1. 数据预处理：

总计16144条记录，ID不为空，且无重复值。

Area: ['.', 'nan', '?']转换为000，取地址第一位，再one-hot

#TODO：提取人口大村等信息。

ck字段：ck\_mapping = {'0': 0, '1': 1, '?': 0, '.': 0}

ck-saveall: fillna(0)

ck-drawall: fillna(0)

ck-percentage: ck-saveall/(ck-saveall+ck-drawall+0.01)

ck-savetime

ck-drawtime

---------------------------------------------

ck-saveavg have 1659 null values:

ck-saveavg has [ 2280 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['296806', '102994', '93544', '70256', '101000', '116813', '13334',

'104118', '518098', '699700',

...

'705402', '51715', '99391', '890000', '46600', '76888', '35850',

'316667', '21948', '3344000'],

dtype='object', length=2213)

---------------------------------------------

ck-drawavg have 1690 null values:

ck-drawavg has [ 2209 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['272796', '11756', '10424', '320880', '7383', '413025', '84325',

'37297', '29558', '27360',

...

'1901', '181750', '25798', '8702', '68130', '38453', '5315', '61289',

'83520', '37505'],

dtype='object', length=2147)

---------------------------------------------

ck-avg have 1632 null values:

ck-avg has [ 2206 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['2229', '48650', '2499', '3384', '670', '1342', '457', '4961', '456',

'2669',

...

'1981', '8956', '415', '1210', '2050', '152', '1257', '5659', '26229',

'313'],

dtype='object', length=1684)

---------------------------------------------

dep-saveall have 1572 null values:

dep-saveall has [ 6688 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index([], dtype='object')

---------------------------------------------

dep-drawall have 1673 null values:

dep-drawall has [ 6706 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index([], dtype='object')

---------------------------------------------

dep-savetime have 1598 null values:

dep-savetime has [ 509 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['140', '168', '189', '132', '159', '173', '139', '125', '181', '311',

...

'349', '9672', '266', '340', '1260', '359', '747', '419', '700', '759'],

dtype='object', length=380)

---------------------------------------------

dep-drawtime have 1774 null values:

dep-drawtime has [ 390 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['134', '96', '120', '102', '182', '152', '137', '180', '107', '161',

...

'590', '279', '2923', '1137', '176', '498', '398', '231', '640', '565'],

dtype='object', length=273)

---------------------------------------------

depsaveavg have 1662 null values:

depsaveavg has [ 6648 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index([], dtype='object')

---------------------------------------------

depdrawavg have 1718 null values:

depdrawavg has [ 6787 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index([], dtype='object')

---------------------------------------------

dep-avg have 1646 null values:

dep-avg has [ 6469 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index([], dtype='object')

---------------------------------------------

dep-9201 have 1640 null values:

dep-9201 has [ 3831 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index([], dtype='object')

---------------------------------------------

fed-9201 have 1634 null values:

fed-9201 has [ 396 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['3470', '3471', '3', '28', '3478', '84', '3808', '27', '3806', '553',

...

'545', '4143521', '186146', '24478', '4807', '4865', '184', '11911',

'136760', '201'],

dtype='object', length=393)

---------------------------------------------

fed-avg have 1622 null values:

fed-avg has [ 362 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['317', '2', '4', '12', '6', '289', '291', '293', '290', '3',

...

'1569', '71', '1296', '15984', '491', '24994127', '394', '65882',

'46911', '418'],

dtype='object', length=358)

---------------------------------------------

comp have 1649 null values:

0 17157

? 3286

1 2401

. 1651

Name: comp, dtype: int64

use following code to generate features from label mapping:

comp\_mapping = {'0': 0, '?': 1, '1': 2, '.': 3}

#TODO: add your logic to handle None

data\_comp, comp\_ohe = one\_hot\_encoder\_column(data, 'comp', comp\_mapping, fill\_na=99)

---------------------------------------------

ck-changame have 1637 null values:

ck-changame has [ 2056 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['36000', '29558', '686108', '287534', '11000', '1484799', '9354485',

'138660', '1179014', '69300',

...

'6373364', '1519114', '429000', '32692702', '588523', '109250',

'1860072', '15137535', '8770043', '1881966'],

dtype='object', length=2007)

---------------------------------------------

dep-changame have 1657 null values:

dep-changame has [ 4272 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index([], dtype='object')

---------------------------------------------

ck-changtime have 1678 null values:

ck-changtime has [ 281 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['91', '88', '36', '64', '63', '68', '34', '78', '35', '53',

...

'303', '362', '126', '492', '129', '456', '216', '268', '207', '175'],

dtype='object', length=230)

---------------------------------------------

dep-changtime have 1627 null values:

dep-changtime has [ 353 ] different values, please create features yourself!

please notice here are some seldom\_indexes: Index(['71', '114', '94', '130', '79', '84', '99', '123', '83', '88',

...

'135', '117', '996', '368', '161', '680', '569', '157', '379', '238'],

dtype='object', length=272)

---------------------------------------------

VV do not have any null values.

VV is numeric column:

count 26144.000000

mean 0.022567

std 0.148522

min 0.000000

25% 0.000000

50% 0.000000

75% 0.000000

max 1.000000

Name: VV, dtype: float64

---------------------------------------------