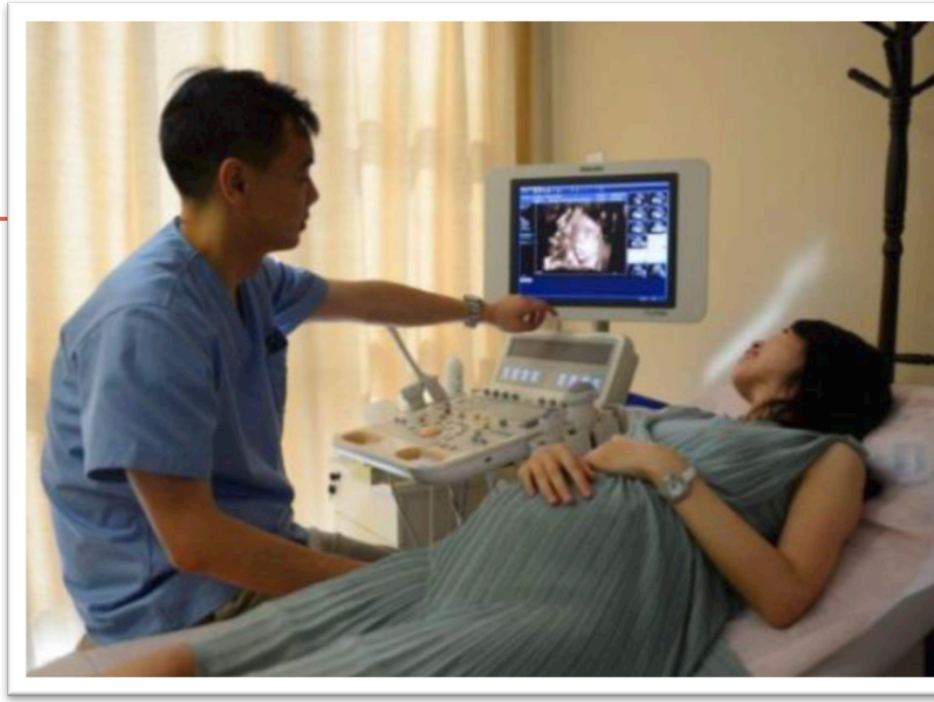
FETAL HEART RATES DETECTION

Juan Zhang 100777497



Business problem

Problem:

Fetal distress during pregnancy is caused by a lack of oxygen. Baby who experience the fetal distress might well risk from brain injury, cerebral palsy and even stillbirth. It's important to monitor fetal health status in every period of time.

