ACKNOWLEDGEMENT

I would like to express my sincere gratitude and appreciation to all those who have contributed to the successful completion of my internship and the preparation of this report. Special thanks go to my supervisor, Prof. Ankita Chauhan, for their guidance, support, and valuable insights throughout my internship journey.

I would also like to extend my thanks to the entire team at Medtoureasy for their warm welcome, cooperation, and for providing me with a conducive learning environment. Their willingness to share knowledge and expertise has been instrumental in my personal and professional growth.

I am indebted to my professors and mentors at Madhuben & Bhanubhai Patel Institute of Technology for their continuous support, encouragement, and for equipping me with the necessary skills and knowledge to excel in my internship.

I am truly grateful to everyone who has played a part in my internship journey, and I acknowledge their contributions with deep gratitude.

ABSTRACT

The growing availability of wearable fitness devices and mobile applications has led to an exponential increase in the collection of personal fitness data. This wealth of information provides a unique opportunity to delve into individuals' physical activity patterns and gain valuable insights into their overall health and well-being. This study aims to analyze fitness data to uncover meaningful patterns, trends, and associations that can contribute to a deeper understanding of personal fitness habits and their impact on health outcomes.

The analysis begins by aggregating and anonymizing a diverse dataset comprising various fitness metrics such as step counts, heart rate, calories burned, distance covered, and sleep patterns. Exploratory data analysis techniques are employed to identify initial trends and outliers, ensuring data quality and integrity. Subsequently, advanced statistical methods, including regression analysis, clustering, and time-series analysis, are applied to unveil patterns and relationships between different variables.

The research investigates potential correlations between fitness data and individual characteristics such as age, gender, body mass index (BMI), and lifestyle factors. By analyzing large-scale data sets, it becomes possible to identify common trends across populations, thereby enabling the development of tailored fitness recommendations and interventions.

The findings from this analysis have the potential to impact various domains, including preventive healthcare, personalized fitness coaching, and the design of user-friendly fitness applications. Understanding how fitness data relates to overall health and well-being will empower individuals to make proactive lifestyle choices and potentially mitigate the risk of chronic diseases associated with sedentary behavior.

Overall, this study aims to harness the power of fitness data analysis to unlock valuable insights into human physical activity patterns, with the goal of promoting healthier and more active lifestyles for individuals across diverse demographics.

Table of Content

1. INTRODUCTION

1.1. ABOUT THE PROJECT

The Project is of Analysis of Fitness Data. With the explosion in fitness tracker popularity, runners all of the world are collecting data with gadgets (smartphones, watches, etc.) to keep themselves motivated. They look for answers to questions like:

- How fast, long, and intense was my run today?
- Have I succeeded with my training goals?
- Am I progressing?
- What were my best achievements?
- How do I perform compared to others?

This data was exported from Runkeeper. The data is a CSV file where each row is a single training activity. In this project, you'll create import, clean, and analyze my data to answer the above questions.

The RunKeeper dataset is a collection of fitness tracking data gathered from the RunKeeper app, which is designed for tracking various outdoor activities such as running, walking, cycling, and hiking. It contains information recorded by users during their workout sessions, including details like distance travelled, duration, speed, pace, elevation, calories burned, and GPS coordinates.

The dataset is typically organized in a tabular format, with each row representing a single activity session and columns representing different attributes and measurements associated with that session. In addition to the numeric data, it may also include categorical information such as activity type, weather conditions, user demographics, and other user-provided tags. It contains 508 rows and 13 columns. The columns are Date, Activity ID, Type, Distance, Duration, Calories, Average Heart Rate, etc. The dataset is commonly used in research and analysis related to fitness tracking, physical activity, and sports analytics.

1.2. OBJECTIVES AND DELIVERABLES

Objectives:

1. Obtain and review raw data
2. Data preprocessing
3. Dealing with missing values
4. Plot running data
5. Running statistics
6. Visualization with averages
7. Did I reach my goals?
8. Am I progressing?
9. Training intensity
10. Detailed summary report
11. Fun facts

Deliverables:

1. Obtain and review raw data:
   a. Documented sources of raw fitness data.
   b. Raw data files or access to the data sources.

2. Data preprocessing:
   a. Cleaned and standardized fitness data.
   b. Removed duplicates and irrelevant data.
   c. Standardized data formats and units.

