CAPSTONE PROJECT REPORT

Fitness Tracking Dashboard

Using Google Colab (Python) & Tableau

PrepInsta

Report & Analysis by

Neeraj Sharma

DAB3018

Under the Guidance of

Sr Rushikesh Konapure

Data Scientist and Instructor @Prepinsta

Acknowledgement

Under the mentorship of ***Sir Rishikesh Konapure***, I analyzed ***FitBit Fitness Tracker*** App data. Extracting insights from minute-level data of 30 Fitbit users, I used Python and Pandas for cleaning, transformation, and analysis. The diverse dataset, generated via Amazon Mechanical Turk, provided trends on user behavior. Deliverables included a concise summary, data source descriptions, cleaning documentations, visualizations, key findings, and high-level content insights.

Fitbit Consumer Behavior Analysis

Objective :

Imagine you are a data analyst at "HealthTrackers Inc.," a fictional company operating in the Fitbit industry. Your company is dedicated to understanding consumer behavior to enhance product offerings and optimize marketing strategies. You have been tasked with analyzing a comprehensive dataset obtained from Fitbit users to uncover trends and insights.

The business objective is to identify key trends, understand their implications for customers, and leverage these insights to shape an effective marketing strategy.

Content

Respondents generated this dataset to a distributed survey via Amazon Mechanical Turk between 03.12.2016 and 05.12.2016. Thirty eligible Fitbit users consented to submit 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 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?

Deliverables:

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of the analysis
5. Supporting visualizations and key findings
6. High-level content recommendations based on the analysis

Tools:

Python for Data Cleaning, Data Transformation, Data Visualisation and Data Analysis,

Pandas Profiling, Tableau, Excel, PowerBI, SQL

Fitbit

- **Summary**

  The FitBit Fitness Tracker App analysis project involved examining data from 18 different sources, including daily activity, calories, hourly steps, etc., collected through a distributed survey. After cleaning and merging the datasets, the analysis focused on identifying trends in daily and hourly activity, minute-level patterns, sleep data, weight logs, and heart rate variations. Utilizing visualizations such as Pandas Profiling, histograms, and scatter plots, the project aimed to provide actionable insights for the marketing team, including potential app feature improvements and correlations between different metrics. The tools used included Python for data processing and visualization, with Pandas Profiling for comprehensive data exploration.

- **Ask Phase**

  Business Task

  Analyze FitBit Fitness Tracker App data to gain insights into how consumers are using 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?

Prepare Phase

In the Prepare phase of the FitBit Fitness Tracker App analysis project, the focus was on setting the groundwork for robust data exploration and analysis. This phase involved:

1. **Data Collection :** Collected data from 18 different files, encompassing various aspects of FitBit app usage, through a distributed survey on Amazon Mechanical Turk. These files included information on daily and hourly activity, calories, sleep, weight logs, and heart rate.
2. **Data Cleaning & Preprocessing :** Conducted thorough data cleaning to address issues such as missing values, outliers, and inconsistencies. Merged relevant datasets to create a consolidated and comprehensive dataset for analysis. Extracted meaningful features from timestamp data to enhance the analysis.
3. **Data Profiling :** Leveraged Pandas Profiling to generate comprehensive data profiles, gaining insights into data distributions, correlations, and potential issues. This step helped in understanding the structure and characteristics of the dataset.
4. **Feature Engineering :** Derived new features or transformed existing ones to enhance the dataset's informativeness for subsequent analysis. This included extracting relevant information from timestamps and creating aggregated metrics.
5. **Data Validation :** Ensured the integrity and consistency of the dataset through validation checks. Verified that the data aligned with the expected patterns and distributions, addressing any anomalies that could impact the analysis.

The Prepare phase laid a solid foundation, ensuring that the dataset was cleaned, organized, and enriched for the subsequent exploration and analysis stages. The goal was to set the stage for meaningful insights and actionable recommendations in the upcoming phases of the project.

Data Sets

1. Uncleaned Data :

Daily Activity, Daily Calories, Daily Steps, Daily Intensities, Heart Rate, Hourly Calories, Hourly Steps, Hourly Intensities, Minute MET, Minute Sleep, Minute Steps, Sleep Day, Weight Log

2. Cleaned & Merged Data :

	id	activity_date	daily_average_heartrate	total_steps	total_distance	tracker_distance	logged_activities_distance	very_active_distance	moderately_active_distance
2	1503960366	2016-04-12		13162	8.5	8.5	0	1.88	0.55
3	1503960366	2016-04-13		10735	6.97	6.97	0	1.57	0.69
4	1503960366	2016-04-14		10460	6.74	6.74	0	2.44	0.4
5	1503960366	2016-04-15		9762	6.28	6.28	0	2.14	1.26
6	1503960366	2016-04-16		12669	8.16	8.16	0	2.71	0.41
7	1503960366	2016-04-17		9705	6.48	6.48	0	3.19	0.78
8	1503960366	2016-04-18		13019	8.59	8.59	0	3.25	0.64
9	1503960366	2016-04-19		15506	9.88	9.88	0	3.53	1.32
10	1503960366	2016-04-20		10544	6.68	6.68	0	1.96	0.48
11	1503960366	2016-04-21		9819	6.34	6.34	0	1.34	0.35
12	1503960366	2016-04-22		12764	8.13	8.13	0	4.76	1.12
13	1503960366	2016-04-23		14371	9.04	9.04	0	2.81	0.87
14	1503960366	2016-04-24		10039	6.41	6.41	0	2.92	0.21
15	1503960366	2016-04-25		15355	9.8	9.8	0	5.29	0.57
16	1503960366	2016-04-26		13755	8.79	8.79	0	2.33	0.92
17	1503960366	2016-04-27		18134	12.21	12.21	0	6.4	0.41
18	1503960366	2016-04-28		13154	8.53	8.53	0	3.54	1.16
19	1503960366	2016-04-29		11181	7.15	7.15	0	1.06	0.5
20	1503960366	2016-04-30		14673	9.25	9.25	0	3.56	1.42
21	1503960366	2016-05-01		10602	6.81	6.81	0	2.29	1.6
22	1503960366	2016-05-02		14727	9.71	9.71	0	3.21	0.57
23	1503960366	2016-05-03		15103	9.66	9.66	0	3.73	1.05
24	1503960366	2016-05-04		11100	7.15	7.15	0	2.46	0.87

