POST OPERATIVE PATIENT DATA

GROUP NUMBER. : 7&8

- Alekhya Bulusu
- Sushma Alimineti
- Amulya Kuntala
- Nagaphani Musunuri

Background :

Post operative care refers to any of your needs after a surgery. This could include cleaning the cuts , pain medication ,dressing cuts, eating or monitoring while sleeping .It is important that it's done right in order for your body to recover properly.

Objective :

To predict the discharge decision based on the attributes given using python programming.

Path :

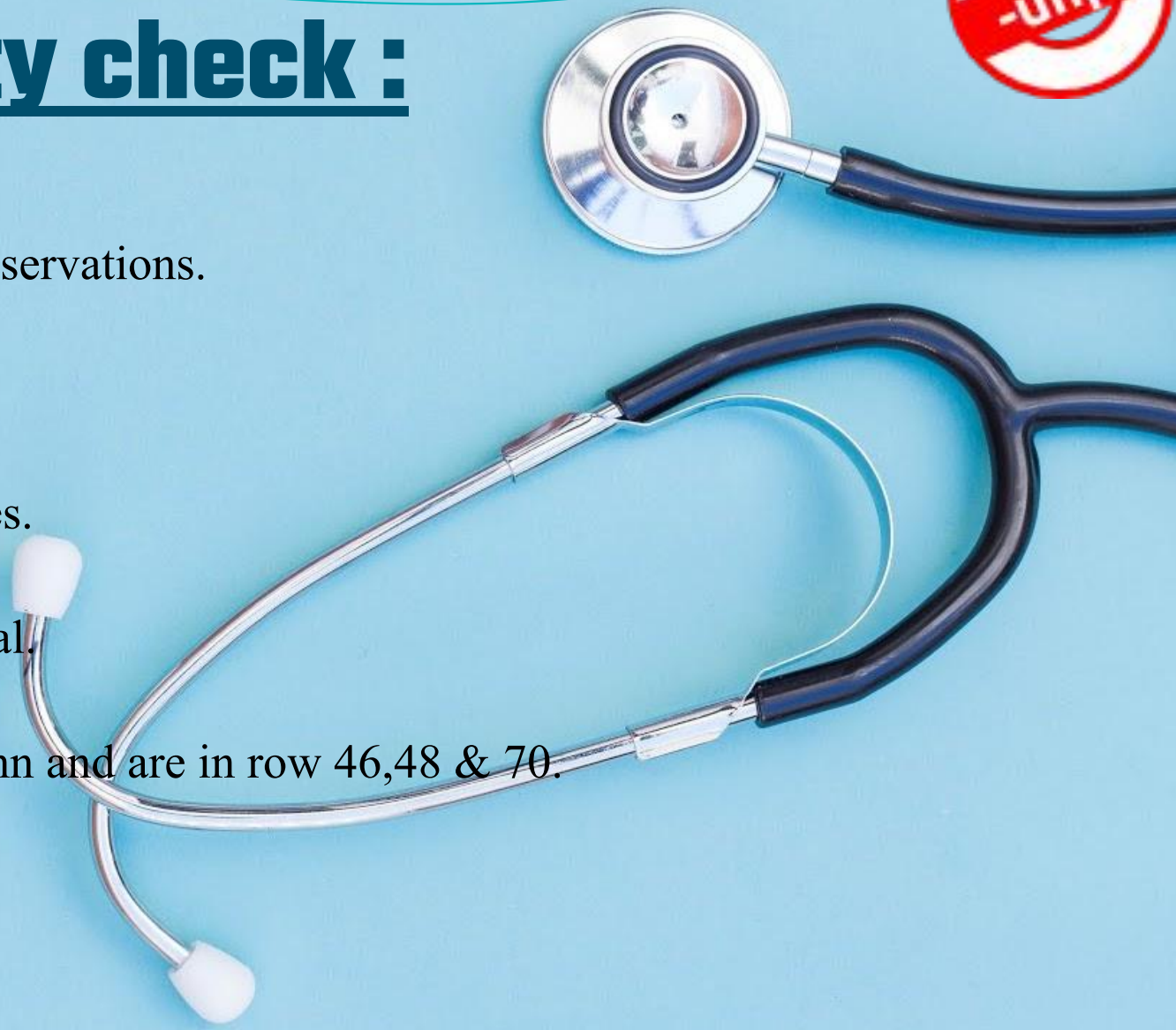
A data set is considered to predict the discharge decision based on the attributes. The data is analyzed using machine learning algorithms, one of which is logistic regression.

