

Capstone Project - Fitbit Consumer Behavior Analysis

Content

This dataset was generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Individual reports can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents the use of different Fitbit trackers and individual tracking behaviors/preferences.

This dataset contains 18 different files like dailyActivity, dailyCalories, hourlySteps, etc...

Business Task:

Analyze FitBit Fitness Tracker App data to gain insights into how consumers use the FitBit app and discover trends and insights for the marketing team.

Business Objectives:

- What are the trends identified?
- How could these trends apply to customers?
- How could these trends help influence marketing strategy?

Data Preparation:

In the initial stages of the project, I commenced by thoroughly analyzing each of the original files containing data on calorie intensity at daily, hourly, and minute levels. I started the data cleaning process by utilizing Google Colab's features. This required looking over the datasets for broad information, closely identifying any cases where values were missing, and standardizing column formats throughout all files. Duplicates were identified and removed in order to preserve data correctness, and any missing values were carefully filled in using filling processes. This painstaking cleaning process created a solid foundation for further analysis, guaranteeing that the datasets were clean and ready for in-depth investigation.

Data after cleaning:

Daily Dataset

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
2	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActive	VeryActive	Moderate	LightActive	Sedentary	VeryActive	FairlyActive	LightlyActive	Sedentary	Calories
2	1.5E+09	#####	13162	8.5	8.5	0	1.88	0.55	6.06	0	25	13	328	728	1985
3	1.5E+09	#####	10735	6.97	6.97	0	1.57	0.69	4.71	0	21	19	217	776	1797
4	1.5E+09	#####	10480	6.74	6.74	0	2.44	0.4	3.91	0	30	11	181	1218	1776
5	1.5E+09	#####	9762	6.28	6.28	0	2.14	1.26	2.83	0	29	34	209	726	1745
6	1.5E+09	#####	12669	8.16	8.16	0	2.71	0.41	5.04	0	36	10	221	773	1863
7	1.5E+09	#####	9705	6.48	6.48	0	3.19	0.78	2.51	0	38	20	164	539	1728
8	1.5E+09	#####	13019	8.59	8.59	0	3.25	0.64	4.71	0	42	16	233	1149	1921
9	1.5E+09	#####	15506	9.88	9.88	0	3.53	1.32	5.03	0	50	31	264	775	2035
10	1.5E+09	#####	10544	6.68	6.68	0	1.96	0.48	4.24	0	28	12	205	818	1786
11	1.5E+09	#####	9819	6.34	6.34	0	1.34	0.35	4.65	0	19	8	211	838	1775
12	1.5E+09	#####	12764	8.13	8.13	0	4.76	1.12	2.24	0	66	27	130	1217	1827
13	1.5E+09	#####	14371	9.04	9.04	0	2.81	0.87	5.36	0	41	21	262	732	1949
14	1.5E+09	#####	10039	6.41	6.41	0	2.92	0.21	3.28	0	39	5	238	709	1788
15	1.5E+09	#####	15355	9.8	9.8	0	5.29	0.57	3.94	0	73	14	216	814	2013
16	1.5E+09	#####	13755	8.79	8.79	0	2.33	0.92	5.54	0	31	23	279	833	1970
17	1.5E+09	#####	18134	12.21	12.21	0	6.4	0.41	5.41	0	78	11	243	1108	2159
18	1.5E+09	#####	13154	8.53	8.53	0	3.54	1.16	3.79	0	48	28	189	782	1898
19	1.5E+09	#####	11181	7.15	7.15	0	1.06	0.5	5.58	0	16	12	243	815	1837
20	1.5E+09	#####	14673	9.25	9.25	0	3.56	1.42	4.27	0	52	34	217	712	1947
21	1.5E+09	5/1/2016	10602	6.81	6.81	0	2.29	1.6	2.92	0	33	35	246	730	1820
22	1.5E+09	5/2/2016	14727	9.71	9.71	0	3.21	0.57	5.92	0	41	15	277	798	2004
23	1.5E+09	5/3/2016	15103	9.66	9.66	0	3.73	1.05	4.88	0	50	24	254	816	1990
24	1.5E+09	5/4/2016	11100	7.15	7.15	0	2.46	0.87	3.82	0	36	22	203	1179	1819
25	1.5E+09	5/5/2016	14070	8.9	8.9	0	2.92	1.08	4.88	0	45	24	250	857	1959
26	1.5E+09	5/6/2016	12159	8.03	8.03	0	1.97	0.25	5.81	0	24	6	289	754	1896
27	1.5E+09	5/7/2016	11992	7.71	7.71	0	2.46	2.12	3.13	0	37	46	175	833	1821

Hourly Dataset:

1	A	B	C	D	E	F
2	Id	ActivityHour	Calories	TotalIntensity	AverageIntensity	StepTotal
2	1.5E+09	#####	81	20	0.333333	373
3	1.5E+09	#####	61	8	0.133333	160
4	1.5E+09	#####	59	7	0.116667	151
5	1.5E+09	#####	47	0	0	0
6	1.5E+09	#####	48	0	0	0
7	1.5E+09	#####	48	0	0	0
8	1.5E+09	#####	48	0	0	0
9	1.5E+09	#####	47	0	0	0
10	1.5E+09	#####	68	13	0.216667	250
11	1.5E+09	#####	141	30	0.5	1864
12	1.5E+09	#####	99	29	0.483333	676
13	1.5E+09	#####	76	12	0.2	360
14	1.5E+09	#####	73	11	0.183333	253
15	1.5E+09	#####	66	6	0.1	221
16	1.5E+09	#####	110	36	0.6	1166
17	1.5E+09	#####	151	58	0.966667	2063
18	1.5E+09	#####	76	13	0.216667	344
19	1.5E+09	#####	83	16	0.266667	489
20	1.5E+09	#####	124	29	0.483333	1386
21	1.5E+09	#####	104	39	0.65	558
22	1.5E+09	#####	132	41	0.683333	1733
23	1.5E+09	#####	100	31	0.516667	684
24	1.5E+09	#####	65	9	0.15	89
25	1.5E+09	#####	81	21	0.35	338
26	1.5E+09	#####	69	14	0.233333	144
27	1.5E+09	#####	48	0	0	0