Fig : Daily Stats

	A	B	C	D	E	F	G
1	id	activity_day	activity_hour	calories	total_intensity	average_intensity	step_total
2	1503960366	2016-04-12	00:00	81	20	0.33	373
3	1503960366	2016-04-12	01:00	61	8	0.13	160
4	1503960366	2016-04-12	02:00	59	7	0.12	151
5	1503960366	2016-04-12	03:00	47	0	0	0
6	1503960366	2016-04-12	04:00	48	0	0	0
7	1503960366	2016-04-12	05:00	48	0	0	0
8	1503960366	2016-04-12	06:00	48	0	0	0
9	1503960366	2016-04-12	07:00	47	0	0	0
10	1503960366	2016-04-12	08:00	68	13	0.22	250
11	1503960366	2016-04-12	09:00	141	30	0.5	1864
12	1503960366	2016-04-12	10:00	99	29	0.48	676
13	1503960366	2016-04-12	11:00	76	12	0.2	360
14	1503960366	2016-04-12	12:00	73	11	0.18	253
15	1503960366	2016-04-12	13:00	66	6	0.1	221
16	1503960366	2016-04-12	14:00	110	36	0.6	1166
17	1503960366	2016-04-12	15:00	151	58	0.97	2063
18	1503960366	2016-04-12	16:00	76	13	0.22	344
19	1503960366	2016-04-12	17:00	83	16	0.27	489
20	1503960366	2016-04-12	18:00	124	29	0.48	1386
21	1503960366	2016-04-12	19:00	104	39	0.65	558
22	1503960366	2016-04-12	20:00	132	41	0.68	1733
23	1503960366	2016-04-12	21:00	100	31	0.52	684
24	1503960366	2016-04-12	22:00	65	9	0.15	89

hourly_stats_data

Fig : Hourly Stats

	A	B	C	D	E	F	G	H
1	id	activity_day	activity_minute	mets	calories	intensity	sleep	steps
2	1503960366	2016-04-12	00:00:00	10	0.79	0	0	0
3	1503960366	2016-04-12	00:01:00	10	0.79	0	0	0
4	1503960366	2016-04-12	00:02:00	10	0.79	0	0	0
5	1503960366	2016-04-12	00:03:00	10	0.79	0	0	0
6	1503960366	2016-04-12	00:04:00	10	0.79	0	0	0
7	1503960366	2016-04-12	00:05:00	12	0.94	0	0	0
8	1503960366	2016-04-12	00:06:00	12	0.94	0	0	0
9	1503960366	2016-04-12	00:07:00	12	0.94	0	0	0
10	1503960366	2016-04-12	00:08:00	12	0.94	0	0	0
11	1503960366	2016-04-12	00:09:00	12	0.94	0	0	0
12	1503960366	2016-04-12	00:10:00	12	0.94	0	0	0
13	1503960366	2016-04-12	00:11:00	12	0.94	0	0	0
14	1503960366	2016-04-12	00:12:00	10	0.79	0	0	0
15	1503960366	2016-04-12	00:13:00	10	0.79	0	0	0
16	1503960366	2016-04-12	00:14:00	12	0.94	0	0	0
17	1503960366	2016-04-12	00:15:00	10	0.79	0	0	0
18	1503960366	2016-04-12	00:16:00	12	0.94	0	0	0
19	1503960366	2016-04-12	00:17:00	10	0.79	0	0	0
20	1503960366	2016-04-12	00:18:00	10	0.79	0	0	0
21	1503960366	2016-04-12	00:19:00	10	0.79	0	0	0
22	1503960366	2016-04-12	00:20:00	12	0.94	0	0	0
23	1503960366	2016-04-12	00:21:00	12	0.94	0	0	0
24	1503960366	2016-04-12	00:22:00	12	0.94	0	0	0

minute_stats_data

Fig : Minute Stats

Analyze and Share Phase

EXPLORATORY DATA ANALYSIS

USING GOOGLE COLAB

Daily Stats

1. Descriptive Numeric Analysis

```
[15] df.describe()
```

	id	daily_average_heartrate	total_steps	total_distance	tracker_distance	logged_activi
count	9.400000e+02	334.000000	940.000000	940.000000	940.000000	
mean	4.855407e+09	75.974042	7637.910638	5.489702	5.475351	
std	2.424805e+09	10.340623	5087.150742	3.924606	3.907276	
min	1.503960e+09	57.870000	0.000000	0.000000	0.000000	
25%	2.320127e+09	67.512500	3789.750000	2.620000	2.620000	
50%	4.445115e+09	75.220000	7405.500000	5.245000	5.245000	
75%	6.962181e+09	81.932500	10727.000000	7.712500	7.710000	
max	8.877689e+09	107.720000	36019.000000	28.030000	28.030000	