A glimpse of the patient data

INDEX	L-CORE	L-SURF	L-O2	L-BP	SURF-STBL	CORE-STBL	BP-STBL	COMFORT	decision ADM-DECS
0	mid	low	excellent	mid	stable	stable	stable	15	A
1	mid	high	excellent	high	stable	stable	stable	10	S
2	high	low	excellent	high	stable	stable	mod-stable	10	A
3	mid	low	good	high	stable	unstable	mod-stable	15	A
4	mid	mid	excellent	high	stable	stable	stable	10	A

- **L-CORE = Patient's internal temperature**
- **L-SURF = Patient's surface temperature**
- **L-O2 = Oxygen saturation**
- **L-BP = Last measurement of BP**
- **SURF-STBL = Stability of patient's surface temp.**
- **CORE-STBL= Stability of patient's core temp.**
- **BP-STBL= Stability of patient's BP.**

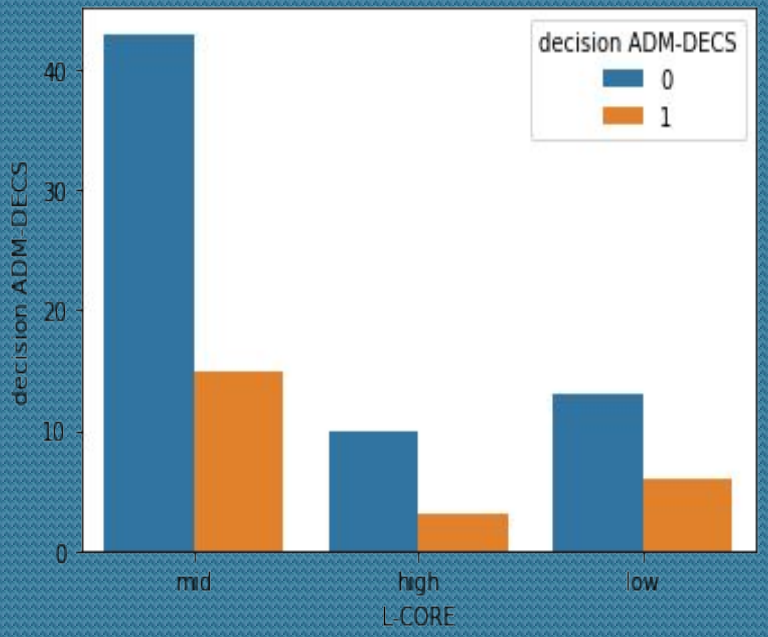
Data and data quality check :