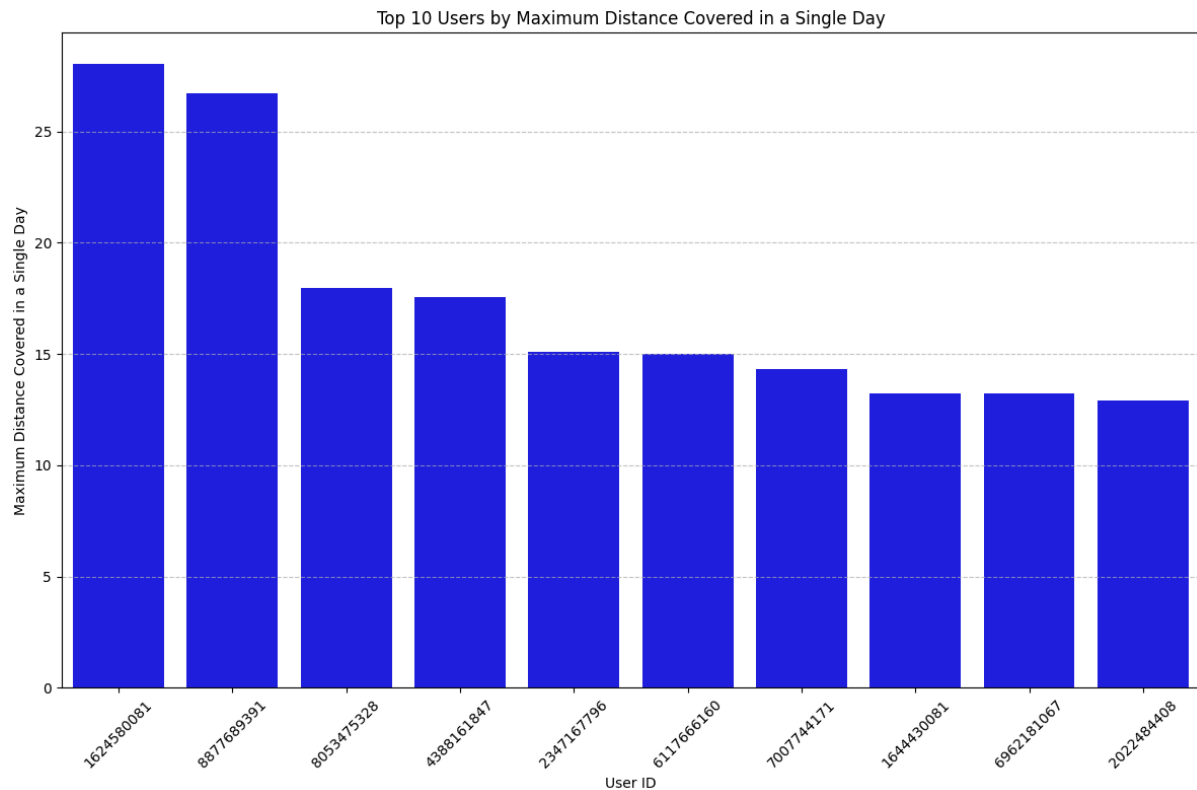
Minute Dataset:

1	A	B	C	D	E	F	G	H	I
2	id	activitydate	activitytime	calories	intensity	METs	value	logid	Steps
2	1.5E+09	#####	0:00:00	0.7865	0	10	3	1.14E+10	0
3	1.5E+09	#####	0:01:00	0.7865	0	10	2	1.14E+10	0
4	1.5E+09	#####	0:02:00	0.7865	0	10	1	1.14E+10	0
5	1.5E+09	#####	0:03:00	0.7865	0	10	1	1.14E+10	0
6	1.5E+09	#####	0:04:00	0.7865	0	10	1	1.14E+10	0
7	1.5E+09	#####	0:05:00	0.9438	0	12	1	1.14E+10	0
8	1.5E+09	#####	0:06:00	0.9438	0	12	1	1.14E+10	0
9	1.5E+09	#####	0:07:00	0.9438	0	12	2	1.14E+10	0
10	1.5E+09	#####	0:08:00	0.9438	0	12	2	1.14E+10	0
11	1.5E+09	#####	0:09:00	0.9438	0	12	2	1.14E+10	0
12	1.5E+09	#####	0:10:00	0.9438	0	12	3	1.14E+10	0
13	1.5E+09	#####	0:11:00	0.9438	0	12	3	1.14E+10	0
14	1.5E+09	#####	0:12:00	0.7865	0	10	3	1.14E+10	0
15	1.5E+09	#####	0:13:00	0.7865	0	10	3	1.14E+10	0
16	1.5E+09	#####	0:14:00	0.9438	0	12	3	1.14E+10	0
17	1.5E+09	#####	0:15:00	0.7865	0	10	2	1.14E+10	0
18	1.5E+09	#####	0:16:00	0.9438	0	12	1	1.14E+10	0
19	1.5E+09	#####	0:17:00	0.7865	0	10	1	1.14E+10	0
20	1.5E+09	#####	0:18:00	0.7865	0	10	1	1.14E+10	0
21	1.5E+09	#####	0:19:00	0.7865	0	10	1	1.14E+10	0
22	1.5E+09	#####	0:20:00	0.9438	0	12	1	1.14E+10	0
23	1.5E+09	#####	0:21:00	0.9438	0	12	1	1.14E+10	0
24	1.5E+09	#####	0:22:00	0.9438	0	12	1	1.14E+10	0
25	1.5E+09	#####	0:23:00	0.9438	0	12	1	1.14E+10	0
26	1.5E+09	#####	0:24:00	0.9438	0	12	1	1.14E+10	0
27	1.5E+09	#####	0:25:00	2.0449	1	26	1	1.14E+10	7

