Fatima Ezzahra Kabba

SEPTEMBRE 2024

Case Study: SQL as a Tool

bellabeat

Presented by Fatima Ezzahra Kabba

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.

Prepared by Fatima Ezzahra Kabba

TABLE
of contents

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

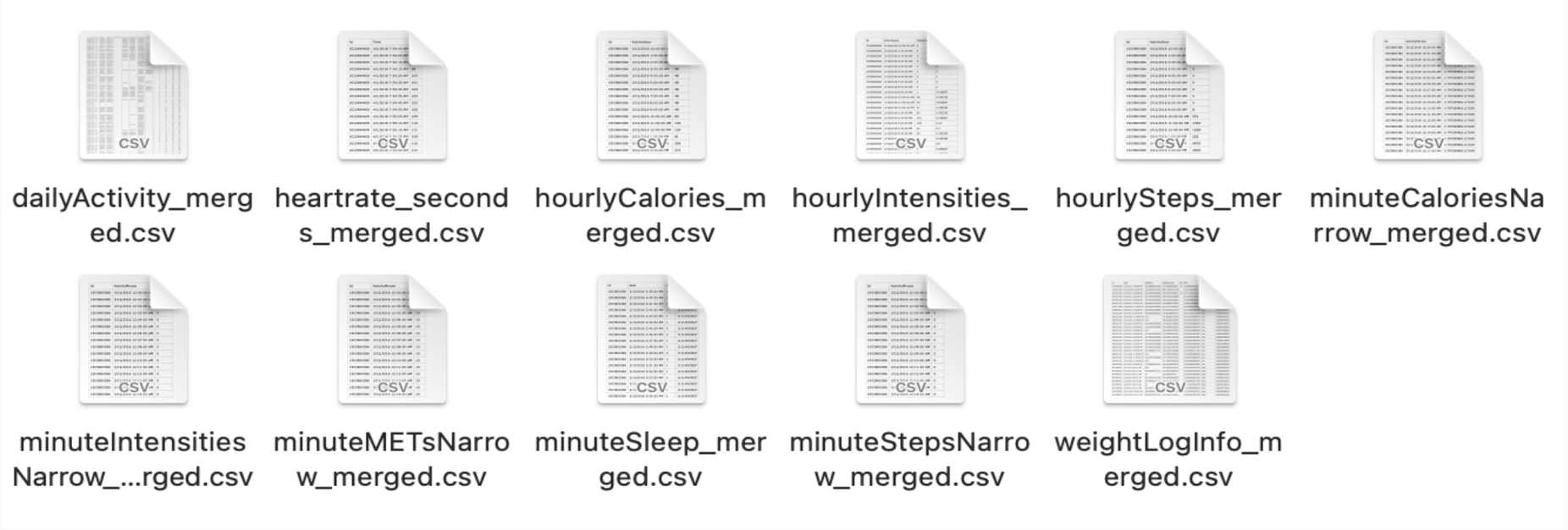
Analyzing data fitness App to unlock new growth opportunities for the company

