

```
In [5]: 1 # #Task 1 - Fitbit Dataset
2
3 # 1. Read this dataset in pandas , mysql and mongodb
4 # 2. while creting a table in mysql dont use manual approach to create it ,always use a automati
5 # ## hint - use csvkit library to automate this task and to Load a data in bulk in you mysql
6 # 3. convert all the dates avaible in dataset to timestamp format in pandas and in sql you to cor
7 # 4 . Find out in this data that how many unique id's we have
8 # 5 . which id is one of the active id that you have in whole dataset
9 # 6 . how many of them have not logged there activity find out in terms of number of ids
10 # 7 . Find out who is the laziest person id that we have in dataset
11 # 8 . Explore over an internet that how much calories burn is required for a healthy person and j
12 # 9. how many person are not a regular person with respect to activity try to find out those
13 # 10 . who is the thired most active person in this dataset find out those in pandas and in sql b
14 # 11 . who is the 5th most laziest person avilable in dataset find it out
15 # 12 . what is a totla acumulative calories burn for a person find out
```

```
In [6]: 1 # 1. Read this dataset in pandas , mysql and mongodb
```

```
In [11]: 1 import mysql.connector as conn
2 from mysql.connector import Error
```

```
In [17]: 1 try:
2     mydb = conn.connect(host='localhost',user='root',password='123456')
3     print(mydb)
4
5     if mydb.is_connected():
6         db_Info = mydb.get_server_info()
7         print("Connected to MySQL Server version ", db_Info)
8         cursor = mydb.cursor()
9         #         cursor.execute("select database();")
10        #         record = cursor.fetchone()
11        #         print("You're connected to database: ", record)
12
13    except Error as e:
14        print("Error while connecting to MySQL", e)
15    # finally:
16    #     if mydb.is_connected():
17    #         cursor.close()
18    #         mydb.close()
19    #         print("MySQL connection is closed")
```

```
<mysql.connector.connection.MySQLConnection object at 0x000001CE983BF0A0>
Connected to MySQL Server version 8.0.26
```

```
In [9]: 1 import csvkit
```

In [16]: 1 !csvsql -i mysql FitBit_data.csv

```
CREATE TABLE `FitBit_data` (
```

C:\Users\user\anaconda3\lib\site-packages\agate\table\from_csv.py:74: RuntimeWarning: Error sniffing CSV dialect: Could not determine delimiter
sys:1: ResourceWarning: unclosed socket <zmq.Socket(zmq.PUSH) at 0x26eef86bd60>

```
    `Id` DECIMAL(38, 0) NOT NULL,  
    `ActivityDate` DATE NOT NULL,  
    `TotalSteps` DECIMAL(38, 0) NOT NULL,  
    `TotalDistance` DECIMAL(38, 17) NOT NULL,  
    `TrackerDistance` DECIMAL(38, 17) NOT NULL,  
    `LoggedActivitiesDistance` DECIMAL(38, 15) NOT NULL,  
    `VeryActiveDistance` DECIMAL(38, 17) NOT NULL,  
    `ModeratelyActiveDistance` DECIMAL(38, 16) NOT NULL,  
    `LightActiveDistance` DECIMAL(38, 17) NOT NULL,  
    `SedentaryActiveDistance` DECIMAL(38, 17) NOT NULL,  
    `VeryActiveMinutes` DECIMAL(38, 0) NOT NULL,  
    `FairlyActiveMinutes` DECIMAL(38, 0) NOT NULL,  
    `LightlyActiveMinutes` DECIMAL(38, 0) NOT NULL,  
    `SedentaryMinutes` DECIMAL(38, 0) NOT NULL,  
    `Calories` DECIMAL(38, 0) NOT NULL
```

```
);
```

C:\Users\user\anaconda3\lib\threading.py:874: ResourceWarning: unclosed file <_io.BufferedWriter name=4>

C:\Users\user\anaconda3\lib\threading.py:874: ResourceWarning: unclosed file <_io.BufferedReader name=5>

C:\Users\user\anaconda3\lib\threading.py:874: ResourceWarning: unclosed file <_io.BufferedReader name=6>

sys:1: ResourceWarning: unclosed socket <zmq.Socket(zmq.PUSH) at 0x26eef87c040>

```
In [14]: 1 try:  
2         cursor.execute("Create database FBit")  
3         print("database created successfully")  
4     except conn.Error as error:  
5         print("Failed to create Database in MySQL: {}".format(error))  
6     finally:  
7         cursor.execute("Use FBit")
```

Failed to create Database in MySQL: 1007 (HY000): Can't create database 'fbit'; database exists

```
In [15]: 1 t = """CREATE TABLE `FitBit_data` (  
2  
3     `Id` DECIMAL NOT NULL,  
4     `ActivityDate` DATE NOT NULL,  
5     `TotalSteps` DECIMAL NOT NULL,  
6     `TotalDistance` DECIMAL NOT NULL,  
7     `TrackerDistance` DECIMAL NOT NULL,  
8     `LoggedActivitiesDistance` DECIMAL NOT NULL,  
9     `VeryActiveDistance` DECIMAL NOT NULL,  
10    `ModeratelyActiveDistance` DECIMAL NOT NULL,  
11    `LightActiveDistance` DECIMAL NOT NULL,  
12    `SedentaryActiveDistance` DECIMAL NOT NULL,  
13    `VeryActiveMinutes` DECIMAL NOT NULL,  
14    `FairlyActiveMinutes` DECIMAL NOT NULL,  
15    `LightlyActiveMinutes` DECIMAL NOT NULL,  
16    `SedentaryMinutes` DECIMAL NOT NULL,  
17    `Calories` DECIMAL NOT NULL  
18 );"""
```

```
In [16]: 1 try:
2         cursor.execute(t)
3         print("Table created successfully")
4     except mysql.connector.Error as error:
5         print("Failed to create table in MySQL: {}".format(error))
6     finally:
7         if mydb.is_connected():
8             cursor.close()
9             mydb.close()
10        print("MySQL connection is closed")
```

