DROWSINESS DATA ANALYSIS

BY CARLO BLOCH

AGENDA

Objectives

Data Overview

Report &

...Visualization

Insights

Recommendations

OBJECTIVE:

TO PERFORM AN EXPLORATORY DATA ANALYSIS (EDA) ON A DATASET COLLECTED FROM WEARABLE DEVICES, INCLUDING VARIOUS PHYSIOLOGICAL PARAMETERS ALONG WITH A 'DROWSINESS' LEVEL, WHICH INDICATES THE LEVEL OF SLEEPINESS BASED ON AN ADAPTED KAROLINSKA SLEEPINESS SCALE (KSS).

DATA OVERVIEW

DATATSET ATTRIBUTES:

- HEARTRATE: HEART RATE READINGS FROM THE SMARTWATCH SENSORS
- PPGGREEN: PPG (*) SENSOR READINGS IN GREEN WAVELENGTH
- PPGRED: PPG SENSOR READINGS IN RED WAVELENGTH
- PPGIR: PPG SENSOR READINGS IN INFRARED WAVELENGTH
- DROWSINESS: LEVEL OF DROWSINESS BASED ON KSS (**)
- PERIOD: MORNING, AFTERNOON, EVENING, NIGHT

(**) KAROLINSKA SLEEPINESS SCALE (KSS)

EXTREMELY ALERT	1
VERY ALERT	2
ALERT	3
RATHER ALERT	4
NEITHER ALERT NOR SLEEPY	5
SOME SIGNS OF SLEEPINESS	6
SLEEPY, BUT NO EFFORT TO KEEP AWAKE	7
SLEEPY, BUT SOME EFFORT TO KEEP AWAKE	8
VERY SLEEPY, GREAT EFFORT TO KEEP AWAKE, FIGHTING SLEEP	9
EXTREMELY SLEEP, CANNOT KEEP AWAKE	10

THE ADAPTED VALUES ARE THREE, 0.0, 1.0 AND 2.0
AND ARE RELATED TO THE ORIGINAL KSS VALUE AS FOLLOWS:
0.0 DROWSINESS EQUIVALENT TO 1-3 KSS
1.0 DROWSINESS EQUIVALENT TO 4-8 KSS
2.0 DROWSINESS EQUIVALENT TO 9-10 KSS

(*) PPG STANDS FOR PHOTOPLETHYSMOGRAPHY.

PPG IS A WAY TO MEASURE BLOOD VOLUME AND WORKS BY MEANS OF LIGHT BEAMS.
THE WAY AND THE TIME THE LIGHT IS REFLECTED BY OUR BODY ALLOWS TO COMPUTE WITH A MATHEMATICAL FUNCTION A VALUE, WHICH IS THE PPG VALUE WE HAVE IN OUR CSV SAMPLE.
THE DIFFERENT COLORS ARE RELATED TO DIFFERENT SENSORS, HAVING DIFFERENT LIGHT FREQUENCY.
A PPG SIGNAL CAN BE AFFECTED BY VARIOUS FACTORS SUCH AS BUT NOT LIMITED TO MOVEMENT, TEMPERATURE, AMBIENT LIGHT, DEVICE VIDEO PROCESSING FILTERS, ETC.
USUALLY, THE GREEN LIGHT IS LESS ACCURATE THAN THE OTHERS, BUT IS THE ONLY WORKING GOOD UNDER STRESS, LIKE WORKING ON WET SKIN OR WITH RUNNING PEOPLE.
THE OTHER TWO SENSORS ARE EQUIVALENT, WITH INFRARED LIGHT ONES MOST ACCURATE IN THE MEASUREMENTS, BUT SOMETIMES GETTING BAD RESULTS WHEN WORKING UNDER STRESS

