

# Case Study: SQL as a Tool



# Case Study: SQL as a Tool

In this document, I utilized **SQL (BigQuery)** to demonstrate my skills in **data manipulation, querying, and statistical analysis**. By importing and analyzing Bellabeat's dataset, I extracted important metrics and performed correlation analysis to uncover meaningful patterns in user behavior.

## Key Skills Demonstrated:

- **Data Import & Management:** Using BigQuery to import large datasets and organize them for analysis.
- **SQL Queries:** Writing SQL queries to merge tables, calculate summary statistics, and identify correlations.
- **Statistical Analysis:** Applying SQL commands to calculate relationships between variables like sleep and activity levels.
- **Data Insight Extraction:** Identifying patterns in the data to offer strategic recommendations for new market opportunities.

# TABLE

## of contents

**01. Introduction**

**02. Company overview**

**03. Scope of the analysis**

**04. The analysis**

# 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

- Google data analytics capstone by Dr.Kabba

explorer

Search BigQuery resources

Showing resources.

SHOW STARRED ONLY

Data canvases

Data preparations

External connections

Bellasta

☆

BMR

☆

Bellaapp

☆

Cars

★

Casestudy

☆

Sales

☆

avocado\_data

☆

demos

☆

employee\_data

☆

Untitled query

BMR

QUERY

SHARE

COPY

SNAPSHOT

<

SCHEMA

DETAILS

PREVIEW

TABLE EXPLORER

PREVIEW

INSIGHTS

>

Row	Id	Date	WeightKg	WeightPounds	BMI	
1	4558609924	04/08/2016	69.40000153	153.0008133	27.13999939	
2	2891001357	04/05/2016	88.40000153	194.8886431	25.03000069	
3	1503960366	04/05/2016	53.29999924	117.5063841	22.96999931	
4	8877689391	04/08/2016	85.0	187.3929229	25.44000053	
5	8877689391	04/01/2016	85.5	188.4952342	25.61000061	
6	8877689391	04/05/2016	86.0	189.5975455	25.76000023	
7	8877689391	04/07/2016	85.09999847	187.6133818	25.48999977	
8	8877689391	04/11/2016	86.09999847	189.8180044	25.79000092	
9	8877689391	04/04/2016	86.59999847	190.9203157	25.94000053	
10	8877689391	04/09/2016	85.40000153	188.2747753	25.55999947	
11	8877689391	04/12/2016	85.80000305	189.1566277	25.68000031	
12	8877689391	04/06/2016	86.30000305	190.258939	25.82999992	
13	2347167796	04/03/2016	63.40000153	139.7730776	24.77000046	
14	2873212765	04/06/2016	56.70000076	125.0021043	21.45000076	

Step 1.Import my data to SQL

Two table was imported in SQL

explorer

Search BigQuery resources

Showing resources.