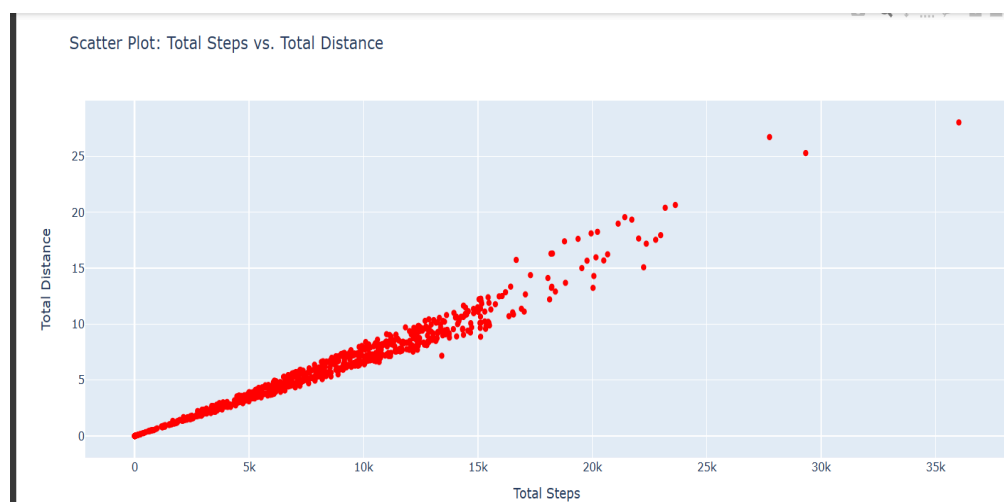
Exploratory Data Analysis:

Daily Dataset :

In this I analyze activity data to identify the top 10 users based on the maximum distance covered in a single day. The provided dataset contains information about users' activities, including their ID, activity date, and total distance covered. By aggregating this data and visualizing it using a bar plot,I was able to identify the users who have achieved the highest maximum distance covered in a single day.Based on data it seems user 1624580081 has coverd maximum steps.

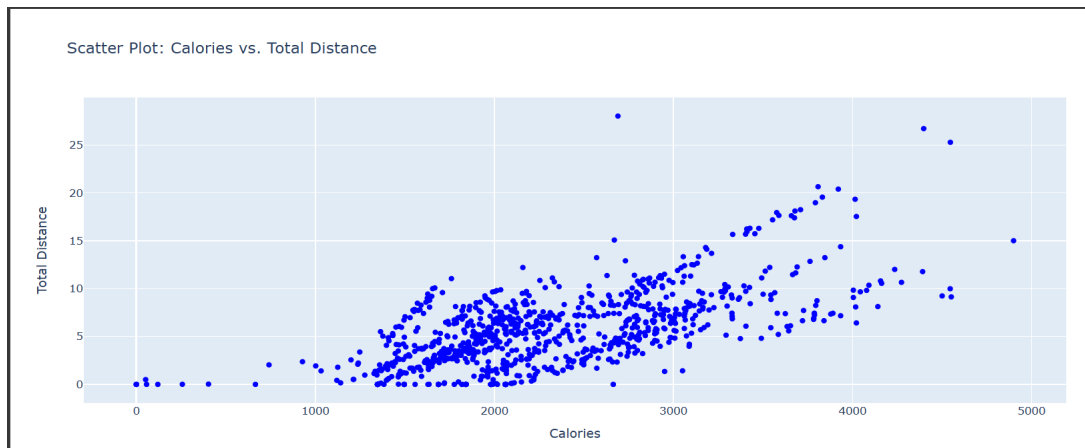


The first scatter plot examines the relationship between total steps and total distance covered.The x-axis represents the total number of steps (TotalSteps), and the y-axis represents the total distance covered (TotalDistance).

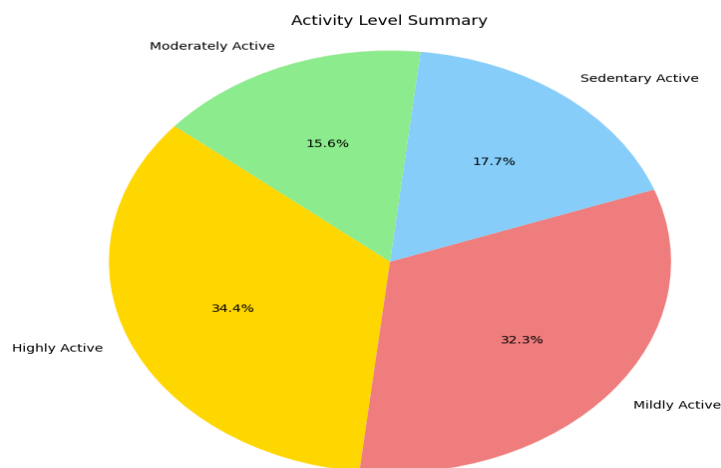


The second scatter plot explores the relationship between calories burned and total distance covered. In this plot, the x-axis represents the number of calories burned (Calories), while the y-axis represents the total distance covered (TotalDistance).

Conclusion: By visualizing these relationships through scatter plots, users can gain insights into how variables such as total steps, calories burned, and total distance covered relate to each other.



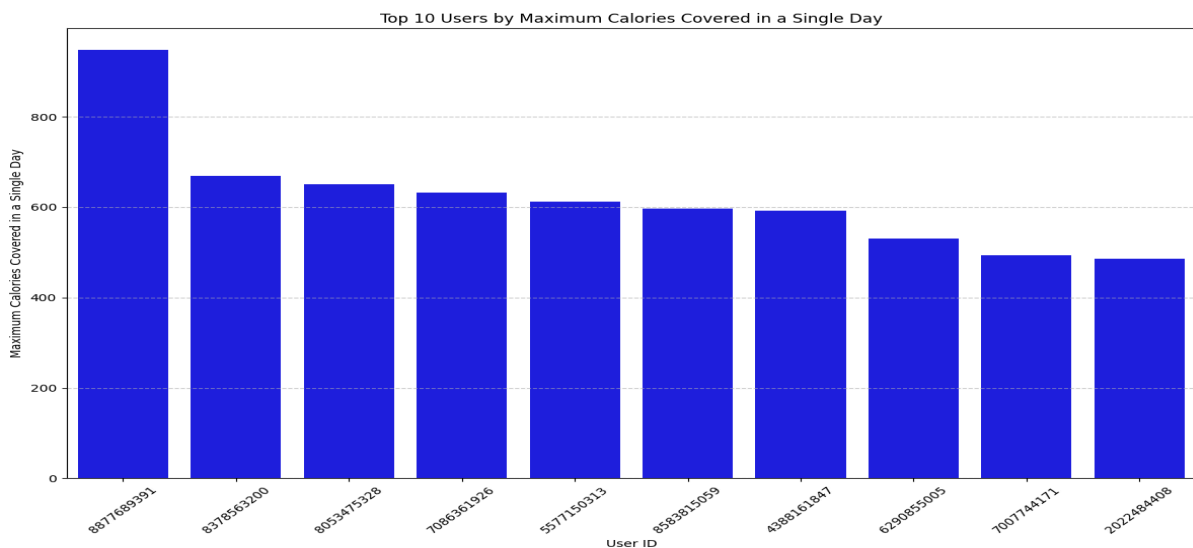
It categorizes users into different activity levels based on their activity metrics such as distances covered and active minutes spent. The summary is presented as a pie chart, with each slice representing a specific activity level. Activity levels include Highly Active, Moderately Active, Mildly Active, Sedentary Active, and Inactive. The chart offers insights into the distribution of users across these activity levels, aiding in understanding overall activity patterns within the dataset.



