In [192]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [193]:
```
train = pd.read_csv('C:/Users/USER/Desktop/DSN Hackathon/Bootcamp Qualificat
test_given = pd.read_csv('C:/Users/USER/Desktop/DSN Hackathon/Bootcamp Quali
sub = pd.read_csv('C:/Users/USER/Desktop/DSN Hackathon/Bootcamp Qualificatio
```

In [171]:
```
1 train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 56000 entries, 0 to 55999
Data columns (total 52 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Applicant_ID    56000 non-null  object
 1   form_field1     53471 non-null  float64
 2   form_field2     52156 non-null  float64
 3   form_field3     55645 non-null  float64
 4   form_field4     55645 non-null  float64
 5   form_field5     55645 non-null  float64
 6   form_field6     42640 non-null  float64
 7   form_field7     50837 non-null  float64
 8   form_field8     42640 non-null  float64
 9   form_field9     47992 non-null  float64
 10  form_field10    55645 non-null  float64
 11  form_field11    24579 non-null  float64
 12  form_field12    46105 non-null  float64
 13  form_field13    50111 non-null  float64
 14  form_field14    56000 non-null  int64
 15  form_field15    33525 non-null  float64
 16  form_field16    42964 non-null  float64
 17  form_field17    44849 non-null  float64
 18  form_field18    45598 non-null  float64
 19  form_field19    55996 non-null  float64
 20  form_field20    55645 non-null  float64
 21  form_field21    40146 non-null  float64
 22  form_field22    35600 non-null  float64
 23  form_field23    27877 non-null  float64
 24  form_field24    42703 non-null  float64
 25  form_field25    50550 non-null  float64
 26  form_field26    48562 non-null  float64
 27  form_field27    46701 non-null  float64
 28  form_field28    55645 non-null  float64
 29  form_field29    55645 non-null  float64
 30  form_field30    30491 non-null  float64
 31  form_field31    16592 non-null  float64
 32  form_field32    50550 non-null  float64
 33  form_field33    54744 non-null  float64
 34  form_field34    55645 non-null  float64
 35  form_field35    32852 non-null  float64
 36  form_field36    54005 non-null  float64
 37  form_field37    50550 non-null  float64
 38  form_field38    55645 non-null  float64
 39  form_field39    51789 non-null  float64
 40  form_field40    12271 non-null  float64
 41  form_field41    17771 non-null  float64
 42  form_field42    54677 non-null  float64
 43  form_field43    55432 non-null  float64
 44  form_field44    50617 non-null  float64
 45  form_field45    24683 non-null  float64
 46  form_field46    40096 non-null  float64
 47  form_field47    56000 non-null  object
 48  form_field48    35111 non-null  float64
 49  form_field49    55645 non-null  float64
```

```
 50  form_field50    44944 non-null  float64
 51  default_status  56000 non-null  object
dtypes: float64(48), int64(1), object(3)
memory usage: 22.2+ MB
```

In [172]:
```
1 test_given.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24000 entries, 0 to 23999
Data columns (total 51 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Applicant_ID  24000 non-null  object
 1   form_field1   22890 non-null  float64
 2   form_field2   22291 non-null  float64
 3   form_field3   23854 non-null  float64
 4   form_field4   23854 non-null  float64
 5   form_field5   23854 non-null  float64
 6   form_field6   18396 non-null  float64
 7   form_field7   21769 non-null  float64
 8   form_field8   18396 non-null  float64
 9   form_field9   20600 non-null  float64
 10  form_field10  23853 non-null  float64
 11  form_field11  10602 non-null  float64
 12  form_field12  19817 non-null  float64
 13  form_field13  21537 non-null  float64
 14  form_field14  24000 non-null  int64
 15  form_field15  14408 non-null  float64
 16  form_field16  18526 non-null  float64
 17  form_field17  19305 non-null  float64
 18  form_field18  19631 non-null  float64
 19  form_field19  24000 non-null  float64
 20  form_field20  23853 non-null  float64
 21  form_field21  17293 non-null  float64
 22  form_field22  15276 non-null  float64
 23  form_field23  11875 non-null  float64
 24  form_field24  18395 non-null  float64
 25  form_field25  21744 non-null  float64
 26  form_field26  20828 non-null  float64
 27  form_field27  20090 non-null  float64
 28  form_field28  23853 non-null  float64
 29  form_field29  23853 non-null  float64
 30  form_field30  13092 non-null  float64
 31  form_field31  7190 non-null   float64
 32  form_field32  21744 non-null  float64
 33  form_field33  23505 non-null  float64
 34  form_field34  23853 non-null  float64
 35  form_field35  14134 non-null  float64
 36  form_field36  23097 non-null  float64
 37  form_field37  21744 non-null  float64
 38  form_field38  23853 non-null  float64
 39  form_field39  22171 non-null  float64
 40  form_field40  5172 non-null   float64
 41  form_field41  7651 non-null   float64
 42  form_field42  23422 non-null  float64
 43  form_field43  23750 non-null  float64
 44  form_field44  21638 non-null  float64
 45  form_field45  10462 non-null  float64
 46  form_field46  17115 non-null  float64
 47  form_field47  24000 non-null  object
 48  form_field48  15078 non-null  float64
 49  form_field49  23854 non-null  float64
```

```
 50  form_field50  19203 non-null  float64
dtypes: float64(48), int64(1), object(2)
memory usage: 9.3+ MB
```

In [173]:
```
1 train.describe()
```

Out[173]:

	form_field1	form_field2	form_field3	form_field4	form_field5	form_field6	forr
count	53471.000000	52156.000000	55645.000000	55645.000000	55645.000000	4.264000e+04	5.0837
mean	3491.795665	0.550737	1.052225	0.851979	1.956317	6.244479e+05	6.8652
std	188.462426	0.820979	2.147768	3.157692	10.512396	1.433422e+06	1.9127
min	2990.000000	0.000000	0.000000	0.000000	0.000000	0.000000e+00	0.0000
25%	3358.000000	0.070788	0.000000	0.000000	0.000000	1.400400e+04	6.8697
50%	3484.000000	0.267575	0.062000	0.000000	0.000000	1.155330e+05	2.7043
75%	3620.000000	0.719512	1.282000	0.000000	0.000000	5.259280e+05	6.9938
max	3900.000000	18.015050	57.371600	91.672200	407.748600	5.313546e+07	2.1587

8 rows × 49 columns

In [174]:
```
1 train.isna().sum()
```

Out[174]:
```
Applicant_ID          0
form_field1        2529
form_field2        3844
form_field3         355
form_field4         355
form_field5         355
form_field6       13360
form_field7        5163
form_field8       13360
form_field9        8008
form_field10        355
form_field11      31421
form_field12       9895
form_field13       5889
form_field14          0
form_field15      22475
form_field16      13036
form_field17      11151
form_field18      10402
form_field19          4
form_field20        355
form_field21      15854
form_field22      20400
form_field23      28123
form_field24      13297
form_field25       5450
form_field26       7438
form_field27       9299
form_field28        355
form_field29        355
form_field30      25509
form_field31      39408
form_field32       5450
form_field33       1256
form_field34        355
form_field35      23148
form_field36       1995
form_field37       5450
form_field38        355
form_field39       4211
form_field40      43729
form_field41      38229
form_field42       1323
form_field43        568
form_field44       5383
form_field45      31317
form_field46      15904
form_field47          0
form_field48      20889
form_field49        355
form_field50      11056
default_status        0
dtype: int64
```

In [175]:
```
1 test_given.isna().sum()
```

Out[175]:
```
Applicant_ID        0
form_field1      1110
form_field2      1709
form_field3       146
form_field4       146
form_field5       146
form_field6      5604
form_field7      2231
form_field8      5604
form_field9      3400
form_field10      147
form_field11    13398
form_field12     4183
form_field13     2463
form_field14        0
form_field15     9592
form_field16     5474
form_field17     4695
form_field18     4369
form_field19        0
form_field20      147
form_field21     6707
form_field22     8724
form_field23    12125
form_field24     5605
form_field25     2256
form_field26     3172
form_field27     3910
form_field28      147
form_field29      147
form_field30    10908
form_field31    16810
form_field32     2256
form_field33      495
form_field34      147
form_field35     9866
form_field36      903
form_field37     2256
form_field38      147
form_field39     1829
form_field40    18828
form_field41    16349
form_field42      578
form_field43      250
form_field44     2362
form_field45    13538
form_field46     6885
form_field47        0
form_field48     8922
form_field49      146
form_field50     4797
dtype: int64
```

In [194]:
```
#To get categorical columns
#train.select_dtypes(object).columns
```

In [195]:
```
#To encode categorical columns
train['form_field47'].replace('charge',1, inplace=True)
train['form_field47'].replace('lending',0, inplace=True)
train['default_status'].replace('yes',1, inplace=True)
train['default_status'].replace('no',0, inplace=True)

test_given['form_field47'].replace('charge',1, inplace=True)
test_given['form_field47'].replace('lending',0, inplace=True)

```

In [196]:
```
#Checking the null values in the train dataset
train_null = train.isna().sum()/len(train)
train_null[train_null>0].sort_values()
```

Out[196]:
```
form_field19    0.000071
form_field49    0.006339
form_field29    0.006339
form_field3     0.006339
form_field4     0.006339
form_field5     0.006339
form_field20    0.006339
form_field10    0.006339
form_field28    0.006339
form_field38    0.006339
form_field34    0.006339
form_field43    0.010143
form_field33    0.022429
form_field42    0.023625
form_field36    0.035625
form_field1     0.045161
form_field2     0.068643
form_field39    0.075196
form_field7     0.092196
form_field44    0.096125
form_field32    0.097321
form_field37    0.097321
form_field25    0.097321
form_field13    0.105161
form_field26    0.132821
form_field9     0.143000
form_field27    0.166054
form_field12    0.176696
form_field18    0.185750
form_field50    0.197429
form_field17    0.199125
form_field16    0.232786
form_field24    0.237446
form_field6     0.238571
form_field8     0.238571
form_field21    0.283107
form_field46    0.284000
form_field22    0.364286
form_field48    0.373018
form_field15    0.401339
form_field35    0.413357
form_field30    0.455518
form_field23    0.502196
form_field45    0.559232
form_field11    0.561089
form_field41    0.682661
form_field31    0.703714
form_field40    0.780875
dtype: float64
```

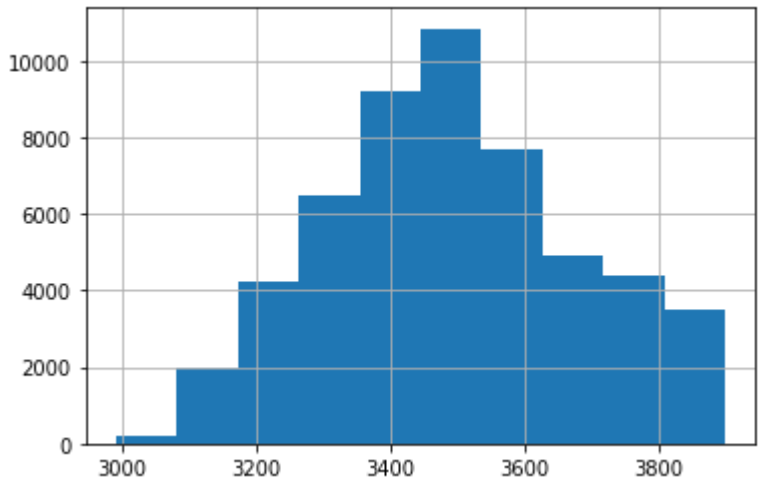
In [197]:
```
#Checking the null values in the test_given dataset

test_given_null = test_given.isna().sum()/len(test_given)
test_given_null[test_given_null>0].sort_values()
```