TARRED ONLY

Data canvases

Data preparations

External connections

Bellasta

☆

BMR

☆

Bellaapp

☆

Cars

★

Casestudy

☆

Sales

☆

avocado\_data

☆

demos

☆

employee\_data

☆

Untitled query

Bellaapp

QUERY

SHARE

COPY

<

SCHEMA

DETAILS

PREVIEW

TABLE EXPLORER

PREVIEW

INSIGHTS

>

Row	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivities	Very
1	1644430081	2016-04-02	20237	14.7100000...	14.7100000...	0.0	2.569
2	1644430081	2016-04-06	8046	5.84999990...	5.84999990...	0.0	1.289
3	2022484408	2016-04-06	15392	10.9300003...	10.9300003...	0.0	3.430
4	2022484408	2016-04-10	10005	7.01999998...	7.01999998...	0.0	0.079
5	3977333714	2016-04-08	10226	6.84000015...	6.84000015...	0.0	
6	4057192912	2016-04-08	8426	6.15999984...	6.15999984...	0.0	0.129
7	4702921684	2016-04-03	12116	9.82999992...	9.82999992...	0.0	
8	4702921684	2016-04-09	14002	11.3599996...	11.3599996...	0.0	0.610
9	5553957443	2016-04-05	11306	7.38000011...	7.38000011...	0.0	2.309
10	6290855005	2016-04-02	9548	7.21999979...	7.21999979...	0.0	2.019
11	6775888955	2016-04-01	7225	5.17999982...	5.17999982...	0.0	1.730
12	6775888955	2016-04-05	9348	6.69999980...	6.69999980...	0.0	1.129
13	6775888955	2016-04-06	11761	8.43000030...	8.43000030...	0.0	1.309
14	6775888955	2016-04-07	13987	10.0299997...	10.0299997...	0.0	2.869

- Google data analytics capstone by Dr.Kabba

weightlog.modified						
Possible Data Loss Some features might be lost if you...						
Save As...						
A1						
fx   Id						
	A	B	C	D	E	F
1	Id	Date	WeightKg	WeightPound	BMI	
2	1503960366	04/05/2016	53.2999992	117.506384	22.9699993	
3	1927972279	04/10/2016	129.600006	285.719105	46.1699982	
4	2347167796	04/03/2016	63.4000015	139.773078	24.7700005	
5	2873212765	04/06/2016	56.7000008	125.002104	21.4500008	
6	2873212765	04/07/2016	57.2000008	126.104416	21.6499996	
7	2891001357	04/05/2016	88.4000015	194.888643	25.0300007	
8	4445114986	3/30/2016 1	92.4000015	203.707134	35.0099983	
9	4558609924	04/08/2016	69.4000015	153.000813	27.1399994	
10	4702921684	04/04/2016	99.699997	219.800869	26.1100006	
11	6962181067	3/30/2016 1	61.5	135.584291	24.0300007	
12	6962181067	3/31/2016 1	61.5	135.584291	24.0300007	
13	6962181067	04/01/2016	60.9000015	134.261521	23.7800007	
14	6962181067	04/02/2016	61.2000008	134.922906	23.8899994	
15	6962181067	04/03/2016	61.5	135.584291	24.0300007	
16	6962181067	04/04/2016	62.4000015	137.568455	24.3500004	
17	6962181067	04/05/2016	61.7000008	136.025218	24.1000004	
18	6962181067	04/06/2016	62.2000008	137.127529	24.2800007	
19	6962181067	04/07/2016	62.2000008	137.127529	24.2800007	
20	6962181067	04/08/2016	61.7000008	136.025218	24.1000004	
21	6962181067	04/09/2016	62.0999985	136.907062	24.2399998	
22	6962181067	04/10/2016	62.5	137.788914	24.3899994	
23	6962181067	04/11/2016	62.2000008	137.127529	24.2800007	
24	6962181067	04/12/2016	62.5	137.788914	24.3899994	
25	8253242879	04/07/2016	75.5999985	166.669467	29.5499992	
26	8877689391	04/01/2016	85.5	188.495234	25.6100006	
27	8877689391	04/04/2016	86.5999985	190.920316	25.9400005	

