

# Project 1 - Mercedes-Benz Greener Manufacturing

Aurore Prevot

## 1. Importation of the librairies

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns

import matplotlib.pyplot as plt
%matplotlib inline

from sklearn.preprocessing import LabelEncoder
from sklearn import model_selection
from sklearn.model_selection import train_test_split
from xgboost import XGBRegressor
from sklearn.metrics import mean_squared_error

import warnings
warnings.filterwarnings("ignore")
```

## 2. Reading the datasets

```
In [2]: data_train = pd.read_csv('Datasets/mercedes_train.csv')
data_train.head()
```

```
Out[2]:
```

	ID	y	X0	X1	X2	X3	X4	X5	X6	X8	...	X375	X376	X377	X378	X379	X380	X382	X383	X384	X385
0	0	130.81	k	v	at	a	d	u	j	o	...	0	0	1	0	0	0	0	0	0	0
1	6	88.53	k	t	av	e	d	y	l	o	...	1	0	0	0	0	0	0	0	0	0
2	7	76.26	az	w	n	c	d	x	j	x	...	0	0	0	0	0	0	1	0	0	0
3	9	80.62	az	t	n	f	d	x	l	e	...	0	0	0	0	0	0	0	0	0	0
4	13	78.02	az	v	n	f	d	h	d	n	...	0	0	0	0	0	0	0	0	0	0

5 rows × 378 columns

```
In [3]: data_train.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4209 entries, 0 to 4208
Columns: 378 entries, ID to X385
dtypes: float64(1), int64(369), object(8)
memory usage: 12.1+ MB
```

```
In [4]: data_train.dtypes
```

```
Out[4]:
```

ID	int64
y	float64
X0	object
X1	object
X2	object
...	
X380	int64
X382	int64
X383	int64
X384	int64
X385	int64

Length: 378, dtype: object

```
In [5]: data_test = pd.read_csv('Datasets/mercedes_test.csv')
data_test.head()
```

```
Out[5]:
```

	ID	X0	X1	X2	X3	X4	X5	X6	X8	X10	...	X375	X376	X377	X378	X379	X380	X382	X383	X384	X385
0	1	az	v	n	f	d	t	a	w	0	...	0	0	0	1	0	0	0	0	0	0
1	2	t	b	ai	a	d	b	g	y	0	...	0	0	1	0	0	0	0	0	0	0
2	3	az	v	as	f	d	a	j	j	0	...	0	0	0	1	0	0	0	0	0	0
3	4	az	l	n	f	d	z	l	n	0	...	0	0	0	1	0	0	0	0	0	0
4	5	w	s	as	c	d	y	i	m	0	...	1	0	0	0	0	0	0	0	0	0

5 rows × 377 columns

```
In [6]: data_test.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4209 entries, 0 to 4208
Columns: 377 entries, ID to X385
dtypes: int64(369), object(8)
memory usage: 12.1+ MB
```

```
In [7]: print(f"Shape of the train dataset: {data_train.shape}")
print(f"Shape of the test dataset: {data_test.shape}")
```

```
Shape of the train dataset: (4209, 378)
Shape of the test dataset: (4209, 377)
```

- The train dataset has 4209 rows and 378 columns of which 8 are categorical and the rest are numerical.
- The test dataset has 4209 rows and 377 columns of which 8 are categorical and the rest are numerical.

Note: To perform the EDA, the two datasets can be combined or they can stay separated.

### 3. Data wrangling, modeling and prediction

#### a. The datasets will be combined

```
In [8]: data_train['Source'] = 'train'
data_test['Source'] = 'test'
print(data_train.head())
print(f"Shape of the train dataset: {data_train.shape}")
print(data_test.head())
print(f"Shape of the test dataset: {data_test.shape}")
```

```
   ID    y  X0 X1  X2 X3 X4 X5 X6 X8  ...  X376  X377  X378  X379  X380  \
0  0  130.81  k  v  at  a  d  u  j  o  ...    0    1    0    0    0
1  6   88.53  k  t  av  e  d  y  l  o  ...    0    0    0    0    0
2  7   76.26  az  w   n  c  d  x  j  x  ...    0    0    0    0    0
3  9   80.62  az  t   n  f  d  x  l  e  ...    0    0    0    0    0
4 13   78.02  az  v   n  f  d  h  d  n  ...    0    0    0    0    0
```

```
   X382  X383  X384  X385  Source
0     0     0     0     0   train
1     0     0     0     0   train
2     1     0     0     0   train
3     0     0     0     0   train
4     0     0     0     0   train
```

[5 rows x 379 columns]

Shape of the train dataset: (4209, 379)

```
   ID  X0 X1  X2 X3 X4 X5 X6 X8  X10  ...  X376  X377  X378  X379  X380  X382  \
0  1  az  v   n  f  d  t  a  w   0  ...    0    0    1    0    0    0
1  2   t  b  ai  a  d  b  g  y   0  ...    0    1    0    0    0    0
2  3  az  v  as  f  d  a  j  j   0  ...    0    0    1    0    0    0
3  4  az  l   n  f  d  z  l  n   0  ...    0    0    1    0    0    0
4  5   w  s  as  c  d  y  i  m   0  ...    0    0    0    0    0    0
```

```
   X383  X384  X385  Source
0     0     0     0   test
1     0     0     0   test
2     0     0     0   test
3     0     0     0   test
4     0     0     0   test
```

[5 rows x 378 columns]

Shape of the test dataset: (4209, 378)

```
In [9]: all_data = pd.concat([data_train, data_test], axis=0, ignore_index=True)
all_data.drop(['ID'], axis=1, inplace=True)
print(all_data.head())
print(all_data.shape)
```

```
   y  X0 X1  X2 X3 X4 X5 X6 X8  X10  ...  X376  X377  X378  X379  X380  \
0  130.81  k  v  at  a  d  u  j  o   0  ...    0    1    0    0    0
1   88.53  k  t  av  e  d  y  l  o   0  ...    0    0    0    0    0
2   76.26  az  w   n  c  d  x  j  x   0  ...    0    0    0    0    0
3   80.62  az  t   n  f  d  x  l  e   0  ...    0    0    0    0    0
4   78.02  az  v   n  f  d  h  d  n   0  ...    0    0    0    0    0
```

```
   X382  X383  X384  X385  Source
0     0     0     0     0   train
1     0     0     0     0   train
2     1     0     0     0   train
3     0     0     0     0   train
4     0     0     0     0   train
```

[5 rows x 378 columns]

(8418, 378)

```
In [10]: all_data.tail()
```

```
Out[10]:
```

	y	X0	X1	X2	X3	X4	X5	X6	X8	X10	...	X376	X377	X378	X379	X380	X382	X383	X384	X385	Source
8413	NaN	aj	h	as	f	d	aa	j	e	0	...	0	0	0	0	0	0	0	0	0	test
8414	NaN	t	aa	ai	d	d	aa	j	y	0	...	1	0	0	0	0	0	0	0	0	test
8415	NaN	y	v	as	f	d	aa	d	w	0	...	0	0	0	0	0	0	0	0	0	test
8416	NaN	ak	v	as	a	d	aa	c	q	0	...	0	1	0	0	0	0	0	0	0	test
8417	NaN	t	aa	ai	c	d	aa	g	r	0	...	0	0	0	0	0	0	0	0	0	test

5 rows x 378 columns

In [11]: `all_data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8418 entries, 0 to 8417
Columns: 378 entries, y to Source
dtypes: float64(1), int64(368), object(9)
memory usage: 24.3+ MB
```

Dropping the columns with null variance

In [12]: `numeric_all_data = all_data._get_numeric_data()`  
`for column in numeric_all_data.columns:`  
 `print(f'{column} variance: {numeric_all_data[column].var(ddof=0)}\n\n')`