REPORT &

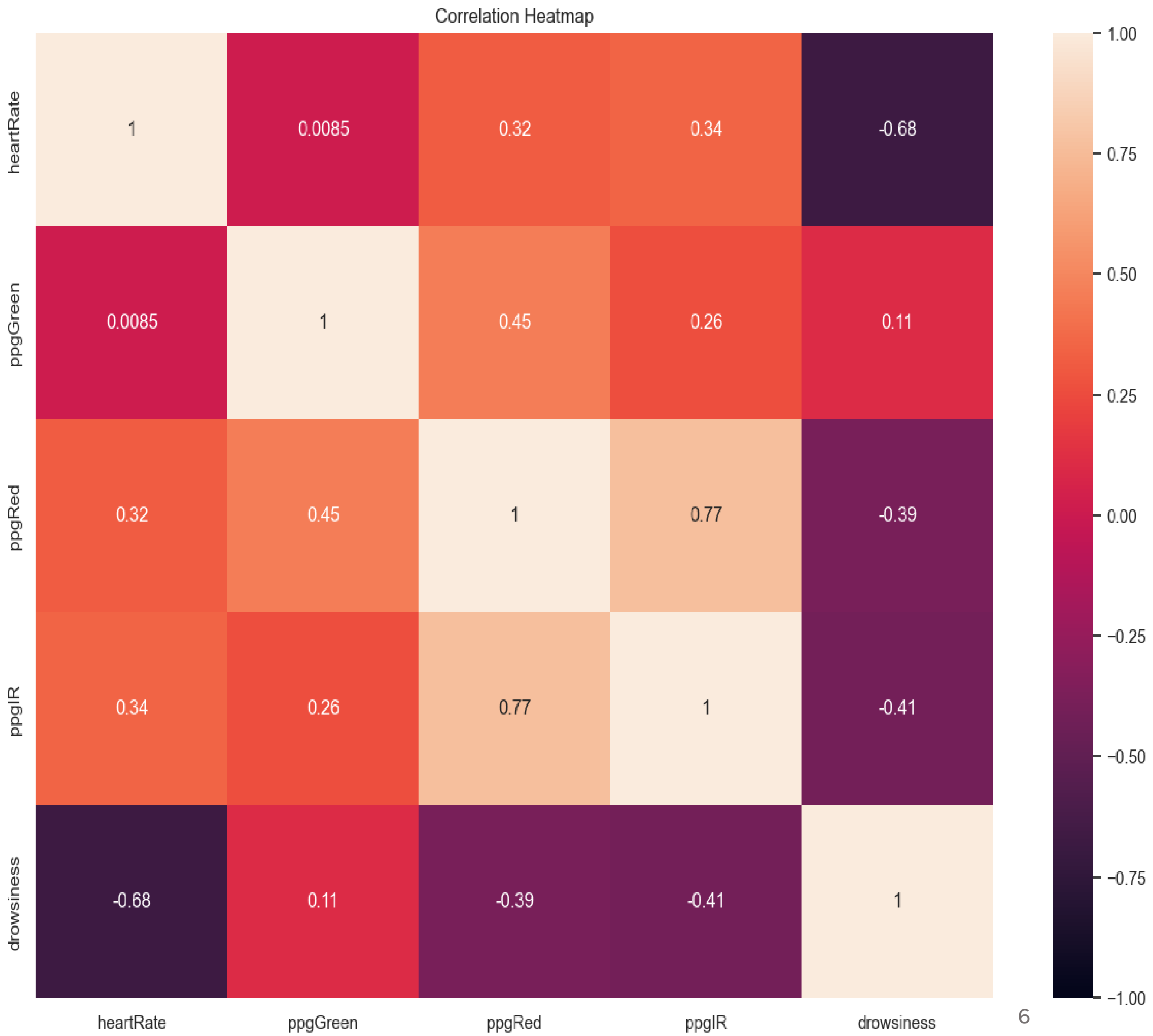
By comparing the different attributes of the records in the dataset within the following *Correlation Heatmap*, we found the following relationship:

I. **heartRate** is *strongly* correlated to **drowsiness**, in a negative versus, i.e. when heartRate increases drowsiness level decreases and viceversa

II. **ppgRed** and **ppgIR** are *strongly* related, in a positive versus, that is when one increases the other also increases, moreover their values overlap, with the only difference that the minimum value of IR sensors is lower than the red sensors and their readings are different because of the different wavelength

III. **ppgRed** is also a *little* related to **ppgGreen** attribute, in a positive versus, although they have different readings and ppgGreen < ppgRed always

