Mid-Program Presentation Present

Presenter's Name: Eleni Yanyan and Alex

04/02/2024

Analysis and Detection of PCOS

Objective: Develop a predictive model for early detection of PCOS and infertility.

Importance: Early diagnosis crucial for effective management.

Dataset: Comprehensive physical and clinical parameters related to PCOS.

Problem Statement

The goal is to develop a predictive model for the early detection of PCOS and infertility-related issues. PCOS is a complex hormonal disorder, and early diagnosis is crucial for effective management.

Tasks

Structure and Features: Understand dataset dimensions and variables.

Missing Values: Identify and handle missing data.

Outliers: Detect and address outliers.

Patterns: Explore patterns in the data.

Data Preprocessing

Encode categorical variables, normalize data.

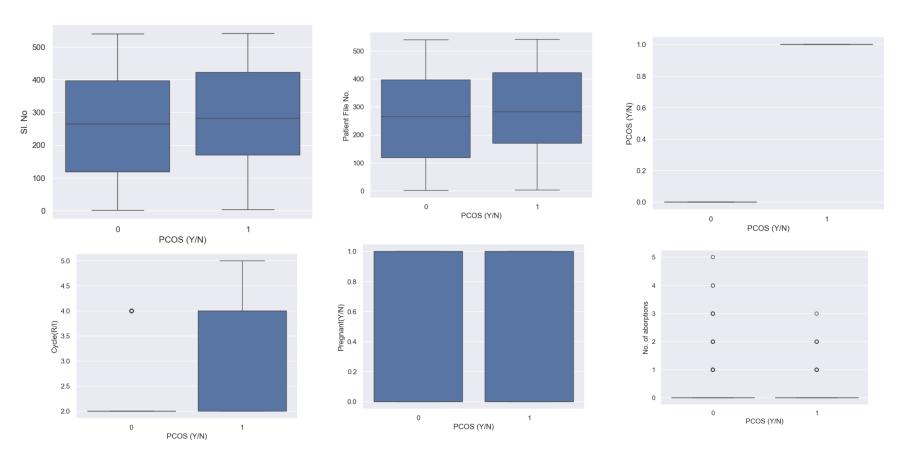
Data

	ss 'pandas.core.frame.Da eIndex: 541 entries, 0 t			21 22	<pre>Hip(inch) Waist(inch)</pre>	541 non-null 541 non-null	int64 int64
	columns (total 44 column	,		23	Waist:Hip Ratio	9 non-null	float64
#	Column	Non-Null Count	Dtype	24	TSH (mIU/L)	541 non-null	float64
				25	AMH(ng/mL)	541 non-null	object
0	Sl. No	541 non-null	int64	26	PRL(ng/mL)	541 non-null	float64
1	Patient File No.	541 non-null	int64	27	Vit D3 (ng/mL)	541 non-null	float64
2	PCOS (Y/N)	541 non-null	int64	28	PRG(ng/mL)	541 non-null	float64
3	Age (yrs)	541 non-null	int64	29	RBS(mg/dl)	541 non-null	float64
4	Weight (Kg)	541 non-null	float64	30	Weight gain(Y/N)	541 non-null	int64
5	Height(Cm)	541 non-null	float64	31	hair growth(Y/N)	541 non-null	int64
6	BMI	241 non-null	float64	32	Skin darkening (Y/N)	541 non-null	int64
7	Blood Group	541 non-null	int64	33	Hair loss(Y/N)	541 non-null	int64
8	Pulse rate(bpm)	541 non-null	int64	34	Pimples(Y/N)	541 non-null	int64
9	RR (breaths/min)	541 non-null	int64	35	Fast food (Y/N)	540 non-null	float64
10	Hb(g/dl)	541 non-null	float64	36	Reg.Exercise(Y/N)	541 non-null	int64
11	Cycle(R/I)	541 non-null	int64	37	BP Systolic (mmHg)	541 non-null	int64
12	Cycle length(days)	541 non-null	int64	38	BP Diastolic (mmHg)	541 non-null	int64
13	Marraige Status (Yrs)	540 non-null	float64	39	Follicle No. (L)	541 non-null	int64
14	Pregnant(Y/N)	541 non-null	int64	40	Follicle No. (R)	541 non-null	int64
15	No. of aborptions	541 non-null	int64	41	Avg. F size (L) (mm)	541 non-null	float64
16	<pre>I beta-HCG(mIU/mL)</pre>	541 non-null	float64	42	Avg. F size (R) (mm)	541 non-null	float64
17	<pre>II beta-HCG(mIU/mL)</pre>	541 non-null	object	43	Endometrium (mm)	541 non-null	float64
18	FSH(mIU/mL)	541 non-null	float64		es: float64(19), int64(2		110004
19	LH(mIU/mL)	541 non-null	float64		ry usage: 186.1+ KB	.5,, 05,600(2)	
20	FSH/LH	9 non-null	float64	memo	Ly abage. 100:11 Kb		

