

HW1_Problem1

September 5, 2018

1 Problem 1

1.0.1 Air Pressure System Failure in Scania Trucks

```
In [1]: import pandas as pd
import numpy as np
import os
from fancyimpute import KNN
# reading testing and training
root_path = os.path.join(os.getcwd(), "DataSet") # root path
path_test = os.path.join(root_path, "aps_failure_test_set.csv") # path of the testing s
path_training = os.path.join(root_path, "aps_failure_training_set.csv") # path of the t

#Reading data sets
print(root_path)
df_test = pd.read_csv(path_test)
df_training = pd.read_csv(path_training)
df_training
```

Using TensorFlow backend.

/Users/francisco/Desktop/Org_Desktop/Fall_2018/Data_Mining/HW/HW1/ipython/DataSet

```
Out[1]:
```

	class	aa_000	ab_000	ac_000	ad_000	ae_000	af_000	ag_000	ag_001	\
0	neg	76698	na	2130706438	280	0	0	0	0	
1	neg	33058	na	0	na	0	0	0	0	
2	neg	41040	na	228	100	0	0	0	0	
3	neg	12	0	70	66	0	10	0	0	
4	neg	60874	na	1368	458	0	0	0	0	
5	neg	38312	na	2130706432	218	0	0	0	0	
6	neg	14	0	6	na	0	0	0	0	
7	neg	102960	na	2130706432	116	0	0	0	0	
8	neg	78696	na	0	na	0	0	0	0	
9	pos	153204	0	182	na	0	0	0	0	
10	neg	39196	na	204	170	0	0	0	0	
11	neg	45912	na	0	454	0	0	0	0	

12	neg	2104	na	36	26	0	0	0	0
13	neg	118950	na	1390	1298	0	0	0	0
14	neg	24416	na	0	na	0	0	0	0
15	neg	14	0	62	34	0	0	0	0
16	neg	31300	0	784	740	0	0	0	0
17	neg	736	2	24	22	16	20	0	0
18	neg	332	na	2130706432	20	0	0	0	0
19	neg	1432	na	2130706440	82	0	0	0	0
20	neg	41212	0	2130706434	104	104	172	0	0
21	neg	14	na	6	6	0	0	0	0
22	neg	157128	na	2130706456	424	0	0	0	0
23	pos	453236	na	2926	na	0	0	0	0
24	neg	58246	na	2130706432	2416	0	0	0	0
25	neg	29394	na	0	na	0	0	0	0
26	neg	8690	na	476	364	0	0	0	0
27	neg	46978	na	334	322	0	0	0	0
28	neg	1870	na	na	na	0	0	0	0
29	neg	12516	0	120	na	0	0	0	0
...
59970	neg	16	0	8	na	0	0	0	0
59971	neg	39026	na	202	168	0	0	0	0
59972	neg	14	na	24	20	0	0	0	0
59973	neg	3248	8	16	10	0	0	0	0
59974	neg	83818	na	552	532	0	0	0	0
59975	neg	40274	na	98	94	0	0	0	0
59976	neg	16978	na	2130706434	1750	0	0	0	0
59977	neg	30320	na	1838	1278	0	0	0	0
59978	neg	38414	na	888	758	0	0	0	0
59979	neg	18	0	2130706432	18	0	0	0	0
59980	neg	562	0	4	4	0	0	0	0
59981	neg	16	na	20	20	0	0	0	0
59982	neg	10628	na	2130706434	98	0	0	0	0
59983	neg	39004	na	90	84	0	0	0	0
59984	neg	33386	na	812	744	0	0	0	0
59985	neg	10792	na	784	608	0	0	0	0
59986	neg	644	na	12	na	0	0	0	0
59987	neg	41330	na	2130706432	744	0	0	0	0
59988	neg	6078	na	52	46	0	0	0	0
59989	neg	61478	na	134	0	0	0	0	0
59990	neg	81354	na	2130706432	156	0	0	0	0
59991	neg	39308	na	452	374	0	0	0	0
59992	neg	14	0	2	2	0	0	0	0
59993	neg	0	2	0	0	0	0	0	0
59994	neg	32	0	0	na	0	0	0	0
59995	neg	153002	na	664	186	0	0	0	0
59996	neg	2286	na	2130706538	224	0	0	0	0
59997	neg	112	0	2130706432	18	0	0	0	0
59998	neg	80292	na	2130706432	494	0	0	0	0

59999	neg	40222	na	698	628	0	0	0	0
	ag_002	...	ee_002	ee_003	ee_004	ee_005	ee_006	ee_007	\
0	0	...	1240520	493384	721044	469792	339156	157956	
1	0	...	421400	178064	293306	245416	133654	81140	
2	0	...	277378	159812	423992	409564	320746	158022	
3	0	...	240	46	58	44	10	0	
4	0	...	622012	229790	405298	347188	286954	311560	
5	0	...	388574	288278	900430	300412	1534	338	
6	0	...	168	48	60	28	0	0	
7	0	...	715518	384948	915978	1052166	1108672	341532	
8	0	...	699290	362510	1190028	1012704	160090	63216	
9	0	...	129862	26872	34044	22472	34362	0	
10	0	...	198386	99614	215734	189966	219996	208724	
11	0	...	495400	278660	544710	438256	249386	73356	
12	9744	...	4776	2330	4634	6534	29980	39372	
13	0	...	1677646	1380302	2531322	426284	17996	11734	
14	0	...	190606	95674	207476	232470	154962	99406	
15	0	...	248	42	48	28	32	10	
16	0	...	798872	112724	51736	7054	6628	27600	
17	0	...	1370	812	2488	2340	2566	20870	
18	0	...	1260	296	878	11774	8	0	
19	0	...	2714	1480	4822	53694	1116	16	
20	0	...	171390	95844	177206	173184	404690	736142	
21	0	...	80	0	0	0	0	0	
22	0	...	1784108	683976	1228736	971366	838200	559330	
23	222	...	7908038	3026002	5025350	2025766	1160638	533834	
24	0	...	670578	293816	464286	336782	229078	163348	
25	0	...	234998	136388	470258	548540	8384	524	
26	0	...	52700	28890	55588	85370	140892	10658	
27	0	...	407008	221836	534336	606554	325542	89624	
28	0	...	7688	5050	33502	40466	1436	34	
29	0	...	133648	53830	92568	72416	70266	111176	
...	
59970	0	...	78	30	56	28	34	42	
59971	0	...	277756	155262	419110	597950	306722	16868	
59972	0	...	388	56	30	0	0	0	
59973	0	...	10464	6202	11372	24842	25718	58364	
59974	0	...	713474	343926	681198	532938	376156	262990	
59975	0	...	313326	182178	466482	587110	271706	38270	
59976	0	...	167080	87242	141522	147958	103198	68364	
59977	0	...	153770	86110	246966	274254	220696	143046	
59978	0	...	391604	150648	269002	200140	149768	117020	
59979	0	...	56	16	34	22	86	10	
59980	0	...	2880	644	1110	526	1162	14756	
59981	0	...	256	10	30	48	30	0	
59982	0	...	553156	99122	21720	7124	1464	616	
59983	0	...	337238	179550	349658	263388	164592	116170	