3. Dealing with missing values:
   a. Identification and handling of missing values in the data.
   b. Imputation techniques used to fill missing values.

4. Plot running data:
   a. Visualizations of running data, such as distance, time, pace, or heart rate, over a specific period.

   b. Line plots, scatter plots, or other relevant visualizations to represent the running data.

5. Running statistics:
   a. Calculation of summary statistics for running data, such as total distance, average pace, maximum heart rate, etc.
   b. Statistical analysis to identify patterns or trends in the running data.

6. Visualization with averages:
   a. Visualizations that include average running data over time, such as weekly or monthly averages.
   b. Comparative visualizations of different running metrics with corresponding averages.

7. Did I reach my goals?
   a. Analysis to determine if fitness goals were achieved.
   b. Comparison of actual performance against set goals or targets.

8. Am I progressing?
   a. Evaluation of fitness progress over time.
   b. Trend analysis to identify improvements or declines in running performance.

9. Training intensity:
   a. Assessment of training intensity based on running data.
   b. Identification of high-intensity or low-intensity training periods.

10. Detailed summary report:
    a. Comprehensive report summarizing the findings of the analysis.
    b. Insights and recommendations based on the analyzed fitness data.

11. Fun facts:

a. Additional interesting and engaging information or insights derived from the data analysis.
b. Unusual or noteworthy patterns, achievements, or statistics related to fitness data.

These deliverables will help provide a comprehensive analysis of the fitness data, including data preprocessing, visualizations, goal achievement evaluation, progress tracking, and a detailed summary report. Additionally, incorporating fun facts will enhance the engagement and interest of the audience in the project result.

2. COMPANY PROFILE

2.1. ABOUT THE COMPANY

MedTourEasy, a global healthcare company, provides you the informational resources needed to evaluate your global options. MedTourEasy provides analytical solutions to our partner healthcare providers globally.

Global health is an area for study, research, and practice that places a priority on improving health and achieving equity in health for all people worldwide. Global health emphasizes transnational health issues, determinants, and solutions; involves many disciplines within and beyond the health sciences and promotes interdisciplinary collaboration; and is a synthesis of population- based prevention with individual-level clinical care.

3. TECHNOLOGY LEARNED

3.1. LANGUAGE AND PLATFORM USED

3.1.1. Language: Python

Python is a high-level, general-purpose programming language known for its simplicity and readability. It has a vast ecosystem of libraries and frameworks that make it suitable for various applications, including data analysis and visualization. Some key features of Python for data analysis are:

- Libraries: Python offers powerful libraries specifically designed for data analysis, such as pandas for data manipulation and analysis, NumPy for numerical computations, Matplotlib for creating visualizations, and scikit-learn for machine learning tasks.
- Easy-to-learn syntax: Python has a clean and intuitive syntax, making it easier for beginners to understand and write code.
- Interoperability: Python can integrate seamlessly with other programming languages, allowing users to leverage functionality from different tools and libraries.
- Active community support: Python has a large and active community of developers, which means there are plenty of resources, tutorials, and forums available to help with any questions or issues.

3.1.2. Platform: Jupyter Notebook

Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, visualizations, and narrative text. It provides an interactive computational environment that supports multiple programming languages, with Python being one of the most commonly used languages. Here are some key features of Jupyter Notebook:

- Interactive environment: Jupyter Notebook provides an interactive environment where you can write and execute code in cells. You can run code snippets and see the results immediately, which makes it ideal for exploratory data analysis.

- Mix of code and documentation: Jupyter Notebook allows you to include documentation, explanations, and visualizations alongside your code. This makes it easy to create comprehensive and self-explanatory data analysis reports.
- Rich output: Jupyter Notebook supports the display of rich media, including images, charts, and interactive widgets, enabling you to create dynamic and interactive visualizations.
- Collaboration and sharing: Jupyter Notebook allows you to share your notebooks with others, making it easy to collaborate on projects. Notebooks can be shared as static files or through platforms like GitHub, allowing for easy replication and collaboration.

