

```
In [16]: """
@Authour:      Adharsh.S
Date created:   20.10.22
language written: python

"""

#importing the necessary package for the usage of

import matplotlib.pyplot as plt # for plotting the file if we needed for visualization
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.preprocessing import StandardScaler
sns.set(color_codes=True) # adds a nice background to the graphs
%matplotlib inline
import pandas as pd
import numpy as np
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
```

```
In [17]: #header_list = ["baseline value "," accelerations","fetal_health "]
df = pd.read_csv("train.csv")#names=header_list)
df.head()
df.shape
df.info()
df.describe()
```

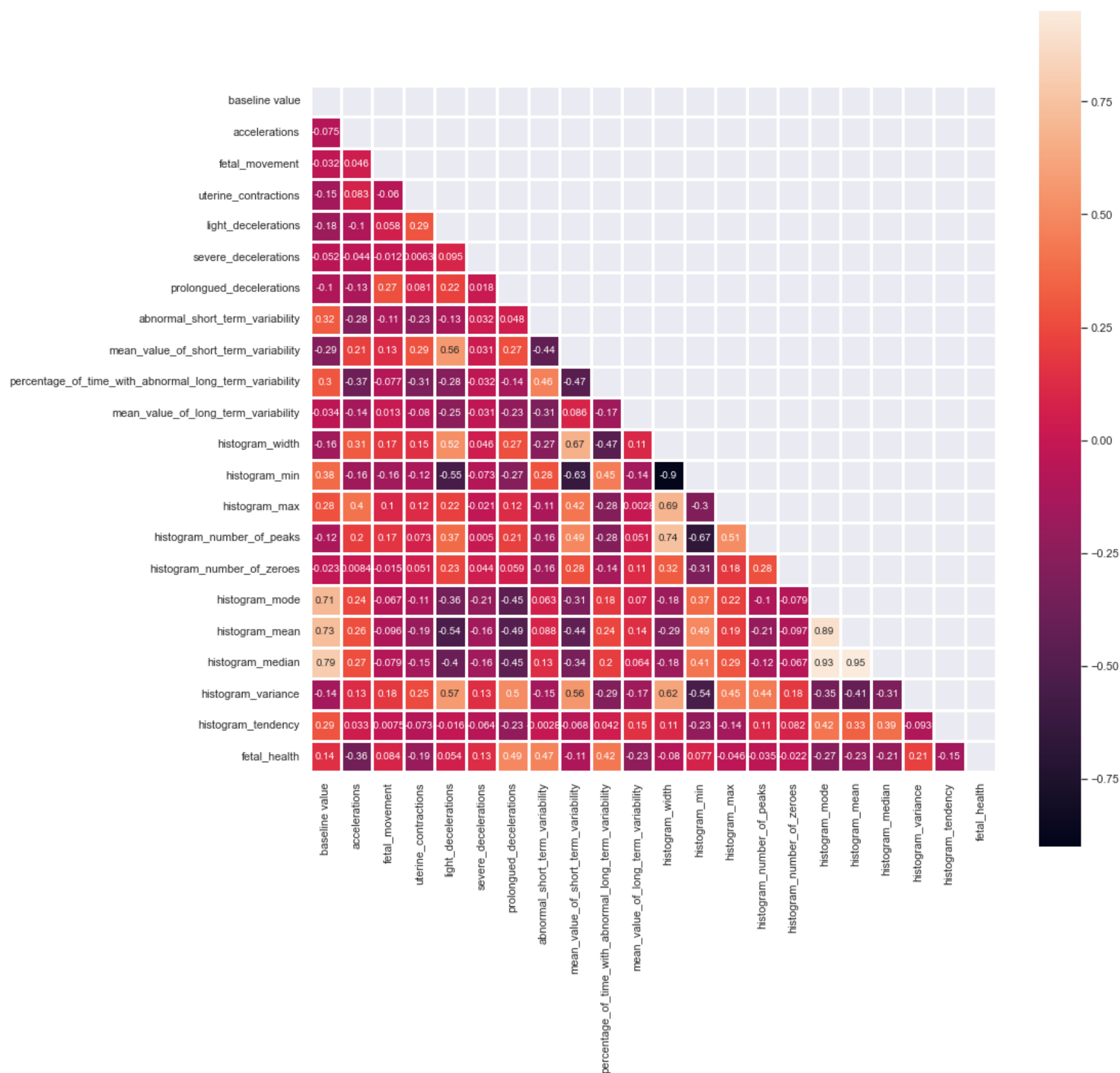
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1700 entries, 0 to 1699
Data columns (total 22 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   baseline value                             1700 non-null   int64
1   accelerations                             1700 non-null   float64
2   fetal_movement                             1700 non-null   float64
3   uterine_contractions                       1700 non-null   float64
4   light_decelerations                       1700 non-null   float64
5   severe_decelerations                       1700 non-null   float64
6   prolonged_decelerations                   1700 non-null   float64
7   abnormal_short_term_variability            1700 non-null   int64
8   mean_value_of_short_term_variability       1700 non-null   float64
9   percentage_of_time_with_abnormal_long_term_variability 1700 non-null   int64
10  mean_value_of_long_term_variability         1700 non-null   float64
11  histogram_width                             1700 non-null   int64
12  histogram_min                              1700 non-null   int64
13  histogram_max                              1700 non-null   int64
14  histogram_number_of_peaks                  1700 non-null   int64
15  histogram_number_of_zeroes                 1700 non-null   int64
16  histogram_mode                             1700 non-null   int64
