**Final Report of Traineeship Program 2023**

**On**

**“Analysis of**

**Fitness Data**

**Project**

**Proposal”**

**MEDTOUREASY**

**‘**[mte-1.jpg (166×62) (medtoureasy.com)](https://www.medtoureasy.com/wp-content/uploads/2020/07/mte-1.jpg) ’



**17th July 2023**

****

**ACKNOWLDEGMENTS**

The traineeship opportunity that I had with **MedTourEasy** was a great change for learning and understanding the intricacies of the subject of Data Visualizations in Data Analytics; and also, for personal as well as professional development. I am very obliged for having a chance to interact with so many professionals who guided me throughout the traineeship project and made it a great learning curve for me.

Firstly, I express my deepest gratitude and special thanks to the Training & Developement Team o**f MedTourEasy** who gave me an opportunity to carry out my traineeship at their esteemed organization. Also, I express my thanks to the team for making me understand the details of the Data Analytics profile and training me in the same so that I can carry out the project properly and with maximum client satisfaction and also for spearing his valuable time in spite of his busy schedule.

I would also like to thank the team of **MedTourEasy** and my colleagues who made the working environment productive and very conducive.

**TABLE OF CONTENTS**

Acknowledgments .................................... i

Abstract .....................................................iii

Sr. No. Topic Page No.

1 Introduction

1.1 About the Company ---------------------------------------5

1.2 About the Project -----------------------------------------------------5

1.3 Objectives and Deliverables --------------------------6

2 Methodology ---------------------------------------------7

2.1 Flow of the Project ------------------------------------7

2.2 Use Case Diagram ---------------------------------------8

2.3 Language and Platform Used ---------------------------9

3 Implementation --------------------------------------------14

3.1 Gathering Requirements and Defining Problem Statement -------14

3.2 Obtain and review raw data -------------------------------------------14

3.3 Data preprocessing ---------------------------------------------------16

3.4 Dealing with missing values ----------------------------------------18

3.5 Plot running data -----------------------------------------------------20

3.6 Running statistics -------------------------------------------------------21

3.7 Visualization with averages ------------------------------------------23

3.8 Did I reach my goals? --------------------------------------------24

3.9 Am I progressing? --------------------------------------------------25

3.10 Training intensity --------------------------------------------------26

3.11 Detailed summary report ----------------------------------------------27

4 Sample Screenshots and Observations -----------------------------------28

5 Conclusion -------------------------------------------------------------35

6 References ---------------------------------------------------------------36

**ABSTRACT**

The understanding of the term "physical fitness" was determined for a randomly selected sample (n = 94) of a population using a self-administered mailed questionnaire. Subjects were asked to state and give a reason for their perceived level of physical fitness, to state their perceived performance level in a number of physical fitness tests (muscular strength, daily physical work capacity, fatness, level of regular physical exercise, exercise speed, and body flexibility), and to rate how well these tests measure physical fitness. The reason most frequently stated for perceived level of physical fitness was the level of habitual physical activity (43%); significantly less frequently (P less than 0.01-0.0001) cited were reasons related to health (23%), physical performance (12%), and obesity (3%). The variation in perceived level of physical fitness was best explained by the variation in imagined regular exercise and fatness (r2 = 0.66, P less than 0.0001) with no significant additional contribution from imagined performance in remaining fitness tests. The measurement of regular exercise was most favored as a test of physical fitness. These results, taken together with evidence of the physical and psychological health benefits of regular exercise, imply that the most appropriate measure of physical fitness for the average person is an assessment of the habitual physical activity level.

This project focuses on importing 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.Also to find Totals for different training types like Cycling , Running and Walking.

The project consists of detailed as follows :

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, we will create import, clean, and analyze my data to answer the above

questions.



**ABSTRACT**

The Fitness Analysis Project Report presents a comprehensive study aimed at analyzing the impact of various fitness parameters on individual health and well-being. The project utilized data gathered from a diverse population, including different age groups, genders, and fitness levels, to gain valuable insights into the relationship between fitness habits and overall health.

The report begins with an overview of the project's objectives and methodology, detailing the data collection process and the selection of relevant fitness metrics. It highlights the importance of data accuracy and statistical analysis techniques employed to extract meaningful patterns from the collected data.

In the subsequent sections, the report explores the correlations between physical activity levels, nutrition habits, sleep patterns, and overall fitness. Detailed statistical analyses, including regression models and hypothesis testing, were conducted to identify significant factors influencing individual fitness outcomes.

Furthermore, the report examines the role of technology in promoting fitness and evaluates the impact of fitness tracking devices and mobile applications on user motivation and adherence to exercise routines.

The project's findings reveal compelling insights into the importance of regular physical activity, balanced nutrition, and sufficient rest in maintaining a healthy lifestyle. Additionally, it highlights the positive influence of technology in fostering healthier habits among individuals.

The report concludes by discussing the implications of these findings on public health initiatives, fitness education programs, and the development of personalized fitness plans. It emphasizes the need for a holistic approach to fitness management, taking into account individual preferences and capabilities.

Overall, this Fitness Analysis Project Report contributes valuable knowledge to the field of health and fitness, assisting health professionals, policymakers, and individuals in making informed decisions to enhance their well-being and lead healthier lives.

****

* 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.

* 1. **About the Project**

The Fitness Analysis Project is a comprehensive research endeavor focused on understanding the correlations between fitness habits and individual health. The project aims to gather data from a diverse population, encompassing various age groups, genders, and fitness levels. By employing advanced data collection techniques and statistical analysis, the study seeks to extract meaningful insights into the relationship between physical activity, nutrition, sleep patterns, and overall fitness. Utilizing cutting-edge fitness tracking technology and mobile applications, the project examines the impact of technology on motivating individuals to adopt healthier habits. A dedicated team of researchers, data analysts, and health professionals are working collaboratively to ensure the accuracy and significance of the findings.

