Final Report of Internship Program 2021
On
“Analysing of Fitness Data”

MEDTOUREASY

26th January 2021

ACKNOWLDEGMENTS

The internship opportunity that I had with MedTourEasy was a great change for learning and understanding various techniques in Data Science and also for personal as well as professional development. I am very obliged for having a Support from many professionals who guided me throughout the internship project and made it a great learning curve for me.

Firstly, I express my deepest gratitude and special thanks to the Training Head of MedTourEasy, Mr. Ankit Hasija who gave me an opportunity to carry out my internship at their esteemed organization. Also, I express my thanks to him for making me understand the details of the Data Scientist profile and training me in the same so that I can carry out the project properly and also for spearing his valuable time in spite of his busy schedule.

I would also like to thank the team of MedTourEasy for making this Intership program very practical and efficient.

TABLE OF CONTENTS

S. No.	Topic
	Visualization with Average
	Did I reach my goal?
	Am I Progressing
	Training Intensity
5.	**Conclusion**

ABSTRACT

There are so many ways to maintain the fitness and that made me think about running styles, training habits, and achievements, then I suddenly realized that I could take an in-depth analytical look at training data. To do some analysis on fitness data using a popular GPS fitness tracker called Runkeeper data is very useful.

Using the Runkeeper app data is great. One key feature: its excellent data export. Anyone who has a smartphone can download the app and analyze their data.

After logging your run, the first step is to export the data from Runkeeper. Then import the data and start exploring to find potential problems. After that, create data cleaning strategies to fix the issues. I exported seven years worth of training data, from 2012 through 2018. Finally, analyze and visualize the clean time-series data in order to gain meaningful insights.

ABOUT THE COMPANY

About the Company MedTourEasy, a global healthcare company, provides you the informational resources needed to evaluate your global options. It helps you find the right healthcare solution based on specific health needs, affordable care while meeting the quality standards that you expect to have in healthcare. MedTourEasy improves access to healthcare for people everywhere. It is an easy to use platform and service that helps patients to get medical second opinions and to schedule affordable, high-quality medical treatment abroad.

ABOUT THE PROJECT

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?

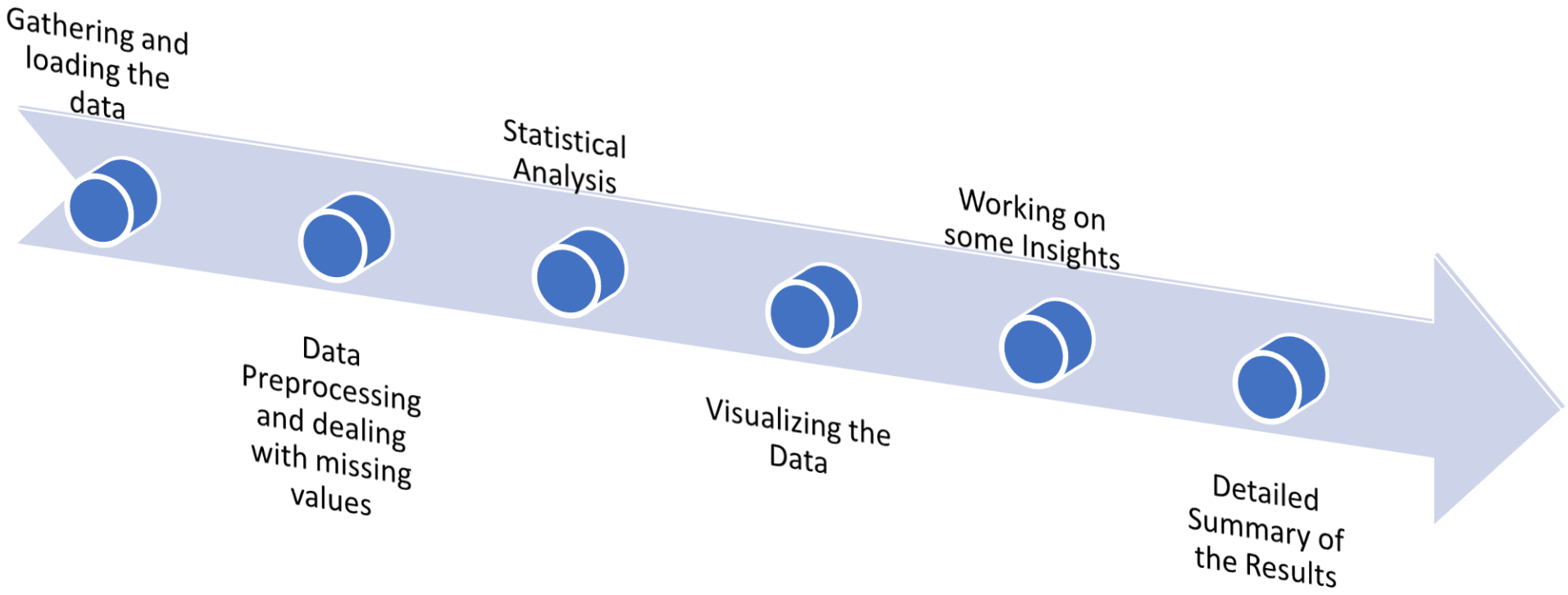
This data was exported from Runkeeper. The data is a CSV file where each row is a single training activity.