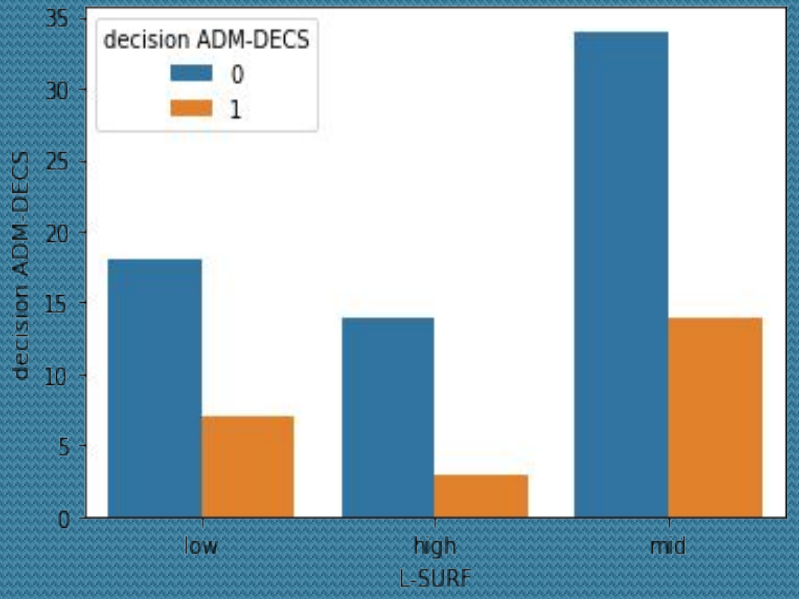


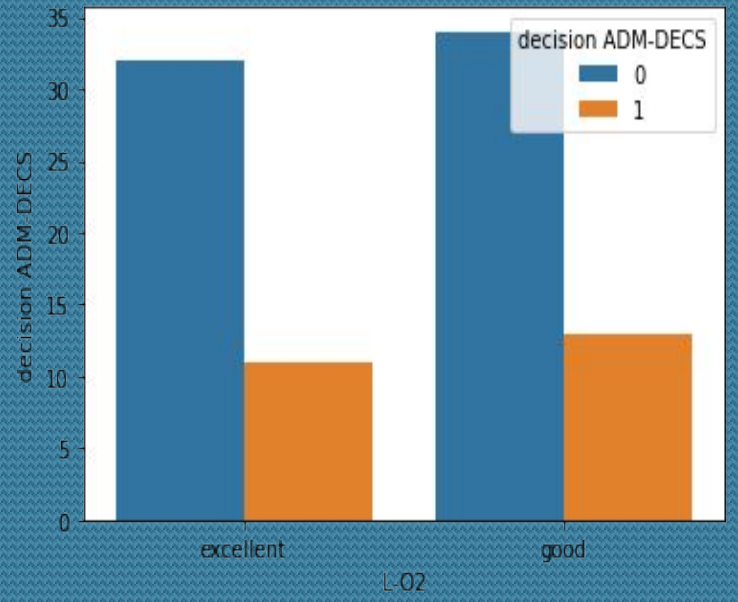
- The given data contains 9 attributes & 90 observations.
- Output variables : Discharge Decision.
- The remaining 8 attributes are input variables.
- The characteristic of attributes are categorical.
- The missing values are in COMFORT column and are in row 46,48 & 70.

