A PROJECT REPORT ON,

STATISTICAL ANALYSIS OF "HEALTH STATUS OF UNIVERSITY TEACHERS. "

Submitted to,

DEPARTMENT OF STATISTICS,

SHIVAJI UNIVERSITY, KOLHAPUR.

For the partial fulfillment of degree

M.Sc.-II in APPLIED STATISTICS AND INFORMATICS 2022-2023

By,

MISS.SHINDE DHANASHRI DADASO

MISS. KAMBLE ANJALI VIJAYKUMAR

MISS. TINGARE SHITAL HARIDAS

Under the Guidance of,

Dr. S. D. Pawar Sir.

CERTIFICATE

This is to certify that this project report on,

STATISTICAL ANALYSIS OF

"HEALTH STATUS OF UNIVERSITY TEACHERS. "

Is submitted by,

Sr.No	Name of Students	PRN
1	MISS.SHINDE DHANASHRI DADASO	2021000550
2	MISS.KAMBLE ANJALI VIJAYKUMAR	2021000487
3	MISS.TINGARE SHITAL HARIDAS	2021000529

As a partial fulfilment for the award of M.Sc., in statistics, under my supervision and guidance as per rules and regulations of the Shivaji University, Kolhapur. During the academic year 2022-2023 and submitted the same. This work represents the bonafide work of these students. To the best of my knowledge, the matter represented here in the project has not been submitted earlier.

Dr. S. D. Pawar

Assistant Professor

Shivaji University, Kolhapur

Prof. Dr. S. B. Mahadik.

Head of Department of Statistics

Shivaji University, Kolhapur

Date:

Place:

ACKNOWLEDGEMENT

We have great pleasure while submitting this project report on**, "TO STUDY THE ANALYSIS ON HEALTH STATUS OF UNIVERSITY TEACHERS ."** in practical fulfillment of **M.Sc.-II.**

We thank **Prof. Dr. S. B. Mahadik** (HOD, Statistics Department) for giving permission for doing this project.

At the outset, we keep on giving deep gratitude towards our project guide **Dr. S.D. Pawar** gave us guidance right from the initial stage of the project and offered us several valuable suggestions for developing this project in a systematic and presentable manner. Also special thanks to **Shri. S. V. Rajguru** (Assistant Professor, Department of Statistics) for giving us valuable insights for analysis.

We would like to express our higher regards to our department of statistics, and also appreciate our respected teachers **Dr. S. D. Pawar, Shri. S. V. Rajguru, Shri. S. K Ganjave and Dr. S. M. Patil.** , Also we would like to thank all non-teaching staff for their help and co-operation. Ultimately, we express our respect to our friends and parents who inspire and support us in this work.

INDEX

INTRODUCTION

In today's fast-paced and demanding academic environment, the health and well-being of university teachers play a crucial role in their professional success and overall quality of life. Recognizing the importance of this issue, conducting a comprehensive analysis of the health status of university teachers is essential to understand their physical, mental, and emotional well-being. Such an analysis can provide valuable insights into the challenges they face and the support systems required to enhance their overall health.

Health check-ups include Height, Weight, Blood pressure, Hemoglobin, Sugar, Urine, Blood group, and past history. Hypertension is commonly seen in teachers, it leading cause of hypertension, heart disease, anemia, and kidney failure. Gaining weight causes so many health issues like high blood pressure, obesity, diabetes, hypertension, etc.

In conclusion, analyzing the health status of university teachers is crucial for understanding and addressing the well-being of these educators. By gaining insights into their physical, mental, and emotional health, educational institutions can implement targeted strategies to support their faculty members, promote a healthy work environment, and ultimately enhance the quality of education provided to students.

AIM

To study the overall Health Status of University Teachers.

OBJECTIVES

- To understand the Age at which Hypertension Occurs.
- To check correlation between height, weight and BMI.
- To check the association between various diseases or complications.
- To check accuracy of Hypertension, High Blood Sugar level, High blood pressure using various machine learning algorithms.

Data Collection Methodology

We have collected secondary data on 166 Teacher's health check-ups from the primary health center of Shivaji University Kolhapur.

The data were chosen from the years 2016 to 2022 excluding 2020, 2021. This data includes records of those teachers who have done Health checkups. That includes Age, Sex, Height, Weight, Blood Group, Blood Pressure, Hemoglobin, Sugar and Other tests.

Statistical tools used for analysis

Statistical Technique

➢ Graphical Presentation:

- Histogram
- Sub-Divided Bar Diagram
- Line Chart
- Box Plot
- Scatter Plot
- Word Cloud

➢ Tests:

- To check independence between Gender and Hypertension

$$\chi^2 = \Sigma \frac{(Oi - Ei)^2}{Ei^2}$$

- To check Proportion of High Blood Sugar Level in Male and Female

$$Z = \frac{(p1 - p2)}{\sqrt{\left(\frac{p1(1-p1)}{n1} + \frac{p2(1-p2)}{n2}\right)}}$$

➢ Association Rule

➢ Statistical Software & Tools:

- MS-Excel
- Ms-Word
- Python
- R Software

Overall summary

Year	Male					Female				
	Height		Weight			Height		Weight		
	Min	Max	Min	Max	Count	Min	Max	Min	Max	Count
2016	121	186	46	109	117	121	166	31	94	25
2017	151	186	49	109	96	121	164	34	80	20
2018	152	183	49	106	71	121	164	35	79	11
2019	151	183	51	105	71	121	164	34	82	15
2022	151	178	55	110	61	121	160	31	88	8