OBJECTIVES AND DELIVERABLES

- 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

METHODOLOGY



LANGUAGE AND PLATFORM USED

- Python
- Jupyter Notebook (IDE)
- Pandas (python library)

  pip install pandas (To Install)

  Import pandas (To Import)
- Matplotlib (python library)

  pip install matplotlib (To Install)

  Import matplotlib.pyplot (To Import)

IMPLEMENTATION

OBTAIN AND REVIEW RAW DATA

- Gathered the exported seven years worth of training data of RunKeeper application, from 2012 through 2018. The data is in CSV file format.
- Import pandas library and use it to read the CSV file data, we store this in a Data frame.
- By using the dataframe_name .head() we can see the preview of the dataset.
- Parse the date so that we can have a clear look through each and every date in an order
- Use .info() to see the summary of the data.

DATA PREPROCESSING

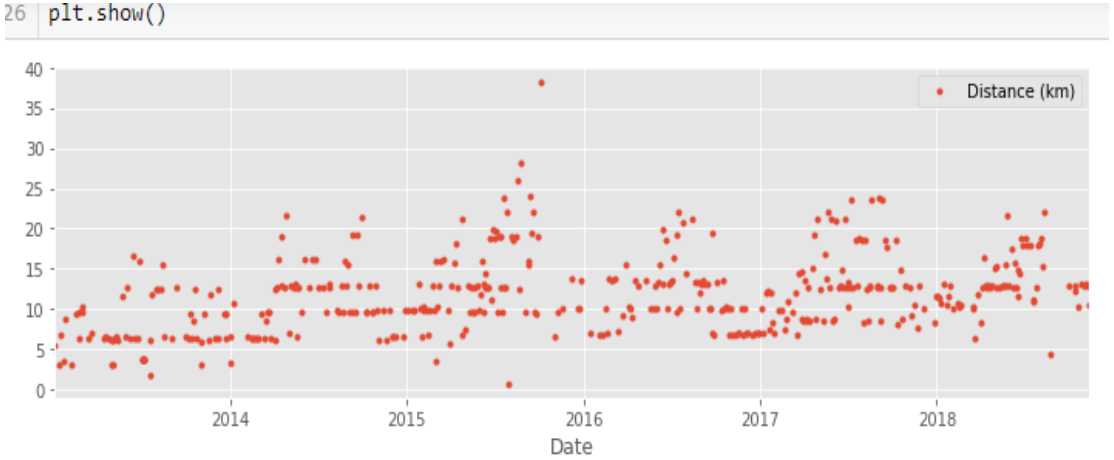
- Seeing the summary there are missing values in some of the columns.
- Remove columns not useful for our analysis.
- Some columns like Route Name is only used once and some don’t use cardiac sensor regularly and friends tagged is never used and few more. So, these columns are not useful for the analysis.
- Store all those columns in a data frame and drop that data frame with column names from the original data frame using .drop(cols, inplace = True).
- Replace the "Other" activity type to "Unicycling“ using .str.replace().
- Then Count missing values by .isnull().sum() to check the total number of null values in each column.

