Human Stress Detection in and through Sleep Report

BY Pratishtha Sharma (IDS2022004) and Gomtesh Jain (IDS2022006)

Table Of Contents

About the data

Start off by reading in the csv file as a dataframe and grabbing a few rows of data to observe.

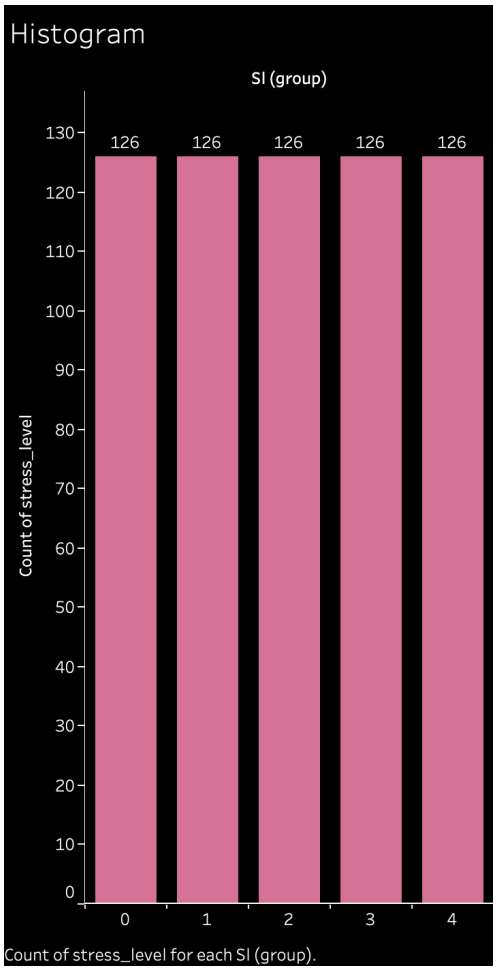
1. Snoring range of the user
2. Respiration rate (The number of breaths you take per minute. The normal respiration rate for an adult at rest is 12 to 20 breaths per minute.)
3. Body temperature (Fahrenheit)
4. Limb movement rate (measured in seconds)
5. Blood oxygen levels (measured in percentages, 90% - 100% is the normal range)
6. Eye movement (when going through REM sleep, our eyes dart around and this is one of the reasons why we may have dreams)
7. Number of hours of sleep
8. heart rate (BPM = beats per minute)
9. Stress Levels (0- low/normal, 1 – medium low, 2- medium, 3-medium high, 4 -high)

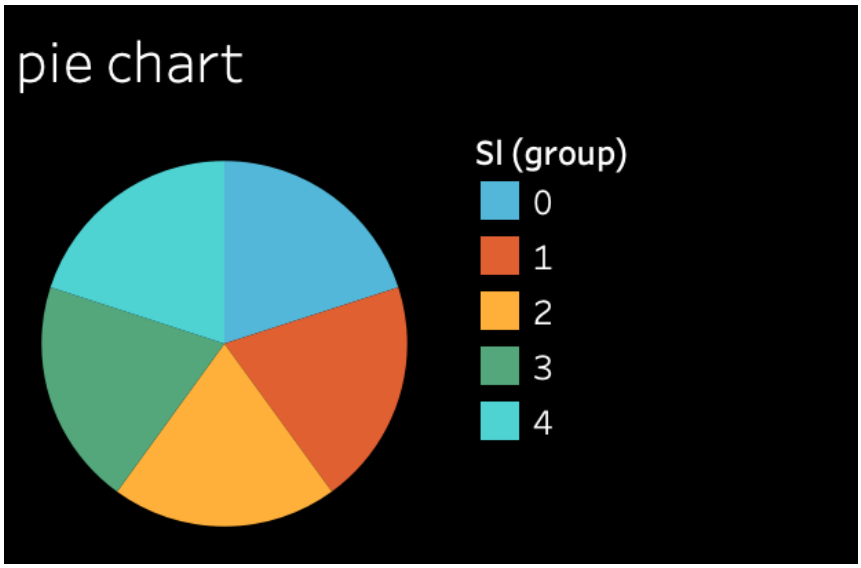
	snoring_rate	respiration_rate	body_temperature	limb_movement	blood_oxygen	eye_movement	sleeping_hours	heart_rate	stress_level
0	93.80	25.680	91.840	16.600	89.840	99.60	1.840	74.20	3
1	91.64	25.104	91.552	15.880	89.552	98.88	1.552	72.76	3
2	60.00	20.000	96.000	10.000	95.000	85.00	7.000	60.00	1
3	85.76	23.536	90.768	13.920	88.768	96.92	0.768	68.84	3
4	48.12	17.248	97.872	6.496	96.248	72.48	8.248	53.12	0

Tableau

Sheets

Sheet 1, 2 - Histogram and Pie chart of stress level





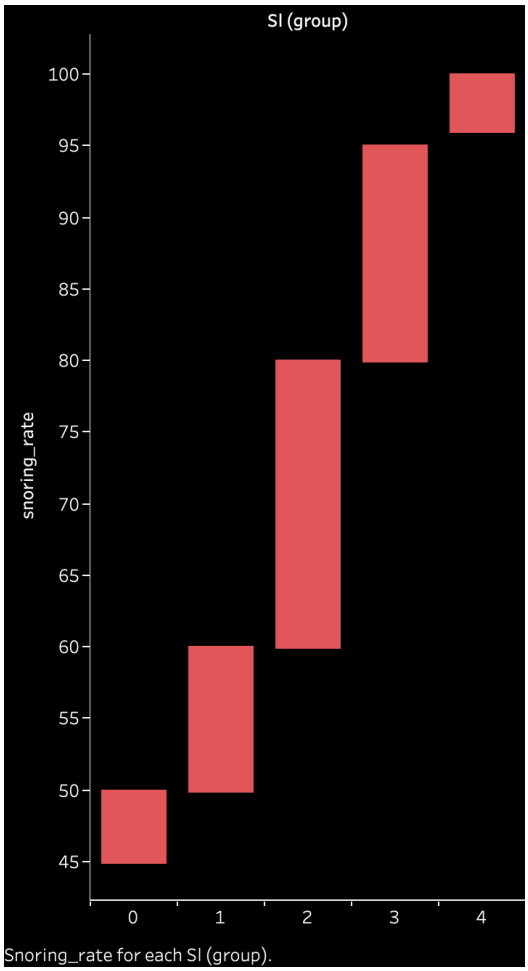
Histogram - A histogram is a bar graph-like representation of data that buckets a range of classes into columns along the horizontal x-axis. The vertical y-axis represents the number count or percentage of occurrences in the data for each column. Columns can be used to visualize patterns of data distributions.

histogram of the stress levels, and we get a uniform distribution. This dataset is very well-balanced and will be easy to use for analysis as we can see every stress level has exact count I.e., 126.

Pie chart- A pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice is proportional to the quantity it represents.

In this , Pie chart also show us same values I.e., 126 and it also show us uniform distribution.

Sheets 3 - Snoring rate for each Stress level



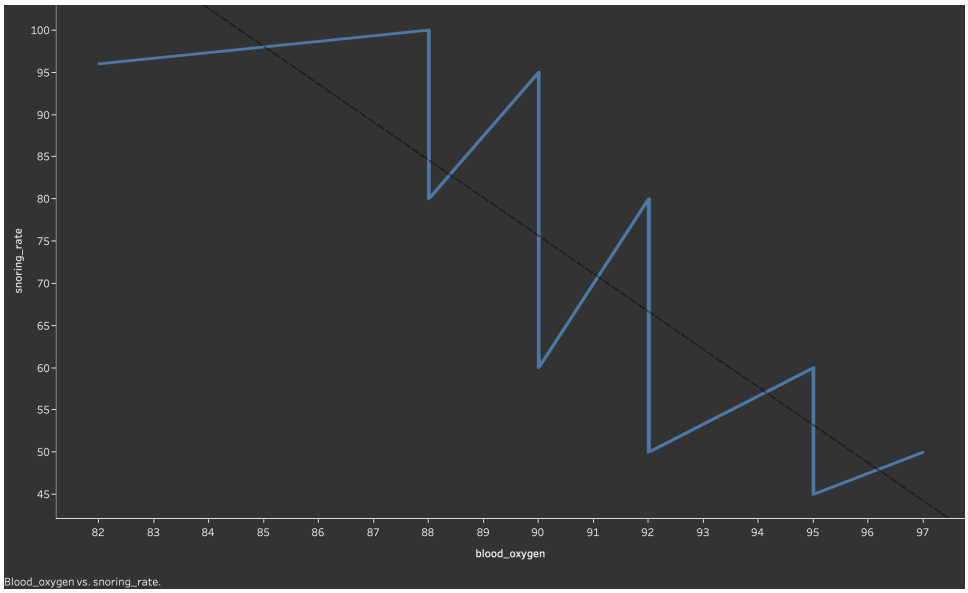
Correlation describes the strength of an association between two variables, and is completely symmetrical, the correlation between A and B is the same as the correlation between B and A.