Full Google colab file: [Link](#)

Hourly Dataset :

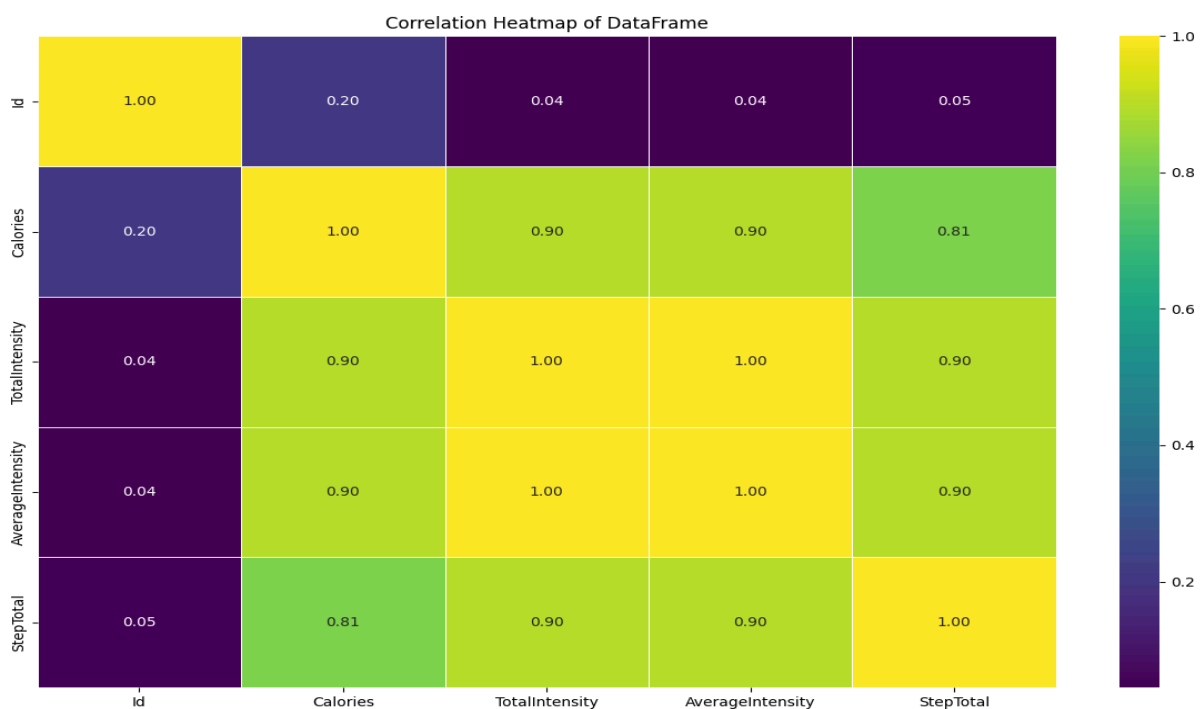
In this I visualize top 10 users based on the maximum Calories intake in a single day. The provided dataset contains information about users' activities, including their ID, activityhour, and Calories covered. By aggregating this data and visualizing it using a bar plot,I was able to identify the users who have achieved the highest maximum Calories covered in a single day.Based on data it seems user 8877689391 has coverd maximum Calories.



A correlation heatmap is generated to visualize the pairwise correlations between selected activity-related metrics within the dataset. The heatmap represents the correlation coefficients between variables, with higher values indicating stronger correlations.

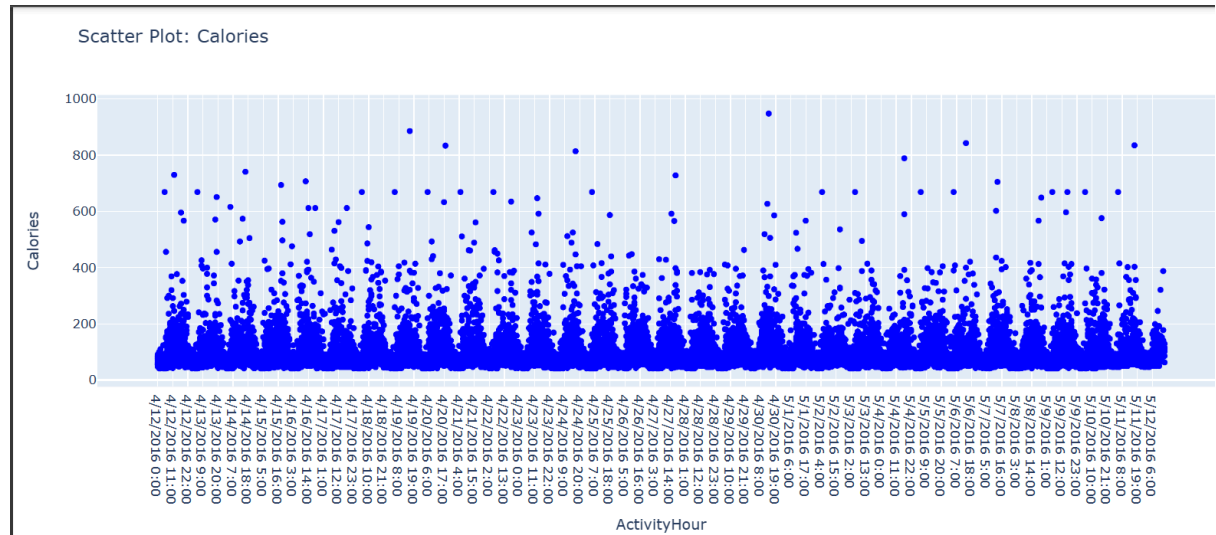
Variables included in the heatmap:

ActivityHour: Time of activity Calories: Number of calories burned TotalIntensity: Overall intensity of activity AverageIntensity: Average intensity of activity StepTotal: Total number of steps taken.