y variance: 160.72851374270064

X10 variance: 0.01589484479879718

X11 variance: 0.00011877895069345627

X12 variance: 0.06913763292531624

X13 variance: 0.05597325047792406

X14 variance: 0.24481764206524154

X15 variance: 0.0005936125176333895

X16 variance: 0.0026066172674931326

X17 variance: 0.008129535071217418

X18 variance: 0.008946763040043453

X19 variance: 0.0944541872254304

X20 variance: 0.12120956569493431

X21 variance: 0.0021337029042236655

X22 variance: 0.07752545611991997

X23 variance: 0.020014923501952876

X24 variance: 0.002133702904223666

X26 variance: 0.006959667539734994

X27 variance: 0.21692251062773144

X28 variance: 0.028816278640375153

X29 variance: 0.039303950550927176

X30 variance: 0.004964415423827063

X31 variance: 0.1796975374528363

X32 variance: 0.010693308939125743

X33 variance: 0.0004749464612734898

X34 variance: 0.004493758947228727

X35 variance: 0.1796975374528363

X36 variance: 0.0036690235052211657

X37 variance: 0.1796975374528363

X38 variance: 0.0311565358396327

X39 variance: 0.0004749464612734898

X40 variance: 0.0013050161605416132

X41 variance: 0.01127407972631704

X42 variance: 0.0008308599596030204

X43 variance: 0.06903657838499126

X44 variance: 0.009063396998399328

X45 variance: 0.18882927785841264

X46 variance: 0.2408779120599428

X47 variance: 0.01185414492392361

X48 variance: 0.02251474272990126

X49 variance: 0.11148304149301404

X50 variance: 0.1710695738989656

X51 variance: 0.20237878521294209

X52 variance: 0.04061058956751367

X53 variance: 0.006373675389616693

X54 variance: 0.039957778083721425

X55 variance: 0.005786977649913656

X56 variance: 0.02251474272990126

X57 variance: 0.01428271371560596

X58 variance: 0.24364629279563116

X59 variance: 0.0009494413452127527

X60 variance: 0.0015419249195105732

X61 variance: 0.04418289085246003

X62 variance: 0.005904373645021041

X63 variance: 0.012085973437882515

X64 variance: 0.23248732295416696

X65 variance: 0.0020154037544478243

X66 variance: 0.026351230867162272

X67 variance: 0.0020154037544478243

X68 variance: 0.07075066716317598

X69 variance: 0.029374611678772033

X70 variance: 0.07345627864940671

X71 variance: 0.09454787541049088

X73 variance: 0.019102511498146752

X74 variance: 0.0014234846518177875

X75 variance: 0.03601957215192574

X76 variance: 0.039957778083721425

X77 variance: 0.011738188331569078

X78 variance: 0.00649093026680713

X79 variance: 0.028369104185156646

X80 variance: 0.04900467839762975

X81 variance: 0.17397919955772517

X82 variance: 0.018531336729307776

X83 variance: 0.001186519445682048

X84 variance: 0.09454787541049088

X85 variance: 0.24266411209368693

X86 variance: 0.0014234846518177873

X87 variance: 0.0009494413452127528

X88 variance: 0.007661926741625115

X89 variance: 0.0013050161605416134

X90 variance: 0.007778871159398278

X91 variance: 0.002133702904223666

X92 variance: 0.0016603369636199697

X93 variance: 0.00023752967780352354

X94 variance: 0.0075449541002685725

X95 variance: 0.0004749464612734898

X96 variance: 0.18603631438159224

X97 variance: 0.0028429051076275308

X98 variance: 0.054084923407686086

X99 variance: 0.009180002733171813

X100 variance: 0.21512716615296926

X101 variance: 0.05972247129534563

X102 variance: 0.006373675389616693

X103 variance: 0.16796926971082854

X104 variance: 0.0022519738304161156

X105 variance: 0.0013050161605416132

X106 variance: 0.01254929178279965

X107 variance: 0.0004749464612734898

X108 variance: 0.01301215855038257

X109 variance: 0.03832130415985703

X110 variance: 0.0011865194456820484

X111 variance: 0.024437065106331504

X112 variance: 0.0020154037544478243

X113 variance: 0.02251474272990126

X114 variance: 0.12434499413259921

X115 variance: 0.20578916779998466

X116 variance: 0.15787787102285258

X117 variance: 0.04461380852364638

X118 variance: 0.23505785637029705

X119 variance: 0.23505785637029705

X120 variance: 0.040719292698937176

X122 variance: 0.007661926741625113

X123 variance: 0.0029610066923196473

X124 variance: 0.0003562521813302013

X125 variance: 0.0039048033208545574

X126 variance: 0.037883838617324396

X127 variance: 0.24991192830803396

X128 variance: 0.03843059998653171

X129 variance: 0.1117481879471639

X130 variance: 0.03843059998653171

X131 variance: 0.02623885866989815

X132 variance: 0.21459272437791133

X133 variance: 0.1084602957120305

X134 variance: 0.02251474272990126

X135 variance: 0.026014029604619775

X136 variance: 0.03984887738721418

X137 variance: 0.24258248949052533

X138 variance: 0.038539867589623

X139 variance: 0.08445986355252166

X140 variance: 0.03832130415985703

X141 variance: 0.015089470734994614

X142 variance: 0.17817752814582744

X143 variance: 0.03744592149745753

X144 variance: 0.15715534728808894

X145 variance: 0.0010679945072390948

X146 variance: 0.038539867589623

X147 variance: 0.02251474272990126

X148 variance: 0.040827967606777295

X150 variance: 0.1657773133304895

X151 variance: 0.07900232567971842

X152 variance: 0.030822975419347424

X153 variance: 0.0009494413452127527

X154 variance: 0.16653405816011252

X155 variance: 0.07195570127956132

X156 variance: 0.20437934336252722

X157 variance: 0.20437934336252717

X158 variance: 0.1782411864381618

X159 variance: 0.013936537353545716

X160 variance: 0.0010679945072390945

X161 variance: 0.15780574565550143

X162 variance: 0.03678819911015557

X163 variance: 0.21346625060588978

X164 variance: 0.06086077074880723

X165 variance: 0.004022650893296172

X166 variance: 0.02926300151825944

X167 variance: 0.0010679945072390948

X168 variance: 0.1958434451376721

X169 variance: 0.0069596675397349955



X170 variance: 0.02488819086322451

X171 variance: 0.22378355085531282

X172 variance: 0.005904373645021041

X173 variance: 0.009879044446555565

X174 variance: 0.018073888889735594

X175 variance: 0.022061246192003576

X176 variance: 0.015549853827239273

X177 variance: 0.04836507555086366

X178 variance: 0.2465282734544611

X179 variance: 0.04439840613521998

X180 variance: 0.13319487972265923

X181 variance: 0.08455651521383795

X182 variance: 0.09153585459119523

X183 variance: 0.003079080053428372

X184 variance: 0.001067994507239095

X185 variance: 0.018073888889735594

X186 variance: 0.24884571188654844

X187 variance: 0.2442918367067007

X189 variance: 0.07475164234442921

X190 variance: 0.00023752967780352343

X191 variance: 0.24929823471081916

X192 variance: 0.0020154037544478243

X194 variance: 0.24885376971960607

X195 variance: 0.011738188331569078

X196 variance: 0.01115798201604556

X197 variance: 0.029932239127584213

X198 variance: 0.02342038107369397

X199 variance: 0.0020154037544478243

X200 variance: 0.0066081569204141785

X201 variance: 0.14567207436169122

X202 variance: 0.1826887857147575

X203 variance: 0.014167349818502598

X204 variance: 0.0004749464612734898

X205 variance: 0.00047494646127348985

X206 variance: 0.018645628130242366

X207 variance: 0.0008308599596030204

X208 variance: 0.05847679521887982

X209 variance: 0.09172495259990286

X210 variance: 0.00023752967780352348

X211 variance: 0.015779876031861265

X212 variance: 0.006256392288842863

X213 variance: 0.0020154037544478243

X214 variance: 0.006373675389616693

X215 variance: 0.09332772756520057

X216 variance: 0.005904373645021041

X217 variance: 0.008129535071217416

X218 variance: 0.21644548973307887

X219 variance: 0.06281888062897498

X220 variance: 0.2464861497562526

X221 variance: 0.008129535071217416

X222 variance: 0.02251474272990126

X223 variance: 0.2468959561870639

X224 variance: 0.21592120844804105

X225 variance: 0.08320082360932102

X226 variance: 0.030822975419347424

X227 variance: 0.0039048033208545574

X228 variance: 0.04104523275170738

X229 variance: 0.042562927724878304

X230 variance: 0.004493758947228727

X231 variance: 0.01635443763070695

X232 variance: 0.03963099132344954

X233 variance: 0.00011877895069345626

X234 variance: 0.1623648704218596

X235 variance: 0.00011877895069345628

X236 variance: 0.0003562521813302013

X237 variance: 0.006021741416545039

X238 variance: 0.07485108814050107

X239 variance: 0.006373675389616693

X240 variance: 0.0031971251909537106

X241 variance: 0.0905886711326536

X242 variance: 0.00777887115939828

X243 variance: 0.007661926741625113

X244 variance: 0.09454787541049088

X245 variance: 0.0008308599596030206

X246 variance: 0.24245920887828135

X247 variance: 0.18262713129684371

X248 variance: 0.001778720784145978

X249 variance: 0.008479944970786052

X250 variance: 0.24703985412697366

X251 variance: 0.23877175537310705

X252 variance: 0.00047494646127348974

X253 variance: 0.0015419249195105732

X254 variance: 0.004493758947228727

X255 variance: 0.016584064705161484

X256 variance: 0.06923865924205791

X257 variance: 0.00011877895069345631

X258 variance: 0.0011865194456820482

X259 variance: 0.00023752967780352345

X260 variance: 0.0004749464612734898

X261 variance: 0.24491975498994364

X262 variance: 0.001067994507239095

X263 variance: 0.03963099132344954

X264 variance: 0.038539867589623

X265 variance: 0.08290959856411957

X266 variance: 0.001186519445682048

X267 variance: 0.00847994497078605

X268 variance: 0.00011877895069345628

X269 variance: 0.0005936125176333895

X270 variance: 0.00047494646127348974

X271 variance: 0.0017787207841459779

X272 variance: 0.03502930539334052

X273 variance: 0.2032561011908038

X274 variance: 0.010460803059165493

X275 variance: 0.19894912591844485

X276 variance: 0.035799710437323856

X277 variance: 0.0014234846518177873

X278 variance: 0.0003562521813302013

X279 variance: 0.03963099132344954

X280 variance: 0.00023752967780352345

X281 variance: 0.003079080053428372

X282 variance: 0.0030790800534283725

X283 variance: 0.12223069494195567

X284 variance: 0.03984887738721419

X285 variance: 0.16480924442023984

X286 variance: 0.048151648813274536

X287 variance: 0.015434800389553182

X288 variance: 0.00023752967780352354

X289 variance: 0.00023752967780352348

X290 variance: 0.00011877895069345631

X291 variance: 0.010460803059165491

X292 variance: 0.009063396998399328

X293 variance: 0.00011877895069345631

X294 variance: 0.1082814710876765

X295 variance: 0.0001187789506934563

X296 variance: 0.0001187789506934563

X297 variance: 0.00011877895069345627

X298 variance: 0.003786927524829557

X299 variance: 0.003786927524829557

X300 variance: 0.16577731333048948

X301 variance: 0.048365075550863654

X302 variance: 0.009063396998399328

X304 variance: 0.06782172246158642

X305 variance: 0.010693308939125743

X306 variance: 0.04245470439437241

X307 variance: 0.0027247752993520267

X308 variance: 0.008246366594657016

X309 variance: 0.006842525556878112

X310 variance: 0.003433130795254216

X311 variance: 0.24039516177786224

X312 variance: 0.003786927524829557

X313 variance: 0.2086391288034595

X314 variance: 0.24640114032308438

X315 variance: 0.02635123086716227

X316 variance: 0.15701050385813564

X317 variance: 0.006608156920414179

X318 variance: 0.0008308599596030205

X319 variance: 0.0004749464612734898

X320 variance: 0.0074279532353286335

X321 variance: 0.1828120098798347

X322 variance: 0.019672980677401

X323 variance: 0.008596691823475482

X324 variance: 0.24364629279563116

X325 variance: 0.005552100988948716

X326 variance: 0.030822975419347424

X327 variance: 0.11263017492807642

X328 variance: 0.036788199110155566

X329 variance: 0.24599099520927253

X330 variance: 0.00011877895069345631

X331 variance: 0.05102338431133413

X332 variance: 0.0007122503504098997

X333 variance: 0.02126654064272211

X334 variance: 0.24816093719456794

X335 variance: 0.002842905107627531

X336 variance: 0.11121764102661365

X337 variance: 0.24952741020793953

X338 variance: 0.00731092414680531

X339 variance: 0.00023752967780352345

X340 variance: 0.02092552919642208

X341 variance: 0.008946763040043453

X342 variance: 0.020811802267155285

X343 variance: 0.0685308823296152

X344 variance: 0.008946763040043453

X345 variance: 0.022967787690464748

X346 variance: 0.04407509087570498

X347 variance: 0.00023752967780352348

X348 variance: 0.048685003980371976

X349 variance: 0.041479424358566844

X350 variance: 0.22439543814319074

X351 variance: 0.20683201509442398

X352 variance: 0.04943051582380603

X353 variance: 0.002370216533025177

X354 variance: 0.1620128517780378

X355 variance: 0.23537561158388426

X356 variance: 0.148870257937022

X357 variance: 0.0017787207841459779

X358 variance: 0.245137514047583

X359 variance: 0.03315256410408383

X360 variance: 0.07195570127956132

X361 variance: 0.034478169368707866

X362 variance: 0.24963420824748425

X363 variance: 0.1877077149899008

X364 variance: 0.0031971251909537106

X365 variance: 0.003079080053428372

X366 variance: 0.001423484651817787

X367 variance: 0.047510691234505756

X368 variance: 0.058996320830116564

X369 variance: 0.00023752967780352345

X370 variance: 0.006139080964485642

X371 variance: 0.012665050810070471

X372 variance: 0.00035625218133020133

X373 variance: 0.016584064705161484

X374 variance: 0.177986383927324

X375 variance: 0.21845979276777447

X376 variance: 0.050599240300161205

X377 variance: 0.21517151951426536

X378 variance: 0.019558943288716935

X379 variance: 0.010577070110937309

X380 variance: 0.008012675324194425

X382 variance: 0.008129535071217418

X383 variance: 0.001067994507239095

X384 variance: 0.0005936125176333895

X385 variance: 0.0015419249195105732

```
In [13]: list_null_var = []
for column in numeric_all_data.columns:
    if numeric_all_data[column].var() <= 0:
        list_null_var.append(column)

print(list_null_var)
print(len(list_null_var))

[]
0
```

- In the combined dataset, there are no column with a null variance.

Checking for null values

```
In [14]: for column in all_data.columns:
    if all_data[column].isnull().sum() != 0:
        print(f'{column} has {all_data[column].isnull().sum()} null values')
```

y has 4209 null values

- Only the y column, the target of the model, has null values. These null values are from the test dataset. So they are no missing values in the dataset.

Checking for unique values

```
In [15]: for column in all_data.columns:
          print(f"Unique values of column {column} :\n{all_data[column].value_counts()}\n\n")
```



```

Unique values of column y :
91.88      7
89.38      7
89.06      7
90.76      7
89.19      6
..
110.90     1
132.60     1
94.08      1
75.89      1
87.48      1
Name: y, Length: 2545, dtype: int64

```

```

Unique values of column X0 :
ak      781
z       695
y       672
ay      612
x       602
t       599
o       515
f       440
w       380
n       362
j       352
az      336
aj      313
s       222
ap      211
al      155
h       139
d       134
e       80
v       76
ai       72
af       69
m        68
at       46
am       46
ba       46
i        43
a        39
u        37
aq       29
aw       27
ax       27
ad       26
b        24
k        23
l        22
r        20
as       16
au       16
bc       12
c         9
ao         9
g         4
aa         2
q         2
an         1
p         1
ae         1
ac         1
ag         1
av         1
ab         1
bb         1
Name: X0, dtype: int64

```

```

Unique values of column X1 :
aa      1659
s      1200
l      1189
b      1188
v      844
r      503
i      392
a      296
c      263
o      163
w      102
z       77
u       77
e       62
m       59
h       56
t       49
y       44
j       44
n       35
f       35

```

```
k      29
p      19
g      15
ab      8
q       6
d       4
Name: X1, dtype: int64
```

Unique values of column X2 :

```
as    3317
ae     974
ai     877
m      715
ak     525
r      308
n      250
s      194
f      172
e      165
aq     135
ay     132
a       91
b       59
t       54
k       50
ag      42
i       40
ao      39
ac      33
z       31
d       24
ap      22
g       17
aw      17
p       14
y       12
h       12
x       12
q       10
at       9
al       9
au       8
ah       7
an       6
av       5
af       5
am       4
ad       4
ab       4
j        3
w        3
aj        1
ax        1
aa        1
ar        1
o         1
c         1
l         1
u         1
Name: X2, dtype: int64
```

Unique values of column X3 :

```
c    3842
f    2159
a     916
d     564
g     513
e     321
b     103
Name: X3, dtype: int64
```

Unique values of column X4 :

```
d    8408
b       5
a       3
c       2
Name: X4, dtype: int64
```

Unique values of column X5 :

```
v    477
r    454
w    449
p    435
n    421
s    419
q    417
ac   412
d    406
af   405
ag   405
```

m	405
l	401
ae	401
ad	398
i	387
ab	376
k	370
j	262
c	252
aa	217
ah	177
o	36
f	13
g	9
x	4
h	3
y	2
b	1
a	1
u	1
t	1
z	1

Name: X5, dtype: int64

Unique values of column X6 :

g	2115
j	2041
d	1214
i	978
l	951
h	408
a	402
k	110
c	78
b	47
f	45
e	29

Name: X6, dtype: int64

Unique values of column X8 :

j	533
e	499
s	499
f	484
n	478
i	471
r	447
a	412
w	388
v	368
b	362
o	332
k	330
m	317
g	267
u	263
h	245
t	236
q	231
y	222
x	215
d	211
p	206
l	201
c	201

Name: X8, dtype: int64

Unique values of column X10 :

0	8282
1	136

Name: X10, dtype: int64

Unique values of column X11 :

0	8417
1	1

Name: X11, dtype: int64

Unique values of column X12 :

0	7789
1	629

Name: X12, dtype: int64

Unique values of column X13 :

