

Ввод [1]:

1

from dataprep.eda import plot, plot\_correlation, create\_report, plot\_missing

2

import pandas as pd

3

from pathlib2 import Path

executed in 3.63s, finished 19:16:23 2021-10-29

NumExpr defaulting to 8 threads.

Ввод [2]:

1

path = Path('../ ../data/2021\_evraz')

executed in 14ms, finished 19:16:24 2021-10-29

chronom

хронометраж - время начала и конца различных операций во время плавки

Ввод [17]:

1

chronom\_train = pd.read\_csv(path.joinpath('chronom\_train.csv'), index\_col=0)

2

chronom\_train

executed in 100ms, finished 19:39:00 2021-10-29

Out[17]:

	NPLV	TYPE_OPER	NOP	VR_NACH	VR_KON	O2
35171	510008	межпл.прост.	межпл. простой	2011-01-01 10:28:13	2011-01-01 10:32:15	NaN
35162	510008	межпл.прост.	межпл. простой	2021-01-01 03:01:07	2021-01-01 03:08:11	NaN
35163	510008	межпл.прост.	Осмотр конвертера	2021-01-01 03:01:19	2021-01-01 03:03:43	NaN
35164	510008	межпл.прост.	Наведение гарнисажа	2021-01-01 03:03:43	2021-01-01 03:05:23	NaN
35165	510008	межпл.прост.	Слив шлака	2021-01-01 03:05:23	2021-01-01 03:06:09	NaN
...	...	...	...	...	...	...
20943	512322	опер	Повалка	2021-04-26 18:31:26	2021-04-26 18:41:10	NaN
20944	512322	опер	Выпуск металла	2021-04-26 18:41:10	2021-04-26 18:47:40	NaN
20945	512322	опер	Отсечка шлака	2021-04-26 18:46:15	2021-04-26 18:46:15	NaN
20946	512322	опер	Прерыв. выпуска	2021-04-26 18:47:40	2021-04-26 18:48:40	NaN
20947	512322	опер	Слив шлака	2021-04-26 18:48:40	2021-04-26 18:48:40	NaN

34406 rows × 6 columns

Ввод [18]:

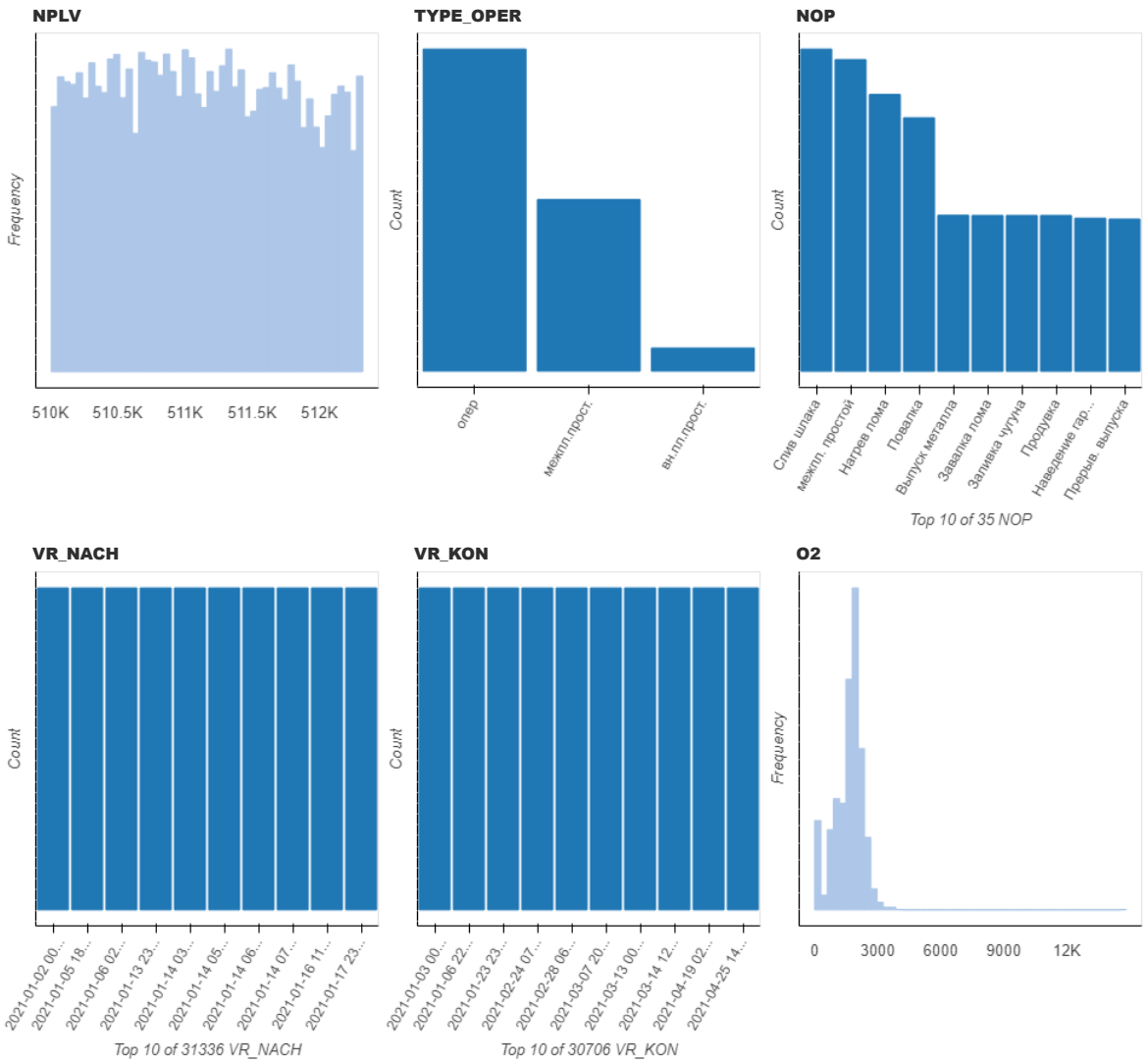
1 plot(chromom\_train)

executed in 1.01s, finished 19:39:05 2021-10-29

Out[18]:

Hide Stats and Insights

Dataset Statistics		Dataset Insights	
Number of Variables	6	O2 has 31490 (91.52%) missing values	Missing
Number of Rows	34406	O2 is skewed	Skewed
Missing Cells	31506	VR_NACH has a high cardinality: 31336 distinct values	High Cardinality
Missing Cells (%)	15.3%	VR_KON has a high cardinality: 30706 distinct values	High Cardinality
Duplicate Rows	0	VR_NACH has constant length 19	Constant Length
Duplicate Rows (%)	0.0%	VR_KON has constant length 19	Constant Length
Total Size in Memory	13.7 MB		
Average Row Size in Memory	418.8 B		
Variable Types	Numerical: 2 Categorical: 4		



Ввод [20]:

```
1 chromom_test = pd.read_csv(path.joinpath('chromom_test.csv'), index_col=0)
2 chromom_test
```

executed in 41ms, finished 19:39:24 2021-10-29

Out[20]:

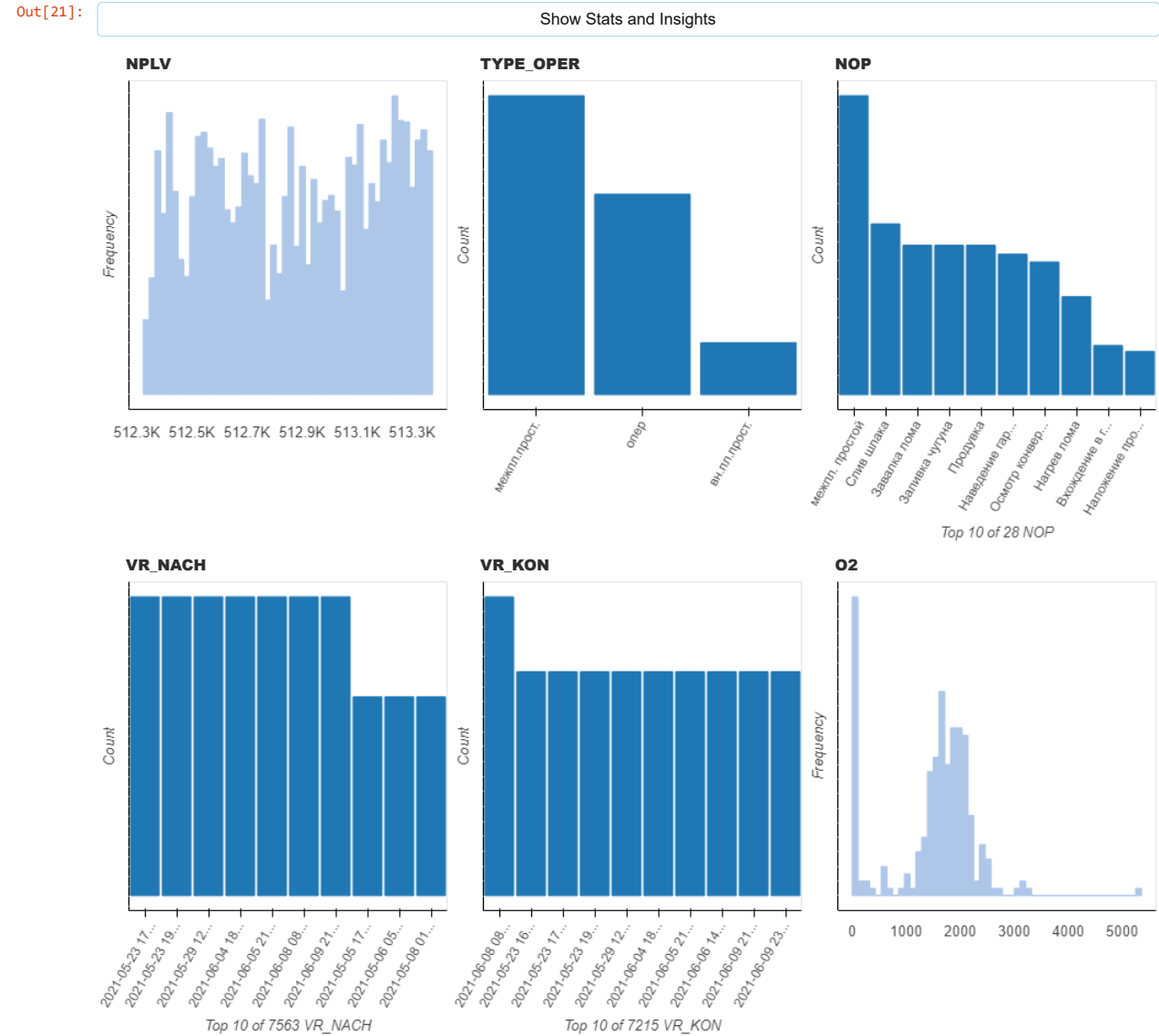
	NPLV	TYPE_OPER	NOP	VR_NACH	VR_KON	O2
0	512324	межпл.прост.	межпл. простой	2011-05-07 21:28:59	2011-05-07 21:33:38	NaN
1	512324	межпл.прост.	межпл. простой	2021-05-05 17:35:55	2021-05-05 17:41:21	NaN
2	512324	межпл.прост.	Осмотр конвертера	2021-05-05 17:36:07	2021-05-05 17:39:07	NaN
3	512324	межпл.прост.	Слив шлака	2021-05-05 17:39:07	2021-05-05 17:41:20	NaN
4	512324	опер	Завалка лома	2021-05-05 17:41:21	2021-05-05 17:43:16	NaN
...	...	...	...	...	...	...
7840	513374	опер	Завалка лома	2021-06-29 22:43:02	2021-06-29 22:43:54	NaN
7841	513374	опер	Нагрев лома	2021-06-29 22:43:54	2021-06-29 22:49:59	NaN
7842	513374	опер	Нагрев лома	2021-06-29 22:43:54	2021-06-29 22:49:59	1797.0
7843	513374	опер	Заливка чугуна	2021-06-29 22:50:07	2021-06-29 22:53:43	NaN
7844	513374	опер	Продувка	2021-06-29 22:53:43	2021-06-29 23:13:47	NaN