59984	0	...	265982	135322	255766	208478	188012	241194
59985	0	...	155506	49046	98602	55236	15122	10052
59986	0	...	5466	2164	6338	5008	5328	4730
59987	0	...	434454	165964	497548	392222	256944	80572
59988	0	...	20474	12750	52018	190226	3504	28
59989	0	...	632856	410326	1130874	535990	205766	39694
59990	0	...	627610	376028	931828	1035236	674626	161266
59991	0	...	312426	161694	323438	292350	250112	157636
59992	0	...	132	44	80	52	54	0
59993	0	...	0	0	0	0	0	0
59994	0	...	258	98	224	166	36	486
59995	0	...	998500	566884	1290398	1218244	1019768	717762
59996	0	...	10578	6760	21126	68424	136	0
59997	0	...	792	386	452	144	146	2622
59998	0	...	699352	222654	347378	225724	194440	165070
59999	0	...	440066	183200	344546	254068	225148	158304

	ee_008	ee_009	ef_000	eg_000
0	73224	0	0	0
1	97576	1500	0	0
2	95128	514	0	0
3	0	0	4	32
4	433954	1218	0	0
5	856	0	0	0
6	0	0	0	0
7	129504	7832	0	0
8	41202	4	0	0
9	0	0	0	0
10	463790	34054	0	0
11	51344	2008	0	0
12	0	0	0	0
13	12506	138	0	0
14	121544	2712	0	0
15	0	0	0	0
16	2	2	0	0
17	38	0	0	0
18	0	0	0	0
19	0	0	0	0
20	9268	6	0	0
21	0	0	0	0
22	826630	18434	0	0
23	493800	6914	0	0
24	372036	31152	0	0
25	208	16	0	0
26	10944	356	0	0
27	6538	10	0	0
28	0	0	0	0
29	592	0	0	0

...
59970	0	0	0	0
59971	18330	2112	0	0
59972	0	0	0	0
59973	120	0	0	0
59974	628964	82484	0	0
59975	6414	242	0	0
59976	56256	1128	0	0
59977	192332	5316	0	0
59978	270838	21082	0	0
59979	0	0	0	0
59980	128	0	0	0
59981	0	0	0	0
59982	114	0	0	0
59983	284594	33338	0	0
59984	138900	3388	0	0
59985	18538	856	0	0
59986	0	0	0	0
59987	2414	0	0	0
59988	14	0	0	0
59989	550	0	0	0
59990	3284	0	0	0
59991	234932	9506	0	0
59992	0	0	0	0
59993	0	0	0	0
59994	0	0	0	0
59995	898642	28588	0	0
59996	0	0	0	0
59997	0	0	0	0
59998	802280	388422	0	0
59999	170384	158	0	0

[60000 rows x 171 columns]

1.1 Visualizing distribution of missing values

```
In [2]: import missingno as msno #library used to visualize missing values. (reusing code)
df_test_replaced_nan = df_test.replace(to_replace='na',value=np.nan) # replacing 'na' with np.nan
df_training_replaced_nan = df_training.replace(to_replace='na',value=np.nan)

print("Test Data Set Missing Values")
print(df_test_replaced_nan.info())
msno.matrix(df_test_replaced_nan)
```

```
Test Data Set Missing Values
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16000 entries, 0 to 15999
Columns: 171 entries, class to eg_000
```

```
dtypes: int64(1), object(170)
memory usage: 20.9+ MB
None
```

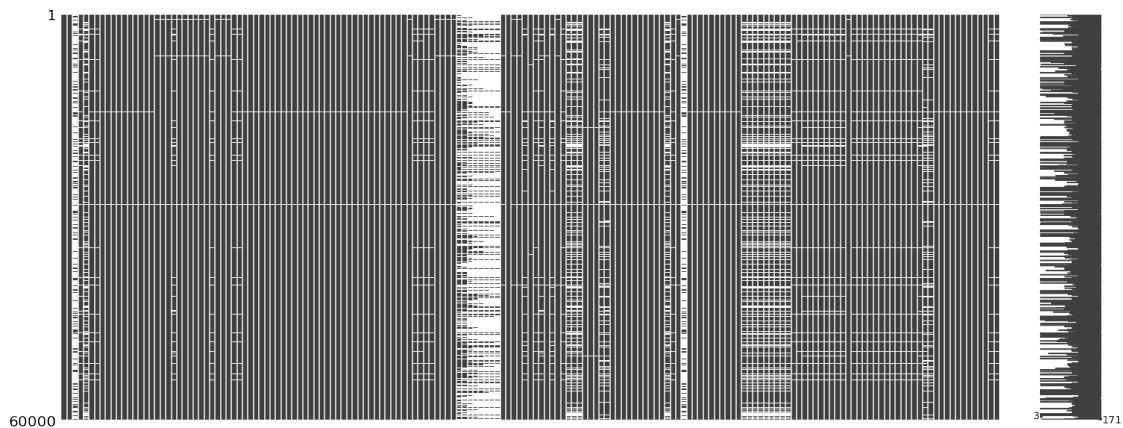
```
Out[2]: <matplotlib.axes._subplots.AxesSubplot at 0x12352cf50>
```

```
In [3]: print("Training Data Set Missing Values")
        print(df_training_replaced_nan.describe())
        print(df_training_replaced_nan.shape)
        msno.matrix(df_training_replaced_nan)
```

Training Data Set Missing Values

```
          aa_000
count  6.000000e+04
mean   5.933650e+04
std    1.454301e+05
min     0.000000e+00
25%     8.340000e+02
50%     3.077600e+04
75%     4.866800e+04
max     2.746564e+06
(60000, 171)
```

```
Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x128dc6d90>
```



1.2 Missing value total

```
In [4]: # function for counting number of na
        #parameters: DataFrame
        def count_na(df):
            total = 0
```

```

    for col in df:
        # If you wish to see all of the missing values for every column then uncomment
        #print("-----")
        #print("Column "+col+" na total: ")
        #print(df[col].isnull().sum())
        total+=df[col].isnull().sum()
        #print("-----")
        #print("\n")
    return total
test_total_missing_values = count_na(df_test_replaced_nan)
training_total_missing_values = count_na(df_training_replaced_nan)
print("The total number of missing values for the testing set is: %d" %test_total_miss)
print("The total number of missing values for the training set is: %d" %training_total)

```

The total number of missing values for the testing set is: 228680
The total number of missing values for the training set is: 850015

1.3 Percentages Function

```

In [5]: from collections import Counter
        # This function gives the percentage of a certain val in a column
        #Parameters:
        #df --> Data Frame
        # col --> column of interest
        # val --> value to count occurrences
        def col_val_percentage(df,col,val=None): # calculate the percentage of missing values
            total = len(df[col])
            if(val == None): # Missing values
                num_val = df[col].isnull().sum()
            else: # Other values
                num_val = Counter(df[col])[val]
            return (num_val/(1.0*total))

        #df_training_replaced_nan['ab_000'].isnull().sum()/(1.0*len(df_training_replaced_nan['ab_000']))
        #col_val_percentage(df_test_replaced_nan,'ab_000','0')

