

Winter Internship



By Deeksha Raina

## Week 8 Capstone Project

Fitness Dataset EDA &  
Dashboard  
Using Google Colab &  
Tableau

Report & Analysis By:  
Deeksha Raina

# Acknowledgement



Under the mentorship of Sir Rushikesh Konapure and with the support of the Preplinsta Team, I analyzed data from the FitBit Fitness Tracker App. Extracting insights from minute-level data of Fitbit users, I utilized Python and Pandas for data cleaning, transformation, and analysis. The diverse dataset, generated via Amazon Mechanical Turk, offered trends in user behavior. Deliverables comprised a concise summary, descriptions of data sources, documentation of cleaning procedures, visualizations, key findings, and high-level analysis and insights.

# DATA ANALYSIS PROCESS



## Data Cleaning & Merging

Checking for null values, dropping duplicate entries, standardized data values & joining DFs using pd.merge. Always make sure to perform data cleaning on the merged df as well.



## EDA using Python

Performing exploratory data analysis wrt various parameters. Finding min, and max values of Total steps & Calories burnt; checking entries for a particular user & analyzing provided data for creating visualizations.



## Visualizations & Trends

I am using Matplotlib & Seaborn to plot graphs & charts thus gathering insights about the data!

Visualizations provide a better understanding of the relationships among the variables. We further analyze the trends present in the data.

# ORIGINAL DATASETS

	ActivityDay	Calories
1503960366	4/12/2016	1985
1503960366	4/13/2016	1797
1503960366	4/14/2016	1776
1503960366	4/15/2016	1745

Daily Calories Dataset

Sedentary	LightlyAct	FairlyActiv	VeryActiv	Sedentary	LightActiv	Moderate	VeryActiv
728	328	13	25	0	6.06	0.55	1.88
776	217	19	21	0	4.71	0.69	1.57
1218	181	11	30	0	3.91	0.4	2.44
726	209	34	29	0	2.83	1.26	2.14

Intensity Dataset

	ActivityDay	StepTotal
03960366	4/12/2016	13162
03960366	4/13/2016	10735
03960366	4/14/2016	10460

Daily Steps Dataset

Id	ActivityHour	Calories
1503960366	4/12/2016 0:00	81
1503960366	4/12/2016 1:00	61
1503960366	4/12/2016 2:00	59

Hourly Calories Dataset

Id	ActivityMinute	METs
1503960366	4/12/2016 0:00	10
1503960366	4/12/2016 0:01	10
1503960366	4/12/2016 0:02	10

Minute METs Dataset

Id	Date	WeightKg	WeightPo	BMI	IsManualF
1503960366	5/2/2016 23:59	52.6	116	22.6	TRUE
1503960366	5/3/2016 23:59	52.6	116	22.6	TRUE
1927972279	4/13/2016 1:08	133.5	294.3	47.5	FALSE
2873212765	4/21/2016 23:59	56.7	125	21.5	TRUE

Weight Dataset

This is just a glimpse of a few original datasets provided to us!

# MERGED DATASETS

See of our df!

Daily Activity Merged Df!

totalSteps	totalDistance	trackersDistance	loggedActivitiesDistance	veryActiveDistance	moderatelyActiveDistance	lightActiveDistance	SedentaryActiveDistance	veryActiveMinutes	fairlyActiveMinutes	lightlyActiveMinutes	SedentaryMinutes	Calories
13162	8.5	8.5	0.0	1.9	0.6	6.1	0.0	25	13	338	728	1985
10735	7.0	7.0	0.0	1.6	0.7	4.7	0.0	21	19	217	776	1797
10460	6.7	6.7	0.0	2.4	0.4	3.9	0.0	30	11	181	1218	1776
9782	6.3	6.3	0.0	2.1	1.3	2.8	0.0	29	34	209	738	1745
12669	8.1	8.1	0.0	2.7	0.4	5.0	0.0	36	10	221	773	1863

df ActivityMinute Calories intensity StepTotal

ID	ActivityMinute	Calories	intensity	StepTotal
1939036	4/12/2016 12:00:00 AM	0.7865	0	373
1939036	4/12/2016 12:00:00 AM	0.7865	0	31
1939036	4/12/2016 12:00:00 AM	0.7865	0	45

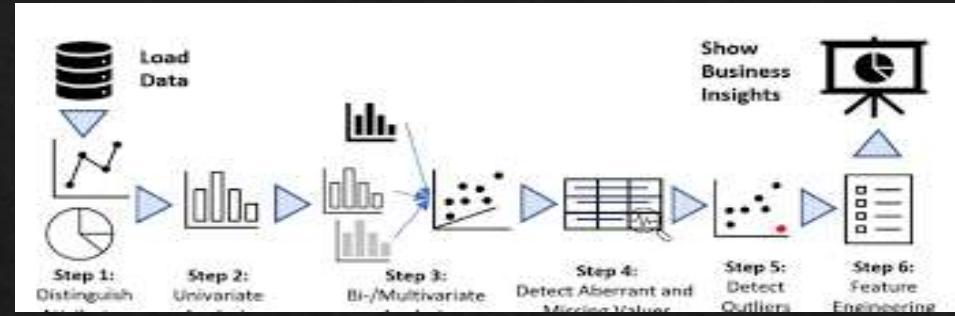
Hourly Merged Df!

\_df.head(2)

ID	ActivityMinute	Calories	intensity	Steps	METS
1	4/12/2016 12:00:00 AM	0.7865	0	0	10
1	4/12/2016 12:01:00 AM	0.7865	0	0	10

Minutes Merged Df!

# EDA USING GOOGLE COLAB



RAW DATA

CORRELATION  
OF VARIABLES

VISUALIZATIONS

DESCRIPTIVE  
ANALYSIS

IDENTIFY  
ATYPICAL  
DATA

GENERATE  
INSIGHTS

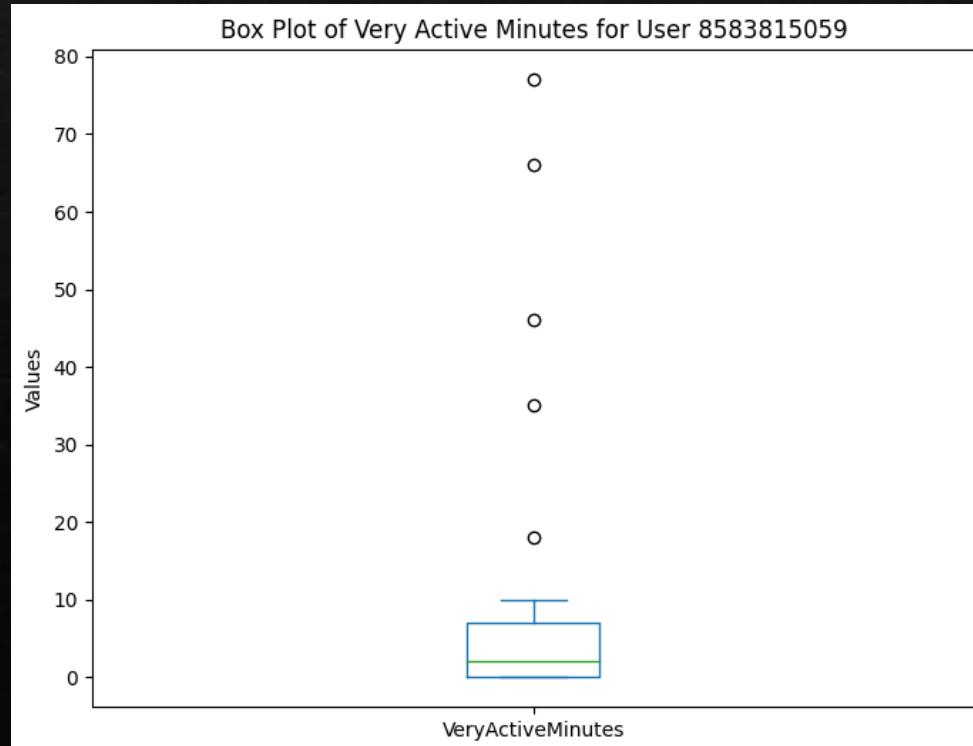
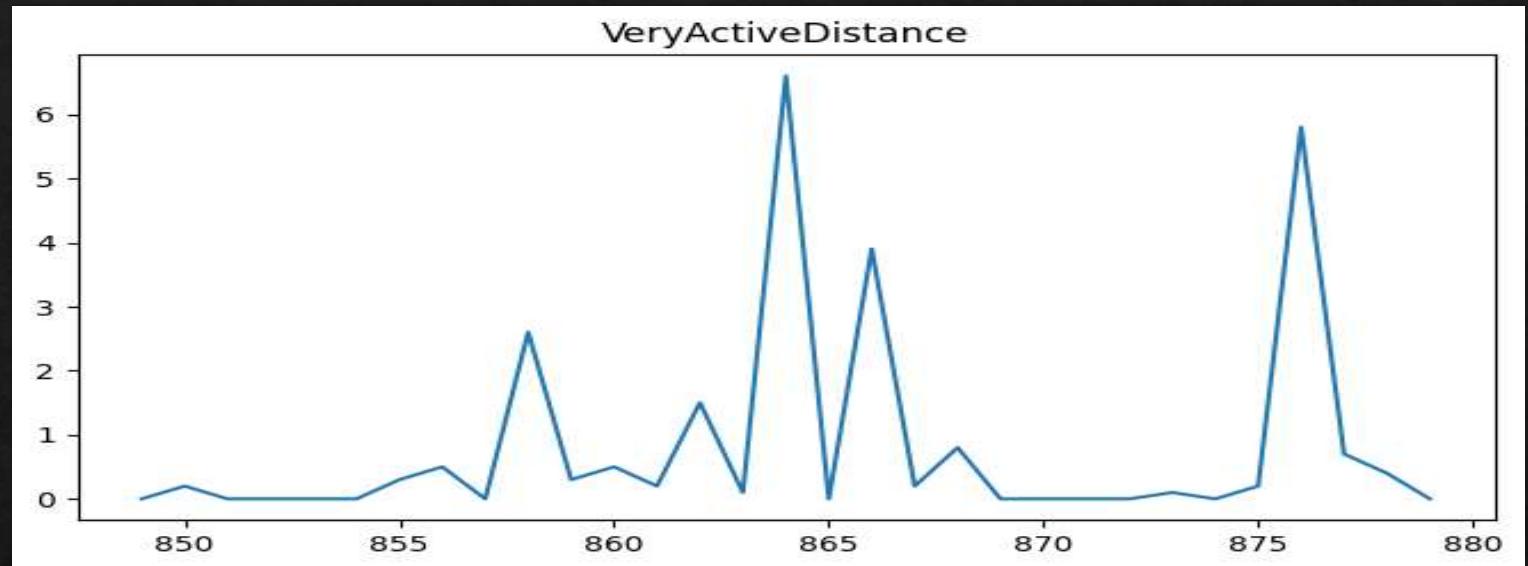
ADJUSTING  
VARIABLE  
TYPES