8 rows × 21 columns

2. Correlation Analysis

```
[16] num_column = df.select_dtypes(include='number')
     matrix = num_column.corr()
     matrix
```

	id	daily_average_heartrate	total_steps	total_distance	tracker_distance
id	1.000000	0.025742	0.185721	0.241000	0.238816
daily_average_heartrate	0.025742	1.000000	0.011922	0.006960	-0.006357
total_steps	0.185721	0.011922	1.000000	0.985369	0.984822
total_distance	0.241000	0.006960	0.985369	1.000000	0.999505
tracker_distance	0.238816	-0.006357	0.984822	0.999505	1.000000
logged_activities_distance	0.187965	0.295305	0.181850	0.188323	0.162572
very_active_distance	0.308691	0.019929	0.740115	0.794582	0.794338
moderately_active_distance	0.026665	-0.013992	0.507105	0.470758	0.470277
light_active_distance	0.019629	-0.006604	0.692208	0.662002	0.661365
sedentary_active_distance	-0.015698	0.009890	0.070505	0.082389	0.074591
very_active_minutes	0.303608	-0.055824	0.667079	0.681297	0.680816

3. USERS WITH MOST ACTIVE AND LESS ACTIVE MINUTES

```
[17] most_active_users = df.sort_values(by=['very_active_minutes', 'fairly_active_minutes'], ascending=False).
     most_active_users
```

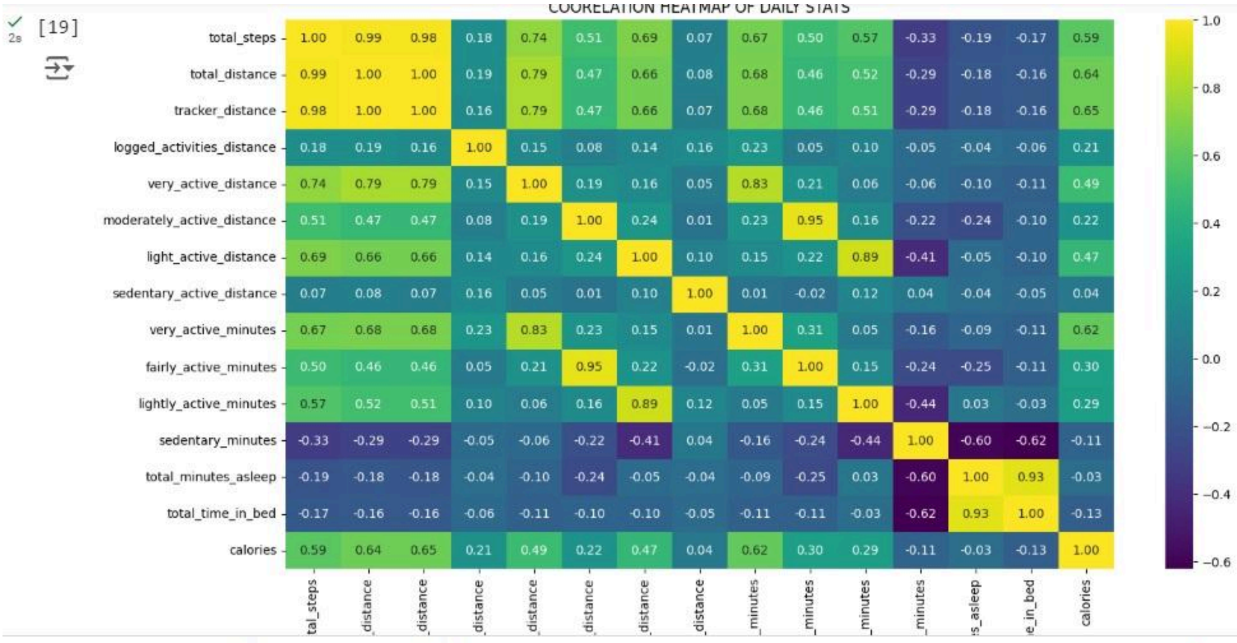
	id	activity_date	very_active_minutes	fairly_active_minutes
579	5577150313	2016-04-24	210	65
585	5577150313	2016-04-30	207	45
572	5577150313	2016-04-17	200	37
586	5577150313	2016-05-01	194	72
50	1624580081	2016-05-01	186	63
571	5577150313	2016-04-16	184	56
827	8378563200	2016-04-21	137	16
771	8053475328	2016-04-15	132	8
780	8053475328	2016-04-24	129	33
779	8053475328	2016-04-23	125	14

```
[18] less_active_users = df[(df['very_active_minutes'] > 0) & (df['fairly_active_minutes'] > 0)]
     less_active_users = less_active_users.sort_values(by=['very_active_minutes', 'fairly_active_minutes']).he
     less_active_users
```