0	7917
1	501

Name: X13, dtype: int64

Unique values of column X14 :  
0 4815  
1 3603  
Name: X14, dtype: int64

Unique values of column X15 :  
0 8413  
1 5  
Name: X15, dtype: int64

Unique values of column X16 :  
0 8396  
1 22  
Name: X16, dtype: int64

Unique values of column X17 :  
0 8349  
1 69  
Name: X17, dtype: int64

Unique values of column X18 :  
0 8342  
1 76  
Name: X18, dtype: int64

Unique values of column X19 :  
0 7529  
1 889  
Name: X19, dtype: int64

Unique values of column X20 :  
0 7230  
1 1188  
Name: X20, dtype: int64

Unique values of column X21 :  
0 8400  
1 18  
Name: X21, dtype: int64

Unique values of column X22 :  
0 7705  
1 713  
Name: X22, dtype: int64

Unique values of column X23 :  
0 8246  
1 172  
Name: X23, dtype: int64

Unique values of column X24 :  
0 8400  
1 18  
Name: X24, dtype: int64

Unique values of column X26 :  
0 8359  
1 59  
Name: X26, dtype: int64

Unique values of column X27 :  
1 5740  
0 2678  
Name: X27, dtype: int64

Unique values of column X28 :  
0 8168  
1 250  
Name: X28, dtype: int64

Unique values of column X29 :  
0 8073  
1 345  
Name: X29, dtype: int64

Unique values of column X30 :  
0 8376  
1 42  
Name: X30, dtype: int64

Unique values of column X31 :  
0 6441  
1 1977  
Name: X31, dtype: int64

Unique values of column X32 :  
0 8327  
1 91  
Name: X32, dtype: int64

Unique values of column X33 :  
0 8414  
1 4  
Name: X33, dtype: int64

Unique values of column X34 :  
0 8380  
1 38  
Name: X34, dtype: int64

Unique values of column X35 :  
0 6441  
1 1977  
Name: X35, dtype: int64

Unique values of column X36 :  
0 8387  
1 31  
Name: X36, dtype: int64

Unique values of column X37 :  
0 6441  
1 1977  
Name: X37, dtype: int64

Unique values of column X38 :  
0 8147  
1 271  
Name: X38, dtype: int64

Unique values of column X39 :  
0 8414  
1 4  
Name: X39, dtype: int64

Unique values of column X40 :  
0 8407  
1 11  
Name: X40, dtype: int64

Unique values of column X41 :  
0 8322  
1 96  
Name: X41, dtype: int64

Unique values of column X42 :  
0 8411  
1 7  
Name: X42, dtype: int64

Unique values of column X43 :  
0 7790  
1 628  
Name: X43, dtype: int64

Unique values of column X44 :  
0 8341  
1 77  
Name: X44, dtype: int64

Unique values of column X45 :  
0 6291  
1 2127  
Name: X45, dtype: int64

Unique values of column X46 :  
1 5013

0 3405  
Name: X46, dtype: int64

Unique values of column X47 :  
0 8317  
1 101  
Name: X47, dtype: int64

Unique values of column X48 :  
0 8224  
1 194  
Name: X48, dtype: int64

Unique values of column X49 :  
0 7342  
1 1076  
Name: X49, dtype: int64

Unique values of column X50 :  
0 6574  
1 1844  
Name: X50, dtype: int64

Unique values of column X51 :  
1 6046  
0 2372  
Name: X51, dtype: int64

Unique values of column X52 :  
0 8061  
1 357  
Name: X52, dtype: int64

Unique values of column X53 :  
0 8364  
1 54  
Name: X53, dtype: int64

Unique values of column X54 :  
0 8067  
1 351  
Name: X54, dtype: int64

Unique values of column X55 :  
0 8369  
1 49  
Name: X55, dtype: int64

Unique values of column X56 :  
0 8224  
1 194  
Name: X56, dtype: int64

Unique values of column X57 :  
0 8296  
1 122  
Name: X57, dtype: int64

Unique values of column X58 :  
1 4880  
0 3538  
Name: X58, dtype: int64

Unique values of column X59 :  
0 8410  
1 8  
Name: X59, dtype: int64

Unique values of column X60 :  
0 8405  
1 13  
Name: X60, dtype: int64

Unique values of column X61 :  
1 8028  
0 390  
Name: X61, dtype: int64

Unique values of column X62 :  
0 8368  
1 50  
Name: X62, dtype: int64

Unique values of column X63 :  
0 8315  
1 103  
Name: X63, dtype: int64

Unique values of column X64 :  
0 5323  
1 3095  
Name: X64, dtype: int64

Unique values of column X65 :  
0 8401  
1 17  
Name: X65, dtype: int64

Unique values of column X66 :  
0 8190  
1 228  
Name: X66, dtype: int64

Unique values of column X67 :  
0 8401  
1 17  
Name: X67, dtype: int64

Unique values of column X68 :  
0 7773  
1 645  
Name: X68, dtype: int64

Unique values of column X69 :  
0 8163  
1 255  
Name: X69, dtype: int64

Unique values of column X70 :  
1 7746  
0 672  
Name: X70, dtype: int64

Unique values of column X71 :  
0 7528  
1 890  
Name: X71, dtype: int64

Unique values of column X73 :  
0 8254  
1 164  
Name: X73, dtype: int64

Unique values of column X74 :  
1 8406  
0 12  
Name: X74, dtype: int64

Unique values of column X75 :  
0 8103  
1 315  
Name: X75, dtype: int64

Unique values of column X76 :  
0 8067  
1 351  
Name: X76, dtype: int64

Unique values of column X77 :  
0 8318  
1 100  
Name: X77, dtype: int64

Unique values of column X78 :  
0 8363  
1 55  
Name: X78, dtype: int64

Unique values of column X79 :  
0 8172  
1 246  
Name: X79, dtype: int64

Unique values of column X80 :  
1 7983  
0 435  
Name: X80, dtype: int64

Unique values of column X81 :  
0 6530  
1 1888  
Name: X81, dtype: int64

Unique values of column X82 :  
0 8259  
1 159  
Name: X82, dtype: int64

Unique values of column X83 :  
0 8408  
1 10  
Name: X83, dtype: int64

Unique values of column X84 :  
0 7528  
1 890  
Name: X84, dtype: int64

Unique values of column X85 :  
0 4930  
1 3488  
Name: X85, dtype: int64

Unique values of column X86 :  
0 8406  
1 12  
Name: X86, dtype: int64

Unique values of column X87 :  
0 8410  
1 8  
Name: X87, dtype: int64

Unique values of column X88 :  
0 8353  
1 65  
Name: X88, dtype: int64

Unique values of column X89 :  
0 8407  
1 11  
Name: X89, dtype: int64

Unique values of column X90 :  
0 8352  
1 66  
Name: X90, dtype: int64

Unique values of column X91 :  
0 8400  
1 18  
Name: X91, dtype: int64

Unique values of column X92 :  
0 8404  
1 14  
Name: X92, dtype: int64

Unique values of column X93 :  
0 8416  
1 2  
Name: X93, dtype: int64

Unique values of column X94 :  
0 8354



```
1      64
Name: X94, dtype: int64
```

```
Unique values of column X95 :
0      8414
1         4
Name: X95, dtype: int64
```

```
Unique values of column X96 :
1      6338
0      2080
Name: X96, dtype: int64
```

```
Unique values of column X97 :
0      8394
1         24
Name: X97, dtype: int64
```

```
Unique values of column X98 :
1      7935
0       483
Name: X98, dtype: int64
```

```
Unique values of column X99 :
0      8340
1         78
Name: X99, dtype: int64
```

```
Unique values of column X100 :
1       5781
0      2637
Name: X100, dtype: int64
```

```
Unique values of column X101 :
1       7881
0        537
Name: X101, dtype: int64
```

```
Unique values of column X102 :
0      8364
1         54
Name: X102, dtype: int64
```

```
Unique values of column X103 :
1      6620
0      1798
Name: X103, dtype: int64
```

```
Unique values of column X104 :
0      8399
1         19
Name: X104, dtype: int64
```

```
Unique values of column X105 :
0      8407
1         11
Name: X105, dtype: int64
```

```
Unique values of column X106 :
0      8311
1       107
Name: X106, dtype: int64
```

```
Unique values of column X107 :
0      8414
1         4
Name: X107, dtype: int64
```

```
Unique values of column X108 :
0      8307
1       111
Name: X108, dtype: int64
```

```
Unique values of column X109 :
0      8082
1        336
Name: X109, dtype: int64
```

Unique values of column X110 :  
0 8408  
1 10  
Name: X110, dtype: int64

Unique values of column X111 :  
1 8207  
0 211  
Name: X111, dtype: int64

Unique values of column X112 :  
0 8401  
1 17  
Name: X112, dtype: int64

Unique values of column X113 :  
0 8224  
1 194  
Name: X113, dtype: int64

Unique values of column X114 :  
0 7193  
1 1225  
Name: X114, dtype: int64

Unique values of column X115 :  
0 5979  
1 2439  
Name: X115, dtype: int64

Unique values of column X116 :  
0 6764  
1 1654  
Name: X116, dtype: int64

Unique values of column X117 :  
0 8024  
1 394  
Name: X117, dtype: int64

Unique values of column X118 :  
1 5238  
0 3180  
Name: X118, dtype: int64

Unique values of column X119 :  
1 5238  
0 3180  
Name: X119, dtype: int64

Unique values of column X120 :  
1 8060  
0 358  
Name: X120, dtype: int64

Unique values of column X122 :  
0 8353  
1 65  
Name: X122, dtype: int64

Unique values of column X123 :  
0 8393  
1 25  
Name: X123, dtype: int64

Unique values of column X124 :  
0 8415  
1 3  
Name: X124, dtype: int64

Unique values of column X125 :  
0 8385  
1 33  
Name: X125, dtype: int64

Unique values of column X126 :  
0 8086  
1 332  
Name: X126, dtype: int64

Unique values of column X127 :  
0 4288  
1 4130  
Name: X127, dtype: int64

Unique values of column X128 :  
1 8081  
0 337  
Name: X128, dtype: int64

Unique values of column X129 :  
0 7339  
1 1079  
Name: X129, dtype: int64

Unique values of column X130 :  
0 8081  
1 337  
Name: X130, dtype: int64

Unique values of column X131 :  
0 8191  
1 227  
Name: X131, dtype: int64

Unique values of column X132 :  
1 5793  
0 2625  
Name: X132, dtype: int64

Unique values of column X133 :  
0 7376  
1 1042  
Name: X133, dtype: int64

Unique values of column X134 :  
0 8224  
1 194  
Name: X134, dtype: int64

Unique values of column X135 :  
0 8193  
1 225  
Name: X135, dtype: int64

Unique values of column X136 :  
1 8068  
0 350  
Name: X136, dtype: int64

Unique values of column X137 :  
1 4934  
0 3484  
Name: X137, dtype: int64

Unique values of column X138 :  
0 8080  
1 338  
Name: X138, dtype: int64

Unique values of column X139 :  
0 7634  
1 784  
Name: X139, dtype: int64

Unique values of column X140 :  
0 8082  
1 336  
Name: X140, dtype: int64

Unique values of column X141 :  
0 8289  
1 129  
Name: X141, dtype: int64

Unique values of column X142 :  
1 6465

0 1953  
Name: X142, dtype: int64

Unique values of column X143 :  
0 8090  
1 328  
Name: X143, dtype: int64

Unique values of column X144 :  
1 6774  
0 1644  
Name: X144, dtype: int64

Unique values of column X145 :  
0 8409  
1 9  
Name: X145, dtype: int64

Unique values of column X146 :  
0 8080  
1 338  
Name: X146, dtype: int64

Unique values of column X147 :  
0 8224  
1 194  
Name: X147, dtype: int64

Unique values of column X148 :  
0 8059  
1 359  
Name: X148, dtype: int64

Unique values of column X150 :  
1 6652  
0 1766  
Name: X150, dtype: int64

Unique values of column X151 :  
0 7690  
1 728  
Name: X151, dtype: int64

Unique values of column X152 :  
0 8150  
1 268  
Name: X152, dtype: int64

Unique values of column X153 :  
0 8410  
1 8  
Name: X153, dtype: int64

Unique values of column X154 :  
0 6641  
1 1777  
Name: X154, dtype: int64

Unique values of column X155 :  
0 7761  
1 657  
Name: X155, dtype: int64

Unique values of column X156 :  
1 6007  
0 2411  
Name: X156, dtype: int64

Unique values of column X157 :  
0 6007  
1 2411  
Name: X157, dtype: int64

Unique values of column X158 :  
0 6464  
1 1954  
Name: X158, dtype: int64

Unique values of column X159 :  
0 8299  
1 119  
Name: X159, dtype: int64

Unique values of column X160 :  
0 8409  
1 9  
Name: X160, dtype: int64

Unique values of column X161 :  
0 6765  
1 1653  
Name: X161, dtype: int64

Unique values of column X162 :  
0 8096  
1 322  
Name: X162, dtype: int64

Unique values of column X163 :  
0 5818  
1 2600  
Name: X163, dtype: int64

Unique values of column X164 :  
0 7870  
1 548  
Name: X164, dtype: int64

Unique values of column X165 :  
0 8384  
1 34  
Name: X165, dtype: int64

Unique values of column X166 :  
0 8164  
1 254  
Name: X166, dtype: int64

Unique values of column X167 :  
0 8409  
1 9  
Name: X167, dtype: int64

Unique values of column X168 :  
0 6168  
1 2250  
Name: X168, dtype: int64

Unique values of column X169 :  
0 8359  
1 59  
Name: X169, dtype: int64

Unique values of column X170 :  
0 8203  
1 215  
Name: X170, dtype: int64

Unique values of column X171 :  
1 5572  
0 2846  
Name: X171, dtype: int64

Unique values of column X172 :  
0 8368  
1 50  
Name: X172, dtype: int64

Unique values of column X173 :  
0 8334  
1 84  
Name: X173, dtype: int64

Unique values of column X174 :  
0 8263  
1 155  
Name: X174, dtype: int64

Unique values of column X175 :  
0 8228  
1 190  
Name: X175, dtype: int64

Unique values of column X176 :  
0 8285  
1 133  
Name: X176, dtype: int64

Unique values of column X177 :  
0 7989  
1 429  
Name: X177, dtype: int64

Unique values of column X178 :  
1 4705  
0 3713  
Name: X178, dtype: int64

Unique values of column X179 :  
0 8026  
1 392  
Name: X179, dtype: int64

Unique values of column X180 :  
0 7086  
1 1332  
Name: X180, dtype: int64

Unique values of column X181 :  
0 7633  
1 785  
Name: X181, dtype: int64

Unique values of column X182 :  
0 7560  
1 858  
Name: X182, dtype: int64

Unique values of column X183 :  
0 8392  
1 26  
Name: X183, dtype: int64

Unique values of column X184 :  
0 8409  
1 9  
Name: X184, dtype: int64

Unique values of column X185 :  
0 8263  
1 155  
Name: X185, dtype: int64

Unique values of column X186 :  
1 4495  
0 3923  
Name: X186, dtype: int64

Unique values of column X187 :  
0 4845  
1 3573  
Name: X187, dtype: int64

Unique values of column X189 :  
1 7733  
0 685  
Name: X189, dtype: int64

Unique values of column X190 :  
0 8416  
1 2  
Name: X190, dtype: int64

Unique values of column X191 :  
0 4432

```
1      3986
Name: X191, dtype: int64
```

```
Unique values of column X192 :
0      8401
1        17
Name: X192, dtype: int64
```

```
Unique values of column X194 :
0      4494
1      3924
Name: X194, dtype: int64
```

```
Unique values of column X195 :
0      8318
1        100
Name: X195, dtype: int64
```

```
Unique values of column X196 :
0      8323
1         95
Name: X196, dtype: int64
```

```
Unique values of column X197 :
0      8158
1        260
Name: X197, dtype: int64
```

```
Unique values of column X198 :
0      8216
1        202
Name: X198, dtype: int64
```

```
Unique values of column X199 :
0      8401
1        17
Name: X199, dtype: int64
```

```
Unique values of column X200 :
0      8362
1         56
Name: X200, dtype: int64
```

```
Unique values of column X201 :
0      6928
1      1490
Name: X201, dtype: int64
```

```
Unique values of column X202 :
0      6393
1      2025
Name: X202, dtype: int64
```

```
Unique values of column X203 :
0      8297
1       121
Name: X203, dtype: int64
```

```
Unique values of column X204 :
0      8414
1         4
Name: X204, dtype: int64
```

```
Unique values of column X205 :
1      8414
0         4
Name: X205, dtype: int64
```

```
Unique values of column X206 :
0      8258
1       160
Name: X206, dtype: int64
```

```
Unique values of column X207 :
0      8411
1         7
Name: X207, dtype: int64
```

