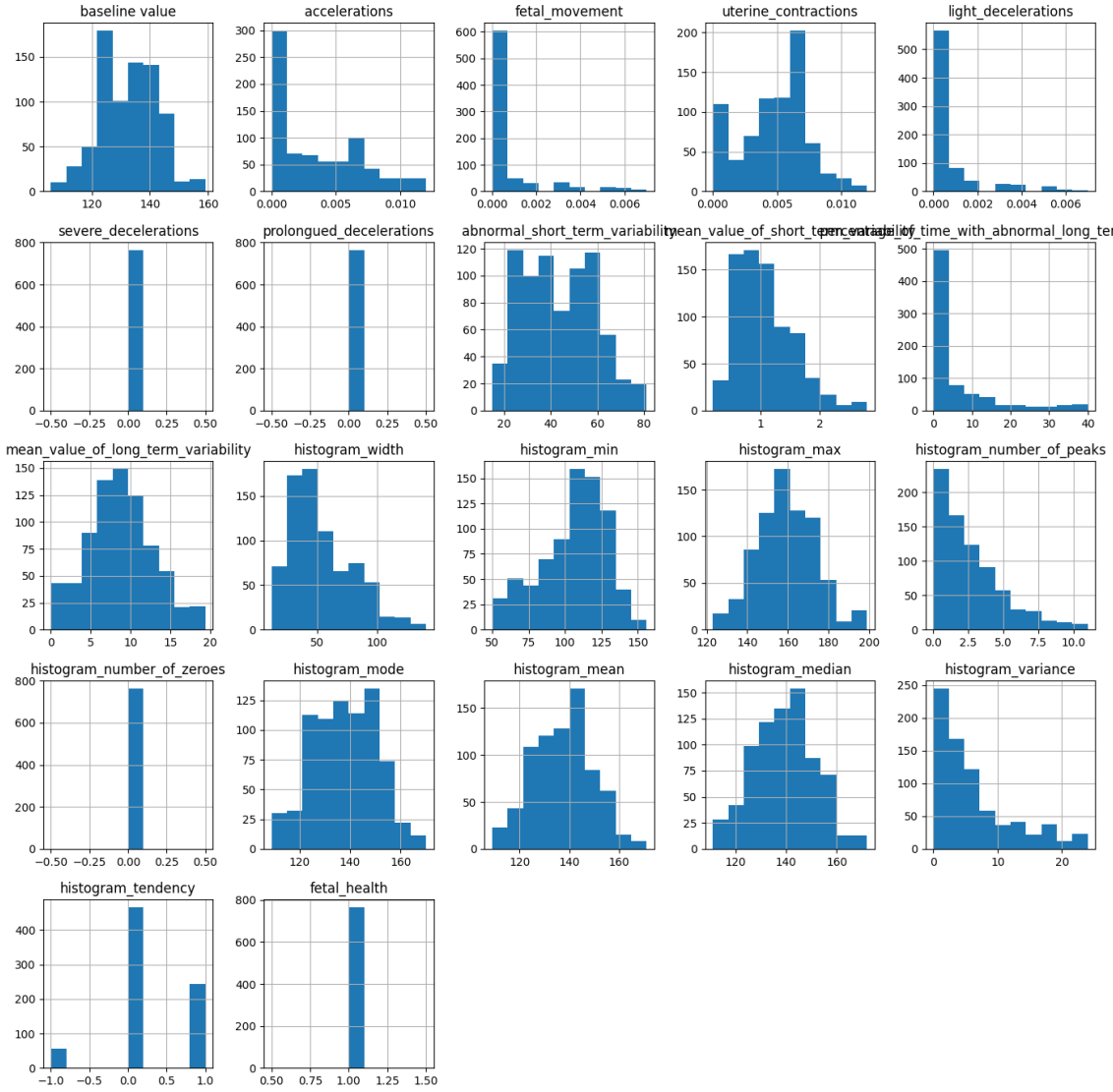
Data Collection and Preprocessing Phase

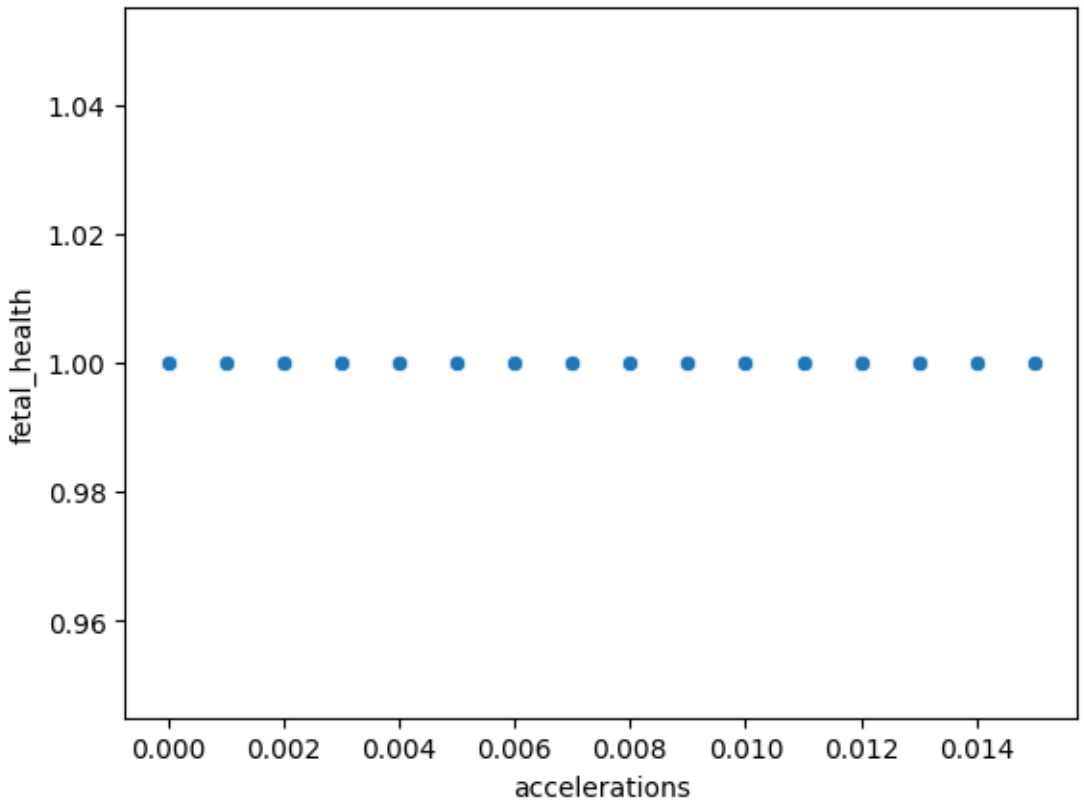
Date	05-07-2024
Team ID	739694
Project Title	FetalAI: Using Machine Learning To Predict And Monitor Fetal Health
Maximum Marks	6 Marks

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.

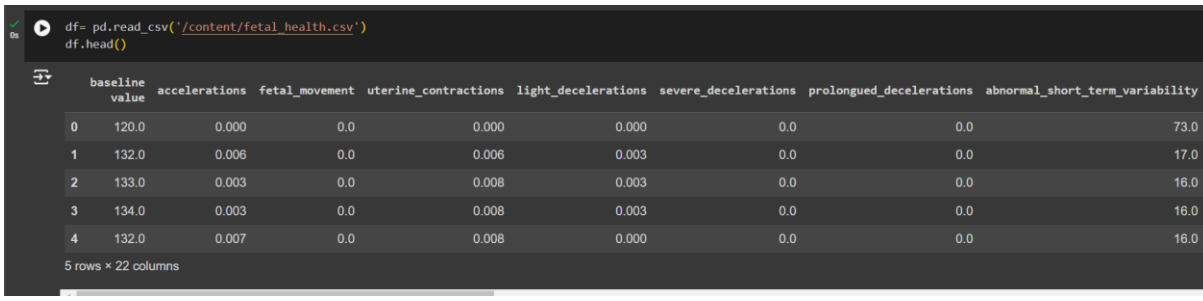
Section	Description
Data Overview	Dimension: 8 rows × 22 columns Descriptive statistics: `df.describe()` baseline value: count 2126.000000, mean 133.303857, std 9.840844, min 106.000000, 25% 126.000000, 50% 133.000000, 75% 140.000000, max 160.000000 accelerations: count 2126.000000, mean 0.003178, std 0.003866, min 0.000000, 25% 0.000000, 50% 0.002000, 75% 0.006000, max 0.019000 fetal_movement: count 2126.000000, mean 0.009481, std 0.046666, min 0.000000, 25% 0.000000, 50% 0.000000, 75% 0.003000, max 0.481000 uterine_contractions: count 2126.000000, mean 0.004366, std 0.002946, min 0.000000, 25% 0.002000, 50% 0.004000, 75% 0.007000, max 0.015000 light_decelerations: count 2126.000000, mean 0.001889, std 0.002960, min 0.000000, 25% 0.000000, 50% 0.000000, 75% 0.003000, max 0.015000 severe_decelerations: count 2126.000000, mean 0.000003, std 0.000057, min 0.000000, 25% 0.000000, 50% 0.000000, 75% 0.000000, max 0.001000 prolongued_decelerations: count 2126.000000, mean 0.000159, std 0.000590, min 0.000000, 25% 0.000000, 50% 0.000000, 75% 0.000000, max 0.005000 abnormal_short_term_variabil: count 2126.000, mean 46.990, std 17.192, min 12.000, 25% 32.000, 50% 49.000, 75% 61.000, max 87.000 8 rows × 22 columns
Univariate Analysis	


baseline value
accelerations
fetal_movement
uterine_contractions
light_decelerations
severe_decelerations
prolongued_decelerations
abnormal_short_term_variability
mean_value_of_short_term_variability
percentage_of_time_with_abnormal_long_term_variability
mean_value_of_long_term_variability
histogram_width
histogram_min
histogram_max
histogram_number_of_peaks
histogram_number_of_zeroes
histogram_mode
histogram_mean
histogram_median
histogram_variance
histogram_tendency
fetal_health

Bivariate Analysis	

Multivariate Analysis	
Outliers and Anomalies	-

Data Preprocessing Code Screenshots

Loading Data	
Handling Missing Data	-
Data Transformation	-
Feature Engineering	Attached the codes in final submission.
Save Processed Data	-