



## **Data Collection and Preprocessing Phase**

Date	05-07-2024
Team ID	739733
Project Title	Fetal AI: Using Machine Learning To Predict And Monitor Fetal Health
Maximum Marks	6 Marks

## **Data Exploration and Preprocessing Report**

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

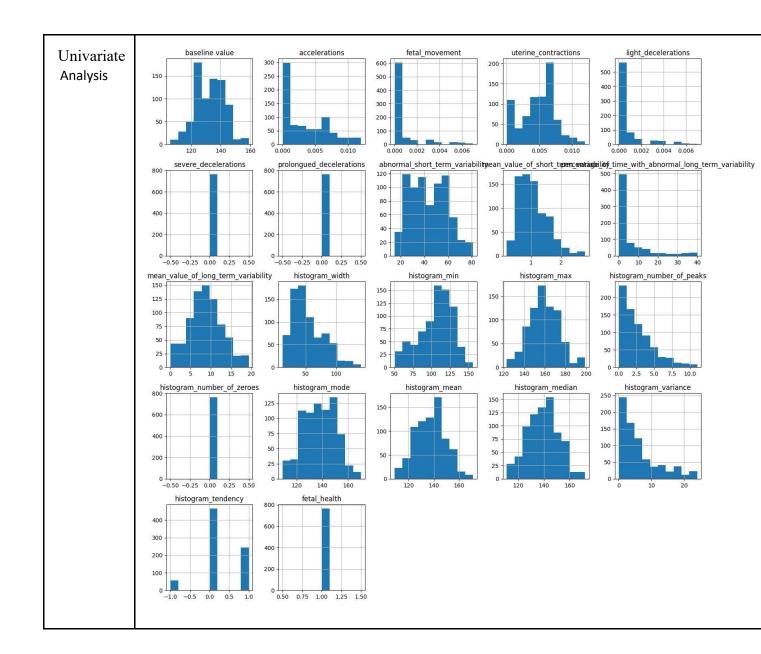




Section	Desc	Description														
	Dime	<u>Dimension:</u>														
	8 row	8 rows × 22 columns														
	Desc	Descriptive statistics:														
	ŏs D	df.desc	ribe <mark>()</mark>													
	<del>_</del>		baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations	severe_decelerations	prolongued_decelerations	abnormal_short_term_variabil						
		count	2126.000000	2126.000000	2126.000000	2126.000000	2126.000000	2126.000000	2126.000000	2126.000						
		mean	133.303857	0.003178	0.009481	0.004366	0.001889	0.000003	0.000159	46.990						
		std	9.840844	0.003866	0.046666	0.002946	0.002960	0.000057	0.000590	17.192						
		min	106.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	12.000						
		25% 50%	126.000000 133.000000	0.000000	0.000000	0.002000	0.000000	0.000000	0.000000	32.000 49.000						
		75%	140.000000	0.006000	0.003000	0.007000	0.003000	0.000000	0.000000	61.000						
		max	160.000000	0.019000	0.481000	0.015000	0.015000	0.001000	0.005000	87.000						
Data		8 rows ×	22 columns													
Overview																

