be used to verify if the data import was successful.

My data source

- Google data analytics capstone course



The screenshot shows the Google Data Studio interface. At the top, there's a toolbar with buttons for 'RUN', 'SAVE', 'DOWNLOAD', 'SHARE', and a status indicator 'Query compl...'. Below the toolbar is a text area containing a SQL query:

```
1 SELECT DISTINCT(Id)
2 FROM `new-project-exercice.Bellasta.Bellaapp`;
```

Below the query editor, there's a section titled 'Query results' with buttons for 'SAVE RESULTS' and 'EXPLORE DATA'. Underneath, there's a tabbed interface with tabs for 'VIEW', 'RESULTS', 'CHART', 'JSON', 'EXECUTION DETAILS', and 'EXECUTION GRAPH'. The 'RESULTS' tab is selected, displaying a table with 9 rows and 2 columns: 'Row' and 'Id'.

Row	Id
1	1644430081
2	2022484408
3	3977333714
4	4057192912
5	4702921684
6	5553957443
7	6290855005
8	6775888955
9	6962181067

1. Count Rows

To get an overview of how large your dataset is, you can count the number of rows in your table:

Purpose: Helps you understand the volume of data you're dealing with and can also be used to verify if the data import was successful.

My data source

- Google data analytics capstone course

Untitled query RUN Query completed.

```
1 SELECT
2 MAX(TotalSteps) AS max_steps,
3 MIN(TotalSteps) AS min_Steps,
4 AVG(TotalSteps) AS avg_steps
5 FROM `new-project-exercice.Bellasta.Bellaapp`;
```

Press Option+F1 for Accessibility Options.

Query results SAVE RESULTS

	JOB INFORMATION	RESULTS	CHART	
Row	max_steps	min_Steps	avg_steps	
1	28497	0	6546.562363238...	

Calculate Maximum, Minimum, and Mean Values

For numeric columns, calculate statistical metrics to understand the range and central tendency:

Purpose: Provides insights into the distribution of numeric data, which is useful for detecting outliers or understanding the spread of the data.



Data cleaning

- **Phase 1: Remove Duplicate Values**

I first removed duplicate values from the dataset to ensure each record is unique

- **Phase 2: Merge Data from Different Sheets**

I then merged data from different sheets into a single sheet using the QUERY function in Google Sheets. The ID was used as the reference or shared value to combine the data accurately

- **Phase 3: Handle Missing Sleep Data**

Since the analysis includes sleep data, I identified and removed rows lacking sleep information to ensure the dataset was complete and relevant for analysis.

Data is clean!

5/6/2016 0:00:00									
A		B		C		D		E	
Id		ActivityDate		TotalSteps		TotalDistance		FairlyActiveMinute	
1	1503960366	4/12/2016	13162	8.5	25	13	19	328	728
2	1503960366	4/13/2016	10735	6.96999979	21	11	34	217	776
3	1503960366	4/14/2016	10460	6.739999771	30	20	209	181	726
4	1503960366	4/15/2016	9762	6.28000021	29	16	221	773	1863
5	1503960366	4/16/2016	12669	8.159999847	36	31	164	539	1728
6	1503960366	4/17/2016	9705	6.480000019	42	20	233	1149	1776
7	1503960366	4/18/2016	13019	8.590000153	50	12	205	818	1745
8	1503960366	4/19/2016	15506	6.679999828	28	8	211	775	1863
9	1503960366	4/20/2016	10544	6.340000153	19	27	130	1921	320
10	1503960366	4/21/2016	9819	8.130000114	66	5	262	1786	377
11	1503960366	4/22/2016	12764	9.039999962	41	14	238	1827	364
12	1503960366	4/23/2016	14371	6.409999847	39	23	216	1949	384
13	1503960366	4/24/2016	10039	8.789999962	73	11	279	2013	449
14	1503960366	4/25/2016	15355	12.21000004	31	28	814	1775	323
15	1503960366	4/26/2016	13755	8.529999733	48	12	833	1877	354
16	1503960366	4/27/2016	18134	7.150000095	16	34	1108	1970	425
17	1503960366	4/28/2016	13154	9.25	52		782	2159	396
18	1503960366	4/29/2016	11181				1898	1970	309
19	1503960366	4/30/2016	14673				1837	2013	296
20	1503960366						1947	2159	264

5/6/2016 0:00:00									
A		B		C		D		E	
Id		ActivityDate		TotalSteps		TotalDistance		FairlyActiveMinute	
392	4319703577	4/21/2016	4500	2.480000019	1	0	19	34	176
393	4319703577	4/22/2016	4935	3.019999981	0	5	0	233	191
394	4319703577	4/23/2016	9259	3.309999943	1	8	19	176	1265
395	4319703577	4/24/2016	10780	6.210000038	0	6	0	546	709
396	4319703577	4/25/2016	10817	6.639999866	8	21	8	692	1886
397	4319703577	4/26/2016	7990	7.230000019	0	47	5	544	1945
398	4319703577	4/27/2016	8221	5.360000134	6	18	8	649	1880
399	4319703577	4/28/2016	9648	5.519999981	6	38	0	680	2314
400	4319703577	4/29/2016	10429	6.239999771	0	0	13	552	2236
401	4319703577	5/1/2016	13658	6.469999738	8	8	288	649	2324
402	4319703577	5/2/2016	9524	7.019999981	27	6	300	680	2367
403	4319703577	5/3/2016	7937	9.489999771	6	0	289	552	2175
	4319703577	5/4/2016	3672	6.420000076	3	0	359	503	546
	4319703577	5/5/2016	10378	5.329999924	0	11	390	493	391
	4319703577	5/6/2016		2.460000038	2	14	709	552	387
	4319703577			6.960000038	3	12	546	503	391
	4319703577			6.369999886			692	546	391