Unique values of column X208 :  
0 7893  
1 525  
Name: X208, dtype: int64

Unique values of column X209 :  
1 7558  
0 860  
Name: X209, dtype: int64

Unique values of column X210 :  
0 8416  
1 2  
Name: X210, dtype: int64

Unique values of column X211 :  
0 8283  
1 135  
Name: X211, dtype: int64

Unique values of column X212 :  
0 8365  
1 53  
Name: X212, dtype: int64

Unique values of column X213 :  
0 8401  
1 17  
Name: X213, dtype: int64

Unique values of column X214 :  
0 8364  
1 54  
Name: X214, dtype: int64

Unique values of column X215 :  
0 7541  
1 877  
Name: X215, dtype: int64

Unique values of column X216 :  
0 8368  
1 50  
Name: X216, dtype: int64

Unique values of column X217 :  
0 8349  
1 69  
Name: X217, dtype: int64

Unique values of column X218 :  
0 5751  
1 2667  
Name: X218, dtype: int64

Unique values of column X219 :  
0 7851  
1 567  
Name: X219, dtype: int64

Unique values of column X220 :  
1 4708  
0 3710  
Name: X220, dtype: int64

Unique values of column X221 :  
0 8349  
1 69  
Name: X221, dtype: int64

Unique values of column X222 :  
0 8224  
1 194  
Name: X222, dtype: int64

Unique values of column X223 :  
1 4678  
0 3740  
Name: X223, dtype: int64



Unique values of column X224 :  
0 5763  
1 2655  
Name: X224, dtype: int64

Unique values of column X225 :  
0 7647  
1 771  
Name: X225, dtype: int64

Unique values of column X226 :  
0 8150  
1 268  
Name: X226, dtype: int64

Unique values of column X227 :  
0 8385  
1 33  
Name: X227, dtype: int64

Unique values of column X228 :  
0 8057  
1 361  
Name: X228, dtype: int64

Unique values of column X229 :  
1 8043  
0 375  
Name: X229, dtype: int64

Unique values of column X230 :  
0 8380  
1 38  
Name: X230, dtype: int64

Unique values of column X231 :  
0 8278  
1 140  
Name: X231, dtype: int64

Unique values of column X232 :  
0 8070  
1 348  
Name: X232, dtype: int64

Unique values of column X233 :  
0 8417  
1 1  
Name: X233, dtype: int64

Unique values of column X234 :  
0 6701  
1 1717  
Name: X234, dtype: int64

Unique values of column X235 :  
0 8417  
1 1  
Name: X235, dtype: int64

Unique values of column X236 :  
0 8415  
1 3  
Name: X236, dtype: int64

Unique values of column X237 :  
0 8367  
1 51  
Name: X237, dtype: int64

Unique values of column X238 :  
1 7732  
0 686  
Name: X238, dtype: int64

Unique values of column X239 :  
0 8364

```
1      54
Name: X239, dtype: int64
```

```
Unique values of column X240 :
0      8391
1       27
Name: X240, dtype: int64
```

```
Unique values of column X241 :
0      7570
1       848
Name: X241, dtype: int64
```

```
Unique values of column X242 :
0      8352
1        66
Name: X242, dtype: int64
```

```
Unique values of column X243 :
0      8353
1        65
Name: X243, dtype: int64
```

```
Unique values of column X244 :
0      7528
1       890
Name: X244, dtype: int64
```

```
Unique values of column X245 :
0      8411
1         7
Name: X245, dtype: int64
```

```
Unique values of column X246 :
0      4940
1      3478
Name: X246, dtype: int64
```

```
Unique values of column X247 :
0      6394
1      2024
Name: X247, dtype: int64
```

```
Unique values of column X248 :
0      8403
1        15
Name: X248, dtype: int64
```

```
Unique values of column X249 :
0      8346
1        72
Name: X249, dtype: int64
```

```
Unique values of column X250 :
1      4667
0      3751
Name: X250, dtype: int64
```

```
Unique values of column X251 :
0      5101
1      3317
Name: X251, dtype: int64
```

```
Unique values of column X252 :
0      8414
1         4
Name: X252, dtype: int64
```

```
Unique values of column X253 :
0      8405
1        13
Name: X253, dtype: int64
```

```
Unique values of column X254 :
0      8380
1         38
Name: X254, dtype: int64
```

Unique values of column X255 :  
0 8276  
1 142  
Name: X255, dtype: int64

Unique values of column X256 :  
0 7788  
1 630  
Name: X256, dtype: int64

Unique values of column X257 :  
0 8417  
1 1  
Name: X257, dtype: int64

Unique values of column X258 :  
0 8408  
1 10  
Name: X258, dtype: int64

Unique values of column X259 :  
0 8416  
1 2  
Name: X259, dtype: int64

Unique values of column X260 :  
0 8414  
1 4  
Name: X260, dtype: int64

Unique values of column X261 :  
0 4809  
1 3609  
Name: X261, dtype: int64

Unique values of column X262 :  
0 8409  
1 9  
Name: X262, dtype: int64

Unique values of column X263 :  
1 8070  
0 348  
Name: X263, dtype: int64

Unique values of column X264 :  
0 8080  
1 338  
Name: X264, dtype: int64

Unique values of column X265 :  
1 7650  
0 768  
Name: X265, dtype: int64

Unique values of column X266 :  
0 8408  
1 10  
Name: X266, dtype: int64

Unique values of column X267 :  
0 8346  
1 72  
Name: X267, dtype: int64

Unique values of column X268 :  
0 8417  
1 1  
Name: X268, dtype: int64

Unique values of column X269 :  
0 8413  
1 5  
Name: X269, dtype: int64

Unique values of column X270 :  
0 8414  
1 4  
Name: X270, dtype: int64

Unique values of column X271 :  
0 8403  
1 15  
Name: X271, dtype: int64

Unique values of column X272 :  
0 8112  
1 306  
Name: X272, dtype: int64

Unique values of column X273 :  
1 6029  
0 2389  
Name: X273, dtype: int64

Unique values of column X274 :  
0 8329  
1 89  
Name: X274, dtype: int64

Unique values of column X275 :  
1 6111  
0 2307  
Name: X275, dtype: int64

Unique values of column X276 :  
0 8105  
1 313  
Name: X276, dtype: int64

Unique values of column X277 :  
0 8406  
1 12  
Name: X277, dtype: int64

Unique values of column X278 :  
0 8415  
1 3  
Name: X278, dtype: int64

Unique values of column X279 :  
0 8070  
1 348  
Name: X279, dtype: int64

Unique values of column X280 :  
0 8416  
1 2  
Name: X280, dtype: int64

Unique values of column X281 :  
0 8392  
1 26  
Name: X281, dtype: int64

Unique values of column X282 :  
0 8392  
1 26  
Name: X282, dtype: int64

Unique values of column X283 :  
0 7218  
1 1200  
Name: X283, dtype: int64

Unique values of column X284 :  
0 8068  
1 350  
Name: X284, dtype: int64

Unique values of column X285 :  
0 6666  
1 1752  
Name: X285, dtype: int64

Unique values of column X286 :  
0 7991

```
1      427
Name: X286, dtype: int64
```

```
Unique values of column X287 :
0      8286
1       132
Name: X287, dtype: int64
```

```
Unique values of column X288 :
0      8416
1         2
Name: X288, dtype: int64
```

```
Unique values of column X289 :
0      8416
1         2
Name: X289, dtype: int64
```

```
Unique values of column X290 :
0      8417
1         1
Name: X290, dtype: int64
```

```
Unique values of column X291 :
0      8329
1         89
Name: X291, dtype: int64
```

```
Unique values of column X292 :
0      8341
1         77
Name: X292, dtype: int64
```

```
Unique values of column X293 :
0      8417
1         1
Name: X293, dtype: int64
```

```
Unique values of column X294 :
0      7378
1     1040
Name: X294, dtype: int64
```

```
Unique values of column X295 :
0      8417
1         1
Name: X295, dtype: int64
```

```
Unique values of column X296 :
0      8417
1         1
Name: X296, dtype: int64
```

```
Unique values of column X297 :
0      8417
1         1
Name: X297, dtype: int64
```

```
Unique values of column X298 :
0      8386
1        32
Name: X298, dtype: int64
```

```
Unique values of column X299 :
0      8386
1        32
Name: X299, dtype: int64
```

```
Unique values of column X300 :
0      6652
1     1766
Name: X300, dtype: int64
```

```
Unique values of column X301 :
0      7989
1        429
Name: X301, dtype: int64
```

Unique values of column X302 :  
0 8341  
1 77  
Name: X302, dtype: int64

Unique values of column X304 :  
1 7802  
0 616  
Name: X304, dtype: int64

Unique values of column X305 :  
0 8327  
1 91  
Name: X305, dtype: int64

Unique values of column X306 :  
0 8044  
1 374  
Name: X306, dtype: int64

Unique values of column X307 :  
0 8395  
1 23  
Name: X307, dtype: int64

Unique values of column X308 :  
0 8348  
1 70  
Name: X308, dtype: int64

Unique values of column X309 :  
0 8360  
1 58  
Name: X309, dtype: int64

Unique values of column X310 :  
0 8389  
1 29  
Name: X310, dtype: int64

Unique values of column X311 :  
1 5034  
0 3384  
Name: X311, dtype: int64

Unique values of column X312 :  
0 8386  
1 32  
Name: X312, dtype: int64

Unique values of column X313 :  
0 5921  
1 2497  
Name: X313, dtype: int64

Unique values of column X314 :  
0 4714  
1 3704  
Name: X314, dtype: int64

Unique values of column X315 :  
0 8190  
1 228  
Name: X315, dtype: int64

Unique values of column X316 :  
0 6776  
1 1642  
Name: X316, dtype: int64

Unique values of column X317 :  
0 8362  
1 56  
Name: X317, dtype: int64

Unique values of column X318 :  
0 8411  
1 7  
Name: X318, dtype: int64

Unique values of column X319 :  
0 8414  
1 4  
Name: X319, dtype: int64

Unique values of column X320 :  
0 8355  
1 63  
Name: X320, dtype: int64

Unique values of column X321 :  
0 6391  
1 2027  
Name: X321, dtype: int64

Unique values of column X322 :  
0 8249  
1 169  
Name: X322, dtype: int64

Unique values of column X323 :  
0 8345  
1 73  
Name: X323, dtype: int64

Unique values of column X324 :  
1 4880  
0 3538  
Name: X324, dtype: int64

Unique values of column X325 :  
0 8371  
1 47  
Name: X325, dtype: int64

Unique values of column X326 :  
0 8150  
1 268  
Name: X326, dtype: int64

Unique values of column X327 :  
0 7329  
1 1089  
Name: X327, dtype: int64

Unique values of column X328 :  
0 8096  
1 322  
Name: X328, dtype: int64

Unique values of column X329 :  
0 4742  
1 3676  
Name: X329, dtype: int64

Unique values of column X330 :  
0 8417  
1 1  
Name: X330, dtype: int64

Unique values of column X331 :  
0 7964  
1 454  
Name: X331, dtype: int64

Unique values of column X332 :  
0 8412  
1 6  
Name: X332, dtype: int64

Unique values of column X333 :  
0 8235  
1 183  
Name: X333, dtype: int64

Unique values of column X334 :  
0 4570

```
1      3848
Name: X334, dtype: int64
```

```
Unique values of column X335 :
0      8394
1        24
Name: X335, dtype: int64
```

```
Unique values of column X336 :
0      7345
1      1073
Name: X336, dtype: int64
```

```
Unique values of column X337 :
1      4392
0      4026
Name: X337, dtype: int64
```

```
Unique values of column X338 :
0      8356
1        62
Name: X338, dtype: int64
```

```
Unique values of column X339 :
0      8416
1         2
Name: X339, dtype: int64
```

```
Unique values of column X340 :
0      8238
1       180
Name: X340, dtype: int64
```

```
Unique values of column X341 :
0      8342
1        76
Name: X341, dtype: int64
```

```
Unique values of column X342 :
0      8239
1       179
Name: X342, dtype: int64
```

```
Unique values of column X343 :
0      7795
1       623
Name: X343, dtype: int64
```

```
Unique values of column X344 :
0      8342
1        76
Name: X344, dtype: int64
```

```
Unique values of column X345 :
0      8220
1       198
Name: X345, dtype: int64
```

```
Unique values of column X346 :
0      8029
1       389
Name: X346, dtype: int64
```

```
Unique values of column X347 :
0      8416
1         2
Name: X347, dtype: int64
```

```
Unique values of column X348 :
1      7986
0       432
Name: X348, dtype: int64
```

```
Unique values of column X349 :
0      8053
1       365
Name: X349, dtype: int64
```