From the above table, we can see that

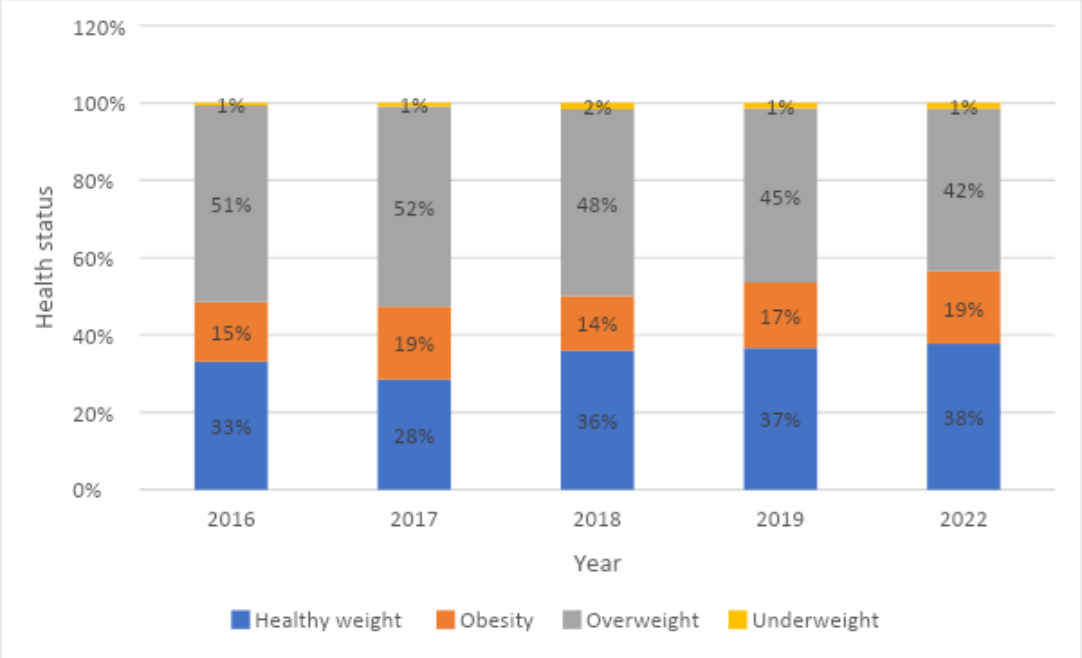
- The number of Male teachers is higher than Female teachers.
- There is data on 142 teachers available in 2016 out of 166 teachers, which is the highest among all years.

Year-wise records of no of Teachers:

Year	No of Patient
1	33
2	39
3	35
4	33
5	28

The 28 Teachers have 5 years of Record of health checkup,33 have 4 years of Record,35 have 3 years of Record,39 have 2 years of Record, and 33 have 1 year of Record.

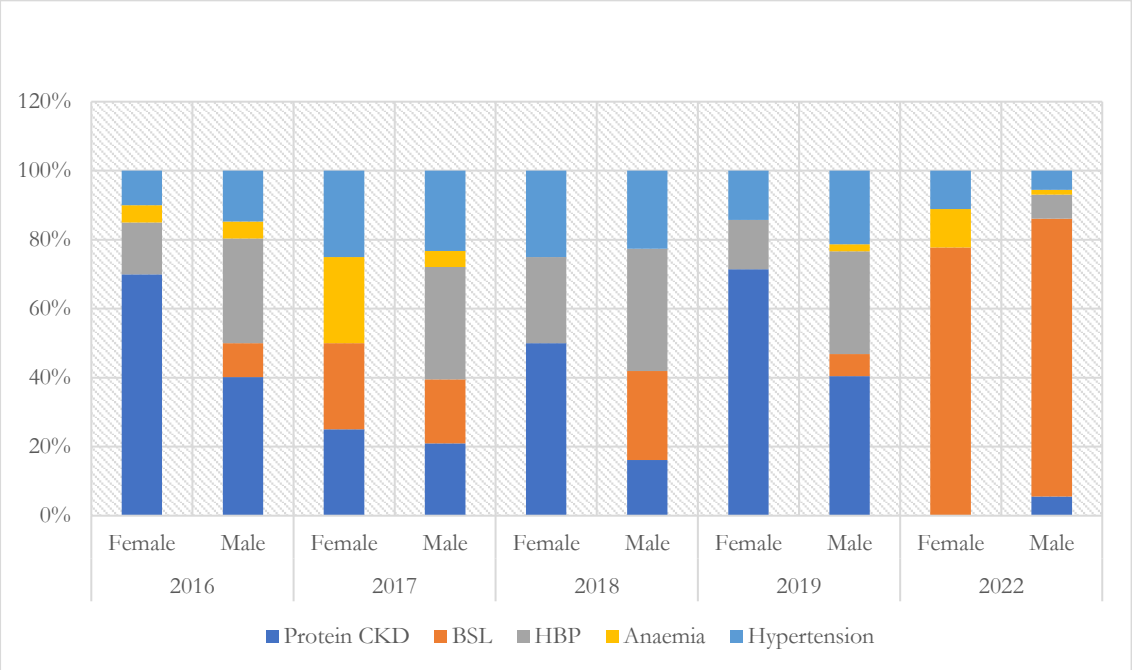
Graphical Representation



From the graph, we see that the count of overweight was high in all years. In 2016 51% of teachers were overweight. The count of obesity was the same in the year 2016 and 2017. In the year 2016, Obesity was 15% and healthy weight was 33%. In 2017 the percentage of overweight is 51% which is slightly higher than in 2016. We can see the count of healthy weight decreases from the year 2016 to 2017. In the year 2022 42% of teachers were overweight and 38% of teachers are in a healthy weight.