4. OUTLINE OF WORK DONE DURING INTERNSHIP

4.1. WEEK-1

- What is Data Science? - Basic Overview and Topics
- Languages of Data science & Packages, APIs, Datasets and Models
- Data Science Lifecycle
- Python Data Structures & Working with Data in Python

4.2. WEEK-2

- Introduction to SQL-Data Science & Accessing Databases using Python
- Importing Datasets & Data Wrangling
- Basic Introduction to Data Visualization Tools & Data Visualization Libraries
- Machine Learning Models
- Types of Machine Learning

4.3. WEEK-3

- Project proposal and final report

5. IMPLEMENTATION AND SCREENSHOTS

5.1. IMPLEMENTATION

1. Obtain and review raw data

   One day, my old running friend and I were chatting about our running styles, training habits, and achievements, when I suddenly realized that I could take an in-depth analytical look at my training. I have been using a popular GPS fitness tracker called Runkeeper for years and decided it was time to analyze my running data to see how I was doing.

   Since 2012, I've been using the Runkeeper app, and it's great. One key feature: its excellent data export. Anyone who has a smartphone can download the app and analyze their data like we will in this notebook.

   After logging your run, the first step is to export the data from Runkeeper (which I've done already). Then import the data and start exploring to find potential problems. After that, create data cleaning strategies to fix the issues. Finally, analyze and visualize the clean time-series data.

   I exported seven years worth of my training data, from 2012 through 2018. The data is a CSV file where each row is a single training activity. Let's load and inspect it.

2. Data preprocessing

   Lucky for us, the column names Runkeeper provides are informative, and we don't need to rename any columns.

   But, we do notice missing values using the info() method. What are the reasons for these missing values? It depends. Some heart rate information is missing because I didn't always use a cardio sensor. In the case of the Notes column, it is an optional field that I sometimes left blank. Also, I only used the Route Name column once, and never used the Friend's Tagged column.

   We'll fill in missing values in the heart rate column to avoid misleading results later, but right now, our first data preprocessing steps will be to:

   - Remove columns not useful for our analysis.
   - Replace the "Other" activity type to "Unicycling" because that was always the "Other" activity.

- Count missing values.

3. Dealing with missing values

   As we can see from the last output, there are 214 missing entries for my average heart rate.We can't go back in time to get those data, but we can fill in the missing values with an average value. This process is called mean imputation. When imputing the mean to fill in missing data, we need to consider that the average heart rate varies for different activities (e.g., walking vs. running). We'll filter the DataFrames by activity type (Type) and calculate each activity's mean heart rate, then fill in the missing values with those means.

4. Plot running data

   Now we can create our first plot! As we found earlier, most of the activities in my data were running (459 of them to be exact). There are only 29, 18, and two instances for cycling, walking, and unicycling, respectively. So for now, let's focus on plotting the different running metrics.An excellent first visualization is a figure with four subplots, one for each running metric (each numerical column). Each subplot will have a different y-axis, which is explained in each legend.

5. Running statistics

   No doubt, running helps people stay mentally and physically healthy and productive at any age. And it is great fun! When runners talk to each other about their hobby, we not only discuss our results, but we also discuss different training strategies.

   You'll know you're with a group of runners if you commonly hear questions like:

   - What is your average distance?
   - How fast do you run?
   - Do you measure your heart rate?
   - How often do you train?

   Let's find the answers to these questions in my data. If you look back at plots in Task 4, you can see the answer to, Do you measure your heart rate? Before 2015: no. To look at the averages, let's only use the data from 2015 through 2018.

In pandas, the resample() method is similar to the groupby() method - with resample() you group by a specific time span. We'll use resample() to group the time series data by a sampling period and apply several methods to each sampling period. In our case, we'll resample annually and weekly.

6. Visualization with averages

   Let's plot the long term averages of my distance run and my heart rate with their raw data to visually compare the averages to each training session. Again, we'll use the data from 2015 through 2018.In this task, we will use matplotlib functionality for plot creation and customization.

7. Did I reach my goals?

   To motivate myself to run regularly, I set a target goal of running 1000 km per year. Let's visualize my annual running distance (km) from 2013 through 2018 to see if I reached my goal each year. Only stars in the green region indicate success.

8. Am I progressing?

   Let's dive a little deeper into the data to answer a tricky question:
   Am I progressing in terms of my running skills?To answer this question, we'll decompose my weekly distance run and visually compare it to the raw data. A red trend line will represent the weekly distance run.We are going to use statsmodels library to decompose the weekly trend.