Out[197]:
```
form_field3     0.006083
form_field4     0.006083
form_field5     0.006083
form_field49    0.006083
form_field38    0.006125
form_field28    0.006125
form_field29    0.006125
form_field20    0.006125
form_field10    0.006125
form_field34    0.006125
form_field43    0.010417
form_field33    0.020625
form_field42    0.024083
form_field36    0.037625
form_field1     0.046250
form_field2     0.071208
form_field39    0.076208
form_field7     0.092958
form_field32    0.094000
form_field37    0.094000
form_field25    0.094000
form_field44    0.098417
form_field13    0.102625
form_field26    0.132167
form_field9     0.141667
form_field27    0.162917
form_field12    0.174292
form_field18    0.182042
form_field17    0.195625
form_field50    0.199875
form_field16    0.228083
form_field6     0.233500
form_field8     0.233500
form_field24    0.233542
form_field21    0.279458
form_field46    0.286875
form_field22    0.363500
form_field48    0.371750
form_field15    0.399667
form_field35    0.411083
form_field30    0.454500
form_field23    0.505208
form_field11    0.558250
form_field45    0.564083
form_field41    0.681208
form_field31    0.700417
form_field40    0.784500
dtype: float64
```

DATA VISUALIZATIONS

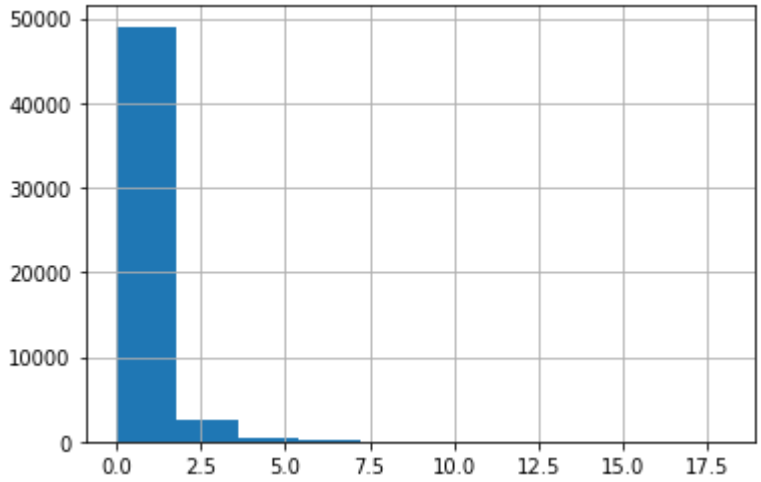
In [105]:
```
1 train['form_field1'].hist()
```

Out[105]: <matplotlib.axes._subplots.AxesSubplot at 0x11f75dc8>



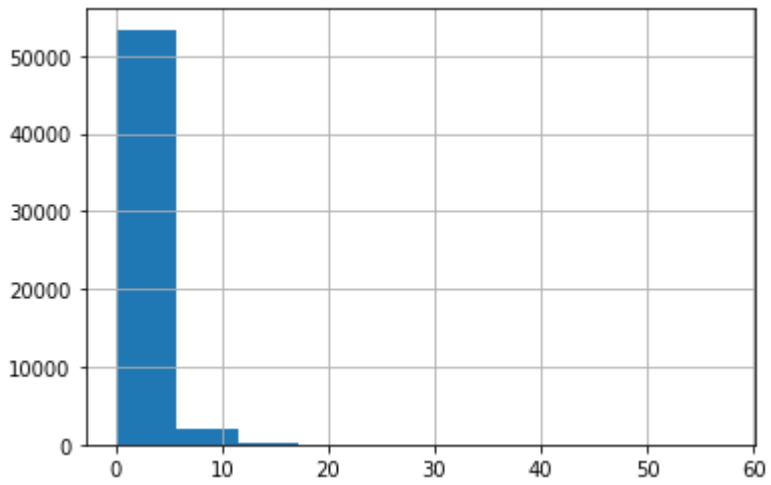
In [106]:
```
1 train['form_field2'].hist()
```

Out[106]: <matplotlib.axes._subplots.AxesSubplot at 0x15a74208>



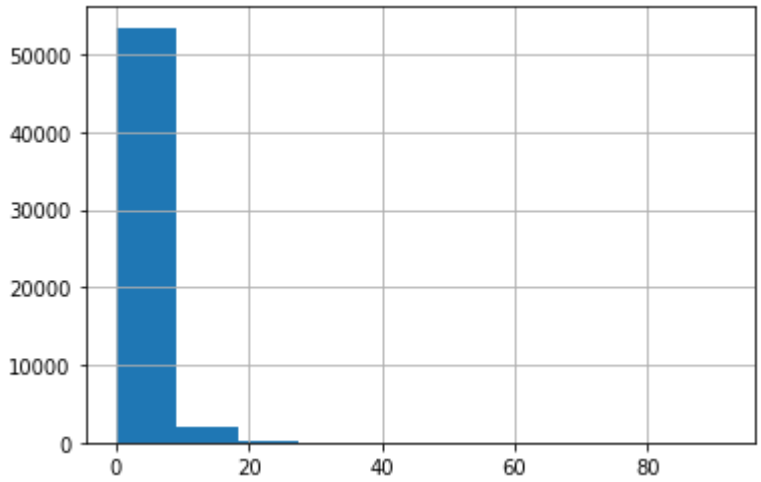
In [107]:
```
1 train['form_field3'].hist()
```

Out[107]: <matplotlib.axes._subplots.AxesSubplot at 0x15f473c8>



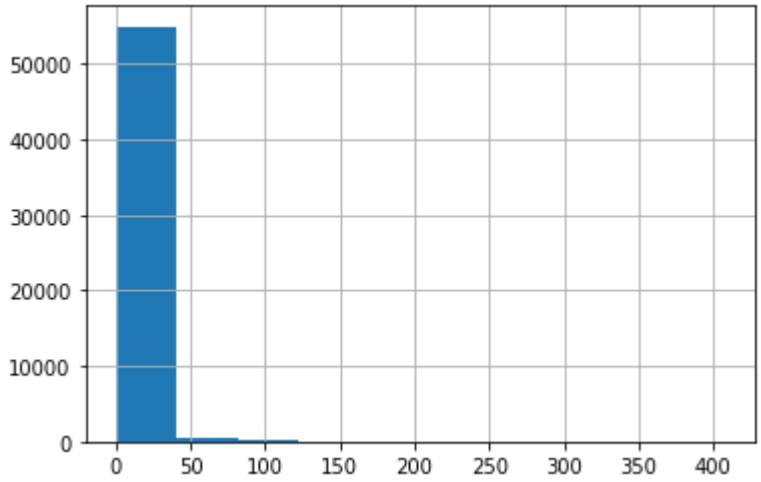
In [108]:
```
1 train['form_field4'].hist()
```

Out[108]: <matplotlib.axes._subplots.AxesSubplot at 0xbd40308>



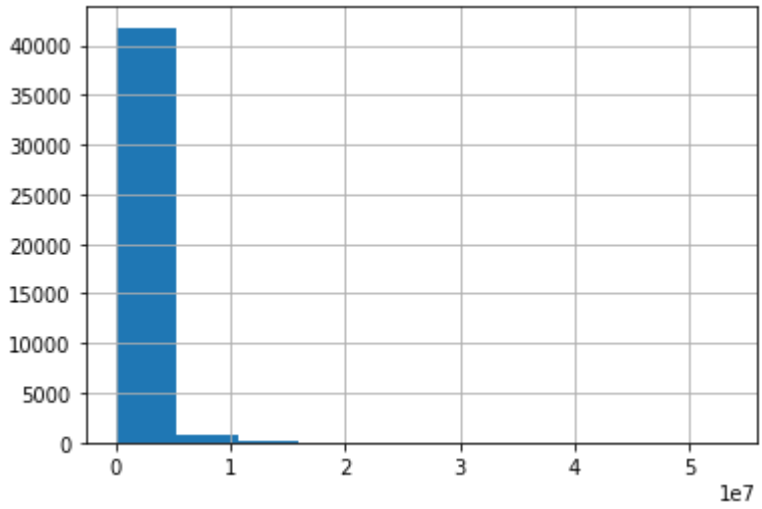
In [109]:
```
1 train['form_field5'].hist()
```

Out[109]: <matplotlib.axes._subplots.AxesSubplot at 0x11ea0988>



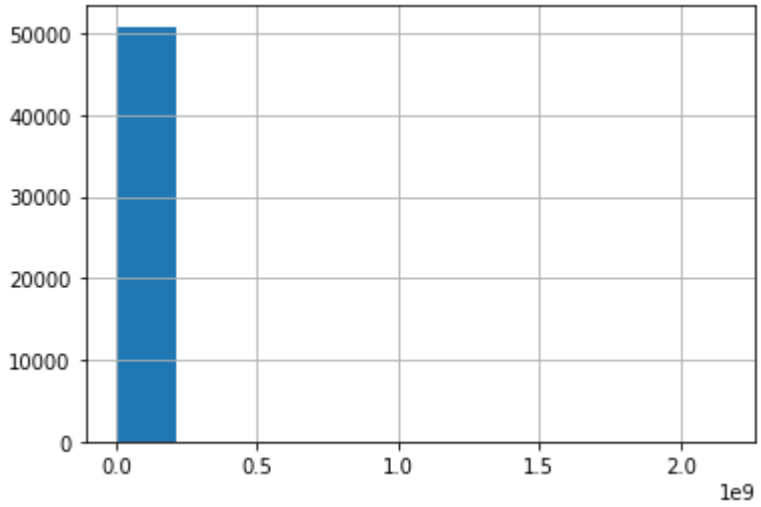
In [110]:
```
1 train['form_field6'].hist()
```

Out[110]: <matplotlib.axes._subplots.AxesSubplot at 0x11f3e148>



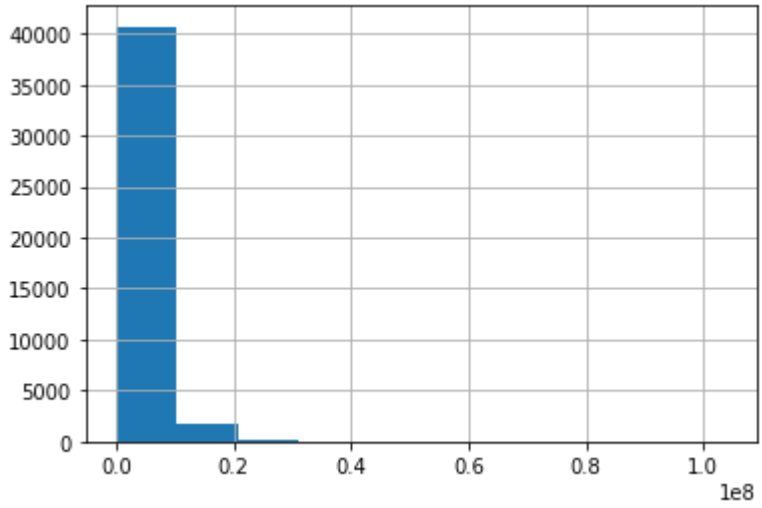
In [111]:
```
1 train['form_field7'].hist()
```

