

Data Collection and Preprocessing Phase

Date	7 July 2024
Team ID	739876
Project Title	FetalAI: Using Machine Learning To Predict And Monitor Fetal Health
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section

Description

Data Overview

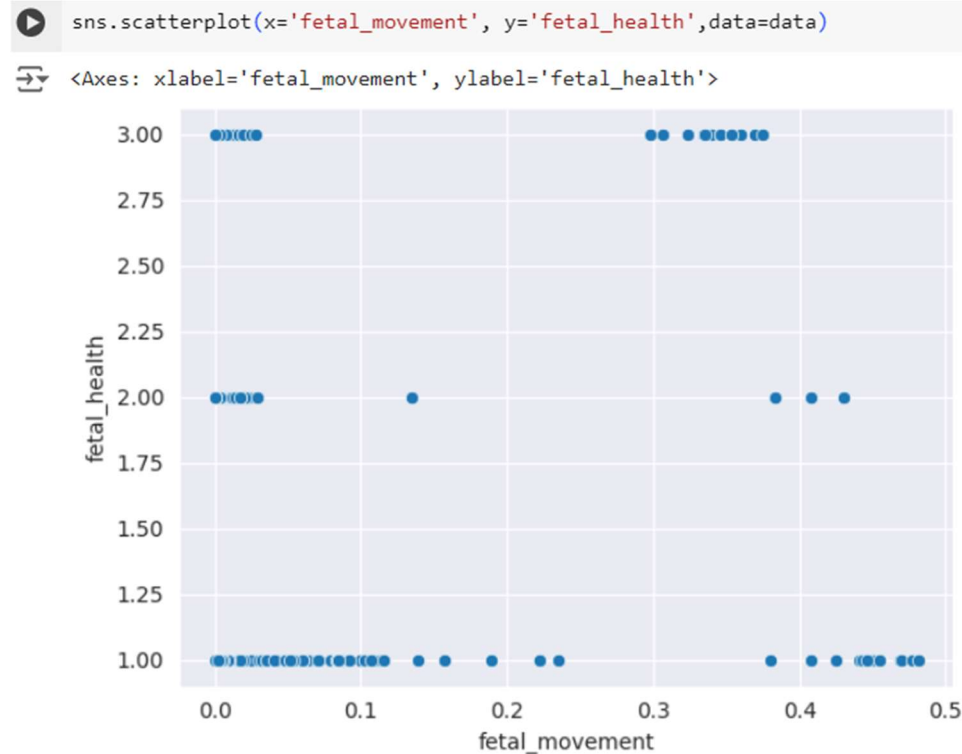
data.describe().T

	count	mean	std	min	25%	50%	75%	max
baseline value	2126.0	133.303857	9.840844	106.0	126.000	133.000	140.000	160.000
accelerations	2126.0	0.003178	0.003866	0.0	0.000	0.002	0.006	0.019
fetal_movement	2126.0	0.009481	0.046666	0.0	0.000	0.000	0.003	0.481
uterine_contractions	2126.0	0.004366	0.002946	0.0	0.002	0.004	0.007	0.015
light_decelerations	2126.0	0.001889	0.002960	0.0	0.000	0.000	0.003	0.015
severe_decelerations	2126.0	0.000003	0.000057	0.0	0.000	0.000	0.000	0.001
prolongued_decelerations	2126.0	0.000159	0.000590	0.0	0.000	0.000	0.000	0.005
abnormal_short_term_variability	2126.0	46.990122	17.192814	12.0	32.000	49.000	61.000	87.000
mean_value_of_short_term_variability	2126.0	1.332785	0.883241	0.2	0.700	1.200	1.700	7.000
percentage_of_time_with_abnormal_long_term_variability	2126.0	9.846660	18.396880	0.0	0.000	0.000	11.000	91.000
mean_value_of_long_term_variability	2126.0	8.187629	5.628247	0.0	4.600	7.400	10.800	50.700
histogram_width	2126.0	70.445908	38.955693	3.0	37.000	67.500	100.000	180.000
histogram_min	2126.0	93.579492	29.560212	50.0	67.000	93.000	120.000	159.000
histogram_max	2126.0	164.025400	17.944183	122.0	152.000	162.000	174.000	238.000
histogram_number_of_peaks	2126.0	4.068203	2.949386	0.0	2.000	3.000	6.000	18.000
histogram_number_of_zeroes	2126.0	0.323612	0.706059	0.0	0.000	0.000	0.000	10.000
histogram_mode	2126.0	137.452023	16.381289	60.0	129.000	139.000	148.000	187.000
histogram_mean	2126.0	134.610536	15.593596	73.0	125.000	136.000	145.000	182.000
histogram_median	2126.0	138.090310	14.466589	77.0	129.000	139.000	148.000	186.000
histogram_variance	2126.0	18.808090	28.977636	0.0	2.000	7.000	24.000	269.000
histogram_tendency	2126.0	0.320320	0.610829	-1.0	0.000	0.000	1.000	1.000
fetal_health	2126.0	1.304327	0.614377	1.0	1.000	1.000	1.000	3.000