A positive correlation is a relationship between two variables that move in tandem—that is, in the same direction. A positive correlation exists when one variable decreases as the other variable decreases, or one variable increases while the other increases.

A negative correlation is a relationship between two variables such that as the value of one variable increases, the other decreases. Correlation is expressed on a range from +1 to -1, known as the correlation coefficent. Values below zero express negative correlation.

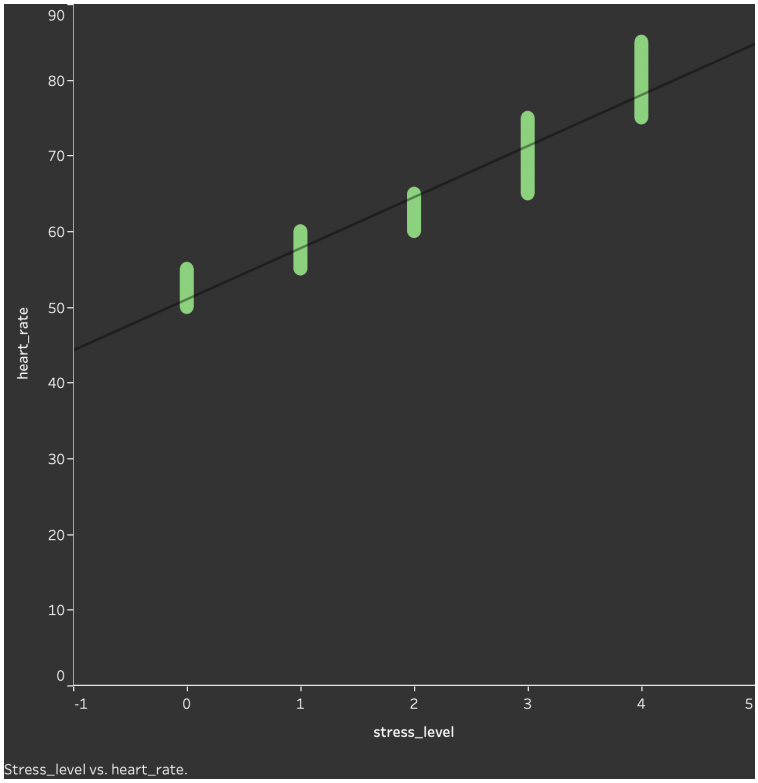
We can see a positive correlation between snoring rate and stress level, as we can see snoring rate increases level of stress also increases, and we have Corr score of 0.975322.

Sheet 4 - Blood Oxygen vs Snoring rate



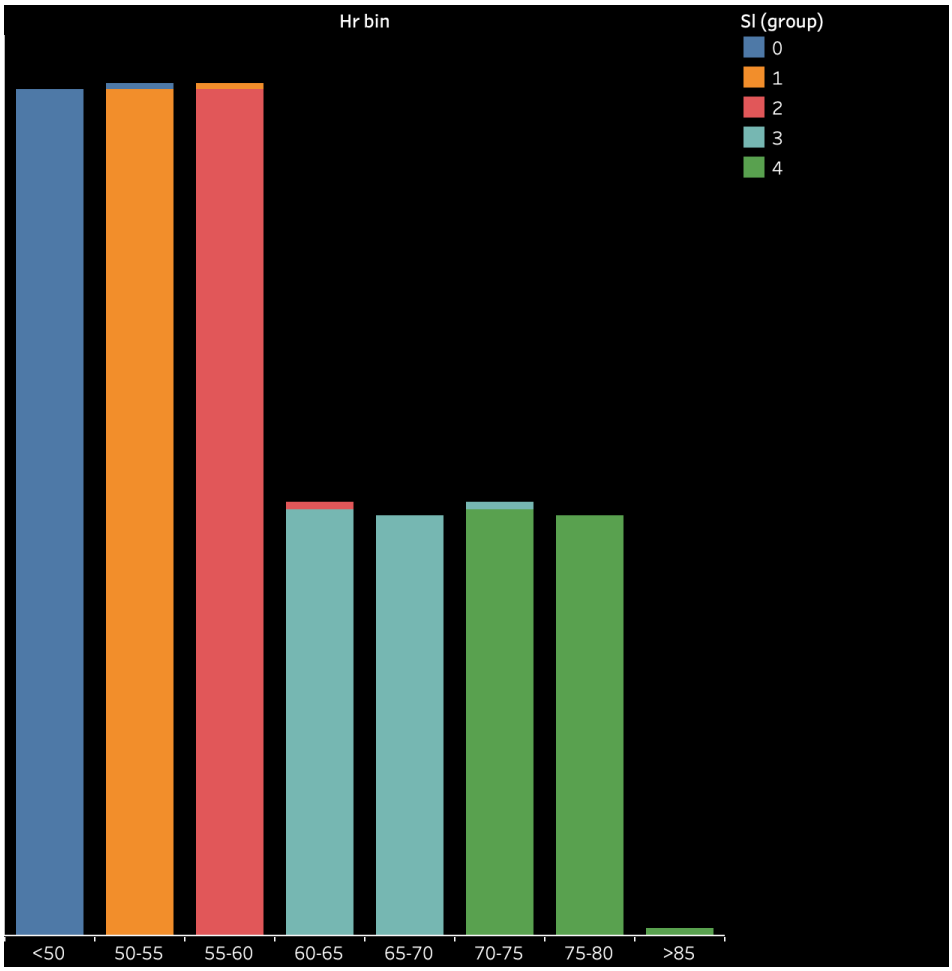
We can see a negative correlation between snoring rate and blood oxygen, as we can see snoring rate increases blood oxygen is also decrease, we have Corr score of -0.903140.

Sheet 5 - Stress level vs Heart rate

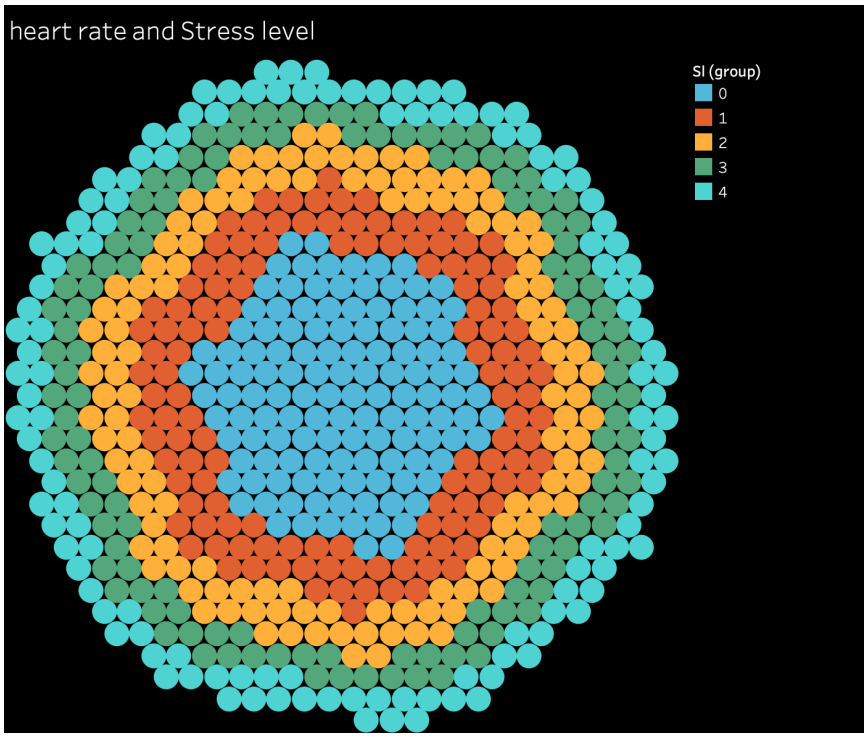


We can see a positive correlation between Heart rate and stress level, as we can see Heart rate increases level of stress is also increases, we have Corr score of 0.963516. They are extremely high.