Unique values of column X350 :  
0 5556  
1 2862  
Name: X350, dtype: int64

Unique values of column X351 :  
0 5958  
1 2460  
Name: X351, dtype: int64

Unique values of column X352 :  
0 7979  
1 439  
Name: X352, dtype: int64

Unique values of column X353 :  
0 8398  
1 20  
Name: X353, dtype: int64

Unique values of column X354 :  
0 6706  
1 1712  
Name: X354, dtype: int64

Unique values of column X355 :  
0 5227  
1 3191  
Name: X355, dtype: int64

Unique values of column X356 :  
0 6886  
1 1532  
Name: X356, dtype: int64

Unique values of column X357 :  
0 8403  
1 15  
Name: X357, dtype: int64

Unique values of column X358 :  
0 4796  
1 3622  
Name: X358, dtype: int64

Unique values of column X359 :  
0 8129  
1 289  
Name: X359, dtype: int64

Unique values of column X360 :  
0 7761  
1 657  
Name: X360, dtype: int64

Unique values of column X361 :  
1 8117  
0 301  
Name: X361, dtype: int64

Unique values of column X362 :  
1 4370  
0 4048  
Name: X362, dtype: int64

Unique values of column X363 :  
1 6310  
0 2108  
Name: X363, dtype: int64

Unique values of column X364 :  
0 8391  
1 27  
Name: X364, dtype: int64

Unique values of column X365 :  
0 8392  
1 26  
Name: X365, dtype: int64

Unique values of column X366 :  
0 8406  
1 12  
Name: X366, dtype: int64

Unique values of column X367 :  
0 7997  
1 421  
Name: X367, dtype: int64

Unique values of column X368 :  
0 7888  
1 530  
Name: X368, dtype: int64

Unique values of column X369 :  
0 8416  
1 2  
Name: X369, dtype: int64

Unique values of column X370 :  
0 8366  
1 52  
Name: X370, dtype: int64

Unique values of column X371 :  
0 8310  
1 108  
Name: X371, dtype: int64

Unique values of column X372 :  
0 8415  
1 3  
Name: X372, dtype: int64

Unique values of column X373 :  
0 8276  
1 142  
Name: X373, dtype: int64

Unique values of column X374 :  
0 6468  
1 1950  
Name: X374, dtype: int64

Unique values of column X375 :  
0 5704  
1 2714  
Name: X375, dtype: int64

Unique values of column X376 :  
0 7968  
1 450  
Name: X376, dtype: int64

Unique values of column X377 :  
0 5780  
1 2638  
Name: X377, dtype: int64

Unique values of column X378 :  
0 8250  
1 168  
Name: X378, dtype: int64

Unique values of column X379 :  
0 8328  
1 90  
Name: X379, dtype: int64

Unique values of column X380 :  
0 8350  
1 68  
Name: X380, dtype: int64

Unique values of column X382 :  
0 8349

```

1      69
Name: X382, dtype: int64

Unique values of column X383 :
0      8409
1         9
Name: X383, dtype: int64

Unique values of column X384 :
0      8413
1         5
Name: X384, dtype: int64

Unique values of column X385 :
0      8405
1         13
Name: X385, dtype: int64

Unique values of column Source :
train    4209
test     4209
Name: Source, dtype: int64

```

- Since 8 columns are categorical, the label encoder will be apply. The ninth categorical column in the "Source" that needs to be kept to separate the train and test datasets before the modeling.

#### Applying Label Encoder

```

In [16]: label_encoder = LabelEncoder()
for column in all_data.columns[:377]:
    if all_data[column].dtypes == 'object':
        all_data[column] = label_encoder.fit_transform(all_data[column])

all_data.head()

```

```

Out[16]:
   y  X0  X1  X2  X3  X4  X5  X6  X8  X10  ...  X376  X377  X378  X379  X380  X382  X383  X384  X385  Source
0  130.81  37  23  20  0  3  27  9  14  0  ...  0  1  0  0  0  0  0  0  0  train
1  88.53  37  21  22  4  3  31  11  14  0  ...  0  0  0  0  0  0  0  0  0  train
2  76.26  24  24  38  2  3  30  9  23  0  ...  0  0  0  0  0  1  0  0  0  train
3  80.62  24  21  38  5  3  30  11  4  0  ...  0  0  0  0  0  0  0  0  0  train
4  78.02  24  23  38  5  3  14  3  13  0  ...  0  0  0  0  0  0  0  0  0  train

```

5 rows × 378 columns

```

In [17]: all_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8418 entries, 0 to 8417
Columns: 378 entries, y to Source
dtypes: float64(1), int64(376), object(1)
memory usage: 24.3+ MB

```

#### Dimensionality reduction

```

In [18]: correlation_matrix = all_data.corr()

```

```

In [19]: list_no_corr = []
for column in correlation_matrix.columns:
    if np.isnan(correlation_matrix.loc["y", column]) == True:
        print(f'y has no correlation with {column}')
        list_no_corr.append(column)

print(list_no_corr)
print(len(list_no_corr))

y has no correlation with X11
y has no correlation with X93
y has no correlation with X107
y has no correlation with X233
y has no correlation with X235
y has no correlation with X268
y has no correlation with X289
y has no correlation with X290
y has no correlation with X293
y has no correlation with X297
y has no correlation with X330
y has no correlation with X347
['X11', 'X93', 'X107', 'X233', 'X235', 'X268', 'X289', 'X290', 'X293', 'X297', 'X330', 'X347']
12

```

- The 12 columns can be removed since they have no correlation with the target.

```
In [20]: all_data.drop(list_no_corr, axis=1, inplace=True)
all_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8418 entries, 0 to 8417
Columns: 366 entries, y to Source
dtypes: float64(1), int64(364), object(1)
memory usage: 23.5+ MB
```

```
In [21]: new_correlation_matrix = all_data.corr()
new_correlation_matrix.iloc[:20, :20]
```

```
Out[21]:
```

	y	X0	X1	X2	X3	X4	X5	X6	X8	X10	X12	X13	X14
y	1.000000	-0.190406	-0.003032	0.078843	-0.150741	-0.015272	-0.039209	0.004252	0.003328	-0.026985	0.089792	0.048276	0.193643
X0	-0.190406	1.000000	-0.286511	-0.151901	-0.090605	-0.003811	0.012749	0.024163	0.054237	0.090390	-0.139645	-0.139466	-0.144014
X1	-0.003032	-0.286511	1.000000	0.104661	0.198627	-0.002193	0.058790	-0.083805	-0.019680	-0.148216	0.105982	0.294716	0.116566
X2	0.078843	-0.151901	0.104661	1.000000	-0.096130	-0.002317	-0.013948	0.070806	-0.050417	0.056485	0.141602	0.231291	-0.055430
X3	-0.150741	-0.090605	0.198627	-0.096130	1.000000	0.009266	0.001079	-0.038184	0.009907	0.015532	0.047579	-0.219658	0.063425
X4	-0.015272	-0.003811	-0.002193	-0.002317	0.009266	1.000000	0.044978	0.039539	-0.019006	0.004192	0.009297	-0.004940	0.003112
X5	-0.039209	0.012749	0.058790	-0.013948	0.001079	0.044978	1.000000	-0.011910	0.010341	-0.010721	0.046590	0.010475	-0.005719
X6	0.004252	0.024163	-0.083805	0.070806	-0.038184	0.039539	-0.011910	1.000000	0.024626	0.082832	-0.091909	-0.050917	0.023021
X8	0.003328	0.054237	-0.019680	-0.050417	0.009907	-0.019006	0.010341	0.024626	1.000000	0.028762	-0.066278	-0.055306	0.028201
X10	-0.026985	0.090390	-0.148216	0.056485	0.015532	0.004192	-0.010721	0.082832	0.028762	1.000000	-0.036416	-0.032236	-0.110850
X12	0.089792	-0.139645	0.105982	0.141602	0.047579	0.009297	0.046590	-0.091909	-0.066278	-0.036416	1.000000	0.249327	-0.245821
X13	0.048276	-0.139466	0.294716	0.231291	-0.219658	-0.004940	0.010475	-0.050917	-0.055306	-0.032236	0.249327	1.000000	-0.079594
X14	0.193643	-0.144014	0.116566	-0.055430	0.063425	0.003112	-0.005719	0.023021	0.028201	-0.110850	-0.245821	-0.079594	1.000000
X15	0.023116	0.020313	-0.013406	-0.009161	-0.010074	0.000798	-0.006433	-0.032273	-0.012437	-0.003124	-0.006928	-0.006133	-0.021088
X16	0.048946	-0.027842	-0.003564	-0.018349	0.008754	-0.028839	-0.025271	0.031184	0.016934	-0.006560	-0.014547	-0.012877	-0.044280
X17	-0.159815	-0.058191	0.122617	0.025315	-0.047906	0.002974	-0.020284	0.040452	0.001436	-0.011650	-0.020823	-0.022869	0.006570
X18	-0.001789	-0.022561	0.048670	-0.065239	-0.013157	0.003123	-0.000955	-0.032566	0.032632	-0.012231	-0.027124	-0.018703	-0.082567
X19	-0.168895	0.211862	-0.232817	-0.328452	-0.069846	0.001104	-0.006815	0.034355	0.040059	-0.044034	-0.097649	-0.086441	-0.297246
X20	-0.095148	0.029729	0.018618	-0.487231	0.061168	-0.004637	0.023529	-0.117944	-0.005906	-0.051945	-0.115192	-0.044282	-0.350649
X21	0.031111	-0.024495	0.063655	-0.016593	-0.028781	0.001514	-0.012332	-0.000819	0.048853	-0.005932	-0.013155	-0.000775	-0.040043

```
In [22]: for row in new_correlation_matrix.index:
list_corr = []
for column in new_correlation_matrix.columns:
if new_correlation_matrix.loc[row, column] >= 0.65 and row != column:
list_corr.append(column)
if len(list_corr) > 0:
print(f'{row} is highly correlated to {list_corr}')
```

X14 is highly correlated to ['X191', 'X223', 'X224', 'X251']  
X15 is highly correlated to ['X260']  
X17 is highly correlated to ['X325', 'X382']  
X18 is highly correlated to ['X125', 'X227']  
X19 is highly correlated to ['X215']  
X21 is highly correlated to ['X184', 'X262', 'X266']  
X28 is highly correlated to ['X29', 'X54', 'X76', 'X148', 'X162', 'X166', 'X232', 'X272', 'X276', 'X279', 'X328']  
X29 is highly correlated to ['X28', 'X54', 'X76', 'X162', 'X166', 'X185', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X31 is highly correlated to ['X35', 'X37']  
X32 is highly correlated to ['X200']  
X33 is highly correlated to ['X39']  
X35 is highly correlated to ['X31', 'X37']  
X37 is highly correlated to ['X31', 'X35']  
X39 is highly correlated to ['X33']  
X44 is highly correlated to ['X302', 'X317']  
X45 is highly correlated to ['X161', 'X202', 'X247', 'X356', 'X377']  
X47 is highly correlated to ['X267']  
X48 is highly correlated to ['X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X198', 'X222', 'X264', 'X306']  
X49 is highly correlated to ['X50', 'X129']  
X50 is highly correlated to ['X49', 'X129']  
X52 is highly correlated to ['X48', 'X66', 'X75', 'X113', 'X134', 'X147', 'X198', 'X222']  
X53 is highly correlated to ['X102', 'X214', 'X239', 'X281', 'X282', 'X308', 'X370']  
X54 is highly correlated to ['X28', 'X29', 'X76', 'X162', 'X166', 'X185', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X57 is highly correlated to ['X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X58 is highly correlated to ['X137', 'X324']  
X60 is highly correlated to ['X248', 'X253', 'X385']  
X61 is highly correlated to ['X111', 'X120']  
X62 is highly correlated to ['X172', 'X216', 'X221']  
X63 is highly correlated to ['X78', 'X174', 'X379']  
X66 is highly correlated to ['X48', 'X52', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X198', 'X222', 'X264', 'X306']  
X67 is highly correlated to ['X213']  
X69 is highly correlated to ['X197', 'X331', 'X346', 'X349', 'X352', 'X367']  
X71 is highly correlated to ['X84', 'X241', 'X244']  
X75 is highly correlated to ['X48', 'X52', 'X66', 'X113', 'X134', 'X147', 'X170', 'X198', 'X222']  
X76 is highly correlated to ['X28', 'X29', 'X54', 'X162', 'X166', 'X185', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X78 is highly correlated to ['X63', 'X379']  
X80 is highly correlated to ['X98', 'X101', 'X111', 'X128', 'X265', 'X304', 'X348']  
X84 is highly correlated to ['X71', 'X241', 'X244']  
X85 is highly correlated to ['X187', 'X194', 'X246', 'X358']  
X87 is highly correlated to ['X245']  
X88 is highly correlated to ['X57', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X89 is highly correlated to ['X245']  
X90 is highly correlated to ['X57', 'X88', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X94 is highly correlated to ['X57', 'X88', 'X90', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X96 is highly correlated to ['X150', 'X363']  
X98 is highly correlated to ['X80', 'X101', 'X128', 'X304', 'X348']  
X99 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X100 is highly correlated to ['X103', 'X144']  
X101 is highly correlated to ['X80', 'X98', 'X128', 'X304', 'X348']  
X102 is highly correlated to ['X53', 'X214', 'X239', 'X281', 'X282', 'X308', 'X370']  
X103 is highly correlated to ['X100', 'X144']  
X108 is highly correlated to ['X371']  
X109 is highly correlated to ['X177']  
X111 is highly correlated to ['X61', 'X80', 'X120', 'X128', 'X348']  
X112 is highly correlated to ['X183', 'X199']  
X113 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X126', 'X130', 'X134', 'X147', 'X170', 'X198', 'X222', 'X264', 'X306']  
X114 is highly correlated to ['X155', 'X360']  
X116 is highly correlated to ['X294']  
X117 is highly correlated to ['X135']  
X118 is highly correlated to ['X119', 'X275', 'X311']  
X119 is highly correlated to ['X118', 'X275', 'X311']  
X120 is highly correlated to ['X61', 'X111']  
X122 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X217', 'X242', 'X243', 'X249', 'X320']  
X125 is highly correlated to ['X18', 'X227']  
X126 is highly correlated to ['X48', 'X66', 'X113', 'X130', 'X134', 'X147', 'X170', 'X179', 'X198', 'X222', 'X264', 'X306']  
X127 is highly correlated to ['X313']  
X128 is highly correlated to ['X80', 'X98', 'X101', 'X111', 'X189', 'X304', 'X348']  
X129 is highly correlated to ['X49', 'X50']  
X130 is highly correlated to ['X48', 'X66', 'X113', 'X126', 'X134', 'X147', 'X170', 'X179', 'X198', 'X222', 'X264']  
X133 is highly correlated to ['X161', 'X202', 'X247', 'X356']  
X134 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X147', 'X170', 'X198', 'X222', 'X264', 'X306']  
X135 is highly correlated to ['X117']  
X136 is highly correlated to ['X189', 'X238', 'X263', 'X265']  
X137 is highly correlated to ['X58', 'X324']  
X138 is highly correlated to ['X140', 'X146', 'X208', 'X368']  
X140 is highly correlated to ['X138', 'X146', 'X208', 'X368']  
X144 is highly correlated to ['X100', 'X103', 'X209']  
X146 is highly correlated to ['X138', 'X140', 'X208', 'X368']  
X147 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X170', 'X198', 'X222', 'X264', 'X306']  
X148 is highly correlated to ['X28']  
X150 is highly correlated to ['X96', 'X363']  
X152 is highly correlated to ['X226', 'X326']  
X154 is highly correlated to ['X283', 'X321', 'X374']  
X155 is highly correlated to ['X114', 'X219', 'X360']  
X159 is highly correlated to ['X274']  
X161 is highly correlated to ['X45', 'X133', 'X202', 'X247', 'X356', 'X377']  
X162 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X166', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X166 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X167 is highly correlated to ['X277']  
X170 is highly correlated to ['X48', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X198', 'X222', 'X228', 'X264', 'X306']  
X172 is highly correlated to ['X62', 'X216', 'X221']  
X174 is highly correlated to ['X63', 'X379']  
X177 is highly correlated to ['X109']  
X178 is highly correlated to ['X250']