Table created successfully
MySQL connection is closed

```
In [21]: 1 try:
2         load = """LOAD DATA INFILE "D:/18_iNeuron/00_S_Assignments/task/FitBit_data.csv"
3                 INTO table fbit.FitBit_data
4                 FIELDS TERMINATED BY ','
5                 ENCLOSED BY '"'
6                 LINES TERMINATED BY '\r\n'
7                 IGNORE 1 LINES;"""
8         cursor.execute(load)
9     except conn.Error as error:
10        print("Failed to create table in MySQL: {}".format(error))
```

```
In [3]: 1 import pandas as pd
```

```
In [4]: 1 fb = pd.read_csv("D:/18_iNeuron/00_S_Assignments/task/FitBit_data.csv")
2       fb.head()
```

```
Out[4]:
```

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | ModeratelyActiveDistance |
|---|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|--------------------------|
| 0 | 1503960366 | 2016-03-25 | 11004 | 7.11 | 7.11 | 0.0 | 2.57 | |
| 1 | 1503960366 | 2016-03-26 | 17609 | 11.55 | 11.55 | 0.0 | 6.92 | |
| 2 | 1503960366 | 2016-03-27 | 12736 | 8.53 | 8.53 | 0.0 | 4.66 | |
| 3 | 1503960366 | 2016-03-28 | 13231 | 8.93 | 8.93 | 0.0 | 3.19 | |
| 4 | 1503960366 | 2016-03-29 | 12041 | 7.85 | 7.85 | 0.0 | 2.16 | |

```
In [84]: 1 fb.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 457 entries, 0 to 456
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Id                                     457 non-null    int64
1   ActivityDate                         457 non-null    object
2   TotalSteps                           457 non-null    int64
3   TotalDistance                        457 non-null    float64
4   TrackerDistance                      457 non-null    float64
5   LoggedActivitiesDistance             457 non-null    float64
6   VeryActiveDistance                  457 non-null    float64
7   ModeratelyActiveDistance            457 non-null    float64
8   LightlyActiveDistance                457 non-null    float64
9   SedentaryActiveDistance              457 non-null    float64
10  VeryActiveMinutes                    457 non-null    int64
11  FairlyActiveMinutes                  457 non-null    int64
12  LightlyActiveMinutes                 457 non-null    int64
13  SedentaryMinutes                     457 non-null    int64
14  Calories                             457 non-null    int64
15  ActivityDate_timeFormat              457 non-null    datetime64[ns]
16  ActivityDate_timeFormat_Day          457 non-null    int64
dtypes: datetime64[ns](1), float64(7), int64(8), object(1)
memory usage: 60.8+ KB
```

```
In [25]: 1 fb.describe()
```

```
Out[25]:
```

| | Id | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | ModeratelyActiveDistance |
|-------|--------------|--------------|---------------|-----------------|--------------------------|--------------------|--------------------------|
| count | 4.570000e+02 | 457.000000 | 457.000000 | 457.000000 | 457.000000 | 457.000000 | |
| mean | 4.628595e+09 | 6546.562363 | 4.663523 | 4.609847 | 0.179427 | 1.180897 | |
| std | 2.293781e+09 | 5398.493064 | 4.082072 | 4.068540 | 0.849232 | 2.487159 | |
| min | 1.503960e+09 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 2.347168e+09 | 1988.000000 | 1.410000 | 1.280000 | 0.000000 | 0.000000 | |
| 50% | 4.057193e+09 | 5986.000000 | 4.090000 | 4.090000 | 0.000000 | 0.000000 | |
| 75% | 6.391747e+09 | 10198.000000 | 7.160000 | 7.110000 | 0.000000 | 1.310000 | |
| max | 8.877689e+09 | 28497.000000 | 27.530001 | 27.530001 | 6.727057 | 21.920000 | |

```
In [26]: 1 fb
```

```
Out[26]:
```

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | ModeratelyActiveDistance |
|-----|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|--------------------------|
| 0 | 1503960366 | 2016-03-25 | 11004 | 7.110000 | 7.110000 | | 0.0 | 2.57 |
| 1 | 1503960366 | 2016-03-26 | 17609 | 11.550000 | 11.550000 | | 0.0 | 6.92 |
| 2 | 1503960366 | 2016-03-27 | 12736 | 8.530000 | 8.530000 | | 0.0 | 4.66 |
| 3 | 1503960366 | 2016-03-28 | 13231 | 8.930000 | 8.930000 | | 0.0 | 3.19 |
| 4 | 1503960366 | 2016-03-29 | 12041 | 7.850000 | 7.850000 | | 0.0 | 2.16 |
| ... | ... | ... | ... | ... | ... | | ... | ... |
| 452 | 8877689391 | 2016-04-08 | 23014 | 20.389999 | 20.389999 | | 0.0 | 11.10 |
| 453 | 8877689391 | 2016-04-09 | 16470 | 8.070000 | 8.070000 | | 0.0 | 0.00 |
| 454 | 8877689391 | 2016-04-10 | 28497 | 27.530001 | 27.530001 | | 0.0 | 21.92 |
| 455 | 8877689391 | 2016-04-11 | 10622 | 8.060000 | 8.060000 | | 0.0 | 1.47 |
| 456 | 8877689391 | 2016-04-12 | 2350 | 1.780000 | 1.780000 | | 0.0 | 0.00 |

