In [2]:
```
import pandas as pd
import seaborn as sns
import numpy as np
import datetime as dt
import matplotlib.pyplot as plt
import re
```

In [8]:
```
df = pd.read_csv("C:/Users/PCC/Downloads/HealthApp_2k.log_structured.csv")
```

In [9]:
```
templates = pd.read_csv("C:/Users/PCC/Downloads/HealthApp_2k.log_templates.csv")
```

In [12]:
```
df
```

Out[12]:

	LineId	Time	Component	Pid	Content	EventId	EventTemplate
0	1	20171223-22:15:29:606	Step_LSC	30002312	onStandStepChanged 3579	E42	onStandStepChanged <*>
1	2	20171223-22:15:29:615	Step_LSC	30002312	onExtend:1514038530000 14 0 4	E39	onExtend:<*> <*> <*> <*>
2	3	20171223-22:15:29:633	Step_StandReportReceiver	30002312	onReceive action: android.intent.action.SCREEN_ON	E41	onReceive action: android.intent.action.SCREEN_ON
3	4	20171223-22:15:29:635	Step_LSC	30002312	processHandleBroadcastAction action:android.in...	E43	processHandleBroadcastAction action:android.in...
4	5	20171223-22:15:29:635	Step_StandStepCounter	30002312	flush sensor data	E12	flush sensor data
...	...	...	...	...	...	...	...
1995	1996	20171224-0:58:53:985	Step_LSC	30002312	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1996	1997	20171224-0:59:7:581	Step_LSC	30002312	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1997	1998	20171224-1:0:0:794	Step_LSC	30002312	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1998	1999	20171224-1:1:0:935	Step_LSC	30002312	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1999	2000	20171224-1:2:35:789	Step_LSC	30002312	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...

2000 rows × 7 columns

In [13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 7 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   LineId         2000 non-null   int64 
 1   Time           2000 non-null   object
 2   Component      2000 non-null   object
 3   Pid            2000 non-null   int64 
 4   Content        2000 non-null   object
 5   EventId        2000 non-null   object
 6   EventTemplate  2000 non-null   object
dtypes: int64(2), object(5)
memory usage: 109.5+ KB
```

In [58]:
```
# Check for missing values and decide how to handle them. You can either remove rows with missing values or impute
# them with appropriate values (mean, median, mode, etc.). To check for missing values:

df.isnull()
```

Out[58]:

	LineId	Time	Component	Content	EventId	EventTemplate
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
1995	False	False	False	False	False	False
1996	False	False	False	False	False	False
1997	False	False	False	False	False	False
1998	False	False	False	False	False	False
1999	False	False	False	False	False	False

In [54]:
```
df.notnull()
```

Out[54]:

	LineId	Time	Component	Content	EventId	EventTemplate
0	True	True	True	True	True	True
1	True	True	True	True	True	True
2	True	True	True	True	True	True
3	True	True	True	True	True	True
4	True	True	True	True	True	True
...	...	...	...	...	...	...
1995	True	True	True	True	True	True
1996	True	True	True	True	True	True
1997	True	True	True	True	True	True
1998	True	True	True	True	True	True
1999	True	True	True	True	True	True

2000 rows × 6 columns

In [56]:
```
df.shape
```

Out[56]:
```
(2000, 6)
```

In [15]:
```
df.head(5)
```

Out[15]:

	LineId	Time	Component	Pid	Content	EventId	EventTemplate
0	1	20171223-22:15:29:606	Step_LSC	30002312	onStandStepChanged 3579	E42	onStandStepChanged <*>
1	2	20171223-22:15:29:615	Step_LSC	30002312	onExtend:1514038530000 14 0 4	E39	onExtend:<*> <*> <*> <*>
2	3	20171223-22:15:29:633	Step_StandReportReceiver	30002312	onReceive action: android.intent.action.SCREEN_ON	E41	onReceive action: android.intent.action.SCREEN_ON
3	4	20171223-22:15:29:635	Step_LSC	30002312	processHandleBroadcastAction action:android.in...	E43	processHandleBroadcastAction action:android.in...
4	5	20171223-22:15:29:635	Step_StandStepCounter	30002312	flush sensor data	E12	flush sensor data

In [57]:
```
df.tail
```

Out[57]:

	LineId	Time	Component	Content	EventId	EventTemplate
1995	1996	20171224-0:58:53:985	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1996	1997	20171224-0:59:7:581	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1997	1998	20171224-1:0:0:794	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1998	1999	20171224-1:1:0:935	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1999	2000	20171224-1:2:35:789	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...

In [17]:
```
df.describe()
```

Out[17]:

	LineId	Pid
count	2000.000000	2000.0
mean	1000.500000	30002312.0
std	577.494589	0.0
min	1.000000	30002312.0
25%	500.750000	30002312.0
50%	1000.500000	30002312.0
75%	1500.250000	30002312.0
max	2000.000000	30002312.0

In [25]:
```
df.count()
```

Out[25]:
```
LineId          2000
Time            2000
Component       2000
Pid             2000
Content         2000
EventId         2000
EventTemplate   2000
dtype: int64
```

In [32]:
```
df.min()