```

1.4 Eliminate columns given percentages

```

In [6]: def get_drop_col(df,na_percent =0.25,zero_percent=0.5):
        drop_ls = [] # list of columns to drop
        for col in df:
            # getting percentages for every column
            (col_na_percent,col_zero_percent) = (col_val_percentage(df,col),col_val_percentage(df,col,'0'))

            if(col_na_percent > na_percent or col_zero_percent > zero_percent):
                drop_ls.append(col)
        print((col_na_percent,col_zero_percent,col_na_percent > na_percent or col_zero_percent > zero_percent))

```

```
return(drop_ls)
```

```
In [7]: df_training_replaced_nan_copy= df_training_replaced_nan.copy()  
        print(df_training_replaced_nan.shape)  
        dropped_ls = get_drop_col(df_training_replaced_nan)  
        dropped_ls  
        df_dropped = df_training_replaced_nan.drop(dropped_ls,axis=1)  
        df_dropped.shape
```

```
(60000, 171)  
(0.0, 0.0, False)  
(0.0, 0.0, False)  
(0.77215, 0.18295, True)  
(0.05558333333333333, 0.14586666666666667, False)  
(0.24768333333333334, 0.03348333333333333, False)  
(0.041666666666666664, 0.9257166666666666, True)  
(0.041666666666666664, 0.9246, True)  
(0.011183333333333333, 0.98555, True)  
(0.011183333333333333, 0.97645, True)  
(0.011183333333333333, 0.93635, True)  
(0.011183333333333333, 0.7815666666666666, True)  
(0.011183333333333333, 0.02175, False)  
(0.011183333333333333, 0.03235, False)  
(0.011183333333333333, 0.020633333333333333, False)  
(0.011183333333333333, 0.26458333333333334, False)  
(0.011183333333333333, 0.41818333333333335, False)  
(0.011183333333333333, 0.6783, True)  
(0.01075, 0.0022166666666666667, False)  
(0.010483333333333334, 0.8931333333333333, True)  
(0.010483333333333334, 0.7819833333333334, True)  
(0.07333333333333333, 0.92045, True)  
(0.0107, 0.62455, True)  
(0.010483333333333334, 0.6215833333333334, True)  
(0.0107, 0.00215, False)  
(0.009816666666666666, 0.0023333333333333335, False)  
(0.0107, 0.0018833333333333334, False)  
(0.009816666666666666, 0.0038333333333333333, False)  
(0.04538333333333333, 0.9154666666666667, True)  
(0.010483333333333334, 0.9891666666666666, True)  
(0.010483333333333334, 0.8899166666666667, True)  
(0.010483333333333334, 0.9885, True)  
(0.041666666666666664, 0.1668, False)  
(0.041683333333333336, 0.16985, False)  
(0.011183333333333333, 0.97935, True)  
(0.011183333333333333, 0.9698666666666667, True)  
(0.011183333333333333, 0.9693666666666667, True)
```


(0.011183333333333333, 0.9679666666666666, True)
 (0.011183333333333333, 0.9526333333333333, True)
 (0.011183333333333333, 0.5056333333333334, True)
 (0.011183333333333333, 0.2767333333333333, False)
 (0.011183333333333333, 0.12386666666666667, False)
 (0.011183333333333333, 0.0824, False)
 (0.011183333333333333, 0.97981666666666667, True)
 (0.011183333333333333, 0.018266666666666667, False)
 (0.011183333333333333, 0.0243, False)
 (0.011183333333333333, 0.021333333333333333, False)
 (0.011183333333333333, 0.019733333333333332, False)
 (0.011183333333333333, 0.019116666666666667, False)
 (0.011183333333333333, 0.018816666666666666, False)
 (0.011183333333333333, 0.14453333333333335, False)
 (0.011183333333333333, 0.75156666666666667, True)
 (0.011183333333333333, 0.9037, True)
 (0.011183333333333333, 0.95741666666666667, True)
 (0.0114666666666666667, 0.01815, False)
 (0.0114666666666666667, 0.018983333333333335, False)
 (0.0114666666666666667, 0.019533333333333333, False)
 (0.0114666666666666667, 0.020066666666666667, False)
 (0.0114666666666666667, 0.0206, False)
 (0.0114666666666666667, 0.022133333333333335, False)
 (0.0114666666666666667, 0.033283333333333333, False)
 (0.0114666666666666667, 0.13396666666666668, False)
 (0.0114666666666666667, 0.41423333333333334, False)
 (0.0114666666666666667, 0.66031666666666667, True)
 (0.01075, 0.0017833333333333334, False)
 (0.045416666666666667, 0.3216, False)
 (0.04545, 0.06716666666666667, False)
 (0.041716666666666666, 0.11585, False)
 (0.041666666666666664, 0.42156666666666665, False)
 (0.0107, 0.0022166666666666667, False)
 (0.0107, 0.00215, False)
 (0.009816666666666666, 0.00195, False)
 (0.009816666666666666, 0.0022833333333333334, False)
 (0.3839, 0.008366666666666666, True)
 (0.45461666666666667, 0.016933333333333335, True)
 (0.65915, 0.020233333333333332, True)
 (0.73348333333333334, 0.0237, True)
 (0.77221666666666667, 0.02505, True)
 (0.7956666666666666, 0.026866666666666667, True)
 (0.81203333333333334, 0.0289, True)
 (0.8210666666666666, 0.030216666666666666, True)
 (0.0121, 0.002, False)
 (0.0027833333333333334, 0.0021833333333333333, False)
 (0.011516666666666666, 0.0017833333333333334, False)
 (0.011516666666666666, 0.0017833333333333334, False)