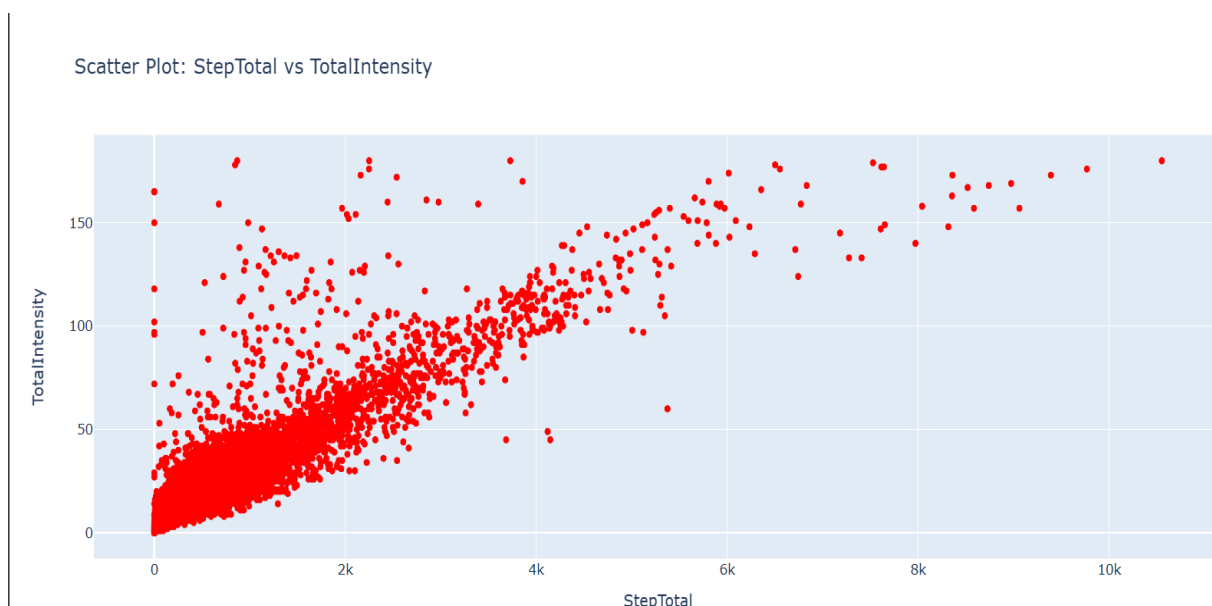


Three scatter plots are generated to visualize relationships between various activity-related metrics:

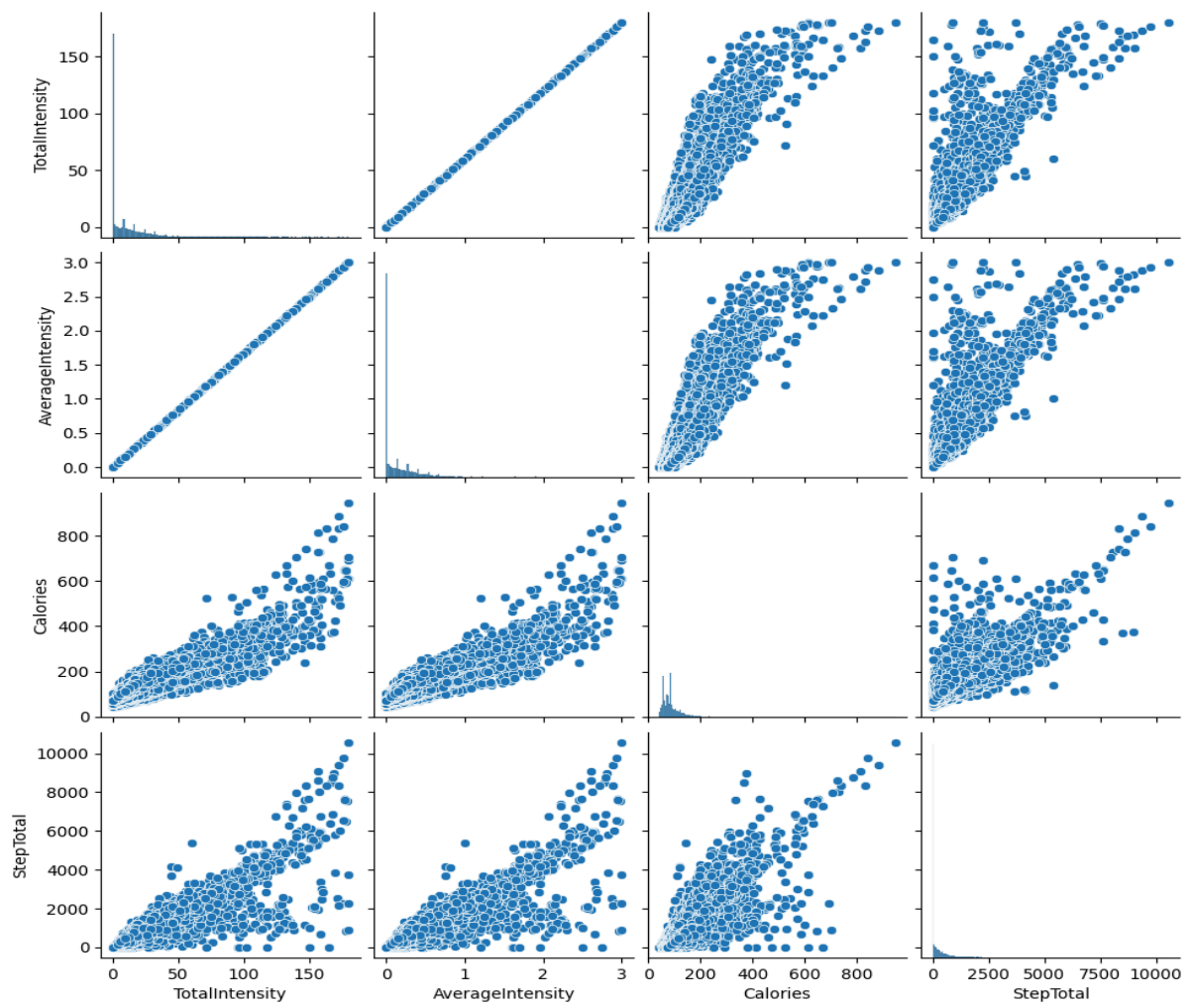
Calories vs. Activity Hour: This scatter plot depicts the relationship between the time of activity (Activity Hour) and the number of calories burned (Calories). Each point represents a data instance, with the color indicating different data points.