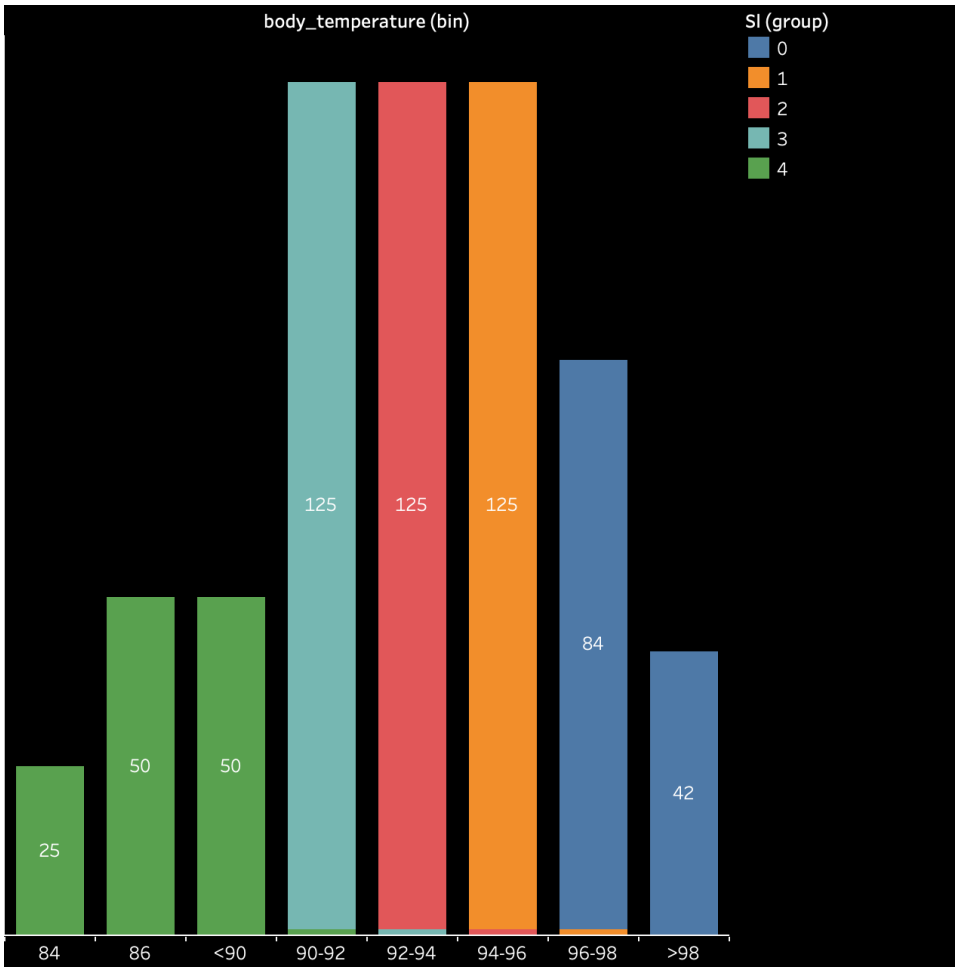
Sheet 6,7 - Heart rate and Stress level



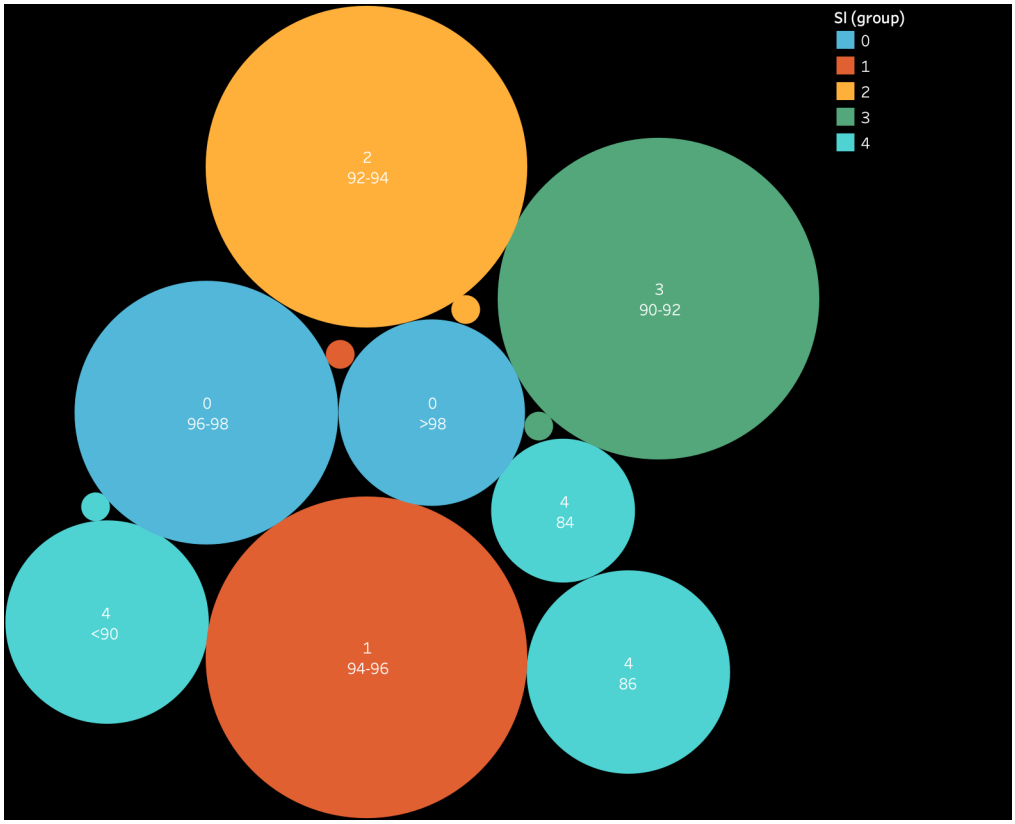
In this sheet, we have divided heart rate into bar of interval 5 and we can see that at <50 heart rate we have stress level zero and at 50-50 we have stress level 2 and so on.