Out[111]: <matplotlib.axes._subplots.AxesSubplot at 0x148a1c08>



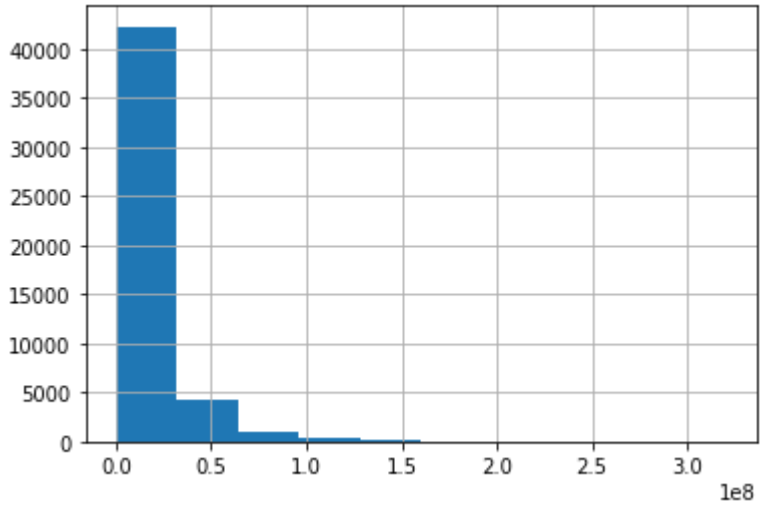
In [112]:
```
1 train['form_field8'].hist()
```

Out[112]: <matplotlib.axes._subplots.AxesSubplot at 0x14c13f08>



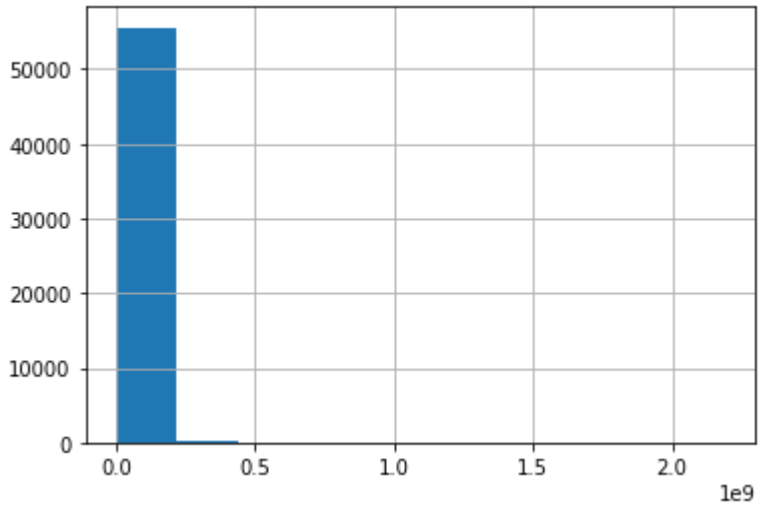
In [113]:
```
1 train['form_field9'].hist()
```

Out[113]: <matplotlib.axes._subplots.AxesSubplot at 0x1537c688>



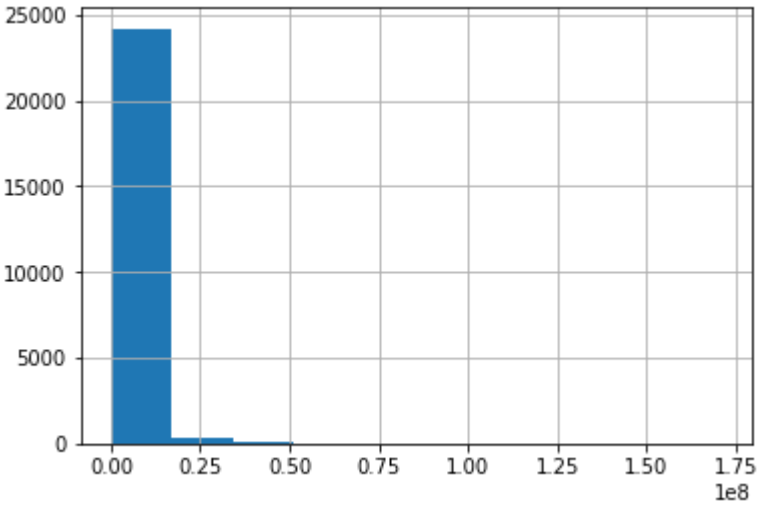
In [114]:
```
1 train['form_field10'].hist()
```

Out[114]: <matplotlib.axes._subplots.AxesSubplot at 0x153e3548>



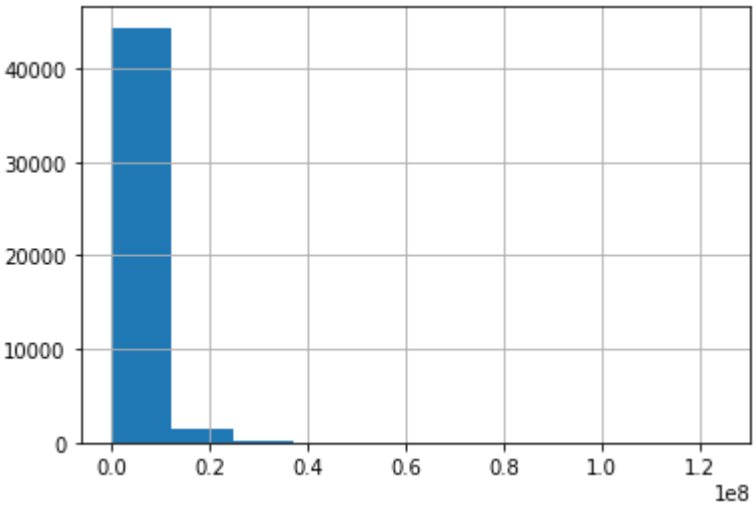
In [115]:
```
1 train['form_field11'].hist()
```

Out[115]: <matplotlib.axes._subplots.AxesSubplot at 0x1609f648>



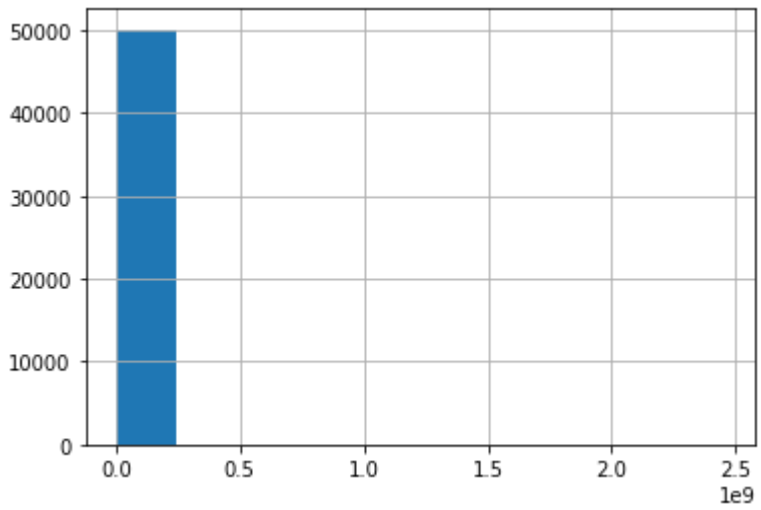
In [116]:
```
1 train['form_field12'].hist()
```

Out[116]: <matplotlib.axes._subplots.AxesSubplot at 0x160f6648>



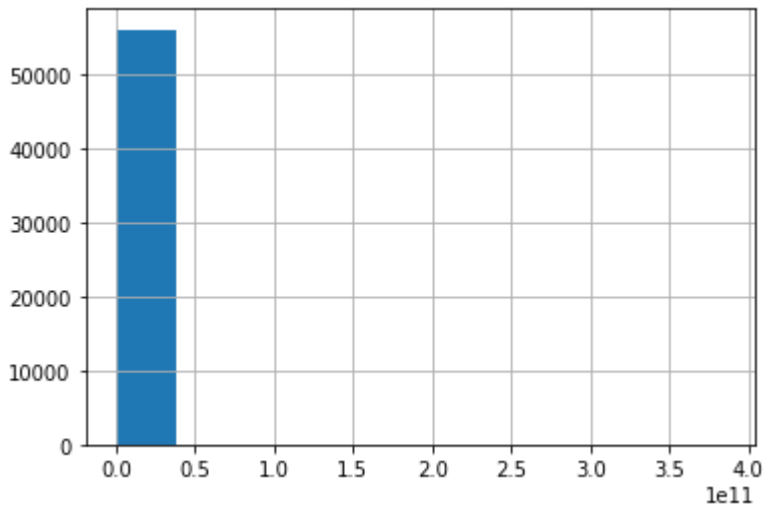
In [117]:
```
1 train['form_field13'].hist()
```

Out[117]: <matplotlib.axes._subplots.AxesSubplot at 0x153ed0c8>



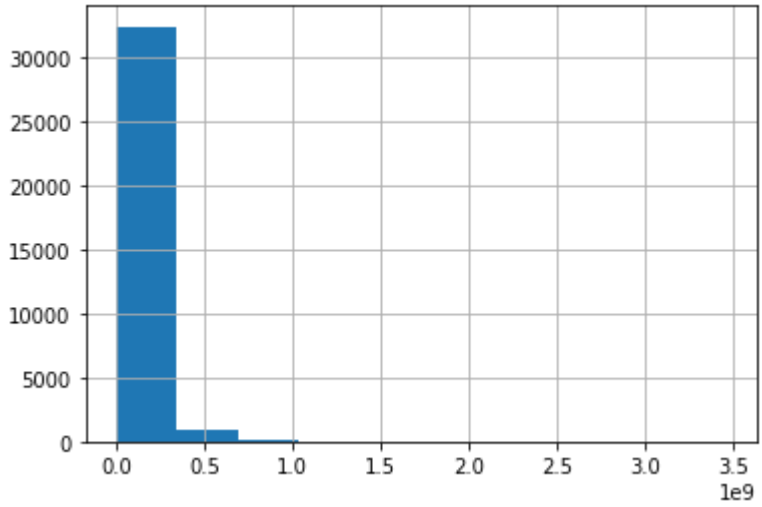
In [118]:
```
1 train['form_field14'].hist()
```

Out[118]: <matplotlib.axes._subplots.AxesSubplot at 0x168a8708>



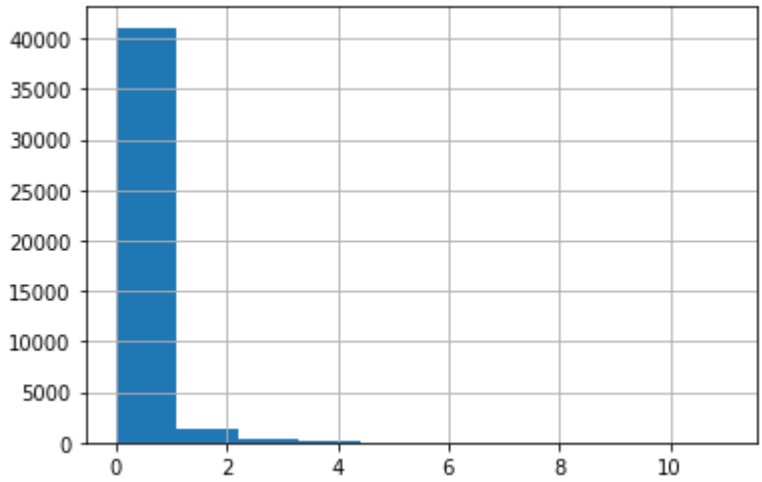
In [119]:
```
1 train['form_field15'].hist()
```

Out[119]: <matplotlib.axes._subplots.AxesSubplot at 0x1691dd48>



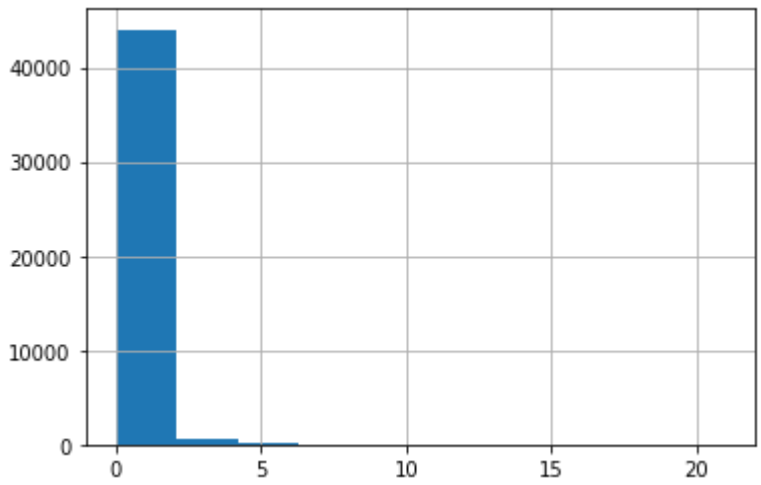
In [120]:
```
1 train['form_field16'].hist()
```