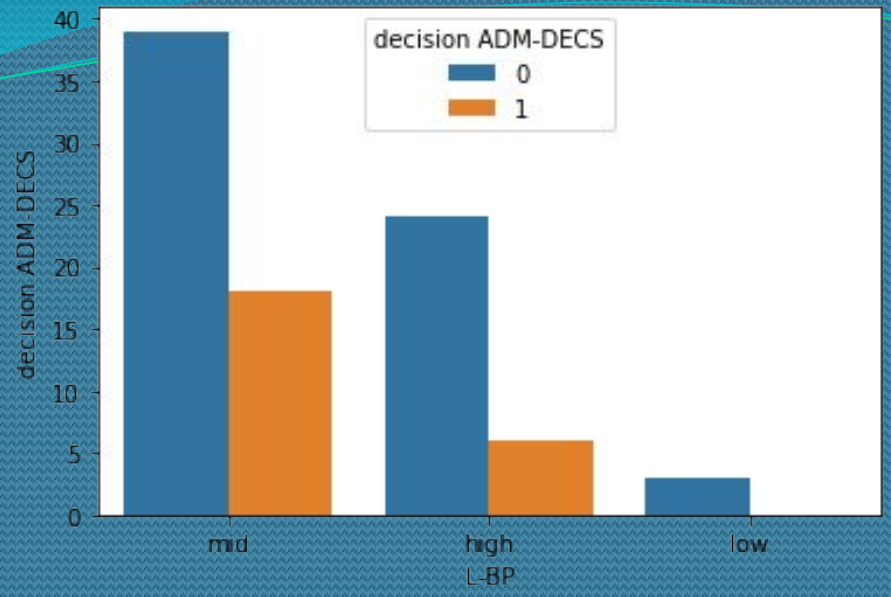
EDA INSIGHTS

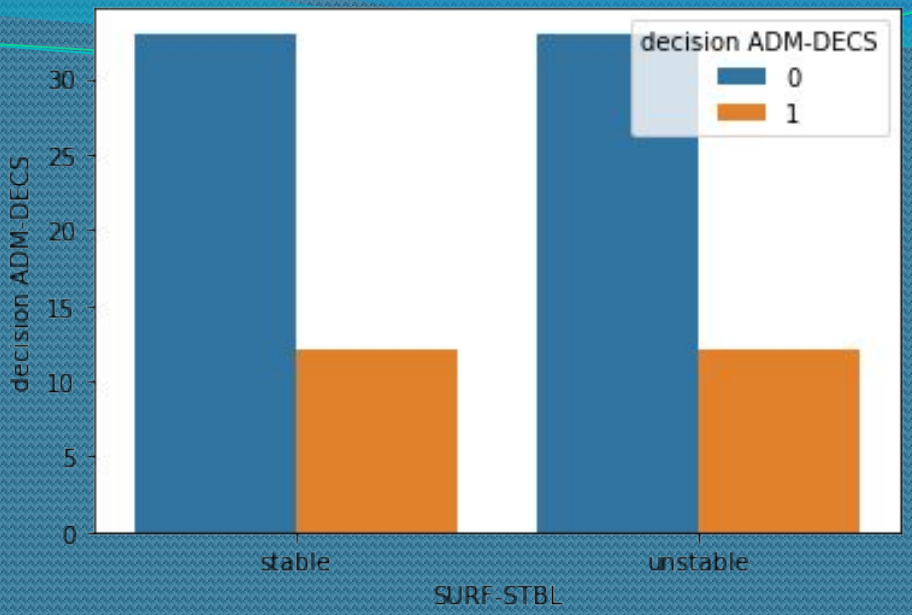
- The missing values in our data are replaced by mode.
- To analyze the data we created different plots.