5/8/2016 12:00:00 AM									
A		B		C		D		E	
Id		ActivityDate		TotalSteps		TotalDistance		FairlyActiveMinute	
506	4319703577	5/8/2016 12:00:00 AM	4500	2.480000019	1	0	19	34	176
507	4319703577	5/9/2016 12:00:00 AM	4935	3.019999981	0	5	0	233	191
508	4319703577	5/10/2016 12:00:00 AM	9259	3.309999943	1	8	19	176	1265
509	4319703577	5/11/2016 12:00:00 AM	10780	6.210000038	0	6	0	546	709
510	4319703577	5/12/2016 12:00:00 AM	10817	6.639999866	8	21	8	692	1886
511	4319703577	5/13/2016 12:00:00 AM	7990	7.230000019	6	18	8	544	1945
512	4319703577	5/14/2016 12:00:00 AM	8221	5.360000134	6	38	0	680	1880
513	4319703577	5/15/2016 12:00:00 AM	9648	5.519999981	0	0	13	552	2236
514	4319703577	5/16/2016 12:00:00 AM	10429	6.239999771	8	8	288	649	2367
515	4319703577	5/17/2016 12:00:00 AM	13658	6.469999738	6	0	359	552	2175
516	4319703577	5/18/2016 12:00:00 AM	9524	7.019999981	27	6	390	503	546
517	4319703577	5/19/2016 12:00:00 AM	7937	9.489999771	3	14	709	552	391
518	4319703577	5/20/2016 12:00:00 AM	10378	6.420000076	0	12	546	503	391
519	4319703577	5/21/2016 12:00:00 AM		2.460000038	2		692	546	391
520	4319703577	5/22/2016 12:00:00 AM		6.960000038	3		709	552	391
521	4319703577	5/23/2016 12:00:00 AM		6.369999886			692	546	391

Data Analyze

- **Phase 1: Initial Data Summary**

I began by calculating a summary of the data. However, the initial summary was not particularly meaningful for the analysis.

- **Phase 2: Advanced Data Analysis with Pivot Tables**

I used pivot tables to better organize and summarize the data by ID.

The pivot table allowed me to structure the data effectively and provided a clearer view for further analysis.

From the pivot table, I calculated new, more meaningful metrics:

Converted Sleep Data: Changed sleep data from minutes to hours per day.

• **Adjusted Activity Data:** Converted active and sedentary hours from minutes to hours to make the data more interpretable

- **Phase 3: Key Metrics and Data Visualization**

Calculate Key Metrics

- Computed averages for active hours, sedentary hours, and sleep hours.
- Calculated the active-to-sedentary ratio.

Visualize Data

- Created charts (e.g., bar and line graphs) to show trends.
- Used scatter plots to explore relationships.

My New Data!

	A	B	C	D	E	F	G
1	ID	SleepHoursDay	Steps per day	Active hoursDay	SedentryhoursDay	CaloriesDay	Active/SedentaryRatio
2	1503960366	6.927956989	12116.74194	20	14.13602151	1816.419355	1.414825239
3	1624580081	9.024731183	5743.903226	4.483333333	20.96236559	1483.354839	0.2138753527
4	1644430081	7.789444444	7282.966667	4.783333333	19.36444444	2811.3	0.2470162956
5	1844505072	7.498924731	2580.064516	0.06666666667	20.11021505	1573.483871	0.003315064831
6	1927972279	7.732258065	916.1290323	0.6833333333	21.95698925	2172.806452	0.03112144956
7	2022484408	6.890860215	11370.64516	18.75	18.54301075	2509.967742	1.011162656
8	2026352035	6.560215054	5566.870968	0.05	11.49032258	1540.645161	0.004351487928
9	2320127002	8.244086022	4716.870968	0.7	20.33494624	1724.16129	0.03442349893
10	2347167796	7.986111111	9519.666667	4.05	11.45277778	2043.444444	0.3536260005
11	2873212765	7.829032258	7555.774194	7.283333333	18.28655914	1916.967742	0.3982888895
12	3372868164	7.921666667	6861.65	3.05	17.95916667	1933.1	0.1698297063
13	3977333714	7.410555556	10984.56667	9.45	11.79222222	1513.666667	0.8013756713
14	4020332650	7.559139785	2267.225806	2.683333333	20.62096774	2385.806452	0.1301264503
15	4057192912	8.925	3838	0.05	20.2875	1973.75	0.002464571781
16	4319703577	7.682758621	7454.758621	1.8	12.77931034	2092.37931	0.1408526713

Data Analyzing

- **Phase 1:** Data aggregation and key

summary

I first removed duplicate values from the dataset to ensure each record is unique

- **Phase 2:** Merge Data from Different Sheets

I then merged data from different sheets into a single sheet using the QUERY function in Google Sheets. The ID was used as the reference or shared value to combine the data accurately

- **Phase 3:** Handle Missing Sleep Data

Since the analysis includes sleep data, I identified and removed rows lacking sleep information to ensure the dataset was complete and relevant for analysis.