457 rows × 9 columns

```
In [31]: 1 import json
```

```
In [34]: 1 # try:
2 #     fb.to_json("D:/18_iNeuron/00_S_Assignments/task/fb_J.json")
3 #     print("JSON file created successfully")
4 # except Exception as e:
5 #     print(e)
```

```
In [ ]: 1 # with open('D:/18_iNeuron/00_S_Assignments/task/fb_J.jsonn') as f:
2 #     fb_J_records=json.load(f)
```

```
In [40]: 1 fb_J = fb.to_dict("records")
```

```
In [27]: 1 import pymongo
```

```
In [35]: 1 client = pymongo.MongoClient("mongodb+srv://muzammil:123456Ab@cluster0.wyo7z.mongodb.net/?retryWrites=true&ssl=true")
2 db = client.test
```

```
In [36]: 1 print(db)
```

Database(MongoClient(host=['cluster0-shard-00-01.wyo7z.mongodb.net:27017', 'cluster0-shard-00-02.wyo7z.mongodb.net:27017', 'cluster0-shard-00-00.wyo7z.mongodb.net:27017'], document_class=dict, tz_aware=False, connect=True, retrywrites=True, w='majority', authsource='admin', replicaset='atlas-ctrmxb-shard-0', tls=True), 'test')

```
In [37]: 1 fb_db = client['FitBit_Database']
        2 fb_table = fb_db['Fitbit_Table']
```

```
In [41]: 1 fb_table.insert_many(fb_J)
```

```
Out[41]: <pymongo.results.InsertManyResult at 0x1ce9cd75e20>
```

```
In [65]: 1 # 3. convert all the dates avaiable in dataset to timestamp format in pandas and in sql you to cor
        2 fb.head()
```

```
Out[65]:
```

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | ModeratelyActiveDistance |
|----------|------------|---------------------|-------------------|----------------------|------------------------|---------------------------------|---------------------------|---------------------------------|
| 0 | 1503960366 | 2016-03-25 | 11004 | 7.11 | 7.11 | 0.0 | 2.57 | |
| 1 | 1503960366 | 2016-03-26 | 17609 | 11.55 | 11.55 | 0.0 | 6.92 | |
| 2 | 1503960366 | 2016-03-27 | 12736 | 8.53 | 8.53 | 0.0 | 4.66 | |
| 3 | 1503960366 | 2016-03-28 | 13231 | 8.93 | 8.93 | 0.0 | 3.19 | |
| 4 | 1503960366 | 2016-03-29 | 12041 | 7.85 | 7.85 | 0.0 | 2.16 | |

```
In [66]: 1 fb['ActivityDate_timeFormat'] = pd.to_datetime(fb["ActivityDate"])
```

```
In [67]: 1 fb.head()
```

```
Out[67]:
```

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | ModeratelyActiveDistance |
|----------|------------|---------------------|-------------------|----------------------|------------------------|---------------------------------|---------------------------|---------------------------------|
| 0 | 1503960366 | 2016-03-25 | 11004 | 7.11 | 7.11 | 0.0 | 2.57 | |
| 1 | 1503960366 | 2016-03-26 | 17609 | 11.55 | 11.55 | 0.0 | 6.92 | |
| 2 | 1503960366 | 2016-03-27 | 12736 | 8.53 | 8.53 | 0.0 | 4.66 | |
| 3 | 1503960366 | 2016-03-28 | 13231 | 8.93 | 8.93 | 0.0 | 3.19 | |
| 4 | 1503960366 | 2016-03-29 | 12041 | 7.85 | 7.85 | 0.0 | 2.16 | |

```
In [8]: 1 fb.dtypes
```

```
Out[8]: Id                                int64
ActivityDate                             object
TotalSteps                               int64
TotalDistance                             float64
TrackerDistance                           float64
LoggedActivitiesDistance                   float64
VeryActiveDistance                        float64
ModeratelyActiveDistance                   float64
LightActiveDistance                       float64
SedentaryActiveDistance                    float64
VeryActiveMinutes                          int64
FairlyActiveMinutes                       int64
LightlyActiveMinutes                      int64
SedentaryMinutes                          int64
Calories                                  int64
ActivityDate_timeFormat                    datetime64[ns]
dtype: object
```

```
In [9]: 1 #4 Find out in this data that how many unique id's we have.  
2 fb.count()
```

```
Out[9]: Id 457  
ActivityDate 457  
TotalSteps 457  
TotalDistance 457  
TrackerDistance 457  
LoggedActivitiesDistance 457  
VeryActiveDistance 457  
ModeratelyActiveDistance 457  
LightActiveDistance 457  
SedentaryActiveDistance 457  
VeryActiveMinutes 457  
FairlyActiveMinutes 457  
LightlyActiveMinutes 457  
SedentaryMinutes 457  
Calories 457  
ActivityDate_timeFormat 457  
dtype: int64
```

```
In [15]: 1 fb['Id'].value_counts()
```

```
Out[15]: 4057192912 32  
4020332650 32  
1503960366 19  
1624580081 19  
2347167796 15  
4702921684 15  
4445114986 15  
6962181067 14  
2320127002 12  
4558609924 12  
8792009665 12  
8378563200 12  
8253242879 12  
7086361926 12  
7007744171 12  
2026352035 12  
5553957443 12  
8877689391 12  
1844505072 12  
2022484408 12  
2873212765 12  
3977333714 12  
1927972279 12  
4319703577 12  
8053475328 11  
5577150313 11  
6290855005 10  
1644430081 10  
6117666160 10  
3372868164 10  
6391747486 9  
6775888955 9  
2891001357 8  
8583815059 8  
4388161847 8  
Name: Id, dtype: int64
```

```
In [17]: 1 v = fb['Id'].value_counts()
```

```
In [18]: 1 type(v)
```

```
Out[18]: pandas.core.series.Series
```

```
In [21]: 1 fb_Uniques_ID = pd.DataFrame([v]).T
```

```
In [22]: 1 fb_Uniques_ID
```

```
Out[22]:
```