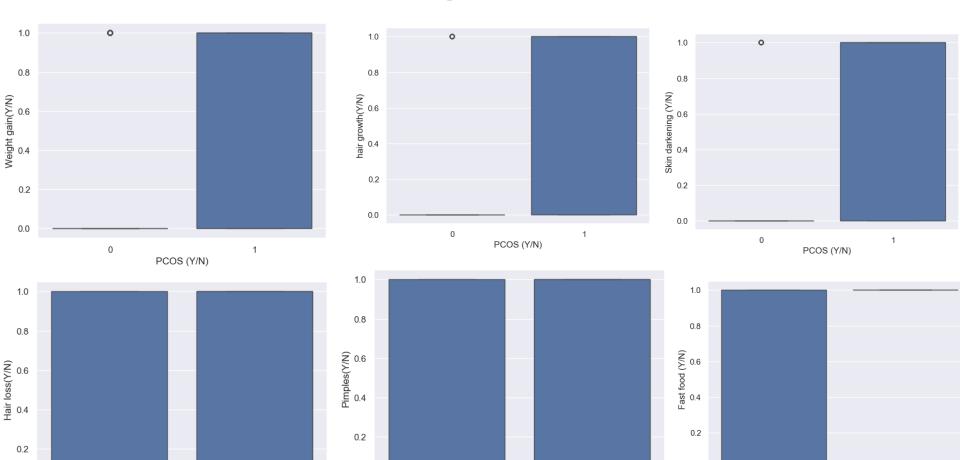
Missing Values

Sl. No	0	Hip(inch)	0
	•	Waist(inch)	0
Patient File No.	0	Waist:Hip Ratio	532
PCOS (Y/N)	0	TSH (mIU/L)	0
Age (yrs)	0	AMH(ng/mL)	1
Weight (Kg)	0	PRL(ng/mL)	0
Height(Cm)	0	Vit D3 (ng/mL)	0
BMI	300	PRG(ng/mL)	0
Blood Group	0	RBS(mg/dl)	0
Pulse rate(bpm)	0	Weight gain(Y/N)	0
RR (breaths/min)	0	hair growth(Y/N)	0
Hb(g/dl)	0	Skin darkening (Y/N)	0
, <u> </u>	-	Hair loss(Y/N)	0
Cycle(R/I)	0	Pimples(Y/N)	0
Cycle length(days)	0	Fast food (Y/N)	1
Marraige Status (Yrs)	1	Reg.Exercise(Y/N)	0
Pregnant(Y/N)	0	BP _Systolic (mmHg)	0
No. of aborptions	0	BP _Diastolic (mmHg)	0
<pre>I beta-HCG(mIU/mL)</pre>	0	Follicle No. (L)	0
II beta-HCG(mIU/mL)	1	Follicle No. (R)	0
FSH(mIU/mL)	0	Avg. F size (L) (mm)	0
	0	Avg. F size (R) (mm)	0
LH(mIU/mL)	- 0	Endometrium (mm)	0
FSH/LH	532	dtype: int64	