weightLogInfo_merged								
Possible Data Loss Some features might be lost if you save this workbook								
Save As...								
A1								
Formatted Deleted Deleted								
	A	B	D	E	F	G	H	J
1	Id	Date	WeightKg	WeightPound	Fat	BMI	IsManualRef	LogId
2	1503960366	#####	53.2999992	117.506384	22	22.9699993	TRUE	1.4599E+12
3	1927972279	#####	129.600006	285.719105		46.1699982	FALSE	1.4603E+12
4	2347167796	#####	63.4000015	139.773078	10	24.7700005	TRUE	1.4597E+12
5	2873212765	#####	56.7000008	125.002104		21.4500008	TRUE	1.46E+12
6	2873212765	#####	57.2000008	126.104416		21.6499996	TRUE	1.4601E+12
7	2891001357	#####	88.4000015	194.888643		25.0300007	TRUE	1.4599E+12
8	4445114986	3/30/2016 1	92.4000015	203.707134		35.0099983	TRUE	1.4594E+12
9	4558609924	#####	69.4000015	153.000813		27.1399994	TRUE	1.4602E+12
10	4702921684	#####	99.6999969	219.800869		26.1100006	TRUE	1.4598E+12
11	6962181067	3/30/2016 1	61.5	135.584291		24.0300007	TRUE	1.4594E+12
12	6962181067	3/31/2016 1	61.5	135.584291		24.0300007	TRUE	1.4595E+12
13	6962181067	#####	60.9000015	134.261521		23.7800007	TRUE	1.4596E+12
14	6962181067	#####	61.2000008	134.922906		23.8899994	TRUE	1.4596E+12
15	6962181067	#####	61.5	135.584291		24.0300007	TRUE	1.4597E+12
16	6962181067	#####	62.4000015	137.568455		24.3500004	TRUE	1.4598E+12
17	6962181067	#####	61.7000008	136.025217		24.1000004	TRUE	1.4599E+12
18	6962181067	#####	62.2000008	137.127529		24.2800007	TRUE	1.46E+12
19	6962181067	#####	62.2000008	137.127529		24.2800007	TRUE	1.4601E+12
20	6962181067	#####	61.7000008	136.025217		24.1000004	TRUE	1.4602E+12
21	6962181067	#####	62.0999985	136.907061		24.2399998	TRUE	1.4602E+12
22	6962181067	#####	62.5	137.788914		24.3899994	TRUE	1.4603E+12
23	6962181067	#####	62.2000008	137.127529		24.2800007	TRUE	1.4604E+12
24	6962181067	#####	62.5	137.788914		24.3899994	TRUE	1.4605E+12
25	8253242879	#####	75.5999985	166.669467		29.5499992	TRUE	1.4601E+12
26	8877689391	#####	85.5	188.495234		25.6100006	FALSE	1.4595E+12
27	8877689391	#####	86.5999985	190.920316		25.9400005	FALSE	1.4598E+12
28	8877689391	#####	86	189.597545		25.7600002	FALSE	1.4598E+12
29	8877689391	#####	86.5999985	190.920316		25.9400005	FALSE	1.4598E+12

Before importing the weight log, the date was formatted, and unnecessary columns were deleted



🏠

✕

BMI

✕

\*Untitled query

✕

daily

✕

🔍Untitled query

▶RUN

📄SAVE

⬇️DOWNLO

1SELECT

2b.Id,

3b.WeightKg,

4b.WeightPounds,

5d.TotalSteps,

6d.ModeratelyActiveDistance,

7d.LightActiveDistance,

8d.SedentaryActiveDistance,

9d.VeryActiveMinutes,

10d.FairlyActiveMinutes,

11d.LightlyActiveMinutes,

12d.SedentaryMinutes,

13d.calories,

14FROM

15`Bellasta.BMI` b

16JOIN

17`Bellasta.daily`d on b.Id=d.Id;

- Google data analytics capstone by Dr.Kabba

Merge the two table into one table using the Id field

Query results

⬇️SAVE RESULTS

📊EXPLORE DATA

⬆️

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION GRAPH	
Row	TotalSteps	ModeratelyActiveDis	LightActiveDistance	SedentaryActiveDist	VeryActiveMinutes	FairlyActiveMinutes		
1	5336	0.0	3.529999971	0.0	0	0		
2	3183	0.0	2.099999905	0.0	0	0		
3	8205	0.0	5.420000076	0.0	0	0		
4	5668	0.0	3.75	0.0	0	0		
5	5512	0.0	3.640000105	0.0	0	0		
6	6263	0.0	4.130000114	0.0	0	0		
7	8828	0.0	5.840000153	0.0	0	0		
8	1260	0.0	0.819999993	0.0	0	0		
9	6176	0.25999999	3.75	0.0	1	7		