DETECTION OF  
MISSING  
VALUES

CONCLUSIONS

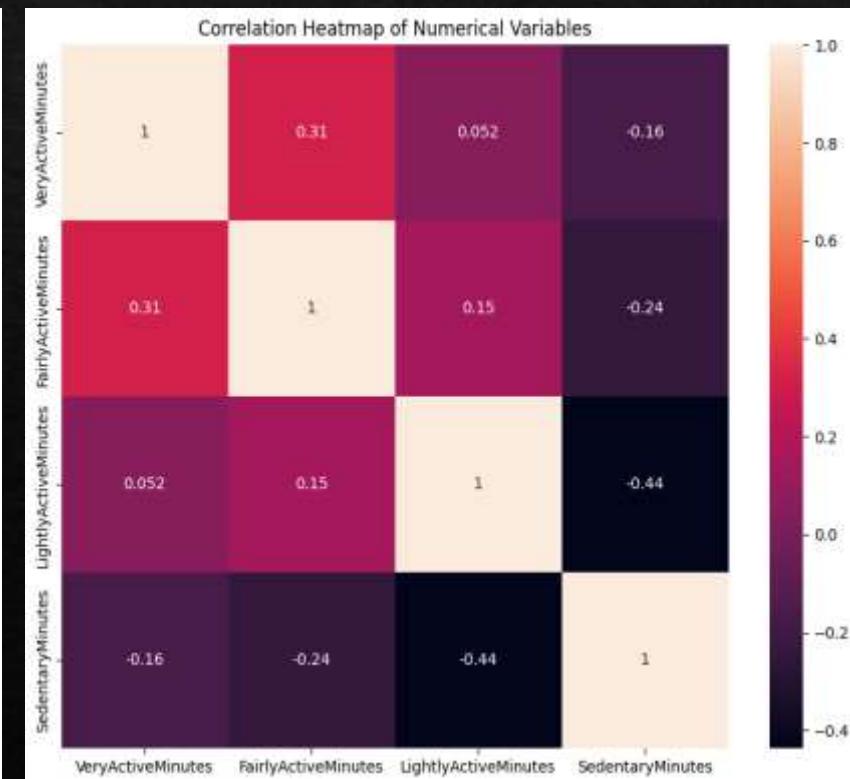
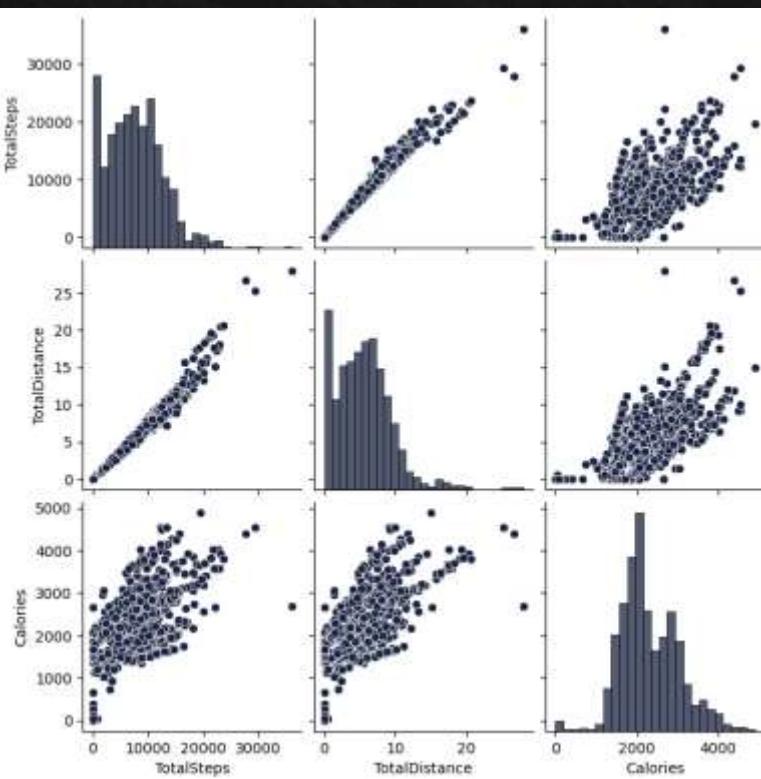
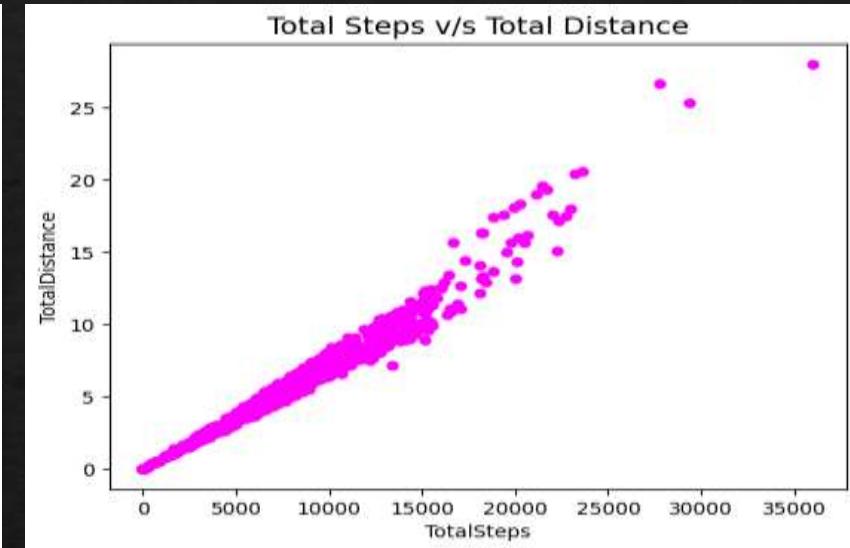
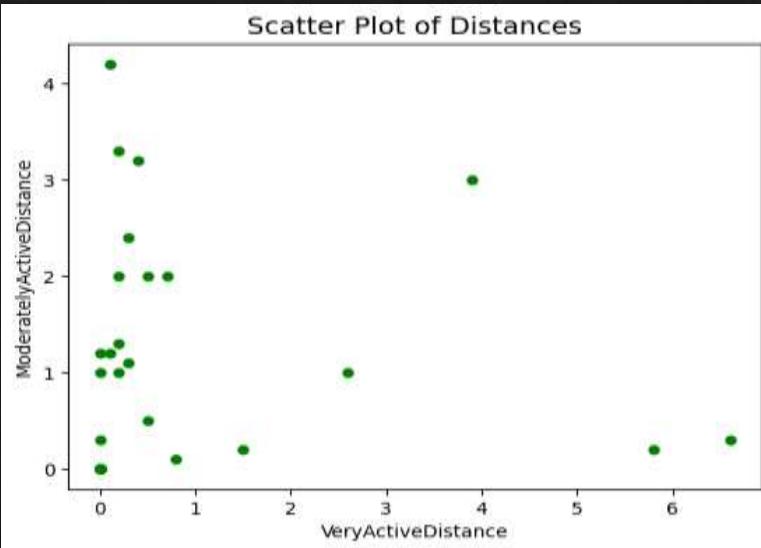
# Daily Activities Merged EDA

- Line charts and 1D Scatter plots are popular choices for visualizations in univariate analysis.
- Boxplots, Kernel plots & Dot plots are other visualizations that provide information & help us analyze univariate data.