Out[120]: <matplotlib.axes._subplots.AxesSubplot at 0x16980348>



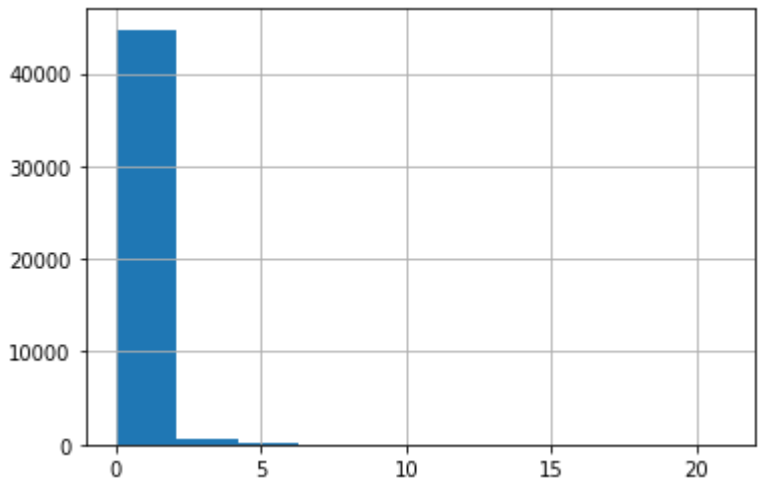
In [121]:
```
1 train['form_field17'].hist()
```

Out[121]: <matplotlib.axes._subplots.AxesSubplot at 0x16a2c9c8>



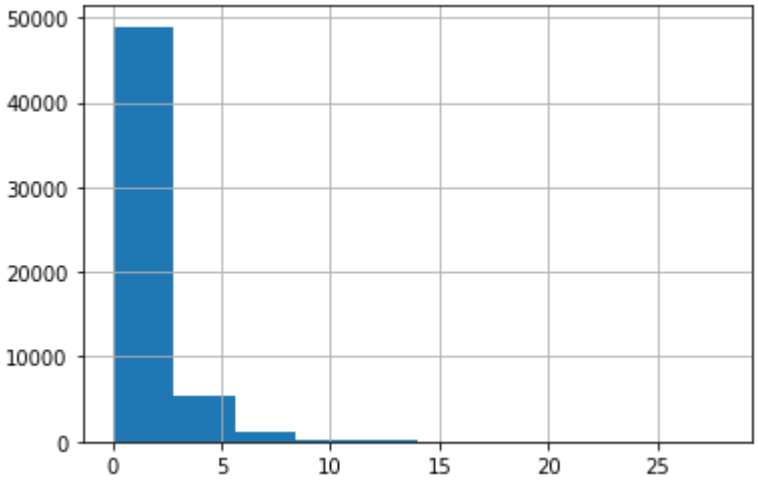
In [122]:
```
1 train['form_field18'].hist()
```

Out[122]: <matplotlib.axes._subplots.AxesSubplot at 0x16ac8348>



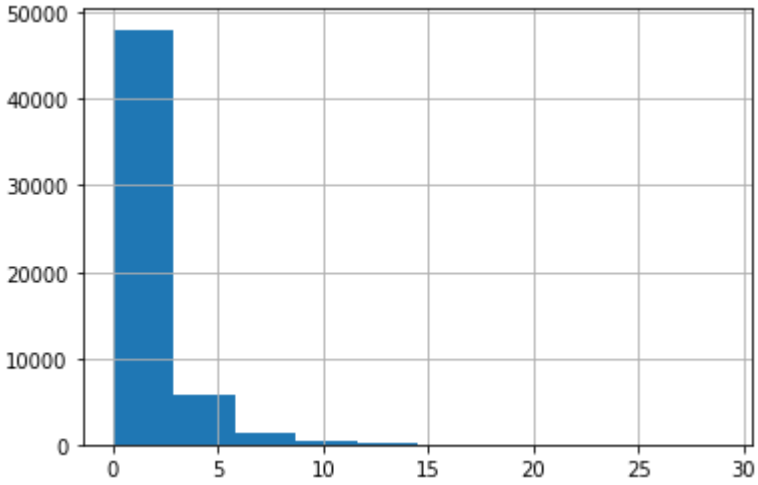
In [123]:
```
1 train['form_field19'].hist()
```

Out[123]: <matplotlib.axes._subplots.AxesSubplot at 0x16b339c8>



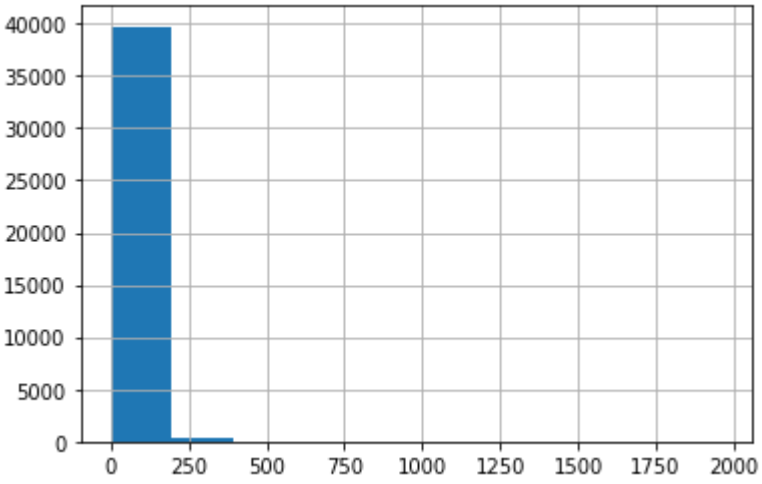
In [124]:
```
1 train['form_field20'].hist()
```

Out[124]: <matplotlib.axes._subplots.AxesSubplot at 0x16bc7f08>



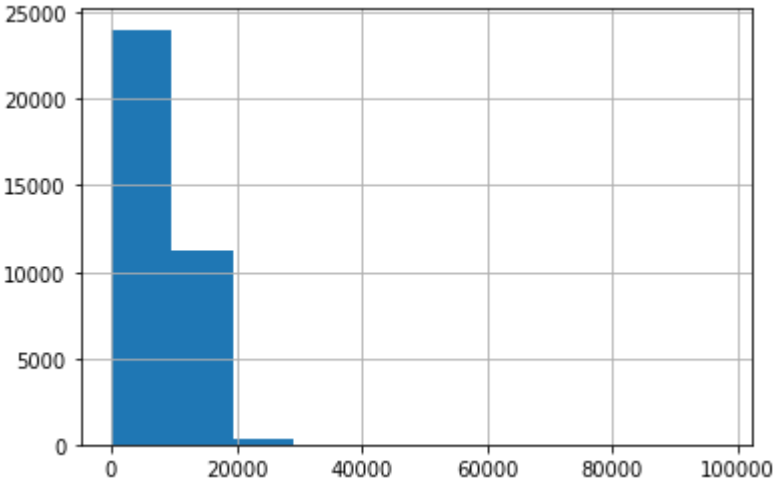
In [125]:
```
1 train['form_field21'].hist()
```

Out[125]: <matplotlib.axes._subplots.AxesSubplot at 0x16b3c388>



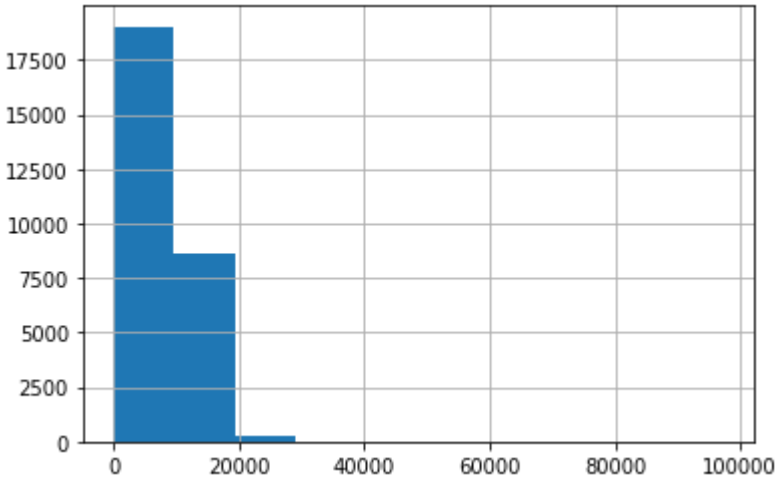
In [126]:
```
1 train['form_field22'].hist()
```

Out[126]: <matplotlib.axes._subplots.AxesSubplot at 0x16caf348>



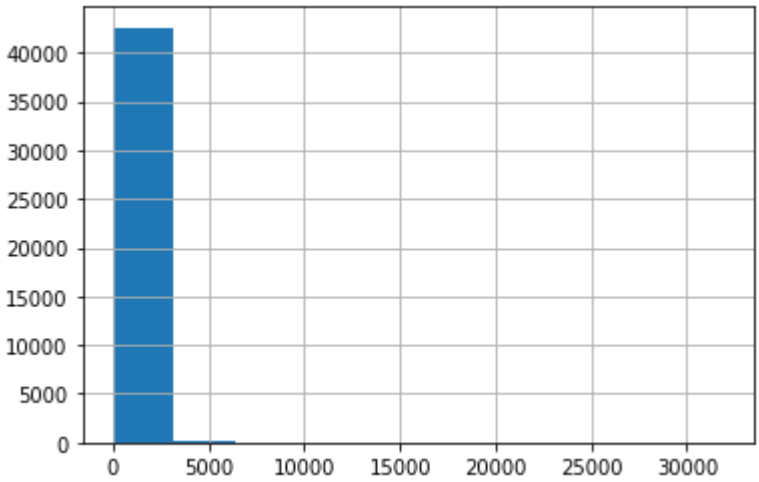
In [127]:
```
1 train['form_field23'].hist()
```

Out[127]: <matplotlib.axes._subplots.AxesSubplot at 0x16d39f48>



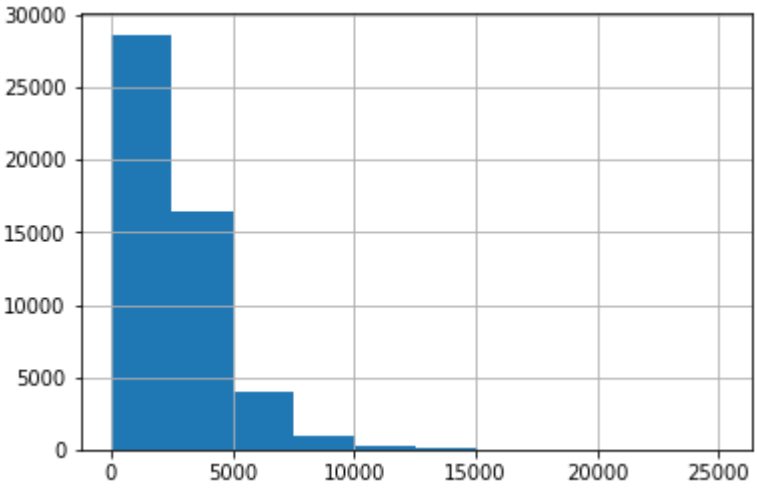
In [128]:
```
1 train['form_field24'].hist()
```