X179 is highly correlated to ['X126', 'X130', 'X264']  
X181 is highly correlated to ['X356']  
X183 is highly correlated to ['X112', 'X199']  
X184 is highly correlated to ['X21', 'X262', 'X266']  
X185 is highly correlated to ['X29', 'X54', 'X76', 'X232', 'X279', 'X378']  
X186 is highly correlated to ['X362']  
X187 is highly correlated to ['X85', 'X194', 'X246', 'X358']  
X189 is highly correlated to ['X128', 'X136', 'X238', 'X263', 'X265']  
X191 is highly correlated to ['X14', 'X224', 'X251']  
X194 is highly correlated to ['X85', 'X187', 'X246', 'X358']  
X197 is highly correlated to ['X69', 'X331', 'X346', 'X349', 'X352', 'X367']  
X198 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X222', 'X264', 'X306']  
X199 is highly correlated to ['X112', 'X183']  
X200 is highly correlated to ['X32']  
X202 is highly correlated to ['X45', 'X133', 'X161', 'X247', 'X356', 'X377']  
X208 is highly correlated to ['X138', 'X140', 'X146', 'X368']  
X209 is highly correlated to ['X144']  
X213 is highly correlated to ['X67']  
X214 is highly correlated to ['X53', 'X102', 'X239', 'X281', 'X282', 'X308', 'X370']  
X215 is highly correlated to ['X19']  
X216 is highly correlated to ['X62', 'X172', 'X221']  
X217 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X242', 'X243', 'X249', 'X320']  
X219 is highly correlated to ['X155', 'X360']  
X221 is highly correlated to ['X62', 'X172', 'X216']  
X222 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X198', 'X264', 'X306']  
X223 is highly correlated to ['X14', 'X251']  
X224 is highly correlated to ['X14', 'X191', 'X251']  
X226 is highly correlated to ['X152', 'X326']  
X227 is highly correlated to ['X18', 'X125']  
X228 is highly correlated to ['X170']  
X230 is highly correlated to ['X254']  
X232 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X185', 'X272', 'X276', 'X279', 'X328', 'X378']  
X236 is highly correlated to ['X339']  
X237 is highly correlated to ['X309', 'X312', 'X317']  
X238 is highly correlated to ['X136', 'X189', 'X263', 'X265']  
X239 is highly correlated to ['X53', 'X102', 'X214', 'X281', 'X282', 'X308', 'X370']  
X240 is highly correlated to ['X364', 'X365', 'X366']  
X241 is highly correlated to ['X71', 'X84', 'X244']  
X242 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X243 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X249', 'X320']  
X244 is highly correlated to ['X71', 'X84', 'X241']  
X245 is highly correlated to ['X87', 'X89']  
X246 is highly correlated to ['X85', 'X187', 'X194', 'X358']  
X247 is highly correlated to ['X45', 'X133', 'X161', 'X202', 'X356', 'X377']  
X248 is highly correlated to ['X60', 'X253', 'X385']  
X249 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X320']  
X250 is highly correlated to ['X178']  
X251 is highly correlated to ['X14', 'X191', 'X223', 'X224']  
X253 is highly correlated to ['X60', 'X248', 'X385']  
X254 is highly correlated to ['X230']  
X255 is highly correlated to ['X308']  
X260 is highly correlated to ['X15']  
X261 is highly correlated to ['X314']  
X262 is highly correlated to ['X21', 'X184', 'X266']  
X263 is highly correlated to ['X136', 'X189', 'X238', 'X265']  
X264 is highly correlated to ['X48', 'X66', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X179', 'X198', 'X222', 'X306']  
X265 is highly correlated to ['X80', 'X136', 'X189', 'X238', 'X263', 'X348']  
X266 is highly correlated to ['X21', 'X184', 'X262']  
X267 is highly correlated to ['X47']  
X272 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X232', 'X276', 'X279', 'X328']  
X274 is highly correlated to ['X159']  
X275 is highly correlated to ['X118', 'X119', 'X311']  
X276 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X277 is highly correlated to ['X167']  
X279 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X185', 'X232', 'X272', 'X276', 'X328', 'X378']  
X281 is highly correlated to ['X53', 'X102', 'X214', 'X239', 'X370']  
X282 is highly correlated to ['X53', 'X102', 'X214', 'X239', 'X370']  
X283 is highly correlated to ['X154', 'X321', 'X374']  
X285 is highly correlated to ['X336']  
X294 is highly correlated to ['X116']  
X295 is highly correlated to ['X296']  
X296 is highly correlated to ['X295']  
X298 is highly correlated to ['X299']  
X299 is highly correlated to ['X298']  
X302 is highly correlated to ['X44', 'X317']  
X304 is highly correlated to ['X80', 'X98', 'X101', 'X128', 'X348']  
X306 is highly correlated to ['X48', 'X66', 'X113', 'X126', 'X134', 'X147', 'X170', 'X198', 'X222', 'X264']  
X308 is highly correlated to ['X53', 'X102', 'X214', 'X239', 'X255', 'X370']  
X309 is highly correlated to ['X237', 'X312', 'X317']  
X311 is highly correlated to ['X118', 'X119', 'X275']  
X312 is highly correlated to ['X237', 'X309', 'X317']  
X313 is highly correlated to ['X127']  
X314 is highly correlated to ['X261']  
X317 is highly correlated to ['X44', 'X237', 'X302', 'X309', 'X312']  
X320 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249']  
X321 is highly correlated to ['X154', 'X283', 'X374']  
X324 is highly correlated to ['X58', 'X137']  
X325 is highly correlated to ['X17', 'X382']  
X326 is highly correlated to ['X152', 'X226']  
X328 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X232', 'X272', 'X276', 'X279', 'X378']  
X331 is highly correlated to ['X69', 'X197', 'X346', 'X349', 'X352', 'X367']  
X333 is highly correlated to ['X373']  
X334 is highly correlated to ['X377']  
X336 is highly correlated to ['X285']

X339 is highly correlated to ['X236']  
 X343 is highly correlated to ['X376']  
 X346 is highly correlated to ['X69', 'X197', 'X331', 'X349', 'X352', 'X367']  
 X348 is highly correlated to ['X80', 'X98', 'X101', 'X111', 'X128', 'X265', 'X304']  
 X349 is highly correlated to ['X69', 'X197', 'X331', 'X346', 'X352', 'X367']  
 X352 is highly correlated to ['X69', 'X197', 'X331', 'X346', 'X349', 'X367']  
 X356 is highly correlated to ['X45', 'X133', 'X161', 'X181', 'X202', 'X247']  
 X358 is highly correlated to ['X85', 'X187', 'X194', 'X246']  
 X360 is highly correlated to ['X114', 'X155', 'X219']  
 X362 is highly correlated to ['X186']  
 X363 is highly correlated to ['X96', 'X150']  
 X364 is highly correlated to ['X240', 'X365', 'X366']  
 X365 is highly correlated to ['X240', 'X364']  
 X366 is highly correlated to ['X240', 'X364']  
 X367 is highly correlated to ['X69', 'X197', 'X331', 'X346', 'X349', 'X352']  
 X368 is highly correlated to ['X138', 'X140', 'X146', 'X208']  
 X370 is highly correlated to ['X53', 'X102', 'X214', 'X239', 'X281', 'X282', 'X308']  
 X371 is highly correlated to ['X108']  
 X373 is highly correlated to ['X333']  
 X374 is highly correlated to ['X154', 'X283', 'X321']  
 X376 is highly correlated to ['X343']  
 X377 is highly correlated to ['X45', 'X161', 'X202', 'X247', 'X334']  
 X378 is highly correlated to ['X29', 'X54', 'X76', 'X162', 'X166', 'X185', 'X232', 'X276', 'X279', 'X328']  
 X379 is highly correlated to ['X63', 'X78', 'X174']  
 X382 is highly correlated to ['X17', 'X325']  
 X385 is highly correlated to ['X60', 'X248', 'X253']

- Some columns are highly correlated together so the features the most correlated to 'y' will be kept.

```
In [23]: features_to_keep = []
for row in new_correlation_matrix.index:
    dict_corr = {row: new_correlation_matrix.loc["y", row]}
    for column in new_correlation_matrix.columns:
        if new_correlation_matrix.loc[row, column] >= 0.65 and row != column:
            dict_corr[column] = new_correlation_matrix.loc["y", column]
    to_keep = max(dict_corr, key=dict_corr.get)
    if to_keep not in features_to_keep:
        features_to_keep.append(to_keep)

print(sorted(features_to_keep))
print(len(features_to_keep))

['X0', 'X1', 'X10', 'X100', 'X102', 'X104', 'X105', 'X106', 'X109', 'X110', 'X115', 'X116', 'X117', 'X12', 'X120', 'X123', 'X124', 'X13', 'X131', 'X132', 'X137', 'X139', 'X141', 'X142', 'X143', 'X144', 'X145', 'X148', 'X150', 'X151', 'X152', 'X153', 'X155', 'X156', 'X157', 'X158', 'X16', 'X160', 'X163', 'X164', 'X165', 'X167', 'X168', 'X169', 'X170', 'X171', 'X173', 'X175', 'X176', 'X178', 'X18', 'X180', 'X181', 'X182', 'X183', 'X185', 'X186', 'X187', 'X189', 'X19', 'X190', 'X192', 'X195', 'X196', 'X2', 'X20', 'X200', 'X201', 'X202', 'X203', 'X204', 'X205', 'X206', 'X207', 'X208', 'X21', 'X210', 'X211', 'X212', 'X213', 'X214', 'X218', 'X22', 'X220', 'X221', 'X225', 'X226', 'X229', 'X23', 'X230', 'X231', 'X234', 'X236', 'X239', 'X24', 'X240', 'X244', 'X245', 'X248', 'X251', 'X252', 'X253', 'X254', 'X255', 'X256', 'X257', 'X258', 'X259', 'X26', 'X260', 'X263', 'X264', 'X265', 'X269', 'X27', 'X270', 'X271', 'X273', 'X274', 'X275', 'X278', 'X28', 'X280', 'X283', 'X284', 'X286', 'X287', 'X288', 'X291', 'X292', 'X295', 'X296', 'X298', 'X299', 'X3', 'X30', 'X300', 'X301', 'X305', 'X307', 'X31', 'X310', 'X313', 'X314', 'X315', 'X316', 'X317', 'X318', 'X319', 'X322', 'X323', 'X325', 'X326', 'X327', 'X329', 'X33', 'X331', 'X332', 'X333', 'X335', 'X336', 'X337', 'X338', 'X34', 'X340', 'X341', 'X342', 'X343', 'X344', 'X345', 'X35', 'X350', 'X351', 'X353', 'X354', 'X355', 'X357', 'X359', 'X36', 'X360', 'X361', 'X364', 'X365', 'X369', 'X37', 'X371', 'X372', 'X375', 'X377', 'X379', 'X38', 'X380', 'X383', 'X384', 'X385', 'X39', 'X4', 'X40', 'X41', 'X42', 'X43', 'X46', 'X47', 'X5', 'X50', 'X51', 'X53', 'X55', 'X56', 'X57', 'X59', 'X6', 'X60', 'X64', 'X65', 'X66', 'X67', 'X68', 'X70', 'X71', 'X73', 'X74', 'X77', 'X79', 'X8', 'X81', 'X82', 'X83', 'X84', 'X86', 'X89', 'X91', 'X92', 'X95', 'X97', 'X98', 'y']
237
```

- We will keep 236 features for modeling.

```
In [24]: features_to_keep.append('Source')
final_data = all_data[features_to_keep]
final_data.head()
```

```
Out[24]:
```

	y	X0	X1	X2	X3	X4	X5	X6	X8	X10	...	X364	X365	X369	X372	X375	X380	X383	X384	X385	Source
0	130.81	37	23	20	0	3	27	9	14	0	...	0	0	0	0	0	0	0	0	0	train
1	88.53	37	21	22	4	3	31	11	14	0	...	0	0	0	0	1	0	0	0	0	train
2	76.26	24	24	38	2	3	30	9	23	0	...	0	0	0	0	0	0	0	0	0	train
3	80.62	24	21	38	5	3	30	11	4	0	...	0	0	0	1	0	0	0	0	0	train
4	78.02	24	23	38	5	3	14	3	13	0	...	0	0	0	0	0	0	0	0	0	train