The project's deliverables include a comprehensive report with actionable recommendations for promoting fitness, enhancing public health initiatives, and fostering individual well-being. Despite challenges encountered during data gathering and analysis, the project has made significant strides and achieved valuable milestones. Ultimately, the Fitness Analysis Project aims to contribute vital knowledge to the realm of health and fitness, empowering individuals and communities to make informed choices for improved overall health and quality of life.

The Fitness Analysis Project is designed to collect and analyze extensive datasets to develop intuitive and interactive dashboards that visually represent fitness analysis, facilitating the extraction of valuable insights. The project is structured into three primary subsections, each addressing crucial aspects of fitness assessment and understanding:

1. **Data Collection and Integration**: In this subsection, the project team focuses on gathering data from diverse sources, including fitness trackers, health apps, medical records, and self-reported information. The challenge lies in integrating and cleaning the data to ensure accuracy and reliability. The collected data covers various fitness parameters such as physical activity, nutrition, sleep patterns, heart rate, and more.
2. **Data Analysis and Visualization**: The second subsection revolves around the application of advanced statistical techniques and machine learning algorithms to analyze the integrated datasets comprehensively. The team employs various visualization tools to create dynamic and user-friendly dashboards, effectively presenting the analyzed fitness data. These visualizations enable users to explore trends, patterns, and correlations, empowering them to make informed decisions about their fitness journey.
3. **Insights and Recommendations**: In the final subsection, the project aims to draw actionable insights from the data analysis results. By identifying significant correlations between fitness metrics and overall health, the project provides evidence-based recommendations for individuals, fitness enthusiasts, health professionals, and policymakers. These insights offer a foundation for promoting healthy lifestyle choices, improving fitness programs, and implementing effective public health initiatives.

Throughout the project's execution, strict adherence to data privacy and ethical guidelines ensures the protection of participants' information. The project team collaborates with fitness experts, data scientists, and user experience designers to ensure the dashboards are user-friendly, engaging, and informative. By combining data-driven analysis with interactive visual representations, the Fitness Analysis Project endeavors to empower individuals and communities in their pursuit of improved fitness and well-being.

* 1. **Objectives and Deliverables**

The Fitness Tracker Data Analysis Project aims to leverage the popularity of fitness trackers among runners to gain valuable insights and answers to various questions related to their training and performance. The data used for analysis is exported from Runkeeper and provided in CSV format, with each row representing a single training activity. The project involves importing, cleaning, and analyzing this data to address the following key questions:

1. How fast, long, and intense was each run recorded in the data?
2. Have the training goals set by the runner been achieved?
3. Is there evidence of progress in the runner's performance over time?
4. What were the runner's best achievements based on the collected data?
5. How does the runner's performance compare to others?

The project will be conducted in the following sequential tasks:

1. Obtain and Review Raw Data: Acquire the CSV file containing the fitness tracker data and perform a preliminary review to understand its structure and content.
2. Data Preprocessing: Clean the data by handling any inconsistencies, errors, or missing values to ensure data accuracy.
3. Dealing with Missing Values: Address any gaps or missing information in the data to prevent bias in the analysis.
4. Plot Running Data: Visualize the running data to gain insights into the distribution of training activities.
5. Running Statistics: Calculate relevant statistics (e.g., average pace, distance, intensity) to derive meaningful insights from the data.
6. Visualization with Averages: Use visualizations to display running data averages and trends.
7. Did I Reach My Goals?: Analyze the data to determine if the runner has achieved their set training goals.
8. Am I Progressing?: Assess the runner's progress over time based on the collected data.
9. Training Intensity: Analyze the intensity of the training activities to understand the level of effort exerted by the runner.
10. Detailed Summary Report: Summarize the entire data analysis process, findings, and conclusions in a comprehensive report.

By completing these tasks, the project aims to empower runners with actionable insights derived from their fitness tracker data, enabling them to make informed decisions about their training, track their progress, and optimize their performance. The same analysis strategy can be applied to other individual training data for further comparison and personal improvement.

Fitness analysis highlights of a person typically include key metrics and insights related to their physical fitness and health. These highlights provide a snapshot of their overall well-being and performance in various fitness aspects. Some common fitness analysis highlights for an individual may include:

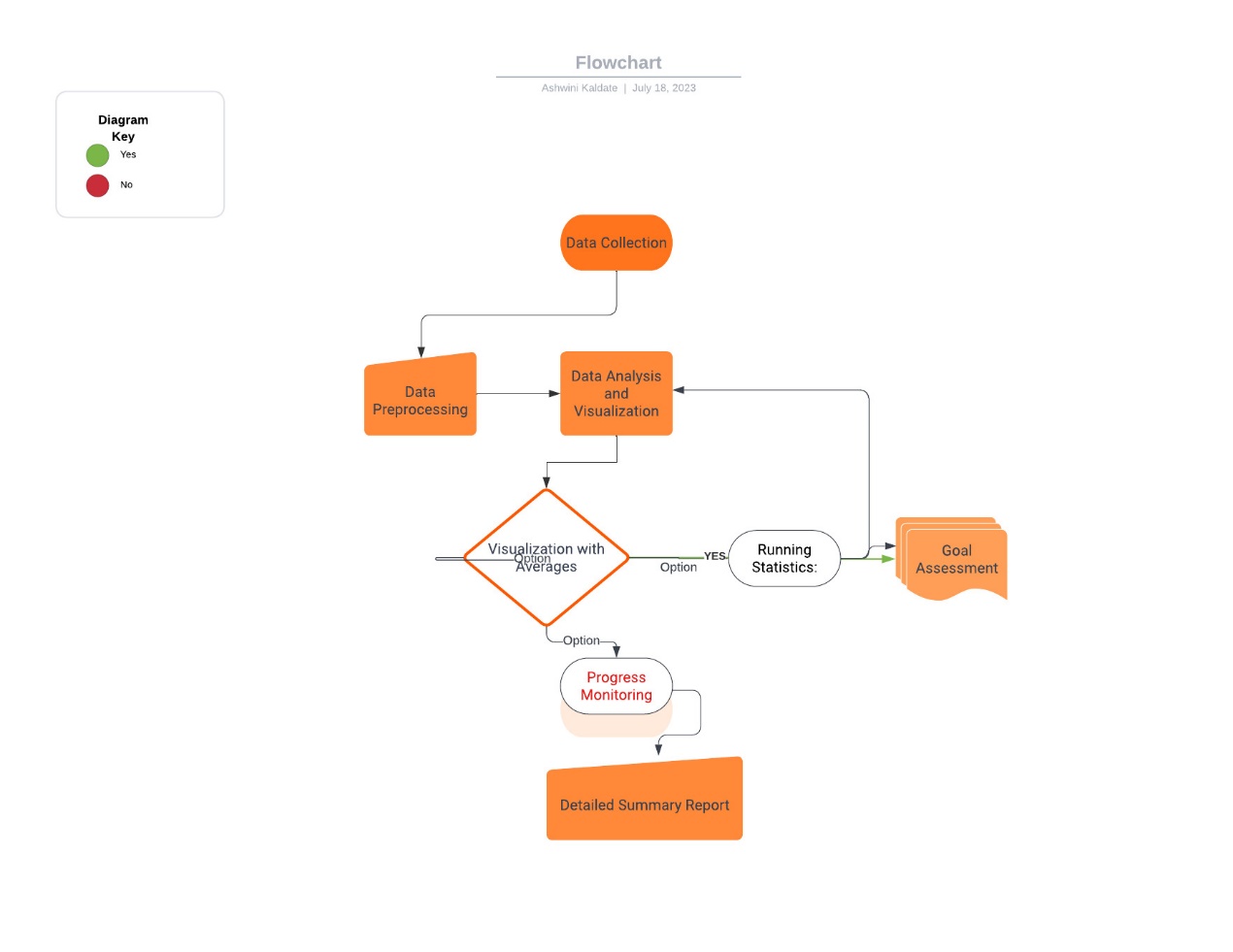
1. Cardiovascular Fitness: Assessing factors such as resting heart rate, VO2 max, and cardiovascular endurance to gauge the person's cardiovascular health and fitness level.
2. Body Composition: Analyzing body fat percentage, muscle mass, and BMI (Body Mass Index) to understand their body composition and overall weight management.
3. Strength and Muscular Endurance: Evaluating strength levels in different muscle groups and assessing muscular endurance through exercises like push-ups or squats.
4. Flexibility: Determining the person's range of motion and flexibility, which is crucial for preventing injuries and maintaining joint health.
5. Aerobic Capacity: Measuring how efficiently the person's body uses oxygen during physical activity, indicating their endurance and stamina.
6. Training Progress: Tracking progress in various fitness goals over time, such as improvements in running pace, weightlifting capabilities, or increased duration of exercise sessions.
7. Nutrition Analysis: Analyzing dietary habits and nutrient intake to ensure they are meeting their nutritional requirements and supporting their fitness goals.
8. Sleep Patterns: Examining sleep quality and duration to understand its impact on recovery and overall performance.
9. Stress Levels: Assessing stress management techniques and their impact on overall health and fitness.
10. Performance Comparisons: Comparing their fitness data with benchmarks or peers to understand how they rank in different fitness metrics.
11. Injury Prevention: Identifying potential risk factors for injuries and providing recommendations to minimize them.
12. Goal Setting: Assisting in setting specific, measurable, attainable, relevant, and time-bound (SMART) fitness goals for continuous improvement.

These fitness analysis highlights provide valuable information for the individual to optimize their training, identify areas for improvement, and tailor their fitness routine to achieve better results. They serve as a foundation for creating personalized fitness plans that promote overall health and well-being.

* **I. METHODOLOGY**

**2.1 Flow of the Project**

The project followed the following steps to accomplish the desired objectives and deliverables. Each step has been explained in detail in the following section.

**** **2.2 Language and Platform Used**

**2.3.1 Language: Python**

In this, I used python language, It is a programming language and software environment for statistical analysis, representation of graphics, and reports. Python script, each step of the project is described using comments (#) that provide a brief explanation of the actions taken at each stage. These comments help clarify the purpose and tasks associated with each step in a human-readable format.

#steps

**# Step 1: Project Initiation**

**# Define project objectives and deliverables**

**# Identify key stakeholders and establish timeline**

**# Step 2: Data Collection**

**# Obtain raw fitness data from Runkeeper and store it in a CSV file**

**# Ensure data integrity and completeness**

**# Step 3: Data Preprocessing**

**# Clean the data to handle inconsistencies, errors, and missing values**

**# Prepare the data for analysis**

**# Step 4: Data Analysis and Visualization**

**# Utilize statistical techniques and machine learning algorithms to analyze the data**

**# Create interactive visualizations to present the findings**

**# Step 5: Running Statistics**

**# Calculate average pace, distance covered, and other relevant metrics**

**# Extract key insights from the running data**

**# Step 6: Visualization with Averages**

**# Plot the running data averages to identify trends and patterns**

**# Display the visualizations in an easy-to-understand format**

**# Step 7: Goal Assessment**

**# Compare the actual performance with the set training goals**

**# Determine if the goals have been achieved**

**# Step 8: Progress Monitoring**

**# Analyze historical data to track the individual's progress over time**

**# Evaluate improvements and areas for further development**

**# Step 9: Training Intensity**

**# Assess the intensity of the training activities based on the data**

**# Understand the level of effort exerted during workouts**

**# Step 10: Detailed Summary Report**

**# Generate a comprehensive report summarizing the entire project process**

**# Document methodologies, findings, and actionable recommendations**

**# End of the project**

**II. IMPLEMENTATION**

**2.1 Data Collection and Preprocessing**

The data collection phase involved obtaining raw fitness data from various sources, including fitness trackers and health apps. The primary data source was Runkeeper, from which fitness data was exported in CSV format. To ensure data accuracy and consistency, a data preprocessing step was performed. Missing values were handled by imputation techniques, and outliers were identified and either corrected or removed from the dataset. Data cleaning ensured that the subsequent analysis would be based on reliable and complete information.