| | Id |
|------------|----|
| 4057192912 | 32 |
| 4020332650 | 32 |
| 1503960366 | 19 |
| 1624580081 | 19 |
| 2347167796 | 15 |
| 4702921684 | 15 |
| 4445114986 | 15 |
| 6962181067 | 14 |
| 2320127002 | 12 |
| 4558609924 | 12 |
| 8792009665 | 12 |
| 8378563200 | 12 |
| 8253242879 | 12 |
| 7086361926 | 12 |
| 7007744171 | 12 |
| 2026352035 | 12 |
| 5553957443 | 12 |
| 8877689391 | 12 |
| 1844505072 | 12 |
| 2022484408 | 12 |
| 2873212765 | 12 |
| 3977333714 | 12 |
| 1927972279 | 12 |
| 4319703577 | 12 |
| 8053475328 | 11 |
| 5577150313 | 11 |
| 6290855005 | 10 |
| 1644430081 | 10 |
| 6117666160 | 10 |
| 3372868164 | 10 |
| 6391747486 | 9 |
| 6775888955 | 9 |
| 2891001357 | 8 |
| 8583815059 | 8 |
| 4388161847 | 8 |

```
In [120]: 1 fb_Uniques_ID.count()
```

```
Out[120]: Id      35
dtype: int64
```

```
In [121]: 1 # 5 . which id is one of the active id that you have in whole dataset
```

```
In [33]: 1 fb1 = fb.groupby(['Id'])['TotalSteps','Calories','LightlyActiveMinutes','SedentaryMinutes'].sum()
```

C:\Users\user\AppData\Local\Temp\ipykernel_14548\2195191442.py:1: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
fb1 = fb.groupby(['Id'])['TotalSteps','Calories','LightlyActiveMinutes','SedentaryMinutes'].sum()

In [34]: 1 fb1.head()

Out[34]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|------------|------------|----------|----------------------|------------------|
| Id | | | | |
| 1503960366 | 221170 | 34128 | 4336 | 15387 |
| 1624580081 | 80299 | 25705 | 2300 | 24288 |
| 1644430081 | 92748 | 29164 | 2277 | 10342 |
| 1844505072 | 43687 | 19391 | 1902 | 12424 |
| 1927972279 | 26170 | 27048 | 1340 | 11435 |

In [110]: 1 (fb.groupby(['Id'])['TotalSteps', 'Calories', 'SedentaryMinutes', 'LightlyActiveMinutes'].sum()).max()

C:\Users\user\AppData\Local\Temp\ipykernel_22836\2390117707.py:1: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
(fb.groupby(['Id'])['TotalSteps', 'Calories', 'SedentaryMinutes', 'LightlyActiveMinutes'].sum()).max()

Out[110]:

| | |
|----------------------|--------|
| TotalSteps | 221170 |
| Calories | 98412 |
| SedentaryMinutes | 43458 |
| LightlyActiveMinutes | 4336 |

dtype: int64

In [111]: 1 # Active ID based on Total Steps
2 fb1.loc[fb1['TotalSteps'] == 221170]

Out[111]:

| | TotalSteps | Calories | LightlyActiveMinutes |
|------------|------------|----------|----------------------|
| Id | | | |
| 1503960366 | 221170 | 34128 | 4336 |

In [113]: 1 # Active ID based on Calories
2 fb1.loc[fb1['Calories'] == 98412]

Out[113]:

| | TotalSteps | Calories | LightlyActiveMinutes |
|------------|------------|----------|----------------------|
| Id | | | |
| 4020332650 | 184851 | 98412 | 4172 |

In [32]: 1 # Active ID based on LightlyActiveMinutes
2 fb1.loc[fb1['LightlyActiveMinutes']==4336]

Out[32]:

| | TotalSteps | Calories | LightlyActiveMinutes |
|------------|------------|----------|----------------------|
| Id | | | |
| 1503960366 | 221170 | 34128 | 4336 |

In [50]: 1 result = fb.groupby('Id')['TotalSteps'].aggregate(['min', 'max']).max()
2 print(result)

min 7413
max 28497
dtype: int64

In [117]: 1
2 fb.loc[fb['TotalSteps'] == 7413]

Out[117]:

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | Mod |
|-----|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|-----|
| 377 | 7007744171 | 2016-04-12 | 7413 | 5.77 | 4.96 | 4.869783 | 4.5 | |


```
In [118]: 1 fb.loc[fb['TotalSteps'] == 28497]
```

Out[118]:

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | Mod |
|-----|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|-------|
| 454 | 8877689391 | 2016-04-10 | 28497 | 27.530001 | 27.530001 | | 0.0 | 21.92 |