7845 rows x 6 columns

Ввод [21]:

```
1 plot(chromom_test)
```

executed in 555ms, finished 19:39:27 2021-10-29



chugun

Химический состав и характеристики чугуна

Ввод [9]:

```
1 chugun_train = pd.read_csv(path.joinpath('chugun_train.csv'))
2 chugun_train
```

executed in 37ms, finished 19:22:52 2021-10-29

Out[9]:

	NPLV	VES	T	SI	MN	S	P	CR	NI	CU	V	TI	DATA_ZAMERA
0	510008	263700.0	1396.0	0.44	0.22	0.023	0.097	0.03	0.01	0.03	0.103	0.084	2021-01-01 03:15:03
1	510009	264500.0	1419.0	0.68	0.20	0.017	0.087	0.02	0.01	0.03	0.084	0.096	2021-01-01 04:23:48
2	510010	263800.0	1384.0	0.56	0.26	0.017	0.096	0.03	0.01	0.03	0.115	0.110	2021-01-01 05:21:40
3	510011	264000.0	1401.0	0.48	0.27	0.018	0.091	0.03	0.01	0.02	0.112	0.110	2021-01-01 06:20:57
4	510012	263300.0	1422.0	0.47	0.23	0.018	0.096	0.02	0.01	0.03	0.083	0.070	2021-01-01 07:23:02
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2058	512318	267200.0	1415.0	0.38	0.28	0.019	0.099	0.02	0.01	0.02	0.081	0.060	2021-04-26 12:53:40
2059	512319	266800.0	1405.0	0.50	0.30	0.017	0.104	0.02	0.01	0.02	0.079	0.081	2021-04-26 14:21:57
2060	512320	276100.0	1398.0	0.61	0.31	0.025	0.115	0.03	0.01	0.03	0.086	0.066	2021-04-26 14:33:29
2061	512321	275800.0	1408.0	0.38	0.27	0.021	0.100	0.02	0.01	0.03	0.076	0.060	2021-04-26 15:50:53
2062	512322	279200.0	1413.0	0.50	0.31	0.019	0.105	0.02	0.01	0.03	0.077	0.067	2021-04-26 17:13:49

2063 rows x 13 columns

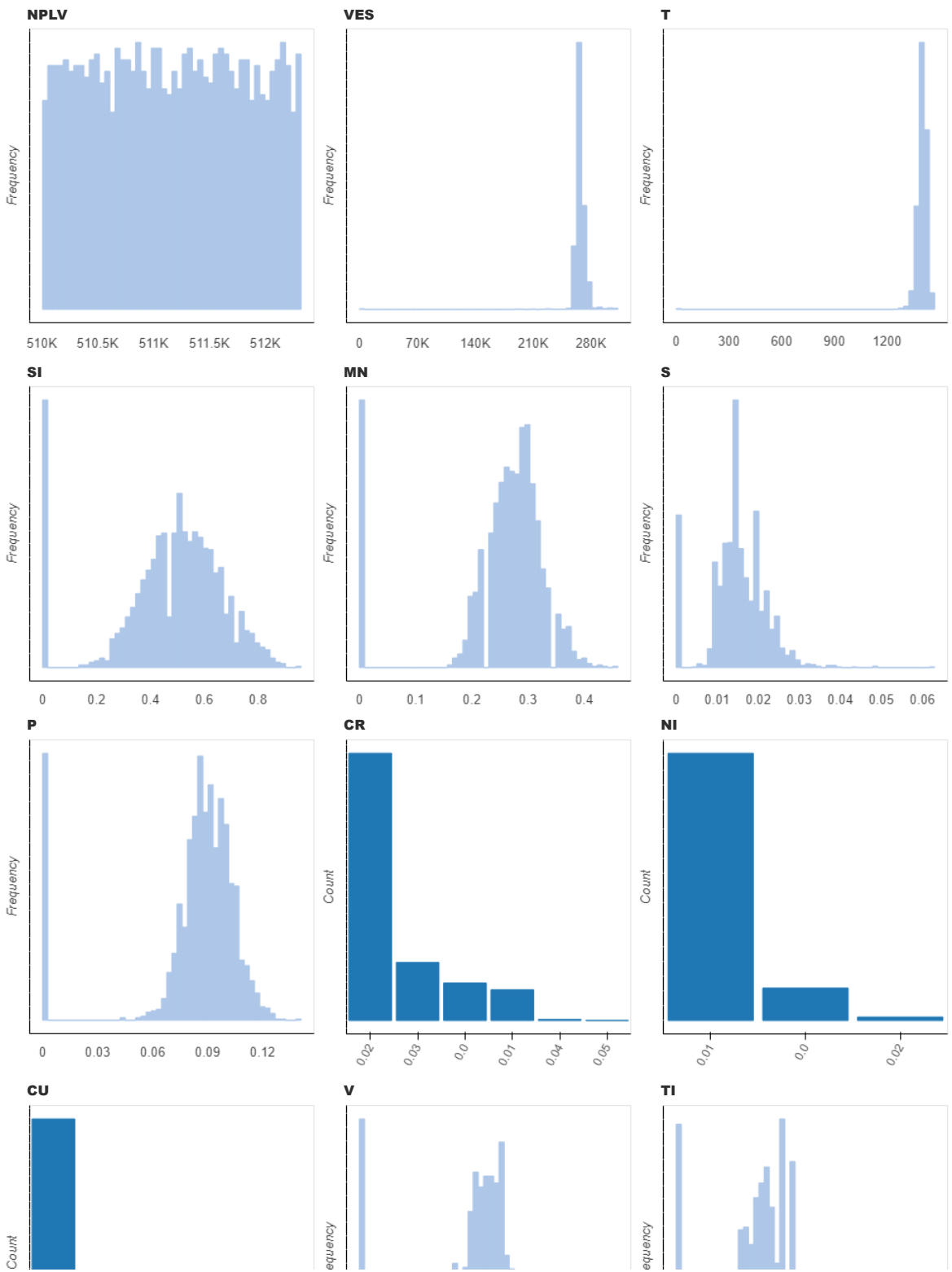
B80d [12]: 1 plot(chugun\_train)

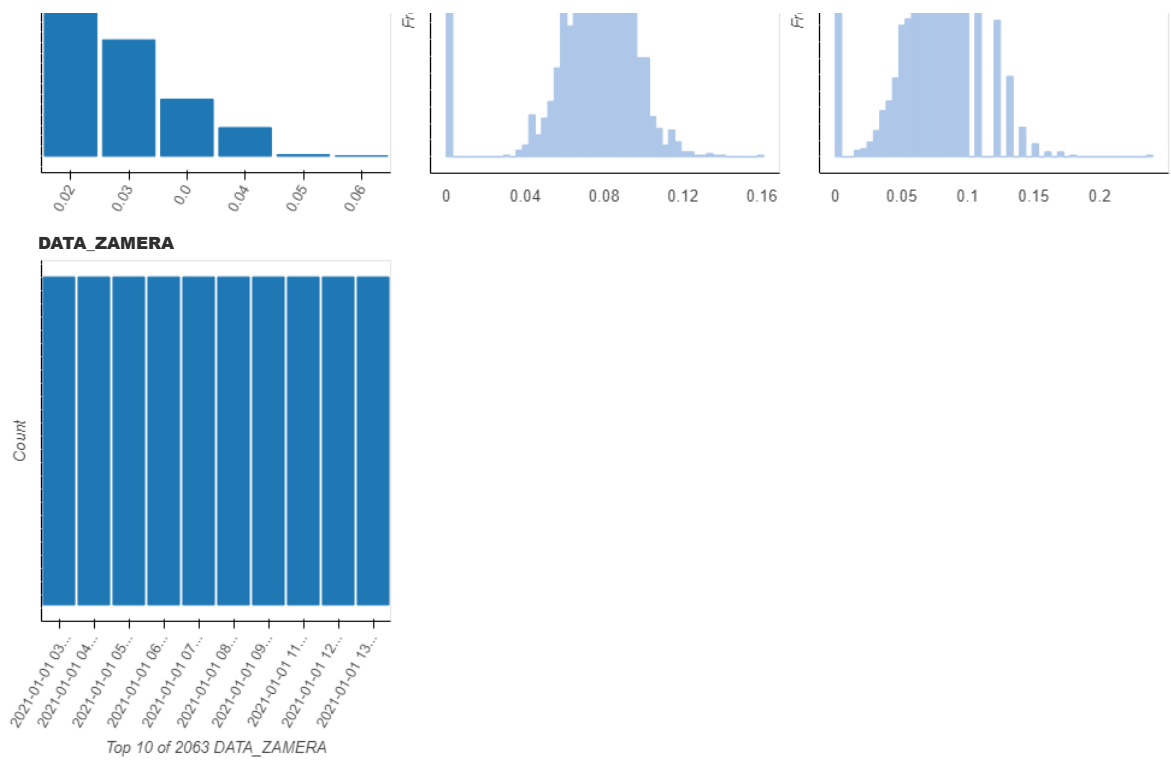
executed in 1.58s, finished 19:23:11 2021-10-29

Out[12]:

Hide Stats and Insights

Dataset Statistics		Dataset Insights	
Number of Variables	13	<b>NPLV</b> is uniformly distributed	Uniform
Number of Rows	2063	<b>VES</b> is skewed	Skewed
Missing Cells	0	<b>T</b> is skewed	Skewed
Missing Cells (%)	0.0%	<b>S</b> is skewed	Skewed
Duplicate Rows	0	<b>DATA_ZAMERA</b> has a high cardinality: 2063 distinct values	High Cardinality
Duplicate Rows (%)	0.0%	<b>DATA_ZAMERA</b> has constant length 19	Constant Length
Total Size in Memory	346.6 KB	<b>DATA_ZAMERA</b> has all distinct values	Unique
Average Row Size in Memory	172.1 B	<b>SI</b> has 195 (9.45%) zeros	Zeros
Variable Types	Numerical: 9 Categorical: 4	<b>MN</b> has 195 (9.45%) zeros	Zeros
		<b>S</b> has 195 (9.45%) zeros	Zeros





```
Bs0d [10]: 1 chugun_test = pd.read_csv(path.joinpath('chugun_test.csv'))
           2 chugun_test
```

executed in 34ms, finished 19:22:53 2021-10-29

Out[10]:

	NPLV	VES	T	SI	MN	S	P	CR	NI	CU	V	TI	DATA_ZAMERA
0	512324	240100.0	1355.0	0.46	0.33	0.027	0.079	0.01	0.01	0.02	0.048	0.030	2021-05-05 17:22:54
1	512327	266400.0	1390.0	0.30	0.33	0.032	0.099	0.01	0.00	0.00	0.050	0.024	2021-05-05 20:00:30
2	512328	270200.0	1373.0	0.00	0.00	0.000	0.000	0.00	0.00	0.00	0.000	0.000	2021-05-05 22:29:21
3	512331	266700.0	1383.0	0.54	0.39	0.028	0.115	0.02	0.01	0.03	0.059	0.042	2021-05-06 01:58:11
4	512333	267400.0	1387.0	0.00	0.00	0.000	0.000	0.00	0.00	0.00	0.000	0.000	2021-05-06 05:27:08
...	...	...	...	...	...	...	...	...	...	...	...	...	...
775	513369	280800.0	1324.0	0.35	0.29	0.035	0.096	0.01	0.01	0.02	0.080	0.062	2021-06-29 16:51:30
776	513370	280600.0	1389.0	0.39	0.31	0.024	0.097	0.01	0.01	0.02	0.093	0.074	2021-06-29 17:35:57
777	513371	279700.0	1332.0	0.30	0.27	0.031	0.099	0.01	0.01	0.02	0.066	0.036	2021-06-29 18:52:34
778	513372	279700.0	1378.0	0.43	0.33	0.026	0.113	0.01	0.01	0.02	0.095	0.070	2021-06-29 19:42:11
779	513374	270600.0	1396.0	0.36	0.33	0.026	0.104	0.02	0.01	0.02	0.092	0.052	2021-06-29 22:06:40

780 rows × 13 columns

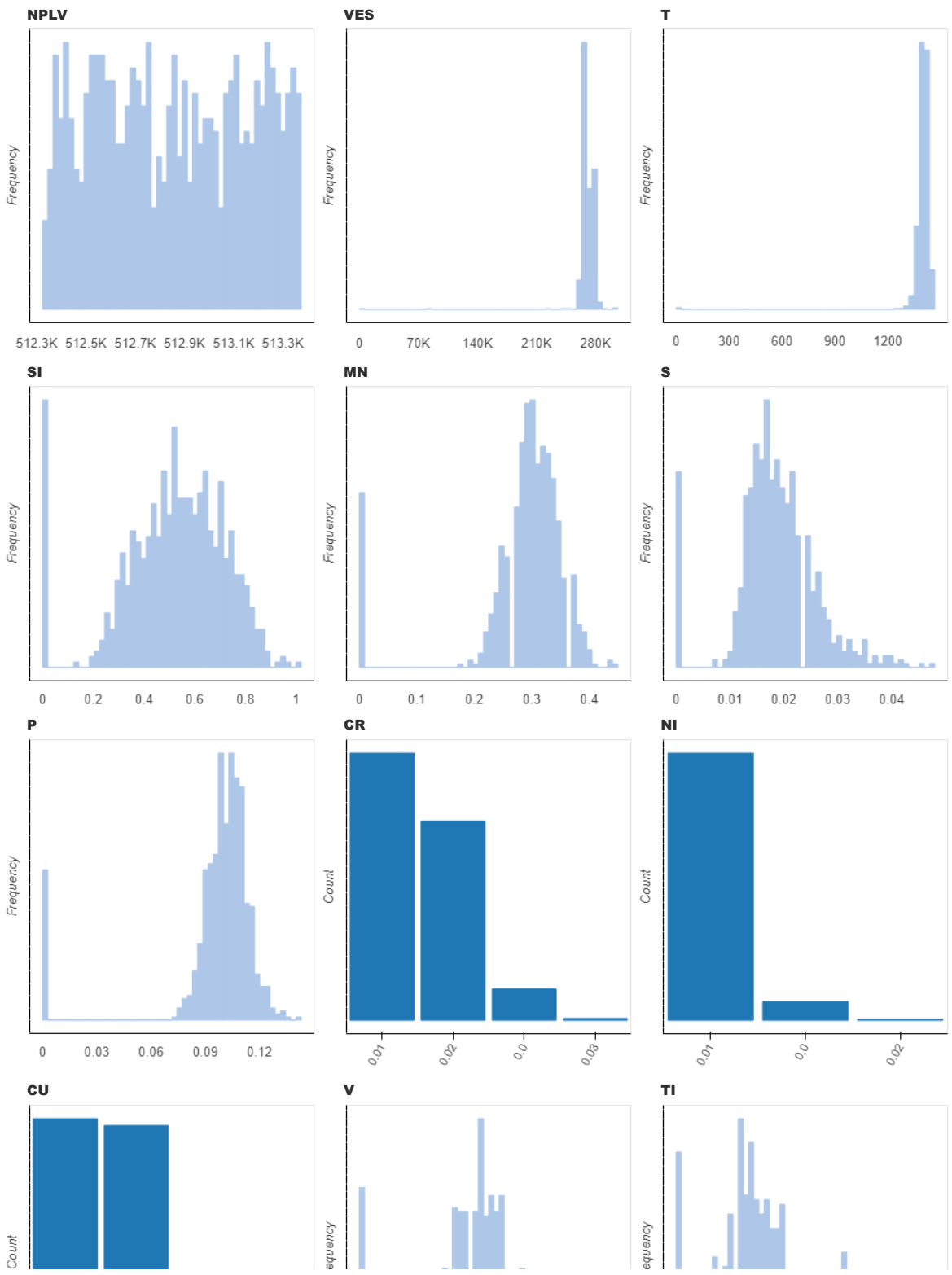
B80d [11]: 1 plot(chugun\_test)

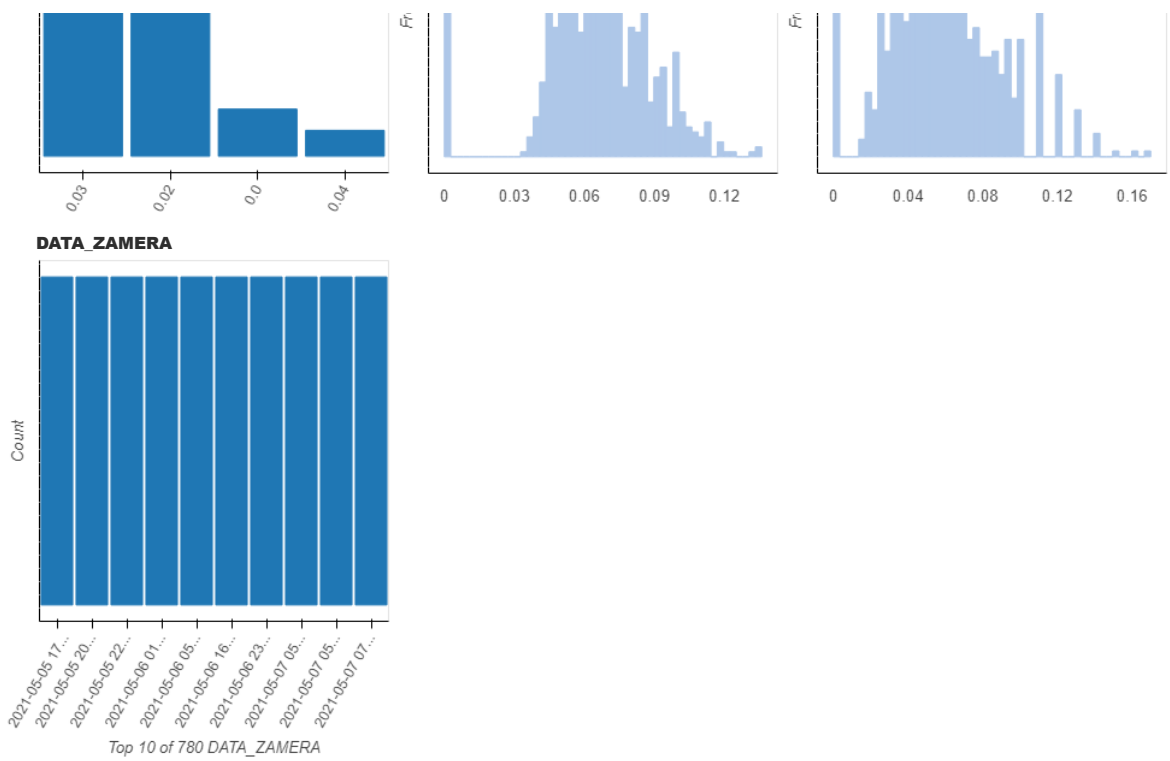
executed in 1.58s, finished 19:22:56 2021-10-29

Out[11]:

Hide Stats and Insights

Dataset Statistics		Dataset Insights	
Number of Variables	13	<b>VES</b> is skewed	Skewed
Number of Rows	780	<b>T</b> is skewed	Skewed
Missing Cells	0	<b>DATA_ZAMERA</b> has a high cardinality: 780 distinct values	High Cardinality
Missing Cells (%)	0.0%	<b>DATA_ZAMERA</b> has constant length 19	Constant Length
Duplicate Rows	0	<b>DATA_ZAMERA</b> has all distinct values	Unique
Duplicate Rows (%)	0.0%	<b>SI</b> has 49 (6.28%) zeros	Zeros
Total Size in Memory	131.1 KB	<b>MN</b> has 49 (6.28%) zeros	Zeros
Average Row Size in Memory	172.2 B	<b>S</b> has 49 (6.28%) zeros	Zeros
Variable Types	Numerical: 9 Categorical: 4	<b>P</b> has 49 (6.28%) zeros	Zeros
		<b>V</b> has 49 (6.28%) zeros	Zeros





## gas

Информация об анализе отходящих газов

Ввод [13]:

```
1 gas_train = pd.read_csv(path.joinpath('gas_train.csv'))
2 gas_train
```

executed in 25.5s, finished 19:23:51 2021-10-29

Out[13]:

	NPLV	Time	V	T	O2	N2	H2	CO2	CO	AR	Т фурмы 1	Т фурмы 2	O2_pressure
0	510008	2021-01-01 03:08:11.437	218263.343750	262.847229	18.722993	80.132247	0.087755	0.163878	9.229025e-03	0.893243	0.000000	0.000000	13.085938
1	510008	2021-01-01 03:08:12.437	218263.343750	262.847229	18.732721	80.138406	0.087959	0.148980	8.390023e-03	0.892948	0.000000	0.000000	13.085938
2	510008	2021-01-01 03:08:13.437	218369.359375	262.152771	18.742449	80.144565	0.088163	0.134082	7.551021e-03	0.892653	0.000000	0.000000	13.085938
3	510008	2021-01-01 03:08:14.437	218475.359375	261.805573	18.752177	80.150724	0.088367	0.119184	6.712018e-03	0.892358	0.000000	0.000000	13.093172
4	510008	2021-01-01 03:08:15.437	218369.359375	260.763885	18.761905	80.156883	0.088571	0.104286	5.873016e-03	0.892063	0.000000	0.000000	13.093172
...	...	...	...	...	...	...	...	...	...	...	...	...	...
6468013	512322	2021-04-26 18:48:35.437	207386.734375	118.402779	1.890000	97.220001	0.090000	0.050000	0.000000e+00	0.750000	23.900463	27.054343	14.424190
6468014	512322	2021-04-26 18:48:36.437	207051.609375	118.402779	2.710000	96.410004	0.080000	0.040000	0.000000e+00	0.770000	23.900463	27.054342	14.424190
6468015	512322	2021-04-26 18:48:37.437	207163.375000	118.402779	3.770000	95.349998	0.090000	0.030000	0.000000e+00	0.770000	23.900463	27.054341	14.438658
6468016	512322	2021-04-26 18:48:38.437	206827.906250	118.750000	3.400000	95.720001	0.090000	0.040000	0.000000e+00	0.750000	23.900463	27.054340	14.445890
6468017	512322	2021-04-26 18:48:39.437	206042.953125	118.750000	2.350000	96.760002	0.090000	0.030000	5.594952e-07	0.770000	23.900463	27.054339	14.445890