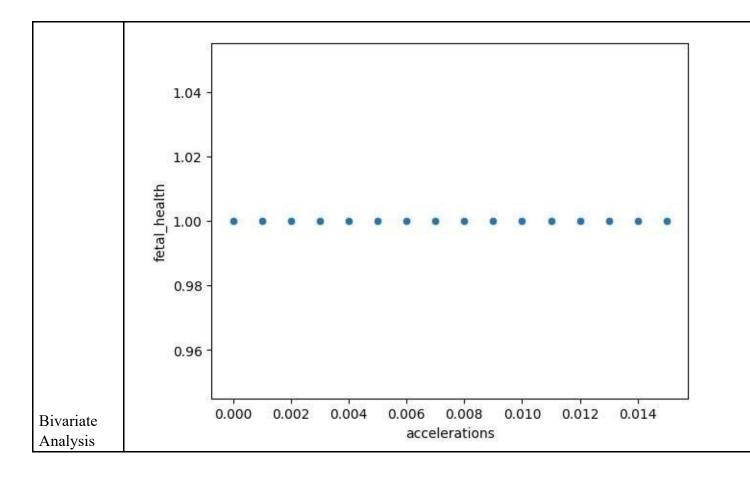
















																								-	1.00
	baseline value	1	0.051	-0.02	0.022	-0.017			0.27	-0.3	0.15	-0.14	-0.025	0.39	0.57	-0.017		0.92	0.92	0.93	-0.088	0.23			
	accelerations -	0.051	1	0.077	0.079	-0.088			-0.067	0.24	-0.36	-0.4	0.34	-0.084	0.49	0.27		0.25	0.3	0.28	0.38	-0.06			
	fetal_movement	-0.02	0.077	1	-0.29	-0.0035			0.35	0.048	-0.047	-0.049	0.17	-0.17	0.03	0.092		-0.036	-0.043	-0.033	0.047	0.08			2.22
	uterine_contractions	0.022	0.079	-0.29	1	0.22			-0.14	0.18	-0.11	-0.12	-0.039	0.097	0.082	0.042		0.057	0.026	0.044	0.2	-0.039			- 0.75
	light_decelerations	-0.017	-0.088	-0.0035	0.22	1			-0.069	0.39	-0.15	-0.1	0.35	-0.38	0.036	0.22		-0.063	-0.19	-0.1	0.61	0.24			
	severe_decelerations -																								
	prolongued_decelerations -																								- 0.50
	abnormal_short_term_variability	0.27	-0.067	0.35	-0.14	-0.069			1	-0.47	0.27	-0.32	-0.11	0.15	0.038	-0.03		0.24	0.21	0.24	-0.11	0.12			
	mean_value_of_short_term_variability								-0.47	1			0.44								0.55				
	percentage_of_time_with_abnormal_long_term_variability									-0.41						-0.14					-0.38				- 0.25
	mean_value_of_long_term_variability										-0.18					0.019					-0.068				
	histogram_width -	-0.025	0.34	0.17	-0.039	0.35			-0.11	0.44	-0.3	0.16	1	-0.84	0.51	0.71		0.058	0.021	0.051	0.63	0.34			- 0.00
	histogram_min	0.39	-0.084	-0.17	0.097	-0.38			0.15	-0.44	0.23	-0.24	-0.84	1	0.037	-0.57		0.35	0.41	0.37	-0.49	-0.45			
	histogram_max -	0.57	0.49	0.03	0.082	0.036			0.038	0.11	-0.18	-0.091	0.51	0.037	1	0.41		0.66	0.69	0.69	0.39	-0.077			
	histogram_number_of_peaks	-0.017	0.27	0.092	0.042	0.22			-0.03	0.34	-0.14	0.019	0.71	-0.57	0.41	1		0.047	0.024	0.054	0.46	0.24			0.25
	histogram_number_of_zeroes																								
	histogram_mode	0.92	0.25	-0.036	0.057	-0.063			0.24	-0.24	0.07	-0.21	0.058	0.35	0.66	0.047		1	0.96	0.98	0.024	0.27			
	histogram_mean	0.92	0.3	-0.043	0.026	-0.19			0.21	-0.32	0.081	-0.19	0.021	0.41	0.69	0.024		0.96	1	0.99	-0.045	0.17			0.50
	histogram_median -	0.93	0.28	-0.033	0.044	-0.1			0.24	-0.27	0.076	-0.22	0.051	0.37	0.69	0.054		0.98	0.99	1	0.0087	0.23			0.50
	histogram_variance	-0.088	0.38	0.047	0.2	0.61			-0.11	0.55	-0.38	-0.068	0.63	-0.49	0.39	0.46		0.024	-0.045	0.0087	1	0.14			
	histogram_tendency -	0.23	-0.06	0.08	-0.039	0.24			0.12	0.013	-0.036	0.052	0.34	-0.45	-0.077	0.24		0.27	0.17	0.23	0.14	1			
	feta <u>l</u> health -																								0.75
	5.4.4.0.0.0.0.0.0.0.0.0	alue -	ions -	nent -	- suoi	- Suoi	- suoi	- Suoi	oility -	oility -	oility -	oility -	idth -	- uiu	- xew	eaks -	- saou	- apoi	- ueau	dian -	- aout	- koue	alth -		
		baseline value	accelerations	fetal_movemer	derine_contraction	ight_decelerat	severe_deceleration	decelerat	rm_variat	rm_variab	rm_variab	m_variab	nistogram_widtl	histogram_mir	histogram_max	ber_of_pe	ber_of_ze	histogram_mode	histogram_mea	histogram_mediar	togram_variance	ogram_tendency	fetal_health		
				fet	uterine	light	severe	prolongued_deceleration	abnormal_short_term_variabilit	short_ter	l_long_ter	f_long_ter	hist	Z	ž	histogram_number_of_peaks	histogram_number_of_zeroes	hist	hist	histor	histogi	histogr			
								bud	abnormal	an_value_of_short_term_variabilit	with_abnormal_long_term_variabilit	nean_value_of_long_term_variabilit				histog	histogi								
										mean	9	mea													
											ntage_of_tin														
Multivariate											perce														
Analysis																									





## **Data Preprocessing Code Screenshots**

Outliers and Anomalies	-								
	os off-	pd.read_csv('/c head()	content/fetal_he						
	<b>⊕</b> ,	baseline value accel	erations fetal	_movement uteri	ne_contractions light_d	ecelerations severe	decelerations prolongue	ed_decelerations abnormal_sh	ort_term_variability
	0		0.000						73.0
	1	132.0	0.006	0.0	0.006	0.003	0.0	0.0	17.0
	3	133.0 134.0	0.003	0.0	0.00B 0.00B	0.003	0.0	0.0	16.0 16.0
	4		0.007		0.008				16.0
	5 row	ws × 22 columns							
T 1' - D-4-									
Loading Data	<u> </u>								
			_						
Handling									
Missing Data	i								
	i								
			_						
Data									
Transformation									
110010101	<u> </u>								
Feature									
			,	~					
Engineering	Attache	ed the co	odes in t	inal subn	nission.				
Save									
Processed Data	<u>-</u>								