9. Training intensity

   Heart rate is a popular metric used to measure training intensity. Depending on age and fitness level, heart rates are grouped into different zones that people can target depending on training goals. A target heart rate during moderate-intensity activities is about 50-70% of maximum heart rate, while during vigorous physical activity it’s about 70-85% of maximum.We'll create a distribution plot of my heart rate data by training intensity. It will be a visual presentation for the number of activities from predefined training zones.

10. Detailed summary report

With all this data cleaning, analysis, and visualization, let's create detailed summary tables of my training.To do this, we'll create two tables. The first table will be a summary of the distance (km) and climb (m) variables for each training activity. The second table will list the summary statistics for the average speed (km/hr), climb (m), and distance (km) variables for each training activity.

11. Fun facts

To wrap up, let’s pick some fun facts out of the summary tables and solve the last exercise. These data (my running history) represent 6 years, 2 months and 21 days. And I remember how many running shoes I went through–7.

FUN FACTS

Average distance: 11.38 km

Longest distance: 38.32 km

Highest climb: 982 m

Total climb: 57,278 m

Total number of km run: 5,224 km

Total runs: 459

Number of running shoes gone through: 7 pairs

The story of Forrest Gump is well known–the man, who for no particular reason decided to go for a "little run." His epic run duration was 3 years, 2 months and 14 days (1169 days). In the picture you can see Forrest’s route of 24,700 km.

FORREST RUN FACTS

Average distance: 21.13 km

Total number of km run: 24,700 km

Total runs: 1169

Number of running shoes gone through: ...

Assuming Forest and I go through running shoes at the same rate, figure out how many pairs of shoes Forrest needed for his run.

5.2. SCREENSHOTS

1. *Obtain and review raw data*

Date	Activity Id	Type	Route Name	Distance (km)	Duration	Average Pace	Average Speed (km/h)	Calories Burned	Climb (m)	Average Heart Rate (bpm)	Friend's Tagged	Notes	GPX File
2017-08-10 18:28:54	7b11bced-0b12-4e02-9717-09587f4c715f	Running	NaN	8.43	43:48	5:12	11.54	596.0	86	147.0	NaN	TomTom MySports Watch	2017-08-10-182854.gpx
2016-05-22 15:31:06	d6e2c045-ce5b-49be-846e-a59df6bd4f0d	Cycling	NaN	39.62	1:48:59	2:45	21.81	877.0	319	NaN	NaN	NaN	2016-05-22-153106.gpx
2018-11-04 16:05:00	c09b2f92-f855-497c-b624-c196b3ef036c	Running	NaN	13.01	1:15:16	5:47	10.37	967.0	171	155.0	NaN	NaN	2018-11-04-160500.gpx

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 508 entries, 2018-11-11 14:05:12 to 2012-08-22 18:53:54
Data columns (total 13 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Activity Id               508 non-null    object
 1   Type                      508 non-null    object
 2   Route Name                1 non-null      object
 3   Distance (km)             508 non-null    float64
 4   Duration                  508 non-null    object
 5   Average Pace              508 non-null    object
 6   Average Speed (km/h)      508 non-null    float64
 7   Calories Burned           508 non-null    float64
 8   Climb (m)                 508 non-null    int64
 9   Average Heart Rate (bpm)  294 non-null    float64
 10  Friend's Tagged           0 non-null      float64
 11  Notes                     231 non-null    object
 12  GPX File                  504 non-null    object
dtypes: float64(5), int64(1), object(7)
memory usage: 55.6+ KB
None
```

2. *Data preprocessing*

```
Running    459
Cycling     29
Walking     18
Other        2
Name: Type, dtype: int64

Type                          0
Distance (km)                 0
Duration                      0
Average Pace                  0
Average Speed (km/h)          0
Climb (m)                     0
Average Heart Rate (bpm)    214
dtype: int64
```

3. *Dealing with missing values – Calculating Mean and then filling missing values with counted means*

```
The sample mean for Average Heart Rate (bpm) for the 'Running' is:  144.985559566787
The sample mean for Average Heart Rate (bpm) for the 'Cycling' is:  124.4