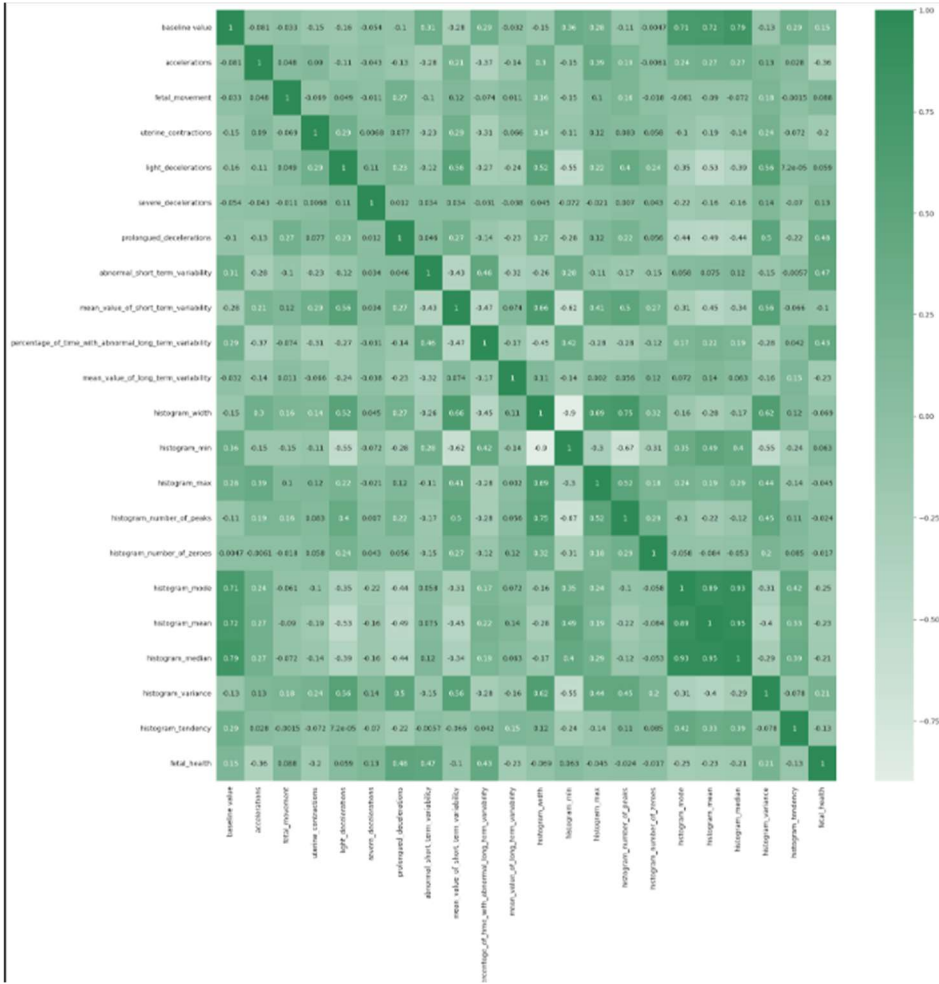
Univariate Analysis



Bivariate Analysis



Multivariate Analysis



Data Preprocessing Code Screenshots

Loading Data

```
[ ] data = pd.read_csv("/content/fetal_health_dataset.csv")

[ ] data.head()

[ ]
```

	baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations	severe_decelerations	prolonged_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
0	120	0.000	0.0	0.000	0.000	0.0	0.0	0.0	73													
1	132	0.006	0.0	0.006	0.003	0.0	0.0	0.0	17													
2	133	0.003	0.0	0.008	0.003	0.0	0.0	0.0	16													
3	134	0.003	0.0	0.008	0.003	0.0	0.0	0.0	16													
4	132	0.007	0.0	0.008	0.000	0.0	0.0	0.0	16													

5 rows x 22 columns

Handling Missing Data

```
data.isnull().sum()
```

baseline_value	0
accelerations	0
fetal_movement	0
uterine_contractions	0
light_decelerations	0
severe_decelerations	0
prolongued_decelerations	0
abnormal_short_term_variability	0
mean_value_of_short_term_variability	0
percentage_of_time_with_abnormal_long_term_variability	0
mean_value_of_long_term_variability	0
histogram_width	0
histogram_min	0
histogram_max	0
histogram_number_of_peaks	0
histogram_number_of_zeroes	0
histogram_mode	0
histogram_mean	0
histogram_median	0
histogram_variance	0
histogram_tendency	0
fetal_health	0
dtype: int64	

```
data.isnull().any()
```

baseline_value	False
accelerations	False
fetal_movement	False
uterine_contractions	False
light_decelerations	False
severe_decelerations	False
prolongued_decelerations	False
abnormal_short_term_variability	False
mean_value_of_short_term_variability	False
percentage_of_time_with_abnormal_long_term_variability	False
mean_value_of_long_term_variability	False
histogram_width	False
histogram_min	False
histogram_max	False
histogram_number_of_peaks	False
histogram_number_of_zeroes	False
histogram_mode	False
histogram_mean	False
histogram_median	False
histogram_variance	False
histogram_tendency	False
fetal_health	False
dtype: bool	

Data Transformation

Scaling Data

```
scale = MinMaxScaler()
X = pd.DataFrame(scale.fit_transform(X), columns=X.columns)
X.head()
```

	prolongued_decelerations	abnormal_short_term_variability	percentage_of_time_with_abnormal_long_term_variability	histogram_variance	histogram_median	mean_value_of_long_term_variability
0	0.0	0.813333	0.472527	0.271375	0.403670	0.047337
1	0.0	0.066667	0.000000	0.044610	0.577982	0.205128
2	0.0	0.053333	0.000000	0.048327	0.559633	0.264300
3	0.0	0.053333	0.000000	0.048327	0.550459	0.453649
4	0.0	0.053333	0.000000	0.040892	0.559633	0.392505

Feature Engineering	Attached the codes in final submission.
Save Processed Data	-