Use MATLAB to have a quick basic look into the training dataset.

A	B	C	D	E
Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...
数值	数值	数值	数值	数值
Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...
10000	3000	8000	3000	3000
8000	8000	2000	3000	2000
5000	15000	8000	3000	3000
35000	8000	8000	5000	5000
16000	2000	2000	500	500
32000	6000	2000	2000	2000
10000	3000	3000	2000	2000
2000	3000	500	500	500
11000	3000	2000	2000	3000
5000	2000	500	2000	2000
2000	500	500	500	500
2000	500	500	500	500
5000	5000	6000	2000	2000
2000	500	500	500	500

Fig. 1

Get:

1. The data was preprocessed already.
2. Too many features.
3. Different feature type.
4. PCA might be necessary.
5. More than 1 label.
6. Dataset not very large.

T	U	V	W	X	Y	Z	AA	AB	AC
train									
Quan_5	Quan_6	Quan_7	Quan_8	Quan_9	Quan_10	Quan_11	Quan_12	Quan_13	Quan_14
数值	数值	数值	数值	数值	数值	数值	数值	数值	数值
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
58	29	13	58	58	38	26	2	38	38
22	13	0	22	22	17	10	0	17	17
9	1	3	9	9	7	0	2	7	7
17	2	8	17	17	10	2	3	10	10
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	1	0	1	1	1	1	0	1	1
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	1	0	4	4	4	1	0	4	4
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	2	0	4	4	3	1	0	3	3
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	0	0	5	5	3	0	0	3	3
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	5	8	14	14	7	1	3	7	7
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
23	14	0	23	23	7	3	0	7	7
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fig. 2

7. Lots of missing values formatted as: ‘NaN’

Use MATLAB to have a quick basic look into the sample submission dataset.

A	B	C	D	E	F	G	H	I	J	K	L	M
samplesubmission												
id	Outcome_M1	Outcome_M2	Outcome_M3	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_M12
数值	数值	数值	数值	数值	数值	数值	数值	数值	数值	数值	数值	数值
d	Outcome_M1	Outcome_M2	Outcome_M3	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_...	Outcome_M12
1	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
2	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
3	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
4	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
5	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
5	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
7	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
8	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
9	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
10	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
11	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
12	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
13	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
14	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
15	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818
16	20620.5059920107	10837.5668449198	5194.36997319...	3494.6380...	2852.1320...	2018.4397...	1751.4409...	1492.7219...	1411.3737...	1320.8841...	1190.6976...	1071.98748043818

Fig. 3

Get:

1. More than 1 response/label needs to be predicted.

Ok, now, some pandas/seaborn based EDA.