```

Out[32]:
```
LineId                                                     1
Time                                      20171223-22:15:29:606
Component                                                 HiH_
Pid                                                   30002312
Content                        getDiffTotalSteps= 1513958400215##0
EventId                                                    E1
EventTemplate    Alarm uploadStaticsToDB totalSteps=<*>:<*>:<*>...
dtype: object
```

In [34]:
```
df.head(5)
```

Out[34]:

	LineId	Time	Component	Pid	Content	EventId	EventTemplate
0	1	20171223-22:15:29:606	Step_LSC	30002312	onStandStepChanged 3579	E42	onStandStepChanged <*>
1	2	20171223-22:15:29:615	Step_LSC	30002312	onExtend:1514038530000 14 0 4	E39	onExtend:<*> <*> <*> <*>
2	3	20171223-22:15:29:633	Step_StandReportReceiver	30002312	onReceive action: android.intent.action.SCREEN_ON	E41	onReceive action: android.intent.action.SCREEN_ON
3	4	20171223-22:15:29:635	Step_LSC	30002312	processHandleBroadcastAction action:android.in...	E43	processHandleBroadcastAction action:android.in...
4	5	20171223-22:15:29:635	Step_StandStepCounter	30002312	flush sensor data	E12	flush sensor data

In [36]:
```
df = df.drop('Pid', axis=1)
```

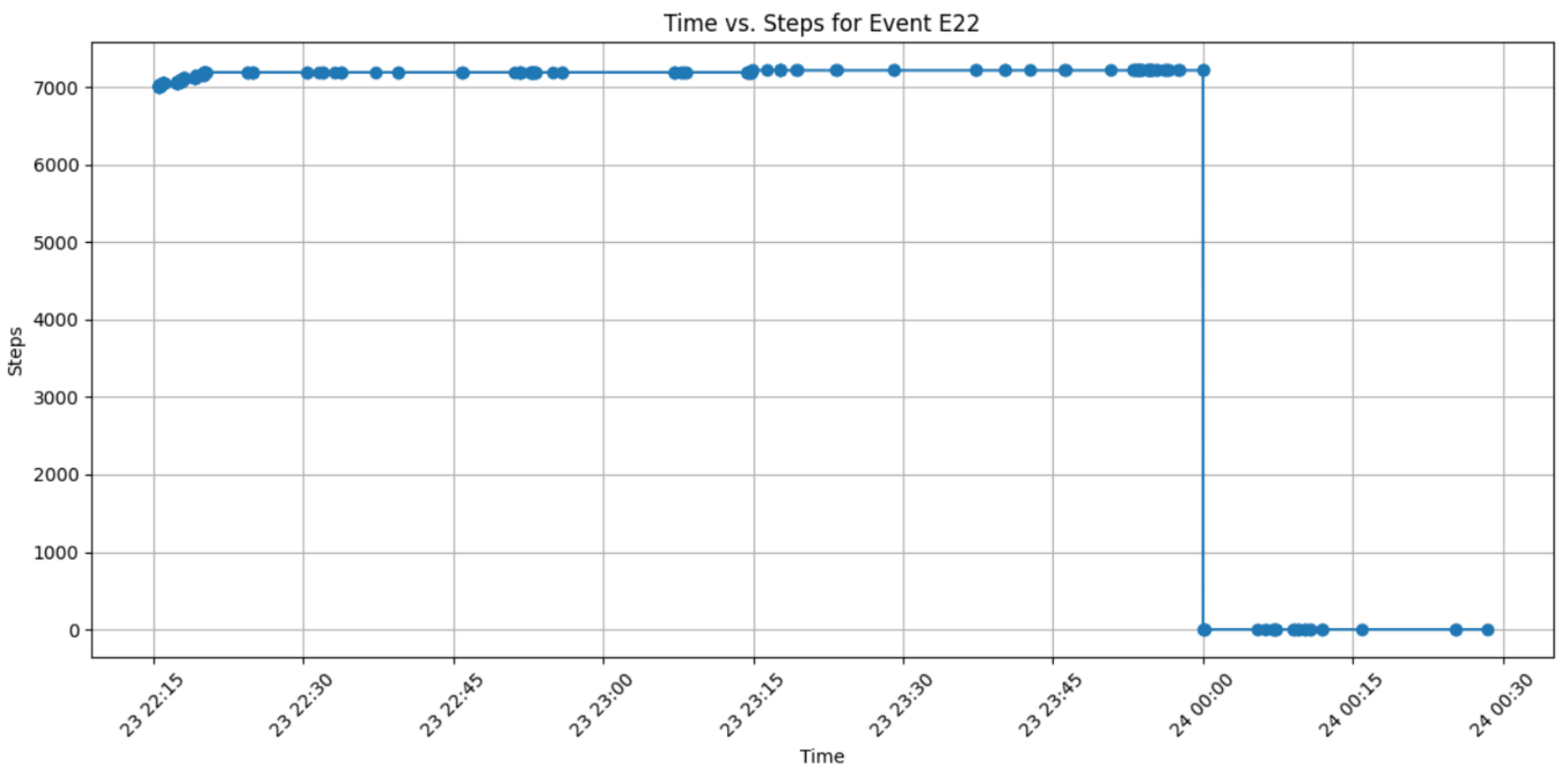
In [37]:
```
df
```

Out[37]:

	LineId	Time	Component	Content	EventId	EventTemplate
0	1	20171223-22:15:29:606	Step_LSC	onStandStepChanged 3579	E42	onStandStepChanged <*>
1	2	20171223-22:15:29:615	Step_LSC	onExtend:1514038530000 14 0 4	E39	onExtend:<*> <*> <*> <*>
2	3	20171223-22:15:29:633	Step_StandReportReceiver	onReceive action: android.intent.action.SCREEN_ON	E41	onReceive action: android.intent.action.SCREEN_ON
3	4	20171223-22:15:29:635	Step_LSC	processHandleBroadcastAction action:android.in...	E43	processHandleBroadcastAction action:android.in...
4	5	20171223-22:15:29:635	Step_StandStepCounter	flush sensor data	E12	flush sensor data
...	...	...	...	...	...	...
1995	1996	20171224-0:58:53:985	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1996	1997	20171224-0:59:7:581	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...

In [39]:
```
dfs23 = df[df['EventId'] == 'E22']
dfs23['Time'] = pd.to_datetime(dfs23['Time'], format='%Y%m%d-%H:%M:%S:%f')
dfs23['Steps'] = dfs23['Content'].str.extract(r'##(\d+)##')
dfs23['Steps'] = dfs23['Steps'].astype(int)
plt.figure(figsize=(12, 6))
plt.plot(dfs23['Time'], dfs23['Steps'], marker='o', linestyle='-')
plt.title('Time vs. Steps for Event E22')
plt.xlabel('Time')
plt.ylabel('Steps')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