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

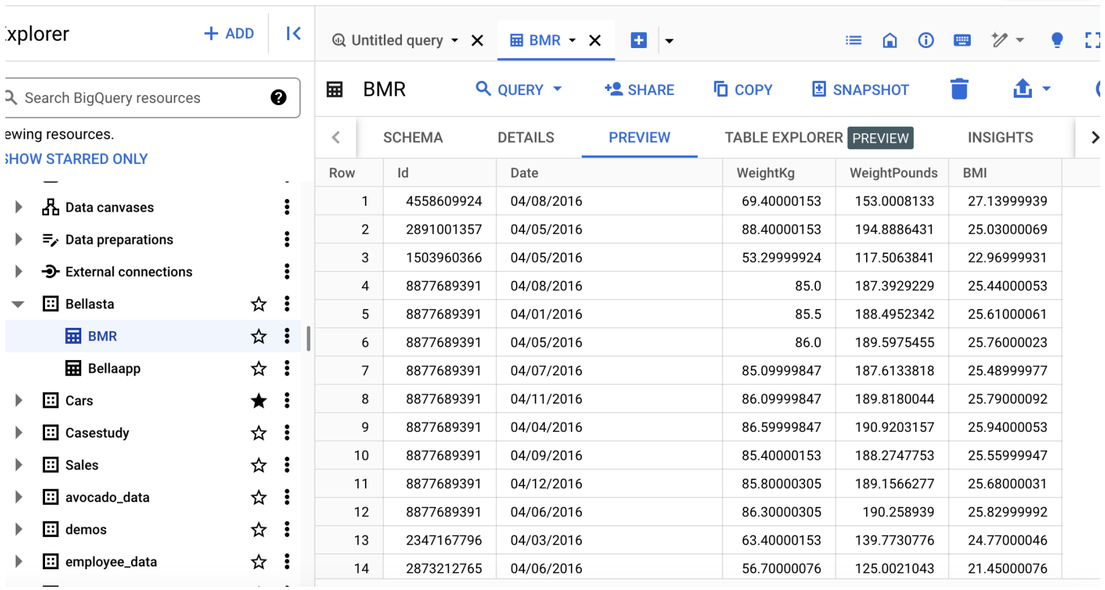


- Google data analytics capstone by Dr.Kabba

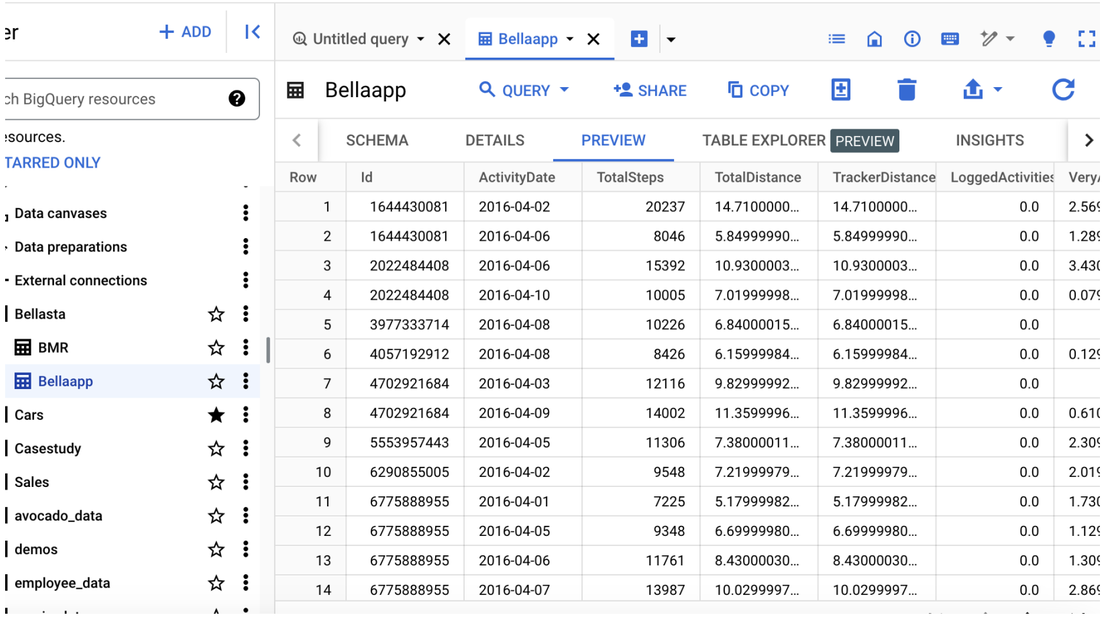
Step 1.Import my data to SQL

Two table was imported in SQL

.

