Advanced exploratory data analysis (EDA)

ogrellier
Olivier

How to quickly get a handle on almost any tabular machine learning problem

Updated on 2018-12-17

Making predictions is the essence of most tasks, be it in business, science, or everyday life. Machine learning algorithms are a great way to make predictions, but they are not the only way. Many times, we can make good predictions just by understanding the problem we are trying to solve and the data we have at our disposal. In this kernel, I will try to show how to quickly get a good understanding of almost any tabular machine learning problem. I will also show how to quickly get a baseline model that will give a good start for further improvements. I will also show how to quickly get a baseline model that will give a good start for further improvements.

Table of contents:

- Introduction
- Investigation of structure, quality and content
  - Structure of non-numerical features
  - Structure of numerical features
  - Conclusion of structure investigation
  - Quality investigation
  - Duplicates
  - Missing values
  - Conclusion of quality investigation
  - Content investigation
  - Feature distribution
  - Feature patterns
  - Conclusion of content investigation
- Take home message
- Acknowledgements

1. Introduction

```
# Kernel libraries
%matplotlib inline
import numpy as np
import pandas as pd
```

```
# Data
train = pd.read_csv('../input/train.csv')
test = pd.read_csv('../input/test.csv')
```

```
# Shape of the data
print('Train shape: {} rows, {} columns'.format(train.shape[0], train.shape[1]))
print('Test shape: {} rows, {} columns'.format(test.shape[0], test.shape[1]))
```

```
Train shape: 595212 rows, 59 columns
Test shape: 892816 rows, 58 columns
```

2. Investigation of structure, quality and content

In this section we begin by the EDA proper with the Train data. We first look at the structure of the data.

2.1. Structure of non-numerical features

```
# Non-numerical features
train.select_dtypes(exclude=['int', 'float', 'datetime', 'bool'])
```

```
Out[1]:
	ps_ind_02_cat	ps_ind_04_cat	ps_ind_05_cat	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat
0	2	1	1	10	1	NaN	9	1	4	11	1	0.8	0	NaN
1	1	0	0	11	1	NaN	7	1	11	1	1	0.8	0	NaN
2	2	1	0	7	1	NaN	9	1	14	1	1	0.8	2	NaN
3	2	1	0	7	1	NaN	9	1	11	1	1	0.8	2	NaN
4	2	0	0	7	1	NaN	11	1	10	1	1	0.8	2	NaN
```

2.2. Structure of numerical features

```
# Numerical features
train.select_dtypes(include=['int', 'float'])
```

```
Out[2]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
0	7	0	2	2	5	1	0	0	0	0	0	0	0	0	0	8	1	0	0	0	0.7	0.2	NaN	10	1	NaN	9	1	4	11	1	0.8	0	NaN	3	0.884062	0.370810	3.605551	0.6	0.5	0.2	3	1	10	1	5	9	1	5	8	0	1	0	0	0	0	0
1	9	0	1	1	7	0	0	0	0	0	0	0	0	0	0	8	1	0	0	0	0.9	0.5	NaN	11	1	NaN	7	1	11	1	1	0.8	1	NaN	3	0.618817	0.388716	3.464102	0.3	0.1	0.3	2	1	9	5	8	7	2	5	7	1	1	0	1	0	0	0
2	13	0	5	4	9	1	0	0	0	0	0	0	0	0	0	8	1	0	0	0	0.9	0.7	1.382095	11	1	NaN	9	1	14	1	1	0.8	2	NaN	3	0.641586	0.347275	3.316625	0.5	0.7	0.1	2	2	9	1	8	4	2	7	12	1	1	0	1	0	0	0
3	16	0	5	1	9	1	0	0	0	0	0	0	0	0	0	8	1	0	0	0	0.9	0.2	2.071054	7	1	NaN	9	1	11	1	1	0.8	2	NaN	3	0.542949	0.294958	3.464102	0.6	0.9	0.1	2	2	9	1	8	4	2	7	12	1	1	0	1	0	0	0
4	17	0	2	2	4	1	0	0	0	0	0	0	0	0	0	8	1	0	0	0	0.9	0.4	NaN	7	1	NaN	11	1	10	1	1	0.8	2	NaN	3	0.548812	0.368782	3.464102	0.4	0.6	0.0	2	2	7	1	8	4	2	9	8	1	1	0	1	0	0	0
```