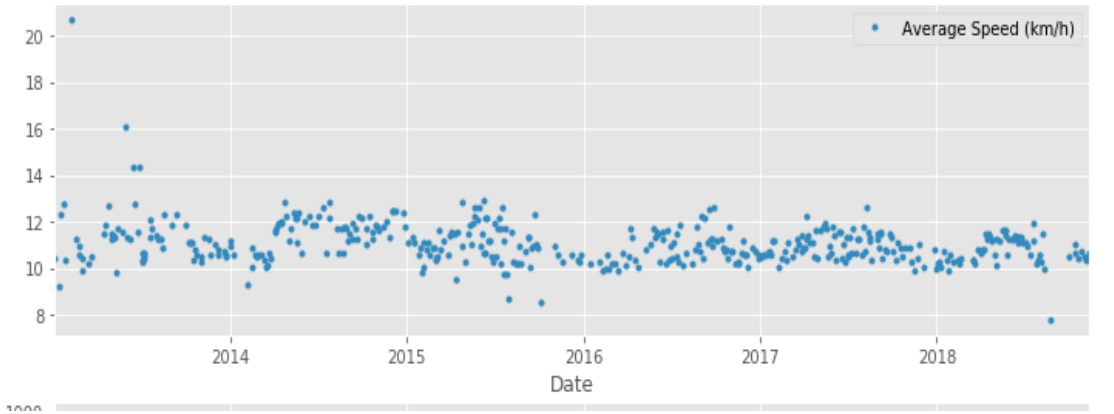
DEALING WITH MISSING VALUES

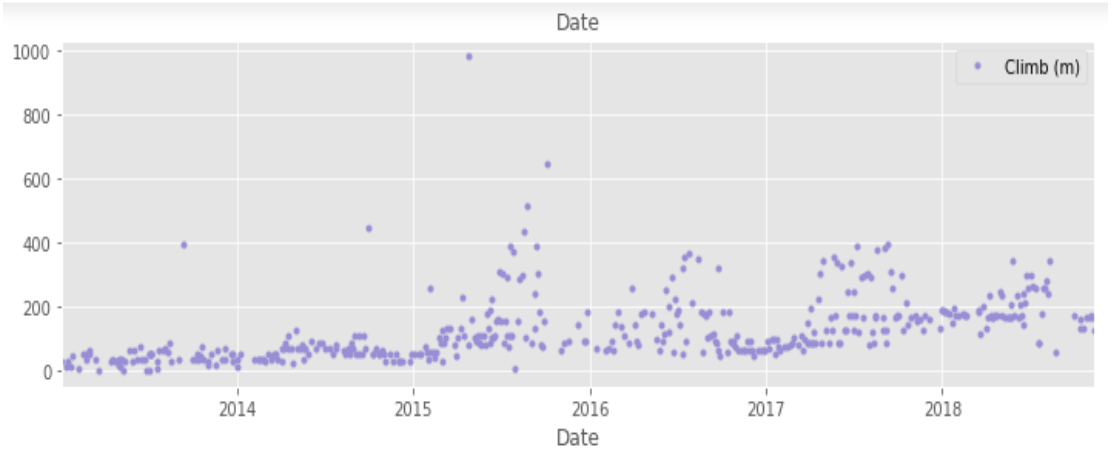
- The last output from the count of null values, there are 214 missing entries for average heart rate.
- We can't go back in time to get those data, but we can fill in the missing values with an average value. This process is called mean imputation. When imputing the mean to fill in missing data, we need to consider that the average heart rate varies for different activities (e.g., walking vs. running).
- filter the DataFrames by activity type and calculate each activity's mean heart rate, then fill in the missing values with those means.
- We use .fillna() method to fill the missing values with the mean of each type
- Finally check if there are any null values found by using .isnull().sum().