…VISUALIZATION

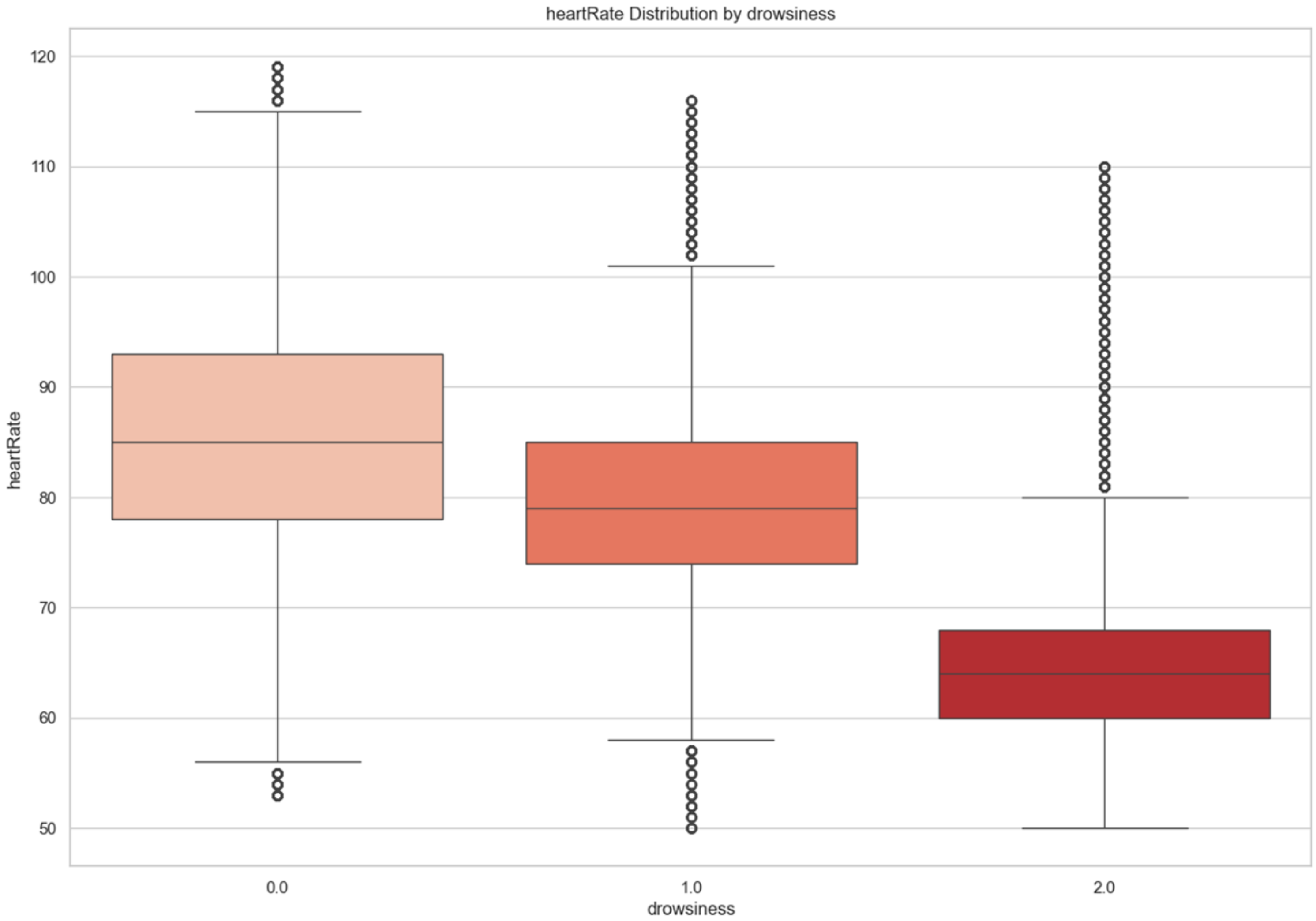


INSIGHTS

FROM POINT II) AND III) WE MAY SUGGEST USING CONFIGURATIONS WITH ONLY TWO SENSORS ACTIVE AT THE SAME TIME:
CONFIG. 1) IR + RED SENSORS
CONFIG. 2) GREEN + RED SENSORS
FROM POINT I) WE CAN STATE THAT WE HAVE SURE A WAY TO ASSOCIATE HEALTHY CARDIOVASCULAR VOLUMES TO DROWSINESS LEVELS.