(0.05428333333333336, 0.0, False)
(0.00788333333333332, 0.0204, False)
(0.04538333333333333, 0.23068333333333332, False)
(0.0726, 0.001816666666666667, False)
(0.0121, 0.001816666666666667, False)
(0.05425, 0.01463333333333333, False)
(0.01126666666666666, 0.0, False)
(0.0417, 0.19741666666666666, False)
(0.24768333333333334, 0.3578, False)
(0.24768333333333334, 0.04858333333333333, False)
(0.24768333333333334, 0.75215, True)
(0.00563333333333333, 0.003716666666666667, False)
(0.00563333333333333, 0.7853666666666667, True)
(0.00563333333333333, 0.00323333333333333, False)
(0.15921666666666667, 0.6131166666666666, True)
(0.16461666666666666, 0.40358333333333335, False)
(0.01145, 0.955, True)
(0.01145, 0.8410666666666666, True)
(0.01145, 0.5089666666666667, True)
(0.01145, 0.04323333333333333, False)
(0.01145, 0.01845, False)
(0.01145, 0.01958333333333335, False)
(0.01145, 0.04543333333333333, False)
(0.01145, 0.15481666666666666, False)
(0.01145, 0.25645, False)
(0.01145, 0.4329666666666667, False)
(0.24768333333333334, 0.2132166666666667, False)
(0.0454, 0.15895, False)
(0.01151666666666666, 0.001783333333333334, False)
(0.77215, 0.22653333333333334, True)
(0.01115, 0.01816666666666668, False)
(0.01115, 0.02091666666666667, False)
(0.01115, 0.01946666666666667, False)
(0.01115, 0.02008333333333335, False)
(0.01115, 0.02038333333333333, False)
(0.01115, 0.01883333333333334, False)
(0.01115, 0.0187, False)
(0.01115, 0.08511666666666666, False)
(0.01115, 0.23405, False)
(0.01115, 0.94045, True)
(0.23013333333333333, 0.03643333333333333, False)
(0.23013333333333333, 0.036, False)
(0.23013333333333333, 0.01715, False)
(0.23013333333333333, 0.0273, False)
(0.23013333333333333, 0.6584833333333333, True)
(0.23013333333333333, 0.1519, False)
(0.23013333333333333, 0.76075, True)
(0.23013333333333333, 0.46191666666666664, False)

```

(0.2301333333333333, 0.01715, False)
(0.041716666666666666, 0.014433333333333333, False)
(0.0454, 0.014333333333333333, False)
(0.0668, 0.9155, True)
(0.0668, 0.9070666666666667, True)
(0.0668, 0.8168, True)
(0.06676666666666667, 0.7607833333333334, True)
(0.06678333333333333, 0.9299333333333333, True)
(0.06678333333333333, 0.9279166666666666, True)
(0.0668, 0.9294833333333333, True)
(0.06681666666666666, 0.9289, True)
(0.011516666666666666, 0.0017833333333333334, False)
(0.0454, 0.22145, False)
(0.04543333333333333, 0.2125, False)
(0.04543333333333333, 0.7400666666666667, True)
(0.04543333333333333, 0.7377833333333333, True)
(0.04545, 0.044366666666666665, False)
(0.04545, 0.04251666666666667, False)
(0.04543333333333333, 0.05466666666666667, False)
(0.04543333333333333, 0.05463333333333333, False)
(0.04538333333333333, 0.6593, True)
(0.0454, 0.6044833333333334, True)
(0.04538333333333333, 0.9497333333333333, True)
(0.04538333333333333, 0.94225, True)
(0.06678333333333333, 0.3574833333333333, False)
(0.17065, 0.021266666666666666, False)
(0.15921666666666667, 0.023633333333333333, False)
(0.011183333333333333, 0.01815, False)
(0.011183333333333333, 0.019283333333333333, False)
(0.011183333333333333, 0.022733333333333335, False)
(0.011183333333333333, 0.02595, False)
(0.011183333333333333, 0.02995, False)
(0.011183333333333333, 0.0469, False)
(0.011183333333333333, 0.0743, False)
(0.011183333333333333, 0.13163333333333332, False)
(0.011183333333333333, 0.288, False)
(0.011183333333333333, 0.53105, True)
(0.0454, 0.95035, True)
(0.04538333333333333, 0.9465666666666667, True)

```

Out[7]: (60000, 108)

1.5 Percentile and Imputation Methods

```

In [8]: def drop_percentile(df,per):
        for col in df:
            #df[col].quantile(per)

```

```

        d_p = df[df[col] <= df[col].quantile(per)]
    return d_p
d = df_training_replaced_nan.drop('class',axis=1)
d =drop_percentile(d.astype(np.float),0.99)

df_training_modified = pd.DataFrame(d)
#df_training_modified
df_training_modified_copy1= df_training_modified.copy(deep=True)
#df_training_modified_copy2= df_training_modified.copy(deep=True)
#df_training_modified_copy3= df_training_modified.copy(deep=True)
#Imputation

def impute_mean(df):
    for col in df:
        df_prime=df[col].dropna() # drop columns
        mean = np.mean(df_prime)
        df_prime = df[col].replace(to_replace=np.nan,value=mean)
        df[col]=df_prime
    return df

#imputation using median
def impute_median(df):
    for col in df:
        # if not the class column
        df_prime=df[col].dropna() # drop columns
        median = np.median(df_prime)
        df_prime = df[col].replace(to_replace=np.nan,value=median)
    return df
def knn_impute(df,k):
    df_filled = pd.DataFrame(KNN(k).complete(df))
    return df_filled

df_mean_imputed = impute_mean(df_training_modified)
df_median_imputed = impute_median(df_training_modified)
df_training_replaced_nan
#df_knn_imputed = knn_impute(df_training_modified_copy3,3)

```

```

Out[8]:
   class aa_000 ab_000 ac_000 ad_000 ae_000 af_000 ag_000 ag_001 \
0    neg  76698   NaN 2130706438   280      0      0      0      0
1    neg  33058   NaN      0    NaN      0      0      0      0
2    neg  41040   NaN   228   100      0      0      0      0
3    neg    12     0    70    66      0     10      0      0
4    neg  60874   NaN  1368   458      0      0      0      0
5    neg  38312   NaN 2130706432   218      0      0      0      0
6    neg    14     0     6    NaN      0      0      0      0

```

7	neg	102960	NaN	2130706432	116	0	0	0	0
8	neg	78696	NaN	0	NaN	0	0	0	0
9	pos	153204	0	182	NaN	0	0	0	0
10	neg	39196	NaN	204	170	0	0	0	0
11	neg	45912	NaN	0	454	0	0	0	0
12	neg	2104	NaN	36	26	0	0	0	0
13	neg	118950	NaN	1390	1298	0	0	0	0
14	neg	24416	NaN	0	NaN	0	0	0	0
15	neg	14	0	62	34	0	0	0	0
16	neg	31300	0	784	740	0	0	0	0
17	neg	736	2	24	22	16	20	0	0
18	neg	332	NaN	2130706432	20	0	0	0	0
19	neg	1432	NaN	2130706440	82	0	0	0	0
20	neg	41212	0	2130706434	104	104	172	0	0
21	neg	14	NaN	6	6	0	0	0	0
22	neg	157128	NaN	2130706456	424	0	0	0	0
23	pos	453236	NaN	2926	NaN	0	0	0	0
24	neg	58246	NaN	2130706432	2416	0	0	0	0
25	neg	29394	NaN	0	NaN	0	0	0	0
26	neg	8690	NaN	476	364	0	0	0	0
27	neg	46978	NaN	334	322	0	0	0	0
28	neg	1870	NaN	NaN	NaN	0	0	0	0
29	neg	12516	0	120	NaN	0	0	0	0
...
59970	neg	16	0	8	NaN	0	0	0	0
59971	neg	39026	NaN	202	168	0	0	0	0
59972	neg	14	NaN	24	20	0	0	0	0
59973	neg	3248	8	16	10	0	0	0	0
59974	neg	83818	NaN	552	532	0	0	0	0
59975	neg	40274	NaN	98	94	0	0	0	0
59976	neg	16978	NaN	2130706434	1750	0	0	0	0
59977	neg	30320	NaN	1838	1278	0	0	0	0
59978	neg	38414	NaN	888	758	0	0	0	0
59979	neg	18	0	2130706432	18	0	0	0	0
59980	neg	562	0	4	4	0	0	0	0
59981	neg	16	NaN	20	20	0	0	0	0
59982	neg	10628	NaN	2130706434	98	0	0	0	0
59983	neg	39004	NaN	90	84	0	0	0	0
59984	neg	33386	NaN	812	744	0	0	0	0
59985	neg	10792	NaN	784	608	0	0	0	0
59986	neg	644	NaN	12	NaN	0	0	0	0
59987	neg	41330	NaN	2130706432	744	0	0	0	0
59988	neg	6078	NaN	52	46	0	0	0	0
59989	neg	61478	NaN	134	0	0	0	0	0
59990	neg	81354	NaN	2130706432	156	0	0	0	0
59991	neg	39308	NaN	452	374	0	0	0	0
59992	neg	14	0	2	2	0	0	0	0
59993	neg	0	2	0	0	0	0	0	0