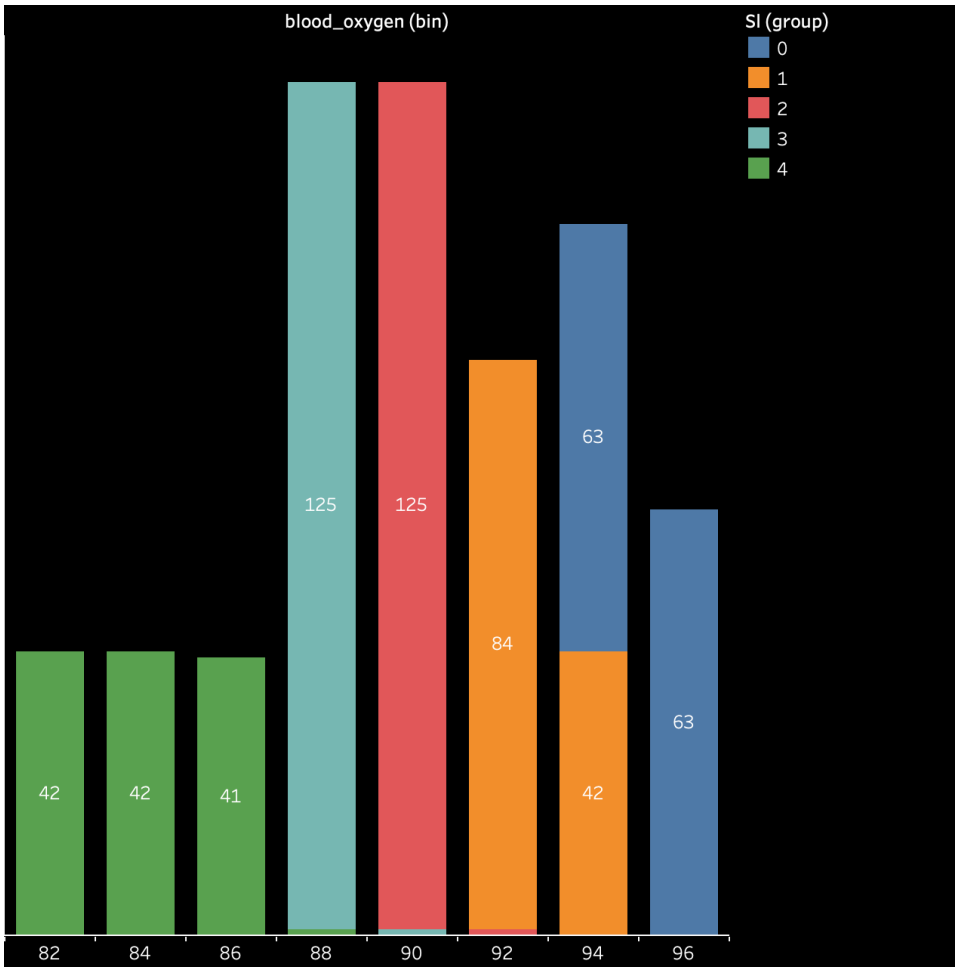
Sheet 8, 9 - Body temp. and stress level



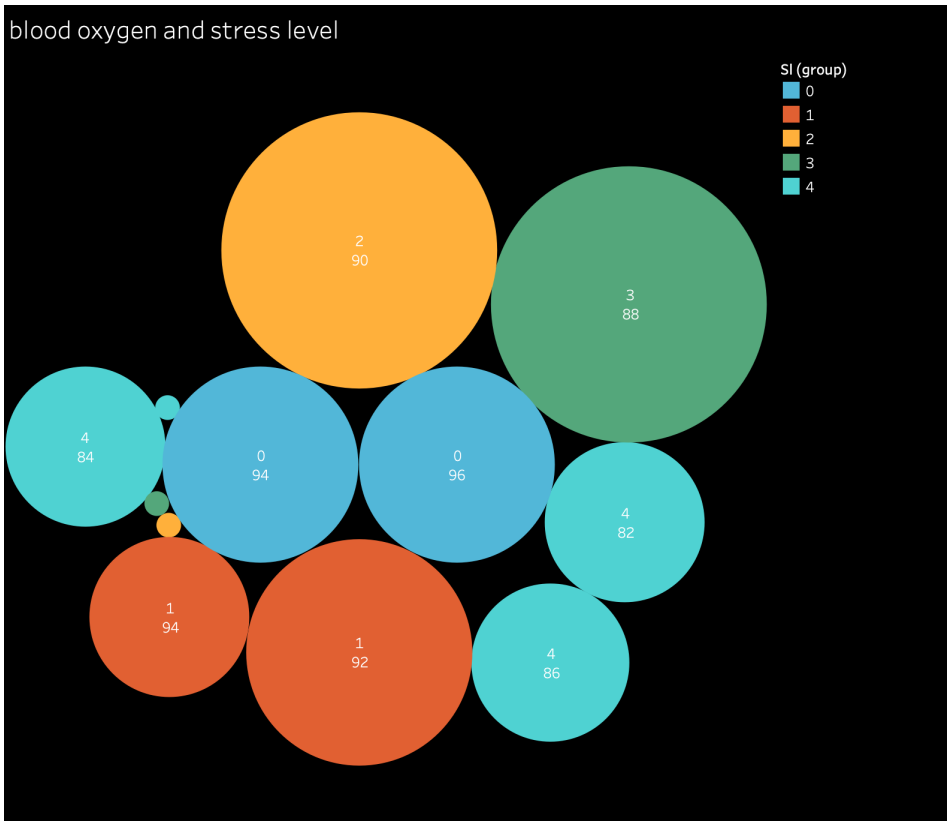
In this sheet, we have divided Body temperature into bars of interval 2 and we can see that at <90 Body temperature we have stress level 4 and at 90-92, we have stress level 3 and so on.



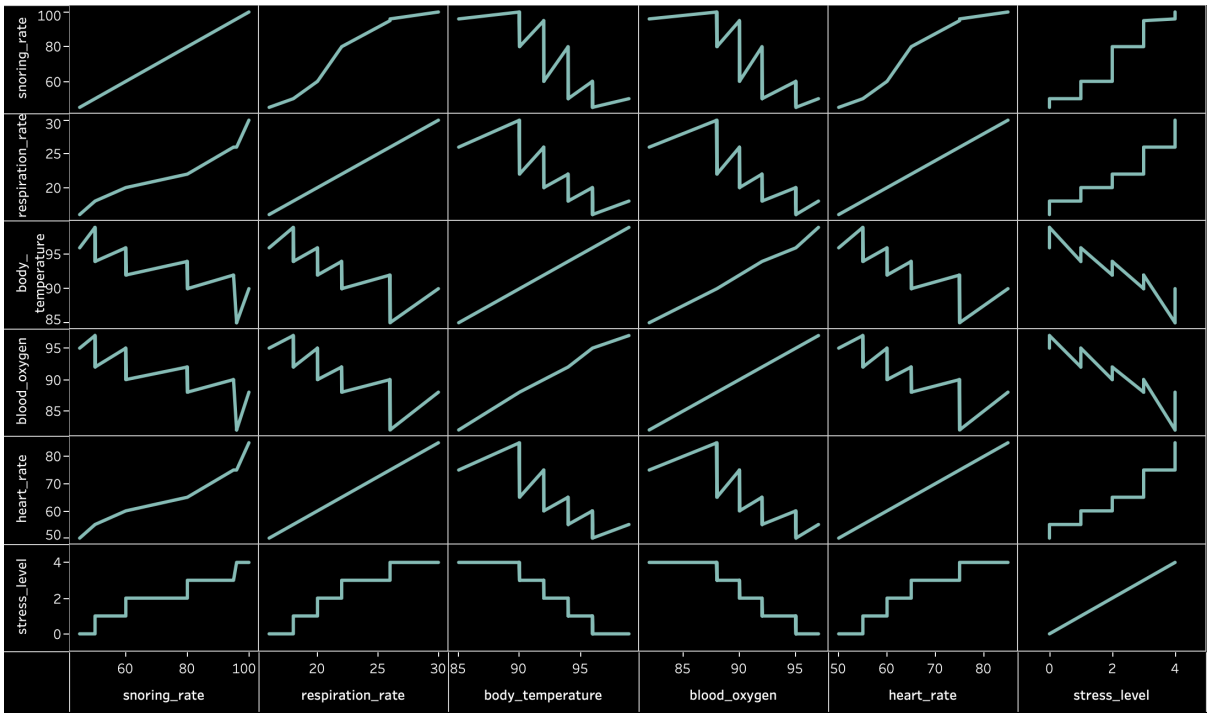
Sheet 10, 11 - blood oxygen and stress level



In this sheet, we have divided blood oxygen into bars of interval 2 and we can see that at <86 blood oxygen we have stress level 4 and at 88-90, we have stress level 3 and so on.



Sheet 12 - Correlation Matrix



In this sheet, we have made a pair plot which tells us about the correlation between the 2 filed and so we understated how correlated they are.