Out[128]: <matplotlib.axes._subplots.AxesSubplot at 0x16d4ed48>



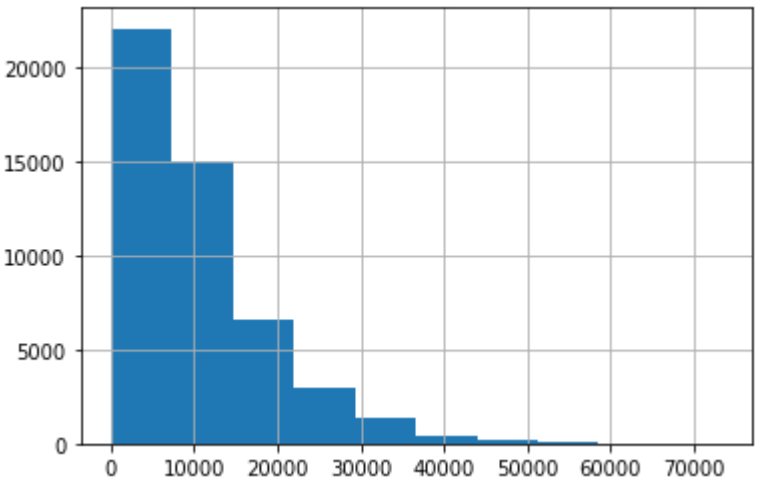
In [129]:
```
1 train['form_field25'].hist()
```

Out[129]: <matplotlib.axes._subplots.AxesSubplot at 0x16e0c1c8>



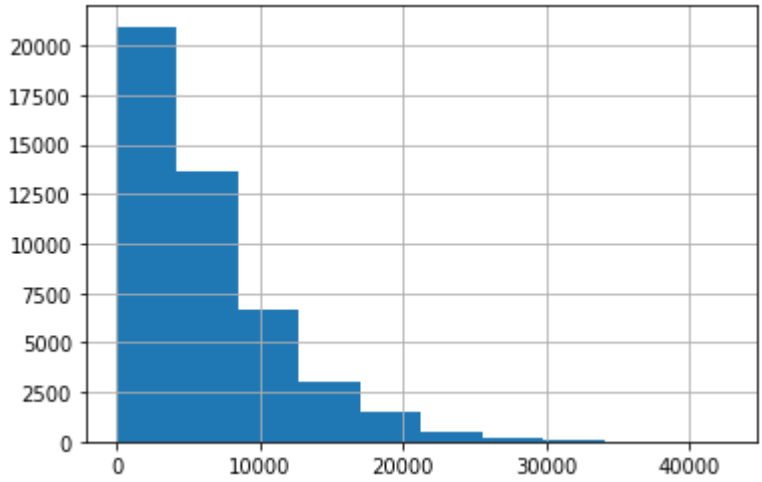
In [130]:
```
1 train['form_field26'].hist()
```

Out[130]: <matplotlib.axes._subplots.AxesSubplot at 0x16e68f88>



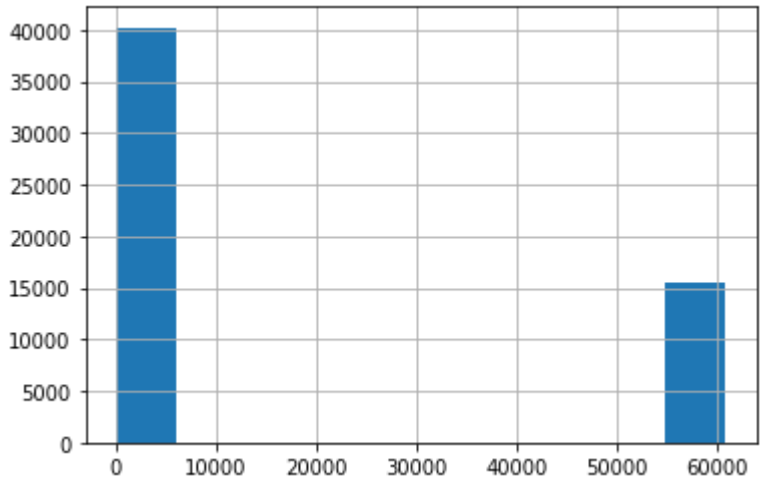
In [131]:
```
1 train['form_field27'].hist()
```

Out[131]: <matplotlib.axes._subplots.AxesSubplot at 0x15e7a608>



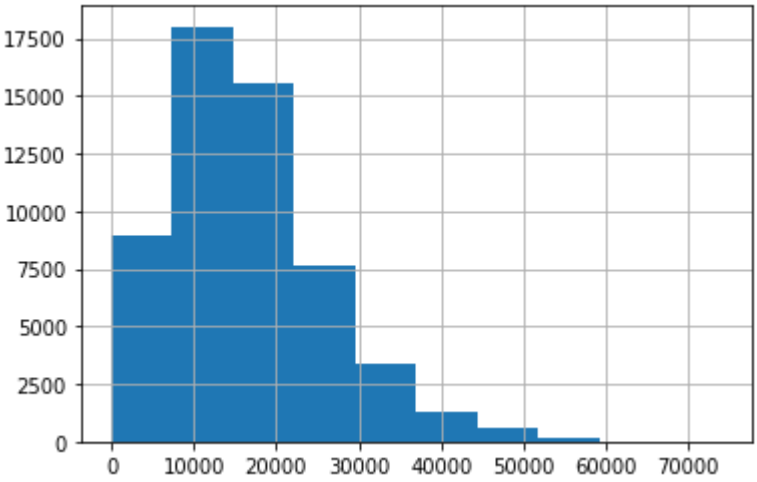
In [132]:
```
1 train['form_field28'].hist()
```

Out[132]: <matplotlib.axes._subplots.AxesSubplot at 0x15b34348>



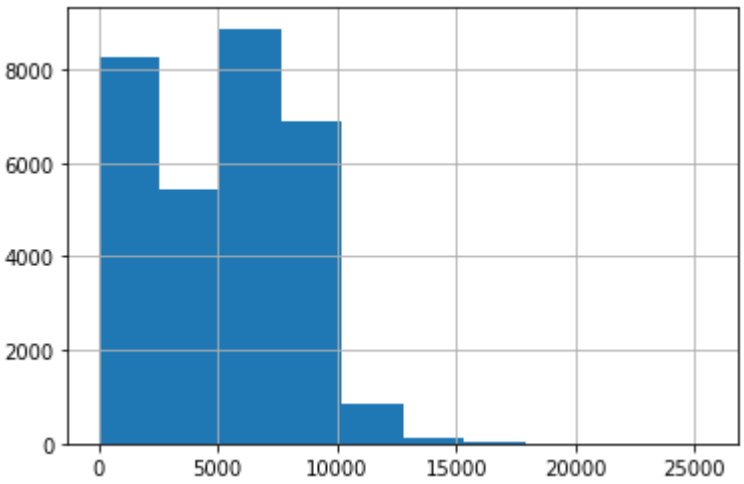
In [133]:
```
1 train['form_field29'].hist()
```

Out[133]: <matplotlib.axes._subplots.AxesSubplot at 0x15c5b688>



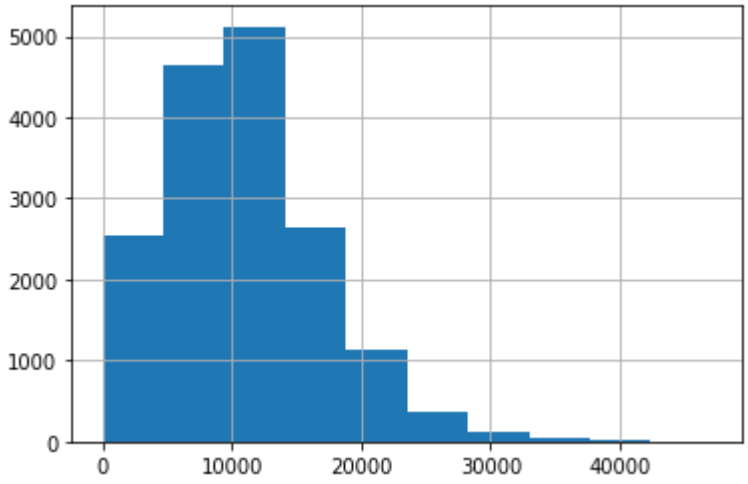
In [134]:
```
1 train['form_field30'].hist()
```

Out[134]: <matplotlib.axes._subplots.AxesSubplot at 0x15c2a6c8>



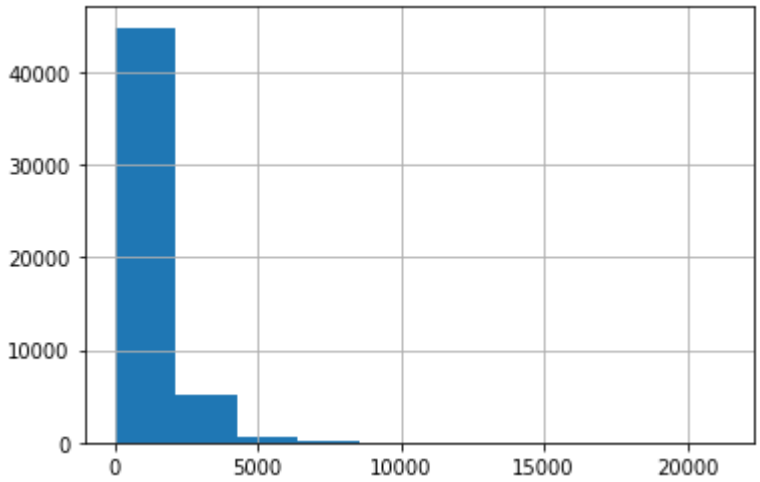
In [135]:
```
1 train['form_field31'].hist()
```

Out[135]: <matplotlib.axes._subplots.AxesSubplot at 0x15b79608>



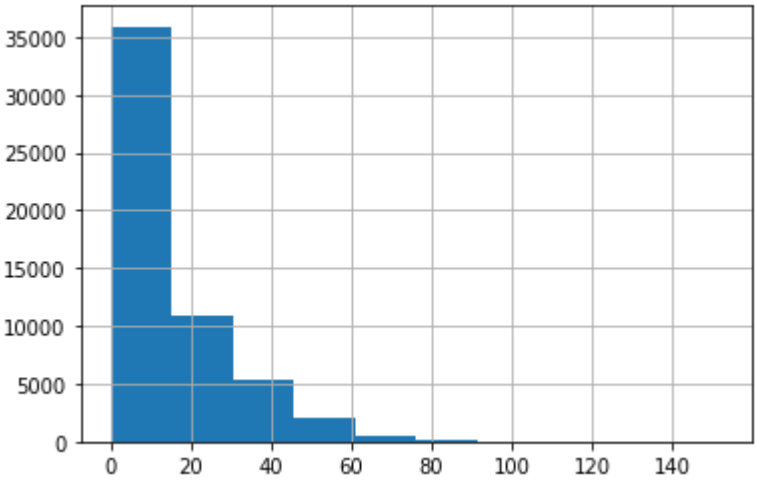
In [136]:
```
1 train['form_field32'].hist()
```