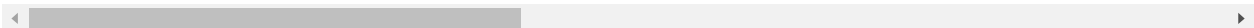


```
In [98]: 1 fb
```

Out[98]:

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | Mod |
|-----|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|-------|
| 0 | 1503960366 | 2016-03-25 | 11004 | 7.110000 | 7.110000 | | 0.0 | 2.57 |
| 1 | 1503960366 | 2016-03-26 | 17609 | 11.550000 | 11.550000 | | 0.0 | 6.92 |
| 2 | 1503960366 | 2016-03-27 | 12736 | 8.530000 | 8.530000 | | 0.0 | 4.66 |
| 3 | 1503960366 | 2016-03-28 | 13231 | 8.930000 | 8.930000 | | 0.0 | 3.19 |
| 4 | 1503960366 | 2016-03-29 | 12041 | 7.850000 | 7.850000 | | 0.0 | 2.16 |
| ... | ... | ... | ... | ... | ... | | ... | ... |
| 452 | 8877689391 | 2016-04-08 | 23014 | 20.389999 | 20.389999 | | 0.0 | 11.10 |
| 453 | 8877689391 | 2016-04-09 | 16470 | 8.070000 | 8.070000 | | 0.0 | 0.00 |
| 454 | 8877689391 | 2016-04-10 | 28497 | 27.530001 | 27.530001 | | 0.0 | 21.92 |
| 455 | 8877689391 | 2016-04-11 | 10622 | 8.060000 | 8.060000 | | 0.0 | 1.47 |
| 456 | 8877689391 | 2016-04-12 | 2350 | 1.780000 | 1.780000 | | 0.0 | 0.00 |

457 rows × 9 columns



```
In [122]: 1 # 6 . how many of them have not Logged there activity find out in terms of number of ids
```

```
In [123]: 1 fb.isna().sum()
```

Out[123]:

| | |
|--------------------------|-------|
| Id | 0 |
| ActivityDate | 0 |
| TotalSteps | 0 |
| TotalDistance | 0 |
| TrackerDistance | 0 |
| LoggedActivitiesDistance | 0 |
| VeryActiveDistance | 0 |
| ModeratelyActiveDistance | 0 |
| LightActiveDistance | 0 |
| SedentaryActiveDistance | 0 |
| VeryActiveMinutes | 0 |
| FairlyActiveMinutes | 0 |
| LightlyActiveMinutes | 0 |
| SedentaryMinutes | 0 |
| Calories | 0 |
| ActivityDate_timeFormat | 0 |
| dtype: | int64 |

```
In [15]: 1 (fb == 0).all(axis=0)
```

```
Out[15]: Id False
ActivityDate False
TotalSteps False
TotalDistance False
TrackerDistance False
LoggedActivitiesDistance False
VeryActiveDistance False
ModeratelyActiveDistance False
LightActiveDistance False
SedentaryActiveDistance False
VeryActiveMinutes False
FairlyActiveMinutes False
LightlyActiveMinutes False
SedentaryMinutes False
Calories False
dtype: bool
```

```
In [13]: 1 fb.loc[fb['LoggedActivitiesDistance'] == 0]
```

```
Out[13]:
```

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | Mod |
|-----|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|-----|
| 0 | 1503960366 | 2016-03-25 | 11004 | 7.110000 | 7.110000 | 0.0 | 2.57 | |
| 1 | 1503960366 | 2016-03-26 | 17609 | 11.550000 | 11.550000 | 0.0 | 6.92 | |
| 2 | 1503960366 | 2016-03-27 | 12736 | 8.530000 | 8.530000 | 0.0 | 4.66 | |
| 3 | 1503960366 | 2016-03-28 | 13231 | 8.930000 | 8.930000 | 0.0 | 3.19 | |
| 4 | 1503960366 | 2016-03-29 | 12041 | 7.850000 | 7.850000 | 0.0 | 2.16 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 452 | 8877689391 | 2016-04-08 | 23014 | 20.389999 | 20.389999 | 0.0 | 11.10 | |
| 453 | 8877689391 | 2016-04-09 | 16470 | 8.070000 | 8.070000 | 0.0 | 0.00 | |
| 454 | 8877689391 | 2016-04-10 | 28497 | 27.530001 | 27.530001 | 0.0 | 21.92 | |
| 455 | 8877689391 | 2016-04-11 | 10622 | 8.060000 | 8.060000 | 0.0 | 1.47 | |
| 456 | 8877689391 | 2016-04-12 | 2350 | 1.780000 | 1.780000 | 0.0 | 0.00 | |

433 rows × 15 columns

```
In [25]: 1 fb.loc[(fb['LoggedActivitiesDistance'] == 0) & (fb['VeryActiveDistance'] == 0) & (fb['ModeratelyActiveDistance'] == 0) & (fb['LightActiveDistance'] == 0) & (fb['SedentaryActiveDistance'] == 0) & (fb['VeryActiveMinutes'] == 0) & (fb['FairlyActiveMinutes'] == 0) & (fb['LightlyActiveMinutes'] == 0)]
```

```
Out[25]:
```

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | Mod |
|-----|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|-----|
| 51 | 1844505072 | 2016-04-04 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 52 | 1844505072 | 2016-04-05 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 59 | 1844505072 | 2016-04-12 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 96 | 2320127002 | 2016-04-01 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 97 | 2320127002 | 2016-04-02 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 402 | 8253242879 | 2016-04-02 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 403 | 8253242879 | 2016-04-03 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 404 | 8253242879 | 2016-04-04 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 405 | 8253242879 | 2016-04-05 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 412 | 8253242879 | 2016-04-12 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | |

68 rows × 15 columns

```
In [ ]: 1 df3.dropna(thresh=3)
```

```
In [20]: 1 fb.groupby("Id")['LoggedActivitiesDistance']==0
```

Out[20]: False

```
In [22]: 1 fb[["LoggedActivitiesDistance", "VeryActiveDistance", 'ModeratelyActiveDistance', "LightActiveDistance", "SedentaryActiveDistance", "VerySedentaryActiveDistance"]]
```

Out[22]:

| | LoggedActivitiesDistance | VeryActiveDistance | ModeratelyActiveDistance | LightActiveDistance | SedentaryActiveDistance | VerySedentaryActiveDistance |
|-----|--------------------------|--------------------|--------------------------|---------------------|-------------------------|-----------------------------|
| 0 | 0.0 | 2.57 | 0.46 | 4.07 | 0.00 | 0.00 |
| 1 | 0.0 | 6.92 | 0.73 | 3.91 | 0.00 | 0.00 |
| 2 | 0.0 | 4.66 | 0.16 | 3.71 | 0.00 | 0.00 |
| 3 | 0.0 | 3.19 | 0.79 | 4.95 | 0.00 | 0.00 |
| 4 | 0.0 | 2.16 | 1.09 | 4.61 | 0.00 | 0.00 |
| ... | ... | ... | ... | ... | ... | ... |
| 452 | 0.0 | 11.10 | 0.63 | 8.62 | 0.00 | 0.00 |
| 453 | 0.0 | 0.00 | 0.02 | 8.02 | 0.00 | 0.00 |
| 454 | 0.0 | 21.92 | 1.12 | 4.46 | 0.00 | 0.00 |
| 455 | 0.0 | 1.47 | 0.15 | 6.37 | 0.01 | 0.00 |
| 456 | 0.0 | 0.00 | 0.00 | 1.78 | 0.00 | 0.00 |

457 rows × 8 columns

```
In [35]: 1 # 7 Find out who is the laziest person id that we have in dataset
2 fb2 = fb.groupby(['Id'])['TotalSteps', 'Calories', 'LightlyActiveMinutes', 'SedentaryMinutes'].sum()
```

C:\Users\user\AppData\Local\Temp\ipykernel_14548\2398762924.py:2: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
fb2 = fb.groupby(['Id'])['TotalSteps', 'Calories', 'LightlyActiveMinutes', 'SedentaryMinutes'].sum()

```
In [38]: 1 fb2.head()
```

Out[38]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|------------|------------|----------|----------------------|------------------|
| Id | | | | |
| 1503960366 | 221170 | 34128 | 4336 | 15387 |
| 1624580081 | 80299 | 25705 | 2300 | 24288 |
| 1644430081 | 92748 | 29164 | 2277 | 10342 |
| 1844505072 | 43687 | 19391 | 1902 | 12424 |
| 1927972279 | 26170 | 27048 | 1340 | 11435 |

```
In [42]: 1 fb2.agg([min,max])
```

Out[42]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|-----|------------|----------|----------------------|------------------|
| min | 0 | 14442 | 0 | 7293 |
| max | 221170 | 98412 | 4336 | 43458 |

```
In [43]: 1 # Laziest person based on SedentaryMinutes
2 fb2.loc[fb2['SedentaryMinutes']==43458]
```

Out[43]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|------------|------------|----------|----------------------|------------------|
| Id | | | | |
| 4057192912 | 60391 | 60913 | 1384 | 43458 |

```
In [44]: 1 # laziest person based on TotalSteps
        2 fb2.loc[fb2['TotalSteps']==0]
```

```
Out[44]:
```

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|--|------------|----------|----------------------|------------------|
| | Id | | | |
| | 4388161847 | 0 | 14442 | 0 |
| | | | | 11074 |

```
In [46]: 1 # laziest person based on Calories
        2 fb2.loc[fb2['Calories']==14442]
```

```
Out[46]:
```

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|--|------------|----------|----------------------|------------------|
| | Id | | | |
| | 4388161847 | 0 | 14442 | 0 |
| | | | | 11074 |

```
In [47]: 1 # 8 .Explore over an internet that how much calories burn is required for a healthy person and fi
        2 fb2.loc[fb2['Calories']>2250]
```

Out[47]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|--|------------|----------|----------------------|------------------|
|--|------------|----------|----------------------|------------------|