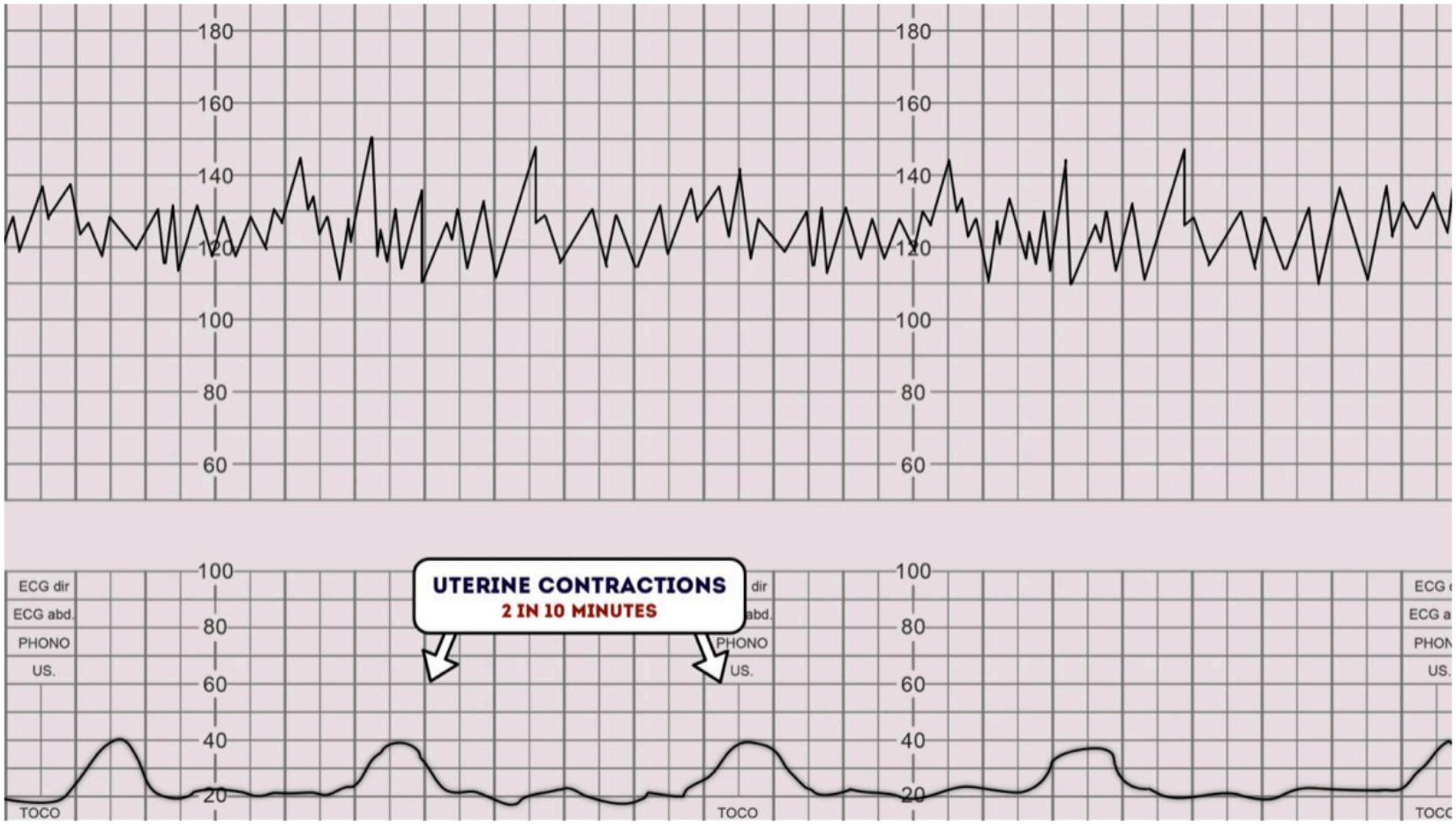
Symptom:

If a fetal is suffering from a lack of oxygen, his or her heart rate will differ from the healthy babies. Their reaction to the uterine contraction is also different.

Solution:

Cardiotocography (CTG) is used during pregnancy to monitor fetal heart rate and uterine contractions. It is most commonly used in the third trimester and its purpose is to monitor fetal well-being and allow early detection of fetal distress. An abnormal CTG may indicate the need for further investigations and potential intervention

Cardiogram



Uterine contractions (CTG)

Data source

- Where it from?

Faculty of Medicine, University of Porto, Portugal. There are 2126 fetal cardiotocograms included in this dataset, the dataset is classified by expert obstetricians

How does it looks?

columns	interpretation
b	start instant
e	end instant
UC	uterine contractions, present in a 10 minute period
LB	baseline rate is the average heart rate of the fetus within a 10-minute window.(normal :110-160 bpm),(bradycardia,<100)
MSTV	The mean value of SHORT term variation of fetal heart rate from one beat to the next
ASTV	percentage of time with abnormal SHORT term variability (SisPorto)
AC	of greater than 15 bpm for greater than 15 seconds. Accelerations occurring alongside uterine contractions is a sign of a healthy fetus. No
DS	severe decelerations<Not significance>
DL	Light decelerations
DP	prolongued decelerations,a deceleration that lasts more than 3 minutes. <abnormal >3 mins >
DR	repetitive decelerations
ALTV	percentage of time with abnormal LONG term variability (SisPorto)
MLTV	The mean value of LONG term variation of fetal heart rate from one beat to the next
FM	fetal movement (SisPorto)
Width	histogram width
Min	low freq. of the histogram

	columns	interpretation
18	Max	high freq. of the histogram
19	Nmax	number of histogram peaks
20	Nzeros	number of histogram zeros
21	Mode	histogram mode
22	Mean	histogram mean
23	Median	histogram median
24	Variance	histogram variance
25	Tendency	histogram tendency: -1=left assymetric; 0=symmetric; 1=right assymetric
26	A	calm sleep
27	B	REM sleep
28	C	calm vigilance
29	D	active vigilance
30	SH	shift pattern (A or Susp with shifts)
31	AD	accelerative/decelerative pattern (stress situation)
32	DE	decelerative pattern (vagal stimulation)
33	LD	largely decelerative pattern
34	FS	flat-sinusoidal pattern (pathological state)
35	SUSP	suspect pattern
36	NSP	Normal=1; Suspect=2; Pathologic=3