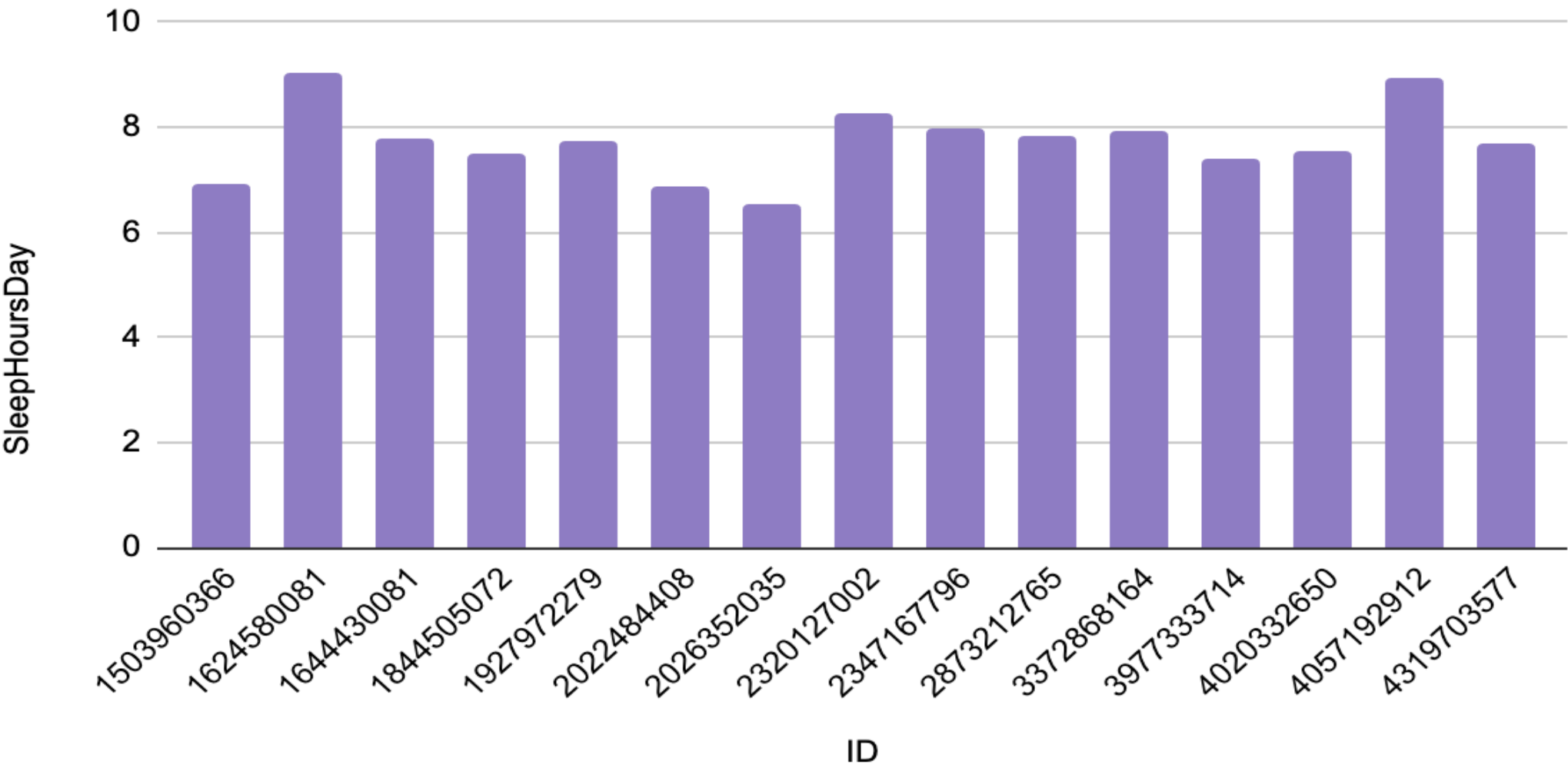
Sleeping Hours per day

Table: Key metrics indicators

Average	7.73
Min	6.56
Max	9.02

Source: Author calculation

SleepHoursDay vs. ID



- The data generally shows that the sleeping hours fall within the recommended range set by the American Association for adults, which is between 7 to 8 hours.