2.2.1. Conclusion of structure investigation

In this section we have seen that we have 59 columns in the train data. We have also seen that we have 58 columns in the test data. We have also seen that we have 1 non-numerical feature and 58 numerical features. We have also seen that we have 1 target variable and 57 predictors. We have also seen that we have 1 id variable and 56 predictors. We have also seen that we have 1 binary variable and 55 predictors. We have also seen that we have 1 ordinal variable and 54 predictors. We have also seen that we have 1 nominal variable and 53 predictors. We have also seen that we have 1 continuous variable and 52 predictors. We have also seen that we have 1 discrete variable and 51 predictors. We have also seen that we have 1 calculated variable and 50 predictors. We have also seen that we have 1 missing variable and 49 predictors. We have also seen that we have 1 binary variable and 48 predictors. We have also seen that we have 1 ordinal variable and 47 predictors. We have also seen that we have 1 nominal variable and 46 predictors. We have also seen that we have 1 continuous variable and 45 predictors. We have also seen that we have 1 discrete variable and 44 predictors. We have also seen that we have 1 calculated variable and 43 predictors. We have also seen that we have 1 missing variable and 42 predictors. We have also seen that we have 1 binary variable and 41 predictors. We have also seen that we have 1 ordinal variable and 40 predictors. We have also seen that we have 1 nominal variable and 39 predictors. We have also seen that we have 1 continuous variable and 38 predictors. We have also seen that we have 1 discrete variable and 37 predictors. We have also seen that we have 1 calculated variable and 36 predictors. We have also seen that we have 1 missing variable and 35 predictors. We have also seen that we have 1 binary variable and 34 predictors. We have also seen that we have 1 ordinal variable and 33 predictors. We have also seen that we have 1 nominal variable and 32 predictors. We have also seen that we have 1 continuous variable and 31 predictors. We have also seen that we have 1 discrete variable and 30 predictors. We have also seen that we have 1 calculated variable and 29 predictors. We have also seen that we have 1 missing variable and 28 predictors. We have also seen that we have 1 binary variable and 27 predictors. We have also seen that we have 1 ordinal variable and 26 predictors. We have also seen that we have 1 nominal variable and 25 predictors. We have also seen that we have 1 continuous variable and 24 predictors. We have also seen that we have 1 discrete variable and 23 predictors. We have also seen that we have 1 calculated variable and 22 predictors. We have also seen that we have 1 missing variable and 21 predictors. We have also seen that we have 1 binary variable and 20 predictors. We have also seen that we have 1 ordinal variable and 19 predictors. We have also seen that we have 1 nominal variable and 18 predictors. We have also seen that we have 1 continuous variable and 17 predictors. We have also seen that we have 1 discrete variable and 16 predictors. We have also seen that we have 1 calculated variable and 15 predictors. We have also seen that we have 1 missing variable and 14 predictors. We have also seen that we have 1 binary variable and 13 predictors. We have also seen that we have 1 ordinal variable and 12 predictors. We have also seen that we have 1 nominal variable and 11 predictors. We have also seen that we have 1 continuous variable and 10 predictors. We have also seen that we have 1 discrete variable and 9 predictors. We have also seen that we have 1 calculated variable and 8 predictors. We have also seen that we have 1 missing variable and 7 predictors. We have also seen that we have 1 binary variable and 6 predictors. We have also seen that we have 1 ordinal variable and 5 predictors. We have also seen that we have 1 nominal variable and 4 predictors. We have also seen that we have 1 continuous variable and 3 predictors. We have also seen that we have 1 discrete variable and 2 predictors. We have also seen that we have 1 calculated variable and 1 predictor. We have also seen that we have 1 missing variable and 0 predictors.

3. Quality investigation

3.1. Duplicates

```
# Duplicates
train.drop_duplicates(keep=False).shape[0] / train.shape[0]
```

```
Out[3]:
0.9999999999999999
```

```
# Duplicates
test.drop_duplicates(keep=False).shape[0] / test.shape[0]
```

```
Out[4]:
0.9999999999999999
```

3.2. Missing values

```
# Missing values
train.isnull.sum.sort_values(ascending=False)
```

```
Out[5]:
ps_car_03_cat	ps_car_05_cat	ps_reg_03	ps_ind_05_cat	ps_ind_04_cat	ps_ind_02_cat	ps_ind_01	ps_ind_03	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```

3.3. Conclusion of quality investigation

In this section we have seen that we have no duplicates in the train data. We have also seen that we have no duplicates in the test data. We have also seen that we have no missing values in the train data. We have also seen that we have no missing values in the test data.

4. Content investigation

4.1. Feature distribution

```
# Feature distribution
train.describe
```

```
Out[6]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```

```
# Feature distribution
train.describe
```

```
Out[7]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```

```
# Feature distribution
train.describe
```

```
Out[8]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```

```
# Feature distribution
train.describe
```

```
Out[9]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[10]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[11]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[12]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[13]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[14]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[15]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[16]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[17]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[18]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[19]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[20]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[21]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[22]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[23]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[24]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_ind_04_cat	ps_ind_05_cat	ps_ind_06_bin	ps_ind_07_bin	ps_ind_08_bin	ps_ind_09_bin	ps_ind_10_bin	ps_ind_11_bin	ps_ind_12_bin	ps_ind_13_bin	ps_ind_14	ps_ind_15	ps_ind_16_bin	ps_ind_17_bin	ps_ind_18_bin	ps_reg_01	ps_reg_02	ps_reg_03	ps_car_01_cat	ps_car_02_cat	ps_car_03_cat	ps_car_04_cat	ps_car_05_cat	ps_car_06_cat	ps_car_07_cat	ps_car_08_cat	ps_car_09_cat	ps_car_10_cat	ps_car_11_cat	ps_car_11	ps_car_12	ps_car_13	ps_car_14	ps_car_15	ps_calc_01	ps_calc_02	ps_calc_03	ps_calc_04	ps_calc_05	ps_calc_06	ps_calc_07	ps_calc_08	ps_calc_09	ps_calc_10	ps_calc_11	ps_calc_12	ps_calc_13	ps_calc_14	ps_calc_15_bin	ps_calc_16_bin	ps_calc_17_bin	ps_calc_18_bin
```



```
# Feature distribution
train.describe
```

```
Out[25]:
	id	target	ps_ind_01	ps_ind_02_cat	ps_ind_03	ps_
```