6468018 rows × 13 columns



Ввод [14]:

1 plot(gas\_train)

executed in 1m 51.4s, finished 19:25:42 2021-10-29

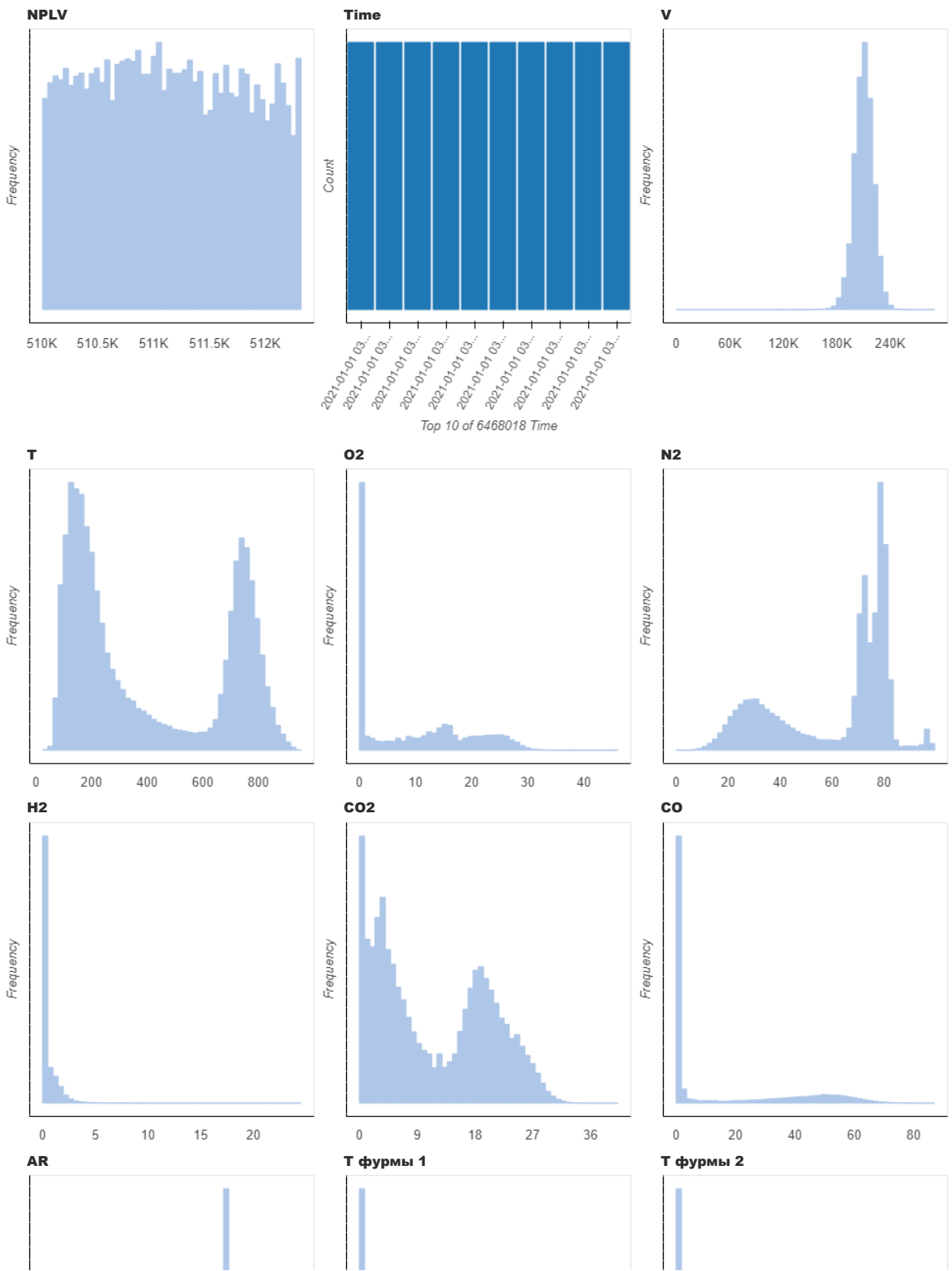
Out[14]:

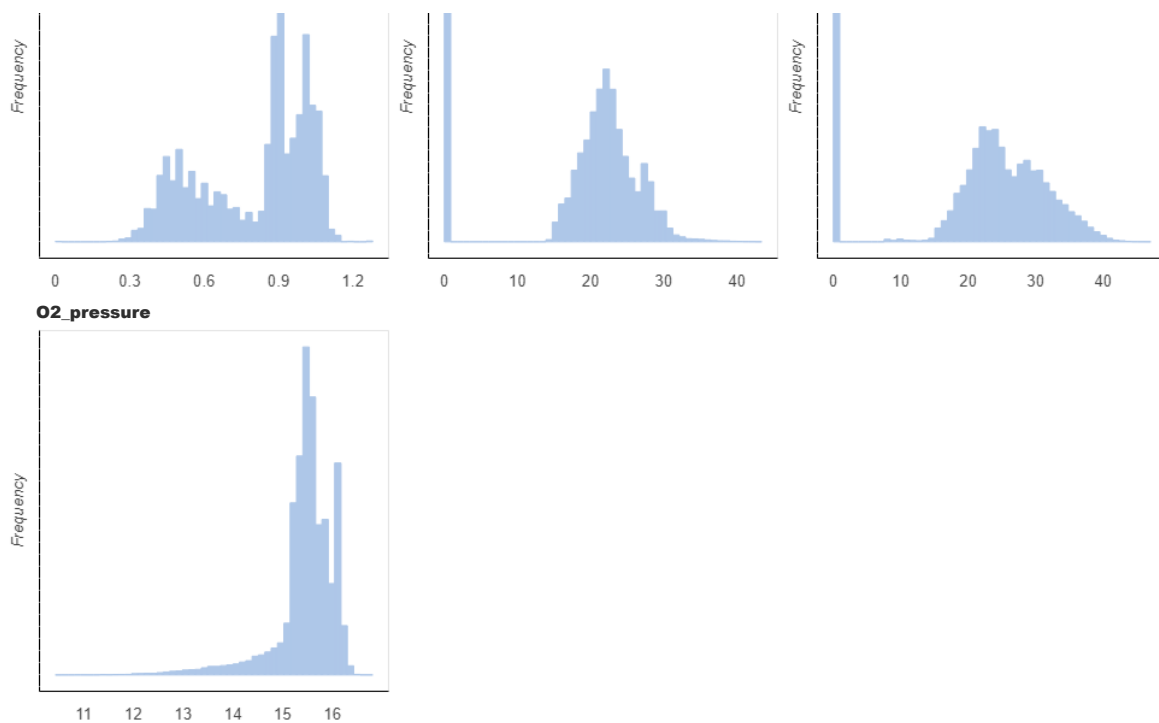
Hide Stats and Insights

Dataset Statistics			Dataset Insights	
Number of Variables	13	<span>V</span> is skewed	Skewed	
Number of Rows	6.468×10 <sup>06</sup>	<span>O2</span> is skewed	Skewed	
Missing Cells	0	<span>N2</span> is skewed	Skewed	
Missing Cells (%)	0.0%	<span>H2</span> is skewed	Skewed	
Duplicate Rows	0	<span>CO</span> is skewed	Skewed	
Duplicate Rows (%)	0.0%	<span>AR</span> is skewed	Skewed	
Total Size in Memory	1.1 GB	<span>T фурмы 1</span> is skewed	Skewed	
Average Row Size in Memory	176.0 B	<span>T фурмы 2</span> is skewed	Skewed	
Variable Types	Numerical: 12 Categorical: 1	<span>O2_pressure</span> is skewed	Skewed	
			<span>Time</span> has a high cardinality: 6468018 distinct values	
			High Cardinality	

1

2





```
Ввод [15]: 1 gas_test = pd.read_csv(path.joinpath('gas_test.csv'))
           2 gas_test
```

executed in 9.06s, finished 19:25:51 2021-10-29

Out[15]:

	NPLV	Time	V	T	O2	N2	H2	CO2	CO	AR	Т фурмы 1	Т фурмы 2	O2_pressure
0	512324	2021-05-05 17:41:21.437	213979.046875	137.326396	18.671576	79.388503	0.153247	0.591429	0.216998	0.980324	26.244215	26.673637	16.124132
1	512324	2021-05-05 17:41:22.437	214303.343750	137.413201	18.678876	79.386573	0.153680	0.585952	0.216933	0.980043	26.244215	26.673768	16.124132
2	512324	2021-05-05 17:41:23.437	214842.750000	137.500007	18.686176	79.384642	0.154113	0.580476	0.216868	0.979762	26.244215	26.673900	16.124132
3	512324	2021-05-05 17:41:24.437	214842.750000	137.586812	18.693477	79.382712	0.154545	0.575000	0.216803	0.979482	26.244215	26.674031	16.124132
4	512324	2021-05-05 17:41:25.437	215273.281250	137.673618	18.700777	79.380781	0.154978	0.569524	0.216739	0.979201	26.244215	26.674163	16.124132
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2515863	513374	2021-06-29 23:29:03.437	232936.171875	141.319443	5.410000	81.769997	0.050000	11.600000	0.000000	1.160000	30.005785	30.295139	16.059027
2515864	513374	2021-06-29 23:29:04.437	232737.328125	141.319443	5.410000	81.769997	0.050000	11.600000	0.000000	1.160000	30.005785	30.295139	16.059027
2515865	513374	2021-06-29 23:29:05.437	232239.515625	142.013885	5.410000	81.769997	0.050000	11.600000	0.000000	1.160000	30.005785	30.295139	16.066261
2515866	513374	2021-06-29 23:29:06.437	232139.796875	142.361115	5.410000	81.769997	0.050000	11.600000	0.000000	1.160000	30.005785	30.295139	16.066261
2515867	513374	2021-06-29 23:29:07.437	231340.687500	142.361115	5.410000	81.769997	0.050000	11.600000	0.000000	1.160000	30.005785	30.295139	16.073496

2515868 rows × 13 columns

Ввод [16]:

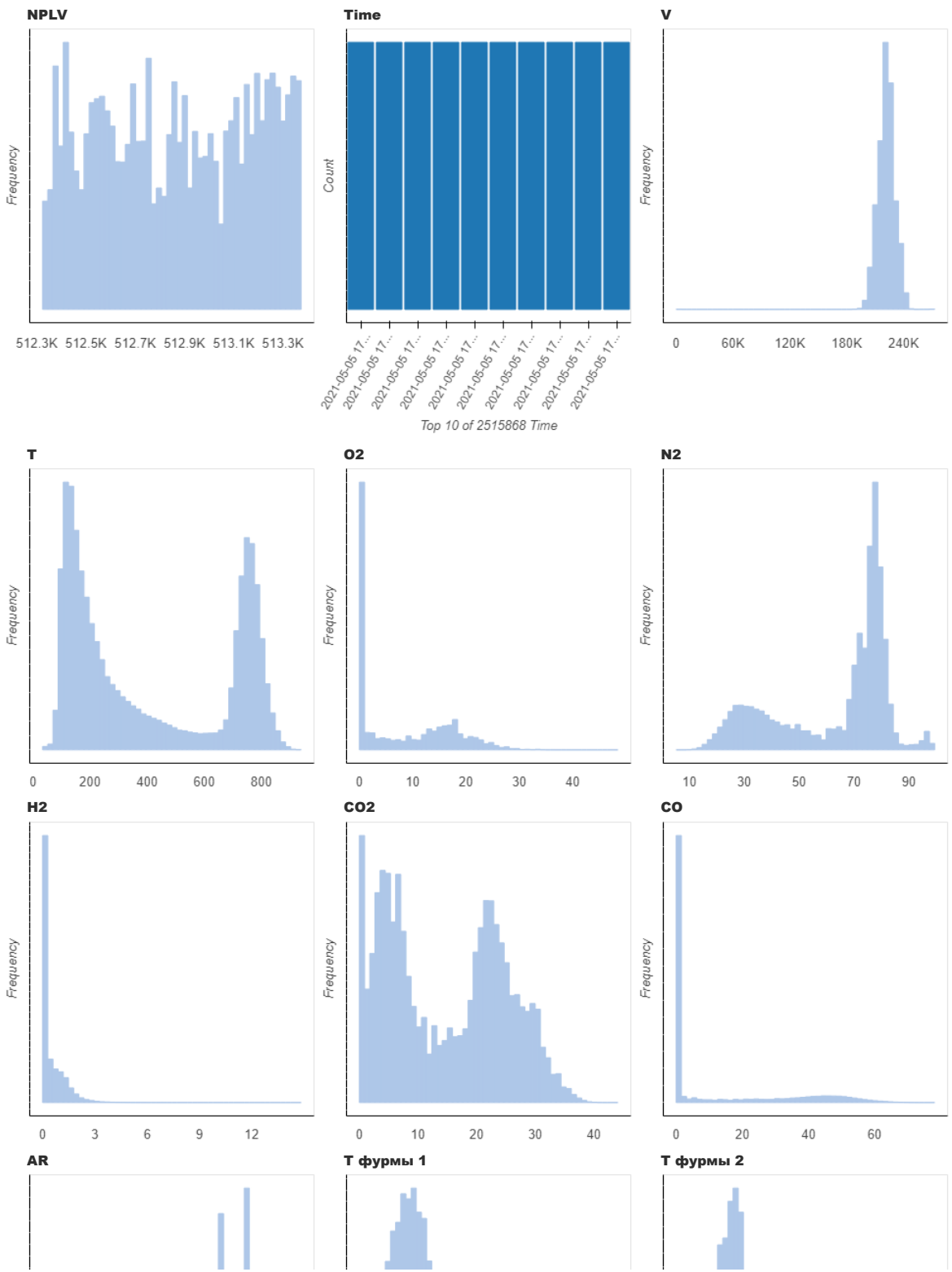
1 plot(gas\_test)

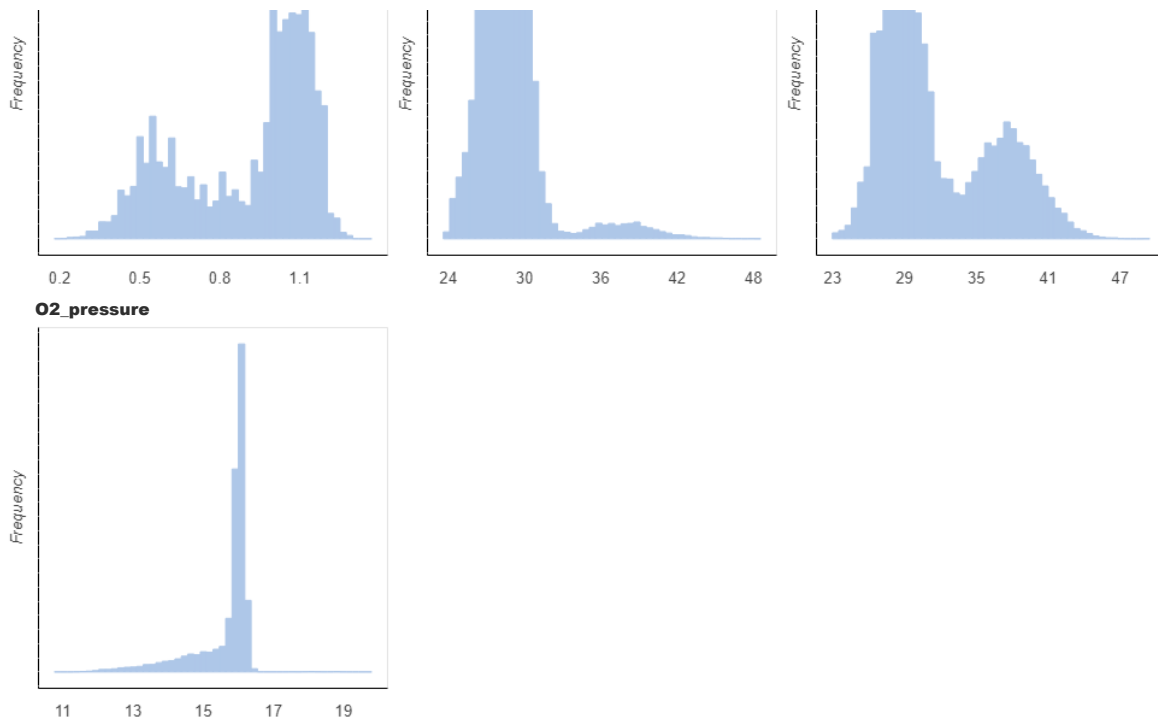
executed in 56.9s, finished 19:26:48 2021-10-29

Out[16]:

Hide Stats and Insights

Dataset Statistics			Dataset Insights	
Number of Variables	13	<span>V</span> is skewed	Skewed	
Number of Rows	2.5159×10 <sup>06</sup>	<span>O2</span> is skewed	Skewed	
Missing Cells	0	<span>N2</span> is skewed	Skewed	
Missing Cells (%)	0.0%	<span>H2</span> is skewed	Skewed	
Duplicate Rows	0	<span>CO</span> is skewed	Skewed	
Duplicate Rows (%)	0.0%	<span>O2_pressure</span> is skewed	Skewed	
Total Size in Memory	422.3 MB	<span>Time</span> has a high cardinality: 2515868 distinct values	High Cardinality	
Average Row Size in Memory	176.0 B	<span>Time</span> has constant length 23	Constant Length	
Variable Types	Numerical: 12 Categorical: 1	<span>Time</span> has all distinct values	Unique	
		<span>O2</span> has 895410 (35.59%) zeros	Zeros	





## lom

Вместе с чугуном в фурму засыпают лом - это часть технологического процесса. Таблица содержит вес и тип ломов, использованных в каждой плавке

```
Ввод [22]: 1 lom_train = pd.read_csv(path.joinpath('lom_train.csv'))
           2 lom_train
```

executed in 34ms, finished 19:42:55 2021-10-29

Out[22]:

	NPLV	VDL	NML	VES
0	510008	4	K	56500
1	510008	8	O	16700
2	510008	13	КП	3000
3	510009	4	K	49800
4	510009	8	O	22800
...	...	...	...	...
6371	512321	8	O	4200
6372	512321	13	КП	3000
6373	512322	4	K	66700
6374	512322	13	КП	6000
6375	512322	48	У2КШ	3400

6376 rows x 4 columns

Ввод [24]:

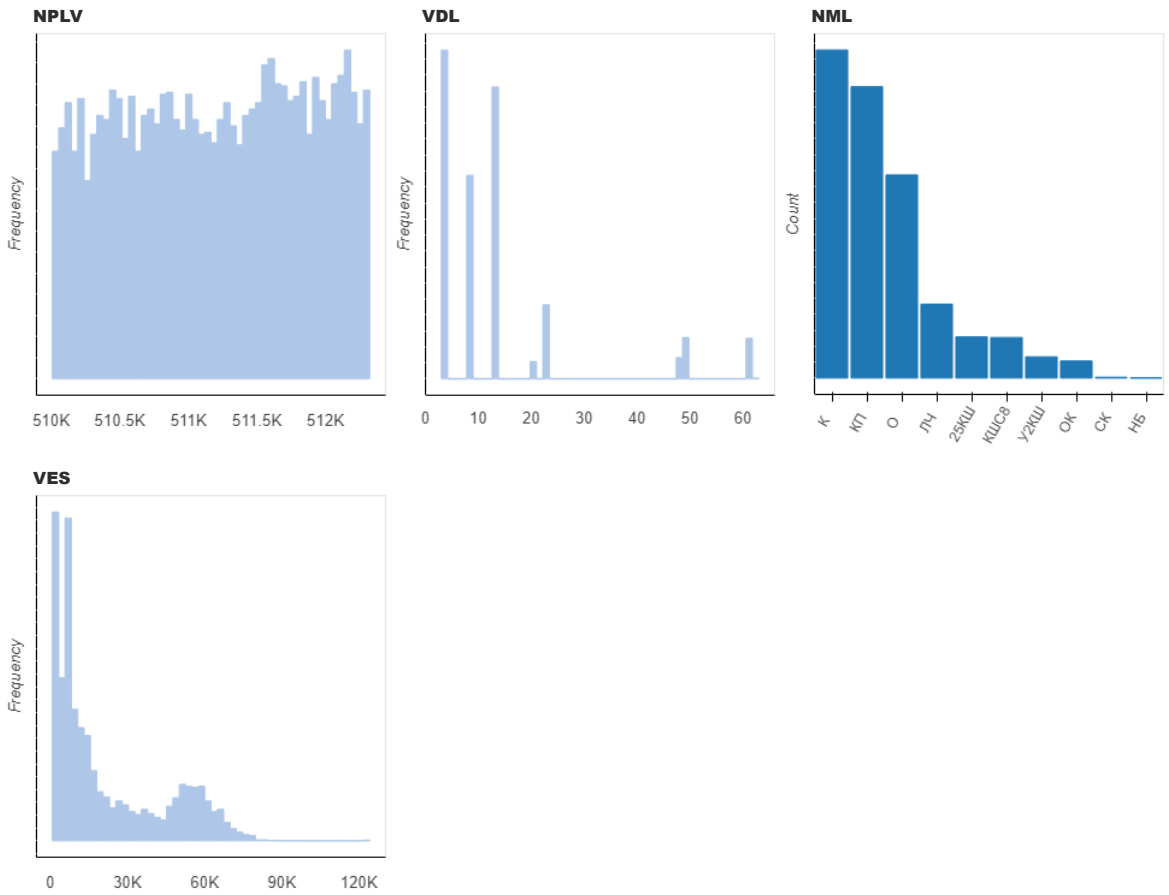
1 plot(lom\_train)

executed in 428ms, finished 19:43:22 2021-10-29

Out[24]:

Hide Stats and Insights

Dataset Statistics		Dataset Insights	
Number of Variables	4	VDL is skewed	Skewed
Number of Rows	6376	VES is skewed	Skewed
Missing Cells	0	NML has constant length 4	Constant Length
Missing Cells (%)	0.0%		
Duplicate Rows	0		
Duplicate Rows (%)	0.0%		
Total Size in Memory	710.0 KB		
Average Row Size in Memory	114.0 B		
Variable Types	Numerical: 3 Categorical: 1		



Ввод [25]:

1 lom\_test = pd.read\_csv(path.joinpath('lom\_test.csv'))  
2 lom\_test

executed in 20ms, finished 19:43:28 2021-10-29

Out[25]:

	NPLV	VDL	NML	VES
0	512324	4	K	26040
1	512324	8	O	17660
2	512324	13	КП	2000
3	512327	8	O	10600
4	512327	23	ЛЧ	8000
...	...	...	...	...
2627	513372	48	У2КУ	500
2628	513374	4	K	54400
2629	513374	13	КП	7000
2630	513374	8	O	9100
2631	513374	23	ЛЧ	2000