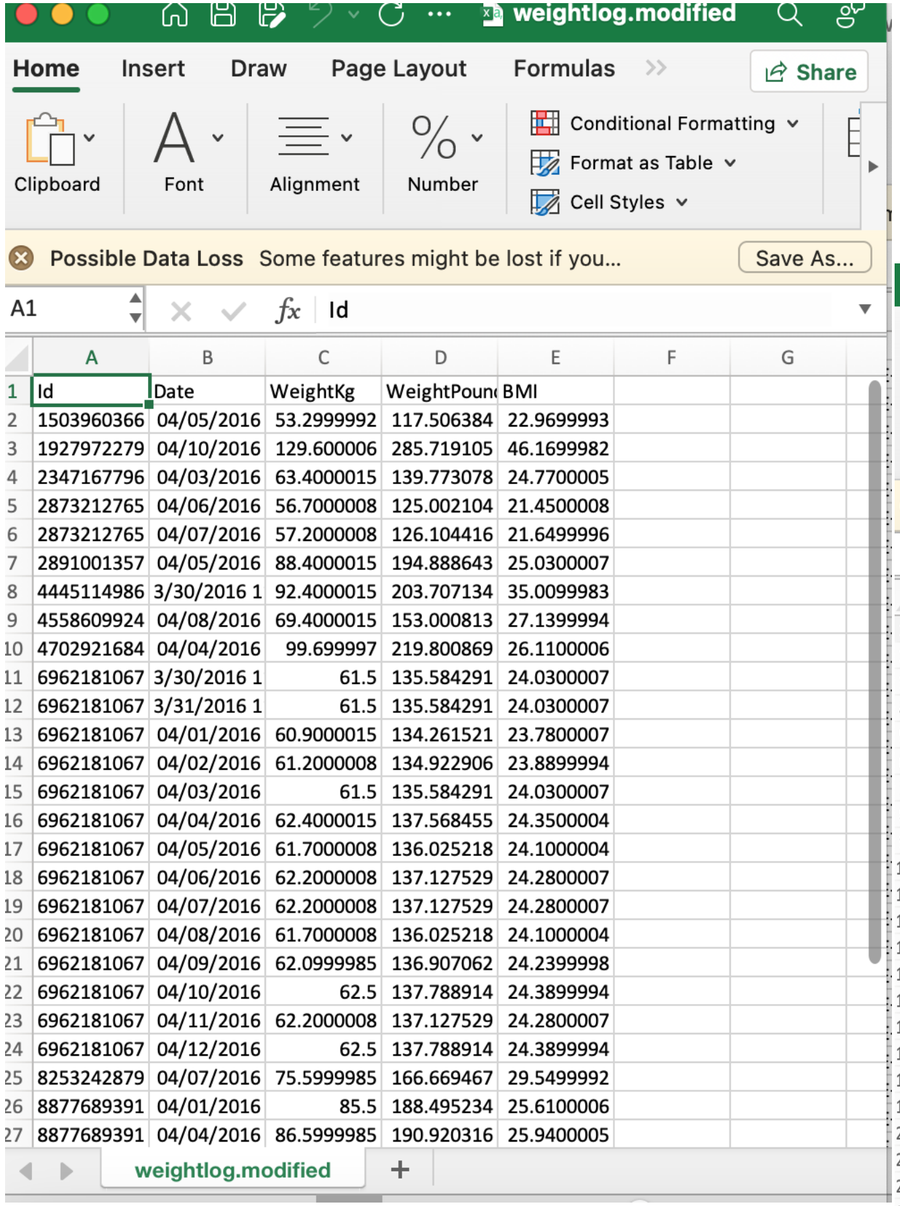


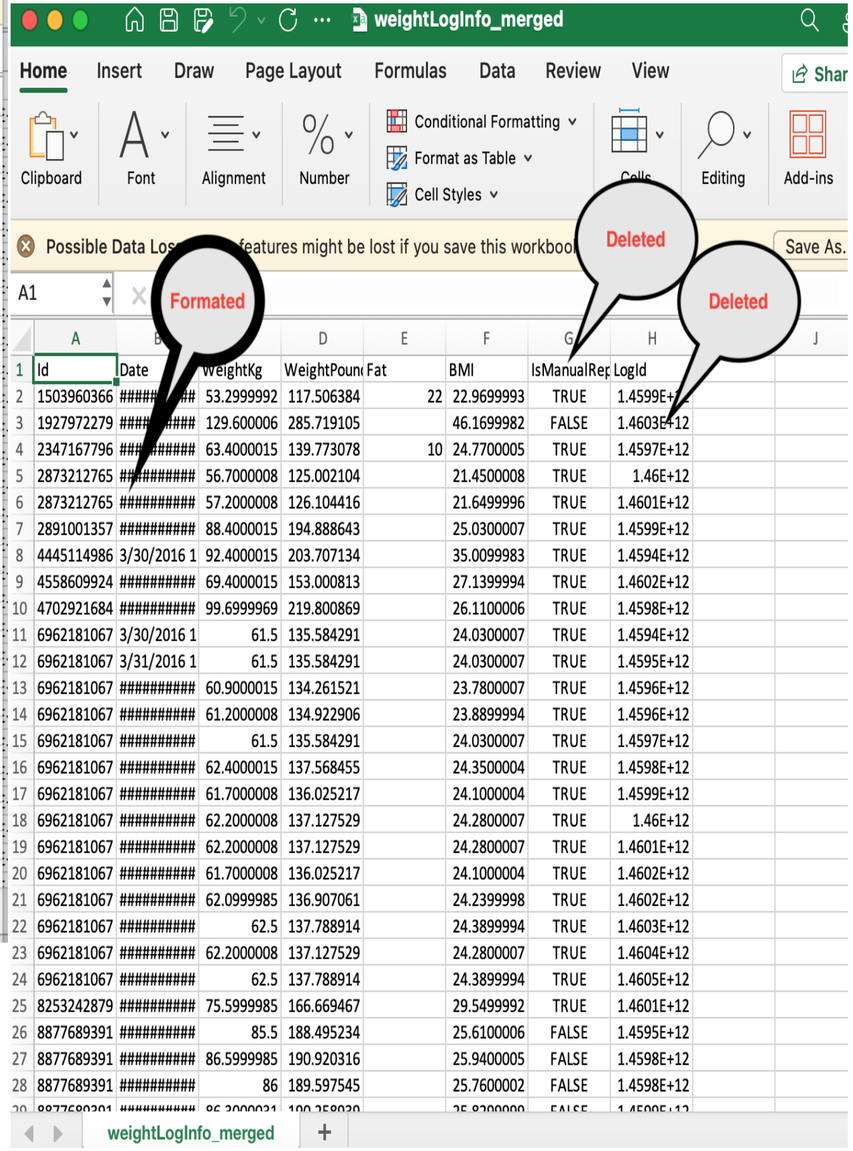
Row	Id	Date	WeightKg	WeightPounds	BMI
1	4558609924	04/08/2016	69.40000153	153.0008133	27.13999939
2	2891001357	04/05/2016	88.40000153	194.8886431	25.03000069
3	1503960366	04/05/2016	53.29999924	117.5063841	22.96999931
4	8877689391	04/08/2016	85.0	187.3929229	25.44000053
5	8877689391	04/01/2016	85.5	188.4952342	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



Row	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivities	Very
1	1644430081	2016-04-02	20237	14.7100000...	14.7100000...	0.0	2.56
2	1644430081	2016-04-06	8046	5.84999990...	5.84999990...	0.0	1.28
3	2022484408	2016-04-06	15392	10.9300003...	10.9300003...	0.0	3.43
4	2022484408	2016-04-10	10005	7.01999998...	7.01999998...	0.0	0.07
5	3977333714	2016-04-08	10226	6.84000015...	6.84000015...	0.0	
6	4057192912	2016-04-08	8426	6.15999984...	6.15999984...	0.0	0.12
7	4702921684	2016-04-03	12116	9.82999992...	9.82999992...	0.0	
8	4702921684	2016-04-09	14002	11.3599996...	11.3599996...	0.0	0.61
9	5553957443	2016-04-05	11306	7.38000011...	7.38000011...	0.0	2.30
10	6290855005	2016-04-02	9548	7.21999979...	7.21999979...	0.0	2.01
11	6775888955	2016-04-01	7225	5.17999982...	5.17999982...	0.0	1.73
12	6775888955	2016-04-05	9348	6.69999980...	6.69999980...	0.0	1.12
13	6775888955	2016-04-06	11761	8.43000030...	8.43000030...	0.0	1.30
14	6775888955	2016-04-07	13987	10.0299997...	10.0299997...	0.0	2.86