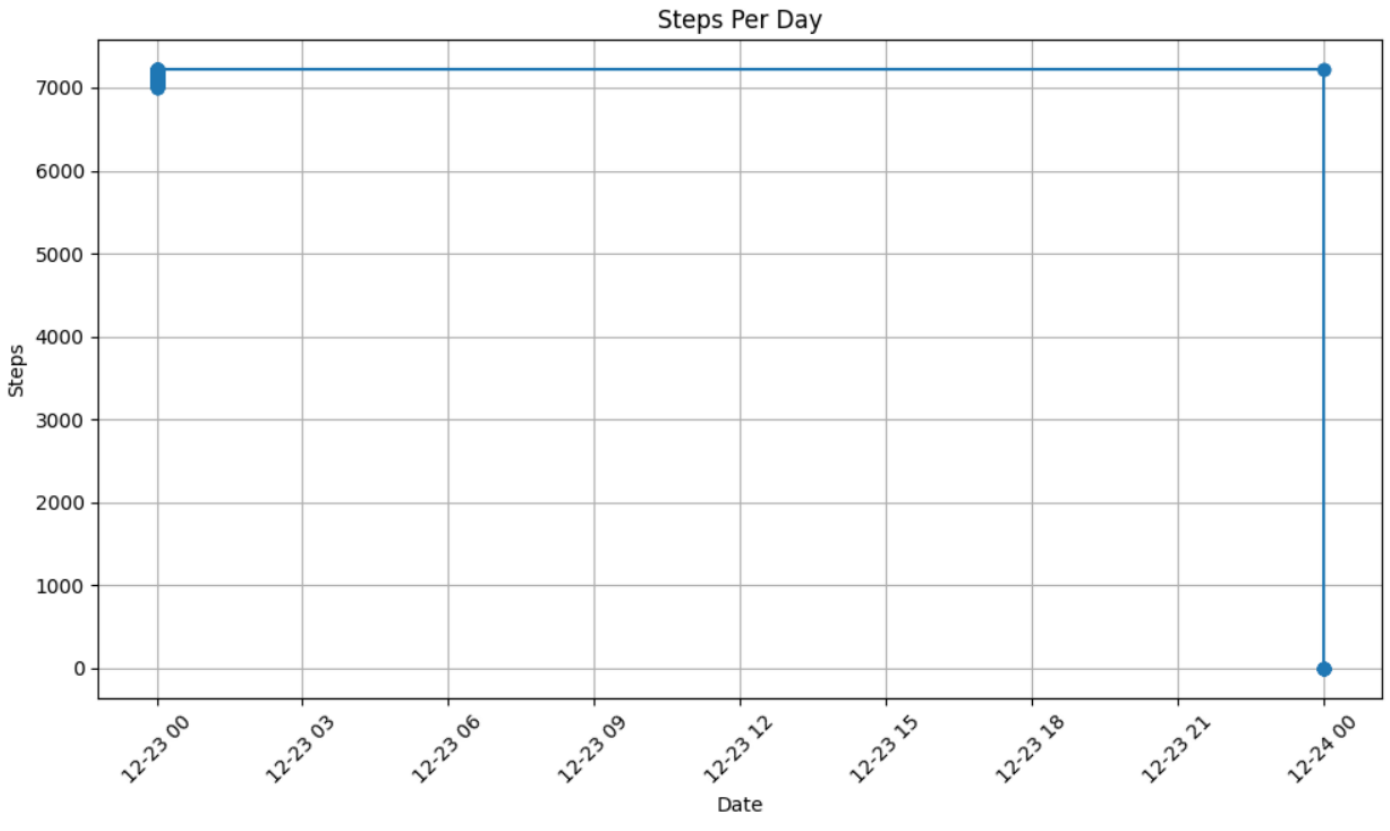
```
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\2267965645.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfs23['Time'] = pd.to_datetime(dfs23['Time'], format='%Y%m%d-%H:%M:%S:%f')
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\2267965645.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfs23['Steps'] = dfs23['Content'].str.extract(r'##(\d+)##')
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\2267965645.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfs23['Steps'] = dfs23['Steps'].astype(int)
```