🏠

×

Bellasta

×

\*Untitled query

×

\*Untitled query

×

+

⋮

🔍Untitled query

▶RUN

📄SAVE

⬇️DOWNLOAD

👤SHARE

⋮

✅This quer..

```
1 CREATE TABLE Bellasta.CombinedData AS
2 SELECT
3     B.Id,
4     B.WeightKg,
5     B.WeightPounds,
6     B.BMI,
7     D.TotalSteps,
8     D.ModeratelyActiveDistance,
9     D.LightActiveDistance,
10    D.SedentaryActiveDistance,
11    D.VeryActiveMinutes,
12    D.FairlyActiveMinutes,
13    D.LightlyActiveMinutes,
14    D.SedentaryMinutes,
15    D.Calories
16 FROM
17     Bellasta.BMI B
18 JOIN
19     Bellasta.daily D ON B.Id = D.Id;
20
```

- Google data analytics capstone by Dr.Kabba

Create a new table from the combined data(Simplified Analysis)

🔍 Search BigQuery resources

ewing resources.

SHOW STARRED ONLY

▶📓 Notebooks

▶🗂️ Data canvases

▶🔧 Data preparations

▶🔗 External connections

▼📊 Bellasta

📊 BMI

📊 CombinedData

📊 daily

▶🚗 Cars

▶📖 Casestudy

▶📈 Sales

▶🥑 avocado\_data

▶📺 demos

▶👥 employee\_data

Results per page: 50 1 – 50 of 436 ⏪ ⏩

- Google data analytics capstone by Dr.Kabba

▼ X 🔍 Untitled query ▼ X 📊 Combine... ata ▼ X 🔍 \*Untitled query ▼ X 🔍 \*Untitled query ▼ X + ▼

Untitled query 

▶ RUN

📄 SAVE ▼

⬇️ DOWNLOAD

👤 SHARE ▼

🕒 SCHEDULE

⚙️ MORE ▼

✅ Query comp

1 CREATE TABLE Bellasta.FilteredData AS

2 SELECT

3     Id,

4     WeightKg,

5     BMI,

6     TotalSteps,

7     VeryActiveMinutes,

8     FairlyActiveMinutes,

9     SedentaryMinutes,

0     Calories

1 FROM `Bellasta.CombinedData`

2

Press Option+F1 for Accessibility Options

Create a New Table with Only the Relevant Data

xplorer

+ ADD

🔍 Search BigQuery resources

Showing resources.

HOW STARRED ONLY

▶ 📓 Notebooks

▶ 📁 Data canvases

▶ ⚙️ Data preparations

▶ ➡️ External connections

▼ 📊 Bellasta

📊 BMI

📊 CombinedData

📊 FilteredData

📊 daily

▶ 🚗 Cars

▶ 📖 Casestudy

▶ 📈 Sales

▶ 🥑 avocado\_data

▶ 📺 demos

< X 📊 Combine... ata ▼ X 🔍 \*Untitled query ▼ X 🔍 \*Untitled query ▼ X 📊 FilteredData ▼ X > + ▼