- Google data analytics capstone by Dr.Kabba





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

- Google data analytics capstone by Dr.Kabba

Merge the two table into one table using the Id field

BMI | *Untitled query | daily

Untitled query RUN SAVE DOWNL

```sql
SELECT
 b.Id,
 b.WeightKg,
 b.WeightPounds,
 d.TotalSteps,
 d.ModeratelyActiveDistance,
 d.LightActiveDistance,
 d.SedentaryActiveDistance,
 d.VeryActiveMinutes,
 d.FairlyActiveMinutes,
 d.LightlyActiveMinutes,
 d.SedentaryMinutes,
 d.calories,
 FROM
 `Bellasta.BMI` b
JOIN
`Bellasta.daily`d on b.Id=d.Id;
```

Query results SAVE RESULTS EXPLORE DATA

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

ow	TotalSteps	ModeratelyActiveDis	LightActiveDistance	SedentaryActiveDista	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

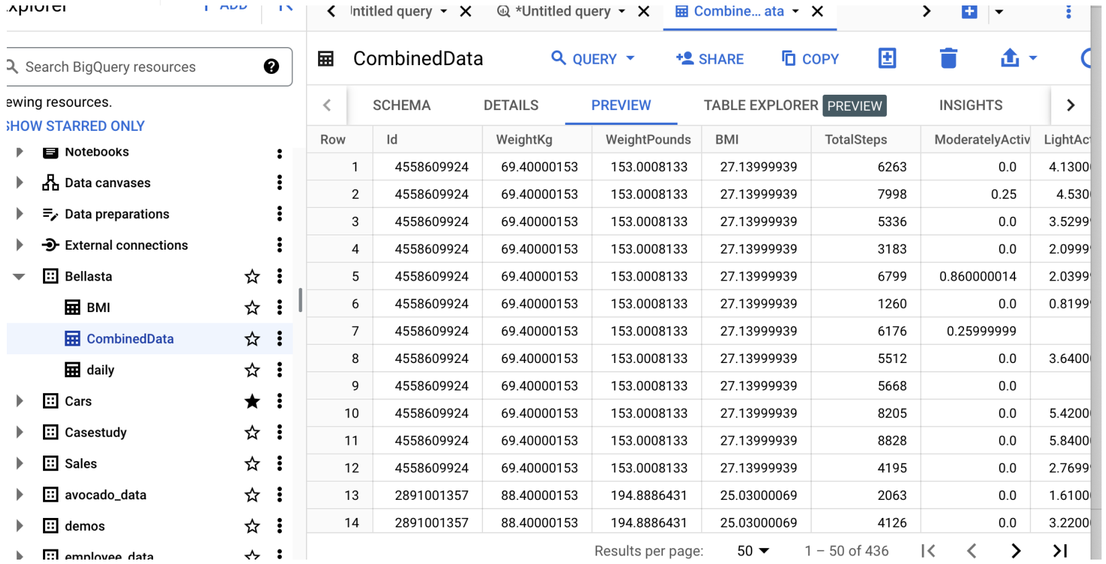
- Google data analytics capstone by Dr.Kabba

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

Bellasta | *Untitled query | *Untitled query

Untitled query | RUN | SAVE | DOWNLOAD | SHARE | This quer..