In [41]:
```
dfs23['Time'] = pd.to_datetime(dfs23['Time'], format='%Y%m%d-%H:%M:%S:%f')

dfs23['Date'] = dfs23['Time'].dt.date

print(dfs23[['Date', 'Steps']])

plt.figure(figsize=(10, 6))
plt.plot(dfs23['Date'], dfs23['Steps'], marker='o')
plt.title('Steps Per Day')
plt.xlabel('Date')
plt.ylabel('Steps')
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()
plt.show()
```

```
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\2098986252.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfs23['Time'] = pd.to_datetime(dfs23['Time'], format='%Y%m%d-%H:%M:%S:%f')
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\2098986252.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfs23['Date'] = dfs23['Time'].dt.date
```

```
           Date  Steps
5    2017-12-23   6993
15   2017-12-23   7007
22   2017-12-23   7008
29   2017-12-23   7009
36   2017-12-23   7011
...         ...    ...
1925 2017-12-24      0
1935 2017-12-24      0
1951 2017-12-24      0
1957 2017-12-24      0
1966 2017-12-24      0

[242 rows x 2 columns]
```




```
dfs23
```

	LineId	Time	Component	Content	EventId	EventTemplate	Steps	Date
5	6	2017-12-23 22:15:29.635	Step_SPUtils	getTodayTotalDetailSteps = 1514038440000##699...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	6993	2017-12-23
15	16	2017-12-23 22:15:29.950	Step_SPUtils	getTodayTotalDetailSteps = 1514038440000##700...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	7007	2017-12-23
22	23	2017-12-23 22:15:30.632	Step_SPUtils	getTodayTotalDetailSteps = 1514038440000##700...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	7008	2017-12-23
29	30	2017-12-23 22:15:31.142	Step_SPUtils	getTodayTotalDetailSteps = 1514038440000##700...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	7009	2017-12-23
36	37	2017-12-23 22:15:32.145	Step_SPUtils	getTodayTotalDetailSteps = 1514038440000##701...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	7011	2017-12-23
...	...	...	...	...	...	...	...	...
1925	1926	2017-12-24 00:11:57.442	Step_SPUtils	getTodayTotalDetailSteps = 1514045400000##0##...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	0	2017-12-24
1935	1936	2017-12-24 00:15:53.285	Step_SPUtils	getTodayTotalDetailSteps = 1514045400000##0##...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	0	2017-12-24
1951	1952	2017-12-24 00:25:16.438	Step_SPUtils	getTodayTotalDetailSteps = 1514045640000##0##...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	0	2017-12-24
1957	1958	2017-12-24 00:25:17.438	Step_SPUtils	getTodayTotalDetailSteps = 1514046240000##0##...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	0	2017-12-24
1966	1967	2017-12-24 00:28:25.453	Step_SPUtils	getTodayTotalDetailSteps = 1514046240000##0##...	E22	getTodayTotalDetailSteps = <*>##<*>##<*>##<*>#...	0	2017-12-24

242 rows × 8 columns

```
dfs23['Date'] = dfs23['Time'].dt.date
dfs23['Hour'] = dfs23['Time'].dt.hour
dfs23['Minute'] = dfs23['Time'].dt.minute

steps_increase_threshold = 1000
min_duration_minutes = 30

workout_sessions = []
current_session = None

for index, row in dfs23.iterrows():
    if row['Steps'] > steps_increase_threshold:
        if current_session is None:
            current_session = {'start_time': row['Time']}
        else:
            current_session['end_time'] = row['Time']
    else:
        if current_session is not None:
            duration = (current_session['end_time'] - current_session['start_time']).total_seconds() / 60
            if duration >= min_duration_minutes:
                workout_sessions.append(current_session)
            current_session = None

if workout_sessions:
    print("Workout sessions:")
    for session in workout_sessions:
        print(f"Start: {session['start_time']} - End: {session['end_time']}")
```

```
Workout sessions:
Start: 2017-12-23 22:15:29.635000 - End: 2017-12-24 00:00:00.234000
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\4098350854.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfs23['Date'] = dfs23['Time'].dt.date
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\4098350854.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfs23['Hour'] = dfs23['Time'].dt.hour
```