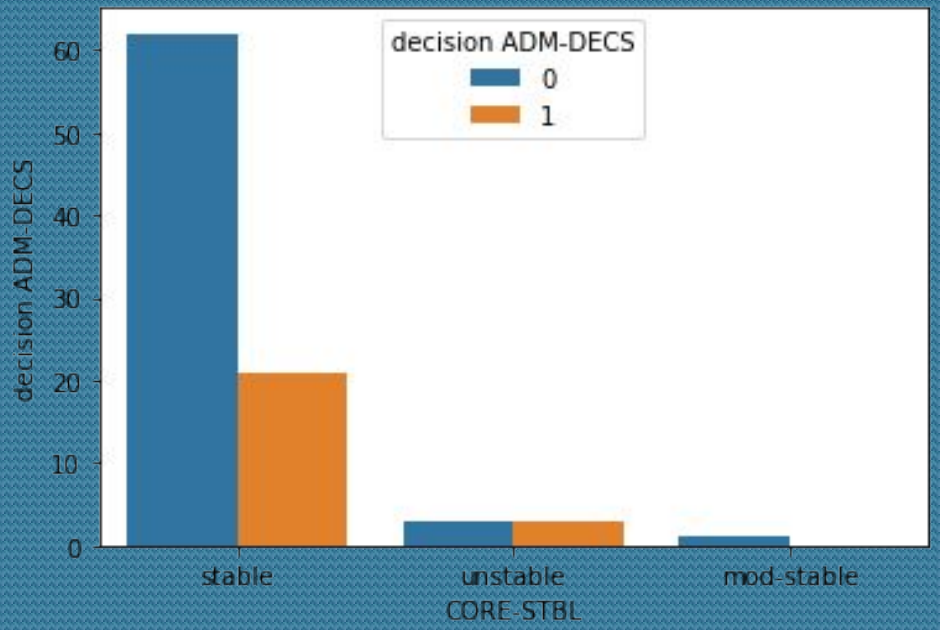


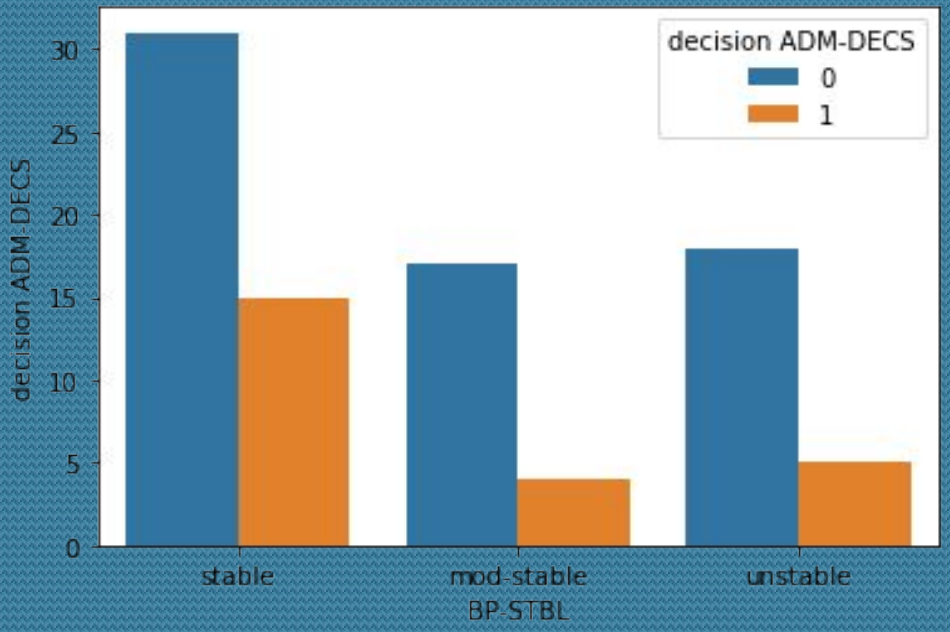




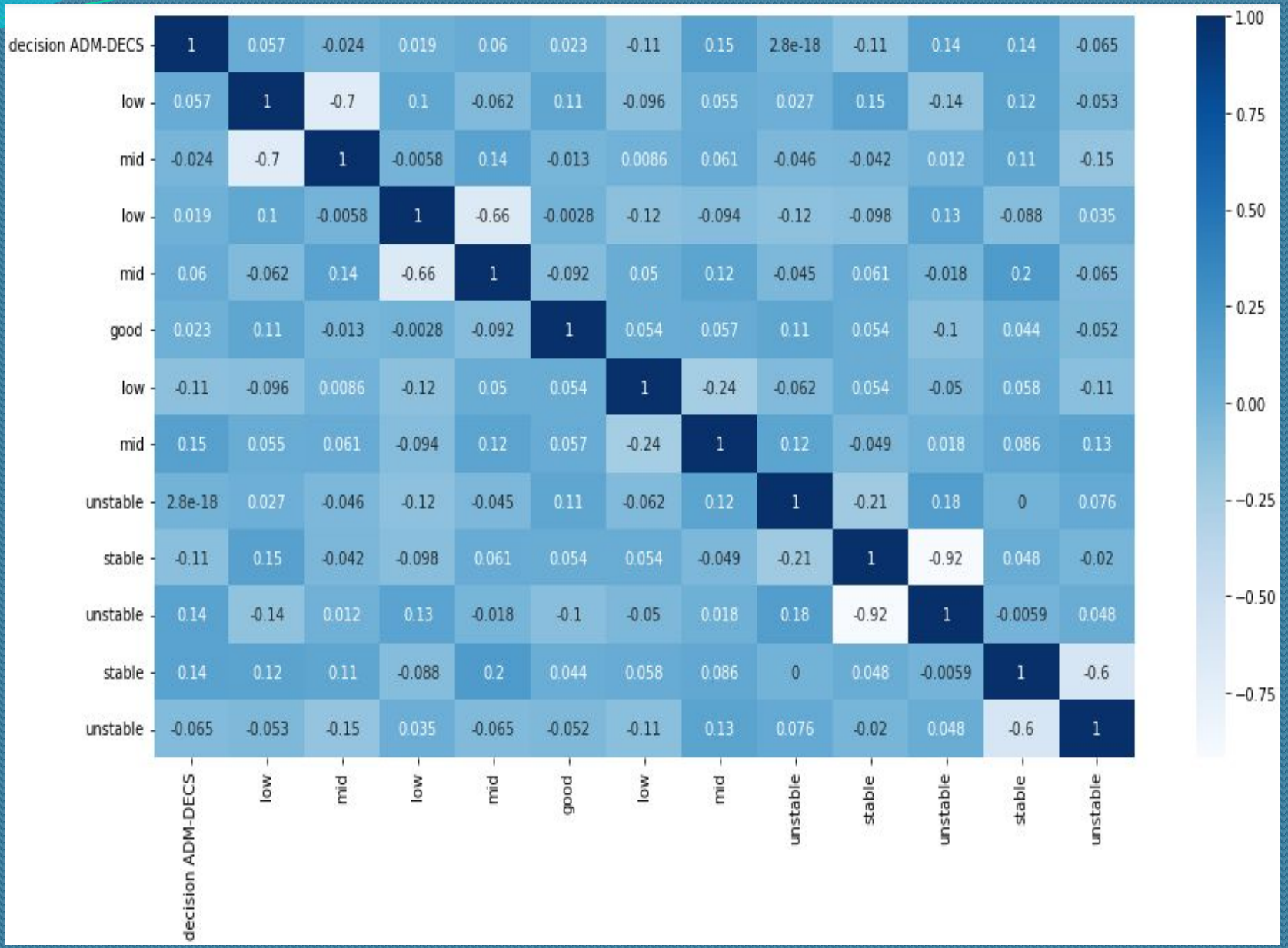

decision ADM-DECS
0
1
decision ADM-DECS
0
5
10
15
20
25
30
35
40
mid
high
low
L-BP


decision ADM-DECS
0
1
decision ADM-DECS
0
5
10
15
20
25
30
stable
unstable
SURF-STBL


decision ADM-DECS
0
1
decision ADM-DECS
0
10
20
30
40
50
60
stable
unstable
mod-stable
CORE-STBL


decision ADM-DECS
0
1
decision ADM-DECS
0
5
10
15
20
25
30
stable
mod-stable
unstable
BP-STBL

HEATMAP :



- Heat map is the graphical representation of correlation matrix representing correlation between different variables.
- Correlation ranges from -1 to 1. So, all the values in the heat map also are ranging from -1 to 1.
- If the value is positive then there is positive correlation between the attributes. If value is negative then there is negative correlation and if the value is 0 there is no correlation i.e. the 2 attributes are independent.

Algorithms:

- Depending on the response, the output variable is a categorical, so we use logistic regression.
- Logistic regression is a machine learning classification algorithm that is used to predict the probability of a categorical dependent variable.
- In this case, the dependent variable is a binary variable that contains data code as 0 & 1.
- We use train and test split method to get best accuracy value.

SPLIT SIZE BETWEEN TRAIN AND TEST	ACCURACY
TRAIN:50% , TEST:50%	75%
TRAIN:75% , TEST:25%	61%
TRAIN:67% , TEST:33%	70%
TRAIN:80% , TEST:20%	83%

Here we took 4 splits with different test and train percentages. As we can observe from the table above, we obtain best accuracy when split size between train and test is 80% and 20% respectively. Hence we obtain the best result at train:80% and test 20%.

Model statistics:

- **Accuracy:**

  It is the proportion of correct predictions over total predictions.