5 rows × 238 columns

```
In [25]: final_data_train = final_data.loc[final_data["Source"] == "train"]
final_data_train.drop(['Source'], axis=1, inplace=True)
final_data_test = final_data.loc[final_data["Source"] == "test"]
final_data_test.drop(['y', "Source"], axis=1, inplace=True)
```

```
In [26]: final_data_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4209 entries, 0 to 4208
Columns: 237 entries, y to X385
dtypes: float64(1), int64(236)
memory usage: 7.6 MB
```

```
In [27]: final_data_test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4209 entries, 4209 to 8417
Columns: 236 entries, X0 to X385
dtypes: int64(236)
memory usage: 7.6 MB
```

Splitting the train dataset

```
In [28]: y_train = final_data_train["y"]
X_train = final_data_train.drop(['y'], axis=1)
```

```
In [29]: X_train.head()
```

```
Out[29]:
```

	X0	X1	X2	X3	X4	X5	X6	X8	X10	X12	...	X361	X364	X365	X369	X372	X375	X380	X383	X384	X385
0	37	23	20	0	3	27	9	14	0	0	...	1	0	0	0	0	0	0	0	0	0
1	37	21	22	4	3	31	11	14	0	0	...	1	0	0	0	0	1	0	0	0	0
2	24	24	38	2	3	30	9	23	0	0	...	1	0	0	0	0	0	0	0	0	0
3	24	21	38	5	3	30	11	4	0	0	...	1	0	0	0	1	0	0	0	0	0
4	24	23	38	5	3	14	3	13	0	0	...	1	0	0	0	0	0	0	0	0	0

5 rows × 236 columns

```
In [30]: y_train.head()
```

```
Out[30]:
```

0	130.81
1	88.53
2	76.26
3	80.62
4	78.02

Name: y, dtype: float64

Creation of a train and test split for the model

```
In [31]: xtrain, xtest, ytrain, ytest = train_test_split(X_train, y_train, test_size=0.3, random_state=67)
```

Model 1: Fitting the XGBoost Regressor to the training dataset with K-fold cross validation

```
In [32]: xgboost_regressor = XGBRegressor(objective='reg:squarederror', n_estimators=10, random_state=67)
xgboost_regressor.fit(xtrain, ytrain)
```

```
Out[32]:
```

▼ XGBRegressor

XGBRegressor(base\_score=0.5, booster='gbtree', callbacks=None, colsample\_bylevel=1, colsample\_bynode=1, colsample\_bytree=1, early\_stopping\_rounds=None, enable\_categorical=False, eval\_metric=None, gamma=0, gpu\_id=-1, grow\_policy='depthwise', importance\_type=None, interaction\_constraints='', learning\_rate=0.300000012, max\_bin=256, max\_cat\_to\_onehot=4, max\_delta\_step=0, max\_depth=6, max\_leaves=0, min\_child\_weight=1, missing=nan, monotone\_constraints=(), n\_estimators=10, n\_jobs=0, num\_parallel\_tree=1, predictor='auto', random\_state=67, reg\_alpha=0, reg\_lambda=1, ...)

```
In [33]: predictions = xgboost_regressor.predict(xtest)
```

```
In [34]: kfold = model_selection.KFold(n_splits=10)

results = model_selection.cross_val_score(xgboost_regressor, X_train, y_train, cv=kfold, scoring='neg_root_mean_squared_error')

print(f'RMSE of model 1: {abs(results.mean())}')

RMSE of model 1: 8.715858578949872
```

Prediction of the test dataset (final\_data\_test)

```
In [35]: test_df1 = pd.DataFrame(xgboost_regressor.predict(final_data_test), columns=['y'])
test_df1.head()
```

```
Out[35]:
```

	y
0	88.388672
1	103.497986
2	88.942810
3	75.290504
4	108.704353

b. The train and test datasets not combined

```
In [36]: data_train.drop(['ID'], axis=1, inplace=True)
data_train.head()
```



```
Out[36]:
```

	y	X0	X1	X2	X3	X4	X5	X6	X8	X10	...	X376	X377	X378	X379	X380	X382	X383	X384	X385	Source	
0	130.81	k	v	at	a	d	u	j	o	0	...	0	1	0	0	0	0	0	0	0	0	train
1	88.53	k	t	av	e	d	y	l	o	0	...	0	0	0	0	0	0	0	0	0	0	train
2	76.26	az	w	n	c	d	x	j	x	0	...	0	0	0	0	0	1	0	0	0	0	train
3	80.62	az	t	n	f	d	x	l	e	0	...	0	0	0	0	0	0	0	0	0	0	train
4	78.02	az	v	n	f	d	h	d	n	0	...	0	0	0	0	0	0	0	0	0	0	train

5 rows × 378 columns

```
In [37]: data_test.drop(['ID'], axis=1, inplace=True)
data_test.head()
```

```
Out[37]:
```

	X0	X1	X2	X3	X4	X5	X6	X8	X10	X11	...	X376	X377	X378	X379	X380	X382	X383	X384	X385	Source	
0	az	v	n	f	d	t	a	w	0	0	...	0	0	1	0	0	0	0	0	0	0	test
1	t	b	ai	a	d	b	g	y	0	0	...	0	1	0	0	0	0	0	0	0	0	test
2	az	v	as	f	d	a	j	j	0	0	...	0	0	1	0	0	0	0	0	0	0	test
3	az	l	n	f	d	z	l	n	0	0	...	0	0	1	0	0	0	0	0	0	0	test
4	w	s	as	c	d	y	i	m	0	0	...	0	0	0	0	0	0	0	0	0	0	test

5 rows × 377 columns

Dropping the columns with null variance

```
In [38]: numeric_data_train = data_train._get_numeric_data()
list_null_var_train = []
for column in numeric_data_train.columns:
    if numeric_data_train[column].var() == 0:
        list_null_var_train.append(column)

print(list_null_var_train)
print(len(list_null_var_train))

['X11', 'X93', 'X107', 'X233', 'X235', 'X268', 'X289', 'X290', 'X293', 'X297', 'X330', 'X347']
12
```

- 12 columns have a null variance so they will be removed from both train and test datasets

```
In [39]: data_train.drop(list_null_var_train, axis=1, inplace=True)
data_train.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4209 entries, 0 to 4208
Columns: 366 entries, y to Source
dtypes: float64(1), int64(356), object(9)
memory usage: 11.8+ MB
```

```
In [40]: data_test.drop(list_null_var_train, axis=1, inplace=True)
data_test.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4209 entries, 0 to 4208
Columns: 365 entries, X0 to Source
dtypes: int64(356), object(9)
memory usage: 11.7+ MB
```

Checking for null values

```
In [41]: n = 0
for column in data_train.columns:
    if data_train[column].isnull().sum() != 0:
        print(f'{column} has {data_train[column].isnull().sum()} null values')
        n += 1
if n == 0:
    print('There are no null value.')

There are no null value.
```

```
In [42]: n = 0
for column in data_test.columns:
    if data_test[column].isnull().sum() != 0:
        print(f'{column} has {data_test[column].isnull().sum()} null values')
        n += 1
if n == 0:
    print('There are no null value.')

There are no null value.
```

- There are no null value in both datasets.

Checking for unique values

```
In [43]: for column in data_train.columns:
```

```
print(f"Unique values of column {column} :\n{data_train[column].value_counts()}\n\n")
```

```

Unique values of column y :
91.88      7
89.38      7
89.06      7
90.76      7
89.19      6
..
110.90     1
132.60     1
94.08      1
75.89      1
87.48      1
Name: y, Length: 2545, dtype: int64

```

```

Unique values of column X0 :
z      360
ak     349
y      324
ay     313
t      306
x      300
o     269
f     227
n     195
w     182
j     181
az     175
aj     151
s     106
ap     103
h      75
d      73
al      67
v      36
af      35
m       34
ai      34
e       32
ba      27
at      25
a       21
ax      19
aq      18
am      18
i       18
u       17
aw      16
l       16
ad      14
au      11
k       11
b       11
r       10
as      10
bc       6
ao       4
c        3
aa       2
q        2
ac       1
g        1
ab       1
Name: X0, dtype: int64

```

```

Unique values of column X1 :
aa     833
s     598
b     592
l     590
v     408
r     251
i     203
a     143
c     121
o      82
w      52
z      46
u      37
e      33
m      32
t      31
h      29
y      23
f      23
j      22
n      19
k      17
p       9
g       6
d       3
q       3
ab      3

```

Name: X1, dtype: int64

Unique values of column X2 :

as	1659
ae	496
ai	415
m	367
ak	265
r	153
n	137
s	94
f	87
e	81
aq	63
ay	54
a	47
t	29
k	25
i	25
b	21
ao	20
ag	19
z	19
d	18
ac	13
g	12
ap	11
y	11
x	10
aw	8
at	6
h	6
al	5
an	5
q	5
av	4
ah	4
p	4
au	3
am	1
j	1
af	1
l	1
aa	1
c	1
o	1
ar	1

Name: X2, dtype: int64

Unique values of column X3 :

c	1942
f	1076
a	440
d	290
g	241
e	163
b	57

Name: X3, dtype: int64

Unique values of column X4 :

d	4205
a	2
b	1
c	1

Name: X4, dtype: int64

Unique values of column X5 :

w	231
v	231
q	220
r	215
s	214
d	214
n	212
p	208
m	208
i	207
ae	205
ag	204
ac	200
ab	197
l	195
af	188
ad	185
k	177
c	131
j	125
aa	112
ah	97
o	20

```
f      7
x      2
y      1
g      1
h      1
u      1
Name: X5, dtype: int64
```

Unique values of column X6 :

```
g    1042
j    1039
d     625
i     488
l     478
a     206
h     190
k      43
c      38
b      28
f      20
e      12
```

Name: X6, dtype: int64

Unique values of column X8 :

```
j    277
s    255
f    243
n    242
i    237
e    225
r    219
a    210
w    196
v    194
b    190
k    176
o    163
m    155
g    130
u    119
t    119
q    117
h    117
y    116
x    105
d    103
l    101
p    100
c    100
```

Name: X8, dtype: int64

Unique values of column X10 :

```
0    4153
1      56
```

Name: X10, dtype: int64

Unique values of column X12 :

```
0    3893
1     316
```

Name: X12, dtype: int64

Unique values of column X13 :

```
0    3965
1     244
```

Name: X13, dtype: int64

Unique values of column X14 :

```
0    2407
1    1802
```

Name: X14, dtype: int64

Unique values of column X15 :

```
0    4207
1         2
```

Name: X15, dtype: int64

Unique values of column X16 :

```
0    4198
1        11
```

Name: X16, dtype: int64

Unique values of column X17 :

```
0    4177
1        32
```

Name: X17, dtype: int64

Unique values of column X18 :  
0 4176  
1 33  
Name: X18, dtype: int64

Unique values of column X19 :  
0 3790  
1 419  
Name: X19, dtype: int64

Unique values of column X20 :  
0 3608  
1 601  
Name: X20, dtype: int64

Unique values of column X21 :  
0 4198  
1 11  
Name: X21, dtype: int64

Unique values of column X22 :  
0 3843  
1 366  
Name: X22, dtype: int64

Unique values of column X23 :  
0 4122  
1 87  
Name: X23, dtype: int64

Unique values of column X24 :  
0 4201  
1 8  
Name: X24, dtype: int64

Unique values of column X26 :  
0 4188  
1 21  
Name: X26, dtype: int64

Unique values of column X27 :  
1 2873  
0 1336  
Name: X27, dtype: int64

Unique values of column X28 :  
0 4072  
1 137  
Name: X28, dtype: int64

Unique values of column X29 :  
0 4028  
1 181  
Name: X29, dtype: int64

Unique values of column X30 :  
0 4190  
1 19  
Name: X30, dtype: int64

Unique values of column X31 :  
0 3231  
1 978  
Name: X31, dtype: int64

Unique values of column X32 :  
0 4162  
1 47  
Name: X32, dtype: int64

Unique values of column X33 :  
0 4208  
1 1  
Name: X33, dtype: int64

Unique values of column X34 :  
0 4186

```
1      23
Name: X34, dtype: int64
```

```
Unique values of column X35 :
0      3231
1       978
Name: X35, dtype: int64
```

```
Unique values of column X36 :
0      4190
1        19
Name: X36, dtype: int64
```

```
Unique values of column X37 :
0      3231
1       978
Name: X37, dtype: int64
```

```
Unique values of column X38 :
0      4069
1       140
Name: X38, dtype: int64
```

```
Unique values of column X39 :
0      4208
1         1
Name: X39, dtype: int64
```

```
Unique values of column X40 :
0      4206
1         3
Name: X40, dtype: int64
```

```
Unique values of column X41 :
0      4161
1        48
Name: X41, dtype: int64
```

```
Unique values of column X42 :
0      4208
1         1
Name: X42, dtype: int64
```

```
Unique values of column X43 :
0      3905
1       304
Name: X43, dtype: int64
```

```
Unique values of column X44 :
0      4161
1        48
Name: X44, dtype: int64
```

```
Unique values of column X45 :
0      3143
1      1066
Name: X45, dtype: int64
```

```
Unique values of column X46 :
1      2514
0      1695
Name: X46, dtype: int64
```

```
Unique values of column X47 :
0      4155
1        54
Name: X47, dtype: int64
```

```
Unique values of column X48 :
0      4115
1        94
Name: X48, dtype: int64
```

```
Unique values of column X49 :
0      3695
1        514
Name: X49, dtype: int64
```