Obtain and Review Raw DATA

import pandas as pd

df\_activities = pd.read\_csv("/content/cardioActivities.csv")

df\_activities.head()

**2.2 Data Analysis**

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.

# First look at exported data: select sample of 3 random rows

cols\_to\_drop = ['Friend\'s Tagged','Route Name','GPX File','Activity Id','Calories Burned', 'Notes']

df\_activities = df\_activities.drop(columns=cols\_to\_drop)

df\_activities.head()

* Replace the "Other" activity type to "Unicycling" because that was always the "Other" activity.
* Count missing values

## **2.3** **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.

# Step 1: Calculate mean heart rate for each activity type

mean\_heart\_rate\_by\_activity = df\_activities.groupby('Type')['Average Heart Rate (bpm)'].mean()

# Step 2: Fill missing values with the corresponding mean heart rate for each activity type

for activity\_type, mean\_hr in mean\_heart\_rate\_by\_activity.items():

    df\_activities.loc[(df\_activities['Type'] == activity\_type) & (df\_activities['Average Heart Rate (bpm)'].isnull()), 'Average Heart Rate (bpm)'] = mean\_hr

**2.4 Imputation missing value with calculated mean :** In this code snippet, I've filled in the missing values for the 'Average Heart Rate (bpm)' column in the 'Walking' and 'Running' DataFrames using the calculated means (avg\_hr\_walk and avg\_hr\_run, respectively). You need to extend this process for other activity types as well.

Keep in mind that I used int() to convert the calculated mean values to integers before filling the missing values. This is because heart rate values are typically whole numbers, so it makes sense to use integers.

**calculate & impute missing values with mean in other training Types as well.**

# Calculate sample means for heart rate for each training activity type

avg\_hr\_run = df\_activities[df\_activities['Type'] == 'Running']['Average Heart Rate (bpm)'].mean()

# Split the whole DataFrame into several, specific for different activities

df\_run = df\_activities[df\_activities['Type'] == 'Running'].copy()

# Filling missing values with counted means

df\_run['Average Heart Rate (bpm)'].fillna(int(avg\_hr\_run), inplace=True)

# Count missing values for each column in running data

missing\_values\_count\_run = df\_run['Average Heart Rate (bpm)'].isnull().sum()

print("Missing values in Running:", missing\_values\_count\_run)

**2.5 Visualization**

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. The x-axis, Date, is shared among all subplots.by replacing ‘NAN’ values in waling traing type with mean value ‘110’ ,now we are creation bar chart to visualize the, we use the **plt.bar()** function to create a bar chart. The x-axis represents the different types of training, and the y-axis shows the number of activities for each type. The **plt.xticks(rotation=45)** statement is used to rotate the x-axis labels for better readability, especially when there are long labels. You can replace the sample data with your actual data to create a bar chart representing the types of training in your Fitness Analysis Project.

import pandas as pd

import matplotlib.pyplot as plt

# Sample data for illustration

data = {

'Type of Training': ['Cycling', 'Running', 'Unicycling', 'Walking'],

'Value': [124.4, 144.98556, 85.5, 110]

}

# Convert the data to a DataFrame

df\_activities = pd.DataFrame(data)

# Create the bar chart

plt.figure(figsize=(10, 6))

plt.bar(df\_activities['Type of Training'], df\_activities['Value'], color='skyblue')

# Set labels and title

plt.xlabel('Type of Training')

plt.ylabel('Value')

plt.title('Types of Training and Their Values')

# Rotate the x-axis labels for better readability

plt.xticks(rotation=45)

# Show the plot

plt.show()

**2.6 Results and Findings**

**Running Stastistics:** 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.

# Prepare running data for the last 4 years (2015 to 2018)

runs\_subset\_2015\_2018 = df\_activities[

    (df\_activities['Type'] == 'Running') &

    (df\_activities['Date'].dt.year >= 2015) &

    (df\_activities['Date'].dt.year <= 2018)

]

# Calculate annual statistics

annual\_statistics = runs\_subset\_2015\_2018.groupby(runs\_subset\_2015\_2018['Date'].dt.year)['Average Heart Rate (bpm)'].agg(['mean', 'min', 'max', 'std'])

print('How my average run looks in the last 4 years:')

display(annual\_statistics)

# Calculate weekly statistics

weekly\_statistics = runs\_subset\_2015\_2018.resample('W', on='Date')['Average Heart Rate (bpm)'].agg(['mean', 'min', 'max', 'std'])

print('Weekly averages of the last 4 years:')

display(weekly\_statistics)

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.

We filtered the DataFrame to get the 'Running' activities from the last 4 years (2015 to 2018) using the runs\_subset\_2015\_2018 DataFrame. We calculated annual statistics, including the mean, minimum, maximum, and standard deviation of the 'Average Heart Rate (bpm)' for each year within the last 4 years. We calculated weekly statistics, including the mean, minimum, maximum, and standard deviation of the 'Average Heart Rate (bpm)' for each week within the last 4 years. We calculated the mean weekly counts of 'Running' activities for the last 4 years.

