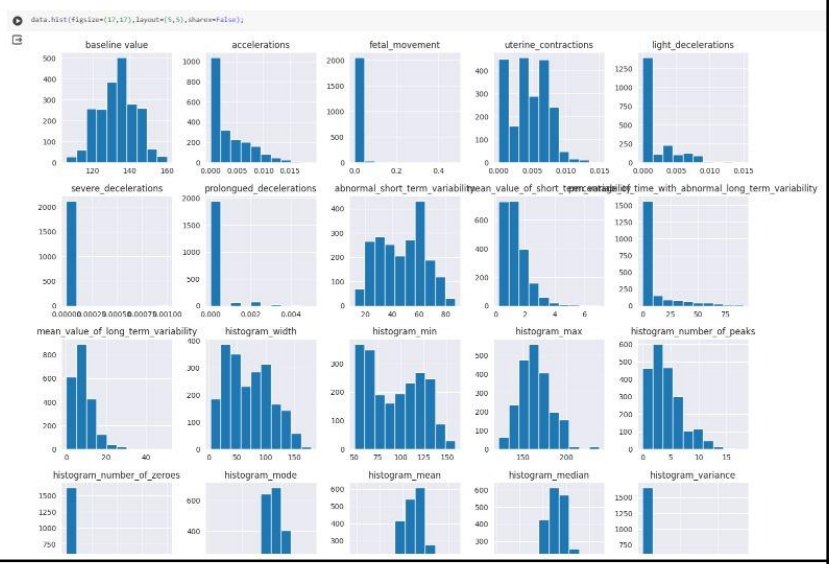
# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 30 April 2024 |
| Team ID | Team - 737850 |
| 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:<br>2126 rows × 22 columns<br>Descriptive statistics:<br><br>`#data.info()`<br><br>`<class 'pandas.core.frame.DataFrame'>`<br>`RangeIndex: 2126 entries, 0 to 2125`<br>`Data columns (total 22 columns):`<br>` #  Column  Non-Null Count  Dtype`<br>`---  ------  --------------  -----`<br>` 0  baseline value  2126 non-null  float64`<br>` 1  accelerations  2126 non-null  float64`<br>` 2  fetal_movement  2126 non-null  float64`<br>` 3  uterine_contractions  2126 non-null  float64`<br>` 4  light_decelerations  2126 non-null  float64`<br>` 5  severe_decelerations  2126 non-null  float64`<br>` 6  prolongued_decelerations  2126 non-null  float64`<br>` 7  abnormal_short_term_variability  2126 non-null  float64`<br>` 8  mean_value_of_short_term_variability  2126 non-null  float64`<br>` 9  percentage_of_time_with_abnormal_long_term_variability  2126 non-null  float64`<br>` 10  mean_value_of_long_term_variability  2126 non-null  float64`<br>` 11  histogram_width  2126 non-null  float64`<br>` 12  histogram_min  2126 non-null  float64`<br>` 13  histogram_max  2126 non-null  float64`<br>` 14  histogram_number_of_peaks  2126 non-null  float64`<br>` 15  histogram_number_of_zeroes  2126 non-null  float64`<br>` 16  histogram_mode  2126 non-null  float64`<br>` 17  histogram_mean  2126 non-null  float64`<br>` 18  histogram_median  2126 non-null  float64`<br>` 19  histogram_variance  2126 non-null  float64`<br>` 20  histogram_tendency  2126 non-null  float64`<br>` 21  fetal_health  2126 non-null  float64`<br>`dtypes: float64(22)`<br>`memory usage: 365.5 KB` |
| Univariate Analysis | |