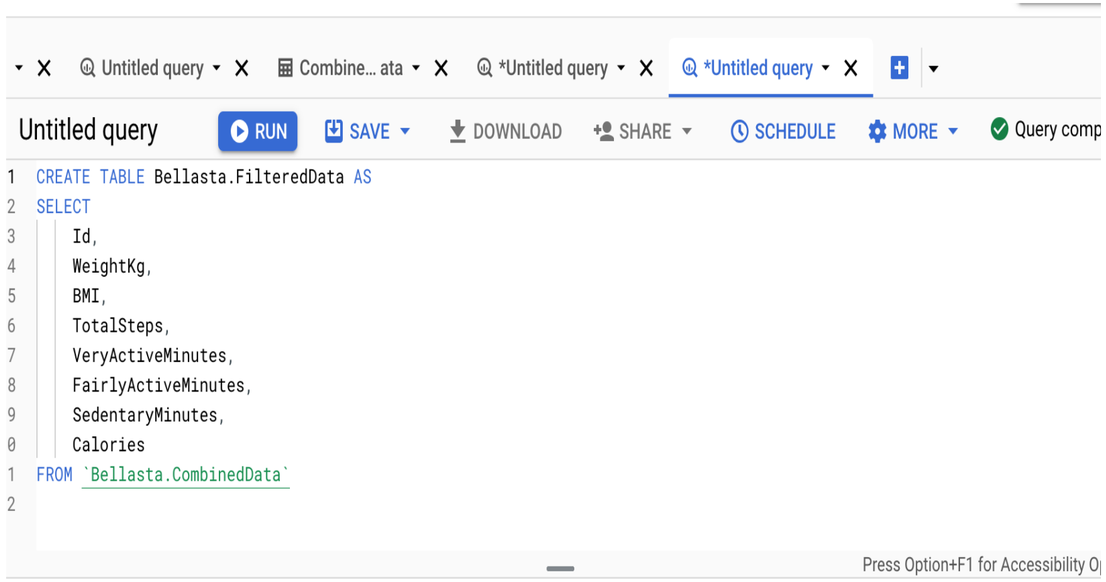
```
1  CREATE TABLE Bellasta.CombinedData AS
2  SELECT
3      B.Id,
4      B.WeightKg,
5      B.WeightPounds,
6      B.BMI,
7      D.TotalSteps,
8      D.ModeratelyActiveDistance,
9      D.LightActiveDistance,
10     D.SedentaryActiveDistance,
11     D.VeryActiveMinutes,
12     D.FairlyActiveMinutes,
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
```