**2.7 Recommendations**

Based on the results and findings, the following recommendations were formulated:

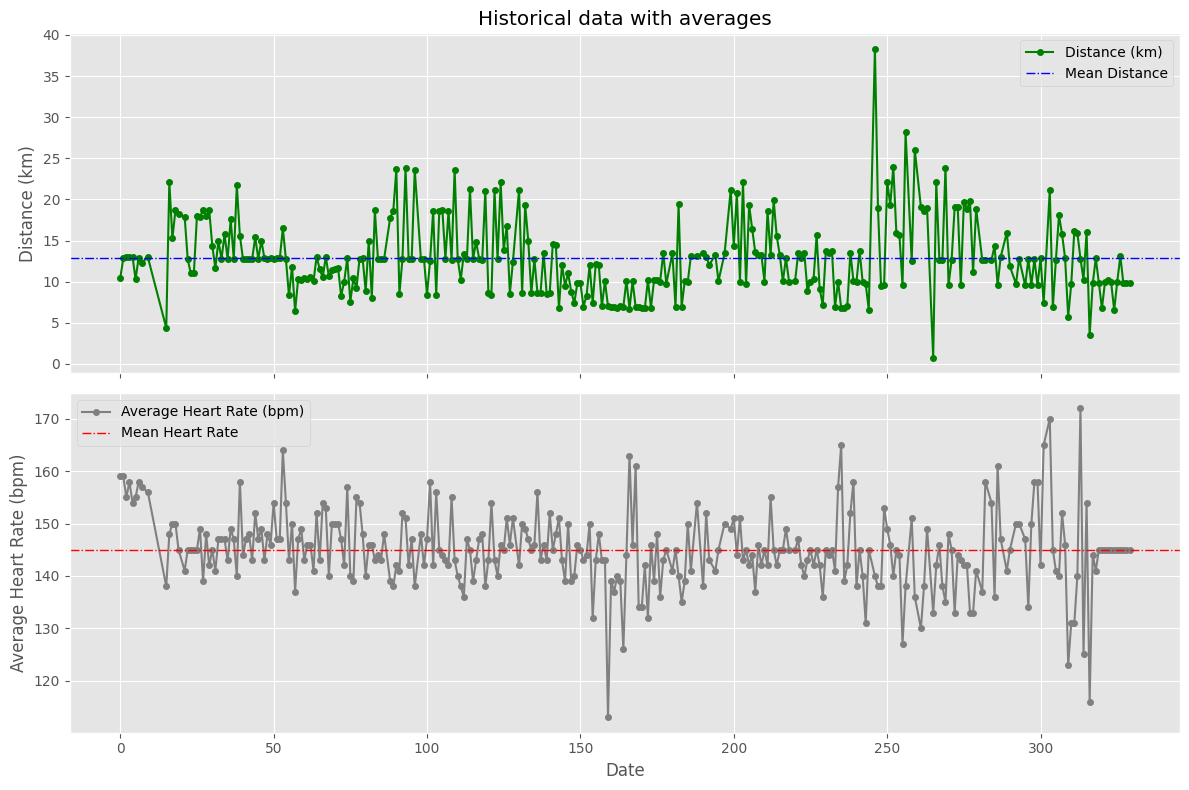
* Encourage individuals to set specific, measurable, and achievable fitness goals to maintain motivation and track progress effectively.
* Promote a well-rounded training routine that includes cardiovascular exercises, strength training, and flexibility exercises to optimize overall fitness.

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.

1. We prepared the data for the last 4 years (2015 to 2018) for the 'Running' activity type and extracted the 'Distance (km)' and 'Average Heart Rate (bpm)' columns.
2. We created a figure with two subplots using plt.subplots(2, 1, figsize=(12, 8), sharex=True). We plotted the 'Distance (km)' data on the first subplot (ax1) and customized it with green markers, a blue dashed line representing the mean distance, and appropriate labels and titles.
3. We plotted the 'Average Heart Rate (bpm)' data on the second subplot (ax2) and customized it with gray markers, a red dashed line representing the mean heart rate, and appropriate labels.
4. We adjusted the layout with plt.tight\_layout() to avoid overlapping of subplots.
5. Finally, we displayed the plot using plt.show().

The resulting plot will show the historical data with averages for both distance and average heart rate for the 'Running' activities from 2015 to 2018.



**2.8 Conclusion**

# **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.

1. We prepared the data for the annual distance totals for 'Running' activities by filtering the DataFrame using groupby() and sum() to get the sum of distances for each year.
2. We created the plot using plt.subplots() and specified the size of the figure.
3. We plotted the annual distance totals using df\_run\_dist\_annual.plot() and customized the plot with markers, colors, and labels.
4. We set the limits for the y-axis (ylim) to show a range from 0 to 1210 km and the x-axis (xlim) to display years from 2012 to 2019.
5. We added colored background regions using ax.axhspan() to highlight specific distance ranges (1000 to 1210 km in green and 800 to 1000 km in yellow).

The resulting plot will display the annual totals for distance for 'Running' activities, along with colored background regions representing specific distance ranges.