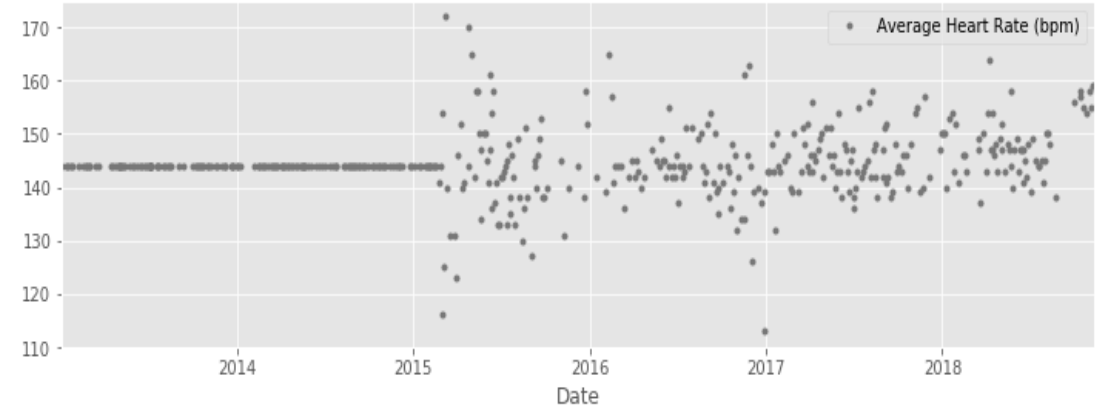
PLOT RUNNING DATA

```
26 plt.show()
```









- The above plots are different metrics of running instance.
- They represent Distance, average Speed, Climb, and Average Heart Rate from the year 2013 to 2018.
- These plots are done by importing the matplotlib and using the ‘ggplot’ style.

RUNNING STATISTICS

```
14 print('How many trainings per week I had on average:', weekly_counts_average)
```

How my average run looks in 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 last 4 years:

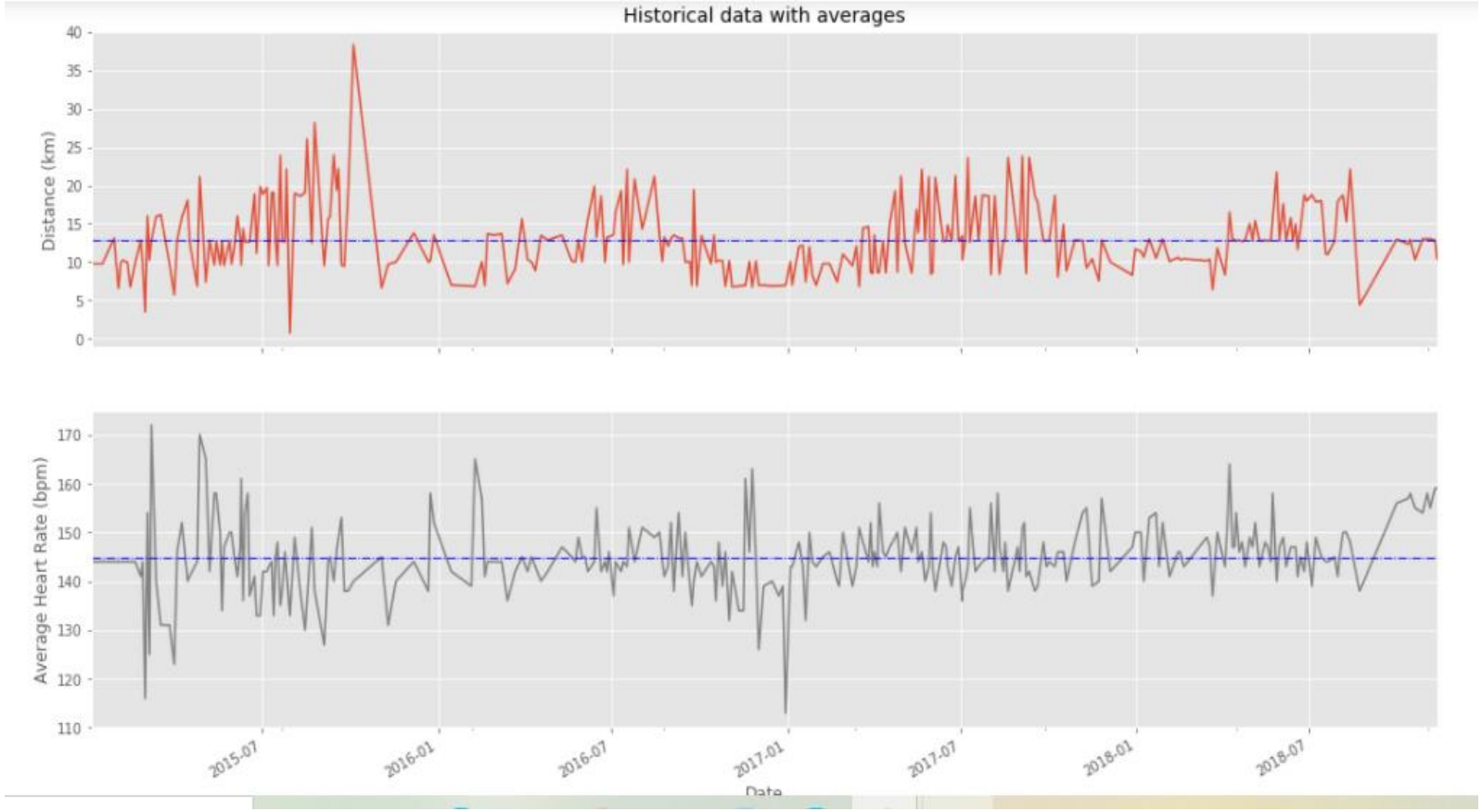
```
Distance (km)              12.518176
Average Speed (km/h)       10.835473
Climb (m)                 158.325444
Average Heart Rate (bpm)  144.801775
dtype: float64
```

How many trainings per week I had on average: 1.5

The Following Output gives :