Sleeping Hours per day

- The maximum recorded value is 9 hours, slightly higher than the recommended range.
- The data also indicates that all tracked individuals went to sleep after 12 AM, which could negatively affect sleep quality.
- The company can leverage this data to enhance its product offerings and increase market share by addressing sleep quality and providing tailored solutions

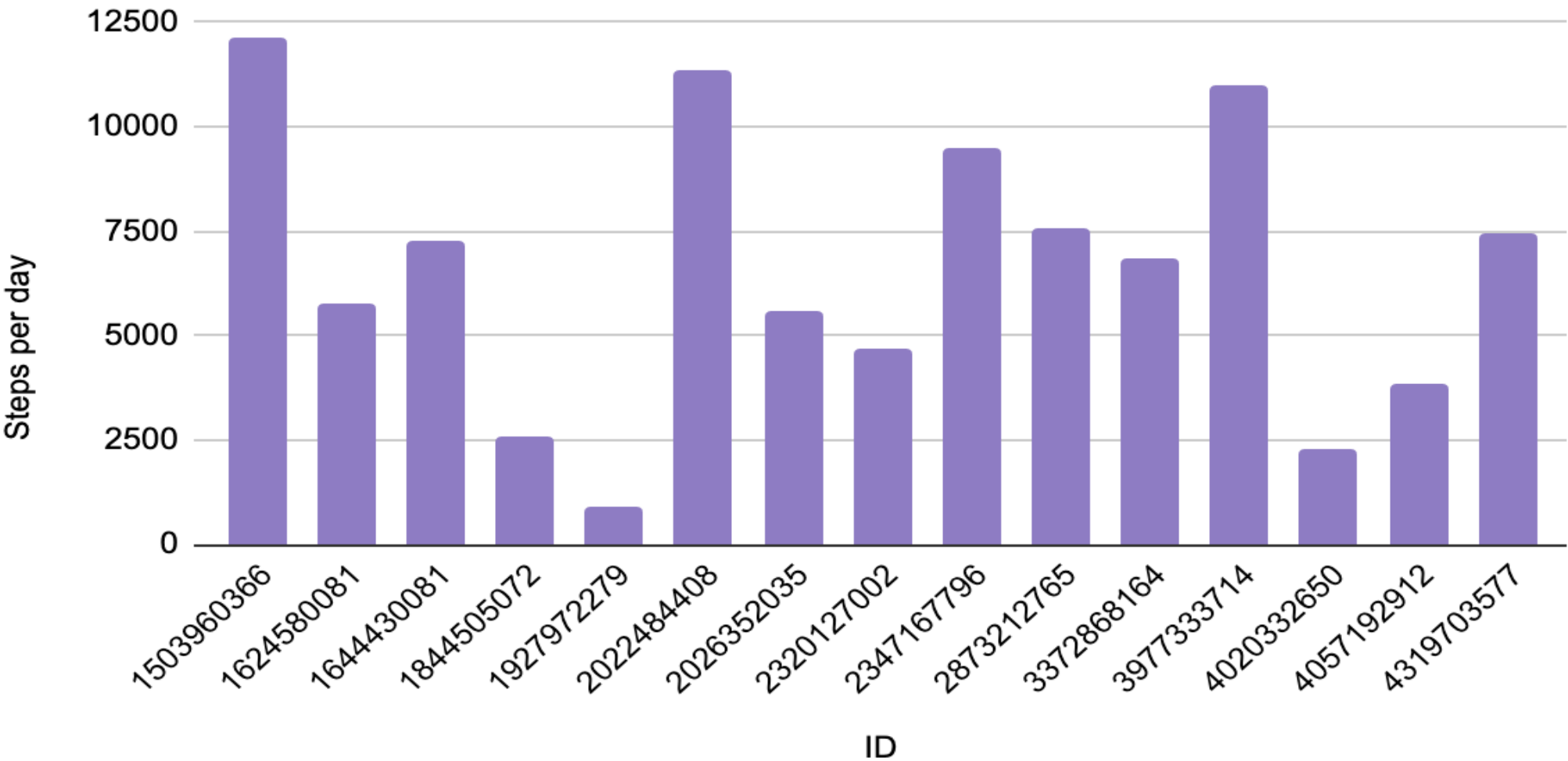
Steeps per day

Table: Key metrics indicators

Average	6585.05
Min	916.12
Max	12116.74

Source: Author calculation

Steps per day vs. ID



-The data generally shows that the average number of steps per day is below the recommended 10,000 steps set by the World Health Organization.

Steeps per day

- The minimum number of steps recorded is 916, which is far below the WHO recommendation. The maximum number of steps recorded is 12,116, which is slightly above the recommended amount.
- The data indicates that there is an opportunity for the company to compete in the area of daily step counts.
- The company can leverage this data in its marketing strategy to highlight the importance of daily steps for consumer health. Additionally, the company could propose options that notify clients when their step counts are below, at, or above the norm.