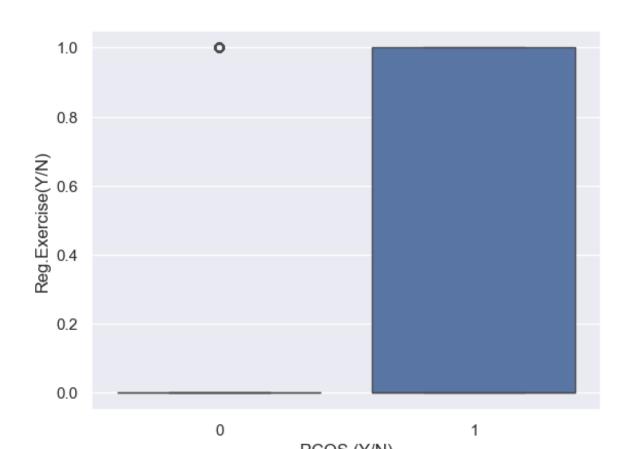
Descriptive Table



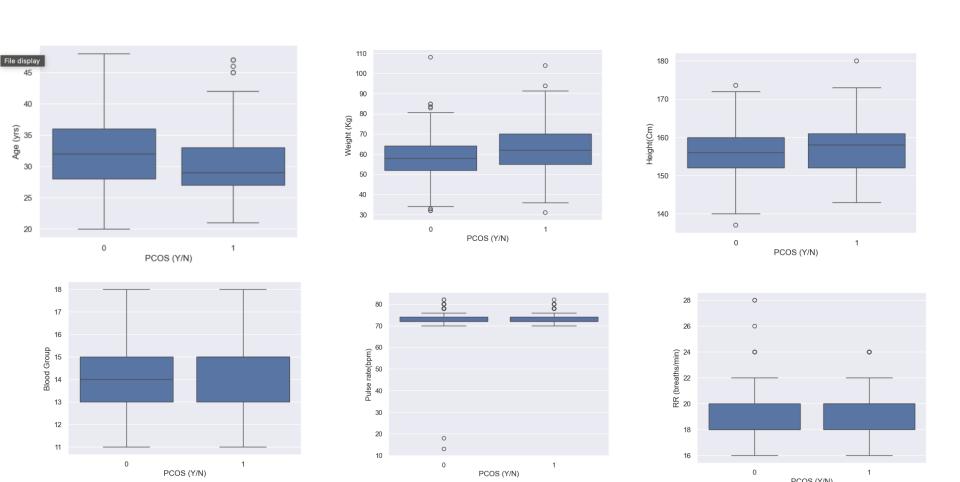
Descriptive Table

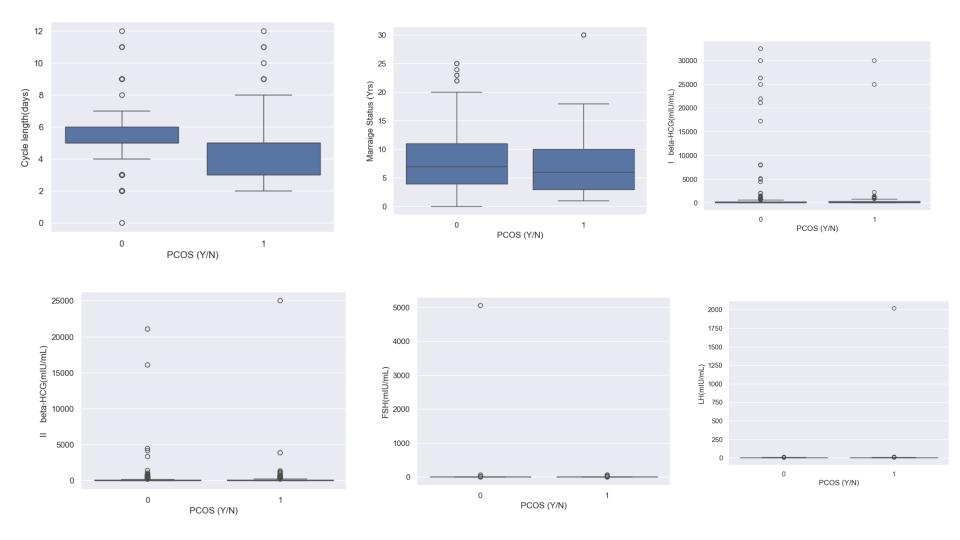


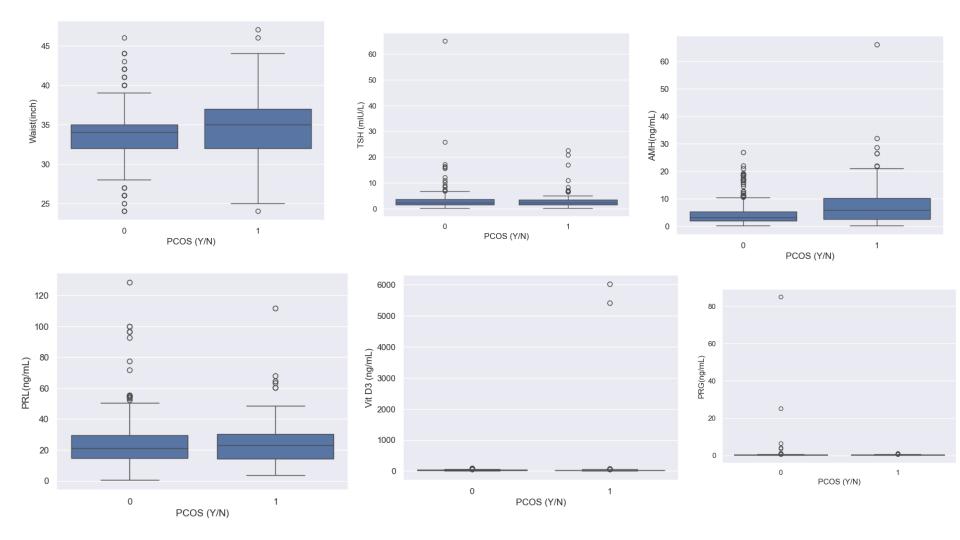
Descriptive Table

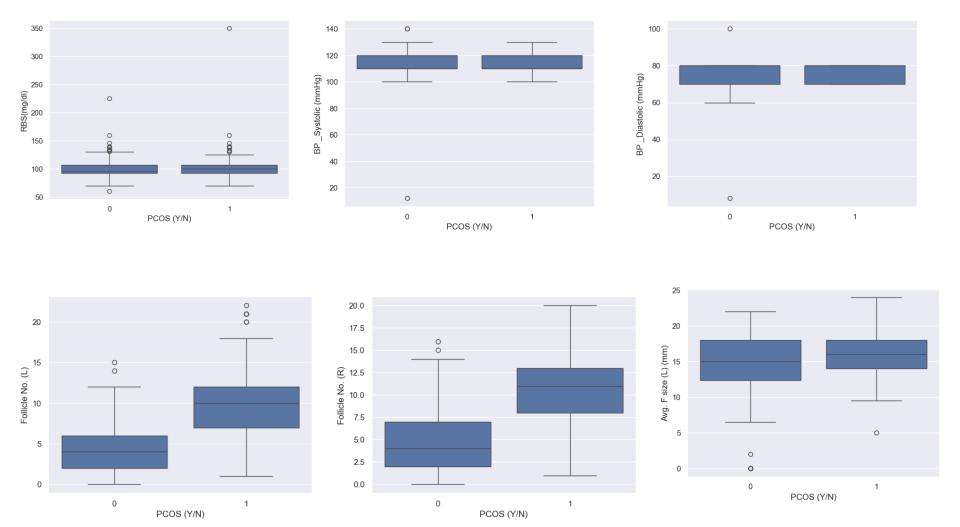


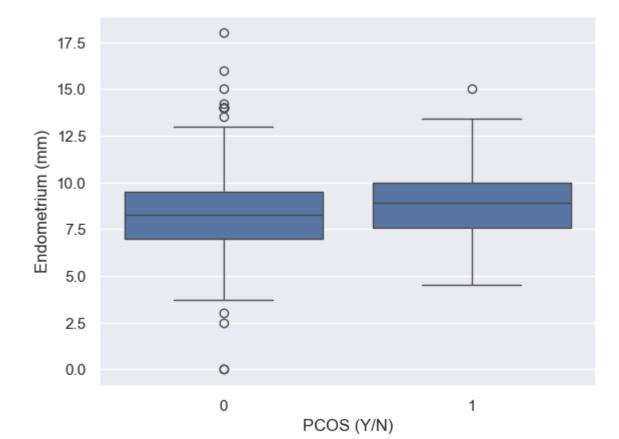
Exploratory Data Analysis











Correlation Matrix of all the features

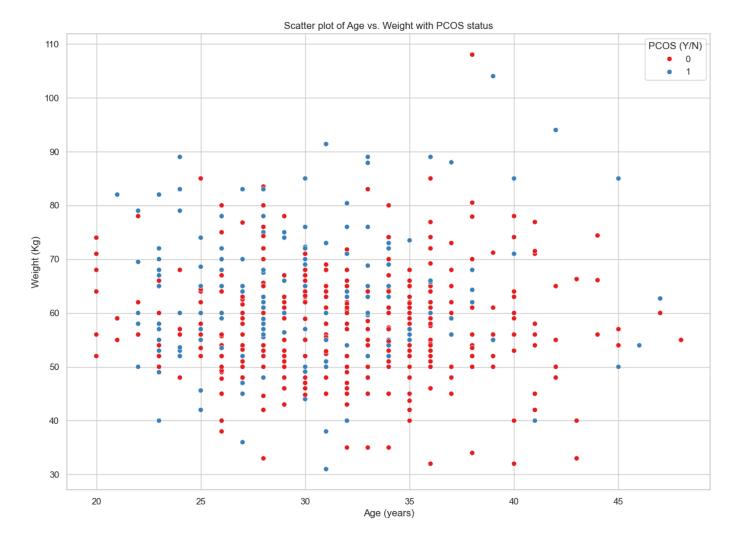
																	(Correla	ation M	Matrix o	of Featu	ıres															
SI. No	1.00	1.00	0.06	0 07	0.08	0.01 0	000	0.04 -	0.02 -0	09 -0	02 -0 1	2 -0.02	0.04	-0.11	-0.02	-0.04	0.02	0.05	0.33	0.05	0 07 0	10 00	0 80	04 -0 0	0.02	0.01	-0.05 (03 0	02 0 1	6 0 1	0 00	0.01	0.01	0.05	0.09 0	29 0	24 0.03
Patient File No.																																					24 0.03
PCOS (Y/N)																																					10 0.11
Age (yrs)																																					08 -0.10
Weight (Kg)																																					07 -0.01
Height(Cm)																								03 0.05													06 -0.05
																																					02 -0.02
Blood Group																																					
Pulse rate(bpm)																																					03 -0.04
RR (breaths/min)																																					02 -0.06
Hb(g/dl)																																					02 -0.07