Split whole DataFrame into several, specific for different activities
```

Date	Type	Distance (km)	Duration	Average Pace	Average Speed (km/h)	Climb (m)	Average Heart Rate (bpm)
2018-11-11 14:05:12	Running	10.44	58:40	5:37	10.68	130	159.0
2018-11-09 15:02:35	Running	12.84	1:14:12	5:47	10.39	168	159.0
2018-11-04 16:05:00	Running	13.01	1:15:16	5:47	10.37	171	155.0
2018-11-01 14:03:58	Running	12.98	1:14:25	5:44	10.47	169	158.0
2018-10-27 17:01:36	Running	13.02	1:12:50	5:36	10.73	170	154.0
...	...	...	...	...	...	...	...
2012-09-08 08:35:02	Running	3.27	15:55	4:52	12.32	15	NaN

Date	Type	Distance (km)	Duration	Average Pace	Average Speed (km/h)	Climb (m)	Average Heart Rate (bpm)
2013-08-15 18:49:50	Walking	2.48	2:23:46	57:56	1.04	67	NaN
2013-08-08 07:56:08	Walking	1.51	15:24	10:11	5.89	6	NaN
2013-06-03 07:04:59	Walking	1.33	11:59	9:03	6.63	5	NaN
2013-04-29 18:48:30	Walking	1.37	22:39	16:30	3.64	10	NaN
2013-04-29 13:10:14	Walking	3.83	38:30	10:04	5.96	25	NaN
2013-04-28 10:56:47	Walking	1.32	13:48	10:30	5.72	5	NaN
2013-04-18 21:48:44	Walking	1.50	24:41	16:28	3.64	9	NaN
2013-03-11 18:27:56	Walking	1.86	16:39	8:56	6.72	16	NaN
2012-11-04 18:59:06	Walking	1.22	12:05	9:54	6.07	10	NaN

Date	Type	Distance (km)	Duration	Average Pace	Average Speed (km/h)	Climb (m)	Average Heart Rate (bpm)
2018-10-06 16:45:02	Cycling	19.63	1:26:26	4:24	13.63	210	79.0
2018-09-16 14:55:03	Cycling	32.61	1:55:15	3:32	16.98	462	118.0
2018-09-01 17:06:15	Cycling	36.89	1:58:39	3:13	18.65	491	122.0
2018-08-28 18:44:33	Cycling	28.17	1:27:07	3:06	19.40	400	111.0
2018-08-25 17:18:32	Cycling	19.41	1:11:33	3:41	16.28	199	124.0
2017-09-22 12:27:14	Cycling	49.18	2:42:32	3:18	18.15	367	NaN
2017-08-17 18:36:00	Cycling	15.53	40:04	2:35	23.25	164	138.0
2017-05-01 17:38:35	Cycling	20.19	54:29	2:42	22.24	204	135.0
2016-10-03 11:47:16	Cycling	23.62	1:12:42	3:05	19.50	301	126.0
2016-09-10 17:13:51	Cycling	13.11	32:47	2:30	23.99	163	136.0