Out[136]: <matplotlib.axes._subplots.AxesSubplot at 0x15beea08>



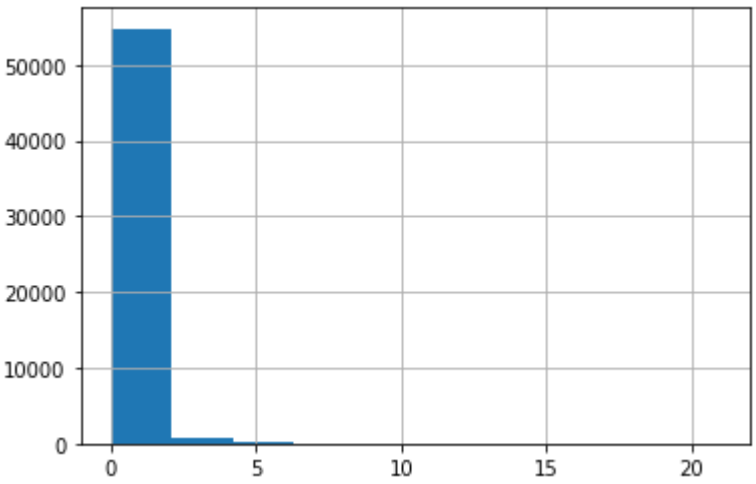
In [137]:
```
1 train['form_field33'].hist()
```

Out[137]: <matplotlib.axes._subplots.AxesSubplot at 0x15def948>



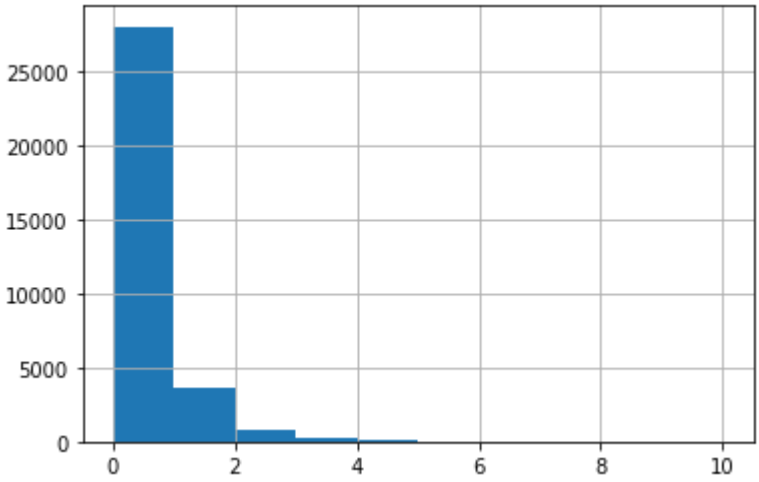
In [138]:
```
1 train['form_field34'].hist()
```

Out[138]: <matplotlib.axes._subplots.AxesSubplot at 0x15ec6548>



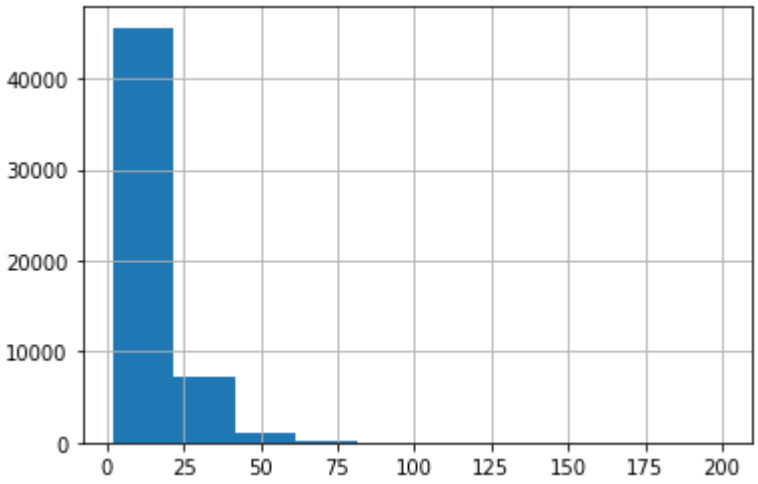
In [139]:
```
1 train['form_field35'].hist()
```

Out[139]: <matplotlib.axes._subplots.AxesSubplot at 0x15f83848>



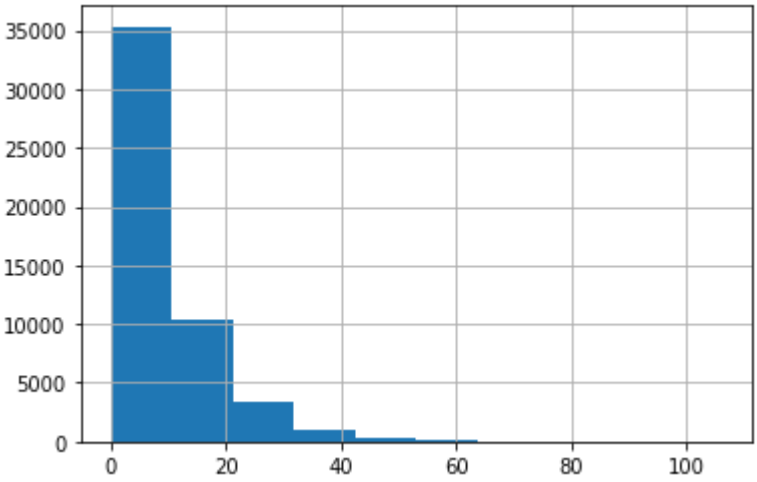
In [140]:
```
1 train['form_field36'].hist()
```

Out[140]: <matplotlib.axes._subplots.AxesSubplot at 0x162c5d48>



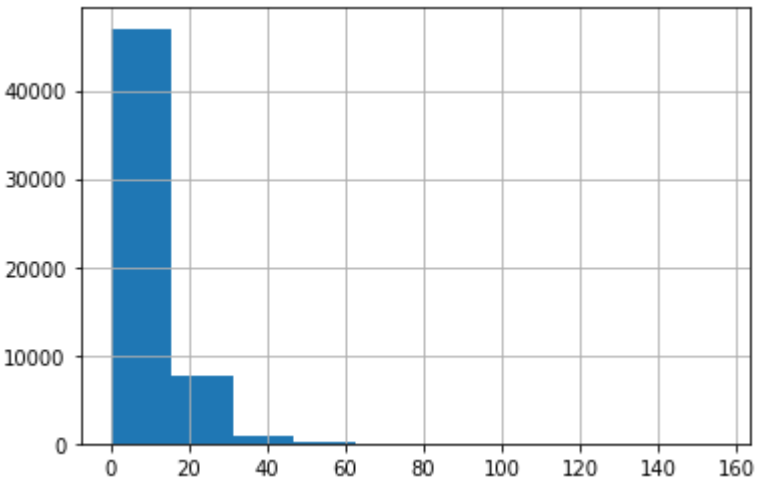
In [141]:
```
1 train['form_field37'].hist()
```

Out[141]: <matplotlib.axes._subplots.AxesSubplot at 0x16f30108>



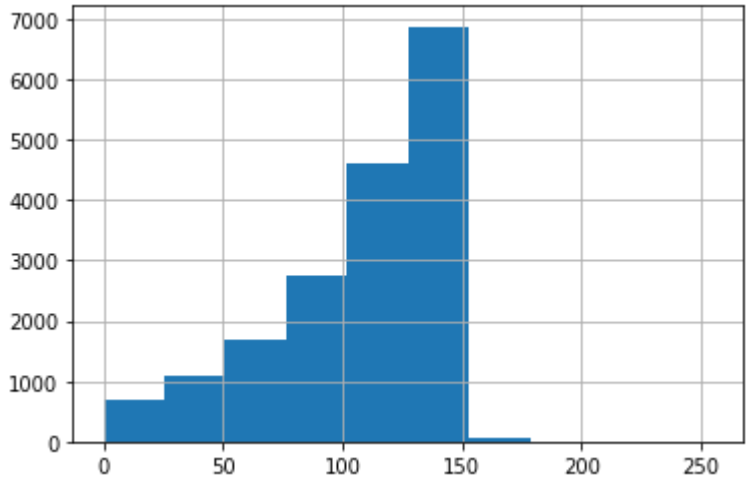
In [142]:
```
1 train['form_field38'].hist()
```

Out[142]: <matplotlib.axes._subplots.AxesSubplot at 0x16f98f88>



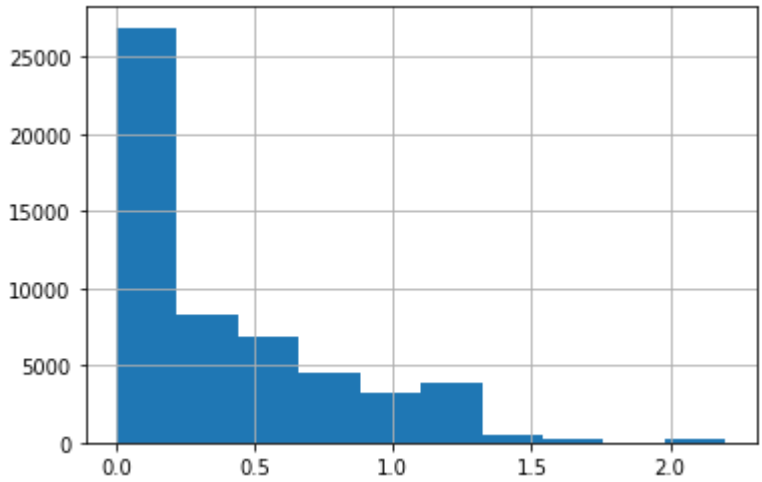
In [143]:
```
1 train['form_field41'].hist()
```

Out[143]: <matplotlib.axes._subplots.AxesSubplot at 0x17043748>



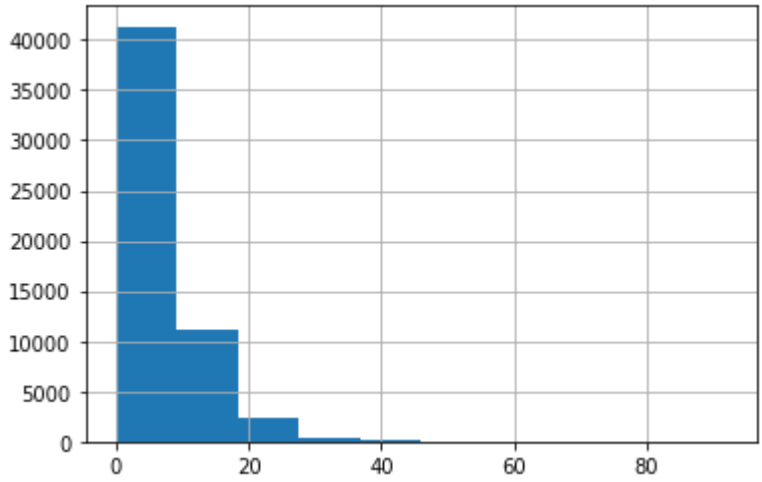
In [144]:
```
1 train['form_field42'].hist()
```

Out[144]: <matplotlib.axes._subplots.AxesSubplot at 0x170b10c8>



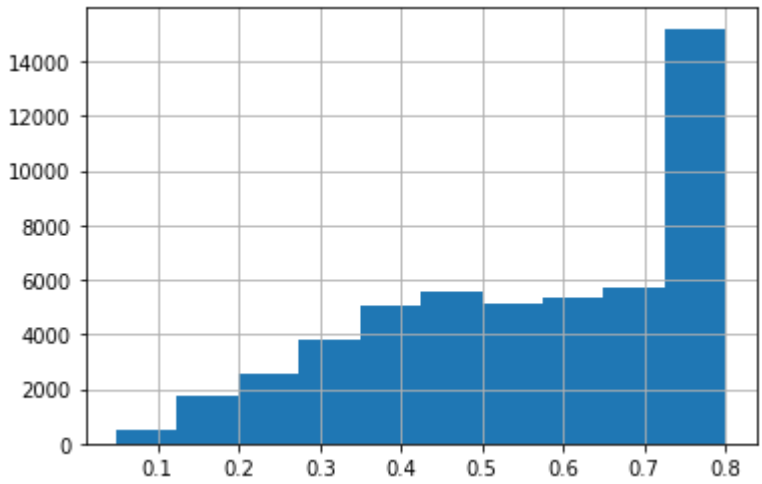
In [145]:
```
1 train['form_field43'].hist()
```