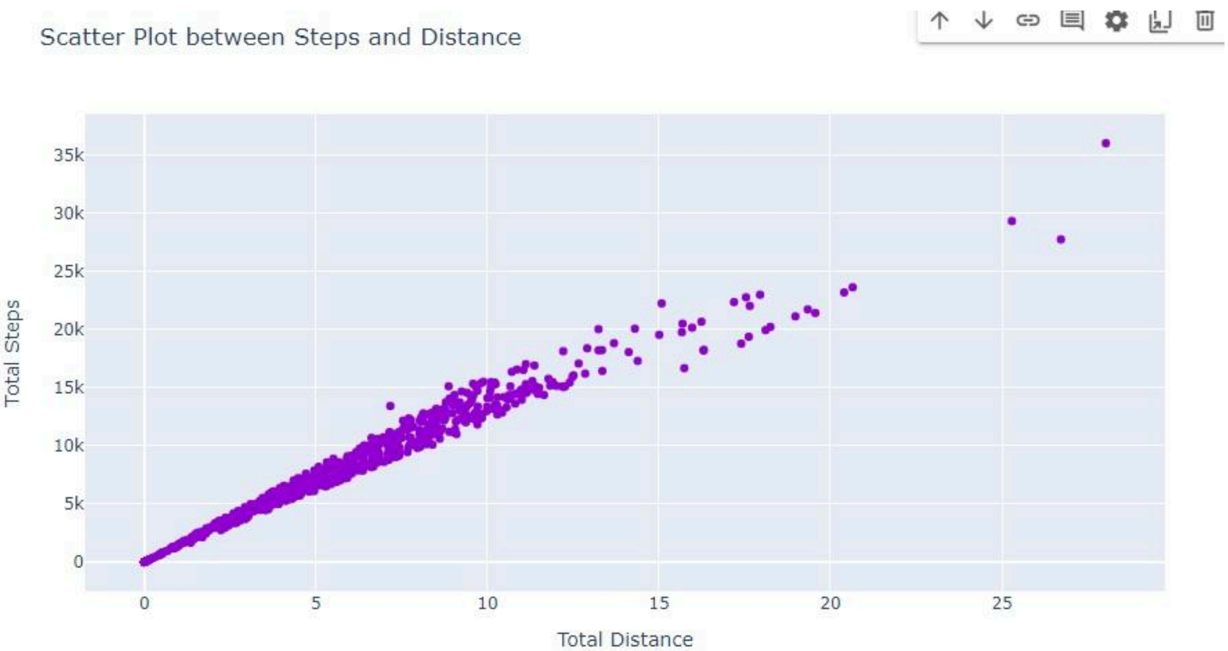
	id	activity_date	very_active_minutes	fairly_active_minutes
393	4319703577	2016-04-24	1	5
135	1927972279	2016-04-24	1	6
271	2873212765	2016-04-18	1	6
286	2873212765	2016-05-03	1	6
82	1644430081	2016-05-02	1	7
252	2347167796	2016-04-17	1	7
508	4702921684	2016-04-15	1	8
519	4702921684	2016-04-26	1	8
383	4319703577	2016-04-14	1	9
387	4319703577	2016-04-18	1	9