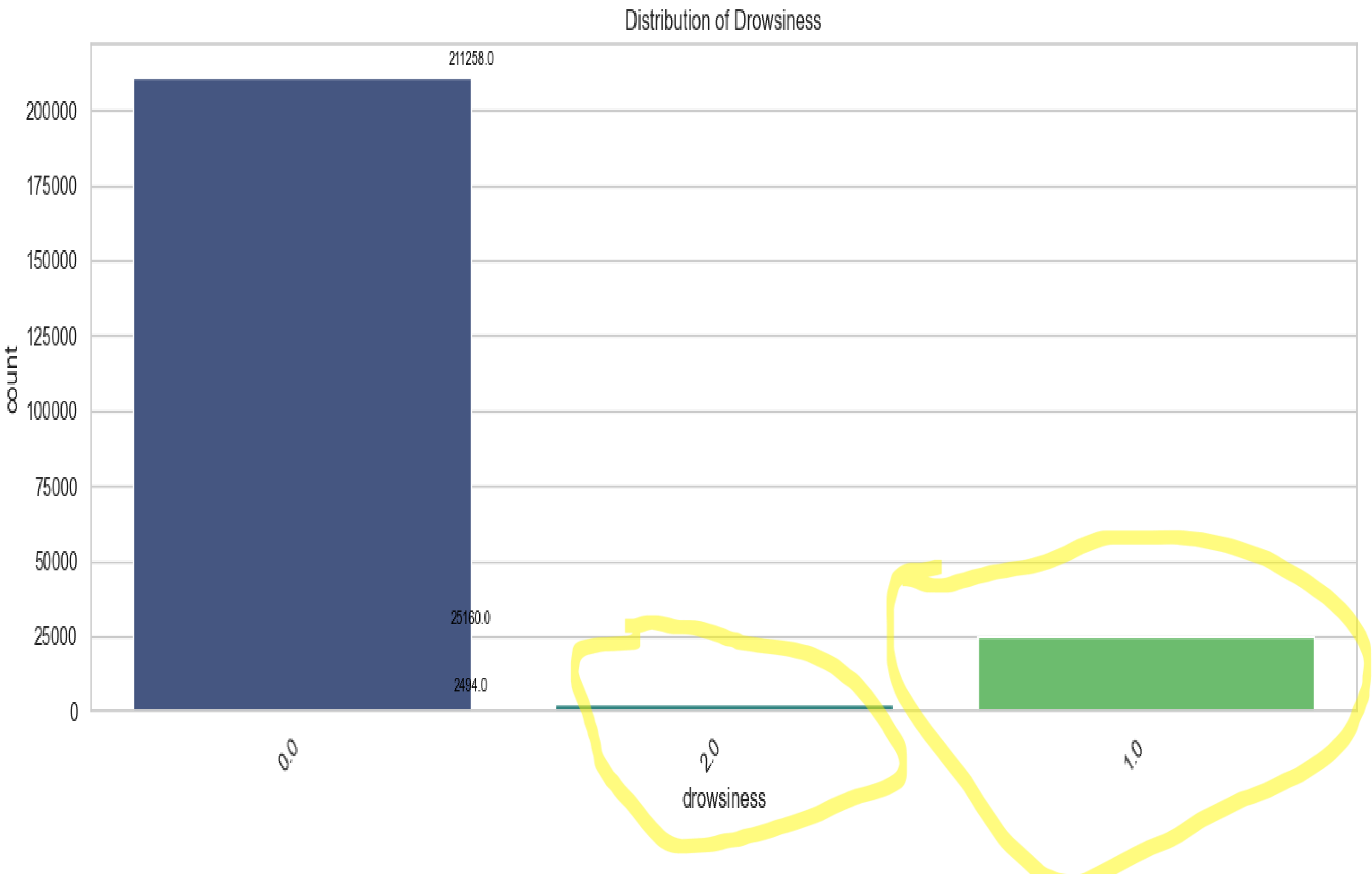
<u>SO, WE MAY ASSUME THAT A HEARTRATE > 100 IS ASSOCIATED TO PEOPLE MAKING PHYSICAL ACTIVITY, LIKE PLAYING SPORTS, HARD WORKING OR WALKING FAST,</u>
<u>I.E. ALL ACTIVITIES WHICH IMPLY PEOPLE BEING AWAKE, ALERT AND NOT SLEEPING</u>