Gender-wise Complications

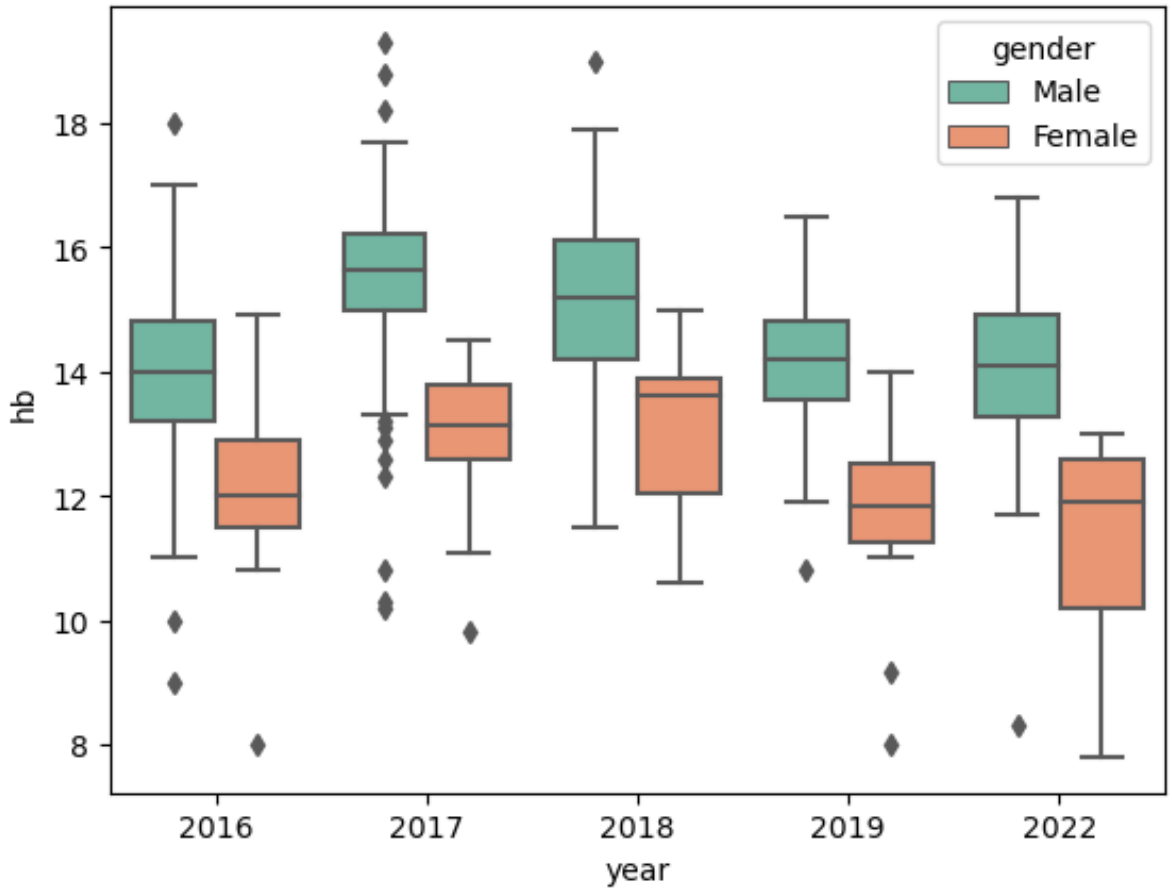
Year	2016		2017		2018		2019		2022	
Complications	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Protein CKD	70%	40%	25%	21%	50%	16%	71%	40%	0%	6%
BSL	0%	10%	25%	19%	0%	26%	0%	6%	78%	81%
HBP	15%	30%	0%	33%	25%	35%	14%	30%	0%	7%
Anemia	5%	5%	25%	5%	0%	0%	0%	2%	11%	1%
Hypertension	10%	15%	25%	23%	25%	23%	14%	21%	11%	6%



Conclusion:

As we saw in the year 2022 the percentage of Blood sugar levels was almost the same in females about 80% as compared to males. In 2019 the protein in urine was high in females about 70% as compared to males in the same year which indicates symptoms of chronic kidney disease.