Data Visualization

1. Heat Map

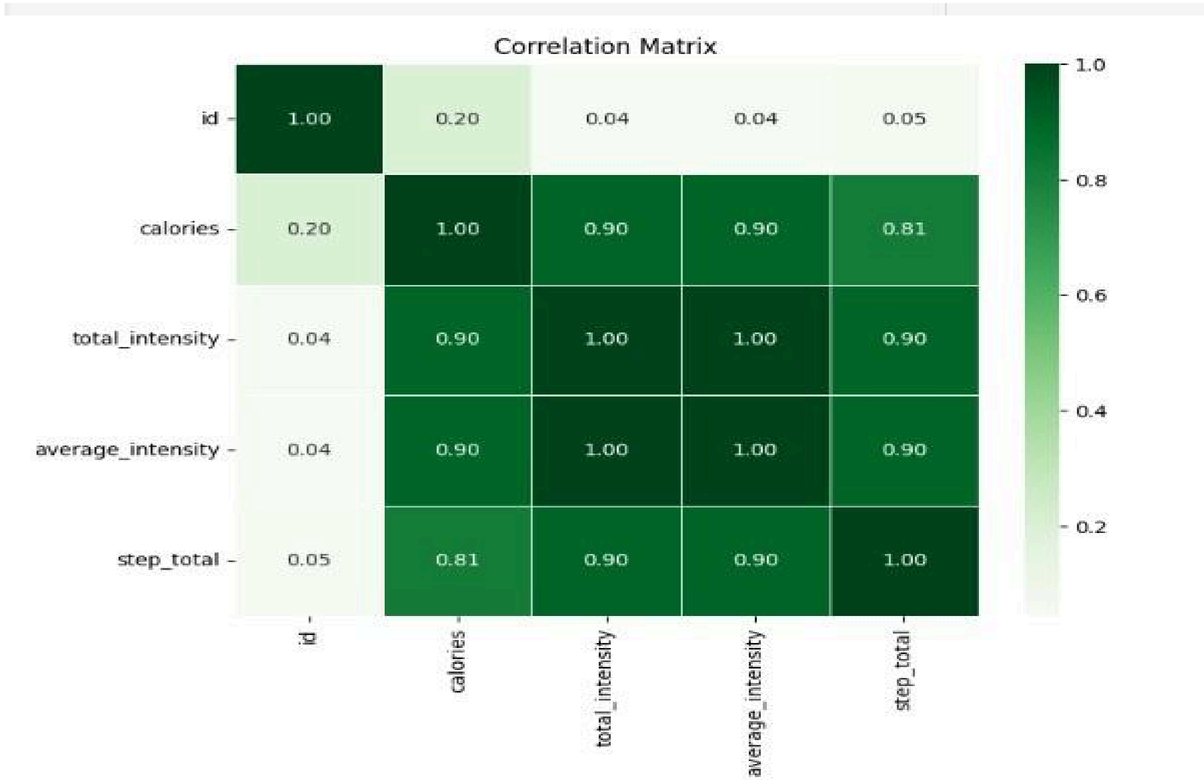


2. Scatter Plot

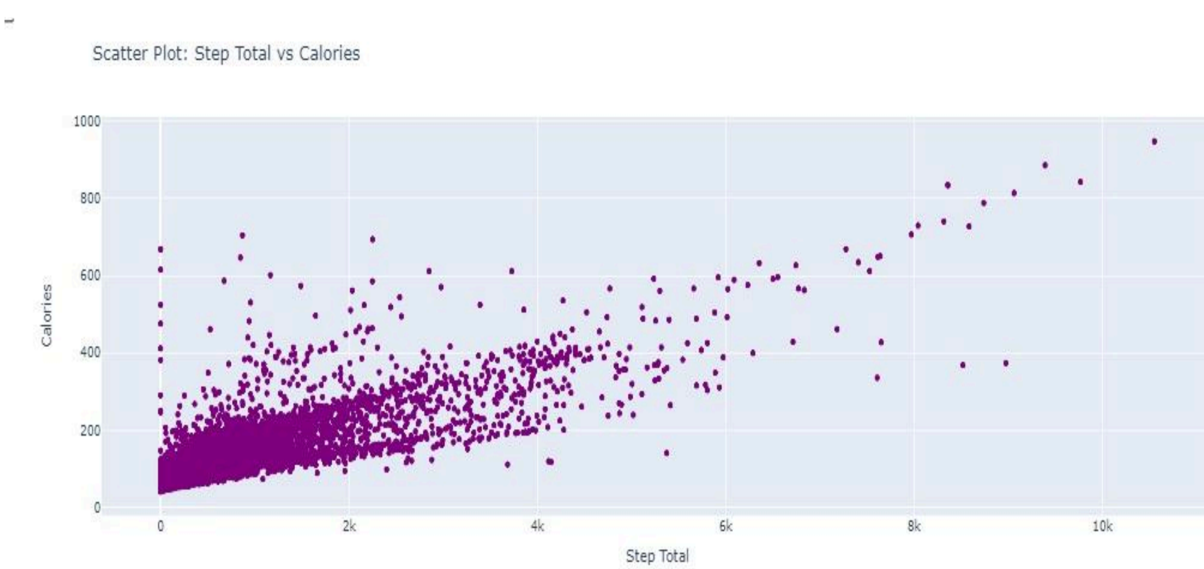


Hourly Stats

1. Heat Map



2. Scatter Plot



Minute Stats

1. TOP & BOTTOM USERS BY MET

```
top[['id', 'activity_day', 'activity_minute', 'mets']]
```

	id	activity_day	activity_minute	mets
411343	2873212765	2016-05-07	07:43:00	157
411344	2873212765	2016-05-07	07:44:00	153
391207	2873212765	2016-04-23	08:07:00	149
231633	2022484408	2016-04-21	16:33:00	146
231634	2022484408	2016-04-21	16:34:00	144
708096	4558609924	2016-05-08	13:36:00	144
231667	2022484408	2016-04-21	17:07:00	144
611678	4388161847	2016-05-02	18:38:00	141
1324347	8877689391	2016-05-11	17:27:00	141
1292634	8877689391	2016-04-19	16:54:00	141

2s

```
bottom[['id', 'activity_day', 'activity_minute', 'mets']]
```

	id	activity_day	activity_minute	mets
126059	1644430081	2016-05-08	23:59:00	6
2404	1503960366	2016-04-13	16:04:00	10
5283	1503960366	2016-04-15	16:03:00	10
5282	1503960366	2016-04-15	16:02:00	10
5281	1503960366	2016-04-15	16:01:00	10
960	1503960366	2016-04-12	16:00:00	10
2399	1503960366	2016-04-13	15:59:00	10
5278	1503960366	2016-04-15	15:58:00	10
5277	1503960366	2016-04-15	15:57:00	10
5285	1503960366	2016-04-15	16:05:00	10

2. TOP & BOTTOM USERS BY CALORIES

```
top[['id', 'activity_day', 'activity_minute', 'calories']]
```

	id	activity_day	activity_minute	calories
891517	6290855005	2016-04-17	09:37:00	19.75
891526	6290855005	2016-04-17	09:46:00	19.75
891535	6290855005	2016-04-17	09:55:00	19.75
891534	6290855005	2016-04-17	09:54:00	19.75
891533	6290855005	2016-04-17	09:53:00	19.75
891532	6290855005	2016-04-17	09:52:00	19.75
891531	6290855005	2016-04-17	09:51:00	19.75
891530	6290855005	2016-04-17	09:50:00	19.75
891528	6290855005	2016-04-17	09:48:00	19.75
891527	6290855005	2016-04-17	09:47:00	19.75

0s

```
bottom[['id', 'activity_day', 'activity_minute', 'calories']]
```

	id	activity_day	activity_minute	calories
450240	3977333714	2016-04-14	00:00:00	0.7
448324	3977333714	2016-04-12	16:04:00	0.7
448323	3977333714	2016-04-12	16:03:00	0.7
449762	3977333714	2016-04-13	16:02:00	0.7
449761	3977333714	2016-04-13	16:01:00	0.7
448320	3977333714	2016-04-12	16:00:00	0.7
448319	3977333714	2016-04-12	15:59:00	0.7
448318	3977333714	2016-04-12	15:58:00	0.7
448317	3977333714	2016-04-12	15:57:00	0.7
448325	3977333714	2016-04-12	16:05:00	0.7

3. TOP & BOTTOM USERS BY STEPS

```
top[['id', 'activity_day', 'activity_minute', 'steps']]
```

	id	activity_day	activity_minute	steps
1111064	8053475328	2016-04-30	15:44:00	220
692566	4558609924	2016-04-27	18:46:00	207
473207	3977333714	2016-04-29	22:47:00	190
480479	3977333714	2016-05-04	23:59:00	189
473203	3977333714	2016-04-29	22:43:00	188
467363	3977333714	2016-04-25	21:23:00	187
1294175	8877689391	2016-04-20	18:35:00	187
1294166	8877689391	2016-04-20	18:26:00	186
480480	3977333714	2016-05-05	00:00:00	186
1294178	8877689391	2016-04-20	18:38:00	185

0s

```
bottom[['id', 'activity_day', 'activity_minute', 'steps']]
```

	id	activity_day	activity_minute	steps
63180	1624580081	2016-04-26	00:00:00	1
642549	4445114986	2016-04-23	14:09:00	1
403090	2873212765	2016-05-01	14:10:00	1
126911	1644430081	2016-05-09	14:11:00	1
521652	4020332650	2016-05-04	14:12:00	1
732133	4702921684	2016-04-24	14:13:00	1
126914	1644430081	2016-05-09	14:14:00	1
347115	2320127002	2016-05-10	14:15:00	1
126917	1644430081	2016-05-09	14:17:00	1
1298238	8877689391	2016-04-23	14:18:00	1