Step Total vs. Total Intensity: This scatter plot displays the association between the total number of steps taken (Step Total) and the overall intensity of activity (Total Intensity). Each point in the plot corresponds to a data instance, with the color denoting different data points. These scatter plots provide visual insights into the relationships between activity metrics, aiding in understanding patterns and correlations within the dataset.



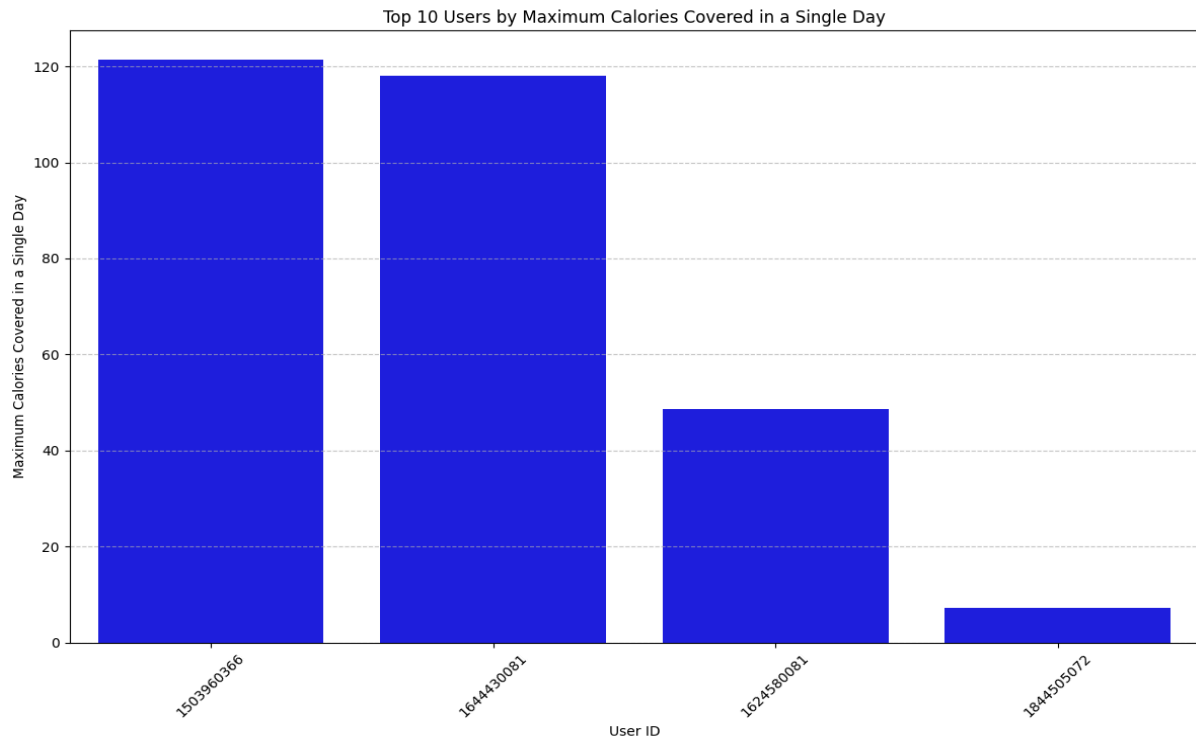
Pairplot:



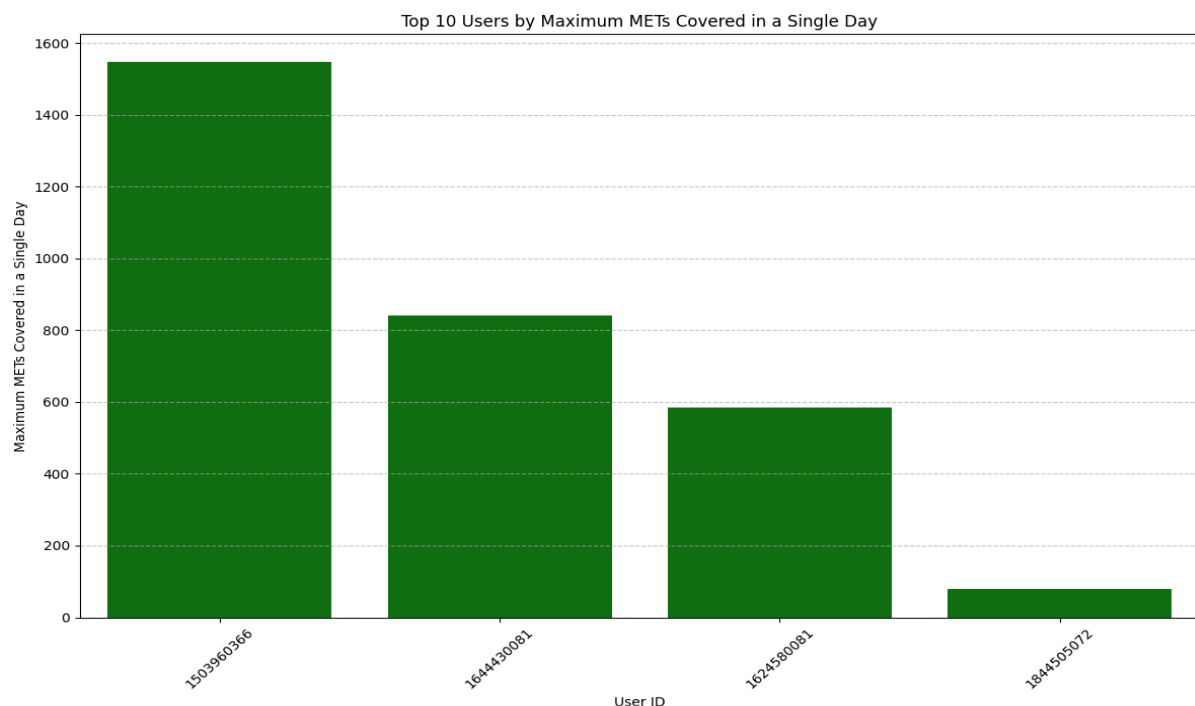
Full google colab file: [Link](#)