Data Acquisition

```
1 import pandas as pd
2 df = pd.read_csv('./fetalHeartRate.csv')
3 df = df.drop(columns=['AC.1', 'FM.1', 'UC.1','DL.1', 'DS.1', 'DP.1'],axis=1)
4 df.shape #(2129, 35)
5 df.head(10)
```

	b	e	AC	FM	UC	DL	DS	DP	DR	LB	...	B	C	D	E	AD	DE	LD	FS	SUSP	NSP
0	240.0	357.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	120.0	...	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	2.0
1	5.0	632.0	4.0	0.0	4.0	2.0	0.0	0.0	0.0	132.0	...	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	−1.0	−1.0	−1.0	1.0
2	177.0	779.0	2.0	0.0	5.0	2.0	0.0	0.0	0.0	133.0	...	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	−1.0	−1.0	−1.0	1.0
3	411.0	1192.0	2.0	0.0	6.0	2.0	0.0	0.0	0.0	134.0	...	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	−1.0	−1.0	−1.0	1.0
4	533.0	1147.0	4.0	0.0	5.0	0.0	0.0	0.0	0.0	132.0	...	1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0
5	0.0	953.0	1.0	0.0	10.0	9.0	0.0	2.0	0.0	134.0	...	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	−1.0	3.0
6	240.0	953.0	1.0	0.0	9.0	6.0	0.0	2.0	0.0	134.0	...	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	−1.0	3.0
7	62.0	679.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	122.0	...	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	3.0
8	120.0	779.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	122.0	...	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	3.0
9	181.0	1192.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	122.0	...	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	3.0

10 rows × 35 columns

Data Cleansing

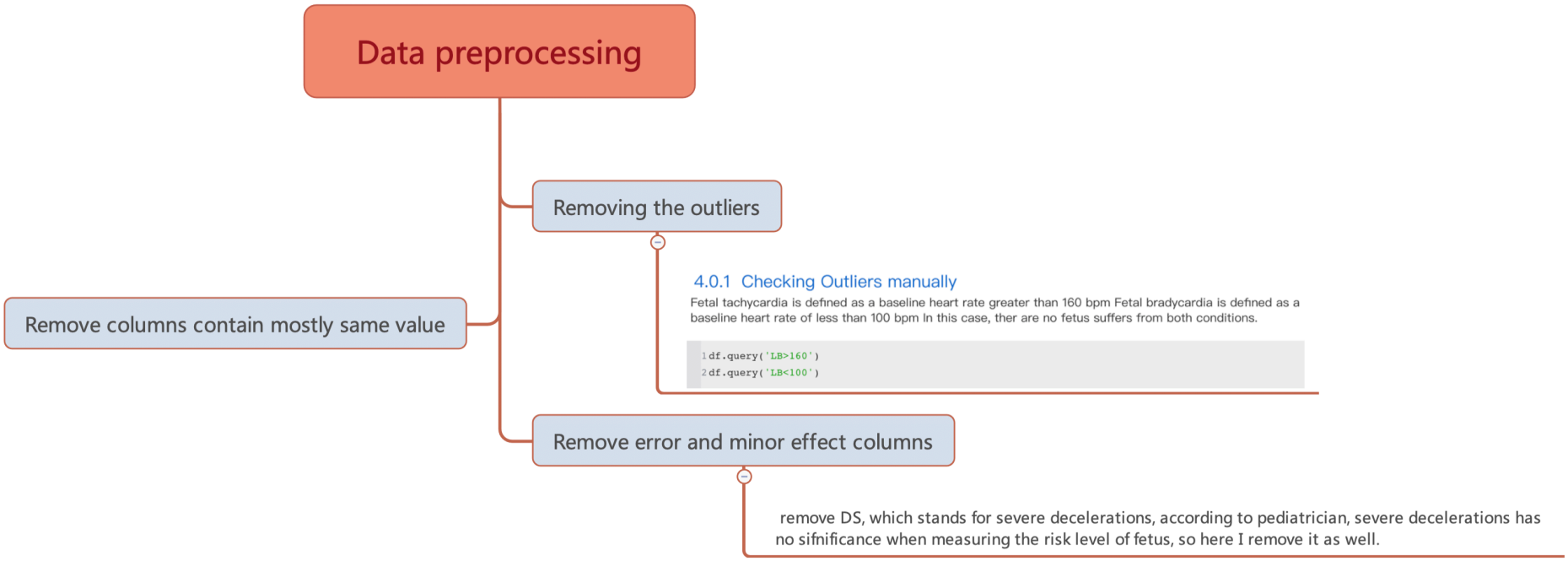
- 1. Considering there are total 2129 simples, and the max number of null value in each column is 3, Direct remove the rows that contain null value won't affect result greatly.
- 2. Datatype: all float number, that is what we want.
- 3. Remove duplicated rows.

```
df = df.drop_duplicates()
df.shape # (2116, 41)
df.head(5)
```