Cycle(R/I)																																					02 0.04
Cycle length(days)																																					01 -0.02
Marraige Status (Yrs)																																					10 -0.11
Pregnant(Y/N)	0.04	0.04		-0.04 - 0.04	0.05	0.05	0.07	0.08	0.08		0.0	5 -0.01	1.00	0.00	0.15	0.17	0.05		0.03 -	0.05 -	0.06 -0	0.0	01 -0.	0.04	-0.01	-0.02	0.07	0.01 0	.03 0.1	1 0.0	80.0	-0.03	0.02	0.05	0.03	0.11 -0	12 -0.01
No. of aborptions	-0.11		0.06	0.22	0.09		0.05 (0.05	0.01 0.	06 0	06 0.0	0.25	0.00	1.00	0.06	0.05			0.08	0.07	0.03		06 0.0	02 -0.0	-0.01	0.12	0.04 -	0.01 -0	0.01 0.0			-0.08	0.07		0.08 -6		12 -0.07
I beta-HCG(mIU/mL)	-0.02 -	0.02	0.03	0.01 0	0.02	0.06		0.02 -	0.09 -0	02 0.0	06 0.0	2 0.11	0.15	0.06	1.00	0.53	0.00 -	0.01	0.02	0.01	0.05 0	.01 -0.	01 -0.	01 -0.0	0 -0 02	-0.06	0.02 (0.05 0	.04 -0.0	01 0.0	3 0.04		0.00	0.05	0.02 0	0.05 0.	07 -0.05
II beta-HCG(mIU/mL)	-0.04 -		0.01	0.04 -	0.00	0.04 4	0.01 -	0.02		09 0.	03 0.0	2 0.11	0.17	0.05	0.53	1.00	0.02 -	0.01	0.02	0.02	0.04 0	.00 -0.	00 -0.	01 0.01	-0.00	0.00	0.02 (0.03 0	.04 -0				0.00	0.06	0.04 4	0.00 0.	04 0.02
FSH(mIU/mL)																																					02 -0.05
LH(mIU/mL)																																					03 0.01
Hip(inch)																																					10 0.03