17  histogram_mean                             1700 non-null   int64
18  histogram_median                           1700 non-null   int64
19  histogram_variance                         1700 non-null   int64
20  histogram_tendency                         1700 non-null   int64
21  fetal_health                               1700 non-null   int64
dtypes: float64(8), int64(14)
memory usage: 292.3 KB
```

Out[17]:

	baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations	severe_decelerations	prolongued_decelerations	abnorma
count	1700.000000	1700.000000	1700.000000	1700.000000	1700.000000	1700.000000	1700.000000	
mean	133.213529	0.003212	0.010211	0.004356	0.001899	0.000004	0.000158	
std	9.873344	0.003888	0.050124	0.002943	0.002976	0.000059	0.000587	
min	106.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	126.000000	0.000000	0.000000	0.002000	0.000000	0.000000	0.000000	
50%	133.000000	0.002000	0.000000	0.004000	0.000000	0.000000	0.000000	
75%	140.000000	0.006000	0.003000	0.006000	0.003000	0.000000	0.000000	
max	159.000000	0.019000	0.481000	0.015000	0.015000	0.001000	0.005000	

8 rows × 22 columns

```
In [36]: Target = df["fetal_health"]
corr = df.corr()
mask = np.zeros_like(corr)
mask[np.triu_indices_from(mask)] = True
with sns.axes_style("dark"):
    f, ax = plt.subplots(figsize=(15, 15))
    ax = sns.heatmap(corr,mask=mask,square=True,linewidths=2.5,cmap="rocket",annot=True)
```



```
In [19]: df_dup = df.drop_duplicates(subset = None , keep = 'first', inplace = False)
```

```
In [20]: print("Count of type 1.0 fetal health in the dataset ",len(df.loc[df["fetal_health"]==1.0]))
print("Count of type 2.0 fetal health in the dataset ",len(df.loc[df["fetal_health"]==2.0]))
print("Count of type 3.0 fetal health in the dataset ",len(df.loc[df["fetal_health"]==3.0]))
```

```
Count of type 1.0 fetal health in the dataset 1323
Count of type 2.0 fetal health in the dataset 236
Count of type 3.0 fetal health in the dataset 141
```

```
In [21]: X = df_dup.iloc[:, :-1]
y = df_dup.iloc[:, -1]
```