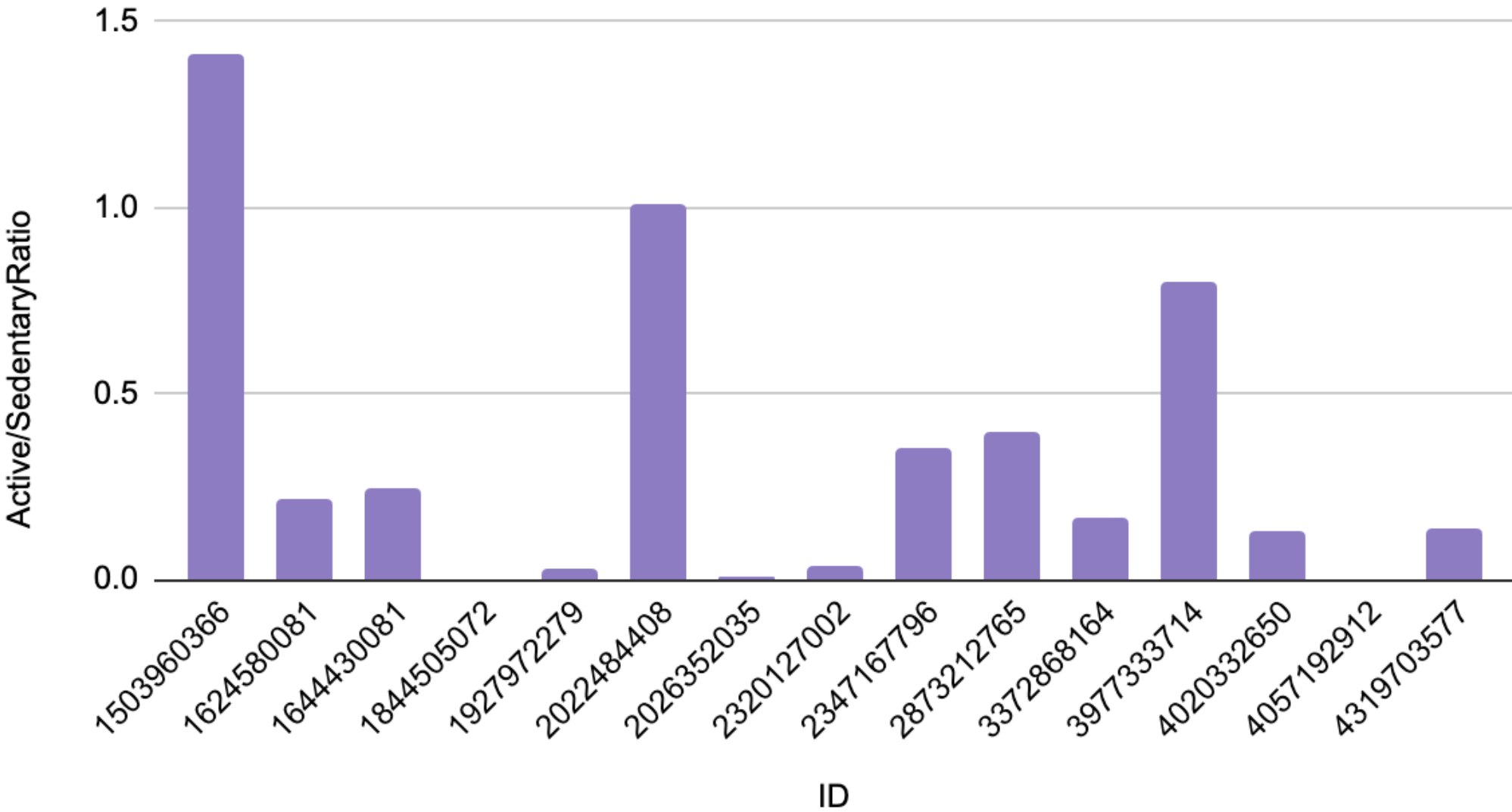
Active/sedentary Ratio

Table: Key metrics indicators

Average	33%
Min	0.2%
Max	141%

Source: Author calculation

Active/SedentaryRatio vs. ID



The data shows that the average active/sedentary ratio is 33%, meaning that 33% of the time people are active, and 67% of the time they are sedentary.

Steeps per day

- This indicates that the population is largely sedentary. The World Health Organization (WHO) notes that sedentary behaviour is strongly associated with the development of cardiovascular disease, which is the leading cause of global mortality and disability.
- The marketing team should focus on this information to promote products that track and monitor activity and send notifications to consumers when they are inactive.