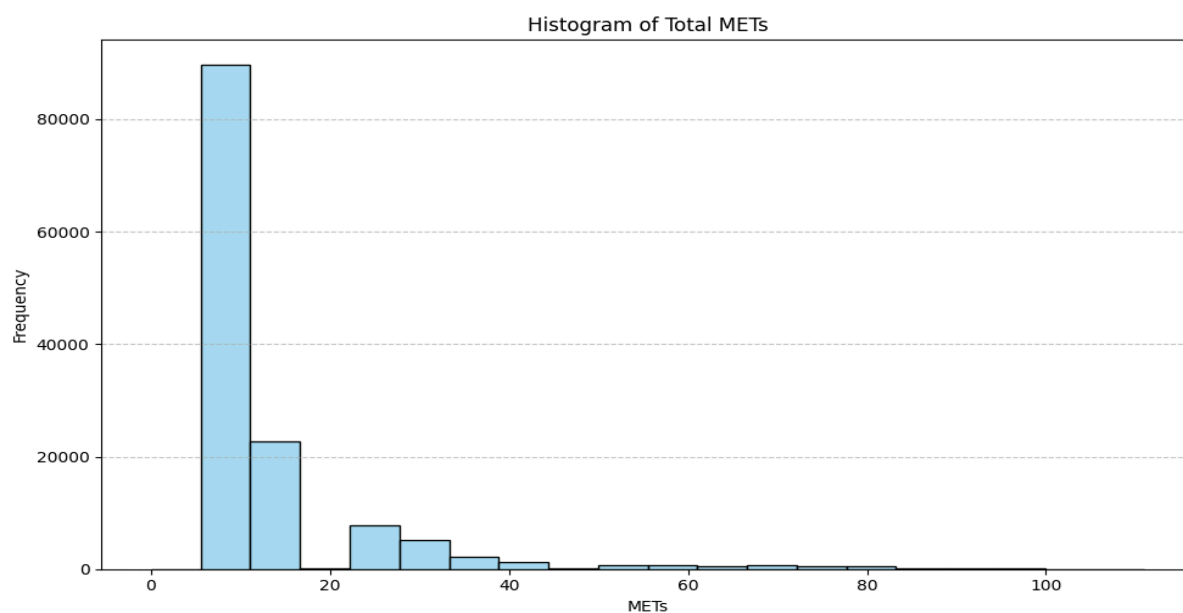
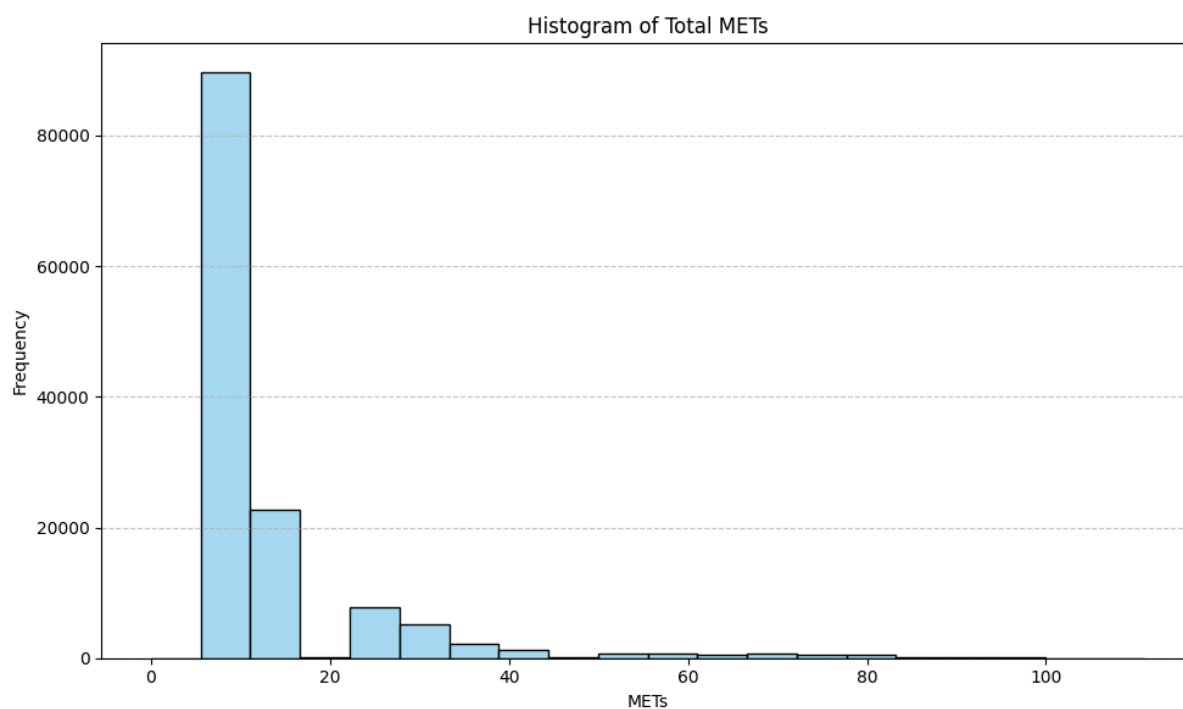
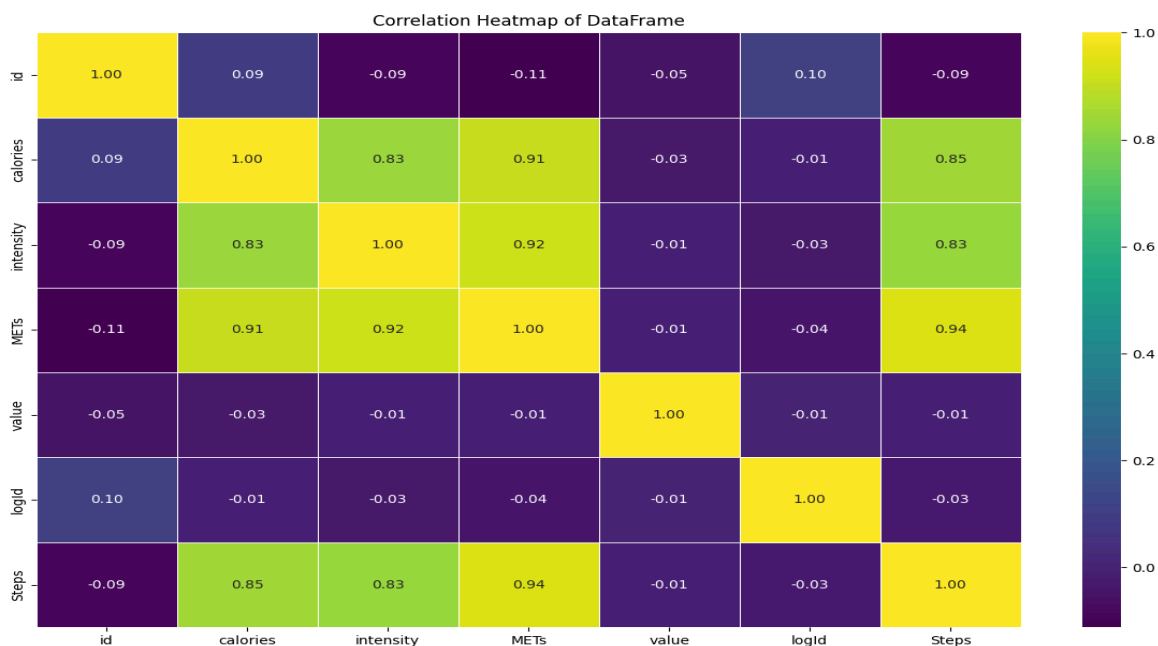
Minute dataset:

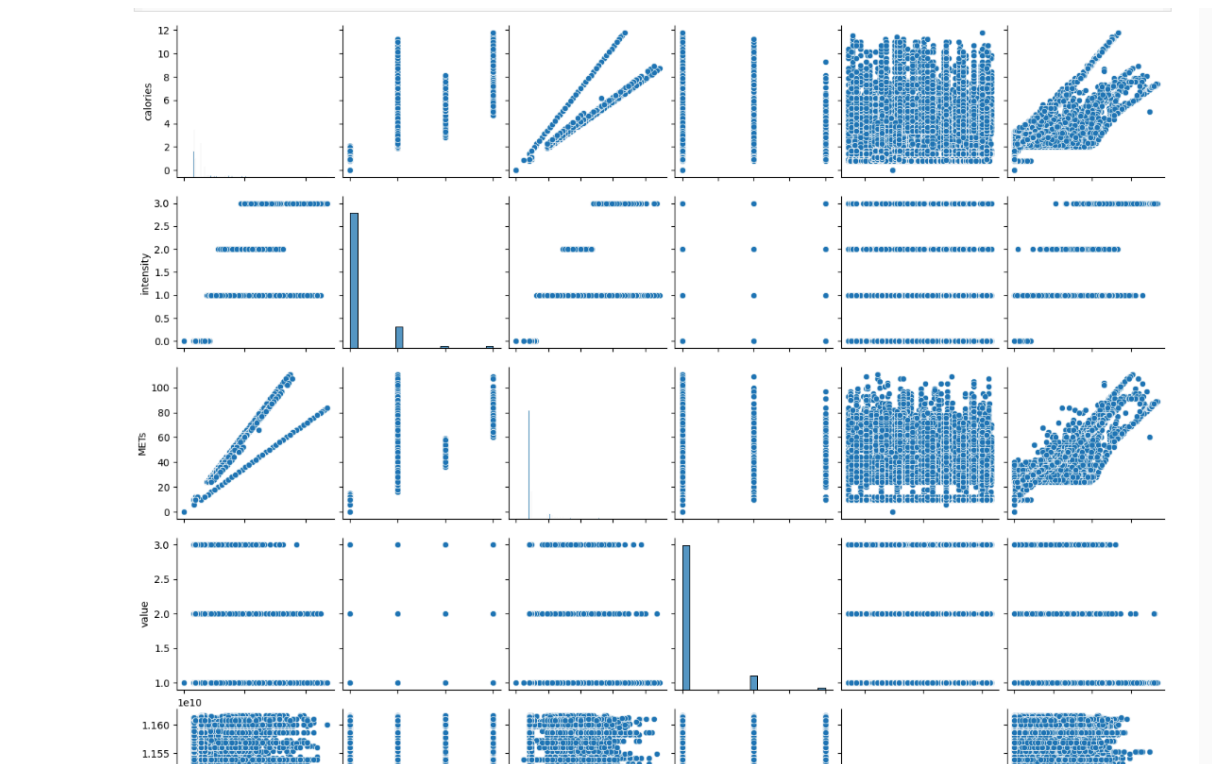
To visualize top 10 users based on the maximum Calories intake in a single day. The provided dataset contains information about users' activities, including their ID, activityminute, and calories covered. By aggregating this data and visualizing it using a bar plot, I was able to identify the users who have achieved the highest maximum Calories covered in a single day. Based on data it seems user 5577150313 has covered maximum Calories.



To visualize top 10 users based on the maximum METs intake in a single day. The provided dataset contains information about users' activities, including their ID, activityminute, and METs covered. By aggregating this data and visualizing it using a bar plot, I was able to identify the users who have achieved the highest maximum METs covered in a single day. Based on data it seems user 1503960366 has covered maximum METs.







Full google colab file: [Link](#)