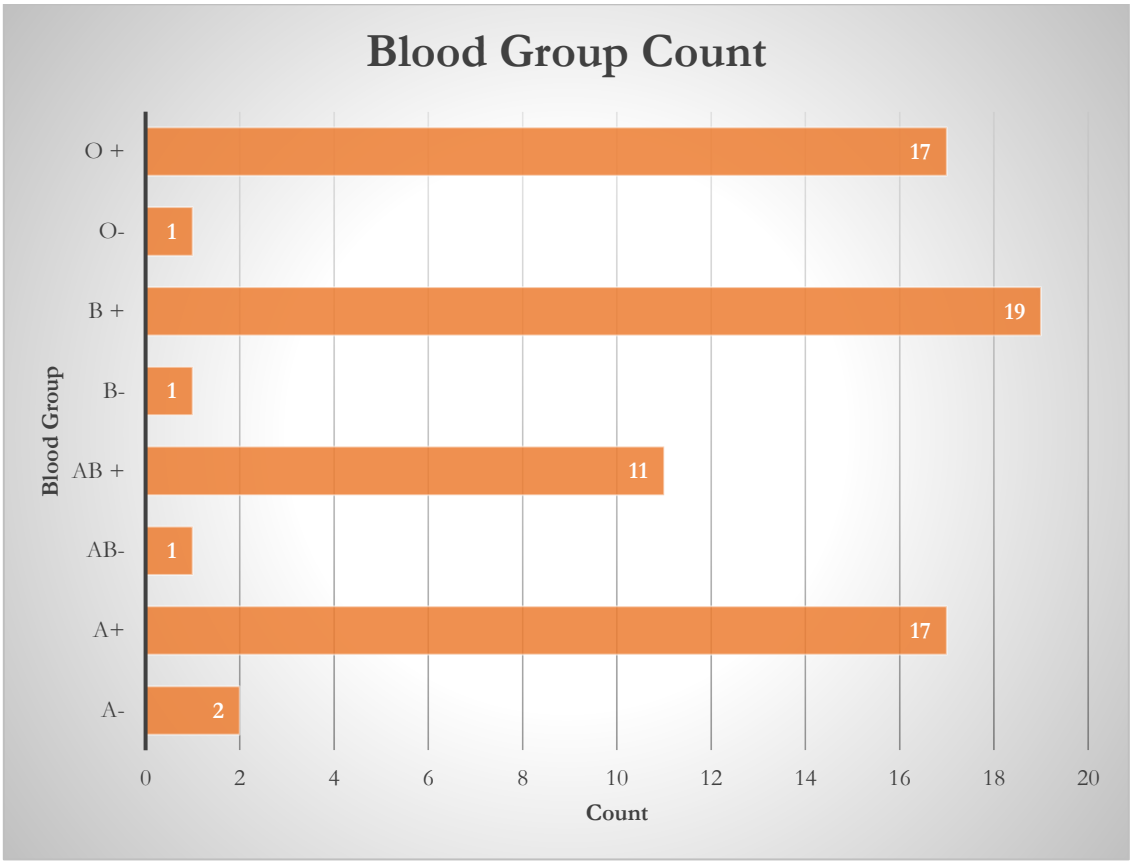
Boxplot of Hemoglobin



Conclusion:

From the above boxplot, we see that male has higher HB than female teachers. The HB of the male was higher in the year 2017 as compared to other years and outliers were present. In the year 2022, the female HB was lower as compared to other years mean was 12.

Barplot of Teachers Blood Group

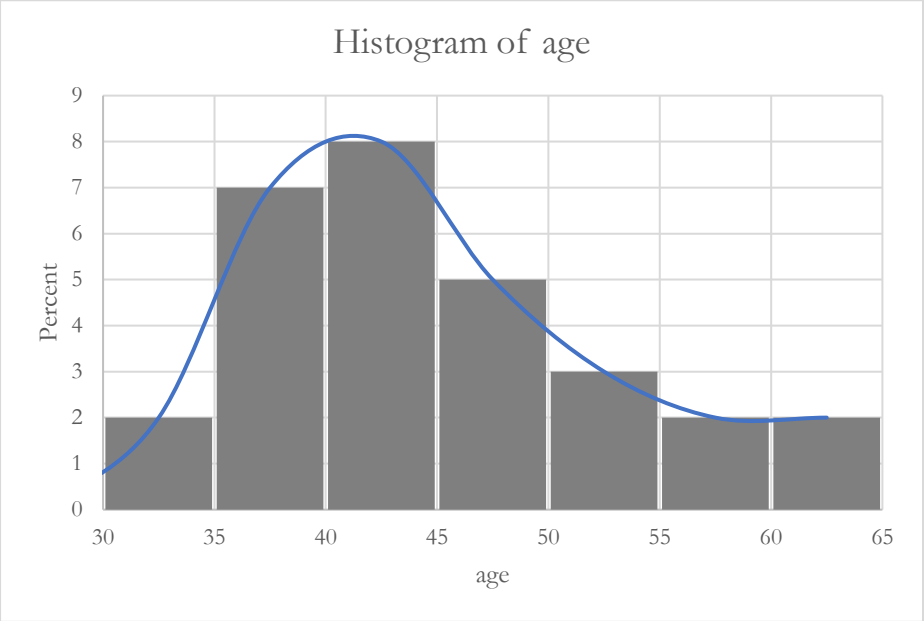


Conclusion:

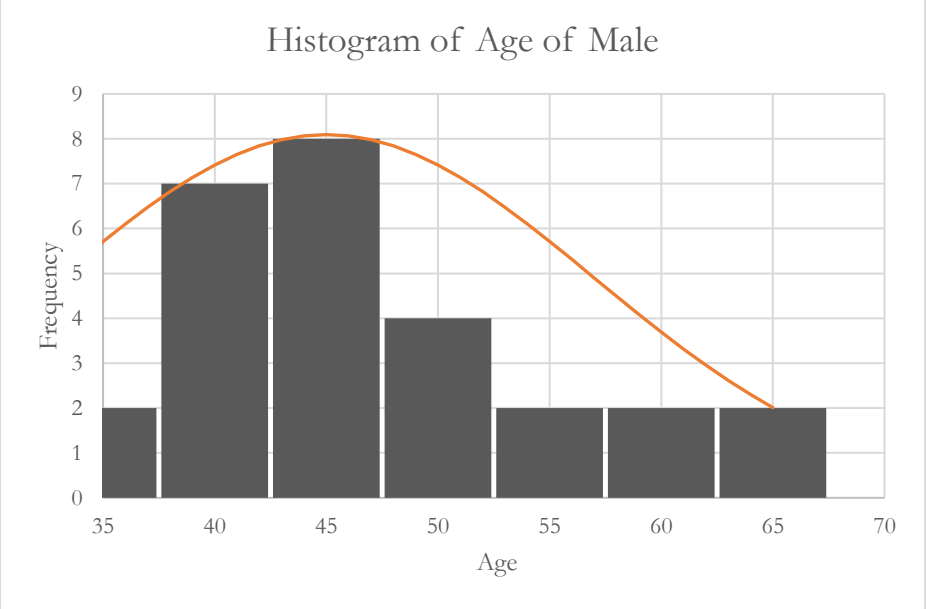
- Above plot shows the count of teachers blood group in year 2022. Here the count of A+ve blood group is maximum which is 17. And the count of O-ve,
- B-ve and AB-ve is 1 which is minimum.

Histogram of Age

The Age at which Hypertension occurs



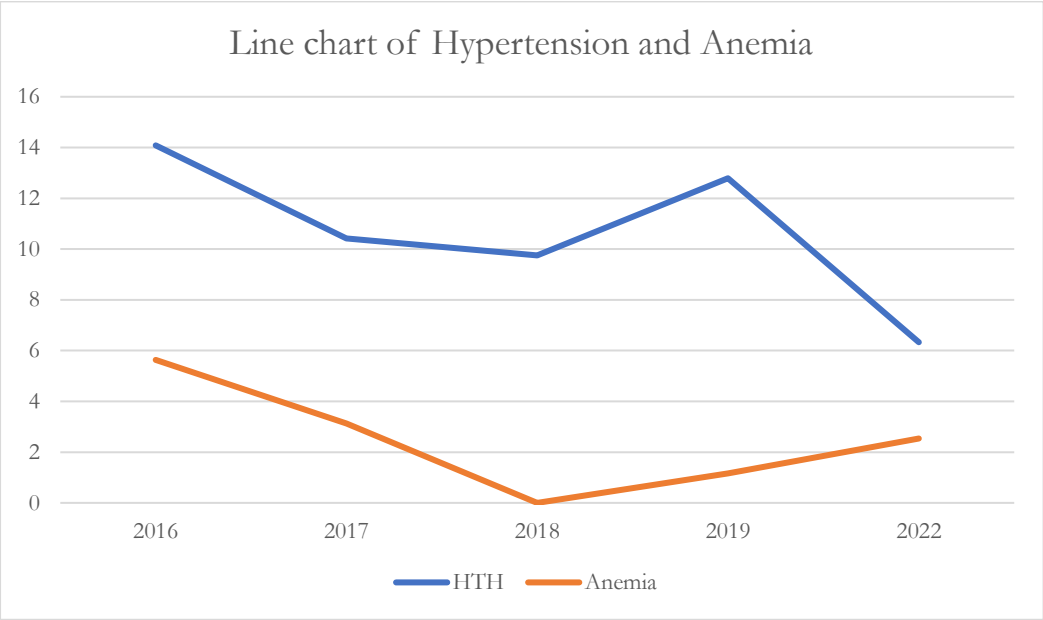
Conclusion: From the above chart we see that the age of Hypertension occurs is 43.



Conclusion: From the above chart we see that the age of hypertension for males is 45.

Comparison Of Hypertension And Anemia:

	2016	2017	2018	2019	2022
HTH(%)	14.08451	10.41667	9.756098	12.7907	6.329114
Anemia (%)	5.633803	3.125	0	1.162791	2.531646



Conclusion:

In the year 2022 out of 100, 6.3291 have an occurrence of Hypertension, and 2.5316 have an occurrence of Anemia.

Statistical Analysis

BMI analysis

- **1. Correlation analysis**

		Height(m)	Weight(kg)	BMI
Height(m)	**Pearson Correlation**	1	0.4963	-0.1741
	N	166	166	166
Weight(kg)	**Pearson Correlation**	0.4963	1	0.7609
	N	166	166	166
BMI	**Pearson Correlation**	-0.1741	0.7609	1
	N	166	166	166

In the above table, N indicates the number of teachers in the datasets. The correlation between Height and Weight is 0.4963. The significant value of correlation is 0.000. Every variable is to correlate with 1. The half diagonal is identical. The following is a Pearson correlation between the Height and Weight is equivalent to the correlation between Weight and Height. The above correlation is Positive. The correlation between Height and BMI is -0.1741 which is a negative correlation. The Pearson correlation between Weight and BMI is 0.7609 which is positive.

To check the association between different diseases(complications):

Antecedents	Consequents	support	confidence
{HBP, BSL, Protein}	{HTH}	0.6	0.5
{HTH, BSL, Protein}	{HBP}	0.6	0.6
{HTH, HBP, BSL}	{Protein}	0.6	0.428571429
{HTH, BSL)	{HBP, Protein}	0.6	0.230769231

50% of those who have HBP, BSL, and Protein also have HTH

60% of those who have HTH, BSL, and Protein also have HBP

42% of those who have HTH, HBP, and BSL also have protein in their urine

23% of those who have both HTH and BSL also have HBP and have protein in their urine.