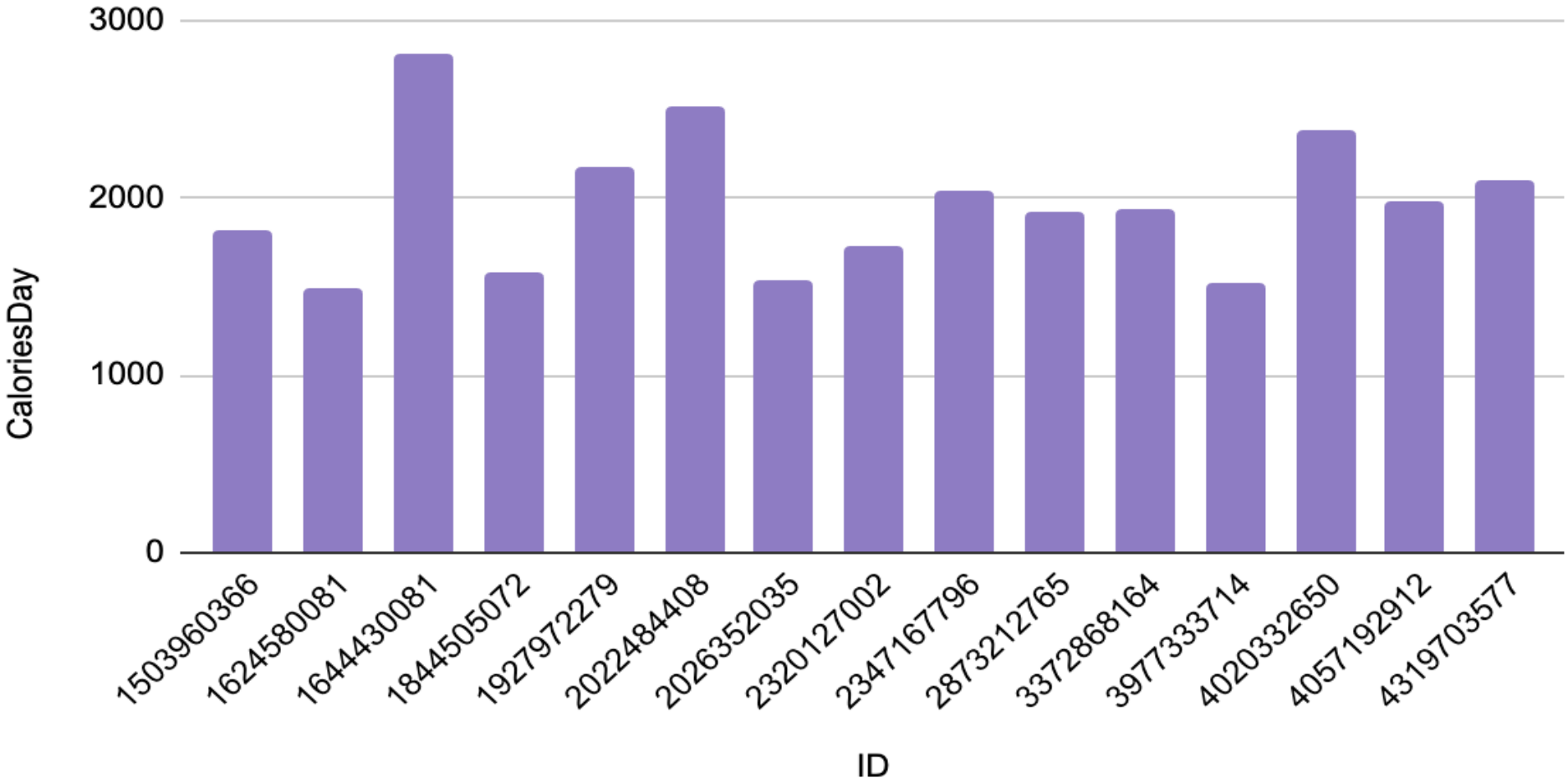
Calories per day

Table: Key metrics indicators

Average	1966.083555
Min	1483.354839
Max	2811.3
Standard deviatin	388.0687567

Source: Author calculation

CaloriesDay vs. ID

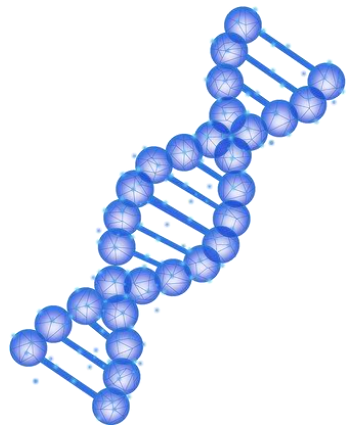


The data shows that the average number of calories consumed by the tracked population is 1966, which is lower than the recommended calories by WHO, which is between 2000 and 2500 daily. This indicates that the population cares about their weight and overall health.