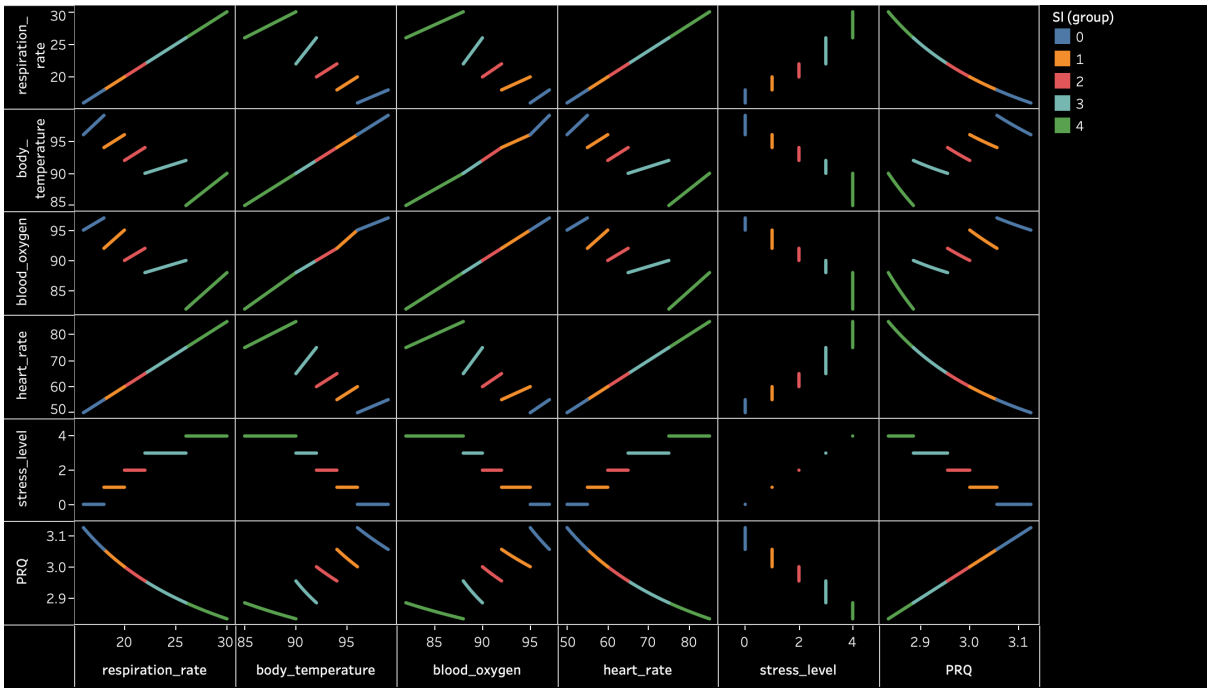
We can also say pairwise relationships between 2 filed and to visualize the correlation of each feature pair in a dataset against the class distribution, since a pair plot visually gives an idea of correlation of each feature pair, it helps us to understand and quickly analyse the correlation matrix (Pearson) of the dataset as well.

PRQ-

A specific and unique aspect of cardiorespiratory activity can be captured by dividing the heart rate (HR) by the respiration rate (RR), giving the pulse-respiration quotient (PRQ = HR/RR). In this review article, we summarize the main findings of studies using and investigating the PRQ.

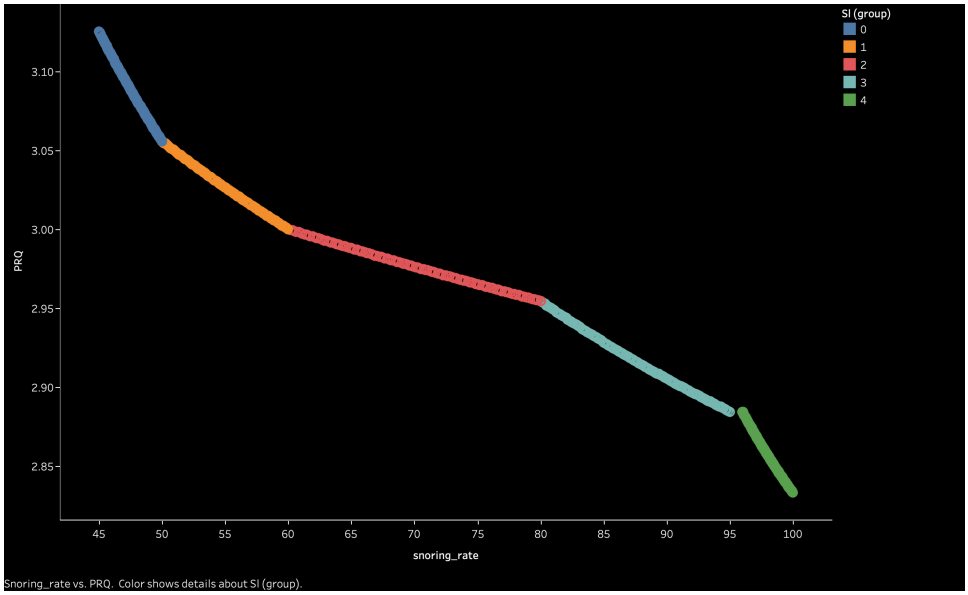
We describe why the PRQ is a powerful parameter that captures complex regulatory states of the cardiorespiratory system, and we highlight the need to re-introduce the use of this parameter in modern studies about human physiology and pathophysiology. We show that the PRQ (i) changes during human development, (ii) is time-dependent (ultradian, circadian, and infradian rhythms), (iii) shows specific patterns during sleep, (iv) changes with physical activity and body posture, (v) is linked with psychophysical and cognitive activity, (vi) is sex-dependent, and (vii) is determined by the individual physiological constitution. Furthermore, we discuss the medical aspects of the PRQ in terms of applications for disease classification and monitoring. Finally, we explain why there should be a revival in the use of the PRQ for basic research about human physiology and for applications in medicine, and we give recommendations for the use of the PRQ in studies and medical applications.

Sheets 13 - Corr with PRQ



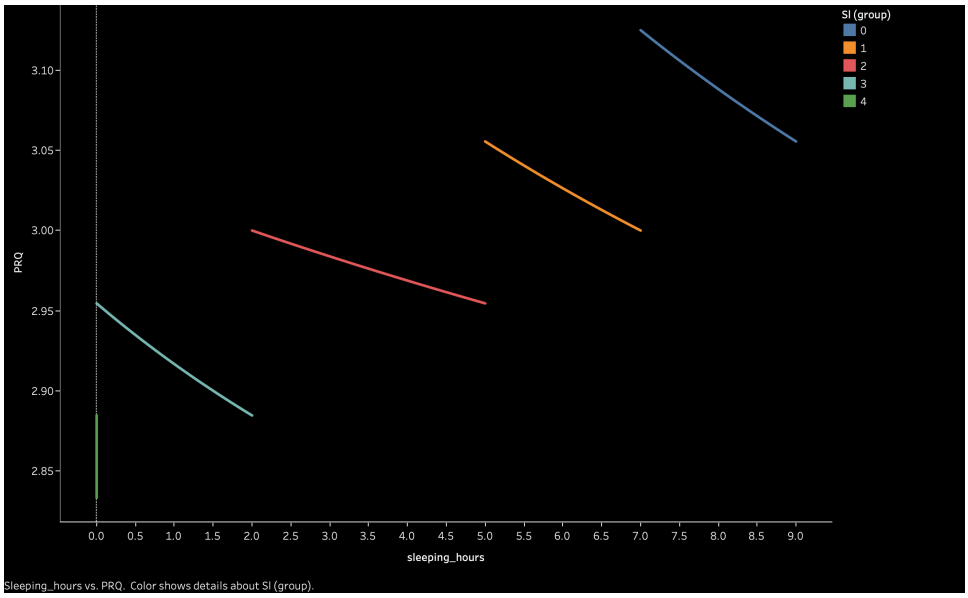
We have again made a pair plot but this time we have added PRQ so we see its correlation with other filed and we add colour that shows us the stress level.

Sheet 14 - PRQ vs Snoring Rate



We can see a negative correlation between snoring rate and PRQ, as we can see snoring rate increases PQR is also decrease, we have Corr score of -0.98.

Sheet 15 - PRQ vs sleeping hours



We can see a Positive correlation between sleeping hours and PRQ, as we can see sleeping hours increase PQR is also increases, and we have a Corr score of 0.91.

Visualization using python

Python is a great language for doing data analysis, primarily because of the fantastic ecosystem of data-centric Python packages. Pandas is one of those packages and makes importing and analyzing data much easier.