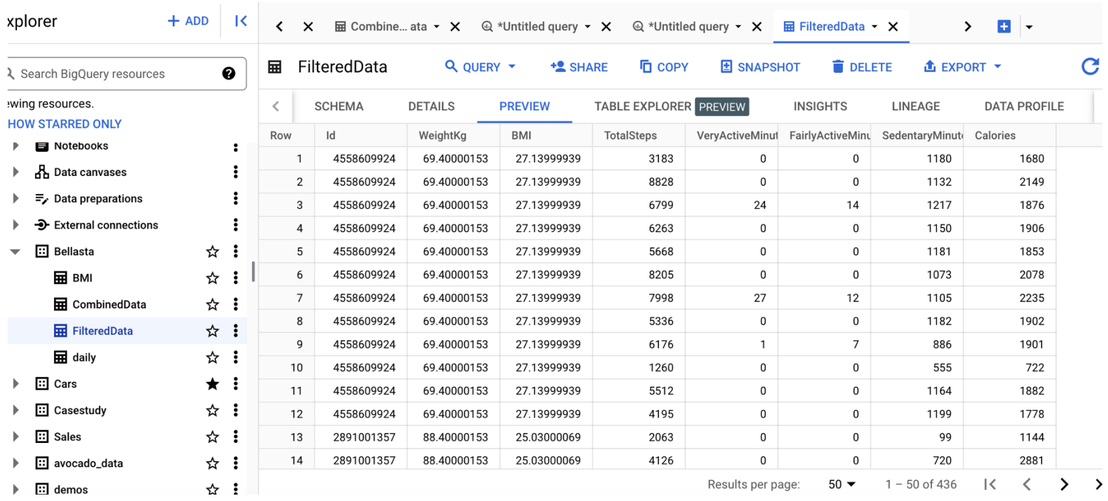


Row	Id	WeightKg	WeightPounds	BMI	TotalSteps	ModeratelyActiv	LightAc
1	4558609924	69.40000153	153.0008133	27.13999939	6263	0.0	4.1300
2	4558609924	69.40000153	153.0008133	27.13999939	7998	0.25	4.530
3	4558609924	69.40000153	153.0008133	27.13999939	5336	0.0	3.5299
4	4558609924	69.40000153	153.0008133	27.13999939	3183	0.0	2.0999
5	4558609924	69.40000153	153.0008133	27.13999939	6799	0.860000014	2.0399
6	4558609924	69.40000153	153.0008133	27.13999939	1260	0.0	0.8199
7	4558609924	69.40000153	153.0008133	27.13999939	6176	0.25999999	
8	4558609924	69.40000153	153.0008133	27.13999939	5512	0.0	3.6400
9	4558609924	69.40000153	153.0008133	27.13999939	5668	0.0	
10	4558609924	69.40000153	153.0008133	27.13999939	8205	0.0	5.4200
11	4558609924	69.40000153	153.0008133	27.13999939	8828	0.0	5.8400
12	4558609924	69.40000153	153.0008133	27.13999939	4195	0.0	2.7699
13	2891001357	88.40000153	194.8886431	25.03000069	2063	0.0	1.6100
14	2891001357	88.40000153	194.8886431	25.03000069	4126	0.0	3.2200

- Google data analytics capstone by Dr.Kabba

Create a New Table with Only the Relevant Data





- Google data analytics capstone by Dr.Kabba

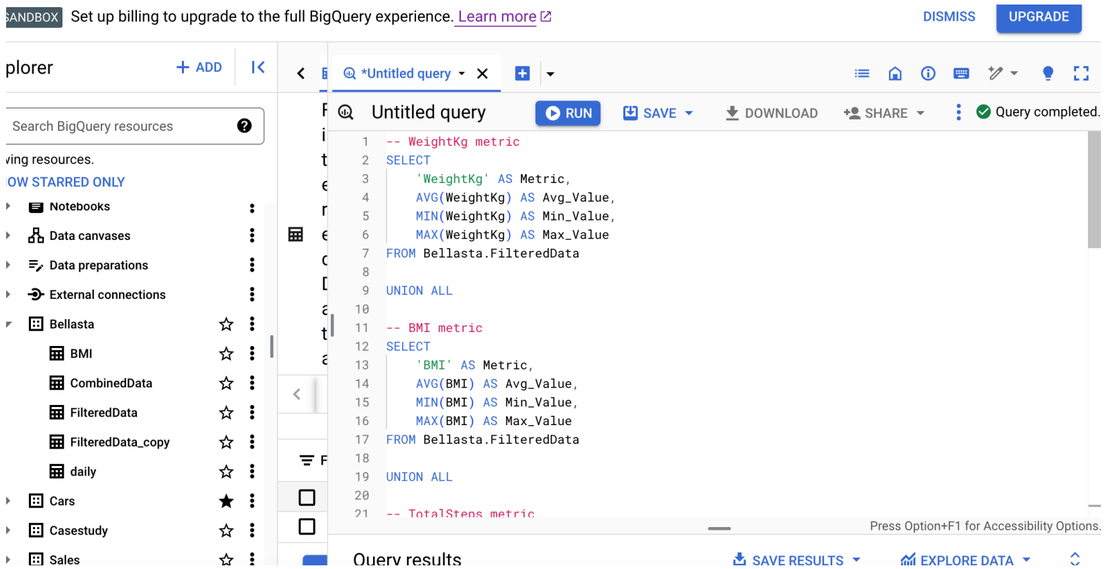
```sql
SELECT
    Id,
    WeightKg,
    BMI,
    TotalSteps,
    VeryActiveMinutes,
    FairlyActiveMinutes,
    SedentaryMinutes,
    Calories
    FROM
    `Bellasta.CombinedData`
```