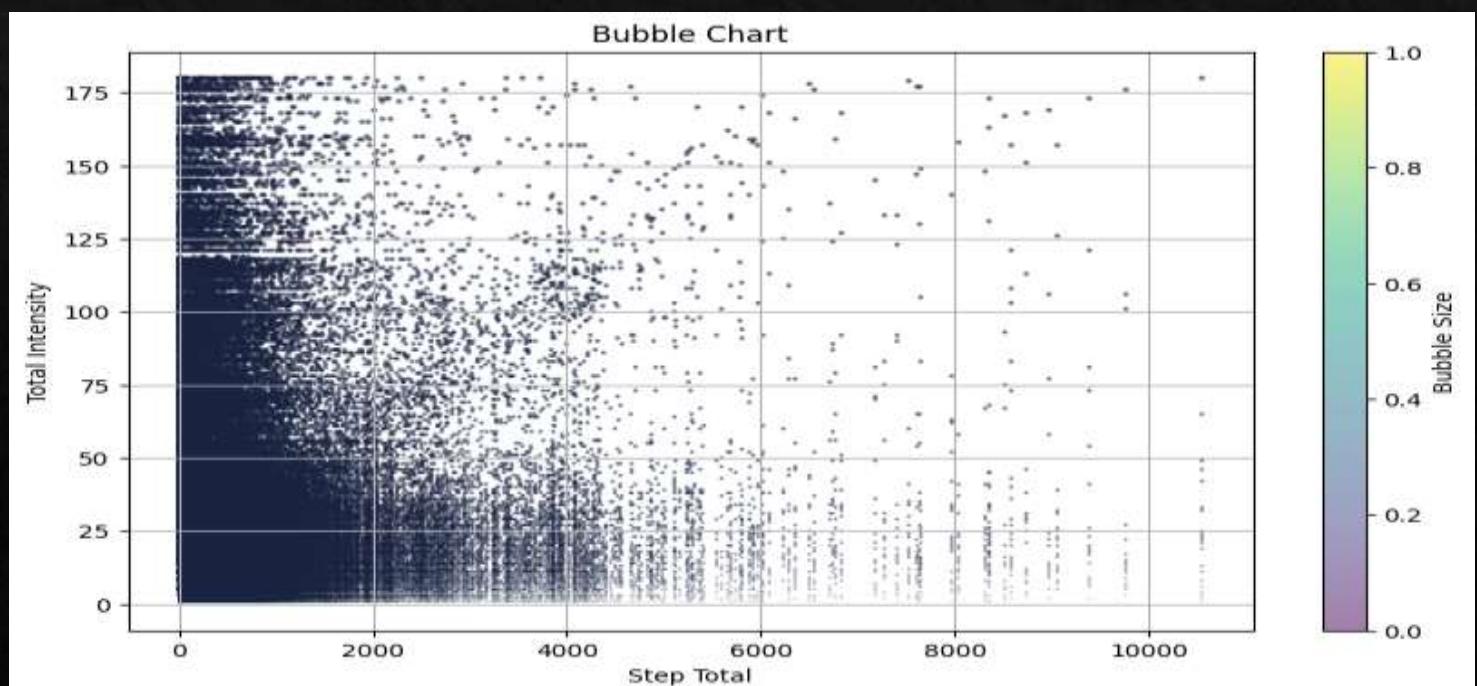
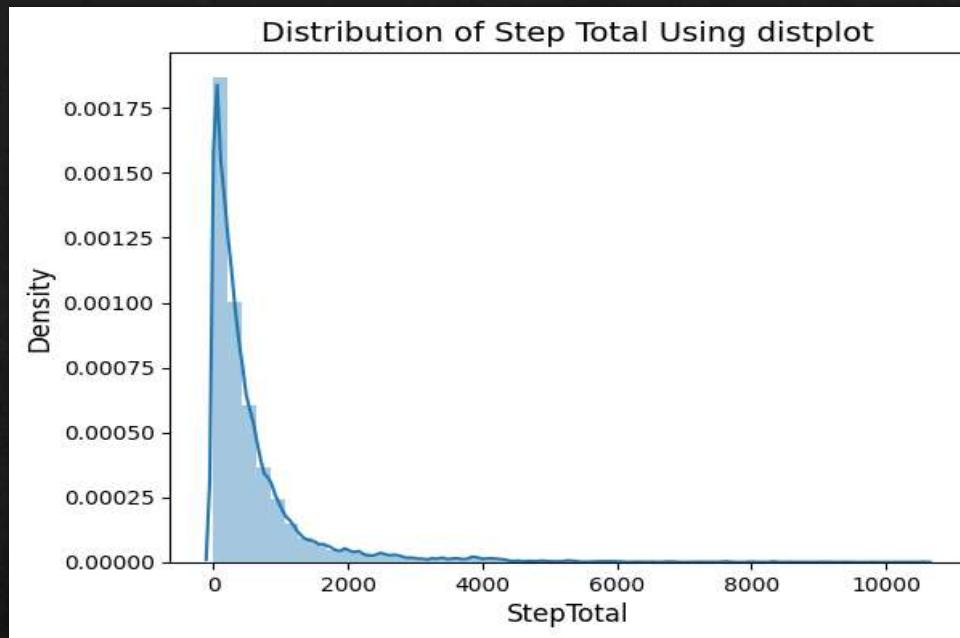
# Daily Activities Merged EDA

- I was able to analyze the trends & patterns for various Distances available in the df using scatter plots.
- Pair plots & heatmaps are preferred visualizations for multi-variate analysis. The closer the value of the matrix is to 1, the stronger is the correlation.



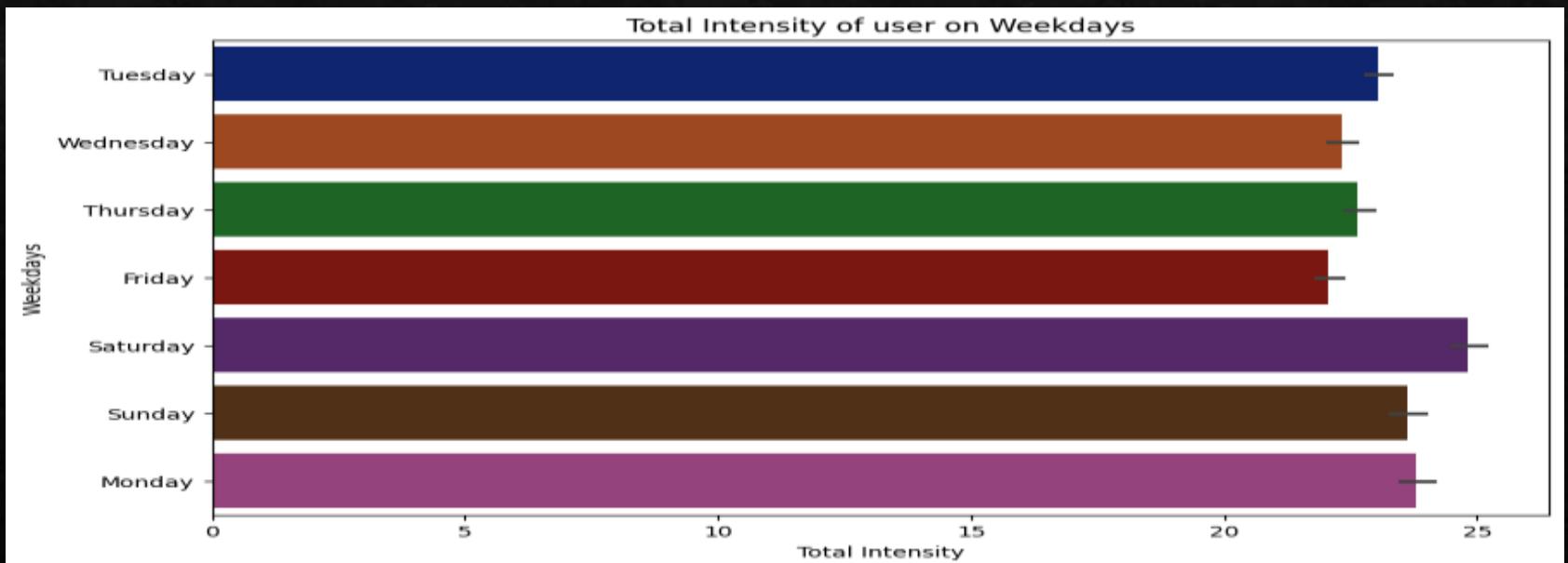
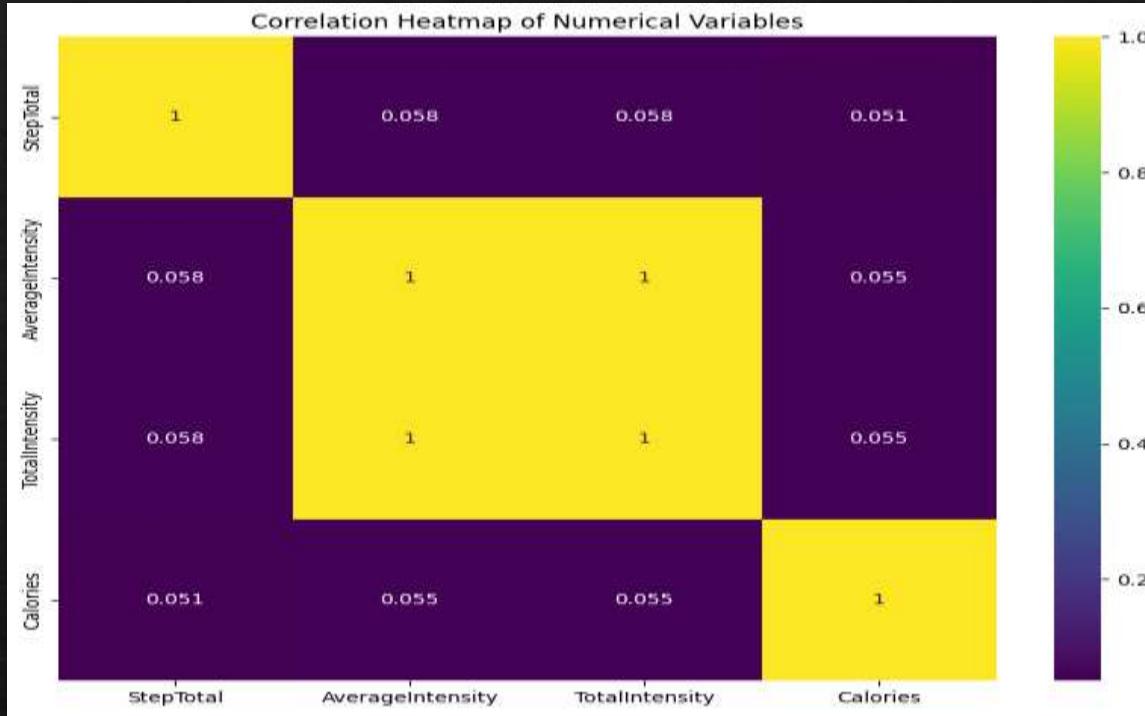
# Hourly Activities Merged EDA