4. CORRELATION ANALYSIS

```
crr = df[['mets', 'calories', 'intensity', 'sleep', 'steps']].corr()
crr
```

	mets	calories	intensity	sleep	steps
mets	1.000000	0.916072	0.797352	-0.046622	0.856944
calories	0.916072	1.000000	0.729724	-0.040229	0.759650
intensity	0.797352	0.729724	1.000000	-0.032558	0.675697
sleep	-0.046622	-0.040229	-0.032558	1.000000	-0.048410
steps	0.856944	0.759650	0.675697	-0.048410	1.000000

5. Heat Map

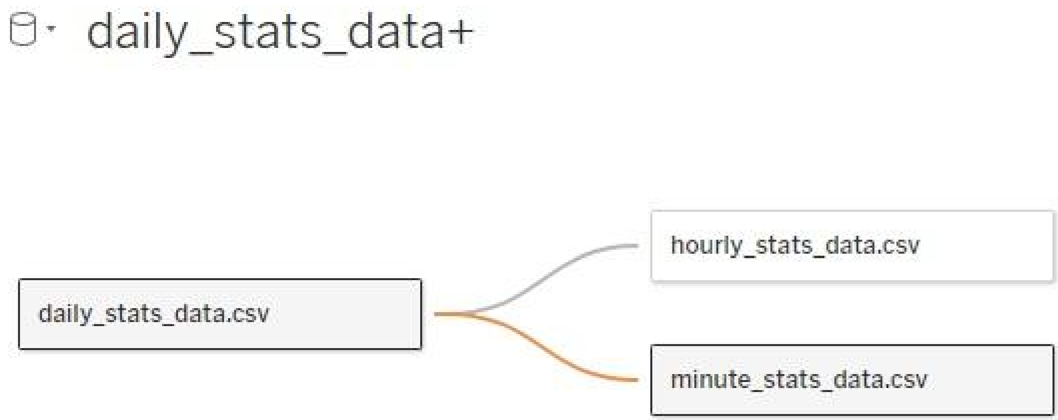
Act Phase

Key Insights

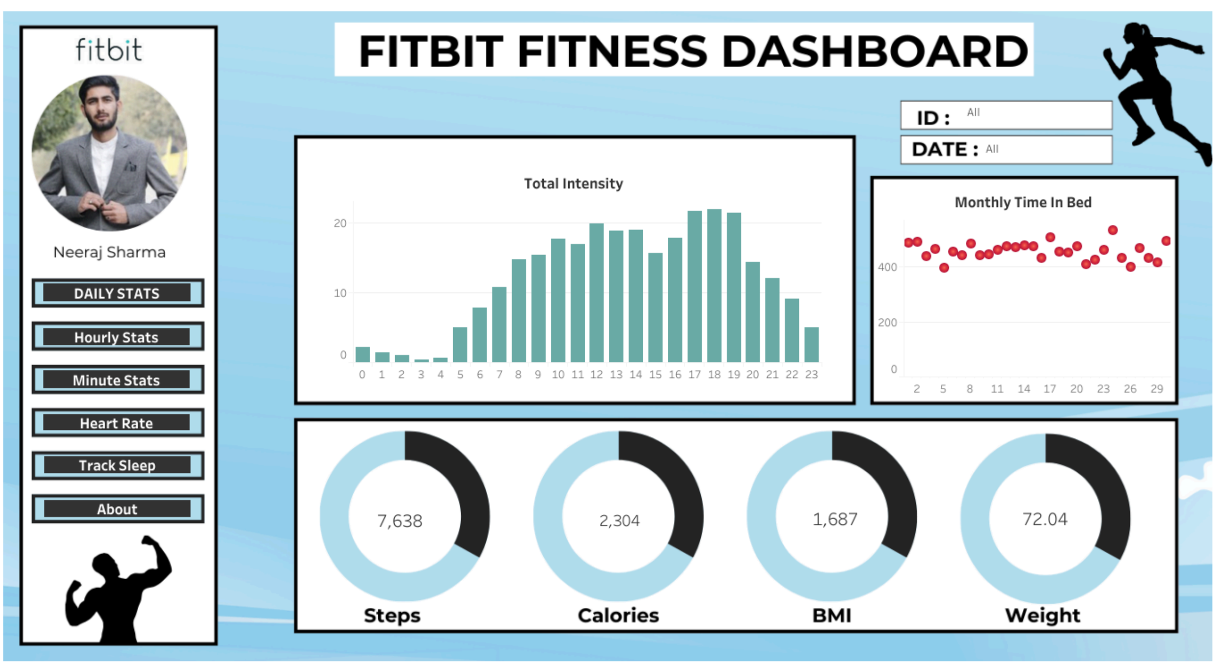
1. In terms of physical activities on a daily basis, Users spent the most time (~ 3.2hrs) and highest distance (3.35km) in the lightly active level.
2. Although users spent 21 minutes on average in the Very Active category, 81% of their day is spent being sedentary which highlights a concern.
3. The average user burns 2307 calories and clocks 7652 steps per day.
4. Users seem to burned a consistent amount of calories throughout the week with the highest burned (2365 calories) on Saturdays and lowest (2204 calories) on Thursdays.
5. The average user burn the highest calories between 5pm-7pm.
6. The highest number of steps clocked (8125 steps) are on Tuesdays and the lowest(6993 steps) are on Sundays.
7. The average user begins their day at 5am and clocked the highest number of steps between 5-7pm. They gradually reduce their activeness from 8pm onwards.
8. There is a strong positive linear relationship between total steps clocked and total calories burned.
9. Users have a consistent sleep schedule with a mean sleep hours of 419.5 minutes (~ 7hrs) across the week. The highest recorded mean time asleep was on Sundays (~ 7.5hrs) and the lowest was on Thursdays (~ 6.7 hrs).
10. 44.3% of users have inadequate sleep hours(<7hours).
11. At least 5 relevant pairs of variables are found to have a strong correlation (r >0.6).