	b	e	AC	FM	UC	DL	DS	DP	DR	LB	...	B	C	D	E	AD	DE	LD	FS	SUSP	NSP
0	240.0	357.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	120.0	...	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	2.0
1	5.0	632.0	4.0	0.0	4.0	2.0	0.0	0.0	0.0	132.0	...	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	−1.0	−1.0	−1.0	1.0
2	177.0	779.0	2.0	0.0	5.0	2.0	0.0	0.0	0.0	133.0	...	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	−1.0	−1.0	−1.0	1.0
3	411.0	1192.0	2.0	0.0	6.0	2.0	0.0	0.0	0.0	134.0	...	−1.0	−1.0	−1.0	−1.0	1.0	−1.0	−1.0	−1.0	−1.0	1.0
4	533.0	1147.0	4.0	0.0	5.0	0.0	0.0	0.0	0.0	132.0	...	1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	1.0

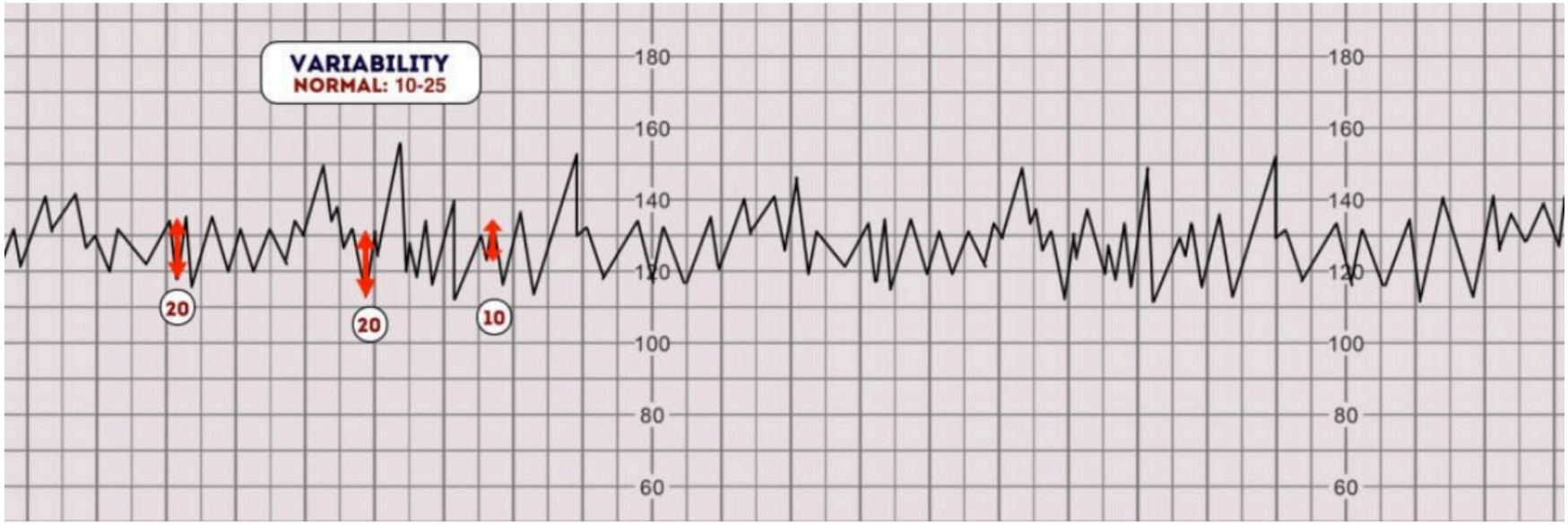
5 rows × 35 columns

Data Preprocessing

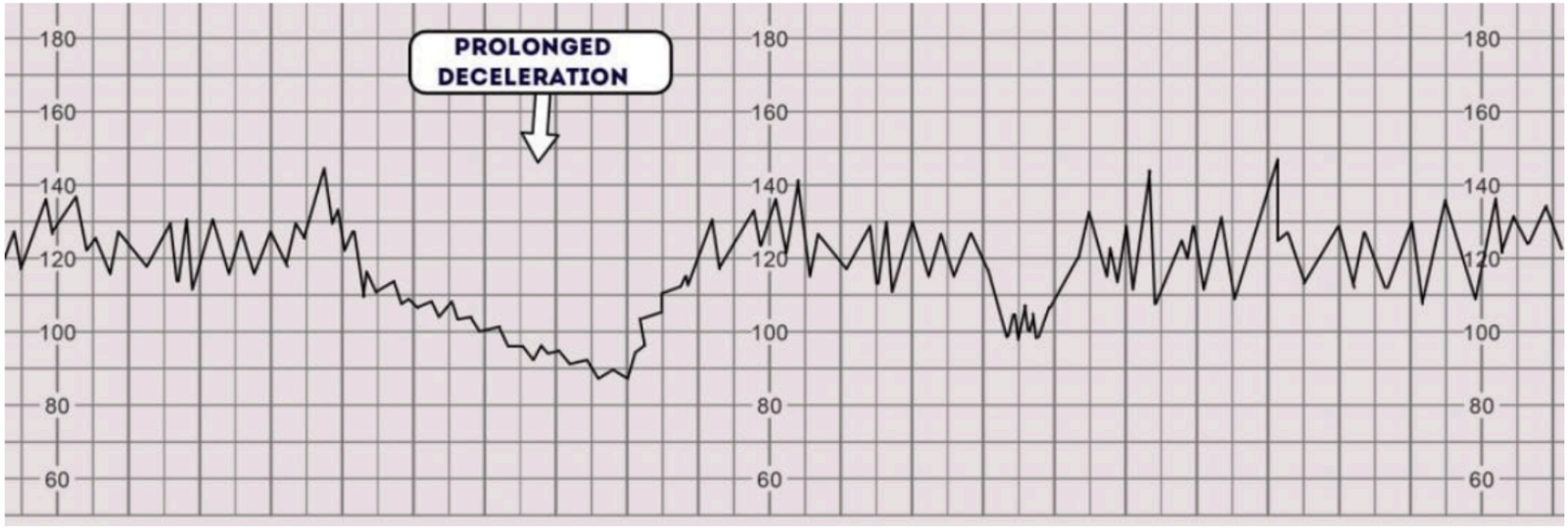


EDA

1.Baseline variability refers to the variation of fetal heart rate from one beat to the next. A healthy fetus baseline variability range from 5 to 25 bpm.
Abnormal:
less than 5 bpm for more than 50 minutes



There are no outliers in Baseline variability.