59994	neg	32	0	0	NaN	0	0	0	0
59995	neg	153002	NaN	664	186	0	0	0	0
59996	neg	2286	NaN	2130706538	224	0	0	0	0
59997	neg	112	0	2130706432	18	0	0	0	0
59998	neg	80292	NaN	2130706432	494	0	0	0	0
59999	neg	40222	NaN	698	628	0	0	0	0

	ag_002	...	ee_002	ee_003	ee_004	ee_005	ee_006	ee_007	\
0	0	...	1240520	493384	721044	469792	339156	157956	
1	0	...	421400	178064	293306	245416	133654	81140	
2	0	...	277378	159812	423992	409564	320746	158022	
3	0	...	240	46	58	44	10	0	
4	0	...	622012	229790	405298	347188	286954	311560	
5	0	...	388574	288278	900430	300412	1534	338	
6	0	...	168	48	60	28	0	0	
7	0	...	715518	384948	915978	1052166	1108672	341532	
8	0	...	699290	362510	1190028	1012704	160090	63216	
9	0	...	129862	26872	34044	22472	34362	0	
10	0	...	198386	99614	215734	189966	219996	208724	
11	0	...	495400	278660	544710	438256	249386	73356	
12	9744	...	4776	2330	4634	6534	29980	39372	
13	0	...	1677646	1380302	2531322	426284	17996	11734	
14	0	...	190606	95674	207476	232470	154962	99406	
15	0	...	248	42	48	28	32	10	
16	0	...	798872	112724	51736	7054	6628	27600	
17	0	...	1370	812	2488	2340	2566	20870	
18	0	...	1260	296	878	11774	8	0	
19	0	...	2714	1480	4822	53694	1116	16	
20	0	...	171390	95844	177206	173184	404690	736142	
21	0	...	80	0	0	0	0	0	
22	0	...	1784108	683976	1228736	971366	838200	559330	
23	222	...	7908038	3026002	5025350	2025766	1160638	533834	
24	0	...	670578	293816	464286	336782	229078	163348	
25	0	...	234998	136388	470258	548540	8384	524	
26	0	...	52700	28890	55588	85370	140892	10658	
27	0	...	407008	221836	534336	606554	325542	89624	
28	0	...	7688	5050	33502	40466	1436	34	
29	0	...	133648	53830	92568	72416	70266	111176	
...	
59970	0	...	78	30	56	28	34	42	
59971	0	...	277756	155262	419110	597950	306722	16868	
59972	0	...	388	56	30	0	0	0	
59973	0	...	10464	6202	11372	24842	25718	58364	
59974	0	...	713474	343926	681198	532938	376156	262990	
59975	0	...	313326	182178	466482	587110	271706	38270	
59976	0	...	167080	87242	141522	147958	103198	68364	
59977	0	...	153770	86110	246966	274254	220696	143046	
59978	0	...	391604	150648	269002	200140	149768	117020	

59979	0	...	56	16	34	22	86	10
59980	0	...	2880	644	1110	526	1162	14756
59981	0	...	256	10	30	48	30	0
59982	0	...	553156	99122	21720	7124	1464	616
59983	0	...	337238	179550	349658	263388	164592	116170
59984	0	...	265982	135322	255766	208478	188012	241194
59985	0	...	155506	49046	98602	55236	15122	10052
59986	0	...	5466	2164	6338	5008	5328	4730
59987	0	...	434454	165964	497548	392222	256944	80572
59988	0	...	20474	12750	52018	190226	3504	28
59989	0	...	632856	410326	1130874	535990	205766	39694
59990	0	...	627610	376028	931828	1035236	674626	161266
59991	0	...	312426	161694	323438	292350	250112	157636
59992	0	...	132	44	80	52	54	0
59993	0	...	0	0	0	0	0	0
59994	0	...	258	98	224	166	36	486
59995	0	...	998500	566884	1290398	1218244	1019768	717762
59996	0	...	10578	6760	21126	68424	136	0
59997	0	...	792	386	452	144	146	2622
59998	0	...	699352	222654	347378	225724	194440	165070
59999	0	...	440066	183200	344546	254068	225148	158304

	ee_008	ee_009	ef_000	eg_000
0	73224	0	0	0
1	97576	1500	0	0
2	95128	514	0	0
3	0	0	4	32
4	433954	1218	0	0
5	856	0	0	0
6	0	0	0	0
7	129504	7832	0	0
8	41202	4	0	0
9	0	0	0	0
10	463790	34054	0	0
11	51344	2008	0	0
12	0	0	0	0
13	12506	138	0	0
14	121544	2712	0	0
15	0	0	0	0
16	2	2	0	0
17	38	0	0	0
18	0	0	0	0
19	0	0	0	0
20	9268	6	0	0
21	0	0	0	0
22	826630	18434	0	0
23	493800	6914	0	0
24	372036	31152	0	0

25	208	16	0	0
26	10944	356	0	0
27	6538	10	0	0
28	0	0	0	0
29	592	0	0	0
...
59970	0	0	0	0
59971	18330	2112	0	0
59972	0	0	0	0
59973	120	0	0	0
59974	628964	82484	0	0
59975	6414	242	0	0
59976	56256	1128	0	0
59977	192332	5316	0	0
59978	270838	21082	0	0
59979	0	0	0	0
59980	128	0	0	0
59981	0	0	0	0
59982	114	0	0	0
59983	284594	33338	0	0
59984	138900	3388	0	0
59985	18538	856	0	0
59986	0	0	0	0
59987	2414	0	0	0
59988	14	0	0	0
59989	550	0	0	0
59990	3284	0	0	0
59991	234932	9506	0	0
59992	0	0	0	0
59993	0	0	0	0
59994	0	0	0	0
59995	898642	28588	0	0
59996	0	0	0	0
59997	0	0	0	0
59998	802280	388422	0	0
59999	170384	158	0	0

[60000 rows x 171 columns]