HEART RATE VS. DROWSINESS



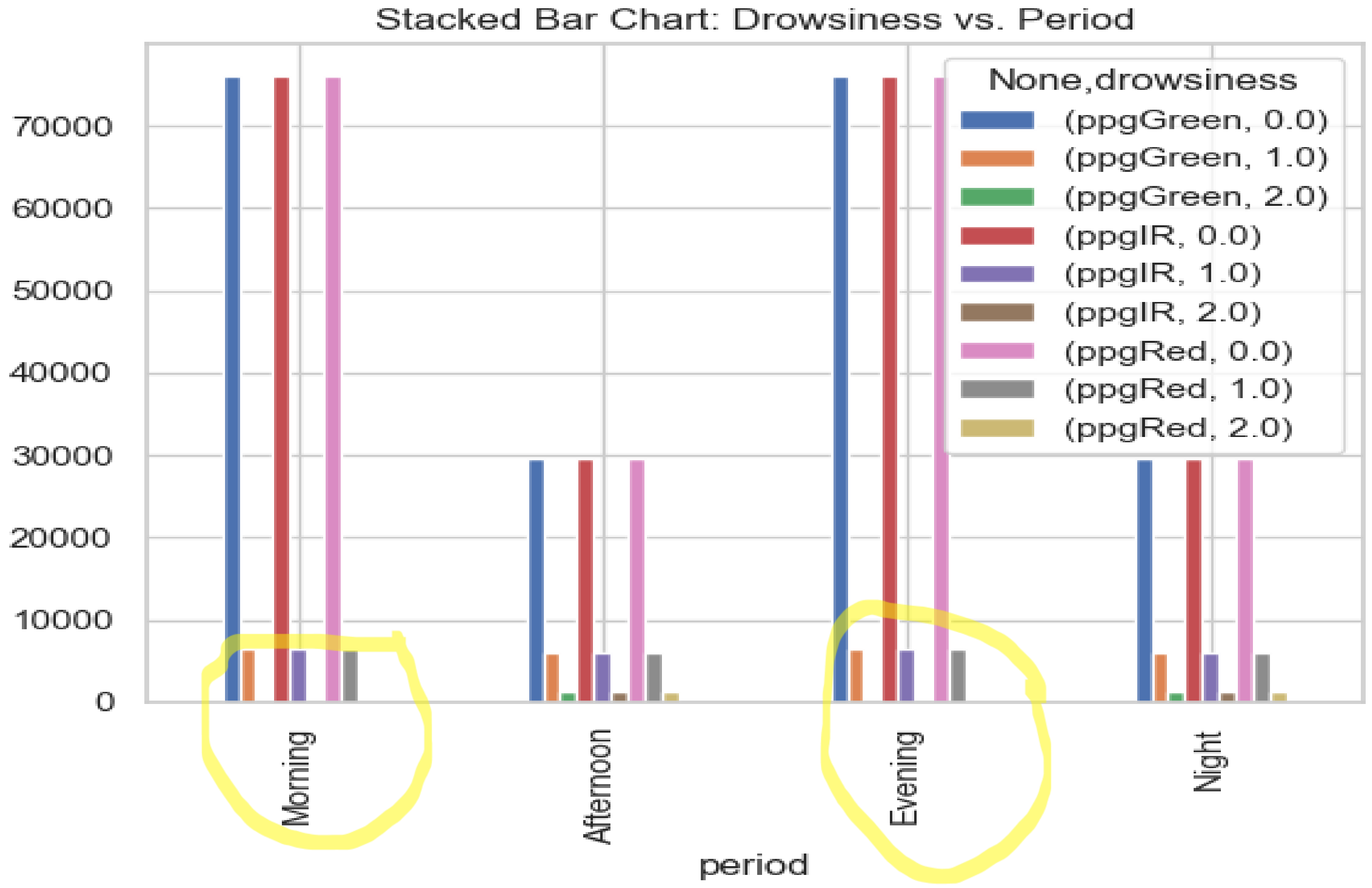
EXPLORING THE DATA I

A) With a heartRate > 100 we find that there are less measurements with drowsiness ≥ 1.0