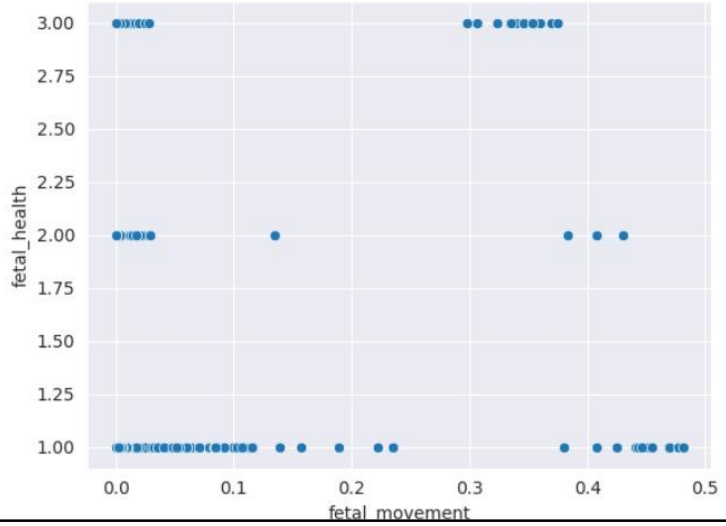
<table>
<tr>
<td></td>
<td>
Univariate analysis


</td>
</tr>
<tr>
<td>Bivariate Analysis</td>
<td>
Bivariate Analysis

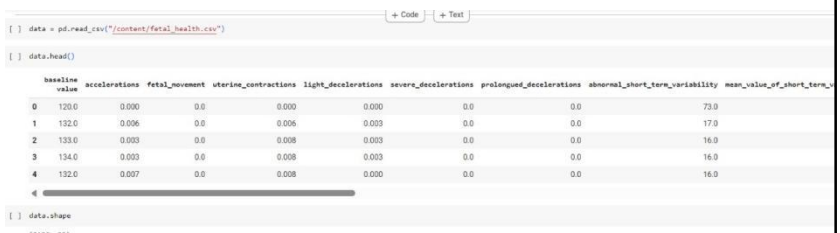
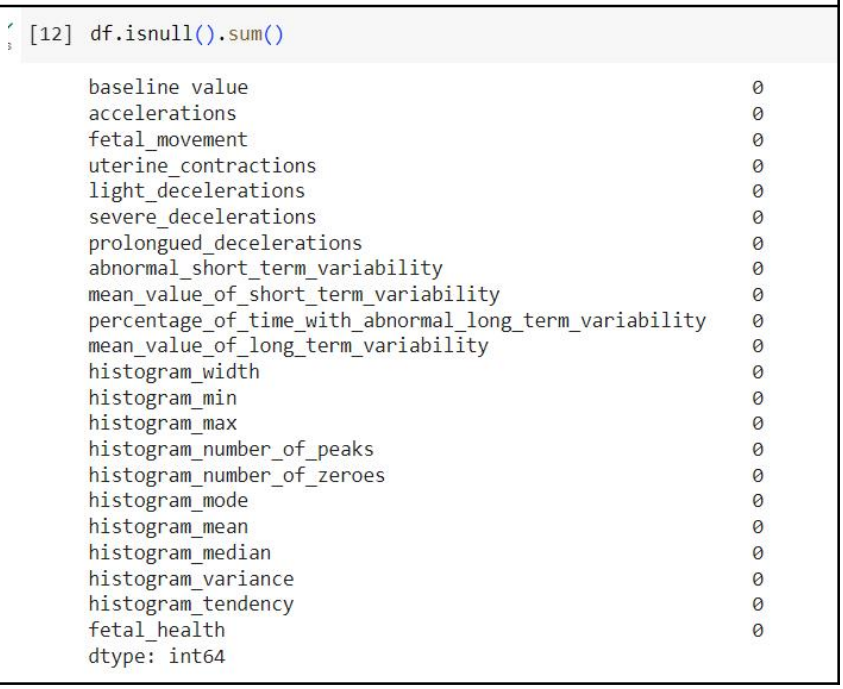
```
import seaborn as sns

sns.scatterplot(x=data.fetal_movement, y=data.fetal_health)
```

```
<Axes: xlabel='fetal_movement', ylabel='fetal_health'>
```



</td>
</tr>
</table>

| | |
|---|---|
| Multivariate Analysis |  |
| Outliers and Anomalies | - |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Handling Missing Data |  |



```
[12] df.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 Transformation | The data is already transformed. |
| Feature Engineering | Attached the codes in final submission. |
| Save Processed Data | - |