| Id | | | | |
|------------|--------|-------|------|-------|
| 1503960366 | 221170 | 34128 | 4336 | 15387 |
| 1624580081 | 80299 | 25705 | 2300 | 24288 |
| 1644430081 | 92748 | 29164 | 2277 | 10342 |
| 1844505072 | 43687 | 19391 | 1902 | 12424 |
| 1927972279 | 26170 | 27048 | 1340 | 11435 |
| 2022484408 | 146099 | 29704 | 3046 | 12707 |
| 2026352035 | 40713 | 16266 | 2032 | 7910 |
| 2320127002 | 37661 | 18385 | 1511 | 14991 |
| 2347167796 | 147001 | 30320 | 3803 | 10258 |
| 2873212765 | 79639 | 20349 | 3303 | 13643 |
| 2891001357 | 6189 | 18187 | 1350 | 8799 |
| 3372868164 | 61275 | 18603 | 2869 | 10851 |
| 3977333714 | 103967 | 16777 | 2483 | 8505 |
| 4020332650 | 184851 | 98412 | 4172 | 34663 |
| 4057192912 | 60391 | 60913 | 1384 | 43458 |
| 4319703577 | 93847 | 23931 | 2984 | 9361 |
| 4388161847 | 0 | 14442 | 0 | 11074 |
| 4445114986 | 64396 | 31617 | 3009 | 12921 |
| 4558609924 | 69423 | 21962 | 2998 | 13024 |
| 4702921684 | 119151 | 42322 | 3701 | 10901 |
| 5553957443 | 100257 | 21630 | 2200 | 7293 |
| 5577150313 | 94690 | 36303 | 1730 | 7324 |
| 6117666160 | 82486 | 20993 | 2999 | 8641 |
| 6290855005 | 16181 | 21656 | 393 | 12896 |
| 6391747486 | 12032 | 15868 | 309 | 11357 |
| 6775888955 | 50031 | 24523 | 1116 | 9161 |
| 6962181067 | 176956 | 29249 | 3620 | 8488 |
| 7007744171 | 147124 | 31525 | 3330 | 12008 |
| 7086361926 | 73247 | 26124 | 1215 | 10666 |
| 8053475328 | 163288 | 31825 | 1698 | 12259 |
| 8253242879 | 28679 | 17556 | 711 | 16228 |
| 8378563200 | 97623 | 40274 | 2027 | 8210 |
| 8583815059 | 24364 | 19130 | 1097 | 10094 |
| 8792009665 | 37139 | 24884 | 1585 | 10723 |
| 8877689391 | 209005 | 41414 | 2892 | 12554 |

```
In [ ]: 1 # 9. how many person are not a regular person with respect to activity try to find out those
```

In [68]:

```
1 fb.head()
```

Out[68]:

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | ModerateDistance |
|---|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|------------------|
| 0 | 1503960366 | 2016-03-25 | 11004 | 7.11 | 7.11 | 0.0 | 2.57 | 0.0 |
| 1 | 1503960366 | 2016-03-26 | 17609 | 11.55 | 11.55 | 0.0 | 6.92 | 0.0 |
| 2 | 1503960366 | 2016-03-27 | 12736 | 8.53 | 8.53 | 0.0 | 4.66 | 0.0 |
| 3 | 1503960366 | 2016-03-28 | 13231 | 8.93 | 8.93 | 0.0 | 3.19 | 0.0 |
| 4 | 1503960366 | 2016-03-29 | 12041 | 7.85 | 7.85 | 0.0 | 2.16 | 0.0 |

In [71]:

```
1 fb['ActivityDate_timeFormat_Day'] = fb['ActivityDate_timeFormat'].dt.day
```

In [72]:

```
1 fb.head()
```

Out[72]:

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | ModerateDistance |
|---|------------|--------------|------------|---------------|-----------------|--------------------------|--------------------|------------------|
| 0 | 1503960366 | 2016-03-25 | 11004 | 7.11 | 7.11 | 0.0 | 2.57 | 0.0 |
| 1 | 1503960366 | 2016-03-26 | 17609 | 11.55 | 11.55 | 0.0 | 6.92 | 0.0 |
| 2 | 1503960366 | 2016-03-27 | 12736 | 8.53 | 8.53 | 0.0 | 4.66 | 0.0 |
| 3 | 1503960366 | 2016-03-28 | 13231 | 8.93 | 8.93 | 0.0 | 3.19 | 0.0 |
| 4 | 1503960366 | 2016-03-29 | 12041 | 7.85 | 7.85 | 0.0 | 2.16 | 0.0 |

In [82]:

```
1 fb.groupby(['Id'])['ActivityDate_timeFormat_Day'].count()
```

Out[82]:

```
Id
1503960366    19
1624580081    19
1644430081    10
1844505072    12
1927972279    12
2022484408    12
2026352035    12
2320127002    12
2347167796    15
2873212765    12
2891001357     8
3372868164    10
3977333714    12
4020332650    32
4057192912    32
4319703577    12
4388161847     8
4445114986    15
4558609924    12
4702921684    15
5553957443    12
5577150313    11
6117666160    10
6290855005    10
6391747486     9
6775888955     9
6962181067    14
7007744171    12
7086361926    12
8053475328    11
8253242879    12
8378563200    12
8583815059     8
8792009665    12
8877689391    12
Name: ActivityDate_timeFormat_Day, dtype: int64
```

```
In [83]: 1 fb.count()
```

```
Out[83]: Id 457
ActivityDate 457
TotalSteps 457
TotalDistance 457
TrackerDistance 457
LoggedActivitiesDistance 457
VeryActiveDistance 457
ModeratelyActiveDistance 457
LightActiveDistance 457
SedentaryActiveDistance 457
VeryActiveMinutes 457
FairlyActiveMinutes 457
LightlyActiveMinutes 457
SedentaryMinutes 457
Calories 457
ActivityDate_timeFormat 457
ActivityDate_timeFormat_Day 457
dtype: int64
```

```
In [ ]: 1 # 10 . who is the thired most active person in this dataset find out those in pandas and in sql k
```

```
In [48]: 1 (fb.groupby(['Id'])['TotalSteps', 'Calories', 'SedentaryMinutes', 'LightlyActiveMinutes'].sum()).max

C:\Users\user\AppData\Local\Temp\ipykernel_14548\2390117707.py:1: FutureWarning: Indexing with multi
ple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
(fb.groupby(['Id'])['TotalSteps', 'Calories', 'SedentaryMinutes', 'LightlyActiveMinutes'].sum()).max
()
```

```
Out[48]: TotalSteps 221170
Calories 98412
SedentaryMinutes 43458
LightlyActiveMinutes 4336
dtype: int64
```

```
In [49]: 1 fb1.aggregate([min,max])
```

```
Out[49]:
```