Antecedent	Consequent	Support	conf	lift
{HTH, HBP}	{BSL}	0.054795	0.26667	1.497436
{BSL}	{HTH, HBP}	0.054795	0.307692	1.497436
{HTH, BSL}	{HBP}	0.054795	1	1.738095
{HBP, BSL}	{HTH}	0.054795	0.2	1.659091

26% who have HTH, and HBP also have sugar

30% who have sugar also have HTH, HBP

100% who have HTH and sugar also have HBP

20% who have HBP and sugar also have HTH

Antecedent	Consequent	Support	conf	lift
{HTH, BSL}	{Protein}	0.054795	0.266667	1.073529
{HTH, Protein}	{BSL}	0.054795	0.4	2.246154
{BSL, protein}	{HTH}	0.054795	0.5	1.659091

26% who have HTH and sugar also have protein in their urine

40% who have HTH and protein in urine also have sugar in their blood

50% who have sugar in their blood and protein in their urine have HTH

HTH-Hypertension, BSL-Blood sugar level, Protein-Protein in urine (Chronic Kidney Disease), HBP-High blood pressure.

Analysis of independence

ii) To test the independence of Hypertension and Gender.

Hypothesis:

H_0: Hypertension and Gender are independent. vs

H_1: Hypertension and Gender are dependent.

Test statistic:

$$\chi^2 = \Sigma \frac{(Oi - Ei)^2}{Ei^2}$$

where Oi represents the observed frequency in each category, Ei represents the expected frequency in each category

Test Procedure:

Gender/Hypertension	Yes	No
Male	11	62
Female	0	14

p-value=0.120184

Decision: Here p-value>0.05 so fails to reject the H_0.

Inference: Hypertension and gender are significantly independent.

Analysis of Proportion

Hypothesis:

H0: Proportion of high blood sugar level is equal in male and female.

H1: Proportion of high blood sugar level is not equal in male and female.

Test statistics:

$$Z=\frac{(p1-p2)}{\sqrt{(\frac{p1(1-p1)}{n1}+\frac{p2(1-p2)}{n2})}}$$

Procedure:

	High	Low	Total
Female	16	7	23
Male	70	61	131

Test statistic: -1.4367902818934637
p-value: 0.15077759256519374

Conclusion:

Here Z value is -1.4367 and p-value is 0.150778 which is greater than 0.05 so we accept H0 that the proportion of high blood sugar level is same in both male and female.

Logistic Regression:

By fitting the logistic model to the data we get,

y- The Hypertension is present or absent.

0=Absent,1=Present,0=Female,1=Male

Model:

y= -6.03197377+0.20860223*age+0.07183546*gender

Accuracy is the proportion of correct prediction over total prediction.
The accuracy with the logistic regression model is 92.85%

Here age has a significant impact on hypertension. Hypertension increases with an increase in age.

Hypertension in males is exp(0.071835) times than female.

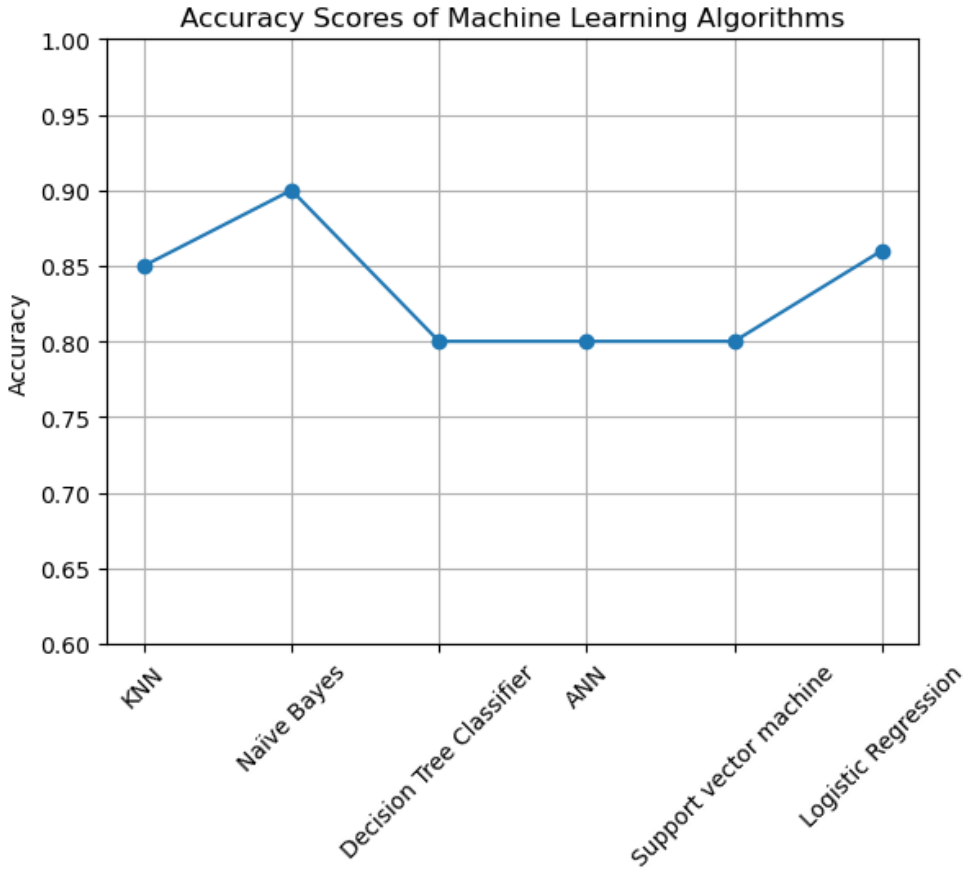
Hypertension Prediction using various machine learning algorithm:

Model:

$$y = \beta_1 * Gender + \beta_2 * Age + \beta_3 * Weight$$

y=1, hypertension present, y=0 hypertension absent

Model	Accuracy
KNN	85.71%
Naïve Bayes	90.48%
Decision Tree Classifier	80.95%
ANN	80.95%
Support vector machine	80.95%
Logistic Regression	86.75%



Conclusion: Almost all the classifier gives better prediction in this scenario. Naïve bayes gives better prediction.