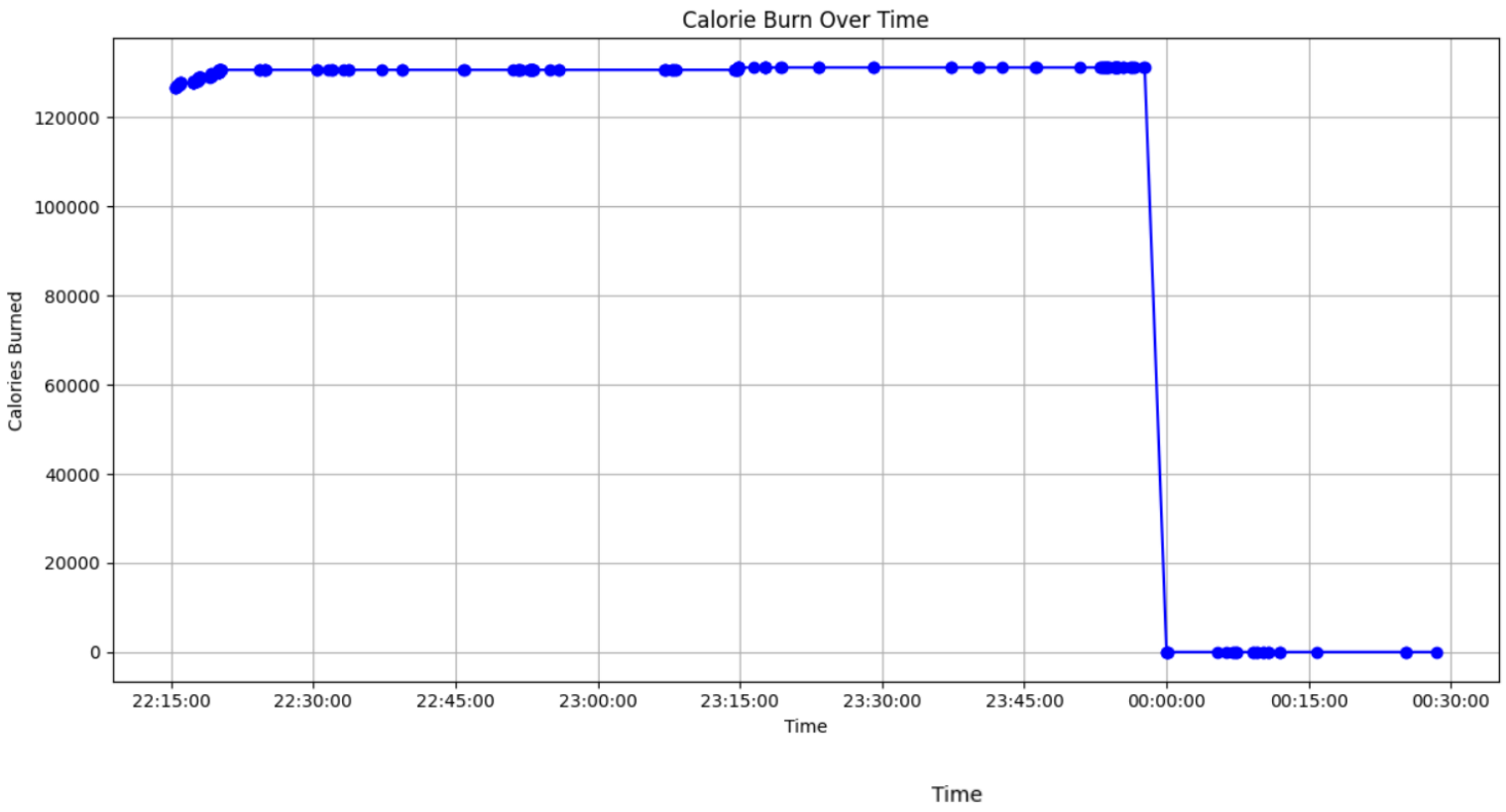
```
dfsd24 = df[df['EventId'] == 'E4']
dfsd24['Calories'] = dfsd24['Content'].str.extract(r'(\d+)')
dfsd24['Calories'] = dfsd24['Calories'].astype(int)
dfsd24.head()
```

```
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\3852311565.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfsd24['Calories'] = dfsd24['Content'].str.extract(r'(\d+)')
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\3852311565.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfsd24['Calories'] = dfsd24['Calories'].astype(int)
```