- ❑ 1D Scatter plots, Distribution Plots & Kernel density plots provide amazing visualizations for univariate analysis.
- ❑ Tried bubble chart visualization to detect the relationship between Total Intensity & Total steps.
- ❑ Plotted bar graphs to help display the visualizations for Total Calories burnt on Weekdays, determine the max total steps & min intensity on different days.



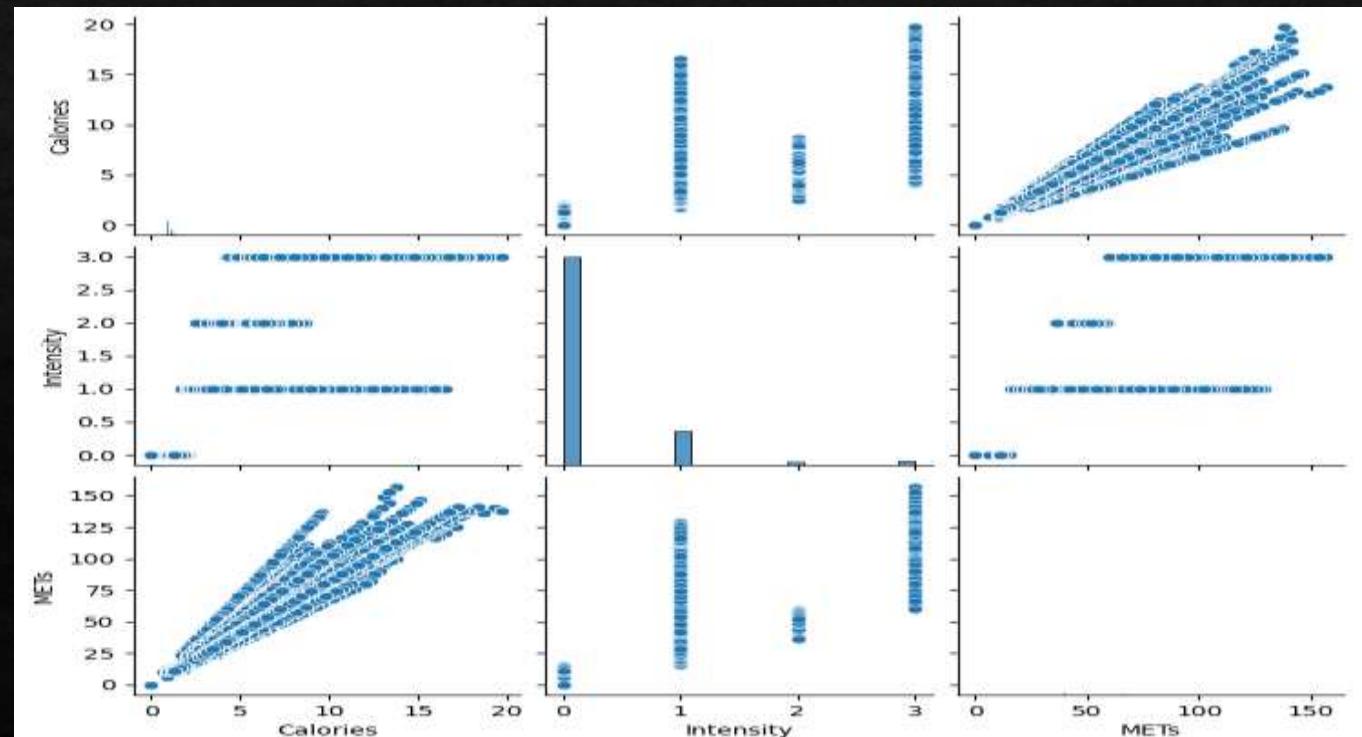
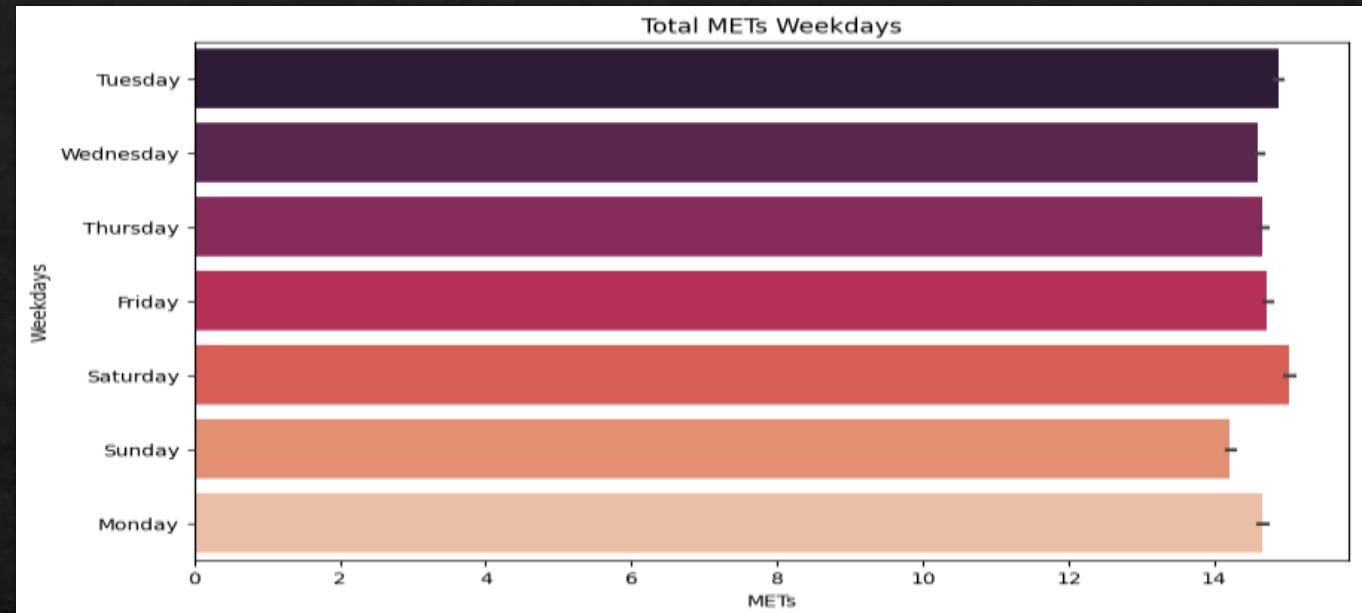
# Hourly Activities Merged EDA

- Line charts provide optimized visualization to determine the trend between numerical columns.
- Plotted correlation matrix heatmaps to help determine the relationships among Calories, Steps & different levels of Intensity.