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|-----|------------|----------|----------------------|------------------|
| min | 0 | 14442 | 0 | 7293 |
| max | 221170 | 98412 | 4336 | 43458 |

```
In [54]: 1 # thired most active person w.r.t Calories
2 fb1['Calories'].nlargest(n=3)
```

```
Out[54]: Id
4020332650 98412
4057192912 60913
4702921684 42322
Name: Calories, dtype: int64
```

```
In [55]: 1 # thired most active person w.r.t TotalSteps
2 fb1['TotalSteps'].nlargest(n=3)
```

```
Out[55]: Id
1503960366 221170
8877689391 209005
4020332650 184851
Name: TotalSteps, dtype: int64
```

```
In [ ]: 1 # 11 . who is the 5th most laziest person available in dataset find it out
```

```
In [56]: 1 fb2.head()
```

Out[56]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|------------|------------|----------|----------------------|------------------|
| Id | | | | |
| 1503960366 | 221170 | 34128 | 4336 | 15387 |
| 1624580081 | 80299 | 25705 | 2300 | 24288 |
| 1644430081 | 92748 | 29164 | 2277 | 10342 |
| 1844505072 | 43687 | 19391 | 1902 | 12424 |
| 1927972279 | 26170 | 27048 | 1340 | 11435 |

```
In [57]: 1 fb2.nsmallest(5, 'TotalSteps')
```

Out[57]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|------------|------------|----------|----------------------|------------------|
| Id | | | | |
| 4388161847 | 0 | 14442 | 0 | 11074 |
| 2891001357 | 6189 | 18187 | 1350 | 8799 |
| 6391747486 | 12032 | 15868 | 309 | 11357 |
| 6290855005 | 16181 | 21656 | 393 | 12896 |
| 8583815059 | 24364 | 19130 | 1097 | 10094 |

```
In [58]: 1 fb2.nsmallest(5, 'Calories')
```

Out[58]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|------------|------------|----------|----------------------|------------------|
| Id | | | | |
| 4388161847 | 0 | 14442 | 0 | 11074 |
| 6391747486 | 12032 | 15868 | 309 | 11357 |
| 2026352035 | 40713 | 16266 | 2032 | 7910 |
| 3977333714 | 103967 | 16777 | 2483 | 8505 |
| 8253242879 | 28679 | 17556 | 711 | 16228 |

In [62]:

```
1 # 12 . what is a totla acumulative calories burn for a person find out
2 fb2
```

Out[62]:

| | TotalSteps | Calories | LightlyActiveMinutes | SedentaryMinutes |
|------------|------------|----------|----------------------|------------------|
| Id | | | | |
| 1503960366 | 221170 | 34128 | 4336 | 15387 |
| 1624580081 | 80299 | 25705 | 2300 | 24288 |
| 1644430081 | 92748 | 29164 | 2277 | 10342 |
| 1844505072 | 43687 | 19391 | 1902 | 12424 |
| 1927972279 | 26170 | 27048 | 1340 | 11435 |
| 2022484408 | 146099 | 29704 | 3046 | 12707 |
| 2026352035 | 40713 | 16266 | 2032 | 7910 |
| 2320127002 | 37661 | 18385 | 1511 | 14991 |
| 2347167796 | 147001 | 30320 | 3803 | 10258 |
| 2873212765 | 79639 | 20349 | 3303 | 13643 |
| 2891001357 | 6189 | 18187 | 1350 | 8799 |
| 3372868164 | 61275 | 18603 | 2869 | 10851 |
| 3977333714 | 103967 | 16777 | 2483 | 8505 |
| 4020332650 | 184851 | 98412 | 4172 | 34663 |
| 4057192912 | 60391 | 60913 | 1384 | 43458 |
| 4319703577 | 93847 | 23931 | 2984 | 9361 |
| 4388161847 | 0 | 14442 | 0 | 11074 |
| 4445114986 | 64396 | 31617 | 3009 | 12921 |
| 4558609924 | 69423 | 21962 | 2998 | 13024 |
| 4702921684 | 119151 | 42322 | 3701 | 10901 |
| 5553957443 | 100257 | 21630 | 2200 | 7293 |
| 5577150313 | 94690 | 36303 | 1730 | 7324 |
| 6117666160 | 82486 | 20993 | 2999 | 8641 |
| 6290855005 | 16181 | 21656 | 393 | 12896 |
| 6391747486 | 12032 | 15868 | 309 | 11357 |
| 6775888955 | 50031 | 24523 | 1116 | 9161 |
| 6962181067 | 176956 | 29249 | 3620 | 8488 |
| 7007744171 | 147124 | 31525 | 3330 | 12008 |
| 7086361926 | 73247 | 26124 | 1215 | 10666 |
| 8053475328 | 163288 | 31825 | 1698 | 12259 |
| 8253242879 | 28679 | 17556 | 711 | 16228 |
| 8378563200 | 97623 | 40274 | 2027 | 8210 |
| 8583815059 | 24364 | 19130 | 1097 | 10094 |
| 8792009665 | 37139 | 24884 | 1585 | 10723 |
| 8877689391 | 209005 | 41414 | 2892 | 12554 |