1.6 Imputation

```
In [15]: knn_impute(df_training_modified_copy1[0:10000],3)
```

Imputing row 1/10000 with 5 missing, elapsed time: 119.658

Imputing row 101/10000 with 7 missing, elapsed time: 119.731

Imputing row 201/10000 with 8 missing, elapsed time: 119.812

Imputing row 301/10000 with 10 missing, elapsed time: 119.891

Imputing row 401/10000 with 5 missing, elapsed time: 119.964

Imputing row 501/10000 with 9 missing, elapsed time: 120.027
Imputing row 601/10000 with 24 missing, elapsed time: 120.094
Imputing row 701/10000 with 22 missing, elapsed time: 120.166
Imputing row 801/10000 with 6 missing, elapsed time: 120.225
Imputing row 901/10000 with 8 missing, elapsed time: 120.285
Imputing row 1001/10000 with 34 missing, elapsed time: 120.347
Imputing row 1101/10000 with 11 missing, elapsed time: 120.427
Imputing row 1201/10000 with 24 missing, elapsed time: 120.500
Imputing row 1301/10000 with 3 missing, elapsed time: 120.571
Imputing row 1401/10000 with 8 missing, elapsed time: 120.640
Imputing row 1501/10000 with 8 missing, elapsed time: 120.706
Imputing row 1601/10000 with 10 missing, elapsed time: 120.777
Imputing row 1701/10000 with 10 missing, elapsed time: 120.849
Imputing row 1801/10000 with 6 missing, elapsed time: 120.915
Imputing row 1901/10000 with 8 missing, elapsed time: 120.980
Imputing row 2001/10000 with 6 missing, elapsed time: 121.055
Imputing row 2101/10000 with 8 missing, elapsed time: 121.118
Imputing row 2201/10000 with 11 missing, elapsed time: 121.189
Imputing row 2301/10000 with 8 missing, elapsed time: 121.251
Imputing row 2401/10000 with 8 missing, elapsed time: 121.330
Imputing row 2501/10000 with 2 missing, elapsed time: 121.396
Imputing row 2601/10000 with 4 missing, elapsed time: 121.469
Imputing row 2701/10000 with 8 missing, elapsed time: 121.555
Imputing row 2801/10000 with 10 missing, elapsed time: 121.629
Imputing row 2901/10000 with 7 missing, elapsed time: 121.685
Imputing row 3001/10000 with 10 missing, elapsed time: 121.754
Imputing row 3101/10000 with 8 missing, elapsed time: 121.832
Imputing row 3201/10000 with 6 missing, elapsed time: 121.895
Imputing row 3301/10000 with 6 missing, elapsed time: 121.960
Imputing row 3401/10000 with 9 missing, elapsed time: 122.041
Imputing row 3501/10000 with 8 missing, elapsed time: 122.111
Imputing row 3601/10000 with 33 missing, elapsed time: 122.174
Imputing row 3701/10000 with 8 missing, elapsed time: 122.242
Imputing row 3801/10000 with 20 missing, elapsed time: 122.316
Imputing row 3901/10000 with 8 missing, elapsed time: 122.388
Imputing row 4001/10000 with 7 missing, elapsed time: 122.461
Imputing row 4101/10000 with 8 missing, elapsed time: 122.534
Imputing row 4201/10000 with 10 missing, elapsed time: 122.606
Imputing row 4301/10000 with 10 missing, elapsed time: 122.671
Imputing row 4401/10000 with 22 missing, elapsed time: 122.735
Imputing row 4501/10000 with 7 missing, elapsed time: 122.813
Imputing row 4601/10000 with 6 missing, elapsed time: 122.880
Imputing row 4701/10000 with 21 missing, elapsed time: 122.947
Imputing row 4801/10000 with 8 missing, elapsed time: 123.014
Imputing row 4901/10000 with 9 missing, elapsed time: 123.082
Imputing row 5001/10000 with 24 missing, elapsed time: 123.150
Imputing row 5101/10000 with 8 missing, elapsed time: 123.218
Imputing row 5201/10000 with 10 missing, elapsed time: 123.284

Imputing row 5301/10000 with 24 missing, elapsed time: 123.369
Imputing row 5401/10000 with 6 missing, elapsed time: 123.434
Imputing row 5501/10000 with 10 missing, elapsed time: 123.516
Imputing row 5601/10000 with 22 missing, elapsed time: 123.588
Imputing row 5701/10000 with 11 missing, elapsed time: 123.664
Imputing row 5801/10000 with 9 missing, elapsed time: 123.726
Imputing row 5901/10000 with 8 missing, elapsed time: 123.793
Imputing row 6001/10000 with 8 missing, elapsed time: 123.862
Imputing row 6101/10000 with 8 missing, elapsed time: 123.929
Imputing row 6201/10000 with 10 missing, elapsed time: 124.008
Imputing row 6301/10000 with 8 missing, elapsed time: 124.090
Imputing row 6401/10000 with 16 missing, elapsed time: 124.161
Imputing row 6501/10000 with 10 missing, elapsed time: 124.228
Imputing row 6601/10000 with 2 missing, elapsed time: 124.299
Imputing row 6701/10000 with 8 missing, elapsed time: 124.369
Imputing row 6801/10000 with 8 missing, elapsed time: 124.439
Imputing row 6901/10000 with 10 missing, elapsed time: 124.531
Imputing row 7001/10000 with 7 missing, elapsed time: 124.619
Imputing row 7101/10000 with 9 missing, elapsed time: 124.718
Imputing row 7201/10000 with 8 missing, elapsed time: 124.812
Imputing row 7301/10000 with 8 missing, elapsed time: 124.908
Imputing row 7401/10000 with 8 missing, elapsed time: 124.997
Imputing row 7501/10000 with 9 missing, elapsed time: 125.095
Imputing row 7601/10000 with 22 missing, elapsed time: 125.185
Imputing row 7701/10000 with 8 missing, elapsed time: 125.272
Imputing row 7801/10000 with 9 missing, elapsed time: 125.363
Imputing row 7901/10000 with 9 missing, elapsed time: 125.432
Imputing row 8001/10000 with 7 missing, elapsed time: 125.512
Imputing row 8101/10000 with 6 missing, elapsed time: 125.590
Imputing row 8201/10000 with 9 missing, elapsed time: 125.671
Imputing row 8301/10000 with 8 missing, elapsed time: 125.754
Imputing row 8401/10000 with 8 missing, elapsed time: 125.840
Imputing row 8501/10000 with 10 missing, elapsed time: 125.928
Imputing row 8601/10000 with 10 missing, elapsed time: 125.993
Imputing row 8701/10000 with 2 missing, elapsed time: 126.077
Imputing row 8801/10000 with 14 missing, elapsed time: 126.144
Imputing row 8901/10000 with 6 missing, elapsed time: 126.235
Imputing row 9001/10000 with 3 missing, elapsed time: 126.369
Imputing row 9101/10000 with 24 missing, elapsed time: 126.438
Imputing row 9201/10000 with 8 missing, elapsed time: 126.556
Imputing row 9301/10000 with 8 missing, elapsed time: 126.707
Imputing row 9401/10000 with 8 missing, elapsed time: 126.772
Imputing row 9501/10000 with 10 missing, elapsed time: 126.858
Imputing row 9601/10000 with 8 missing, elapsed time: 126.928
Imputing row 9701/10000 with 8 missing, elapsed time: 127.009
Imputing row 9801/10000 with 11 missing, elapsed time: 127.083
Imputing row 9901/10000 with 10 missing, elapsed time: 127.151

```

Out [15]:
      0      1      2      3      4      5      6      7      \
0    76698.0  0.000000  2.130706e+09  280.000000  0.0  0.0  0.0  0.0
1    33058.0  0.000000  0.000000e+00  202.179617  0.0  0.0  0.0  0.0
2    41040.0  0.000000  2.280000e+02  100.000000  0.0  0.0  0.0  0.0
3    60874.0  0.000000  1.368000e+03  458.000000  0.0  0.0  0.0  0.0
4    38312.0  0.000000  2.130706e+09  218.000000  0.0  0.0  0.0  0.0
5      14.0  0.000000  6.000000e+00   9.548815  0.0  0.0  0.0  0.0
6   102960.0  0.000000  2.130706e+09  116.000000  0.0  0.0  0.0  0.0
7    78696.0  0.000000  0.000000e+00  318.858178  0.0  0.0  0.0  0.0
8   153204.0  0.000000  1.820000e+02  571.839738  0.0  0.0  0.0  0.0
9    39196.0  2.006923  2.040000e+02  170.000000  0.0  0.0  0.0  0.0
10   45912.0  1.297001  0.000000e+00  454.000000  0.0  0.0  0.0  0.0
11    2104.0  0.000000  3.600000e+01   26.000000  0.0  0.0  0.0  0.0
12   118950.0  0.000000  1.390000e+03  1298.000000  0.0  0.0  0.0  0.0
13    24416.0  1.320639  0.000000e+00  101.883984  0.0  0.0  0.0  0.0
14      14.0  0.000000  6.200000e+01   34.000000  0.0  0.0  0.0  0.0
15   31300.0  0.000000  7.840000e+02  740.000000  0.0  0.0  0.0  0.0
16     736.0  2.000000  2.400000e+01   22.000000  16.0  20.0  0.0  0.0
17     332.0  1.298821  2.130706e+09   20.000000  0.0  0.0  0.0  0.0
18    1432.0  1.511578  2.130706e+09   82.000000  0.0  0.0  0.0  0.0
19   41212.0  0.000000  2.130706e+09  104.000000  104.0  172.0  0.0  0.0
20      14.0  2.000000  6.000000e+00    6.000000  0.0  0.0  0.0  0.0
21   157128.0  0.906431  2.130706e+09   424.000000  0.0  0.0  0.0  0.0
22  453236.0  0.000000  2.926000e+03  1436.822953  0.0  0.0  0.0  0.0
23   58246.0  2.102336  2.130706e+09  2416.000000  0.0  0.0  0.0  0.0
24   29394.0  1.411876  0.000000e+00   312.387672  0.0  0.0  0.0  0.0
25    8690.0  0.728846  4.760000e+02   364.000000  0.0  0.0  0.0  0.0
26   46978.0  1.487173  3.340000e+02   322.000000  0.0  0.0  0.0  0.0
27    1870.0  1.412071  6.238866e+08    64.495403  0.0  0.0  0.0  0.0
28   12516.0  0.000000  1.200000e+02    31.121676  0.0  0.0  0.0  0.0
29   86240.0  22.845158  1.520000e+02   146.000000  0.0  0.0  0.0  0.0
...      ...      ...      ...      ...      ...      ...      ...
9970  39192.0  0.000000  1.960000e+02   194.000000  0.0  0.0  0.0  0.0
9971  41198.0  1.199743  1.254000e+03  1096.000000  0.0  0.0  0.0  0.0
9972  31074.0  0.000000  3.108000e+03  2874.000000  0.0  0.0  0.0  0.0
9973     28.0  0.659314  2.800000e+01   26.000000  0.0  0.0  0.0  0.0
9974  117644.0  1.499645  0.000000e+00  1416.641064  0.0  0.0  0.0  0.0
9975   52432.0  0.000000  5.300000e+02   496.000000  0.0  0.0  0.0  0.0
9976   1418.0  0.692615  9.600000e+01    96.000000  0.0  0.0  0.0  0.0
9977   5580.0  3.964341  0.000000e+00   142.128082  0.0  0.0  0.0  0.0
9978  38270.0  0.000000  2.130706e+09   110.000000  0.0  0.0  0.0  0.0
9979  40994.0  0.000000  0.000000e+00   453.884465  0.0  0.0  0.0  0.0
9980     340.0  6.000000  4.200000e+01    22.000000  0.0  0.0  0.0  0.0
9981     726.0  0.000000  2.130706e+09    14.000000  0.0  0.0  0.0  0.0
9982   31660.0  2.452672  2.860000e+02   286.000000  0.0  0.0  0.0  0.0
9983      8.0  0.000000  1.600000e+01    16.000000  0.0  0.0  0.0  0.0
9984   30746.0  0.503362  2.130706e+09   888.000000  0.0  0.0  0.0  0.0
9985   41612.0  0.537650  2.130706e+09   356.000000  0.0  0.0  0.0  0.0