FITBIT FITNESS DASHBOARD

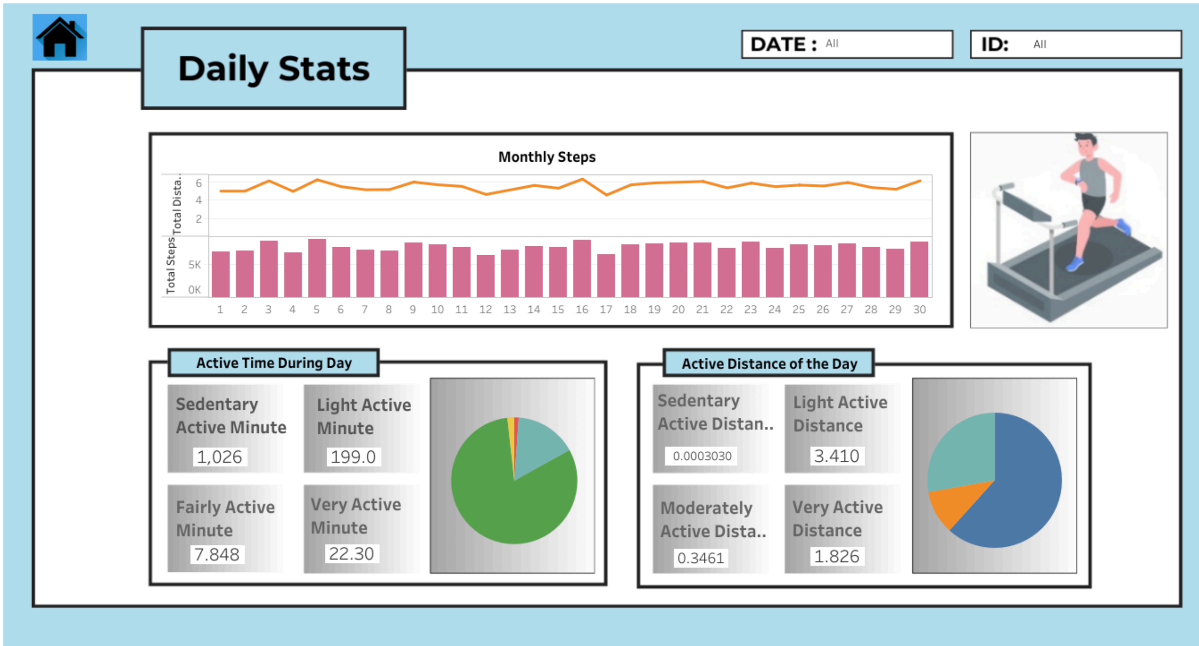
- IMPORTING DATASETS IN TABLEAU



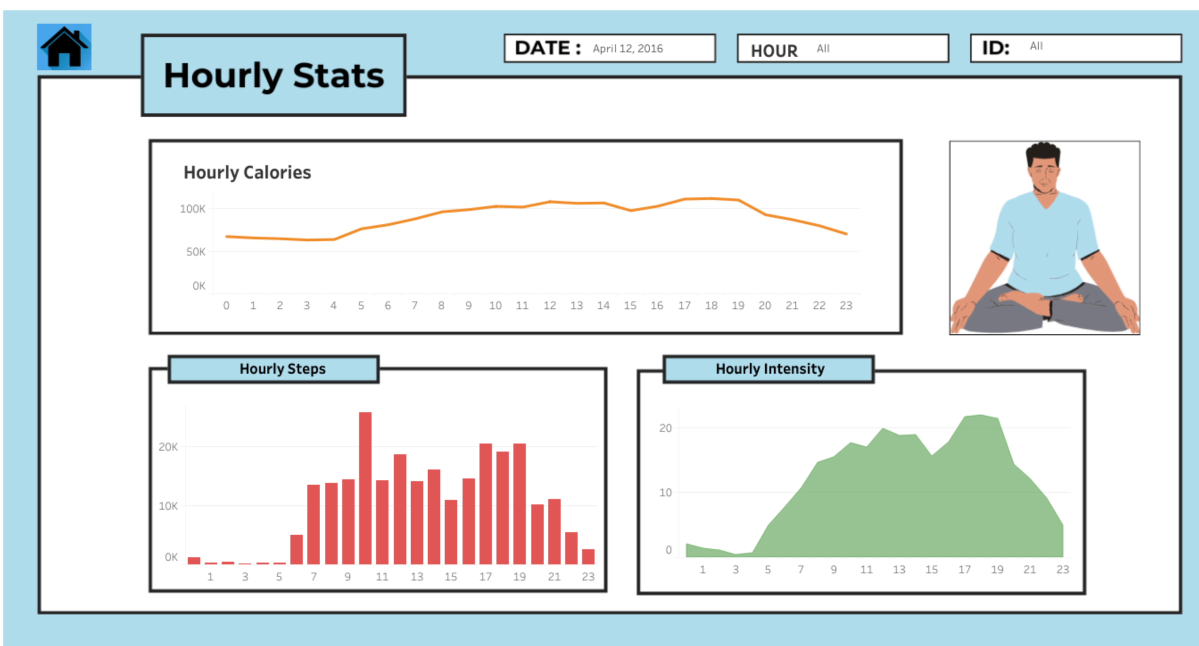
Home Page



Daily Stats

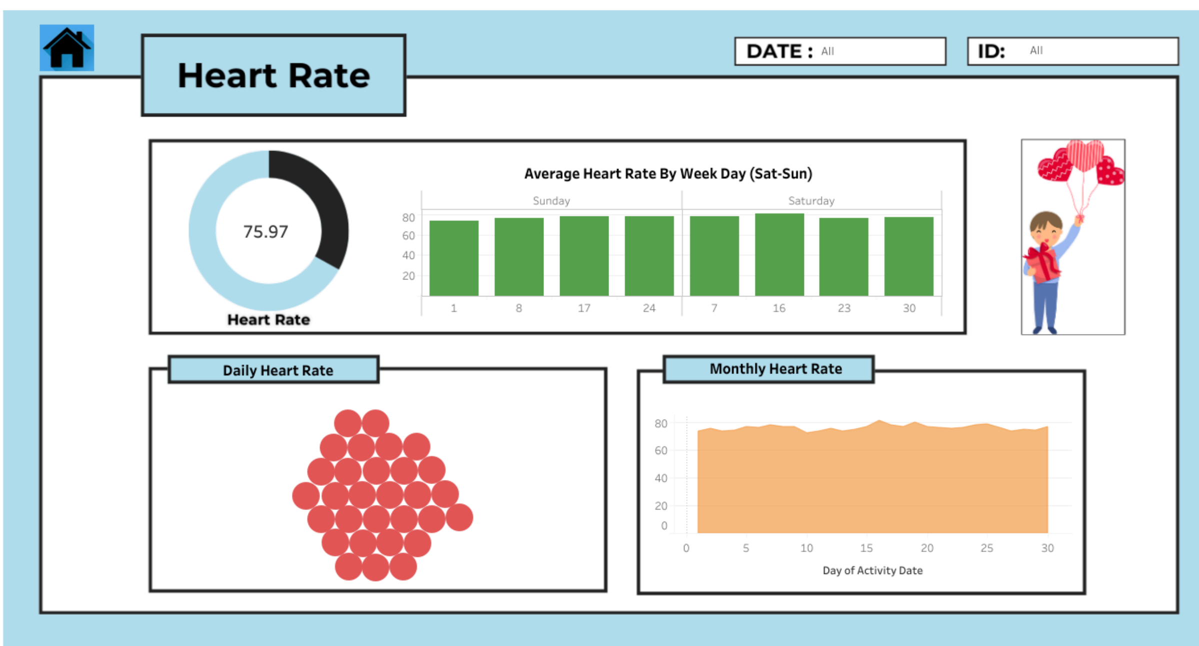


Hourly Stats



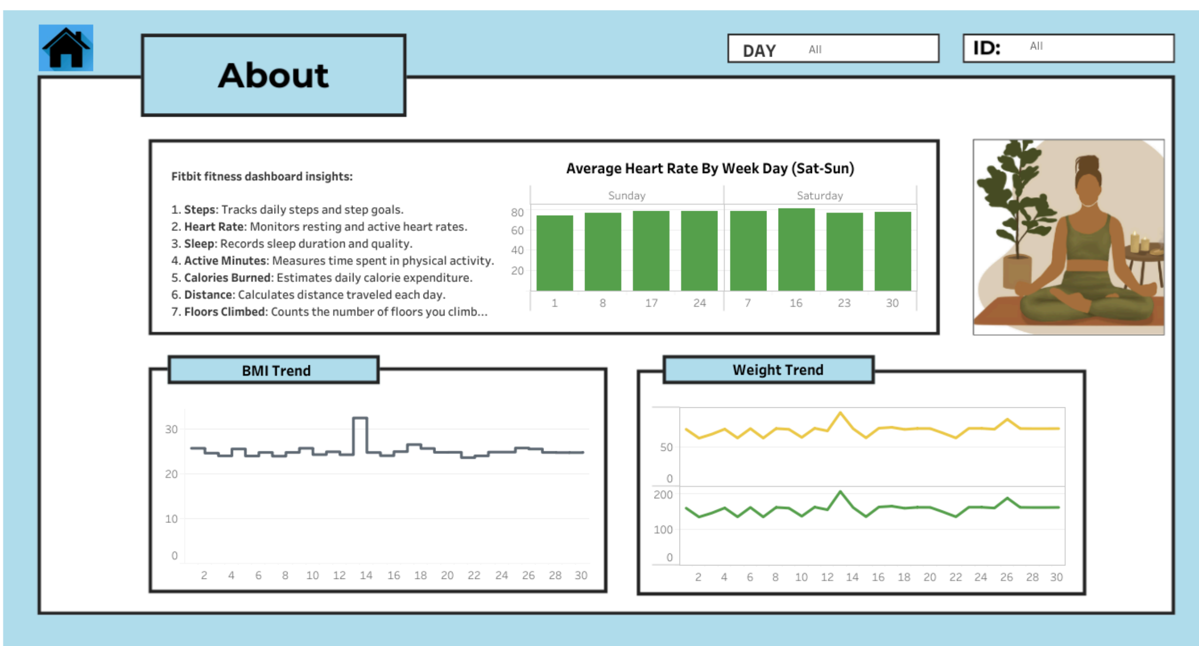
Minute Stats

Heart Rate



Track Sleep

About



Conclusion

The dashboard's intuitive design ensures that complex data is presented in an accessible manner, empowering users to make informed decisions about their fitness routines. Additionally, the ability to drill down into specific metrics helps users set realistic goals, monitor progress, and adjust their activities accordingly. Overall, this Fitness Dashboard serves as a powerful tool for both fitness enthusiasts and professionals, providing actionable insights that drive healthier lifestyles.