Displaying basic statistical details

```
df.describe()
```

	snoring_rate	respiration_rate	body_temperature	limb_movement	blood_oxygen	eye_movement	sleeping_hours	heart_rate	stress_level
count	630.000000	630.000000	630.00000	630.000000	630.000000	630.000000	630.000000	630.000000	630.000000
mean	71.600000	21.800000	92.80000	11.700000	90.900000	88.500000	3.700000	64.500000	2.000000
std	19.372833	3.966111	3.52969	4.299629	3.902483	11.893747	3.054572	9.915277	1.415337
min	45.000000	16.000000	85.00000	4.000000	82.000000	60.000000	0.000000	50.000000	0.000000
25%	52.500000	18.500000	90.50000	8.500000	88.500000	81.250000	0.500000	56.250000	1.000000
50%	70.000000	21.000000	93.00000	11.000000	91.000000	90.000000	3.500000	62.500000	2.000000
75%	91.250000	25.000000	95.50000	15.750000	94.250000	98.750000	6.500000	72.500000	3.000000
max	100.000000	30.000000	99.00000	19.000000	97.000000	105.000000	9.000000	85.000000	4.000000

We have viewed some basic statistical details like percentile, mean, std etc. of data.

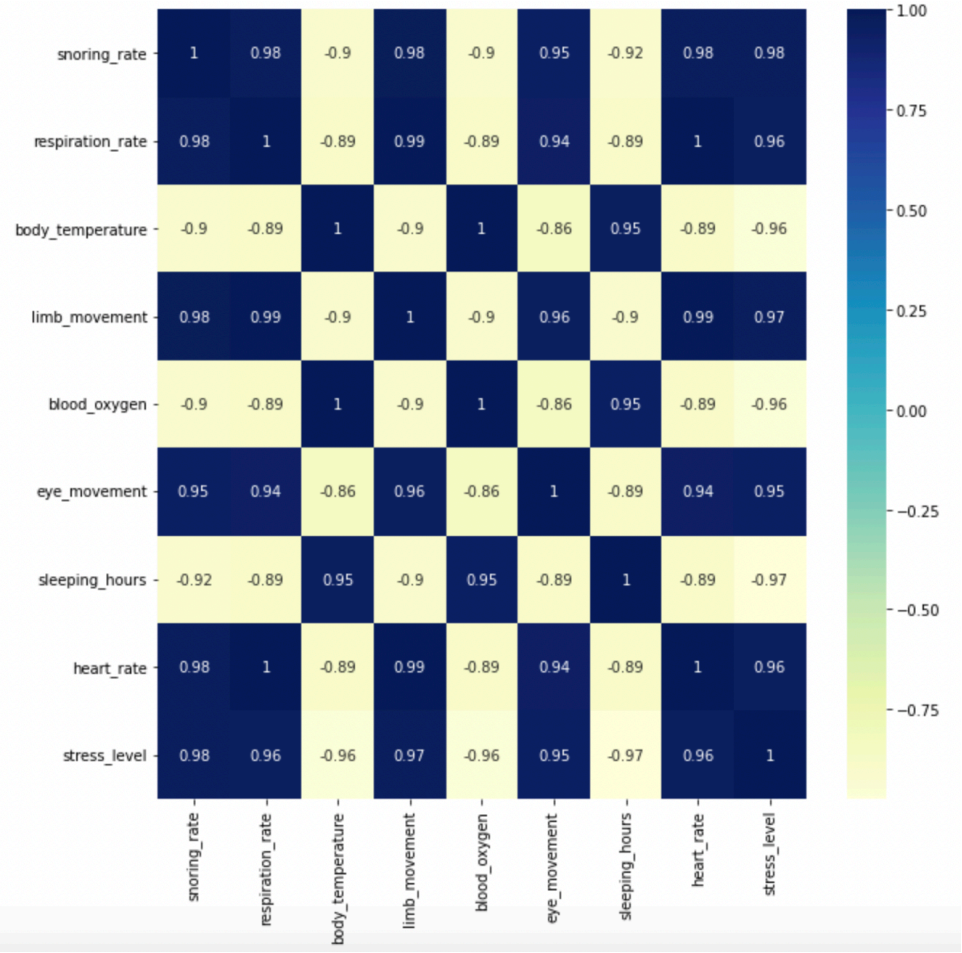
Check if data has null values

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 630 entries, 0 to 629
Data columns (total 9 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   snoring_rate      630 non-null    float64
 1   respiration_rate  630 non-null    float64
 2   body_temperature  630 non-null    float64
 3   limb_movement     630 non-null    float64
 4   blood_oxygen      630 non-null    float64
 5   eye_movement      630 non-null    float64
 6   sleeping_hours    630 non-null    float64
 7   heart_rate        630 non-null    float64
 8   stress_level      630 non-null    int64  
dtypes: float64(8), int64(1)
memory usage: 44.4 KB
```

We can see we do not have any null values and all attributes have float values except stress level it has int and distinct value from here we can say stress level has a categorical type. I.e., Stress Levels (0- low/normal, 1 – medium-low, 2- medium, 3 medium high, 4 -high)

Corr and heat map



Correlation Heatmaps are a type of plot that visualize the strength of relationships between numerical variables. Correlation plots are used to understand which variables are related to each other and the strength of this relationship.

We have made a heat map that shows the correlation between all the attributes. With the use of his we have made interesting tableau sheets.

Accuracy

	Model name	Accuracy scores
1	Logistic Regression	100.00
2	Gaussian Naive Bayes	100.00
4	SVC	100.00
6	KNN or k-Nearest Neighbors	100.00
3	Linear SVC	99.21
5	Random Forest	98.41
0	Decision Tree	96.83
8	Gradient Boosting	96.83
7	Stochastic Gradient Descent	90.48

We have applied Classification models like Decision Trees, Logistic Regression, Gaussian Naive Bayes, Linear SVC, SVC, Random Forest, KNN or k-Nearest Neighbors, Stochastic Gradient Descent and Gradient Boosting. we get best accuracy with Logistic Regression, Gaussian Naive Bayes, SVC and KNN or k-Nearest Neighbors for more you can refer above.