```
Filling missing values with counted means
```

Date	Type	Distance (km)	Duration	Average Pace	Average Speed (km/h)	Climb (m)	Average Heart Rate (bpm)
2013-08-15 18:49:50	Walking	2.48	2:23:46	57:56	1.04	67	110.0
2013-08-08 07:56:08	Walking	1.51	15:24	10:11	5.89	6	110.0
2013-06-03 07:04:59	Walking	1.33	11:59	9:03	6.63	5	110.0
2013-04-29 18:48:30	Walking	1.37	22:39	16:30	3.64	10	110.0
2013-04-29 13:10:14	Walking	3.83	38:30	10:04	5.96	25	110.0
2013-04-28 10:56:47	Walking	1.32	13:48	10:30	5.72	5	110.0
2013-04-18 21:48:44	Walking	1.50	24:41	16:28	3.64	9	110.0

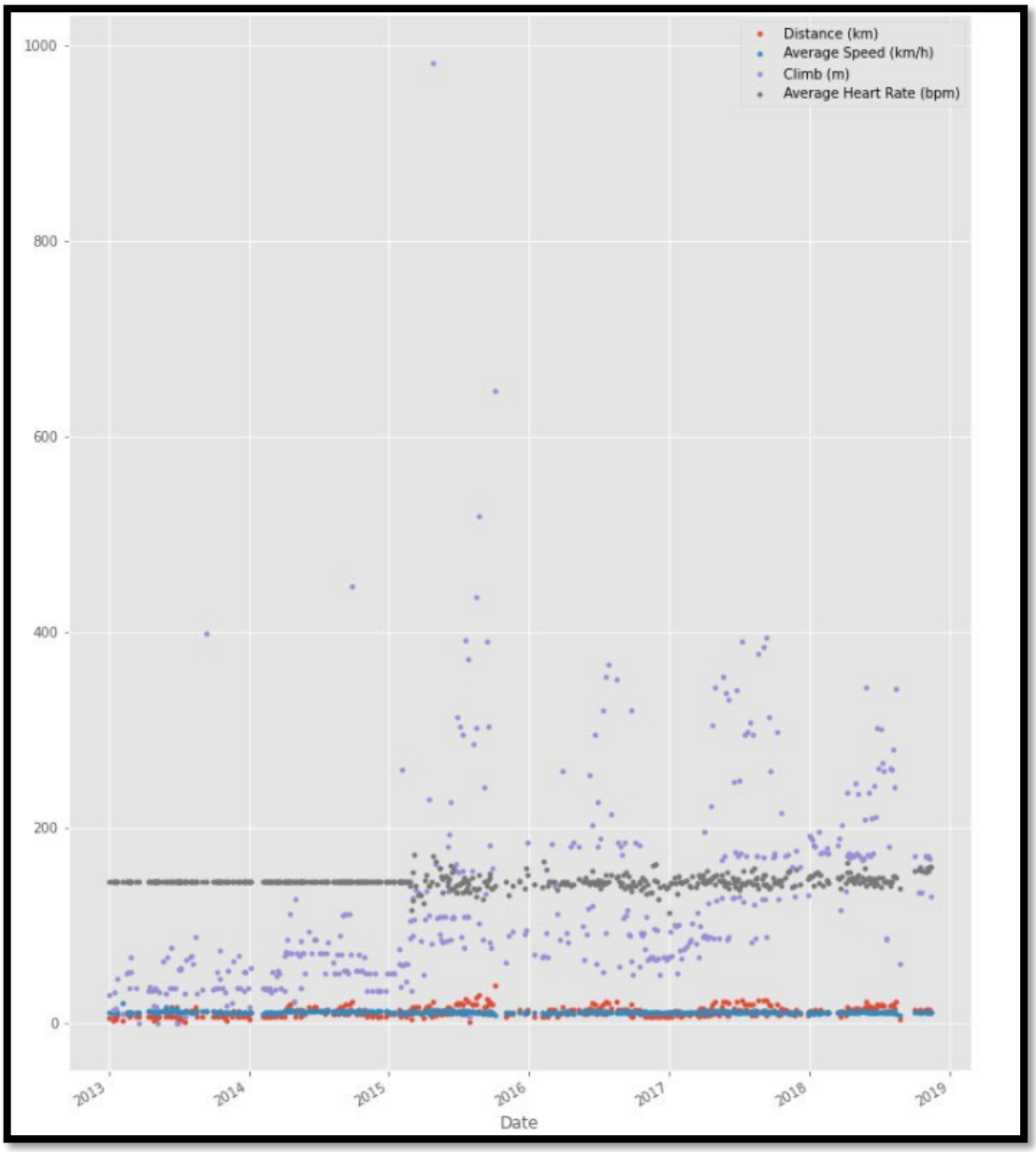
Date	Type	Distance (km)	Duration	Average Pace	Average Speed (km/h)	Climb (m)	Average Heart Rate (bpm)
2018-11-11 14:05:12	Running	10.44	58:40	5:37	10.68	130	159.0
2018-11-09 15:02:35	Running	12.84	1:14:12	5:47	10.39	168	159.0
2018-11-04 16:05:00	Running	13.01	1:15:16	5:47	10.37	171	155.0
2018-11-01 14:03:58	Running	12.98	1:14:25	5:44	10.47	169	158.0
2018-10-27 17:01:36	Running	13.02	1:12:50	5:36	10.73	170	154.0
...	...	...	...	...	...	...	...
2012-09-08 08:35:02	Running	3.27	15:55	4:52	12.32	15	144.0
2012-09-04 19:12:17	Running	6.26	32:35	5:12	11.53	34	144.0

Date	Type	Distance (km)	Duration	Average Pace	Average Speed (km/h)	Climb (m)	Average Heart Rate (bpm)
2018-10-06 16:45:02	Cycling	19.63	1:26:26	4:24	13.63	210	79.0
2018-09-16 14:55:03	Cycling	32.61	1:55:15	3:32	16.98	462	118.0
2018-09-01 17:06:15	Cycling	36.89	1:58:39	3:13	18.65	491	122.0
2018-08-28 18:44:33	Cycling	28.17	1:27:07	3:06	19.40	400	111.0
2018-08-25 17:18:32	Cycling	19.41	1:11:33	3:41	16.28	199	124.0
2017-09-22 12:27:14	Cycling	49.18	2:42:32	3:18	18.15	367	124.0
2017-08-17 18:36:00	Cycling	15.53	40:04	2:35	23.25	164	138.0
2017-05-01 17:38:35	Cycling	20.19	54:29	2:42	22.24	204	135.0
2016-10-03 11:47:16	Cycling	23.62	1:12:42	3:05	19.50	301	126.0
2016-09-10 17:13:51	Cycling	13.11	32:47	2:30	23.99	163	136.0