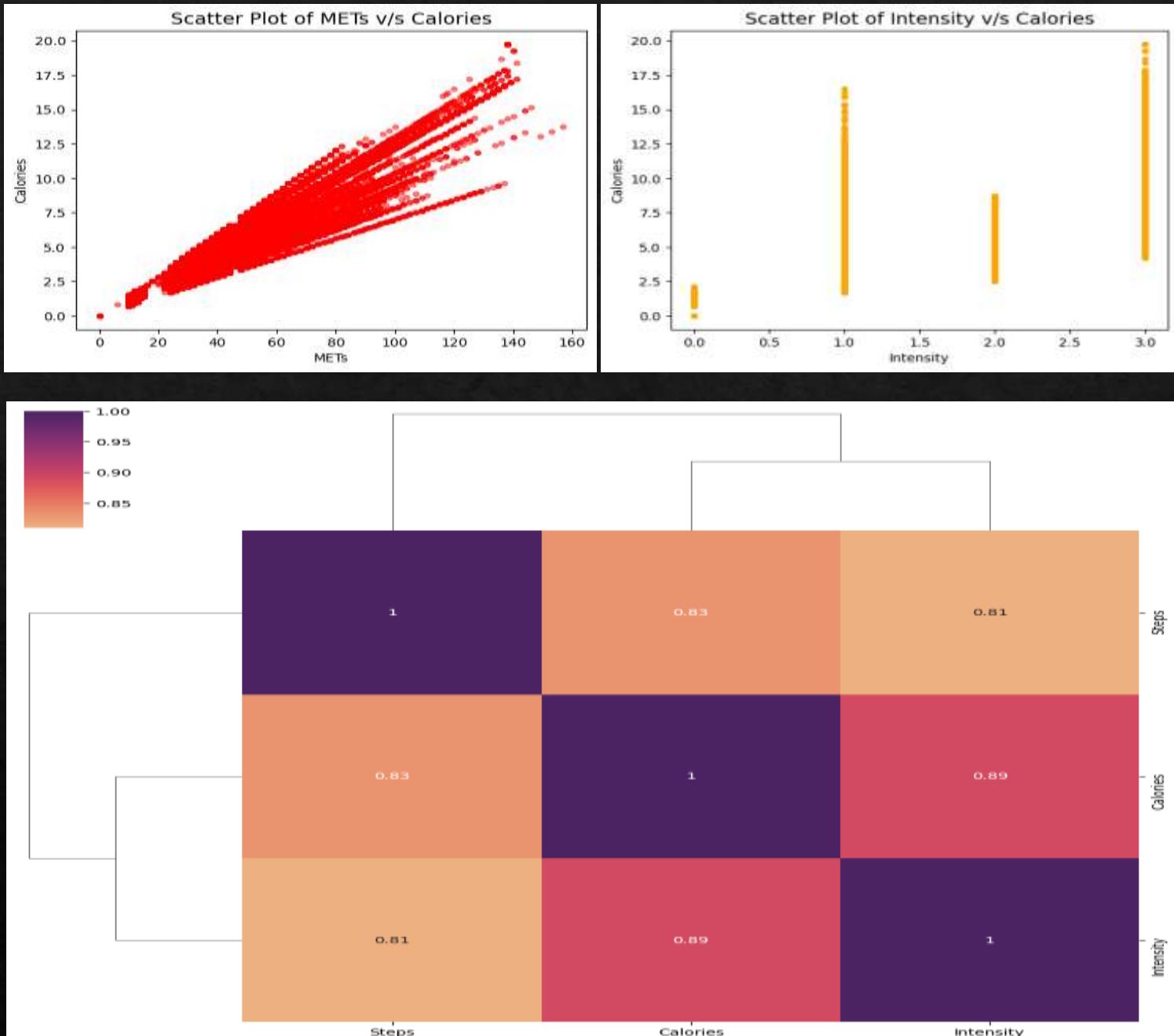
# Minutes Activities Merged EDA

- Plotted Distribution Plots for univariate Analysis.
- Created Bar charts to display the max & min METs on Weekdays.
  - Clustered heatmaps help understand clustering & correlation among Calories, Steps & Intensity.



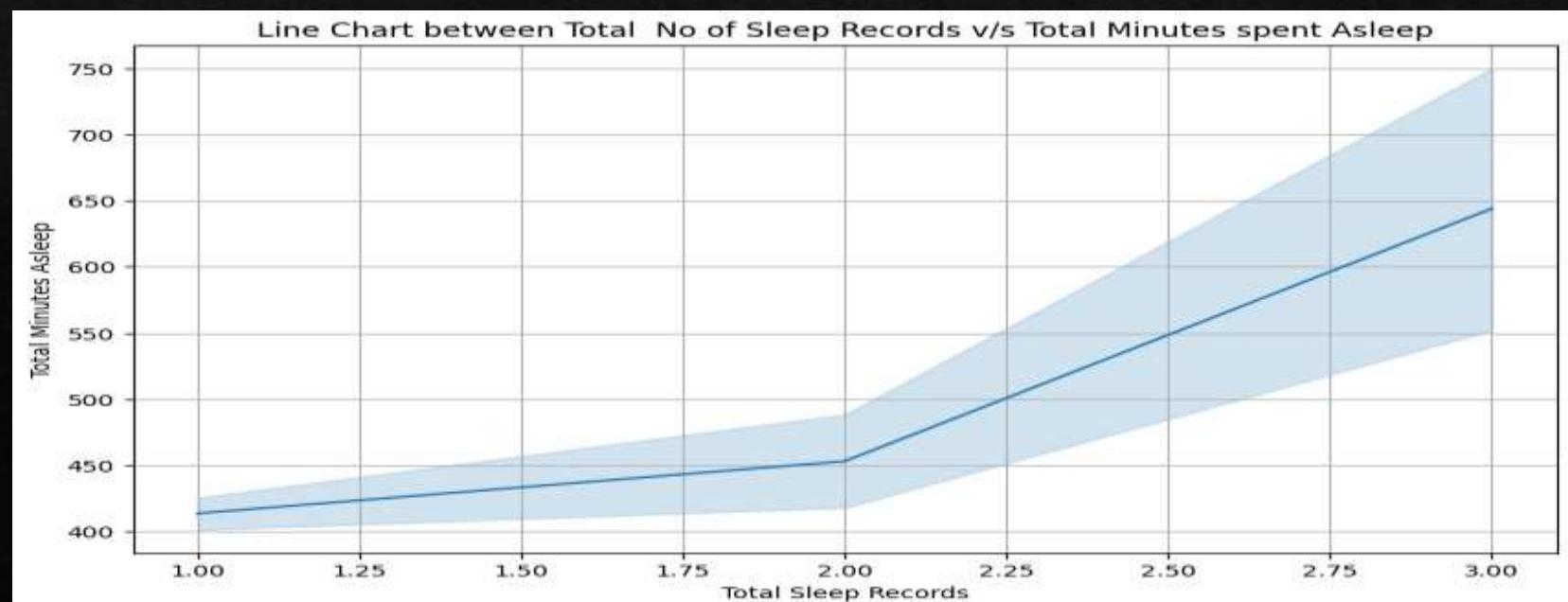
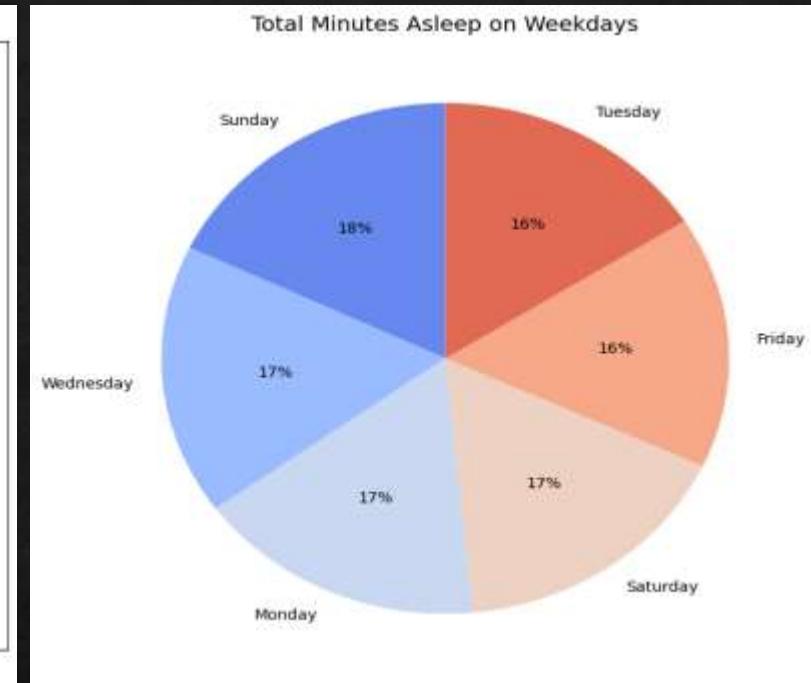
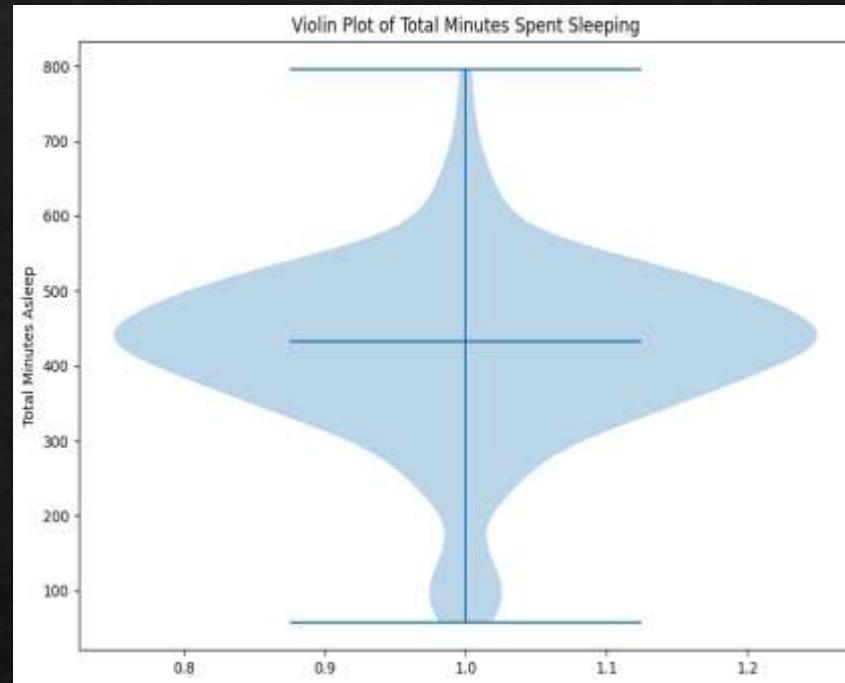
# Minutes Activities Merged EDA

- Plotted Scatter Plots for bivariate analysis.
- Created Pair Plots to determine relationships & dependency among various columns such as Calories, METs & Intensity.
  - Plotted heat maps to help understand the correlation among Calories, METs, Steps & Intensity.