Cross validation Mean scores

	Model name	Cross validation mean scores
1	Logistic Regression	100.000000
2	Gaussian Naive Bayes	100.000000
4	SVC	100.000000
6	KNN or k-Nearest Neighbors	100.000000
3	Linear SVC	99.365079
5	Random Forest	98.941799
0	Decision Tree	98.571429
8	Gradient Boosting	98.359788
7	Stochastic Gradient Descent	94.074074

Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model

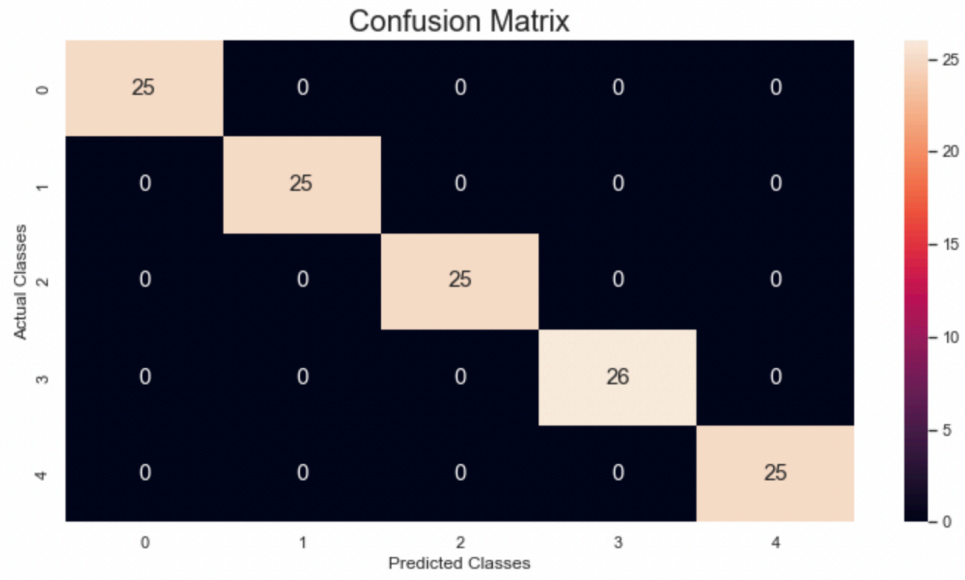
is expected to perform in general when used to make predictions on data not used during the training of the model.

Classification report

	0	1	2	3	4	accuracy	macro avg	weighted avg
precision	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
recall	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
f1-score	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
support	25.0	25.0	25.0	26.0	25.0	1.0	126.0	126.0

classification report is a performance evaluation metric in machine learning. It is used to show the precision, recall, F1 Score, and support of your trained classification model.

Confusion Matrix with Predicted Classes vs Actual Classes



A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm.

Story

Story Page 1

In this above dashboard we have created Stress level characterization with respect to Heart rate and body temperature and blood oxygen as it gives us the idea about linking all these graphs using their respective stress level

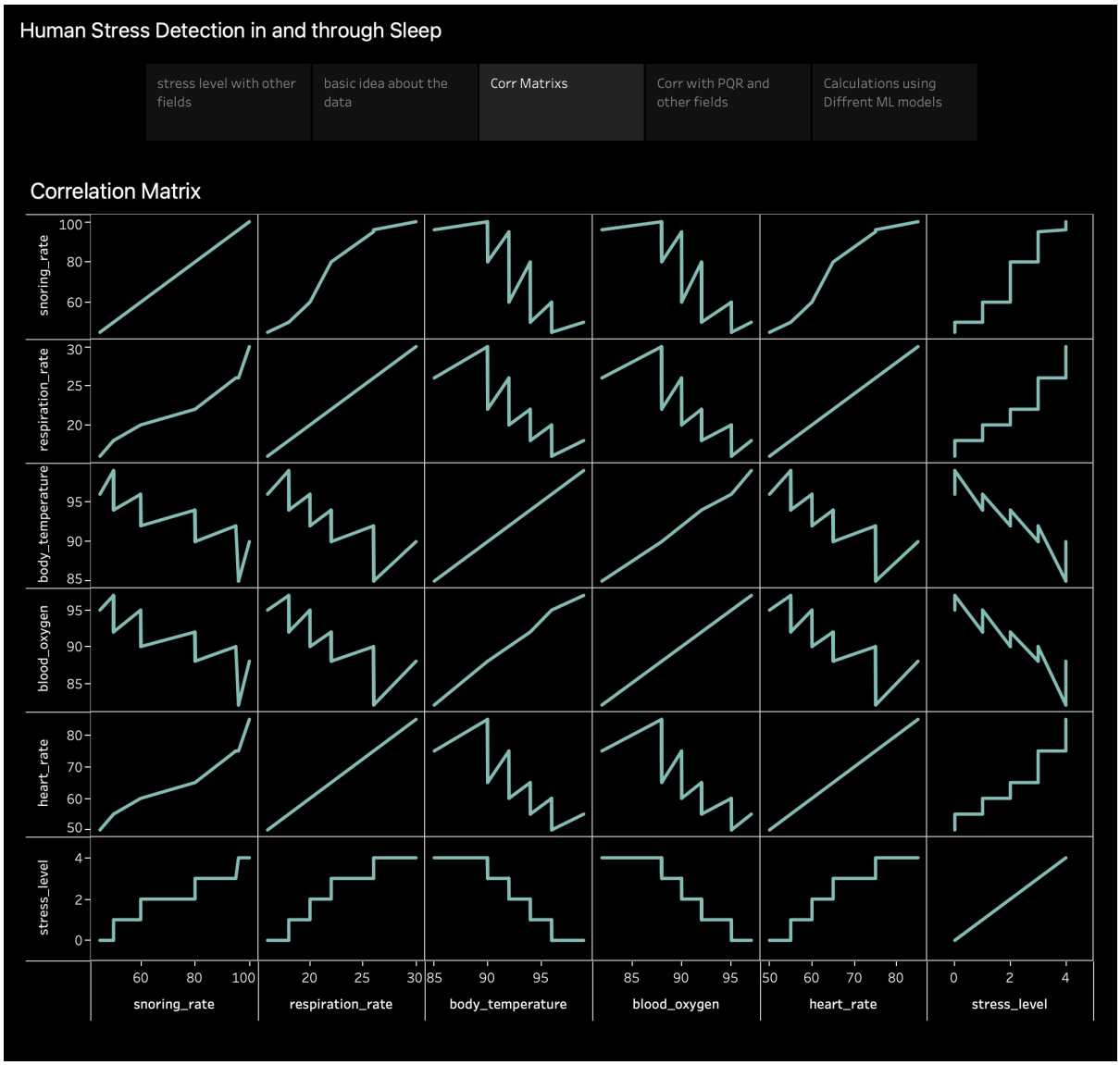
We have used research paper to dived filed like heart rate, body temperature and blood oxygen into bar groups (bin) so we can Cleary see the difference Stress level characterization

Story Page 2

In this dashboard, we have created a histogram of the stress levels, and we get a uniform distribution. This dataset is very well-balanced and will be easy to use for analysis as we can see every stress level has exact count I.e., 126.

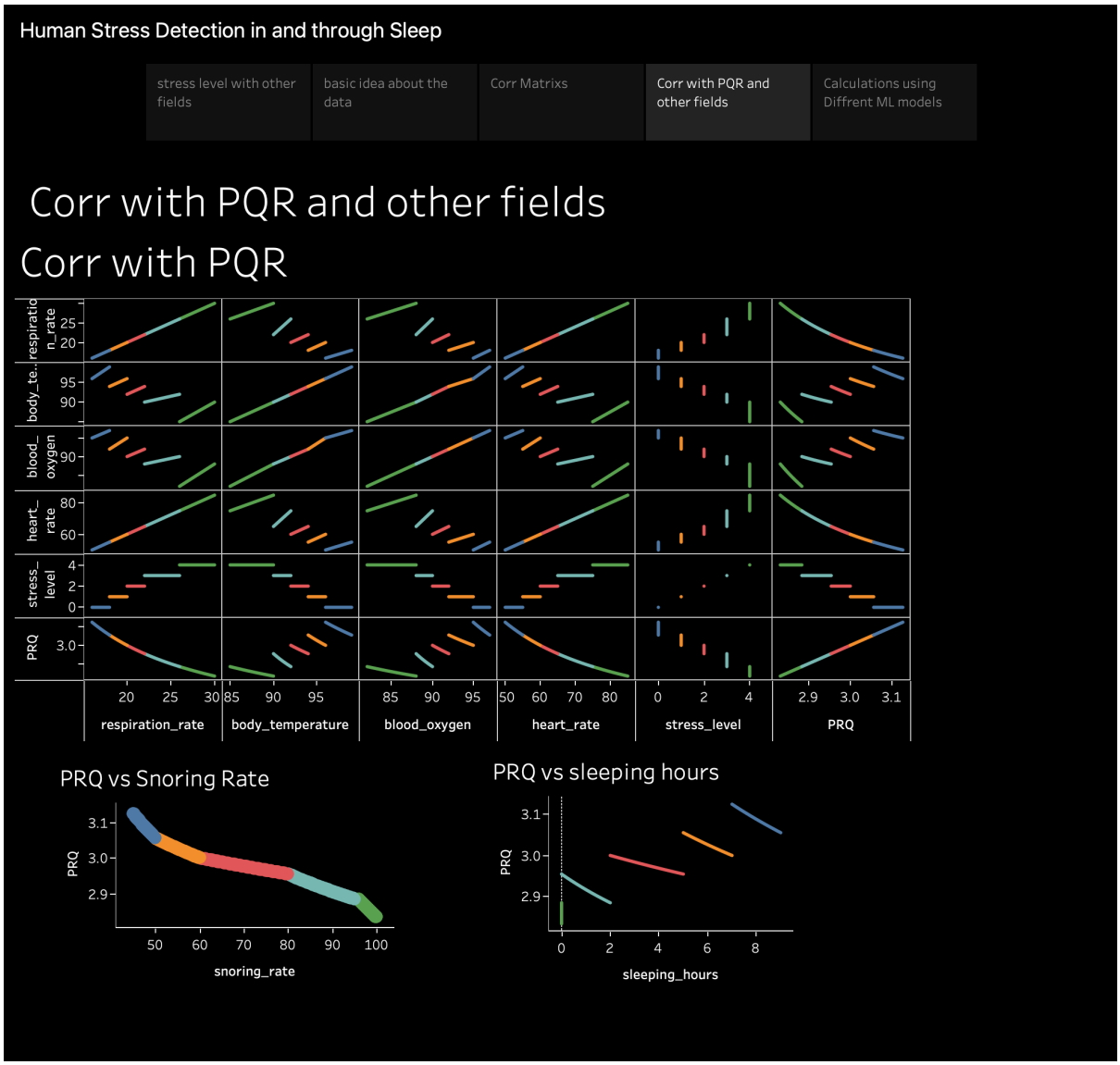
Then we have plotted a correlation graph between different filed.

Story Page 3



In this dashboard, we have created pair plot between different filed.

Story Page 4



In this dashboard, we have made a field i.e., PRQ

So, we have created a pair plot and we have plotted a correlation graph between PRQ and snoring rate that shows us negative correlation and a correlation graph between PRQ and sleeping hour that show us positive correlation.

Story Page 5

	Model name	Accuracy scores
1	Logistic Regression	100.00
2	Gaussian Naive Bayes	100.00
4	SVC	100.00
6	KNN or k-Nearest Neighbors	100.00
3	Linear SVC	99.21
5	Random Forest	98.41
0	Decision Tree	96.83
8	Gradient Boosting	96.83
7	Stochastic Gradient Descent	90.48

	0	1	2	3	4	accuracy	macro avg	weighted avg
precision	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
recall	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