	LineId	Time	Component	Content	EventId	EventTemplate	Calories
8	9	20171223-22:15:29:645	Step_ExtSDM	calculateCaloriesWithCache totalCalories=126775	E4	calculateCaloriesWithCache totalCalories=<*>	126775
17	18	20171223-22:15:29:959	Step_ExtSDM	calculateCaloriesWithCache totalCalories=126797	E4	calculateCaloriesWithCache totalCalories=<*>	126797
24	25	20171223-22:15:30:639	Step_ExtSDM	calculateCaloriesWithCache totalCalories=126818	E4	calculateCaloriesWithCache totalCalories=<*>	126818
31	32	20171223-22:15:31:157	Step_ExtSDM	calculateCaloriesWithCache totalCalories=126861	E4	calculateCaloriesWithCache totalCalories=<*>	126861
38	39	20171223-22:15:32:156	Step_ExtSDM	calculateCaloriesWithCache totalCalories=126882	E4	calculateCaloriesWithCache totalCalories=<*>	126882

```
dfsd24['Time'] = pd.to_datetime(dfsd24['Time'], format='%Y%m%d-%H:%M:%S:%f')

plt.figure(figsize=(12, 6))
plt.plot(dfsd24['Time'], dfsd24['Calories'], marker='o', linestyle='-', color='b')
plt.title('Calorie Burn Over Time')
plt.xlabel('Time')
plt.ylabel('Calories Burned')
plt.grid(True)
plt.tight_layout()

plt.gca().xaxis.set_major_formatter(plt.matplotlib.dates.DateFormatter('%H:%M:%S'))

plt.show()
```

```
C:\Users\PCC\AppData\Local\Temp\ipykernel_14684\2321828199.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfsd24['Time'] = pd.to_datetime(dfsd24['Time'], format='%Y%m%d-%H:%M:%S:%f')
```



.]:
```
# To remove rows with missing values:
df.dropna()
```

.]:

	LineId	Time	Component	Content	EventId	EventTemplate
0	1	20171223-22:15:29:606	Step_LSC	onStandStepChanged 3579	E42	onStandStepChanged <*>
1	2	20171223-22:15:29:615	Step_LSC	onExtend:1514038530000 14 0 4	E39	onExtend:<*> <*> <*> <*>
2	3	20171223-22:15:29:633	Step_StandReportReceiver	onReceive action: android.intent.action.SCREEN_ON	E41	onReceive action: android.intent.action.SCREEN_ON
3	4	20171223-22:15:29:635	Step_LSC	processHandleBroadcastAction action:android.in...	E43	processHandleBroadcastAction action:android.in...
4	5	20171223-22:15:29:635	Step_StandStepCounter	flush sensor data	E12	flush sensor data
...	...	...	...	...	...	...
1995	1996	20171224-0:58:53:985	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1996	1997	20171224-0:59:7:581	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1997	1998	20171224-1:0:0:794	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1998	1999	20171224-1:1:0:935	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...
1999	2000	20171224-1:2:35:789	Step_LSC	processHandleBroadcastAction action:android.in...	E44	processHandleBroadcastAction action:android.in...

2000 rows × 6 columns

]:
```
# Check for null values
df.isnull().values.any()
```

]: False

Conclusions

From the given observations app loads detailed steps data when event E22 occures. On further analysis (refer to cell 3) the following can be concluded: on 23-12-2017(from 10:15PM to 23:57PM) person's total step count = 7214 on 24-12-2017(from 00:00AM to 00:29AM) person's total step count = 0 Person's average number of steps per day = 23898 (approx) Workout sessions (Refer to cell 4): Start: 2017-12-23 22:15:29.635000 - End: 2017-12-24 00:00:00.234000 On 23-12-2017 after 11:14PM there is no change in steps or there isn't any movements so the person might be resting or put their phone on rest

Calorie Counter

From the given observations app loads detailed calorie data when event E4 occures. On further analysis (refer to cell 7, 8) the following can be concluded: on 23-12-2017(from 10:15PM to 23:57PM) person's total(cumulative) calories = 131208 on 24-12-2017(from 00:00AM to 00:29AM) person's total calories = 0

Confirmations

On 23-12-2017 there is no any major change in steps and calories from 11:14 PM, this confirms that person is resting or put thier phone on rest