2632 rows × 4 columns

Ввод [26]:

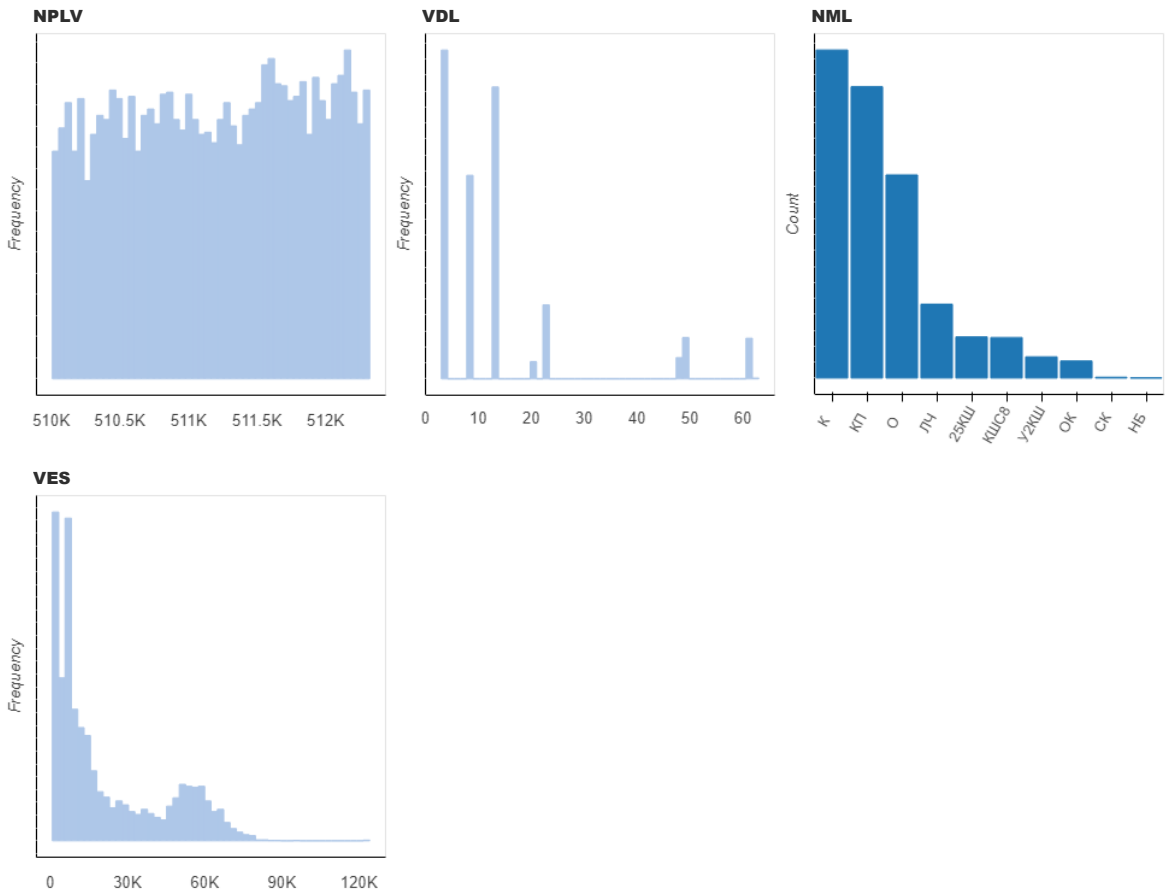
1 plot(lom\_train)

executed in 456ms, finished 19:43:37 2021-10-29

Out[26]:

Hide Stats and Insights

Dataset Statistics		Dataset Insights	
Number of Variables	4	<b>VDL</b> is skewed	Skewed
Number of Rows	6376	<b>VES</b> is skewed	Skewed
Missing Cells	0	<b>NML</b> has constant length 4	Constant Length
Missing Cells (%)	0.0%		
Duplicate Rows	0		
Duplicate Rows (%)	0.0%		
Total Size in Memory	750.7 KB		
Average Row Size in Memory	120.6 B		
Variable Types	Numerical: 3 Categorical: 1		



**plavki**

Основная информация по плавке - характеристики плавки (марка металла, направление разливки) и оборудования

Ввод [27]:

```
1 plavki_train = pd.read_csv(path.joinpath('plavki_train.csv'))
2 plavki_train
```

executed in 27ms, finished 19:43:43 2021-10-29

Out[27]:

	NPLV	plavka_VR_NACH	plavka_VR_KON	plavka_NMZ	plavka_NAPR_ZAD	plavka_STFUT	plavka_TIPE_FUR	plavka_ST_FURM	plavka_TIPE_GOL	plavka_ST_GOL
0	510008	2021-01-01 03:08:11	2021-01-01 03:51:10	C255	МНЛЗ	971	цилиндрическая	11	5 сопловая	11
1	510009	2021-01-01 04:00:44	2021-01-01 05:07:28	C255	МНЛЗ	972	цилиндрическая	12	5 сопловая	12
2	510010	2021-01-01 05:12:29	2021-01-01 06:00:53	Ст3пс/Э	Изп	973	цилиндрическая	13	5 сопловая	13
3	510011	2021-01-01 06:13:48	2021-01-01 07:08:39	Св-08A.z02	Изп	974	цилиндрическая	14	5 сопловая	14
4	510012	2021-01-01 07:13:44	2021-01-01 08:01:59	SC2M/ЭТ	МНЛС	975	цилиндрическая	15	5 сопловая	15
...	...	...	...	...	...	...	...	...	...	...
2132	512318	2021-04-26 13:04:26	2021-04-26 13:55:50	C071TM.z01/ЭТ	МНЛС	3281	коническая	22	5 сопловая	56
2133	512319	2021-04-26 14:10:20	2021-04-26 15:14:23	C071TM.z01/ЭТ	МНЛС	3282	коническая	23	5 сопловая	57
2134	512320	2021-04-26 15:21:37	2021-04-26 16:16:42	40X.1	МНЛЗ	3283	коническая	24	5 сопловая	58
2135	512321	2021-04-26 16:22:37	2021-04-26 17:23:37	40X.1	МНЛЗ	3284	коническая	25	5 сопловая	59
2136	512322	2021-04-26 17:28:00	2021-04-26 18:48:40	40X.1	МНЛЗ	3285	коническая	26	5 сопловая	60

2137 rows x 10 columns

Ввод [28]:

1 plot(plavki\_train)

executed in 1.09s, finished 19:43:58 2021-10-29

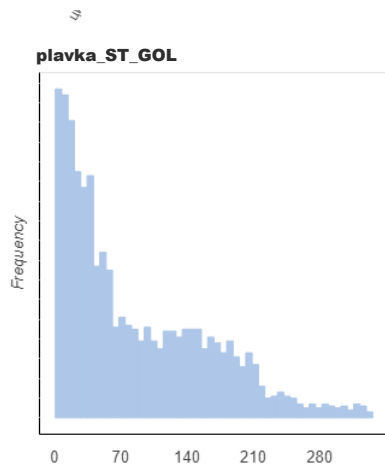
Out[28]:

Hide Stats and Insights

Dataset Statistics			Dataset Insights
Number of Variables	10	<code>NPLV</code> is skewed	Skewed
Number of Rows	2137	<code>plavka_STFUT</code> is skewed	Skewed
Missing Cells	0	<code>plavka_ST_FU...</code> is skewed	Skewed
Missing Cells (%)	0.0%	<code>plavka_VR_NA...</code> has a high cardinality: 2063 distinct values	High Cardinality
Duplicate Rows	0	<code>plavka_VR_KON</code> has a high cardinality: 2063 distinct values	High Cardinality
Duplicate Rows (%)	0.0%	<code>plavka_NMZ</code> has a high cardinality: 66 distinct values	High Cardinality
Total Size in Memory	1.3 MB	<code>plavka_VR_NA...</code> has constant length 19	Constant Length
Average Row Size in Memory	616.8 B	<code>plavka_VR_KON</code> has constant length 19	Constant Length
Variable Types	Numerical: 4 Categorical: 6	<code>plavka_NMZ</code> has constant length 15	Constant Length
		<code>plavka_TIPE_...</code> has constant length 20	Constant Length







Ввод [29]:

```
1 plavki_test = pd.read_csv(path.joinpath('plavki_test.csv'))
2 plavki_test
```

executed in 24ms, finished 19:44:03 2021-10-29

Out[29]:

	NPLV	plavka_VR_NACH	plavka_VR_KON	plavka_NMZ	plavka_NAPR_ZAD	plavka_STFUT	plavka_TIPE_FUR	plavka_ST_FURM	plavka_TIPE_GOL	plavka_ST_GOL
0	512324	2021-05-05 17:41:21	2021-05-05 18:46:07	Ст3пс/Э	Иэл	2	цилиндрическая	2	4-conn x54	2
1	512327	2021-05-05 21:07:01	2021-05-05 22:02:26	C121TM/ЭТ	МНЛС	5	цилиндрическая	5	4-conn x54	5
2	512328	2021-05-05 22:03:12	2021-05-05 23:39:32	Ст3пс/Э	Иэл	6	цилиндрическая	6	4-conn x54	6
3	512331	2021-05-06 02:21:46	2021-05-06 03:49:18	25Г2С	Иэл	9	цилиндрическая	9	4-conn x54	9
4	512333	2021-05-06 05:22:03	2021-05-06 06:25:51	C091TM.z01/ ЭТ	МНЛС	11	цилиндрическая	11	4-conn x54	11
...	...	...	...	...	...	...	...	...	...	...
775	513369	2021-06-29 17:08:33	2021-06-29 17:55:50	Ст3сп/Т	МНЛЗ	1047	цилиндрическая	2	601-5	2
776	513370	2021-06-29 18:03:00	2021-06-29 18:49:50	Ст3сп/Т	МНЛЗ	1048	цилиндрическая	3	601-5	3
777	513371	2021-06-29 18:55:21	2021-06-29 19:53:39	Ст3сп/Т	МНЛЗ	1049	цилиндрическая	4	601-5	4
778	513372	2021-06-29 20:18:40	2021-06-29 20:59:18	Ст3сп/Т	МНЛЗ	1050	цилиндрическая	5	601-5	5
779	513374	2021-06-29 22:43:02	2021-06-29 23:29:10	Ст3сп/Т	МНЛЗ	1052	цилиндрическая	7	601-5	7