f1-score	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
support	25.0	25.0	25.0	26.0	25.0	1.0	126.0	126.0

	Model name	Cross validation mean scores
1	Logistic Regression	100.000000
2	Gaussian Naive Bayes	100.000000
4	SVC	100.000000
6	KNN or k-Nearest Neighbors	100.000000
3	Linear SVC	99.365079
5	Random Forest	98.941799
0	Decision Tree	98.571429
8	Gradient Boosting	98.359788
7	Stochastic Gradient Descent	94.074074

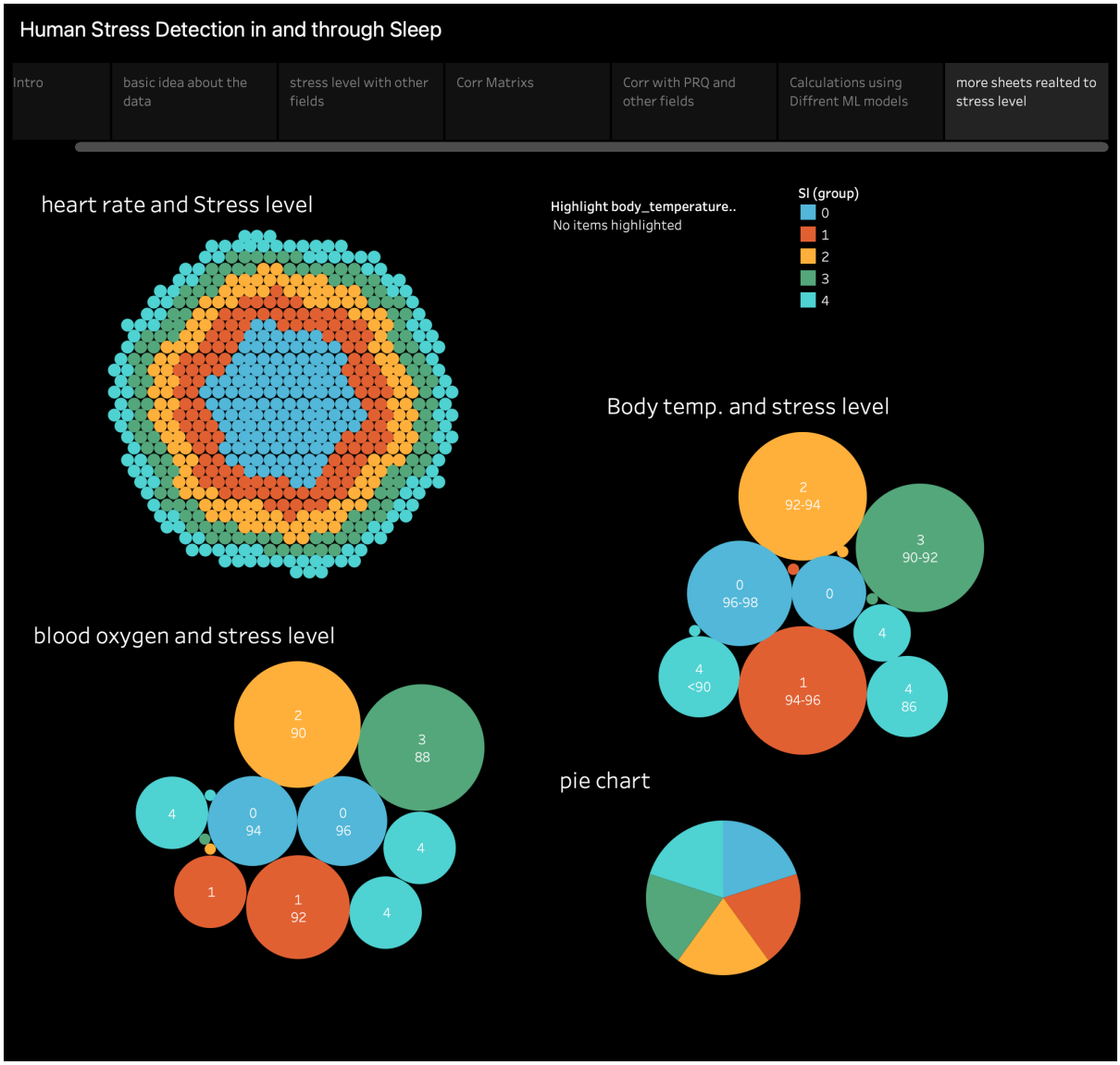
In this dashboard we have used python and applied different Classification models like Decision Tree, Logistic Regression, Gaussian Naive Bayes, Linear SVC, SVC, Random Forest, KNN or k-Nearest Neighbors, Stochastic Gradient Descent and Gradient Boosting.

we get best accuracy with Logistic Regression, Gaussian Naive Bayes, SVC and KNN or k-Nearest Neighbors for more you can refer above.

we get best Cross validation mean scores with Logistic Regression, Gaussian Naive Bayes, SVC and KNN or k-Nearest Neighbors for more you can refer above.

Then we have made classification report and plot Confusion Matrix.

Story Page 6



In this dashboard we have made it attractive and we can filter according to body temperature and different stress level and added Pie chart so it will show us count of stress level.

Future Scope of PRQ : Re-exploration

We need to re-introduce the use of this parameter into modern studies about human physiology and pathophysiology. In particular, we show that the PRQ (i) changes during human development, (ii) is time-dependent (ultradian, circadian, and infradian rhythms), (iii) shows specific patterns during sleep, (iv) changes with physical activity and body posture, (v) is linked with psychophysical and cognitive activity, (vi) is sex-dependent, and (vii) is determined by the individual physiological constitution. Furthermore, we discuss the medical aspects of the PRQ in terms of applications for disease classification and monitoring. Finally, we explain why there should be a revival in the use of the PRQ for basic research about human physiology and for applications in medicine, and we give recommendations for the use of the PRQ in studies and medical applications.

REFERENCES

1. SaYoPillow: Blockchain-Integrated Privacy-Assured IoMT Framework for Stress Management Considering Sleeping Habits
2. https://www.kaggle.com/datasets/laavanya/human-stress-detection-in-and-through-sleep?select=SaYoPillow.csv
3. https://www.researchgate.net/publication/342975717_SaYoPillow_A_Blockchain-Enabled_Privacy-Assured_Framework_for_Stress_Detection_Prediction_and_Control_Considering_Sleeping_Habits_in_the_IoMT
4. https://www.frontiersin.org/articles/10.3389/fphys.2019.00371/full#:~:text=A%20specific%20and%20unique%20aspect,using%20and%20investigating%20the%20PRQ'
5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6465339/#:~:text=A%20specific%20and%20unique%20aspect,using%20and%20investigating%20the%20PRQ