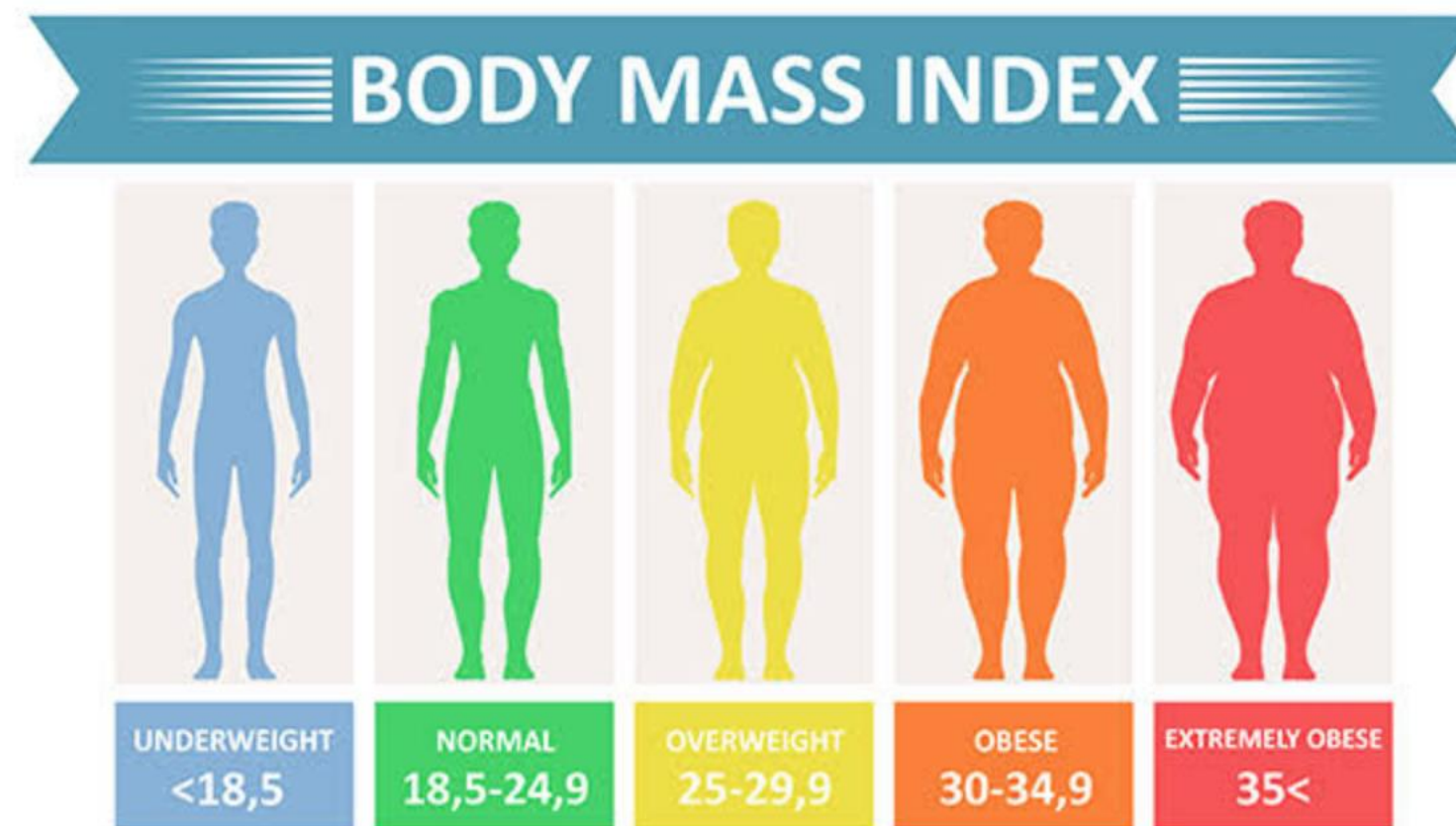
Calories per day

- The minimum value is 1483, which is below the calorie requirement; this could indicate a goal driven by individuals who are trying to lose weight. The maximum value is 2811.3, which is above the norm.
- The marketing team should use this data to target individuals who are interested in health by selling more accessories and introducing features for calorie limits for those who want to lose, maintain, or even gain weight.

Something is missing!!!



We have steps data, sleeping data, calories data!
We need the weight !BMI!



My data before!

The magic of Pivot
Tables!

J10 | fx

	A	B	C	D	E	F	G	H	
1	Id	Date	WeightKg	WeightPounds	Fat	BMI	IsManualReport	LogId	
2	1503960366	5/2/2016 11:59:5	52.59999847	115.9631465	22	22.64999962	TRUE	1462233599000	
3	1503960366	5/3/2016 11:59:5	52.59999847	115.9631465		22.64999962	TRUE	1462319999000	
4	1927972279	4/13/2016 1:08:5	133.5	294.31712		47.54000092	FALSE	1460509732000	
5	2873212765	4/21/2016 11:59	56.70000076	125.0021043		21.45000076	TRUE	1461283199000	
6	2873212765	5/12/2016 11:59	57.29999924	126.3248746		21.69000053	TRUE	1463097599000	
7	4319703577	4/17/2016 11:59	72.40000153	159.6146812	25	27.45000076	TRUE	1460937599000	
8	4319703577	5/4/2016 11:59:5	72.30000305	159.3942223		27.37999916	TRUE	1462406399000	
9	4558609924	4/18/2016 11:59	69.69999695	153.66219		27.25	TRUE	1461023999000	
10	4558609924	4/25/2016 11:59	70.30000305	154.984977		27.45999908	TRUE	1461628799000	
11	4558609924	5/1/2016 11:59:5	69.90000153	154.1031246		27.31999969	TRUE	1462147199000	
12	4558609924	5/2/2016 11:59:5	69.19999695	152.5598787		27.04000092	TRUE	1462233599000	
13	4558609924	5/9/2016 11:59:5	69.09999847	152.3394198		27	TRUE	1462838399000	
14	5577150313	4/17/2016 9:17:5	90.69999695	199.9592651		28	FALSE	1460884675000	
15	6962181067	4/12/2016 11:59	62.5	137.7889139		24.38999939	TRUE	1460505599000	
16	6962181067	4/13/2016 11:59	62.09999847	136.9070615		24.23999977	TRUE	1460591999000	
17	6962181067	4/14/2016 11:59	61.70000076	136.0252175		24.10000038	TRUE	1460678399000	
18	6962181067	4/15/2016 11:59	61.5	135.5842912		24	TRUE	1460764799000	
19	6962181067	4/16/2016 11:59	62	136.6866026		24.20999908	TRUE	1460851199000	
20	6962181067	4/17/2016 11:59	61.40000153	135.3638323		23.95999908	TRUE	1460937599000	