Waist(inch)																																					02 0.01
TSH (mIU/L)																																					09 0.01
AMH(ng/mL)																																					10 0.10
PRL(ng/mL)																																					07 0.03
																																					01 -0.03
Vit D3 (ng/mL)																																					
PRG(ng/mL)																																					04 -0.05
RBS(mg/dl)																																					01 -0.02
Weight gain(Y/N)	0.01																																				04 0.04
hair growth(Y/N)									0.01 0.																												00 0.06
Skin darkening (Y/N)									0.00 0.																												02 0.10
Hair loss(Y/N)																								0.0													04 0.16
Pimples(Y/N)																																					02 0.03
Fast food (Y/N)																																					04 0.14
Reg.Exercise(Y/N)	0.00	0.00	0.07		0.07	0.01 0	0.06	0.04 -	0.05 0.	01 -0.	0.0	3 0.06	-0.08	-0.04	0.04		0.02	0.08	0.03 -	0.07	0.00 0	.12 0.0	0.	0.00	-0.03		0.01	0.11 -0	.01 0.0	9 0.0	1.00	0.03		0.01	0.05 0	0.03 0.	0.08
BP _Systolic (mmHg)	0.01	0.01	0.01	0.07	0.03	0.07 (0.03	0.03 (0.02 0.	05 0.	05 -0.0	0.03						0.03	0.01	0.03	0.05 -0	.00 -0.	01 0.0	0.0	0.05	0.01	-0.04	0.01 0	.06 0.0	6 0.0	0.03	1.00	0.10	0.04	0.03 0	0.05 0.	04 -0.02