```
Total missing values for all columns is:

Type                          0
Distance (km)                 0
Duration                      0
Average Pace                  0
Average Speed (km/h)          0
Climb (m)                     0
Average Heart Rate (bpm)      0
dtype: int64
```

4. Plot running data



5. *Running statistics –*

- What is your average distance?
- How fast do you run?
- Do you measure your heart rate?
- How often do you train?

How my average run looks in the last 4 years:

Date	Distance (km)	Average Speed (km/h)	Climb (m)	Average Heart Rate (bpm)
2015-12-31	13.602805	10.998902	160.170732	143.353659
2016-12-31	11.411667	10.837778	133.194444	143.388889
2017-12-31	12.935176	10.959059	169.376471	145.247059
2018-12-31	13.339063	10.777969	191.218750	148.125000

Weekly averages of the last 4 years:

Date	Distance (km)	Average Speed (km/h)	Climb (m)	Average Heart Rate (bpm)
2015-01-04	9.780000	11.120000	51.0	144.0
2015-01-11	NaN	NaN	NaN	NaN
2015-01-18	9.780000	11.230000	51.0	144.0
2015-01-25	NaN	NaN	NaN	NaN
2015-02-01	9.893333	10.423333	58.0	144.0
...	...	...	...	...
2018-10-14	12.620000	10.840000	146.5	157.5
2018-10-21	10.290000	10.410000	133.0	155.0
2018-10-28	13.020000	10.730000	170.0	154.0
2018-11-04	12.995000	10.420000	170.0	156.5
2018-11-11	11.640000	10.535000	149.0	159.0

202 rows × 4 columns

How many trainings per week I had on average: 1.5

6. CONCLUSION AND FUTURE WORK

The Analysis of Fitness Data project successfully analyzed fitness data using Python and Jupyter Notebook. It provided insights into performance, goal achievement, progress tracking, and training intensity. The project generated visualizations, a detailed summary report, and incorporated fun facts to engage users.

The future scope for the Analysis of Fitness Data project includes integrating additional data sources, exploring machine learning techniques, implementing real-time data analysis, developing a dedicated mobile application with social sharing features, enhancing visualization techniques, analyzing long-term trends, and incorporating user feedback for customization options. These advancements would provide a more comprehensive view of fitness performance, personalized insights, improved user experiences, and the ability to adapt to individual preferences and goals.

7. REFERENCES

➢ DATA COLLECTION

The following websites have been referred to obtain the input data and statistics:

a. https://runkeeper.com/cms/

b. https://www.kaggle.com/datasets/yassershrief/cardio-activities

➢ PROGRAMMING REFERENCES

The following websites have been referred for Python tutorials:

a. https://www.w3schools.com/datascience/

b. https://www.javatpoint.com/data-science

c. https://www.geeksforgeeks.org/data-science-tutorial/https://www.tutorialspoint.com/python_data_science/index.htm