Remove Unnecessary Columns in SQL

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	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 RESULTS

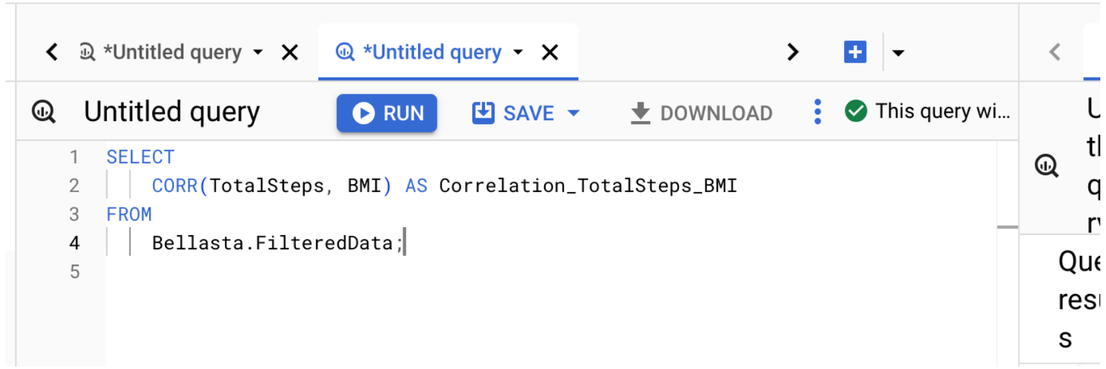
EXPLORE DAT

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

w	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

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

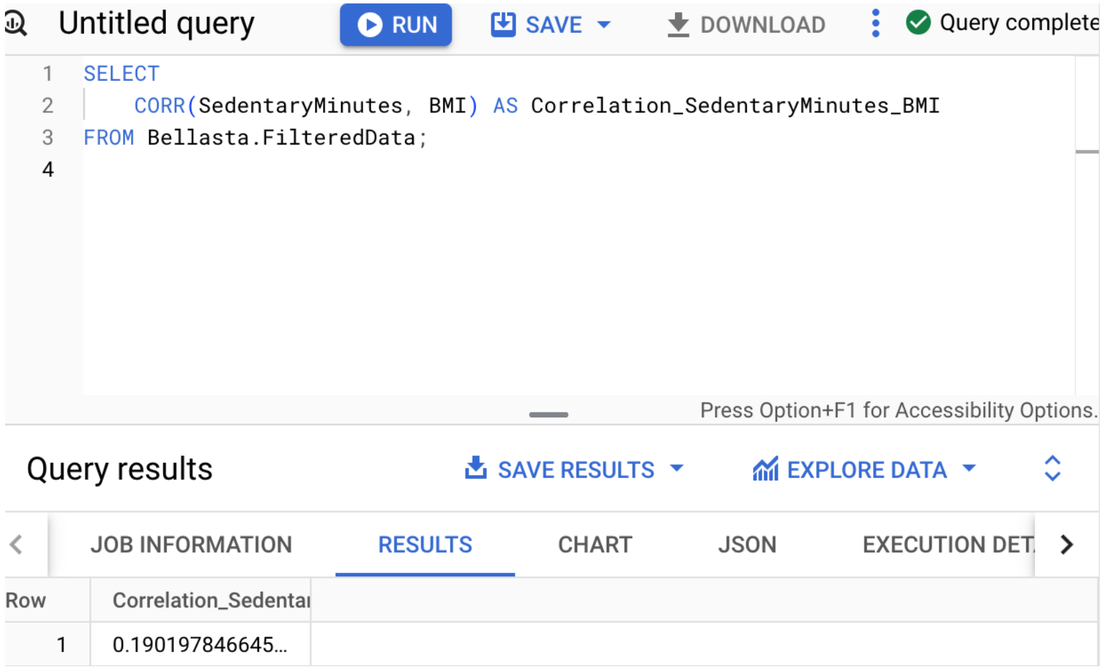
- Google data analytics capstone by Dr.KABBA



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



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.