```

9986	308726.0	0.598090	3.026000e+03	890.261512	0.0	0.0	0.0	0.0
9987	30392.0	1.607311	0.000000e+00	512.383717	0.0	0.0	0.0	0.0
9988	165068.0	0.000000	0.000000e+00	2981.407782	0.0	0.0	0.0	0.0
9989	39750.0	0.000000	2.130706e+09	412.000000	0.0	0.0	0.0	0.0
9990	41618.0	1.952613	0.000000e+00	175.558617	0.0	0.0	0.0	0.0
9991	34360.0	0.000000	7.600000e+02	0.000000	0.0	0.0	0.0	0.0
9992	1750.0	0.590932	4.400000e+01	44.000000	0.0	0.0	0.0	0.0
9993	11184.0	6.415715	1.500000e+02	300.962286	0.0	0.0	0.0	0.0
9994	1212.0	0.000000	1.000000e+01	6.000000	0.0	0.0	0.0	0.0
9995	57798.0	0.000000	0.000000e+00	163.070190	0.0	0.0	0.0	0.0
9996	26262.0	0.000000	1.840000e+02	182.000000	0.0	0.0	0.0	0.0
9997	28.0	0.000000	4.000000e+01	38.000000	0.0	0.0	0.0	0.0
9998	116656.0	3.277061	0.000000e+00	818.753106	0.0	0.0	0.0	0.0
9999	2008.0	0.000000	7.600000e+01	70.000000	0.0	0.0	0.0	0.0

	8	9	...	160	161	162	163	\
0	0.0	0.0	...	1240520.0	493384.0	721044.0	469792.0	
1	0.0	0.0	...	421400.0	178064.0	293306.0	245416.0	
2	0.0	0.0	...	277378.0	159812.0	423992.0	409564.0	
3	0.0	0.0	...	622012.0	229790.0	405298.0	347188.0	
4	0.0	0.0	...	388574.0	288278.0	900430.0	300412.0	
5	0.0	0.0	...	168.0	48.0	60.0	28.0	
6	0.0	0.0	...	715518.0	384948.0	915978.0	1052166.0	
7	0.0	0.0	...	699290.0	362510.0	1190028.0	1012704.0	
8	0.0	11804.0	...	129862.0	26872.0	34044.0	22472.0	
9	0.0	0.0	...	198386.0	99614.0	215734.0	189966.0	
10	0.0	0.0	...	495400.0	278660.0	544710.0	438256.0	
11	9744.0	13148.0	...	4776.0	2330.0	4634.0	6534.0	
12	0.0	0.0	...	1677646.0	1380302.0	2531322.0	426284.0	
13	0.0	0.0	...	190606.0	95674.0	207476.0	232470.0	
14	0.0	0.0	...	248.0	42.0	48.0	28.0	
15	0.0	20386.0	...	798872.0	112724.0	51736.0	7054.0	
16	0.0	1144.0	...	1370.0	812.0	2488.0	2340.0	
17	0.0	0.0	...	1260.0	296.0	878.0	11774.0	
18	0.0	0.0	...	2714.0	1480.0	4822.0	53694.0	
19	0.0	0.0	...	171390.0	95844.0	177206.0	173184.0	
20	0.0	0.0	...	80.0	0.0	0.0	0.0	
21	0.0	0.0	...	1784108.0	683976.0	1228736.0	971366.0	
22	222.0	323436.0	...	7908038.0	3026002.0	5025350.0	2025766.0	
23	0.0	0.0	...	670578.0	293816.0	464286.0	336782.0	
24	0.0	0.0	...	234998.0	136388.0	470258.0	548540.0	
25	0.0	0.0	...	52700.0	28890.0	55588.0	85370.0	
26	0.0	0.0	...	407008.0	221836.0	534336.0	606554.0	
27	0.0	0.0	...	7688.0	5050.0	33502.0	40466.0	
28	0.0	21886.0	...	133648.0	53830.0	92568.0	72416.0	
29	0.0	0.0	...	814318.0	411480.0	783304.0	689466.0	
...	
9970	0.0	2112.0	...	705544.0	335824.0	312610.0	126874.0	