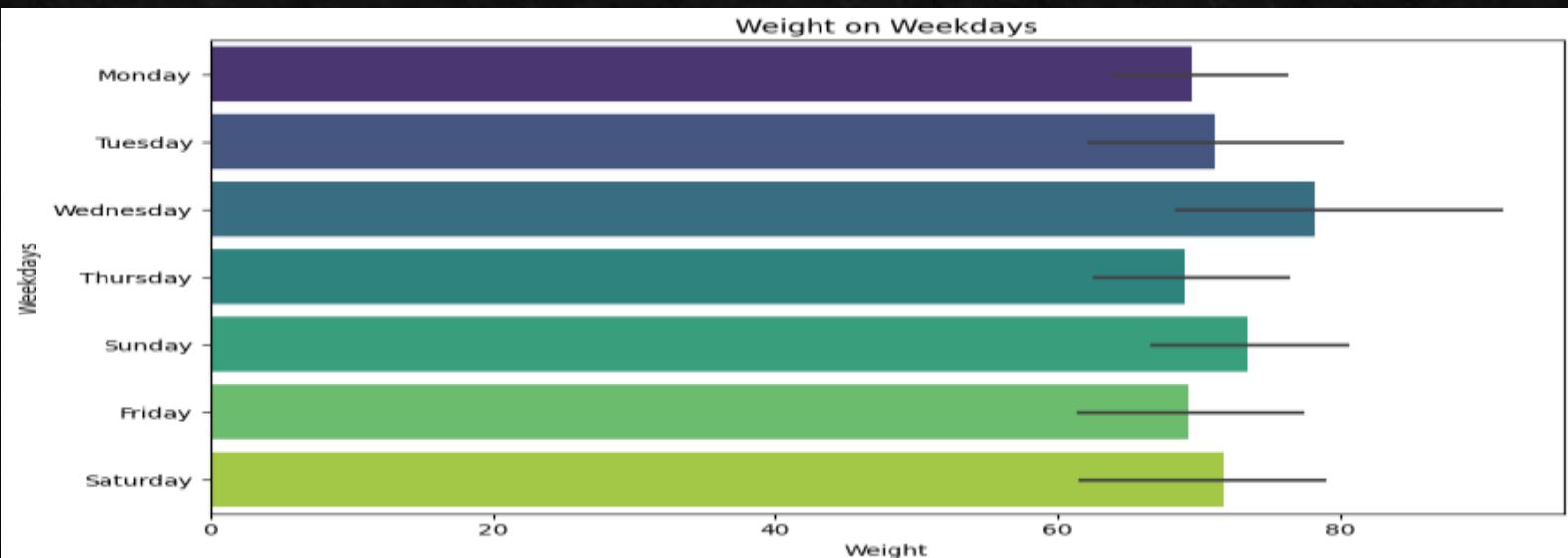
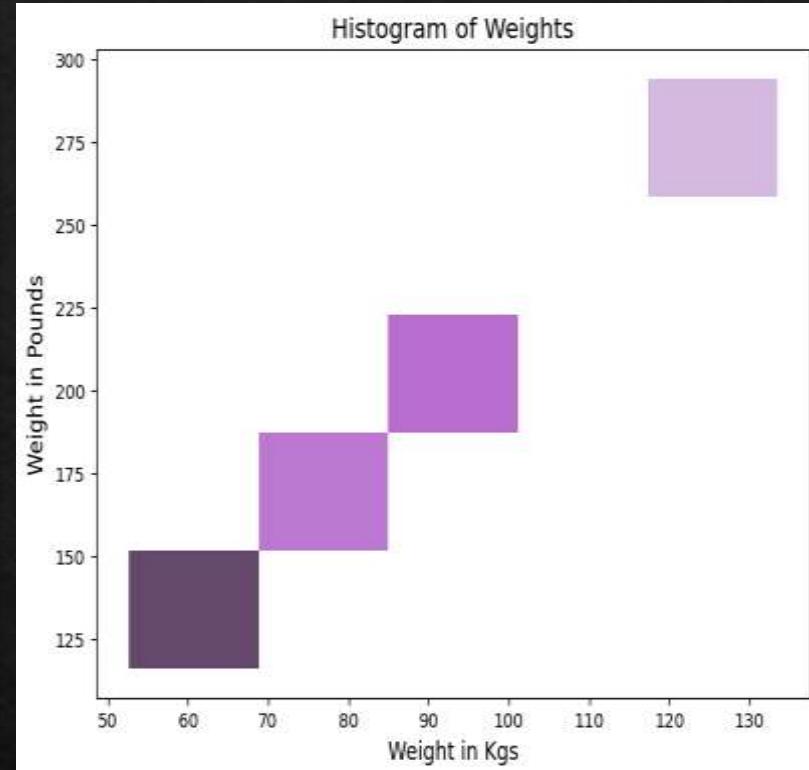
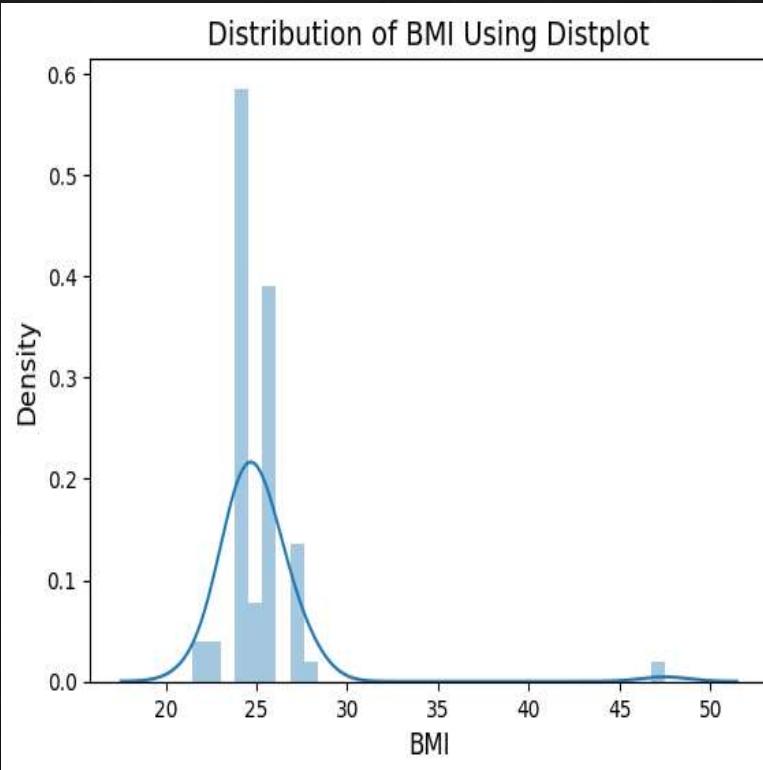
# Sleep Dataset EDA

- In sleep datasets, I used violin plot along with scatter plots & line charts for univariate analysis. Violin plot is a hybrid of a boxplot & a kernel density plot.
- I also applied the knowledge & distribution gathered from a pie chart for data on different days.
- Other visualizations used for analyzing sleep data were line charts, area charts, bar graphs & correlation matrices for looking further into the dependency among the sleep records.