For prolonged deceleration, If it lasts between 2–3 minutes it is classed as non–reassuring. If it lasts longer than 3 minutes it is immediately classed as abnormal. In this case, there are samples that prolonged deceleration are zero, it's impossible for fetus that have no heat beat deceleration. There are about 91% samples have 0 prolonged deceleration. Let's remove this column.

```
df.query('DP>3') # 1 rows × 33 columns
cols = df.query('DP==0')['DP'].count()
# 1938 rows
rows = df.shape[0]
print(cols / rows) # 0.9158790170132325
df = df.drop(columns=['DP'],axis=1)
df.head(5)
```

Alternative way to select features

Using sklearn RFE moduel to estimate top k important features.

```
from sklearn.feature_selection import RFE
from sklearn.linear_model import LinearRegression
from sklearn.metrics import accuracy_score
import seaborn as sns
from sklearn import metrics
from sklearn.model_selection import train_test_split

def getRankedColumns(df):↔

keepColums =getRankedColumns(df)
def scoreHeapMap(y_predict,y_test):↔
def model_and_score(keepColums,df):↔
```

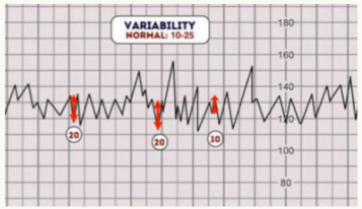
Why Logistic Regression

Why logistic regression

Muti-class task

NSP Score		
	Normal	1
	Suspect	2
	Pathologic	3

Features in this case independent from eachother.



Feature dimention is not too big , here less than 30 features.

Where to deploy diagnostic system