BP _Diastolic (mmHg)																																					02 -0.01
Follicle No. (L)	0.05	0.05	0.60	0.11	0.17	0.11 0	0.06	0.04	0.07 0.	06 0.	30 -0.0		0.05		0.05	0.06	-0.00 -	0.00	0.12	0.13	0.03 0	.20 -0.	01 0.0	0.02	0.04	0.25	0.31 (0.32 0	.10 0.1	8 0.2	0.01	0.04	0.03	1.00	0.80 0	0.25 0.	15 0.08
Follicle No. (R)	0.09	0.09	0.65	0.16	0.12	0.08	0.01	0.05 (0.01 0.	07 0.	25 -0.1	16 -0.09	0.03		0.02	0.04		0.00	0.10	0.09	0.02 0	.19 -0.	01 0.0	0.03	0.03	0.26	0.27 (0.32 0	.06 0.2	0 0.2	0.05	0.03	0.04	0.80	1.00 0	0.16 0.	19 0.08
Avg. F size (L) (mm)	0.29	0.29	0.13	0.02	0.02			0.05 4	0.03 0	03 0.	03 -0 0	5 -0.07	-0.11		0.05	-0.00	0.01	0.04	0.08	0.03	0.09 0	.13 0.0	0.0	01 -0.0		0.02	0.03 (0.04	03 0.0	3 0.0	0.03	0.05	0.04	0.25	0.16	.00 0.	52 -0.02
Avg. F size (R) (mm)	0.24	0.24	0.10	0.08 4	0.07	0.06 4	0.02		0.02 0	02 0.0	02 -0.0	01 -0.10	-0.12	-0.12	0.07	0.04	0.02	0.03	0.10	0.02	0.09 0	.10 0.0	07 -0.	01 -0.0	0.01		0.00	0.02		2 0.0	1 0.04	0.04	0.02	0.15	0.19	0.52	00 -0.05
Endometrium (mm)																																					05 1.00
														7.0																-							
	2	8	Î	(yrs)	(Kg)	F	dn	Ê	û i	(F) (F)	S S	(S	Î	S	F	7	3	1	£	F	7	3 3	1 =	()	=	Ê	Î	2	2 2	Î	Î	(G)	6	2	(3)	E T	Ê
			3	5	Ξ.	Ö	Gro	pp	E	6	da)	(Yrs)	3	- 25	J/m	E.	E	E	i i	Ĕ	2	E 8		Ę	ò	3	3	3	3	S)	2	Ŧ	E	o'	0.	(mm)	(mm)
