STRAVA FITNESS DATA ANALYSIS

```
--PREPARE PHASE--
--1.I will be using a dataset provided by MObius for the analysis of the usage of fitbit
device tracker.
--2.I will checking the data consistency, duplicates and Null values
--3.I will join the tables that have relating data
--4.I will use aggregate functions to count, sum or average columns from the tables
--5.I will analyze the data in SQL &
--6. I will perform data visualization IN POWER BI
--
--checking the columns of each table--

SELECT *
FROM dailyActivity_merged
SELECT *
FROM dailyCalories_merged
SELECT *
FROM dailyIntensities_merged
SELECT *
FROM heartrate_seconds_merged
SELECT *
FROM hourlyCalories_merged
SELECT *
FROM hourlyIntensities_merged
SELECT *
FROM hourlySteps_merged
SELECT *
FROM minuteCaloriesNarrow_merged
SELECT *
FROM minuteMETsNarrow_merged
SELECT *
FROM minuteSleep_merged
SELECT *
FROM sleepDay_merged
SELECT *
FROM weightLogInfo_merged

---ASK PHASE

--How many people were selected to take part in the data collection process?--

SELECT COUNT(DISTINCT(ID)) AS POPULATION
FROM [dbo].[dailyActivity_merged]
--33


--How many days was the activity carried out?--

SELECT COUNT(DISTINCT(ActivityDate)) AS NumberofDays
FROM [dbo].[dailyActivity_merged]
--31


--What distance was covered by each ID?--
```

```sql
SELECT Id,Sum(TotalDistance) AS Distance
FROM [dbo].[dailyActivity_merged]
GROUP BY Id
ORDER BY Distance DESC


--PROCESS PHASE--

--Checking any NULL value in the tables Activity,Sleep & Weight info Logged?--

SELECT *
FROM [dbo].[dailyActivity_merged]
WHERE Id IS NULL

--there is no NULL values in activity table--

SELECT *
FROM [dbo].[sleepDay_merged]
WHERE Id IS NULL

---there is no NULL values in sleepDay table--

SELECT *
FROM [dbo].[weightLogInfo_merged]
WHERE Id IS NULL

----there is no NULL values in weight table--


--checking for Duplicate IDs in the datasets--


SELECT COUNT(DISTINCT (Id))
FROM [dbo].[dailyActivity_merged]
--33

SELECT COUNT(DISTINCT (Id))
FROM [dbo].[dailyCalories_merged]
--33

SELECT COUNT(DISTINCT (Id))
FROM [dbo].[dailyIntensities_merged]
--33

--check activity with calories--

SELECT COUNT(dailyActivity_merged.Id)

FROM dailyActivity_merged

INNER JOIN dailyCalories_merged ON dailyActivity_merged.Id = dailyCalories_merged.Id

AND dailyActivity_merged.ActivityDate = dailyCalories_merged.ActivityDay

AND dailyActivity_merged.Calories = dailyCalories_merged.Calories
--940
```

```sql
--check activity with steps--

SELECT COUNT(dailyActivity_merged.Id)

FROM dailyActivity_merged

INNER JOIN minuteStepsNarrow_merged ON dailyActivity_merged.Id =
minuteStepsNarrow_merged.Id

AND dailyActivity_merged.ActivityDate = minuteStepsNarrow_merged.ActivityMinute

AND dailyActivity_merged.TotalSteps = minuteStepsNarrow_merged.Steps
--72


--check activity with intensities--

SELECT COUNT(dailyActivity_merged.Id)

FROM dailyActivity_merged

INNER JOIN dailyIntensities_merged ON dailyActivity_merged.Id =
dailyIntensities_merged.Id

AND dailyActivity_merged.ActivityDate = dailyIntensities_merged.ActivityDay

AND dailyActivity_merged.LightlyActiveMinutes =
dailyIntensities_merged.LightlyActiveMinutes

AND dailyActivity_merged.FairlyActiveMinutes =
dailyIntensities_merged.FairlyActiveMinutes

AND dailyActivity_merged.VeryActiveMinutes = dailyIntensities_merged.VeryActiveMinutes

AND dailyActivity_merged.LightActiveDistance =
dailyIntensities_merged.LightActiveDistance

AND dailyActivity_merged.ModeratelyActiveDistance =
dailyIntensities_merged.ModeratelyActiveDistance

AND dailyActivity_merged.VeryActiveDistance = dailyIntensities_merged.VeryActiveDistance

--940


--After Analysis, I find out calories, intensities and steps data’s column all match with
activity data.