9971	0.0	0.0 ...	194732.0	82488.0	196336.0	232152.0
9972	0.0	0.0 ...	595178.0	153610.0	187594.0	89276.0
9973	0.0	0.0 ...	668.0	60.0	18.0	26.0
9974	0.0	2.0 ...	1968174.0	488932.0	750872.0	482506.0
9975	0.0	0.0 ...	570110.0	268342.0	522602.0	414046.0
9976	0.0	0.0 ...	2904.0	1014.0	3040.0	53042.0
9977	0.0	0.0 ...	27058.0	12834.0	35874.0	175812.0
9978	0.0	0.0 ...	380516.0	177494.0	338906.0	304596.0
9979	0.0	94.0 ...	271578.0	144146.0	309750.0	364188.0
9980	0.0	0.0 ...	2502.0	1190.0	2034.0	6196.0
9981	0.0	0.0 ...	1566.0	2200.0	3720.0	2824.0
9982	0.0	0.0 ...	338282.0	160698.0	293218.0	281940.0
9983	0.0	0.0 ...	74.0	8.0	20.0	30.0
9984	0.0	0.0 ...	276798.0	143390.0	530966.0	431068.0
9985	0.0	0.0 ...	357856.0	223048.0	467216.0	488868.0
9986	0.0	0.0 ...	2136598.0	926574.0	1827952.0	1695934.0
9987	0.0	0.0 ...	210808.0	90378.0	172596.0	182692.0
9988	0.0	0.0 ...	288954.0	152914.0	355964.0	299488.0
9989	0.0	0.0 ...	405120.0	220770.0	405850.0	317390.0
9990	0.0	0.0 ...	175858.0	73758.0	161792.0	275762.0
9991	0.0	0.0 ...	366532.0	182780.0	366832.0	343904.0
9992	0.0	0.0 ...	10884.0	2582.0	3408.0	48344.0
9993	0.0	10422.0 ...	119566.0	41798.0	56016.0	47928.0
9994	0.0	11274.0 ...	3976.0	2694.0	8334.0	9808.0
9995	0.0	0.0 ...	419766.0	226036.0	514554.0	481152.0
9996	0.0	0.0 ...	240826.0	161830.0	401666.0	258648.0
9997	0.0	0.0 ...	512.0	8.0	32.0	2.0
9998	0.0	0.0 ...	1072650.0	564716.0	1820688.0	1100498.0
9999	0.0	0.0 ...	5616.0	2664.0	10962.0	70824.0

	164	165	166	167	168	169
0	339156.0	157956.0	73224.0	0.0	0.0	0.0
1	133654.0	81140.0	97576.0	1500.0	0.0	0.0
2	320746.0	158022.0	95128.0	514.0	0.0	0.0
3	286954.0	311560.0	433954.0	1218.0	0.0	0.0
4	1534.0	338.0	856.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0
6	1108672.0	341532.0	129504.0	7832.0	0.0	0.0
7	160090.0	63216.0	41202.0	4.0	0.0	0.0
8	34362.0	0.0	0.0	0.0	0.0	0.0
9	219996.0	208724.0	463790.0	34054.0	0.0	0.0
10	249386.0	73356.0	51344.0	2008.0	0.0	0.0
11	29980.0	39372.0	0.0	0.0	0.0	0.0
12	17996.0	11734.0	12506.0	138.0	0.0	0.0
13	154962.0	99406.0	121544.0	2712.0	0.0	0.0
14	32.0	10.0	0.0	0.0	0.0	0.0
15	6628.0	27600.0	2.0	2.0	0.0	0.0
16	2566.0	20870.0	38.0	0.0	0.0	0.0

17	8.0	0.0	0.0	0.0	0.0	0.0
18	1116.0	16.0	0.0	0.0	0.0	0.0
19	404690.0	736142.0	9268.0	6.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0
21	838200.0	559330.0	826630.0	18434.0	0.0	0.0
22	1160638.0	533834.0	493800.0	6914.0	0.0	0.0
23	229078.0	163348.0	372036.0	31152.0	0.0	0.0
24	8384.0	524.0	208.0	16.0	0.0	0.0
25	140892.0	10658.0	10944.0	356.0	0.0	0.0
26	325542.0	89624.0	6538.0	10.0	0.0	0.0
27	1436.0	34.0	0.0	0.0	0.0	0.0
28	70266.0	111176.0	592.0	0.0	0.0	0.0
29	481154.0	285904.0	381200.0	25594.0	0.0	0.0
...
9970	212506.0	172170.0	1106.0	8.0	0.0	0.0
9971	269840.0	251924.0	435940.0	20454.0	0.0	0.0
9972	54258.0	27544.0	23552.0	122.0	0.0	0.0
9973	22.0	0.0	0.0	0.0	0.0	0.0
9974	279364.0	92324.0	56728.0	20.0	0.0	0.0
9975	230436.0	125822.0	195024.0	18584.0	0.0	0.0
9976	3242.0	96.0	6.0	0.0	0.0	0.0
9977	576.0	116.0	2022.0	1598.0	0.0	0.0
9978	222388.0	138520.0	155286.0	1336.0	0.0	0.0
9979	313306.0	196544.0	216248.0	8050.0	0.0	0.0
9980	2170.0	304.0	0.0	0.0	0.0	0.0
9981	8516.0	12266.0	78.0	18.0	0.0	0.0
9982	191014.0	107878.0	45742.0	936.0	0.0	0.0
9983	34.0	78.0	0.0	0.0	0.0	0.0
9984	28444.0	4936.0	3582.0	214.0	0.0	0.0
9985	302092.0	82914.0	23114.0	24.0	0.0	0.0
9986	1345096.0	1042386.0	3123724.0	666360.0	0.0	0.0
9987	139520.0	195354.0	258604.0	17844.0	0.0	0.0
9988	256348.0	199566.0	430024.0	15620.0	0.0	0.0
9989	202976.0	107760.0	151412.0	6.0	0.0	0.0
9990	243622.0	794378.0	10722.0	0.0	0.0	0.0
9991	213896.0	68196.0	41390.0	144.0	0.0	0.0
9992	9472.0	132.0	344.0	56.0	0.0	0.0
9993	54274.0	144122.0	2996.0	0.0	0.0	0.0
9994	10674.0	17530.0	40.0	0.0	0.0	0.0
9995	391556.0	277872.0	257100.0	4008.0	0.0	0.0
9996	113434.0	48668.0	22472.0	148.0	0.0	0.0
9997	0.0	0.0	0.0	0.0	0.0	0.0
9998	387584.0	119906.0	109084.0	9328.0	0.0	0.0
9999	678.0	90.0	62.0	0.0	0.0	0.0

[10000 rows x 170 columns]