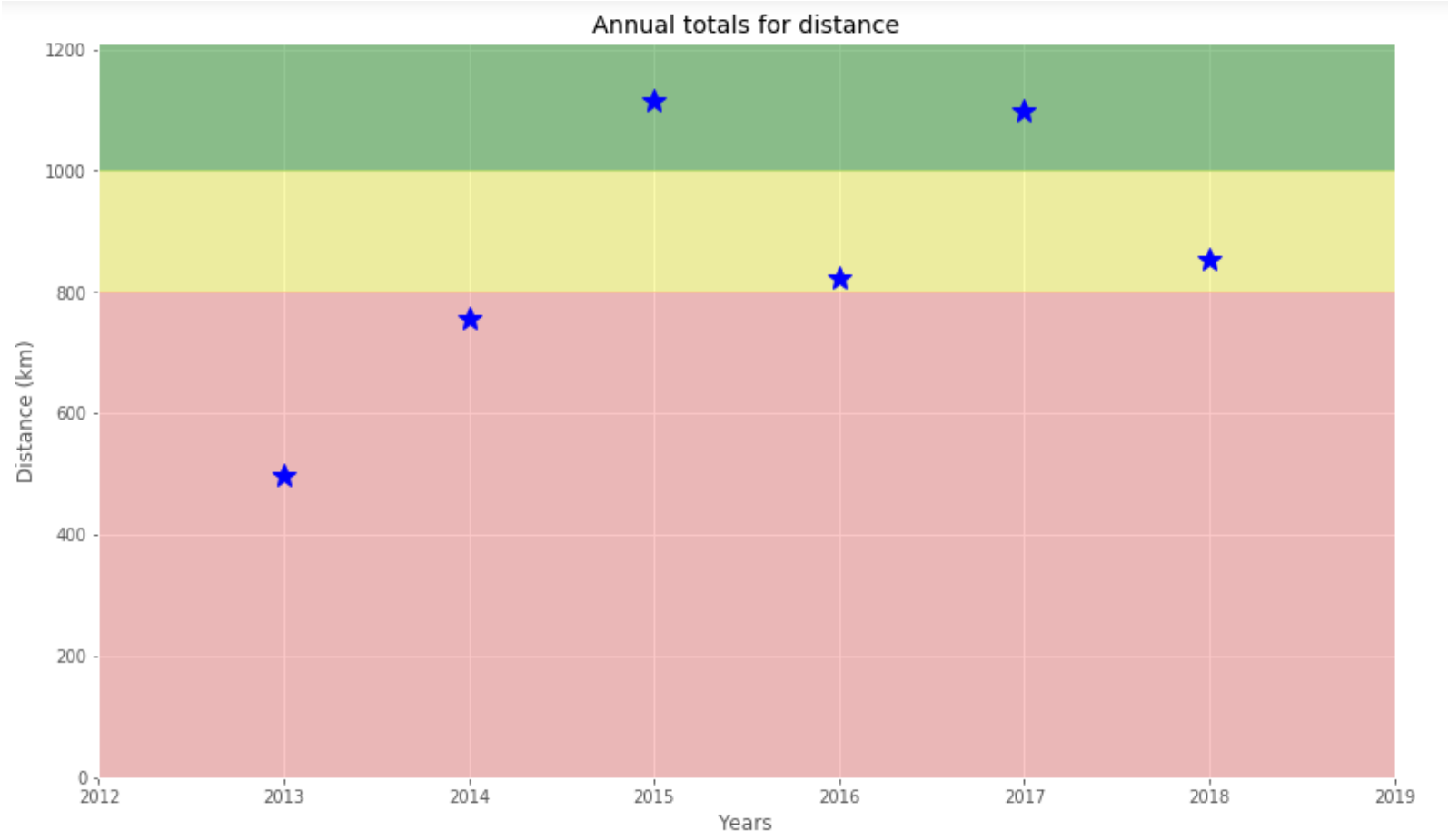
- The Annual Average Run for the last 4 Years according to the given data (2015 to 2018).
- The Weekly Average Run for last 4 years.
- The Average of how often did you train per week.

VISUALIZATION WITH AVERAGES



- This plot Shows the long term averages of the distance run and the heart rate from the year 2015 to 2018.