	(I)	Ē	S	ge	H	E .	9	rate(bpm)	ths	HD(g/di)	cycle(ror)	SI	ant(Y/N)	d o	딑	딑	글	5	Hip(inch)	st(E.	E S	(00)	gn)	E)	gain(Y/N)	€	ning (Y/N)	loss(Y/N)	p	Se	Ē	Ē	8	ž	5 6	E
		t	PCOS (Y/N)	A	Weight	Height(Cm)	Blood	22	69	- 6	g 5	Status	E S	ap	3,0	3	5	LH(mlU/mL)	I	Waist(inch)	TSH (mIU/L	AMH(ng/mL)	2	PRG(ng/mL)	RBS(mg/dl)	b	So.	in in	은 형	food	ō	0	0	8	93	9	2
		atient File	ů.		5	T	m	Pulse	(breaths/min)		9	S	9	Jo.	ICG(mIU/mL)	ŏ	FSH(mIU/mL	$\dot{\Box}$		>	1	δ 2	: 5	d.	œ	Fig.	6	darke	Hair	ast	×e	stolic	astolic	-	i i	size	net
		0						P	RR		90	aige	ď.	Š	Ŧ	Ŧ	EL.						>			Weight	ie	da	-	Fa	D C	Sys	98	Œ.	ŭ.	ш ц	
								-	CC.		Š	Ta		Z	e te	beta-HCG(mIU/mL)										>	-	S			Reg	1				Avg.	Endo
											170	Ma			-10	D												Š				BP.	0.			4	
												-			_	=																-	ш				

1.0

- 0.8

- 0.6

0.0



Next Steps

Model Selection:

 Choose appropriate machine learning models for binary classification, such as Logistic Regression, Decision Trees, Random Forests, or Support Vector Machines.

Model Training and Evaluation:

- Split the data into training and testing sets.
- Train the selected model on the training set.
- Evaluate the model's performance on the testing set using appropriate classification metrics (accuracy, precision, recall, F1 score).