Health Status prediction by using various machine learning algorithm:

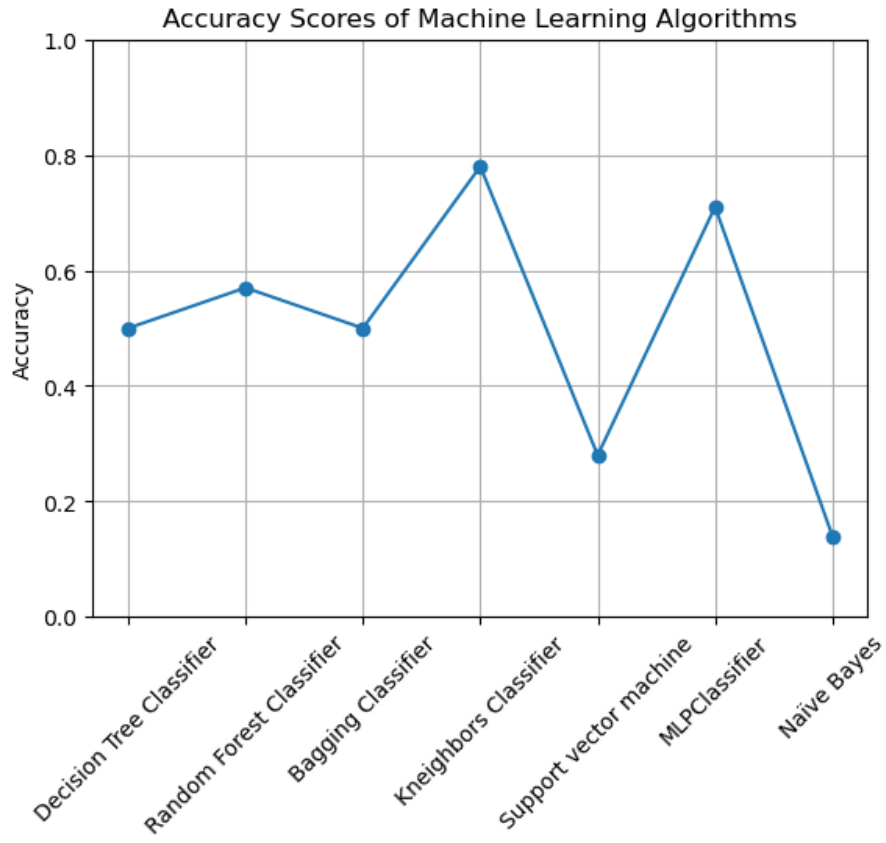
Model:

$$y = \beta_1 * Gender + \beta_2 * Height + \beta_3 * Weight$$

female=0, male=1,

y, 0-underweight, 1-healthy weight, 2-overweight, 3-obesity

Decision Tree Classifier	50.00%
Random Forest Classifier	57.14%
Bagging Classifier	50.00%
KNN	78.57%
SVC	28.57%
MLPClassifier	71.43%
Naïve Bayes	14.29%



Conclusion: KNN and MLPClassifier give better prediction in this scenario.

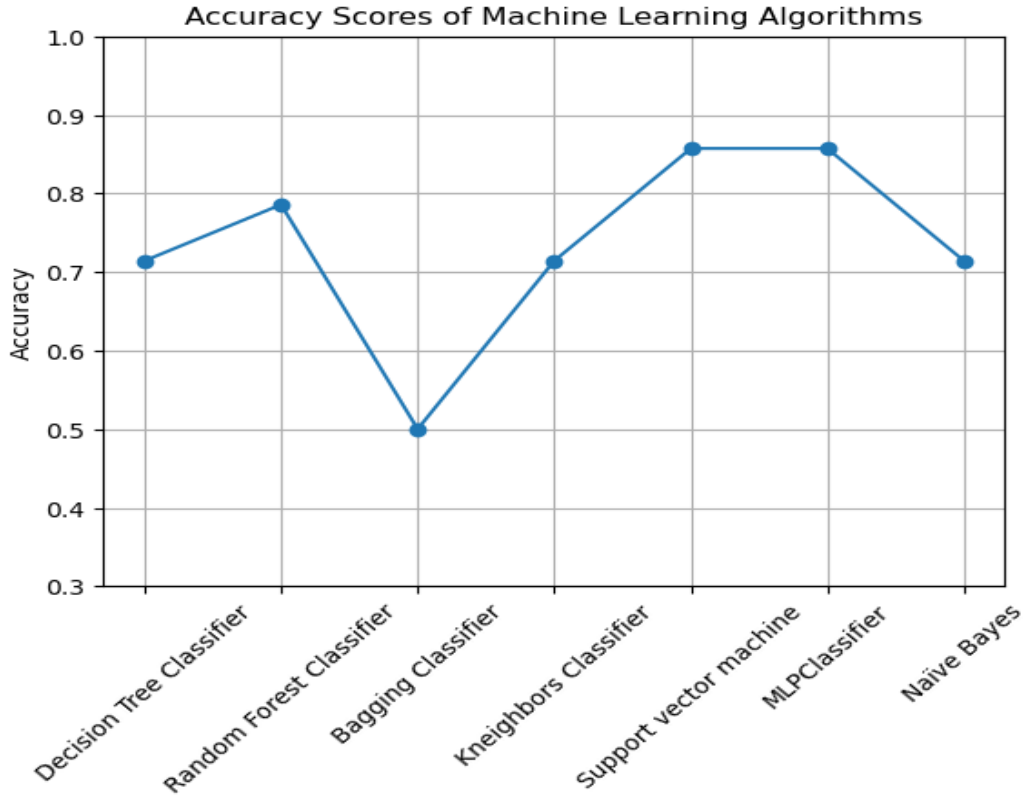
High Blood Sugar Level prediction by using various machine learning algorithm:

Model:

$$y = \beta_1 * Age + \beta_2 * Bp + \beta_3 * BMI$$

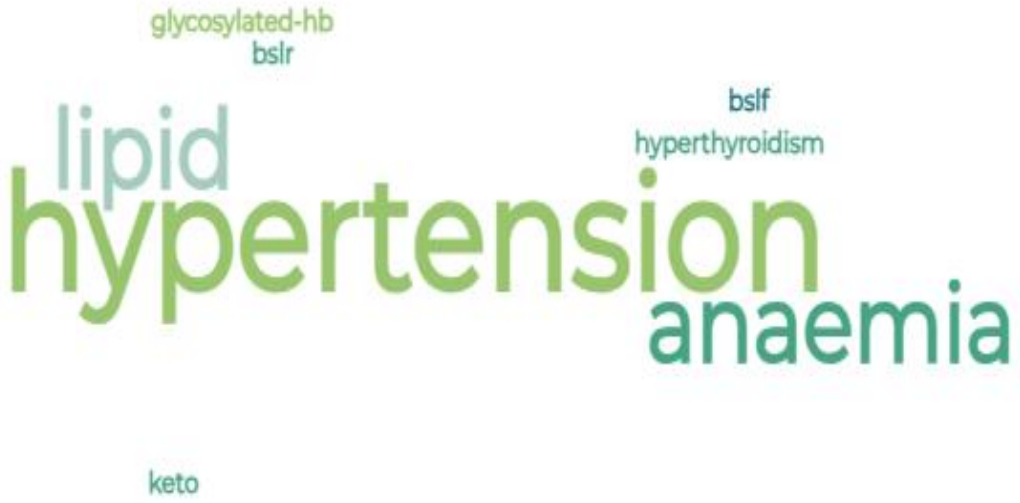
y :1-high, 0-normal

Model	Accuracy
DecisionTreeClassifier	71.43%
RandomForestClassifier	78.57%
BaggingClassifier	50.00%
KNeighborsClassifier	71.43%
SVC	85.71%
MLPClassifier	85.71%
GaussianNB	71.43%



Conclusion: Almost all the classifier gives better prediction in this scenario. Support Vector Machine and MLP Classifier gives better prediction

Wordcloud of complications seen among teachers:



Conclusion:

The above word cloud conclude that Hypertension occurs most probably among the university teachers.

The frequency of Anemia is next higher to the Hypertension.

CONCLUSION

- The count of overweight was high in all years. In 2016 51% of teachers were overweight. In 2022 there were about 19% teachers are obese.
- As we saw in the year 2022 the percentage of Blood sugar levels was almost the same in females about 80% as compared to males.
- The age of Hypertension occurs is 43.
- There is positive correlation between weight and BMI.
- Hypertension and gender are significantly independent.
- The proportion of high blood sugar level is same in both male and female.
- Hypertension occurs most probably among the university teachers.

LIMITATIONS

- This study is conducted only for Teachers who are working in Shivaji University Campus.
- This study is done for only 5 years including the years from 2016 to 2019 and 2022.
- This health camp are conducted once in a year so for more than one camp the results can be changed.
- The study area is restricted for Campus Teachers of Shivaji University, the sample size was moderate.

References:

- Analysis of Body Mass Index Based on Correlation and Regression By Dr. S. Kavitha, T. Sabhanayagham, R. Thenmozhi.

- Python for Data Analytics by Wes McKinney.

Appendix

Boxplot of Haemoglobin:

```
import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

import seaborn as sns

s=pd.read_excel("Book1.xlsx")

fx=sns.boxplot(x='year',y='hb',data=s,hue='gender',palette='Set2')
```

Machine Learning Accuracy plots:

```
algorithm_names = ['KNN','Naïve Bayes','Decision Tree
Classifier','ANN','Support vector machine','Logistic Regression']

accuracy_scores = [0.85, 0.90, 0.80,0.80,0.80,0.86]

# Plotting the accuracy scores

# Plotting the accuracy scores

x_pos = np.arange(len(algorithm_names))

plt.plot(x_pos, accuracy_scores, marker='o')

plt.xticks(x_pos, algorithm_names,rotation=45)

plt.ylabel('Accuracy')

plt.title('Accuracy Scores of Machine Learning Algorithms')

plt.ylim(0.6, 1.0)  # Set the y-axis limits between 0 and 1

plt.grid(True)  # Add grid lines

plt.show()
```

THANK YOU