Out[145]: <matplotlib.axes._subplots.AxesSubplot at 0x1713a7c8>



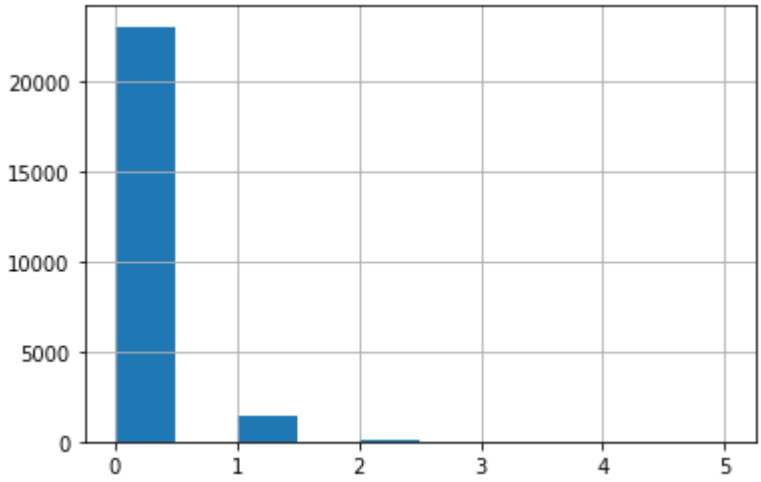
In [146]:
```
1 train['form_field44'].hist()
```

Out[146]: <matplotlib.axes._subplots.AxesSubplot at 0x1713af48>



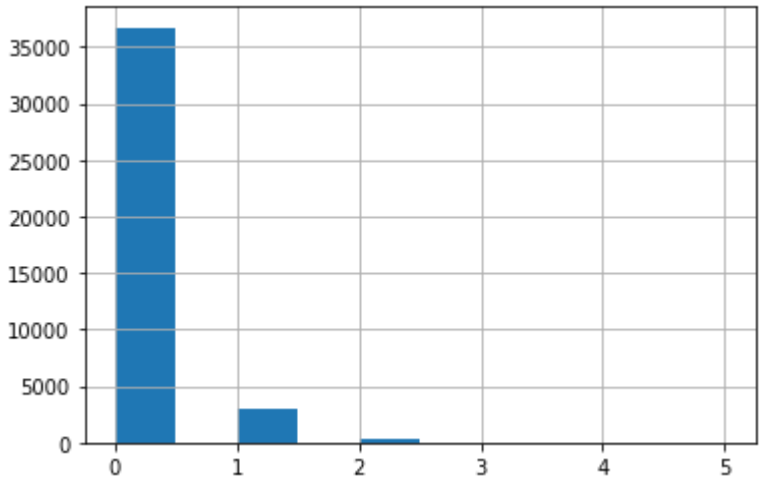
In [147]:
```
1 train['form_field45'].hist()
```

Out[147]: <matplotlib.axes._subplots.AxesSubplot at 0x17262808>



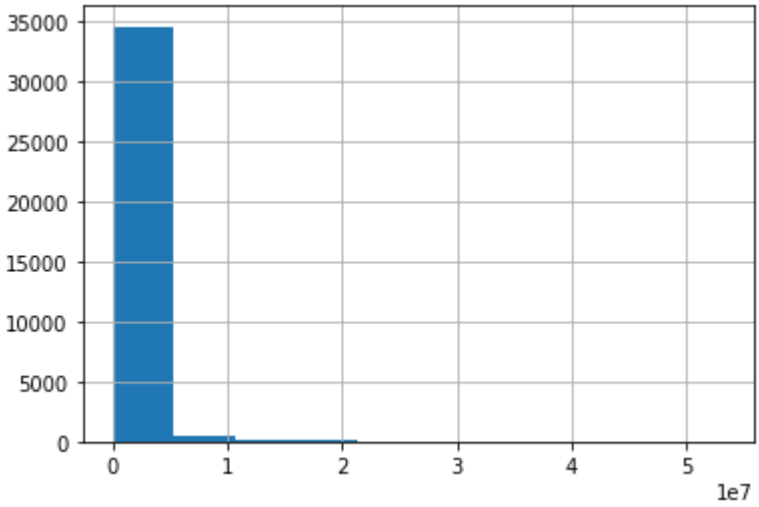
In [148]:
```
1 train['form_field46'].hist()
```

Out[148]: <matplotlib.axes._subplots.AxesSubplot at 0x172d9d88>



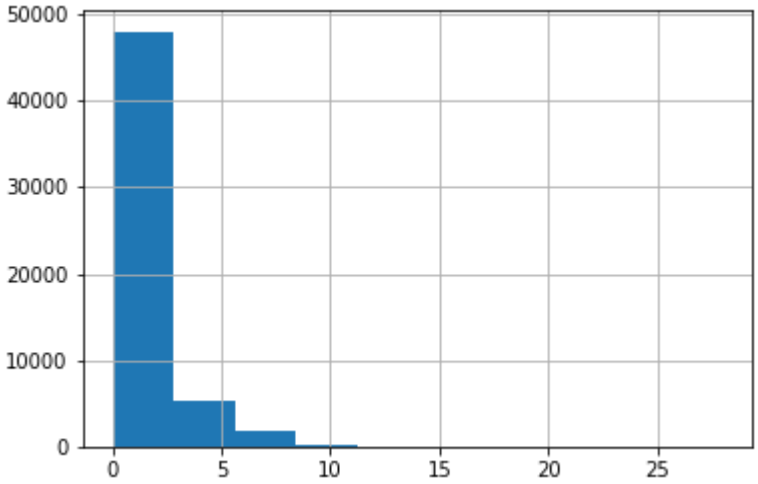
In [149]:
```
1 train['form_field48'].hist()
```

Out[149]: <matplotlib.axes._subplots.AxesSubplot at 0x1735eb08>



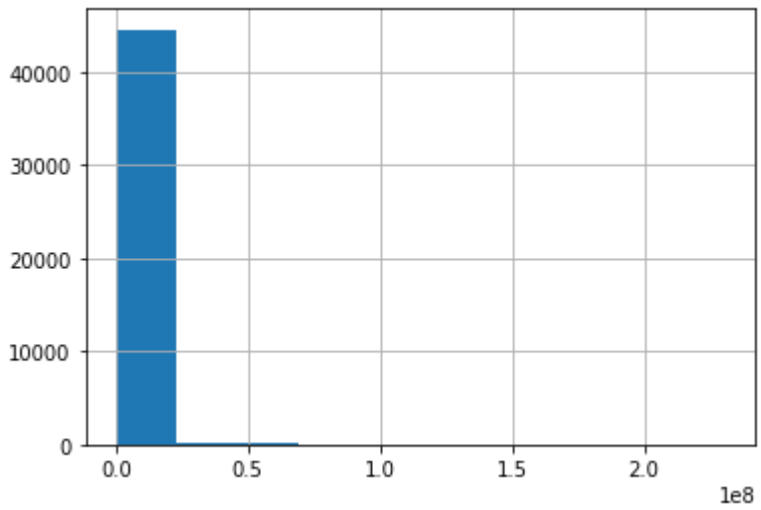
In [150]:
```
1 train['form_field49'].hist()
```

Out[150]: <matplotlib.axes._subplots.AxesSubplot at 0x173c1dc8>



In [158]:
```
1 train['form_field50'].hist()
```

Out[158]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x456c7c48>
```



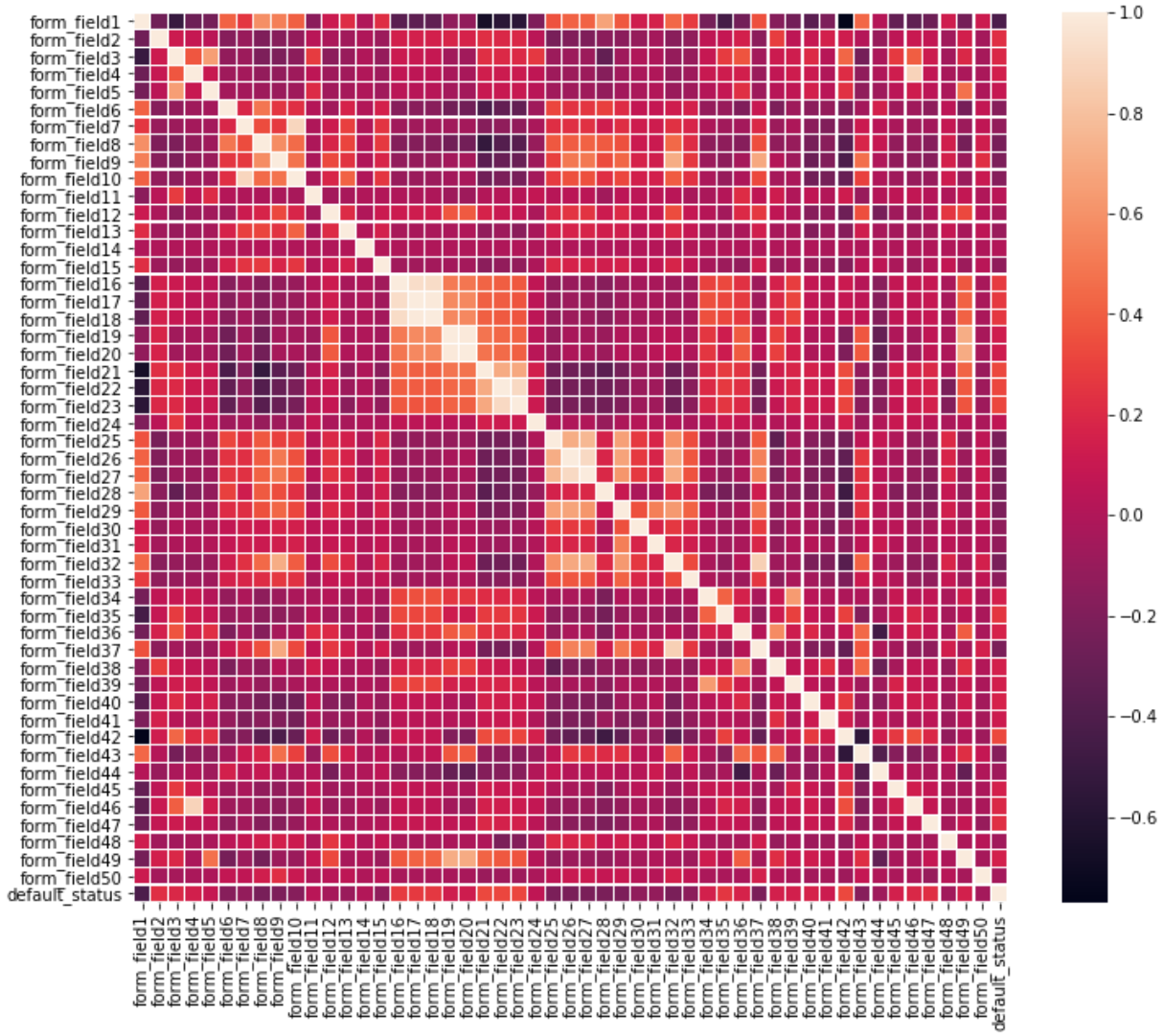
In [156]:
```
1 #features = train.select_dtypes(exclude = object).columns.drop(['default_sta
2 #subplot = plt.subplots(figsize=(20,15))
3 #for i,j in enumerate(features):
4 #    plt.subplot(4,2,i+1)
5 #    plt.subplots_adjust(hspace=0.5)
6 #    sns.countplot(x=j, data=train)
7 #    plt.xticks(rotation=90)
```

In [157]:
```
1 train.columns
```

Out[157]:
```
Index(['Applicant_ID', 'form_field1', 'form_field2', 'form_field3',
       'form_field4', 'form_field5', 'form_field6', 'form_field7',
       'form_field8', 'form_field9', 'form_field10', 'form_field11',
       'form_field12', 'form_field13', 'form_field14', 'form_field15',
       'form_field16', 'form_field17', 'form_field18', 'form_field19',
       'form_field20', 'form_field21', 'form_field22', 'form_field23',
       'form_field24', 'form_field25', 'form_field26', 'form_field27',
       'form_field28', 'form_field29', 'form_field30', 'form_field31',
       'form_field32', 'form_field33', 'form_field34', 'form_field35',
       'form_field36', 'form_field37', 'form_field38', 'form_field39',
       'form_field40', 'form_field41', 'form_field42', 'form_field43',
       'form_field44', 'form_field45', 'form_field46', 'form_field47',
       'form_field48', 'form_field49', 'form_field50', 'default_status'],
      dtype='object')
```

In [152]:
```
corr_mat = train.corr()
f,ax = plt.subplots(figsize = (12,10))
sns.heatmap(corr_mat, ax=ax, linewidths = 0.1)
```

Out[152]: <matplotlib.axes._subplots.AxesSubplot at 0x174de248>



In [198]:
```
#To fill null values
new_features = train.select_dtypes(exclude = object).columns.drop(['default_