📊 FilteredData 🔍 QUERY 👤 SHARE 📄 COPY 📷 SNAPSHOT 🗑️ DELETE 📄 EXPORT ▼ ↻

<

SCHEMA

DETAILS

PREVIEW

TABLE EXPLORER

PREVIEW

INSIGHTS

LINEAGE

DATA PROFILE

Row	Id	WeightKg	BMI	TotalSteps	VeryActiveMinut	FairlyActiveMinu	SedentaryMinut	Calories	
1	4558609924	69.40000153	27.13999939	3183	0	0	1180	1680	
2	4558609924	69.40000153	27.13999939	8828	0	0	1132	2149	
3	4558609924	69.40000153	27.13999939	6799	24	14	1217	1876	
4	4558609924	69.40000153	27.13999939	6263	0	0	1150	1906	
5	4558609924	69.40000153	27.13999939	5668	0	0	1181	1853	
6	4558609924	69.40000153	27.13999939	8205	0	0	1073	2078	
7	4558609924	69.40000153	27.13999939	7998	27	12	1105	2235	
8	4558609924	69.40000153	27.13999939	5336	0	0	1182	1902	
9	4558609924	69.40000153	27.13999939	6176	1	7	886	1901	
10	4558609924	69.40000153	27.13999939	1260	0	0	555	722	
11	4558609924	69.40000153	27.13999939	5512	0	0	1164	1882	
12	4558609924	69.40000153	27.13999939	4195	0	0	1199	1778	
13	2891001357	88.40000153	25.03000069	2063	0	0	99	1144	
14	2891001357	88.40000153	25.03000069	4126	0	0	720	2881	

Results per page: 50 ▼ 1 – 50 of 436 |< < > >



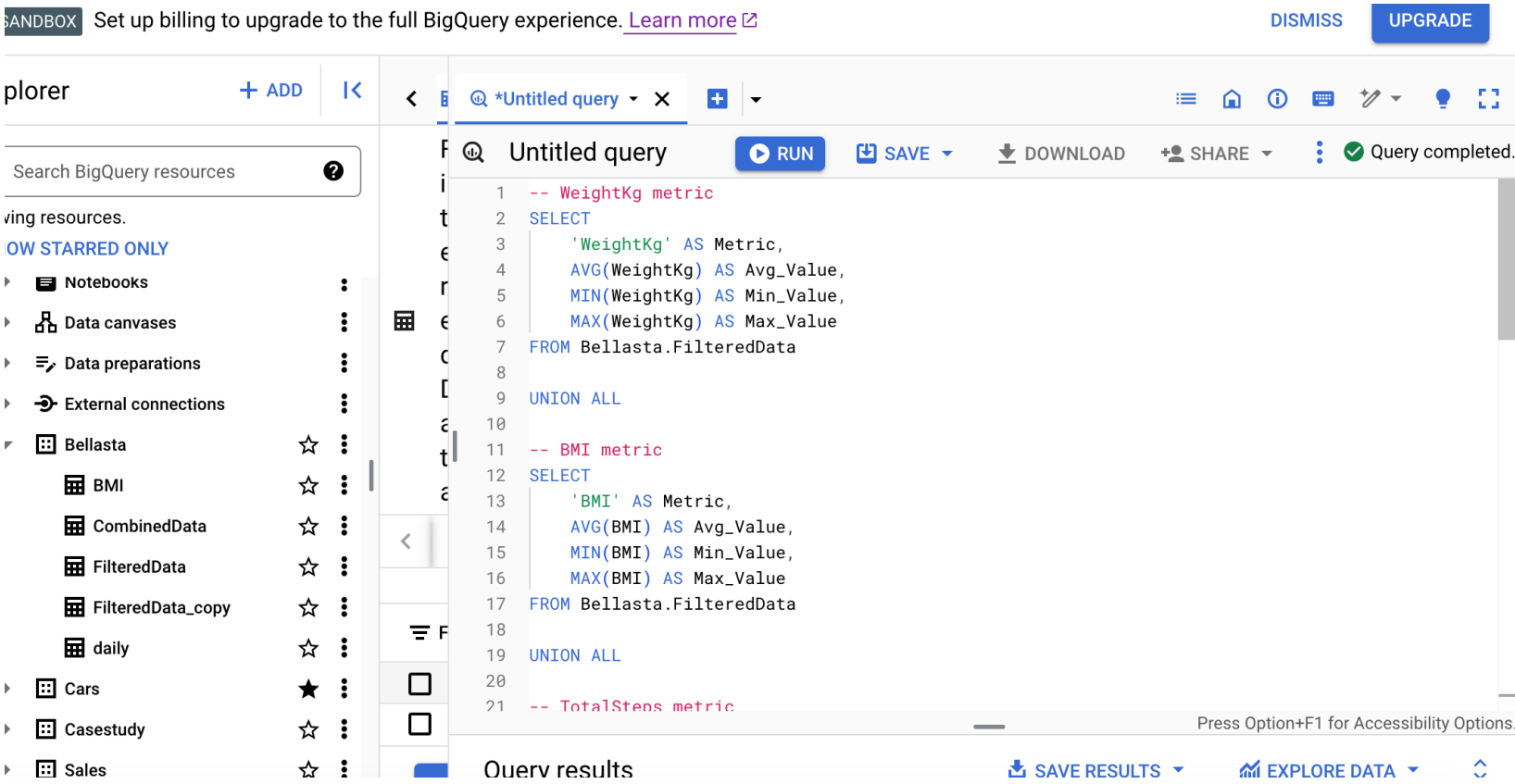
- Google data analytics capstone by Dr.Kabba

```
SELECT
  Id,
  WeightKg,
  BMI,
  TotalSteps,
  VeryActiveMinutes,
  FairlyActiveMinutes,
  SedentaryMinutes,
  Calories
FROM
  `Bellasta.CombinedData`
```