- **Code for finding accuracy:**

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,predictions)
```

- **Output: 83% accuracy**

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.88	0.93	0.90	15
1	0.50	0.33	0.40	3
ACCURACY			0.83	18
MACRO AVG	0.69	0.63	0.65	18
WEIGHTED AVG	0.81	0.83	0.82	18

Summary:

- Logistic regression is used to explain the relationship between target variable and one or more independent variables which may be numerical, categorical or ratio level.

- The classification task of the data is to determine where the patients in a post operative recovery state should be send to next, because hypothermia is significant concern after surgery.

RECOMMENDATIONS:

- Based on the promising results, we conclude with the following recommendations for future work on risk and warning scores for post operative patient deterioration detection as follows :
  - Continuous monitoring
  - Feature extraction
  - Physiological modelling
  - Personal health data, Lab data
  - Combine pre, intra & post-operative data
  - Temporal behavior

Appendix:

- **Codes for logistic regression :**

```
from sklearn.linear_model import LogisticRegression
predictions=logreg.predict(X_test)
from sklearn.metrics import classification_report
classification_report(y_test,predictions)
```

- **Confusion Matrix :** It describes the performance of a classification model on a set of test data for which the true values are known.

- **Code :**

```
from sklearn.metrics import confusion_matrix
z=confusion_matrix(y_test,predictions)
```

- **Output :**

```
array([[14, 1],
       [ 2, 1]])
```

References:

Github link:

- Alekhya Bulusu: https://github.com/alekhyabulusu/unp.project-.git
- Sushma Alimineti:
- Amulya Kuntala:
- Nagaphani Musunuri: https://github.com/phanimusunuri1234/UNP_Assignment/blob/main/unp.ipynb

THANK YOU!