for i in new_features:
    train[i].fillna(-999, inplace=True)
    test_given[i].fillna(-999, inplace=True)
```

In [199]:
```
#Performing some operations in the dataset

train['form_field7-8'] = train['form_field7']-train['form_field8']
train['form_field14-13'] = train['form_field14']-train['form_field13']

test_given['form_field7-8'] = test_given['form_field7']-test_given['form_fie
test_given['form_field14-13'] = test_given['form_field14']-test_given['form_

```

In [200]:
```
#Dropping some columns
new_features1 = train.select_dtypes(exclude = object).columns.drop(['default
```

In [201]:
```
test_given_new = test_given[new_features1]
test_given_new.head()
```

Out[201]:

ı_field9	form_field10	...	form_field41	form_field42	form_field43	form_field44	form_field45	form_fie
40710.0	0.0	...	-999.0000	0.825000	1.01	0.800000	-999.0	
10098.0	18308285.0	...	18.8415	0.507694	4.04	0.623248	1.0	
-999.0	0.0	...	-999.0000	-999.000000	0.00	-999.000000	-999.0	-9
24437.0	493641.0	...	-999.0000	0.916663	2.02	0.464224	-999.0	-9
60523.0	5752921.0	...	-999.0000	0.234047	23.23	0.726688	0.0	

In [202]:
```
X = train[new_features1]
Y = train['default_status']

X.head()

```

Out[202]:

	form_field1	form_field2	form_field3	form_field4	form_field5	form_field6	form_field7	form_field
0	3436.0	0.28505	1.6560	0.0	0.000	0.0	10689720.0	252072
1	3456.0	0.67400	0.2342	0.0	0.000	0.0	898979.0	497531
2	3276.0	0.53845	3.1510	0.0	6.282	-999.0	956940.0	-999
3	3372.0	0.17005	0.5050	0.0	0.000	192166.0	3044703.0	385499
4	3370.0	0.77270	1.1010	0.0	0.000	1556.0	214728.0	214728

5 rows × 50 columns

USING CATBOOST CLASSIFIER

In [205]:
```
from sklearn.model_selection import StratifiedKFold

n_splits = 10
params = {'n_estimators':4000, 'learning_rate':0.01, 'max_depth':10, 'object
          'random_seed': 42,'early_stopping_rounds':200,'use_best_model':Tru
fold=StratifiedKFold(n_splits)
```

In [207]:
```
from sklearn.metrics import roc_auc_score

def metric(Y,predictions):
    return roc_auc_score(Y,predictions, labels=[0,1])

from catboost import CatBoostClassifier
from sklearn.metrics import roc_auc_score

scores_list = []
test_oofs = []
score = 0

for i, (tr_index, vr_index) in enumerate(fold.split(X,Y)):
    X_train, Y_train = X.loc[tr_index, new_features1], Y.loc[tr_index]
    X_valid, Y_valid = X.loc[vr_index,new_features1], Y.loc[vr_index]

    cbc= CatBoostClassifier(**params)
    cbc.fit(X_train,Y_train,eval_set=[(X_valid,Y_valid)],verbose=100)

    predictions = cbc.predict_proba(X_valid)[:,1]
    auc = metric(Y_valid, predictions)
    scores_list.append(auc)
    score += auc/n_splits

    test_prediction1 = cbc.predict_proba(test_given_new[new_features1])[:,1]
    test_oofs.append(test_prediction1)
```

```
m 55s
800:    test: 0.8410015 best: 0.8410015 (800)   total: 5m 4s    remaining: 20
m 14s
900:    test: 0.8413199 best: 0.8414056 (879)   total: 5m 41s   remaining: 19
m 33s
1000:   test: 0.8416087 best: 0.8416214 (999)   total: 6m 18s   remaining: 18
m 54s
1100:   test: 0.8418366 best: 0.8418776 (1070)  total: 6m 56s   remaining: 18
m 15s
1200:   test: 0.8418791 best: 0.8419665 (1153)  total: 7m 33s   remaining: 17
m 36s
1300:   test: 0.8419166 best: 0.8420427 (1258)  total: 8m 10s   remaining: 16
m 57s
1400:   test: 0.8421586 best: 0.8421641 (1398)  total: 8m 47s   remaining: 16
m 18s
1500:   test: 0.8421051 best: 0.8421884 (1419)  total: 9m 25s   remaining: 15
m 40s
1600:   test: 0.8421877 best: 0.8422163 (1592)  total: 10m 2s   remaining: 15
m 2s
1700:   test: 0.8421944 best: 0.8423339 (1649)  total: 10m 39s  remaining: 14
```

In [208]:
```
oof_prediction = pd.DataFrame(test_oofs).T
test_oofs
```

Out[208]:
```
[array([0.30064969, 0.39629273, 0.34561782, ..., 0.27259501, 0.53008168,
        0.20333594]),
 array([0.31205168, 0.37870135, 0.34996592, ..., 0.27079236, 0.56755226,
        0.20349947]),
 array([0.32024978, 0.38571125, 0.38667258, ..., 0.24888701, 0.54661887,
        0.1680194 ]),
 array([0.31122036, 0.44493738, 0.37852077, ..., 0.25588928, 0.5123027 ,
        0.21498019]),
 array([0.31929262, 0.43424696, 0.38164075, ..., 0.26533628, 0.56405951,
        0.1992833 ]),
 array([0.31567073, 0.41889806, 0.39495826, ..., 0.27314625, 0.54402306,
        0.20183657]),
 array([0.31360262, 0.38035741, 0.40355185, ..., 0.23128475, 0.54014177,
        0.17963178]),
 array([0.27255024, 0.353732  , 0.39118801, ..., 0.24726046, 0.55516458,
        0.20348043]),
 array([0.28608311, 0.38907548, 0.34261088, ..., 0.25775327, 0.52758092,
        0.16837781]),
 array([0.29685325, 0.36250858, 0.35447588, ..., 0.25554551, 0.56431049,
        0.17796779])]
```

In [210]:
```
sub['default_status'] = np.mean(test_oofs, axis = 0)

sub.to_csv('Catboost_Sub.csv', index=False)
```

In [211]:
```
from sklearn.model_selection import StratifiedKFold

n_splits = 10
params1 = {'n_estimators':4000, 'learning_rate':0.01, 'max_depth':10,
           'seed':42,'num_leaves':18}
fold=StratifiedKFold(n_splits)
```

In [214]:

```
from sklearn.metrics import roc_auc_score

def metric(Y,pred1):
    return roc_auc_score(Y,pred1, labels=[0,1])

from sklearn.metrics import roc_auc_score
import lightgbm
scores_list = []
test_oofs = []
score = 0

for i, (tr_index, vr_index) in enumerate(fold.split(X,Y)):
    X_train, Y_train = X.loc[tr_index, new_features1], Y.loc[tr_index]
    X_valid, Y_valid = X.loc[vr_index,new_features1], Y.loc[vr_index]

    lgm= lightgbm.LGBMClassifier(**params1)
    lgm.fit(X_train,Y_train,eval_set=[(X_valid,Y_valid)],verbose=100)

    pred1 = lgm.predict_proba(X_valid)[:,1]
    auc = metric(Y_valid, pred1)
    scores_list.append(auc)
    score += auc/n_splits

    pred_test1 = lgm.predict_proba(test_given_new[new_features1])[:,1]
    test_oofs.append(pred_test1)
```

```
[1500]  valid_0's binary_logloss: 0.400958
[1600]  valid_0's binary_logloss: 0.40089
[1700]  valid_0's binary_logloss: 0.4008
[1800]  valid_0's binary_logloss: 0.400706
[1900]  valid_0's binary_logloss: 0.400656
[2000]  valid_0's binary_logloss: 0.400563
[2100]  valid_0's binary_logloss: 0.400393
[2200]  valid_0's binary_logloss: 0.400303
[2300]  valid_0's binary_logloss: 0.400222
[2400]  valid_0's binary_logloss: 0.400169
[2500]  valid_0's binary_logloss: 0.40017
[2600]  valid_0's binary_logloss: 0.400137
[2700]  valid_0's binary_logloss: 0.400233
[2800]  valid_0's binary_logloss: 0.400442
[2900]  valid_0's binary_logloss: 0.400471
[3000]  valid_0's binary_logloss: 0.400612
[3100]  valid_0's binary_logloss: 0.400772
[3200]  valid_0's binary_logloss: 0.40089
[3300]  valid_0's binary_logloss: 0.400996
```

In [215]:
```
oof_prediction1 = pd.DataFrame(test_oofs).T
test_oofs
```

Out[215]:
```
[array([0.30469309, 0.25447681, 0.3481874 , ..., 0.24746639, 0.52314119,
        0.17937296]),
 array([0.28708262, 0.30436785, 0.38946688, ..., 0.2154949 , 0.51350394,
        0.14166591]),
 array([0.35096576, 0.23800835, 0.4174566 , ..., 0.24020376, 0.54250429,
        0.20220741]),
 array([0.30620504, 0.26059375, 0.42210557, ..., 0.21022249, 0.47013571,
        0.17564038]),
 array([0.35195616, 0.33717443, 0.45785169, ..., 0.23272943, 0.50223915,
        0.17171385]),
 array([0.37921353, 0.24830387, 0.41576192, ..., 0.25018693, 0.5387501 ,
        0.15351318]),
 array([0.35236506, 0.28356338, 0.39992498, ..., 0.22726671, 0.53045979,
        0.15572922]),
 array([0.275134  , 0.25313952, 0.39561433, ..., 0.22570366, 0.55438874,
        0.17585192]),
 array([0.28519913, 0.20056165, 0.36741162, ..., 0.21936776, 0.50512513,
        0.15778132]),
 array([0.32532081, 0.16358596, 0.37690145, ..., 0.25431869, 0.53507163,
        0.19013936])]
```

In [217]:
```
sub['default_status'] = np.mean(test_oofs, axis = 0)

sub.to_csv('Lightgbm_Sub.csv', index=False)
```

BLENDING OF THE TWO RESULTS

In [218]:
```
pred_1= pd.read_csv('Lightgbm_Sub.csv')['default_status']
pred_2 = pd.read_csv('Catboost_Sub.csv')['default_status']
#pred3 = pd.read_csv('Real3.csv')['default_status']
```

In [220]:
```
sub_pred = (pred_1*0.4) + (pred_2*0.6)
sub['default_status'] = sub_pred
sub.to_csv('Blended_submission.csv', index = False)
```

In [ ]:
```

```