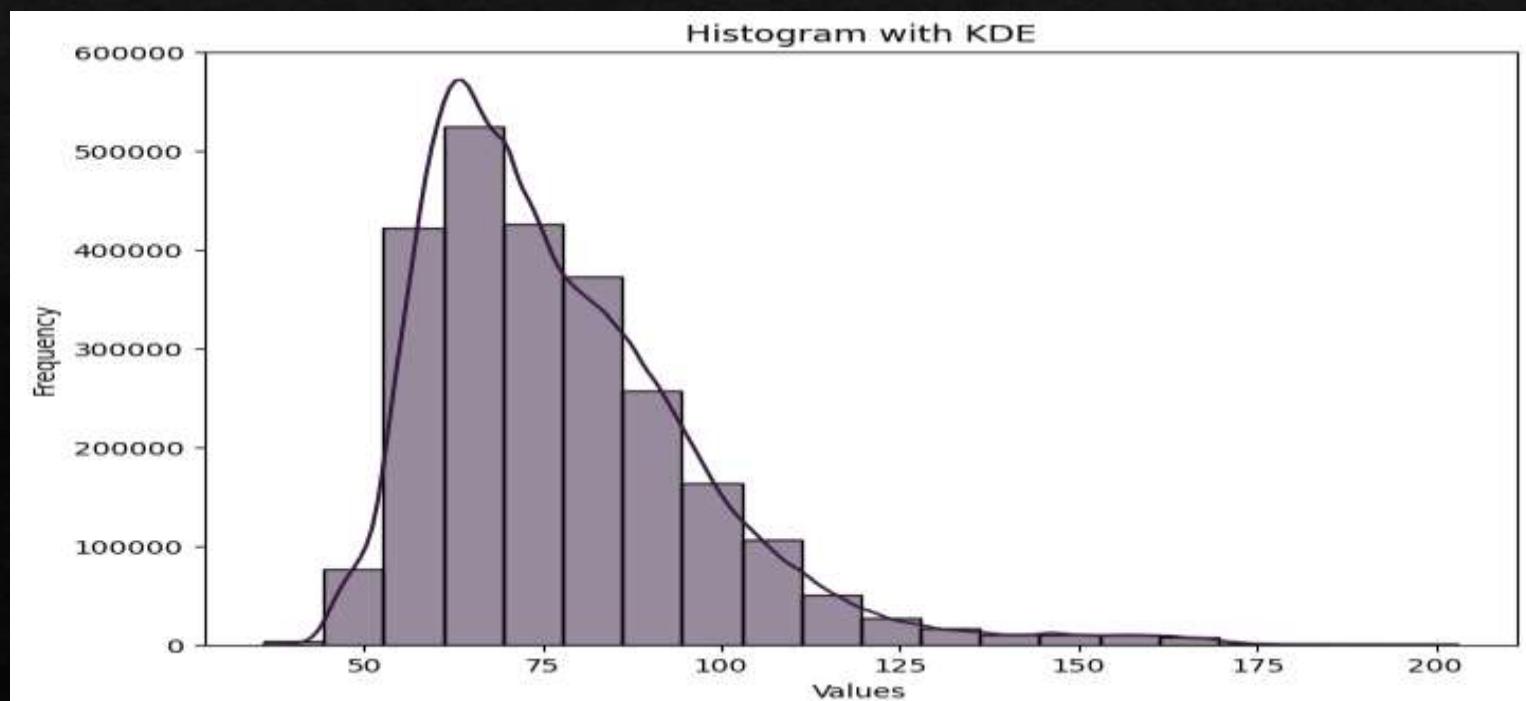
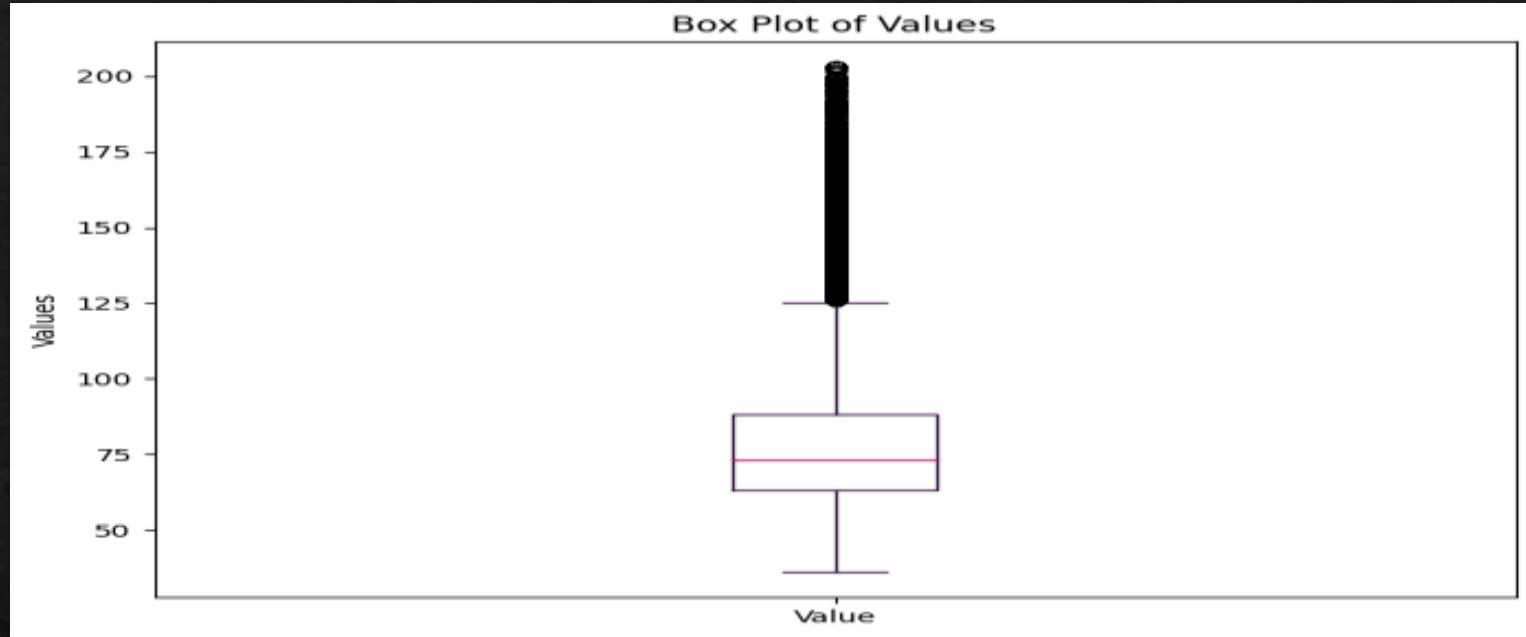
# Weight Dataset EDA

- Weight dataset consists of columns such as Weights in kg & pounds, BMI and Manual Report (Boolean values ).
- Applied the concept of distribution plots, KDE & scatter plots for univariate analysis.
- Other visualizations used for analyzing weight data were histograms, bar graphs, correlation matrices & pair plots for looking further into the dependency among the weight records



# Heart Rate Dataset EDA

- In heart rate datasets, we have 1 numerical column – Value. I have tried to understand the spread & density for that column using histograms and KDE plots.
- Boxplots are also used to detect the presence of outliers in the data frame.
- Bar charts have helped us identify the max & min total values on Weekdays signifying Saturday as the day when heart rate values are highest.