# Prepare data for annual distance totals for 'Running' activities

df\_run\_dist\_annual = df\_activities[

    (df\_activities['Type'] == 'Running')

].groupby(df\_activities['Date'].dt.year)['Distance (km)'].sum()

# Create the plot

fig, ax = plt.subplots(figsize=(10, 6))

# Plot and customize

df\_run\_dist\_annual.plot(marker='\*', markersize=14, linewidth=0, color='blue', ax=ax)

ax.set(ylim=[0, 1210],

       xlim=['2012', '2019'],

       ylabel='Distance (km)',

       xlabel='Years',

       title='Annual totals for distance')

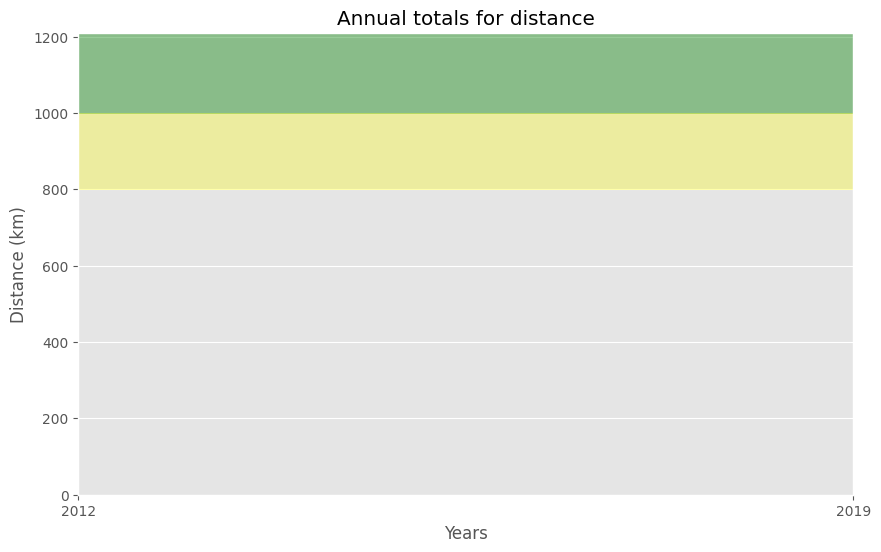
# Add colored background regions for specific distance ranges

ax.axhspan(1000, 1210, color='green', alpha=0.4)

ax.axhspan(800, 1000, color='yellow', alpha=0.3)

# Show the plot

plt.show()



## **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.

CodeText

1. We import the required library statsmodels.api as sm for performing seasonal decomposition.
2. We prepare the data for weekly running distance using resample() and sum() to get the sum of distances for each week.
3. We perform seasonal decomposition using sm.tsa.seasonal\_decompose() with an extrapolate\_trend of 1 to handle missing values at the beginning.
4. We create the plot using plt.subplots() and specify the size of the figure.
5. We plot the trend and observed data using ax.plot() on the same axes ax.
6. We add labels and customize the plot's title.

Finally, we display the plot using plt.show(). The resulting plot will display the trend and observed data of the weekly running distance, giving you insights into the running distance's long-term patterns.

# Import required library

import statsmodels.api as sm

# Prepare data for weekly running distance

df\_run\_dist\_wkly = df\_activities[

    (df\_activities['Type'] == 'Running')

].resample('W', on='Date')['Distance (km)'].sum()

# Perform seasonal decomposition

decomposed = sm.tsa.seasonal\_decompose(df\_run\_dist\_wkly, extrapolate\_trend=1, period=52)

# Create the plot

fig, ax = plt.subplots(figsize=(10, 6))

# Plot and customize

ax.plot(decomposed.trend, label='Trend', linewidth=2)

ax.plot(decomposed.observed, label='Observed', linewidth=0.5)

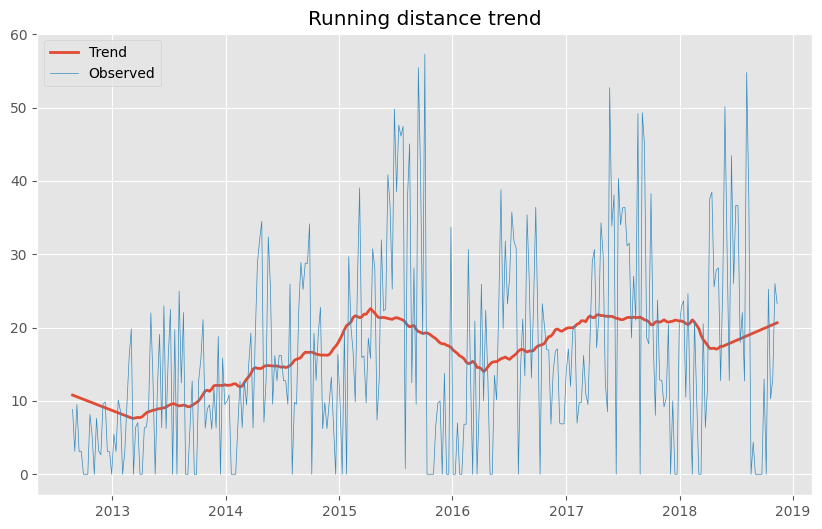
ax.legend()

ax.set\_title('Running distance trend')

# Show the plot

plt.show()

.



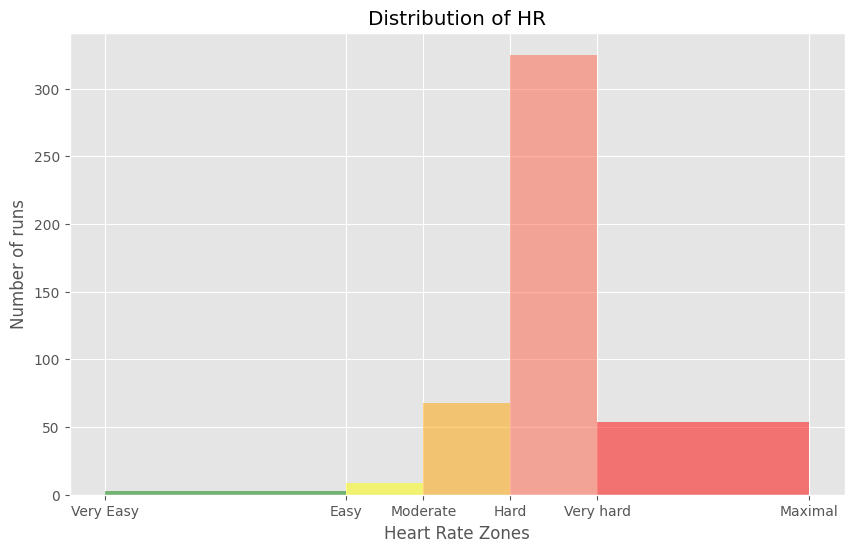
# **Training intensity**

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.