EXPLORING THE DATA II

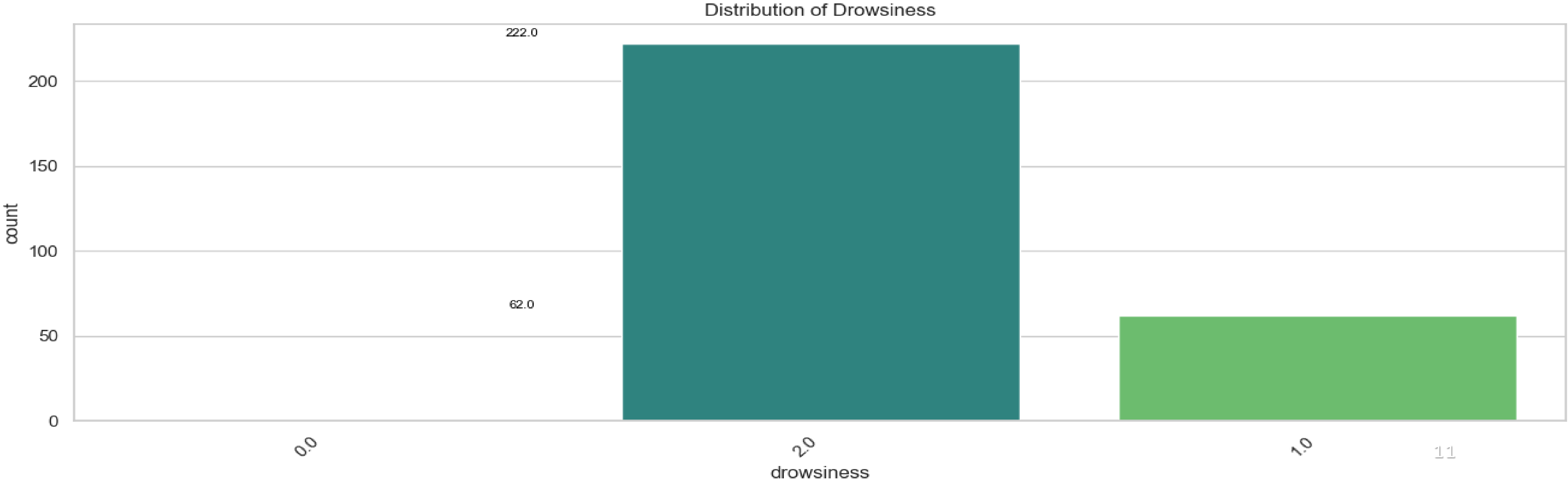
B) When heartRate > 100, drowsiness 2.0 it is missing in **Morning_data** and **Evening_data** samples, for all PPG measurements type



EXPLORING THE DATA III

When heartRate ≤ 50 we find that drowsiness ≥ 1.0 always, so normally it never happens for people being awake to have low heart rates.

Heart rates less than 50 but is a serious problem in any case and should be investigated



STATISTICS

We tested our hypoteses by means of CHI2 method or T-Test on

1. Sleepy subset, that is data with heartRate > 100 and drowsiness = 2.0 and
2. Awake subset, that is data with heartRate ≤ 50 and drowsiness != 2.0

```
1. CHI2 on data-subset where Drowsiness is 2.0 and hearRate is greater then 100
col: heartRate, Chi-Square 414459831.36999047, p-value 0.0, Degree of Freedom 7479
col: ppgGreen, Chi-Square 414459831.36999047, p-value 0.0, Degree of Freedom 7479
col: ppgRed, Chi-Square 414459831.36999047, p-value 0.0, Degree of Freedom 7479
col: ppgIR, Chi-Square 414459831.36999047, p-value 0.0, Degree of Freedom 7479
2. T-TEST on data-subset where Drowsiness is not 2.0 and hearRate is less then or equal to 50
heartRate: t-statistics = nan, p_value = nan
ppgGreen: t-statistics = 356.0146, p_value = 0.0000
ppgRed: t-statistics = 362.2490, p_value = 0.0000
ppgIR: t-statistics = 489.5751, p_value = 0.0000
drowsiness: t-statistics = -inf, p_value = 0.0000
```

RECOMMENDATIONS

According to the table here beside we can suggest focusing on 2 scenarios, where health issues may happen, or health risk is higher:

- **Health-risk 1**, people sleeping or tired and sleepy with a very high heart rate (hearRate ≥ 100) are at risk and should be alerted
- **Health-risk 2**, 50 ≤ heartRate ≤ 60 can happen only at rest, but a heartRate < 50 is to low and we should alert the user

Data-exploit	Result	NOTE
heartRate>100, period Morning or Evening	Drowsiness != 2.0	In *Morning* and *Evening* periods, we shall expect most of the people **being awake**
heartRate>100	Drowsiness≥1.0 is less frequent	This means that we may expect most of the people having a heartRate **lower** than 100 when sleeping or resting
heartRate≤50	Drowsiness != 0.0	Very low heart rates normally happen only when people is **NOT** completely awake
ppgRed corr. to ppgGreen or ppgIR	We can **use just 2 sensors** at the same time, instead of three	Devices can be configured in two setup using 2 sensors together, one with IR and **Red** sensors, one with Green and **Red** sensors
heartRate strong related to drowsiness levels	We may assume that with high hearth rates people are awake	We assumed the with heartRate>100 people are **phisically active**

THANKS!

Carlo Bloch