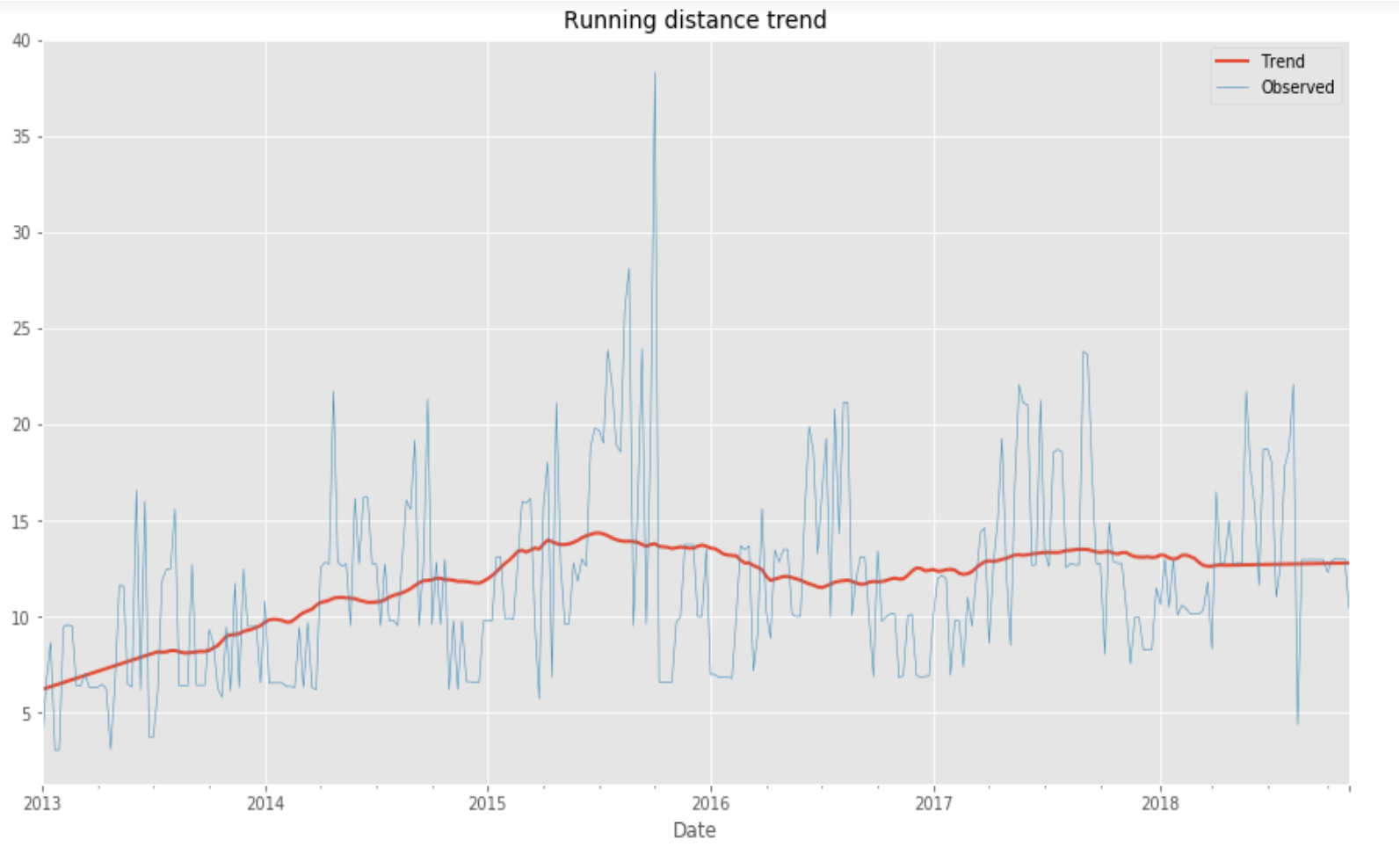
DID I REACH MY GOALS?



- In this case the person set a target goal of running 1000 km per year. So, This visualization from 2013 to 2018 shows that he reached his target for two years 2015 and 2017.