--In this case I will use activity data to continue the process

--Finding duplicates of row from activity, sleep and weight

SELECT Id, ActivityDate, COUNT(*)

FROM dailyActivity_merged

GROUP BY Id, ActivityDate
```

```
HAVING COUNT(*) > 1

--no duplicates-

SELECT Id, Date, COUNT(*)

FROM weightLogInfo_merged

GROUP BY Id, Date

HAVING COUNT(*) > 1

--no duplicates-

SELECT Id, SleepDay, COUNT(*)

FROM sleepDay_merged

GROUP BY Id, SleepDay

HAVING COUNT(*) > 1

--no duplicates-

--ANALYZE AND SHARE PHASE--

--Find how active people use device

SELECT Id, Count(Id) as total_activity,

CASE

WHEN COUNT(Id) BETWEEN 21 and 31 THEN 'Active User'

WHEN COUNT(Id) BETWEEN 11 and 20 THEN 'Moderate User'

WHEN COUNT(Id) BETWEEN 0 and 10 THEN 'Light User'

END AS user_activity_level

FROM dailyActivity_merged

Group by Id

ORDER BY total_activity DESC


 --I find out 93% of fitbit users are active user who use the tracker 21-31 days a month.
 --6% of fitbit users are moderate user who use the tracker 11-20 days a month.
 --0.43% of fitbit users are light user who use the tracker 0-10 days a month.


 --Find the average of activity minutes per week

 SELECT AVG(VeryActiveMinutes) AS VeryActive,
```

```
       AVG(FairlyActiveMinutes) AS FailyActive,
       AVG(LightlyActiveMinutes) AS LightlyActive,
       AVG(SedentaryMinutes) AS SedentaryMinutes
FROM dailyActivity_merged


--Finding average total step vs calories

SELECT Id,

AVG(Calories) as avg_calories,

AVG(TotalSteps) as avg_total_step

FROM dailyActivity_merged

GROUP BY Id

ORDER BY avg_calories DESC


--There exist a positive relationship betweens steps and amount of calories burnt.
--Meaning the more the steps a user made the more the amount of calories they burnt.



--finding average weight with respect to sleep minutes

SELECT sleepDay_merged.Id,

AVG(sleepDay_merged.TotalMinutesAsleep) as avg_sleepingtime,

AVG(weightLogInfo_merged.BMI) as avg_BMI

FROM sleepDay_merged
FULL JOIN weightLogInfo_merged
ON sleepDay_merged.Id=weightLogInfo_merged.Id

GROUP BY sleepDay_merged.Id

ORDER BY avg_BMI DESC

--High BMI is visible in users with averagly more sleeping time  (high body fatness)
--But majority of users did not give enough data on weight and BMI


--finding average sleep time and amount of calories burnt

SELECT sleepDay_merged.Id,

AVG(sleepDay_merged.TotalMinutesAsleep) as avg_sleepingtime,

AVG(dailyCalories_merged.Calories) as avg_Calories

FROM sleepDay_merged
FULL JOIN dailyCalories_merged
ON sleepDay_merged.Id=dailyCalories_merged.Id
```

```
GROUP BY sleepDay_merged.Id

ORDER BY avg_Calories DESC

--There is major relationship between the number of minutes a user slept and the amount
of calories they burnt
--meaning longer sleeping users burnt the least calories and vice versa


--CONCLUSION--

--1. The company can consider collecting data from many users or come up with an APP
linked to their internal data source
--and to every device to help in collection of sufficient data.
--A population of only 33 users is quite small and this narrows the confidence of making
broader decisions
--2.I can recommend that the most active users be rewarded as most loyal customers for
their
--loyalty and continued use of the devices
--3.The company can  consider tracking the performance data for other products i.e.
spring bottle,Time,BellabeatAPP etc
--4.The company can consider using internal data.
--5.The company can also analyze the sales generated by the fitbit and ascertain whether
they are making profits.
--6.I would finally recommend the dataset to have gender specified so that the
variability in the outcome can be grouped by gender.
```