# TABLEAU DASHBOARDS VIZ

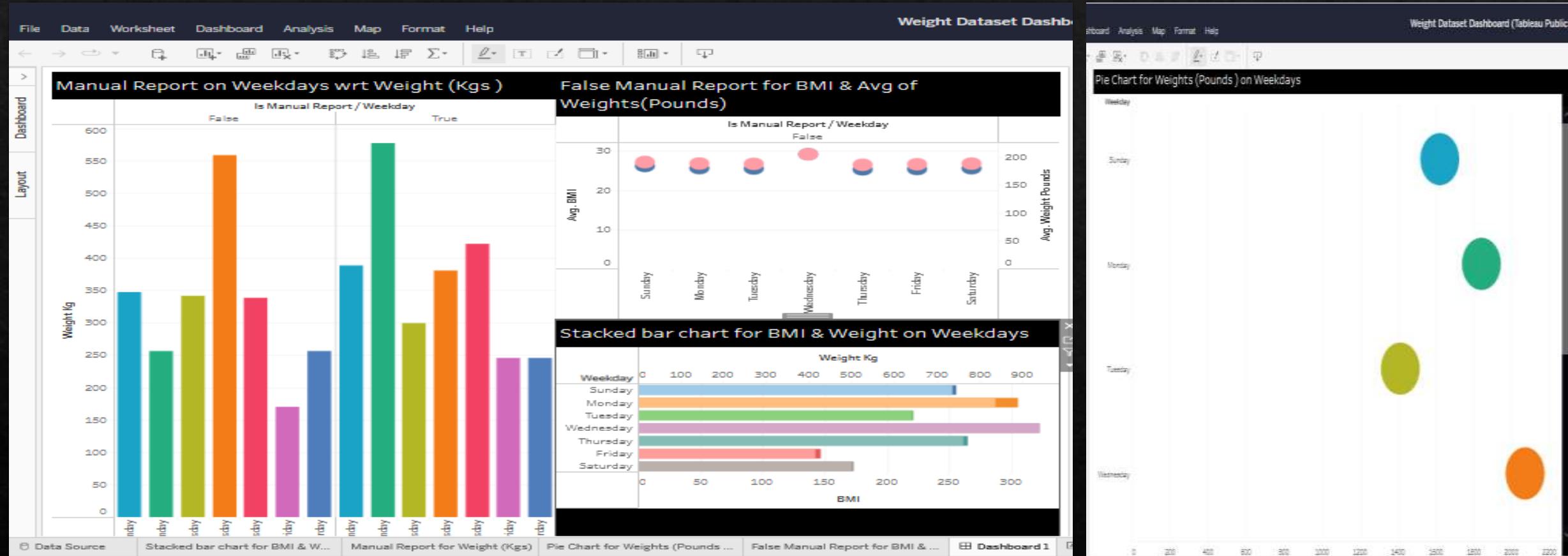
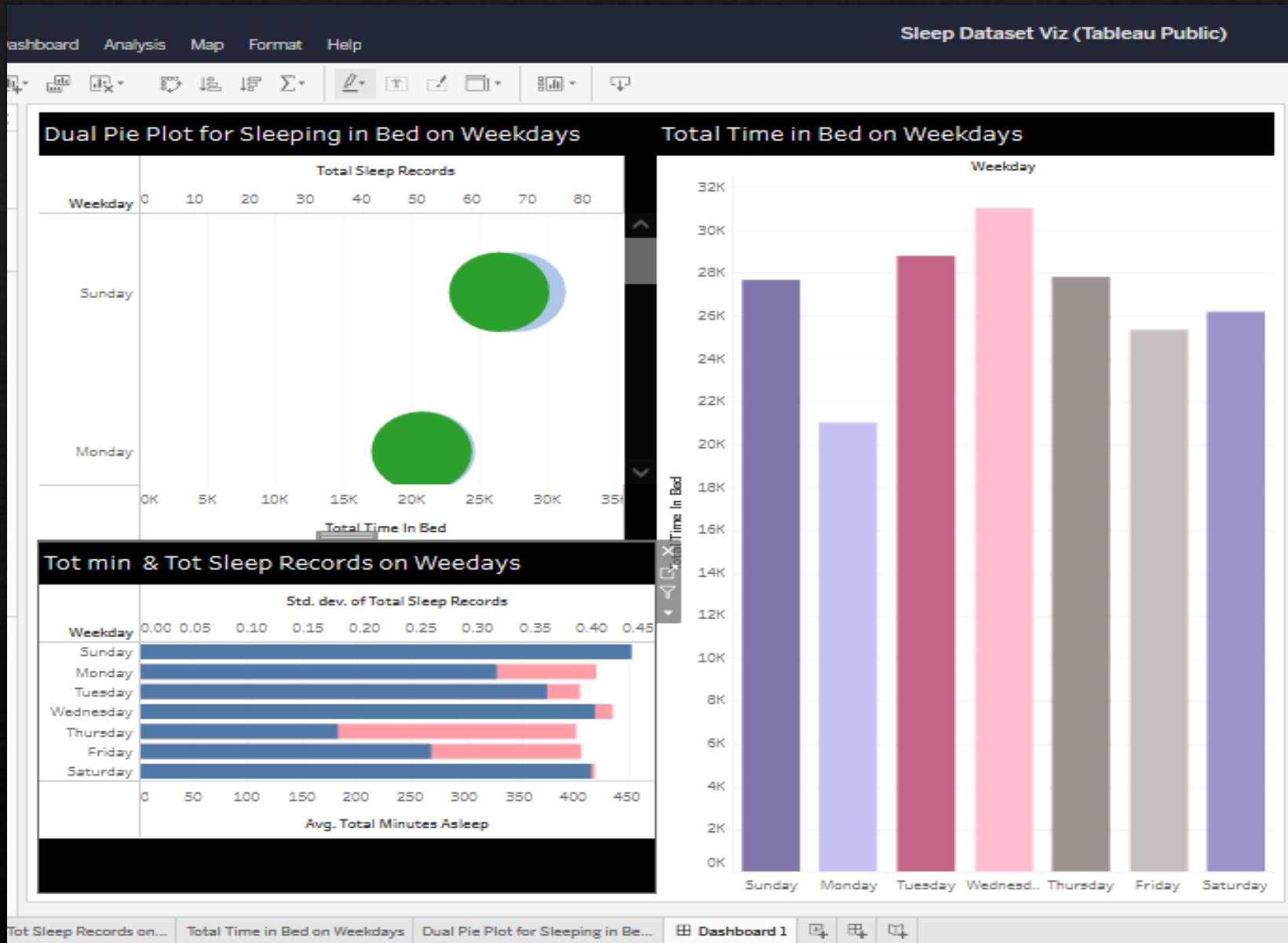


Tableau for Weight Data Viz:

<https://public.tableau.com/app/profile/deeksha.raina/viz/WeightDatasetDashboard/Dashboard1>

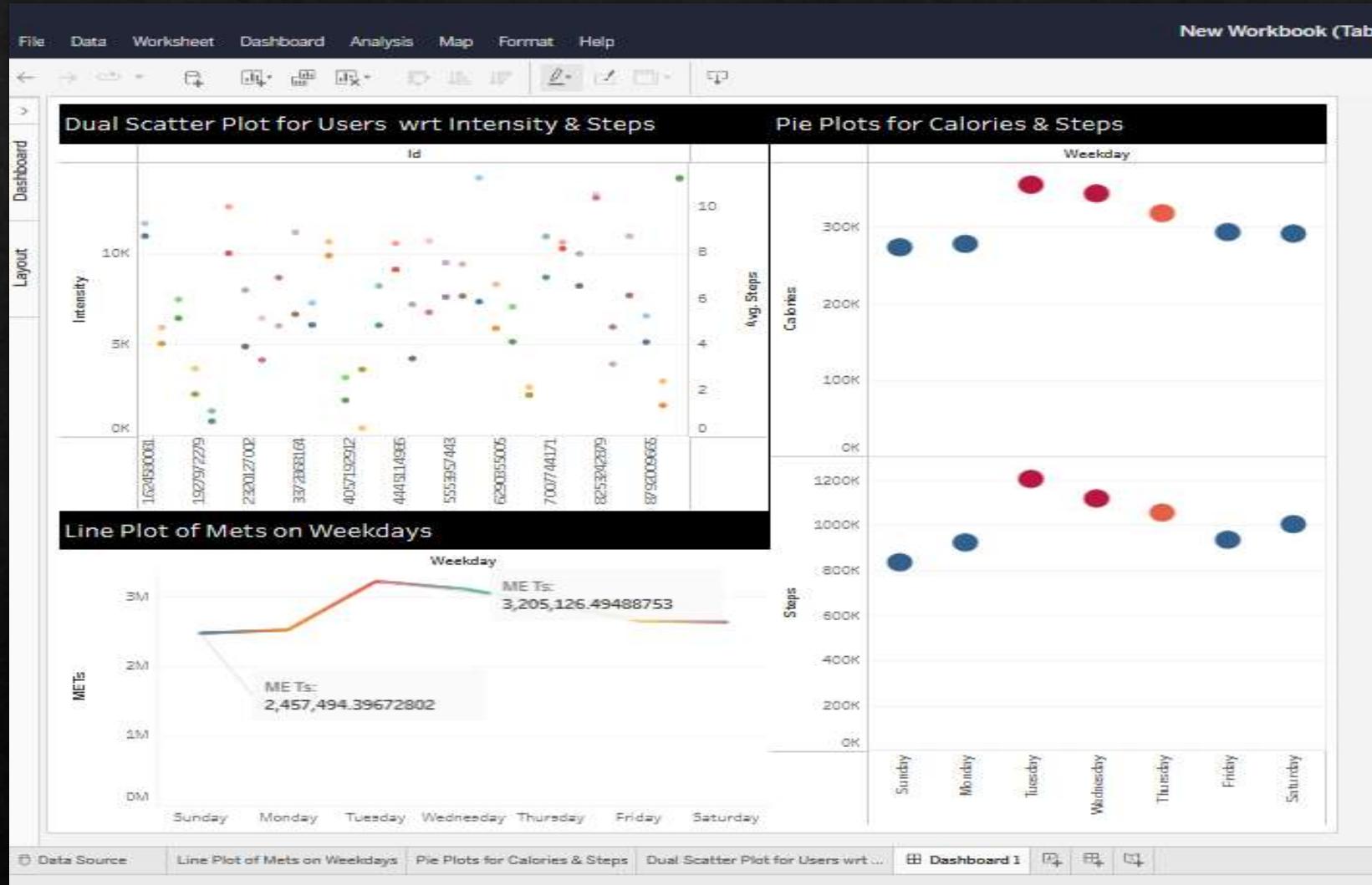
By Deeksha Raina

Tableau for Sleep Data Viz : <https://public.tableau.com/authoring/SleepDatasetViz/Dashboard1#1>



By Deeksha Raina

Tableau Link for Minute Data:  
<https://public.tableau.com/authoring/MinuteDatasetViz/Dashboard1#1>



By Deeksha Raina

# Tableau Viz for Hourly Data



By Deeksha Raina



# Insights & Conclusions



- ❖ From the data provided to us, we could help analyze various aspects of the fitness of users wrt different parameters.
- ❖ Explore activity patterns over time (daily, minute, hourly) to identify trends, peak activity times, and overall activity levels.
- ❖ Perform correlation analysis to identify relationships between different variables and understand how they influence each other.
- ❖ Analyze whether users are meeting their fitness goals based on their activity levels, calorie expenditure, and sleep patterns.
- ❖ To draw conclusions based on a fitness dataset containing various attributes we need to analyze the data and insights presented. We may evaluate the overall fitness levels of users based on metrics such as BMI, weight, total steps, and activity intensity. This can help identify whether users are meeting recommended fitness guidelines and maintaining a healthy lifestyle.
- ❖ Analyzing activity patterns over time to identify trends and peak activity periods. This could indicate whether users are consistently engaging in physical activity or if there are periods of inactivity.
- ❖ Segment users based on their fitness behaviors, activity levels, or other relevant factors to tailor recommendations or interventions. This can help personalize the fitness experience and improve engagement.