Remove Unnecessary Columns in SQL

JOB INFORMATION		RESULTS		CHART		JSON		EXECUTION DETAILS		EXECUTION GRAPH	
N	Id	WeightKg	BMI	TotalSteps	VeryActiveMinutes	FairlyActiveMinutes	SedentaryMinutes	C			
1	4558609924	69.40000153	27.13999939	6263	0	0	1150				
2	4558609924	69.40000153	27.13999939	7998	27	12	1105				
3	4558609924	69.40000153	27.13999939	5336	0	0	1182				
4	4558609924	69.40000153	27.13999939	3183	0	0	1180				
5	4558609924	69.40000153	27.13999939	6799	24	14	1217				
6	4558609924	69.40000153	27.13999939	1260	0	0	555				
7	4558609924	69.40000153	27.13999939	6176	1	7	886				
8	4558609924	69.40000153	27.13999939	5512	0	0	1164				
9	4558609924	69.40000153	27.13999939	5668	0	0	1181				
10	4558609924	69.40000153	27.13999939	8205	0	0	1073				

- Google data analytics capstone by Dr.Kabba



Do some summay statictics

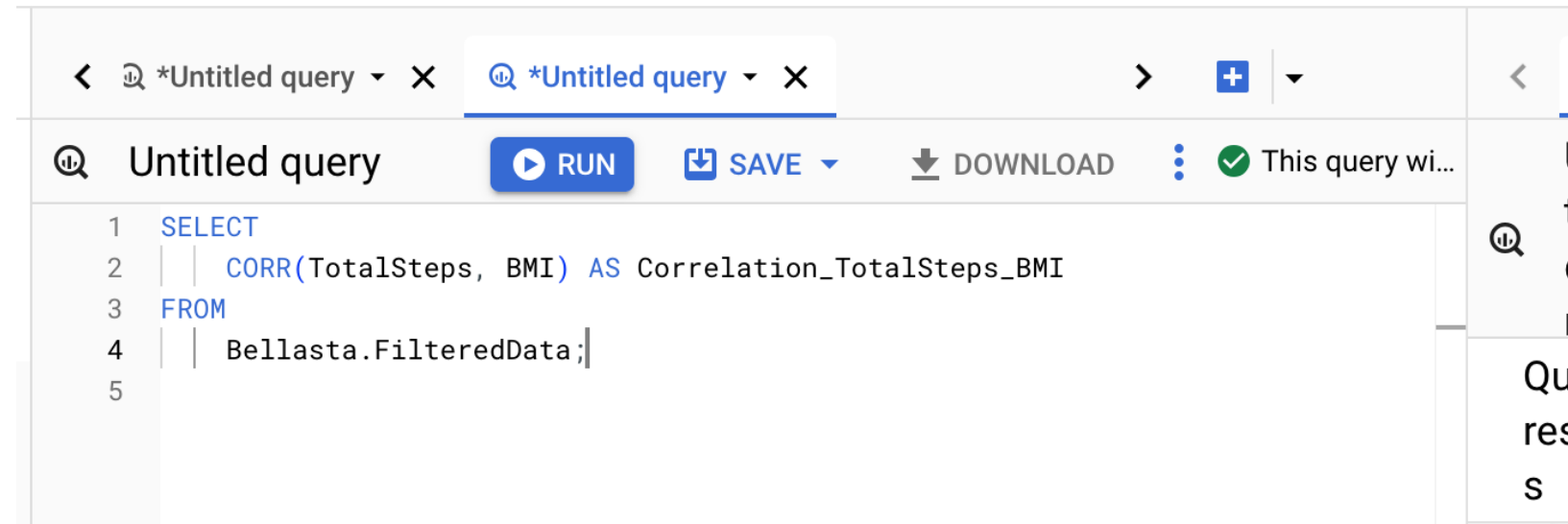
Query results

SAVE RESULTSEXPLORE DATA

	JOB INFORMATION	RESULTS	CHART	JSON	EXECUTION DETAILS
	Metric	Avg_Value	Min_Value	Max_Value	
1	TotalSteps	11926.34403669...	0.0	28497.0	
2	BMI	25.65165144871...	21.45000076	46.16999817	
3	SedentaryMinutes	821.2316513761...	32.0	1440.0	
4	Calories	2399.197247706...	0.0	4562.0	
5	FairlyActiveMinutes	20.97935779816...	0.0	660.0	
6	VeryActiveMinutes	35.25229357798...	0.0	128.0	
7	WeightKg	72.48027571360...	53.29999924	129.6000061	

<b>Metric</b>	<b>Avg_Value</b>	<b>Min_Value</b>	<b>Max_Value</b>
<b>TotalSteps</b>	11926.34	0	28497
<b>BMI</b>	25.65	21.45	46.16
<b>SedentaryMinutes</b>	821.23	32	1440
<b>Calories</b>	2399.19	0	4562
<b>FairlyActiveMinutes</b>	20.97	0	660
<b>VeryActiveMinutes</b>	35.25	0	128
<b>WeightKg</b>	72.48	53.29	129.60