780 rows × 10 columns

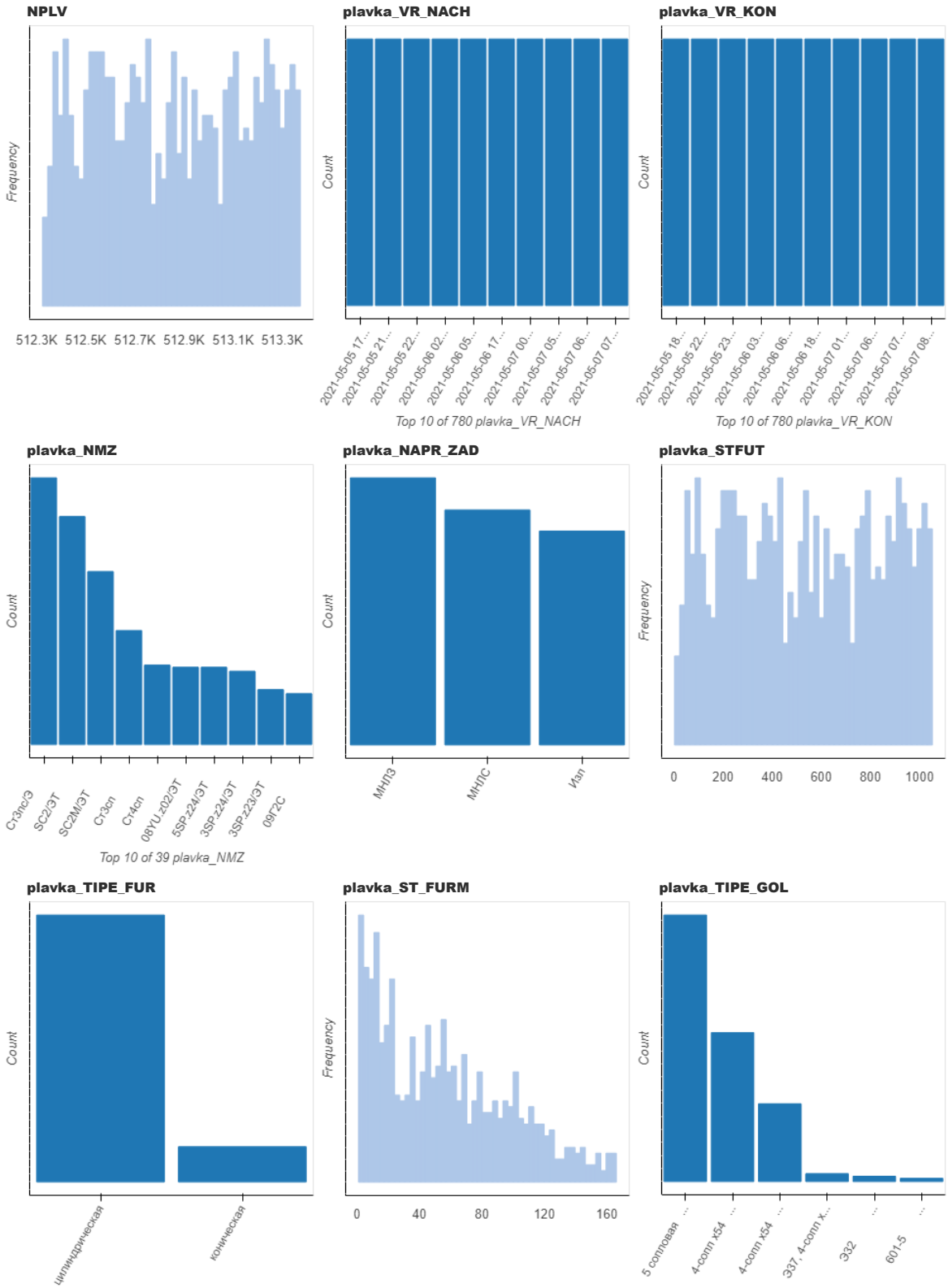
```
Ввод [30]: 1 plot(plavki_test)
```

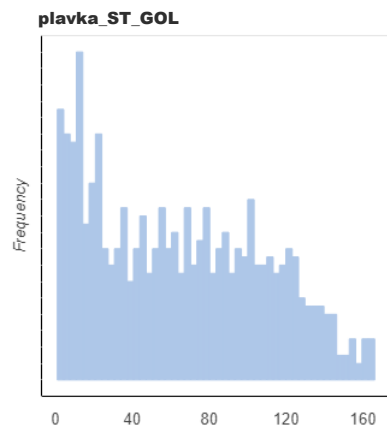
executed in 1.42s, finished 19:44:13 2021-10-29

Out[30]:

Hide Stats and Insights

Dataset Statistics		Dataset Insights	
Number of Variables	10	plavka_VR_NA_ has a high cardinality: 780 distinct values	High Cardinality
Number of Rows	780	plavka_VR_KON has a high cardinality: 780 distinct values	High Cardinality
Missing Cells	0	plavka_VR_NA_ has constant length 19	Constant Length
Missing Cells (%)	0.0%	plavka_VR_KON has constant length 19	Constant Length
Duplicate Rows	0	plavka_NMZ has constant length 15	Constant Length
Duplicate Rows (%)	0.0%	plavka_TIPE_ has constant length 20	Constant Length
Total Size in Memory	469.4 KB	plavka_VR_NA_ has all distinct values	Unique
Average Row Size in Memory	616.3 B	plavka_VR_KON has all distinct values	Unique
Variable Types	Numerical: 4 Categorical: 6		





## produv

Таблица содержит основные параметры продувки - мгновенный расход кислорода и положение (наклон) фурмы

```
Ввод [31]: 1 produv_train = pd.read_csv(path.joinpath('produv_train.csv'))
           2 produv_train
```

executed in 5.30s, finished 19:44:24 2021-10-29

Out[31]:

	NPLV	SEC	RAS	POL
0	510008	2021-01-01 03:18:26	382.000000	3.920000
1	510008	2021-01-01 03:18:28	382.000000	3.920000
2	510008	2021-01-01 03:18:30	553.000000	3.920000
3	510008	2021-01-01 03:18:32	701.000000	3.920000
4	510008	2021-01-01 03:18:34	813.000000	3.920000
...	...	...	...	...
4729797	512322	2021-05-05 16:30:46	363.996249	4.850045
4729798	512322	2021-05-05 16:30:48	363.996999	4.850036
4729799	512322	2021-05-05 16:30:50	363.997749	4.850027
4729800	512322	2021-05-05 16:30:52	363.998499	4.850018
4729801	512322	2021-05-05 16:30:54	363.999250	4.850009

4729802 rows × 4 columns

Ввод [32]:

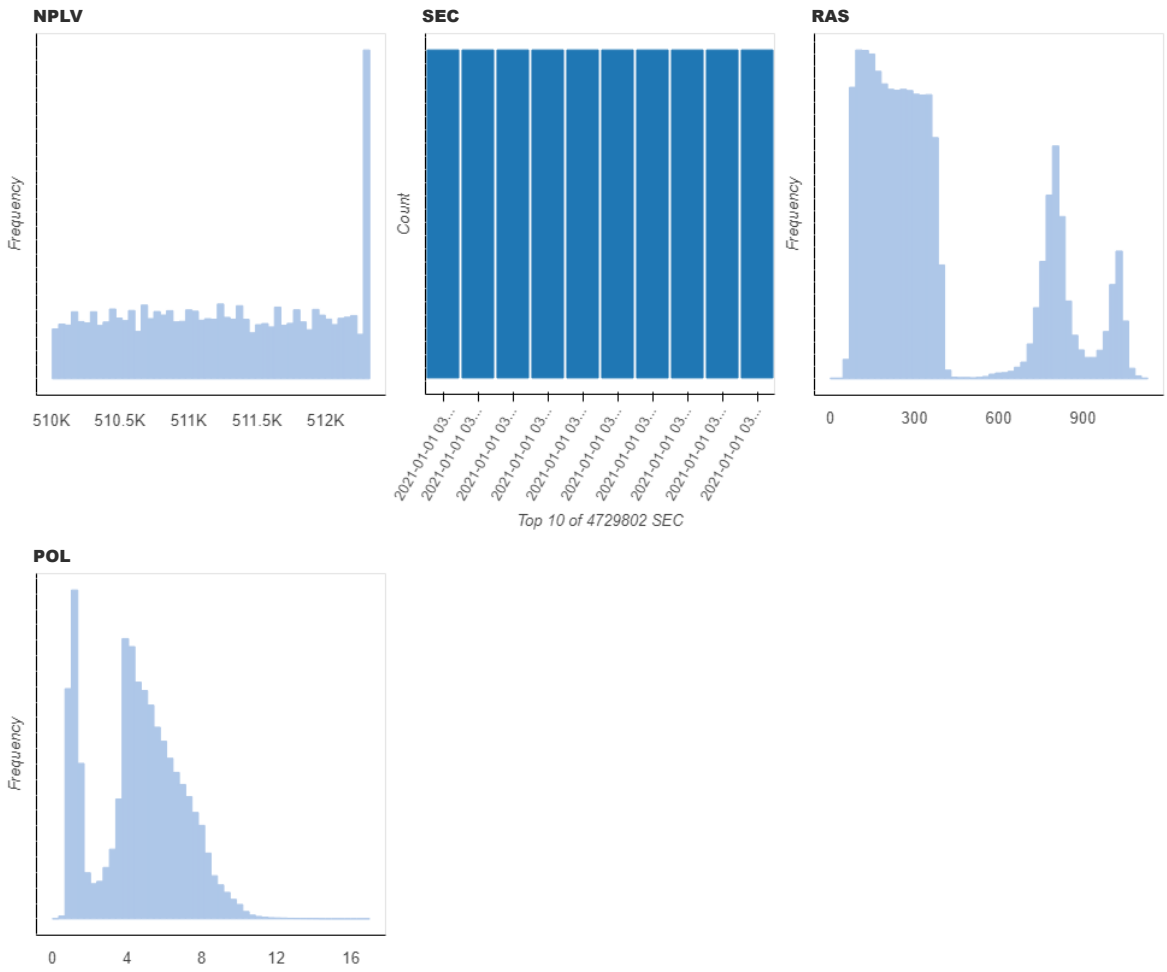
```
1 plot(produv_train)
```

executed in 53.2s, finished 19:45:24 2021-10-29

Out[32]:

Hide Stats and Insights

Dataset Statistics		Dataset Insights	
Number of Variables	4	NPLV is skewed	Skewed
Number of Rows	4.7298×10 <sup>06</sup>	SEC has a high cardinality: 4729802 distinct values	High Cardinality
Missing Cells	0	SEC has constant length 19	Constant Length
Missing Cells (%)	0.0%	SEC has all distinct values	Unique
Duplicate Rows	0		
Duplicate Rows (%)	0.0%		
Total Size in Memory	451.1 MB		
Average Row Size in Memory	100.0 B		
Variable Types	Numerical: 3 Categorical: 1		



Ввод [33]:

```
1 produv_test = pd.read_csv(path.joinpath('produv_test.csv'))
2 produv_test
```

executed in 6.20s, finished 19:45:31 2021-10-29

Out[33]:

	NPLV	SEC	RAS	POL
0	512324	2021-05-05 17:43:18	181.0	4.20
1	512324	2021-05-05 17:43:20	215.0	4.20
2	512324	2021-05-05 17:43:22	243.0	4.20
3	512324	2021-05-05 17:43:24	265.0	4.20
4	512324	2021-05-05 17:43:26	273.0	4.20
...	...	...	...	...
1692372	513374	2021-06-29 23:13:36	186.0	2.94
1692373	513374	2021-06-29 23:13:38	152.0	2.94
1692374	513374	2021-06-29 23:13:40	124.0	2.94
1692375	513374	2021-06-29 23:13:42	101.0	2.94
1692376	513374	2021-06-29 23:13:44	82.0	2.94

1692377 rows × 4 columns

Ввод [34]:

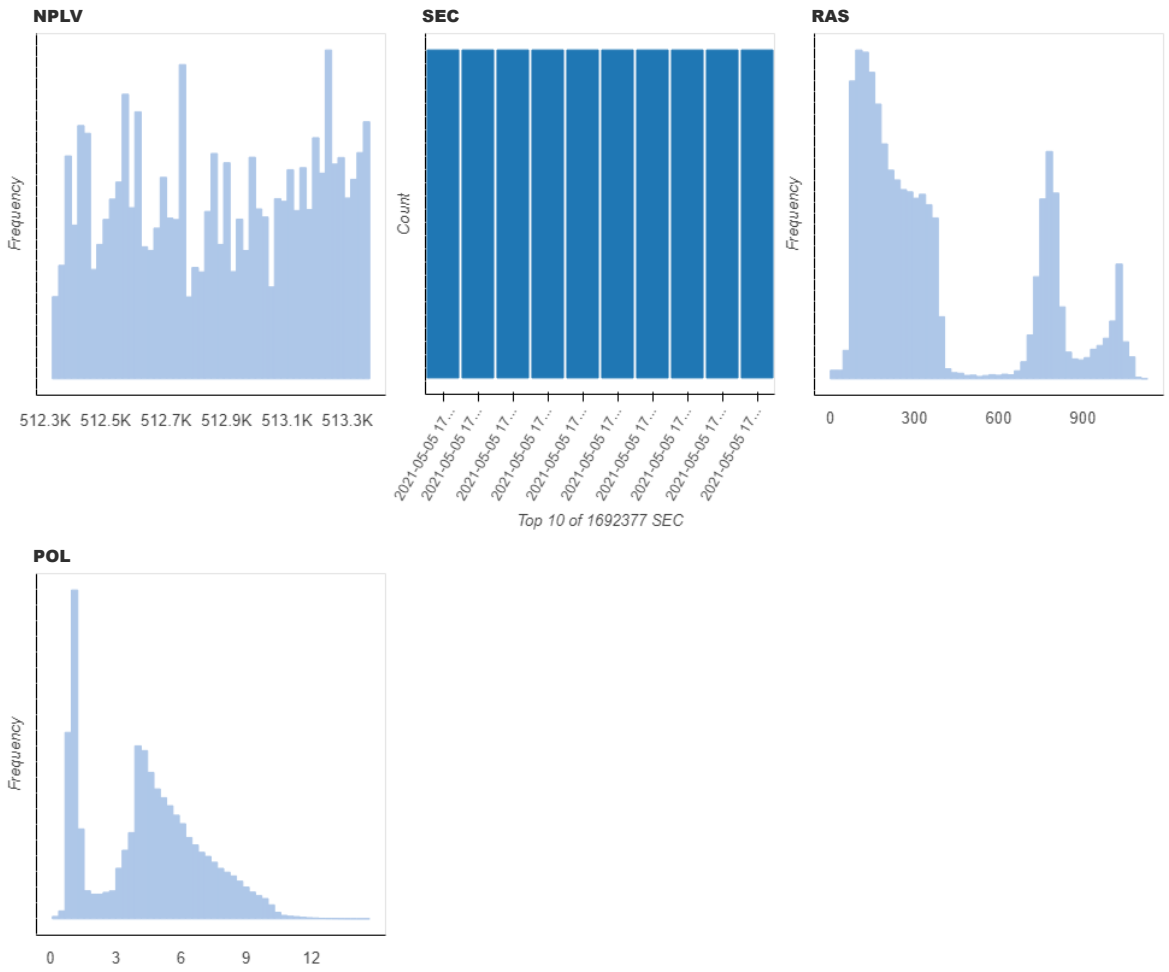
1 plot(produv\_test)

executed in 30.3s, finished 19:46:02 2021-10-29

Out[34]:

Hide Stats and Insights

Dataset Statistics		Dataset Insights	
Number of Variables	4	POL is skewed	Skewed
Number of Rows	1.6924×10 <sup>06</sup>	SEC has a high cardinality: 1692377 distinct values	High Cardinality
Missing Cells	0	SEC has constant length 19	Constant Length
Missing Cells (%)	0.0%	SEC has all distinct values	Unique
Duplicate Rows	0		
Duplicate Rows (%)	0.0%		
Total Size in Memory	161.4 MB		
Average Row Size in Memory	100.0 B		
Variable Types	Numerical: 3 Categorical: 1		



sip

Основная информация по плавке - характеристики плавки (марка металла, направление разливки) и оборудования

Ввод [35]:

```
1 sip_train = pd.read_csv(path.joinpath('sip_train.csv'))
2 sip_train
```

executed in 94ms, finished 19:46:02 2021-10-29

Out[35]:

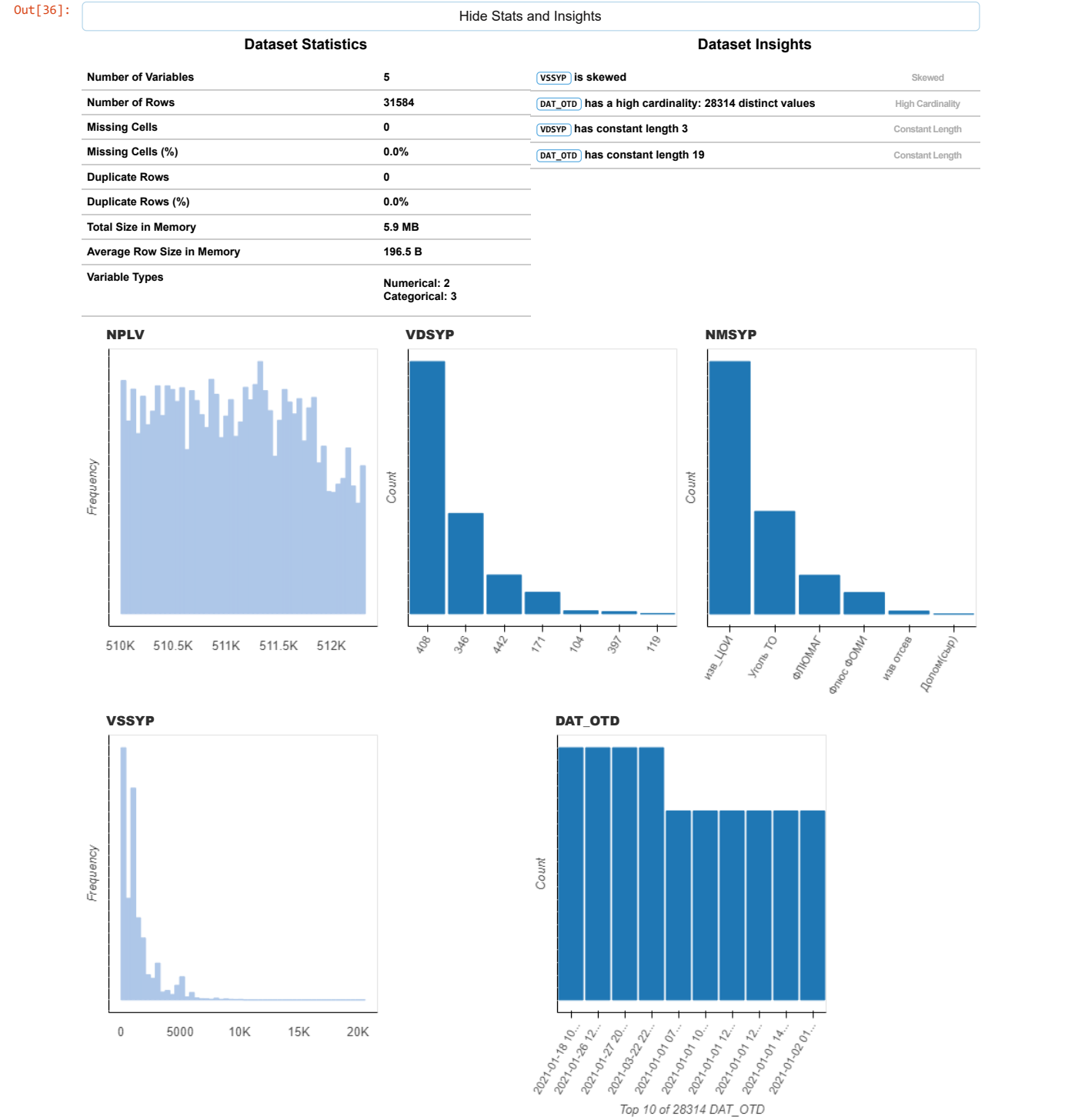
	NPLV	VDSYP	NMSYP	VSSYP	DAT_OTD
0	510008	346	Уголь ТО	570	2021-01-01 03:03:53
1	510008	346	Уголь ТО	220	2021-01-01 03:04:10
2	510008	408	изв_ЦОИ	7300	2021-01-01 03:08:17
3	510008	346	Уголь ТО	270	2021-01-01 03:09:03
4	510008	346	Уголь ТО	430	2021-01-01 03:09:20
...	...	...	...	...	...
31579	512322	408	изв_ЦОИ	30	2021-04-26 18:28:06
31580	512322	408	изв_ЦОИ	490	2021-04-26 18:28:07
31581	512322	408	изв_ЦОИ	20	2021-04-26 18:28:22
31582	512322	408	изв_ЦОИ	970	2021-04-26 18:30:26
31583	512322	408	изв_ЦОИ	10	2021-04-26 18:30:42

31584 rows × 5 columns

Ввод [36]:

```
1 plot(sip_train)
```

executed in 811ms, finished 19:46:03 2021-10-29



Ввод [37]:

```
1 sip_test = pd.read_csv(path.joinpath('sip_test.csv'))
2 sip_test
```

executed in 44ms, finished 19:46:03 2021-10-29

Out[37]:

	NPLV	VDSYP	NMSYP	VSSYP	DAT_OTD
0	512324	408	изв_ЦОИ	3020	2021-05-05 17:38:59
1	512324	442	ФЛЮМАГ	1670	2021-05-05 17:42:07
2	512324	408	изв_ЦОИ	2810	2021-05-05 17:42:07
3	512324	346	Уголь ТО	140	2021-05-05 17:43:39
4	512324	346	Уголь ТО	70	2021-05-05 17:43:49
...	...	...	...	...	...
10457	513374	346	Уголь ТО	630	2021-06-29 22:44:40
10458	513374	346	Уголь ТО	2310	2021-06-29 22:46:17
10459	513374	408	изв_ЦОИ	10010	2021-06-29 22:46:23
10460	513374	346	Уголь ТО	670	2021-06-29 22:48:22
10461	513374	408	изв_ЦОИ	1990	2021-06-29 23:07:10

10462 rows × 5 columns

Ввод [38]:

```
1 plot(sip_test)
```

executed in 604ms, finished 19:46:04 2021-10-29

Out[38]:

Hide Stats and Insights

Dataset Statistics

Number of Variables	5	VSSYP is skewed	Skewed
Number of Rows	10462	DAT_OTD has a high cardinality: 9494 distinct values	High Cardinality
Missing Cells	0	DAT_OTD has constant length 19	Constant Length
Missing Cells (%)	0.0%		
Duplicate Rows	0		
Duplicate Rows (%)	0.0%		
Total Size in Memory	2.0 MB		
Average Row Size in Memory	196.4 B		
Variable Types	Numerical: 2 Categorical: 3		

Dataset Insights

NPLV



VDSYP



NMSYP



VSSYP



DAT\_OTD



Top 10 of 9494 DAT\_OTD

target

целевые значения

Ввод [39]:

```
1 target_train = pd.read_csv(path.joinpath('target_train.csv'))
2 target_train
```

executed in 16ms, finished 19:46:04 2021-10-29

Out[39]:

	NPLV	TST	C
0	510008	1690	0.060
1	510009	1683	0.097
2	510010	1662	0.091
3	510011	1609	0.410
4	510012	1682	0.120
...	...	...	...
2058	512318	1626	0.145
2059	512319	1643	0.087
2060	512320	1615	0.141
2061	512321	1654	0.270
2062	512322	1630	0.183

2063 rows × 3 columns

Ввод [40]:

```
1 plot(target_train)
```

executed in 390ms, finished 19:46:05 2021-10-29