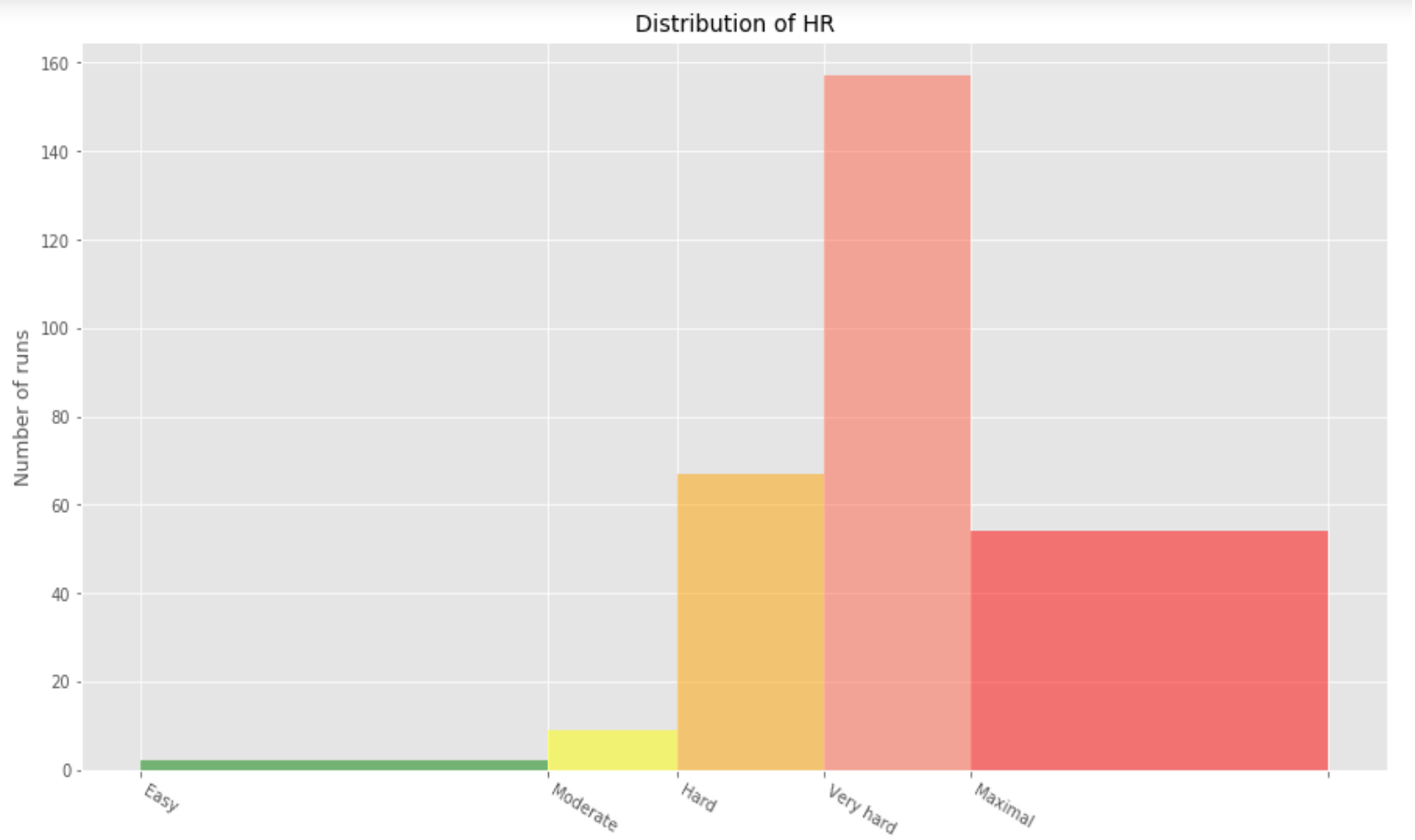
- The stars in the green region indicate success.

AM I PROGRESSING?



- This plot that decompose weekly distance run.
- A red trend line will represent the weekly distance run.
- use statsmodels library to decompose the weekly trend.
- This Trend shows that he made the some progress in 2018 compared to 2107 but the highest is in the year of 2015.

TRAINING INTENSITY



- Heart rate is a popular metric used to measure training intensity.
- This plot represents the training intensity by some levels.
- This is based on the number of runs, here its very hard distribution of heart rate when it is about 150 to 160 runs.

CONCLUSION

Based on the analysis of my Runkeeper Fitness data, The total Distance of running is 5224 kms . Based on this data I further used some analysis to discover an insight which is of 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 Forrest’s route of 24,700 km . This analysis is to find out how many pairs of shoes will be needed to complete the run in the same rate. So, 33 pairs of shoes are needed for the 24,700km run by assuming 7 pairs of shoes for 5224 kms. Therefore this analysis will be helpful in many ways mainly to know the progress that they make and motivates to further improve.