- Google data analytics capstone by Dr.KABBA



Correlation TotalSteps BMI
----------------------------

-0.31282700030
----------------

The negative value (-0.3128) suggests a **negative relationship** between TotalSteps and BMI. This means that as TotalSteps increases, BMI tends to decrease, though the relationship is not very strong.

- Google data analytics capstone course by Dr.KABBA

Untitled query

RUN

SAVE

DOWNLOAD

Query complete

1

2

3

4

SELECT

CORR(SedentaryMinutes, BMI) AS Correlation\_SedentaryMinutes\_BMI

FROM Bellasta.FilteredData;

Press Option+F1 for Accessibility Options.

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION

RESULTS

CHART

JSON

EXECUTION DET.

Row	Correlation_SedentaryMinutes_BMI
1	0.190197846645...

Correlation SedentaryMenutes_-BMI
0.1901978466455

The positive value of 0.1902 indicates a slight positive relationship between SedentaryMinutes and BMI. This means that, as the number of sedentary minutes increases, BMI tends to increase slightly as well.

# Data Analysis Workflow

---

- **Phase 1: Data importation**

Imported raw data into the SQL environment.

- **Phase 2: Data Merging**

Merged multiple datasets into a single table for comprehensive analysis.

- **Phase 3: Data Transformation**

Created a new table from the joined data  
Filtered the data to include only relevant information

- **Phase 4: Data Cleaning**

- Removed unnecessary columns to streamline the dataset.

# Data Analysis Workflow

---

- **Phase 5: Analysis**

Conducted summary statistics to gain insights into the data.

Calculated correlations to understand relationships between variables.