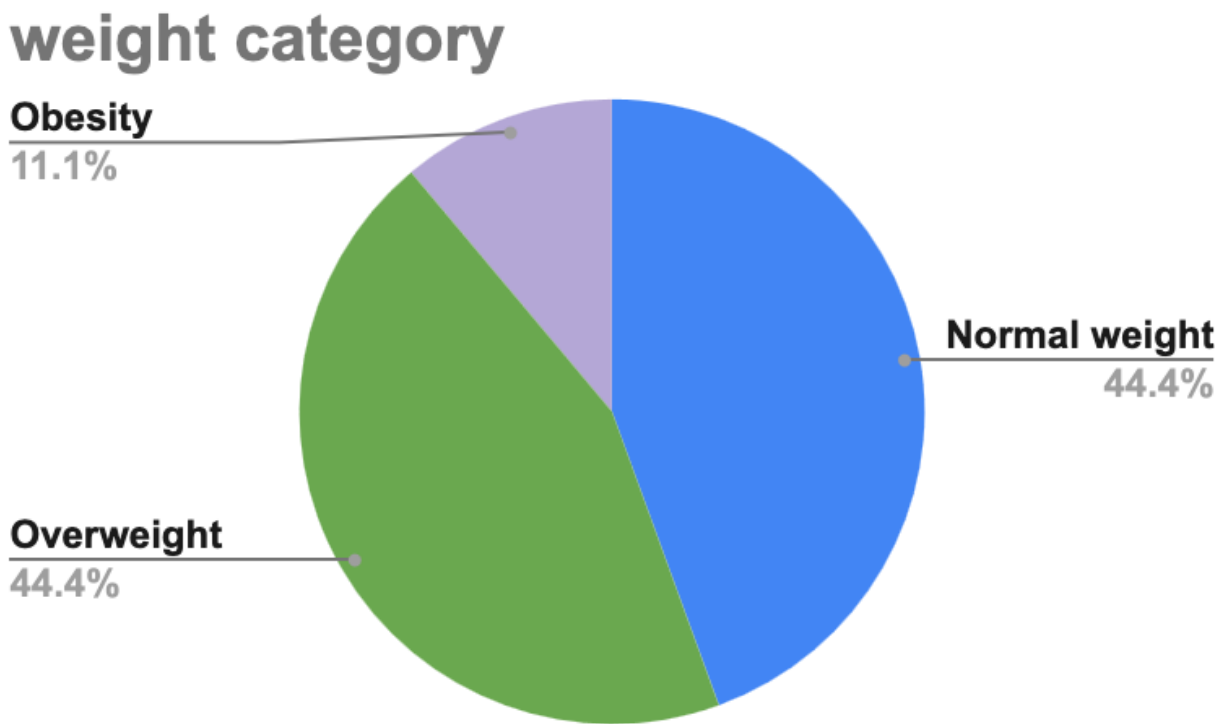
A	B	C	D
<i>Id</i>	AVERAGE of WeightKg	AVERAGE of WeightPounds	AVERAGE of BMI
1503960366	52.59999847	115.9631465	22.64999962
1927972279	133.5	294.31712	47.54000092
2873212765	57	125.6634894	21.57000065
4319703577	72.35000229	159.5044517	27.41499996
4558609924	69.63999939	153.529918	27.21399994
5577150313	90.69999695	199.9592651	28
6962181067	61.55333379	135.7018721	24.02799975
8877689391	85.14583429	187.7144324	25.48708336
Grand Total	72.03582137	158.8118014	25.18522379

Calories per day

Table: Key metrics indicators

Average	27.98801052
Min	21.57000065
Max	47.54000092
Standard deviation	8.236065624

Source: Author calculation



The analysis of Body Mass Index (BMI), which estimates a person’s health status based on height and weight, indicates that 44.4% of the population is overweight, 11.1% are obese, and only 44.4% of the population are normal weight. The mean BMI of 28 confirms this situation. The data confirm that a significant percentage of the population is overweight. The marketing team can use this data to improve the health of new consumers by selling tools that track calories, BMI, and other indicators that could help.

Conclusions

Key finding

- The sleeping hours fall within the recommended range set by the American Association for adults, which is between 7 to 8
- The average number of steps per day is below the recommended 10,000 steps set by the World Health Organization.
- The active-to-sedentary ratio indicate that people are sedentary for 67% of there Time.
- the average number of calories consumed by the tracked population is 1966, which is lower than the recommended calories by WHO
- The analysis of Body Mass Index (BMI), indicates that 44.4% of the population is overweight, 11.1% are obese, the data confirm that a significant percentage of the population is overweight.

Recommendations

Feature Enhancement: Develop app features to increase physical activity and reduce sedentary behavior.

Holistic Wellness: Integrate nutrition and weight management tracking with personalized recommendations.

Advanced Analytics: Offer detailed health reports and insights based on activity, caloric intake, and BMI data.

Support Services: Provide in-app health coaching and community support to address obesity and wellness needs.