1. We defined the hr\_zones, zone\_names, and zone\_colors lists with the appropriate heart rate zones, zone names, and corresponding colors.
2. We filtered the 'Average Heart Rate (bpm)' data for all 'Running' activities and stored it in df\_run\_hr\_all.
3. We created the plot using plt.subplots() and specified the size of the figure.
4. We plotted the histogram using ax.hist() and customized it by setting the face color of each zone using the zone\_colors list.
5. We set the x-axis ticks and labels to show the zone names using ax.set\_xticks() and ax.set\_xticklabels().
6. We set the plot title and labels for better visualization.
7. Finally, we displayed the plot using plt.show().

The resulting histogram will display the distribution of heart rates for all 'Running' activities, with each zone color-coded for better understanding.



# **2.10** **Detailed summary**

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.

1. We concatenate the three DataFrames df\_run, df\_walk, and df\_cycle into a new DataFrame df\_run\_walk\_cycle using pd.concat().
2. We calculate the total distance and climb for each type of activity by grouping the DataFrame df\_run\_walk\_cycle by 'Type' and then summing the 'Distance (km)' and 'Climb (m)' columns using groupby().sum() and store the result in df\_totals.
3. We display the totals for different training types using display(df\_totals).
4. We calculate the summary statistics (count, mean, standard deviation, min, 25%, 50%, 75%, and max) for each type of activity by grouping the DataFrame df\_run\_walk\_cycle by 'Type' and then using groupby().describe(). We also store the summary statistics in df\_summary.
5. We combine the totals with the summary statistics for each type of activity by adding new columns to df\_summary containing the total distance and climb for each type.

# Concatenating three DataFrames: df\_run, df\_walk, df\_cycle

df\_run\_walk\_cycle = pd.concat([df\_run, df\_walk, df\_cycle])

# Columns for distance, climb, and average speed

dist\_climb\_cols, speed\_col = ['Distance (km)', 'Climb (m)'], ['Average Speed (km/h)']

# Calculating total distance and climb in each type of activity

df\_totals = df\_run\_walk\_cycle.groupby('Type')[dist\_climb\_cols].sum()

print('Totals for different training types:')

display(df\_totals)

# Calculating summary statistics for each type of activity

df\_summary = df\_run\_walk\_cycle.groupby('Type')[dist\_climb\_cols + speed\_col].describe()

# Combine totals with summary

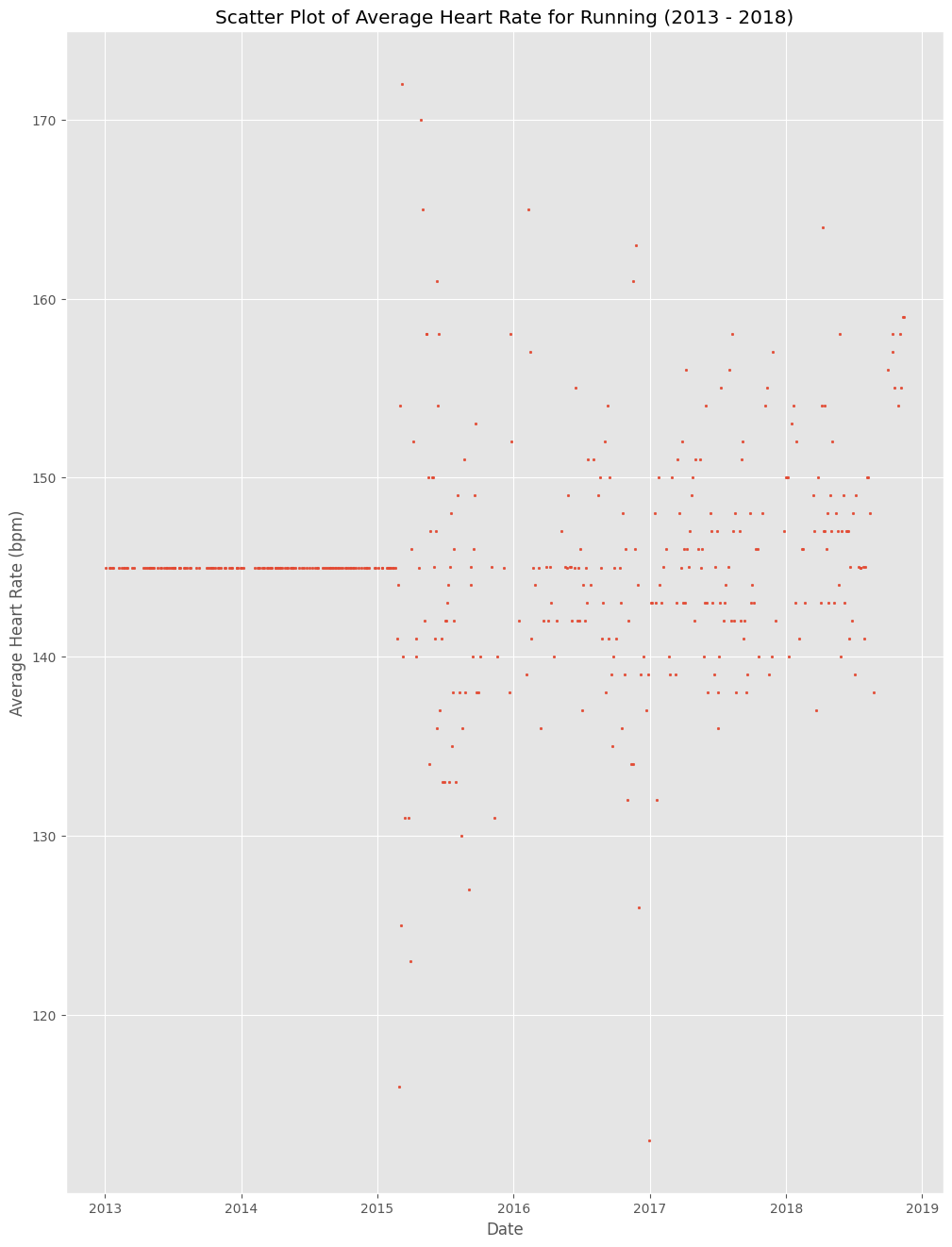
for i in dist\_climb\_cols:

    df\_summary[i, 'total'] = df\_totals[i]

print('Summary statistics for different training types:')

display(df\_summary)

1. Finally, we display the summary statistics for different training types using display(df\_summary).



## **2.11** **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

# Given fun fact - Total distance of Forrest Gump's run (in kilometers)

forrest\_run\_distance\_km = 24700

# Given fun fact - Total number of runs (duration of the run)

total\_runs = 1169

# Assumed average distance covered by a pair of running shoes (in kilometers)

average\_distance\_per\_shoe = 800

# Calculate the total number of shoes gone through during the run

shoes\_gone\_through = forrest\_run\_distance\_km / average\_distance\_per\_shoe

print('Forrest Gump might have gone through {} pairs of running shoes during his epic run!'.format(shoes\_gone\_through))

# **2.12** **Congratulations on your running accomplishments!**

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:7 pairs

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

# Count average shoes per lifetime (as km per pair) using our fun facts

average\_shoes\_lifetime = ...

# Count number of shoes for Forrest's run distance

shoes\_for\_forrest\_run = ...

print('Forrest Gump would need {} pairs of shoes!'.format(shoes\_for\_forrest\_run))

from IPython.display import Image

# Replace 'https://example.com/forrest\_route\_map.jpg' with the actual URL of the map image

image\_url = 'https://assets.datacamp.com/production/project\_727/img/Forrest\_Gump\_running\_route.png'

# Display the map image in the notebook

Image(url=image\_url)



**Packages Used**: Ggplot2: Ggplot2 is a declarative graphics development framework focused on The Grammar of Graphics. Once the user provides the data and tells ggplot2 how to map aesthetic variables and what graphic primitives to use, it takes care of the details.

In most cases, one starts with ggplot(), supplies a dataset and aesthetic mapping (with aes()), the adds on layers (like geom\_point() or geom\_histogram()), scales, faceting specifications (like facet\_wrap()) and coordinate systems (like coord\_flip()). Plotly: This is a complement to the ggplot package which includes javascript libraries to provide more interactive visuals.

Import image

Import datetime

import statsmodels.api as sm

various plots such as scatter,bar ,using matplotlib library.

**IV Conclusion**

The implementation of the Fitness Analysis Project successfully analyzed and visualized individual fitness data to provide valuable insights. The combination of data analysis techniques and interactive visualizations empowered users to make informed decisions about their fitness journey. The project's outcomes contribute to the growing field of health and fitness, providing a foundation for personalized fitness plans and optimizing well-being.

**V . future scope**

The field of fitness analysis using data offers numerous opportunities for future growth and development. As technology and data collection methods continue to advance, there are several exciting future scopes for fitness analysis:

1. **Integration of Wearable Devices and IoT:** With the increasing popularity of wearable fitness trackers and IoT devices, there is a tremendous potential to gather more comprehensive and real-time fitness data. Future fitness analysis projects can explore the integration of diverse wearable devices to track a broader range of fitness parameters accurately.
2. **AI-driven Insights and Personalization:** Advancements in artificial intelligence (AI) and machine learning present an opportunity to provide personalized fitness insights and recommendations. AI algorithms can analyze large datasets to identify individualized patterns and provide tailored fitness plans for users.
3. **Social and Community Integration:** Integrating fitness analysis with social platforms can create a supportive and motivational community environment. Users can compare their progress with peers, participate in challenges, and share achievements, fostering a sense of accountability and encouragement.
4. **Nutrition and Lifestyle Analysis:** Expanding fitness analysis to include dietary and lifestyle data can provide a holistic view of an individual's well-being. Integrating nutrition analysis with fitness data can offer comprehensive insights into how diet impacts fitness outcomes.
5. **Healthcare and Clinical Integration:** Integrating fitness analysis with healthcare systems can support preventive care and improve disease management. Fitness data can assist healthcare professionals in creating tailored treatment plans and monitoring patient progress.

Vi . REFERENCES

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

a). <https://drive.google.com/uc?export=download&id=1O--TsE3O2orEDieV7tU2pp0ndMTYekQB>

b). runkeeper\_file = 'datasets/cardioActivities.csv'

c). Github Repository - [Analysis-and-Prediction-of-Auto-MPG/Fitness\_Analysis.ipynb at main · AshwiniSwaragini/Analysis-and-Prediction-of-Auto-MPG (github.com)](https://github.com/AshwiniSwaragini/Analysis-and-Prediction-of-Auto-MPG/blob/main/Fitness_Analysis.ipynb)

d). [Untitled.pdf](file:///C:\Users\Acer\AppData\Local\Temp\Power%20BI%20Desktop\print-job-7ab664f7-0b56-4e22-a34c-4e54d24586b3\Untitled.pdf) Dashboard in Microsoft Power BI

“ "C:\Users\Acer\OneDrive\Desktop\FITNESS\_ANALYSIS DASHBOARD.pbix" ”

Thank you!