Unique values of column X50 :  
0 3308  
1 901  
Name: X50, dtype: int64

Unique values of column X51 :  
1 3038  
0 1171  
Name: X51, dtype: int64

Unique values of column X52 :  
0 4031  
1 178  
Name: X52, dtype: int64

Unique values of column X53 :  
0 4180  
1 29  
Name: X53, dtype: int64

Unique values of column X54 :  
0 4026  
1 183  
Name: X54, dtype: int64

Unique values of column X55 :  
0 4187  
1 22  
Name: X55, dtype: int64

Unique values of column X56 :  
0 4120  
1 89  
Name: X56, dtype: int64

Unique values of column X57 :  
0 4153  
1 56  
Name: X57, dtype: int64

Unique values of column X58 :  
1 2420  
0 1789  
Name: X58, dtype: int64

Unique values of column X59 :  
0 4206  
1 3  
Name: X59, dtype: int64

Unique values of column X60 :  
0 4203  
1 6  
Name: X60, dtype: int64

Unique values of column X61 :  
1 4015  
0 194  
Name: X61, dtype: int64

Unique values of column X62 :  
0 4184  
1 25  
Name: X62, dtype: int64

Unique values of column X63 :  
0 4161  
1 48  
Name: X63, dtype: int64

Unique values of column X64 :  
0 2630  
1 1579  
Name: X64, dtype: int64

Unique values of column X65 :  
0 4200  
1 9  
Name: X65, dtype: int64



Unique values of column X66 :  
0 4095  
1 114  
Name: X66, dtype: int64

Unique values of column X67 :  
0 4201  
1 8  
Name: X67, dtype: int64

Unique values of column X68 :  
0 3900  
1 309  
Name: X68, dtype: int64

Unique values of column X69 :  
0 4083  
1 126  
Name: X69, dtype: int64

Unique values of column X70 :  
1 3872  
0 337  
Name: X70, dtype: int64

Unique values of column X71 :  
0 3773  
1 436  
Name: X71, dtype: int64

Unique values of column X73 :  
0 4125  
1 84  
Name: X73, dtype: int64

Unique values of column X74 :  
1 4206  
0 3  
Name: X74, dtype: int64

Unique values of column X75 :  
0 4057  
1 152  
Name: X75, dtype: int64

Unique values of column X76 :  
0 4026  
1 183  
Name: X76, dtype: int64

Unique values of column X77 :  
0 4156  
1 53  
Name: X77, dtype: int64

Unique values of column X78 :  
0 4185  
1 24  
Name: X78, dtype: int64

Unique values of column X79 :  
0 4103  
1 106  
Name: X79, dtype: int64

Unique values of column X80 :  
1 3986  
0 223  
Name: X80, dtype: int64

Unique values of column X81 :  
0 3245  
1 964  
Name: X81, dtype: int64

Unique values of column X82 :  
0 4137

```
1      72
Name: X82, dtype: int64
```

```
Unique values of column X83 :
0      4204
1         5
Name: X83, dtype: int64
```

```
Unique values of column X84 :
0      3773
1       436
Name: X84, dtype: int64
```

```
Unique values of column X85 :
0      2491
1      1718
Name: X85, dtype: int64
```

```
Unique values of column X86 :
0      4203
1         6
Name: X86, dtype: int64
```

```
Unique values of column X87 :
0      4205
1         4
Name: X87, dtype: int64
```

```
Unique values of column X88 :
0      4179
1        30
Name: X88, dtype: int64
```

```
Unique values of column X89 :
0      4206
1         3
Name: X89, dtype: int64
```

```
Unique values of column X90 :
0      4178
1        31
Name: X90, dtype: int64
```

```
Unique values of column X91 :
0      4202
1         7
Name: X91, dtype: int64
```

```
Unique values of column X92 :
0      4205
1         4
Name: X92, dtype: int64
```

```
Unique values of column X94 :
0      4178
1        31
Name: X94, dtype: int64
```

```
Unique values of column X95 :
0      4208
1         1
Name: X95, dtype: int64
```

```
Unique values of column X96 :
1      3191
0      1018
Name: X96, dtype: int64
```

```
Unique values of column X97 :
0      4191
1        18
Name: X97, dtype: int64
```

```
Unique values of column X98 :
1      3967
0        242
Name: X98, dtype: int64
```

Unique values of column X99 :  
0 4173  
1 36  
Name: X99, dtype: int64

Unique values of column X100 :  
1 2905  
0 1304  
Name: X100, dtype: int64

Unique values of column X101 :  
1 3938  
0 271  
Name: X101, dtype: int64

Unique values of column X102 :  
0 4180  
1 29  
Name: X102, dtype: int64

Unique values of column X103 :  
1 3302  
0 907  
Name: X103, dtype: int64

Unique values of column X104 :  
0 4201  
1 8  
Name: X104, dtype: int64

Unique values of column X105 :  
0 4199  
1 10  
Name: X105, dtype: int64

Unique values of column X106 :  
0 4154  
1 55  
Name: X106, dtype: int64

Unique values of column X108 :  
0 4147  
1 62  
Name: X108, dtype: int64

Unique values of column X109 :  
0 4039  
1 170  
Name: X109, dtype: int64

Unique values of column X110 :  
0 4205  
1 4  
Name: X110, dtype: int64

Unique values of column X111 :  
1 4103  
0 106  
Name: X111, dtype: int64

Unique values of column X112 :  
0 4197  
1 12  
Name: X112, dtype: int64

Unique values of column X113 :  
0 4115  
1 94  
Name: X113, dtype: int64

Unique values of column X114 :  
0 3594  
1 615  
Name: X114, dtype: int64

Unique values of column X115 :  
0 3007  
1 1202  
Name: X115, dtype: int64

Unique values of column X116 :  
0 3381  
1 828  
Name: X116, dtype: int64

Unique values of column X117 :  
0 4002  
1 207  
Name: X117, dtype: int64

Unique values of column X118 :  
1 2619  
0 1590  
Name: X118, dtype: int64

Unique values of column X119 :  
1 2619  
0 1590  
Name: X119, dtype: int64

Unique values of column X120 :  
1 4031  
0 178  
Name: X120, dtype: int64

Unique values of column X122 :  
0 4179  
1 30  
Name: X122, dtype: int64

Unique values of column X123 :  
0 4198  
1 11  
Name: X123, dtype: int64

Unique values of column X124 :  
0 4207  
1 2  
Name: X124, dtype: int64

Unique values of column X125 :  
0 4196  
1 13  
Name: X125, dtype: int64

Unique values of column X126 :  
0 4045  
1 164  
Name: X126, dtype: int64

Unique values of column X127 :  
0 2125  
1 2084  
Name: X127, dtype: int64

Unique values of column X128 :  
1 4034  
0 175  
Name: X128, dtype: int64

Unique values of column X129 :  
0 3693  
1 516  
Name: X129, dtype: int64

Unique values of column X130 :  
0 4034  
1 175  
Name: X130, dtype: int64

Unique values of column X131 :  
0 4097  
1 112  
Name: X131, dtype: int64

Unique values of column X132 :  
1 2898

0 1311  
Name: X132, dtype: int64

Unique values of column X133 :  
0 3686  
1 523  
Name: X133, dtype: int64

Unique values of column X134 :  
0 4115  
1 94  
Name: X134, dtype: int64

Unique values of column X135 :  
0 4095  
1 114  
Name: X135, dtype: int64

Unique values of column X136 :  
1 4026  
0 183  
Name: X136, dtype: int64

Unique values of column X137 :  
1 2449  
0 1760  
Name: X137, dtype: int64

Unique values of column X138 :  
0 4037  
1 172  
Name: X138, dtype: int64

Unique values of column X139 :  
0 3828  
1 381  
Name: X139, dtype: int64

Unique values of column X140 :  
0 4039  
1 170  
Name: X140, dtype: int64

Unique values of column X141 :  
0 4149  
1 60  
Name: X141, dtype: int64

Unique values of column X142 :  
1 3242  
0 967  
Name: X142, dtype: int64

Unique values of column X143 :  
0 4048  
1 161  
Name: X143, dtype: int64

Unique values of column X144 :  
1 3401  
0 808  
Name: X144, dtype: int64

Unique values of column X145 :  
0 4203  
1 6  
Name: X145, dtype: int64

Unique values of column X146 :  
0 4037  
1 172  
Name: X146, dtype: int64

Unique values of column X147 :  
0 4115  
1 94  
Name: X147, dtype: int64

Unique values of column X148 :  
0 4020  
1 189  
Name: X148, dtype: int64

Unique values of column X150 :  
1 3335  
0 874  
Name: X150, dtype: int64

Unique values of column X151 :  
0 3849  
1 360  
Name: X151, dtype: int64

Unique values of column X152 :  
0 4073  
1 136  
Name: X152, dtype: int64

Unique values of column X153 :  
0 4206  
1 3  
Name: X153, dtype: int64

Unique values of column X154 :  
0 3330  
1 879  
Name: X154, dtype: int64

Unique values of column X155 :  
0 3887  
1 322  
Name: X155, dtype: int64

Unique values of column X156 :  
1 3019  
0 1190  
Name: X156, dtype: int64

Unique values of column X157 :  
0 3019  
1 1190  
Name: X157, dtype: int64

Unique values of column X158 :  
0 3242  
1 967  
Name: X158, dtype: int64

Unique values of column X159 :  
0 4152  
1 57  
Name: X159, dtype: int64

Unique values of column X160 :  
0 4204  
1 5  
Name: X160, dtype: int64

Unique values of column X161 :  
0 3377  
1 832  
Name: X161, dtype: int64

Unique values of column X162 :  
0 4037  
1 172  
Name: X162, dtype: int64

Unique values of column X163 :  
0 2932  
1 1277  
Name: X163, dtype: int64

Unique values of column X164 :  
0 3946  
1 263  
Name: X164, dtype: int64

Unique values of column X165 :  
0 4190  
1 19  
Name: X165, dtype: int64

Unique values of column X166 :  
0 4069  
1 140  
Name: X166, dtype: int64

Unique values of column X167 :  
0 4205  
1 4  
Name: X167, dtype: int64

Unique values of column X168 :  
0 3069  
1 1140  
Name: X168, dtype: int64

Unique values of column X169 :  
0 4181  
1 28  
Name: X169, dtype: int64

Unique values of column X170 :  
0 4107  
1 102  
Name: X170, dtype: int64

Unique values of column X171 :  
1 2767  
0 1442  
Name: X171, dtype: int64

Unique values of column X172 :  
0 4184  
1 25  
Name: X172, dtype: int64

Unique values of column X173 :  
0 4168  
1 41  
Name: X173, dtype: int64

Unique values of column X174 :  
0 4136  
1 73  
Name: X174, dtype: int64

Unique values of column X175 :  
0 4115  
1 94  
Name: X175, dtype: int64

Unique values of column X176 :  
0 4137  
1 72  
Name: X176, dtype: int64

Unique values of column X177 :  
0 3998  
1 211  
Name: X177, dtype: int64

Unique values of column X178 :  
1 2346  
0 1863  
Name: X178, dtype: int64

Unique values of column X179 :  
0 4007  
1 202  
Name: X179, dtype: int64

Unique values of column X180 :  
0 3544

```
1      665
Name: X180, dtype: int64
```

```
Unique values of column X181 :
0      3814
1       395
Name: X181, dtype: int64
```

```
Unique values of column X182 :
0      3762
1       447
Name: X182, dtype: int64
```

```
Unique values of column X183 :
0      4192
1        17
Name: X183, dtype: int64
```

```
Unique values of column X184 :
0      4203
1         6
Name: X184, dtype: int64
```

```
Unique values of column X185 :
0      4130
1         79
Name: X185, dtype: int64
```

```
Unique values of column X186 :
1      2256
0      1953
Name: X186, dtype: int64
```

```
Unique values of column X187 :
0      2439
1      1770
Name: X187, dtype: int64
```

```
Unique values of column X189 :
1      3853
0       356
Name: X189, dtype: int64
```

```
Unique values of column X190 :
0      4208
1         1
Name: X190, dtype: int64
```

```
Unique values of column X191 :
0      2227
1      1982
Name: X191, dtype: int64
```

```
Unique values of column X192 :
0      4199
1        10
Name: X192, dtype: int64
```

```
Unique values of column X194 :
0      2256
1      1953
Name: X194, dtype: int64
```

```
Unique values of column X195 :
0      4160
1         9
Name: X195, dtype: int64
```

```
Unique values of column X196 :
0      4166
1         3
Name: X196, dtype: int64
```

```
Unique values of column X197 :
0      4073
1        136
Name: X197, dtype: int64
```



Unique values of column X198 :  
0 4112  
1 97  
Name: X198, dtype: int64

Unique values of column X199 :  
0 4197  
1 12  
Name: X199, dtype: int64

Unique values of column X200 :  
0 4181  
1 28  
Name: X200, dtype: int64

Unique values of column X201 :  
0 3461  
1 748  
Name: X201, dtype: int64

Unique values of column X202 :  
0 3193  
1 1016  
Name: X202, dtype: int64

Unique values of column X203 :  
0 4138  
1 71  
Name: X203, dtype: int64

Unique values of column X204 :  
0 4208  
1 1  
Name: X204, dtype: int64

Unique values of column X205 :  
1 4208  
0 1  
Name: X205, dtype: int64

Unique values of column X206 :  
0 4128  
1 81  
Name: X206, dtype: int64

Unique values of column X207 :  
0 4208  
1 1  
Name: X207, dtype: int64

Unique values of column X208 :  
0 3944  
1 265  
Name: X208, dtype: int64

Unique values of column X209 :  
1 3782  
0 427  
Name: X209, dtype: int64

Unique values of column X210 :  
0 4208  
1 1  
Name: X210, dtype: int64

Unique values of column X211 :  
0 4146  
1 63  
Name: X211, dtype: int64

Unique values of column X212 :  
0 4186  
1 23  
Name: X212, dtype: int64

Unique values of column X213 :  
0 4201  
1 8  
Name: X213, dtype: int64

Unique values of column X214 :  
0 4180  
1 29  
Name: X214, dtype: int64

Unique values of column X215 :  
0 3794  
1 415  
Name: X215, dtype: int64

Unique values of column X216 :  
0 4184  
1 25  
Name: X216, dtype: int64

Unique values of column X217 :  
0 4178  
1 31  
Name: X217, dtype: int64

Unique values of column X218 :  
0 2894  
1 1315  
Name: X218, dtype: int64

Unique values of column X219 :  
0 3925  
1 284  
Name: X219, dtype: int64

Unique values of column X220 :  
1 2362  
0 1847  
Name: X220, dtype: int64

Unique values of column X221 :  
0 4175  
1 34  
Name: X221, dtype: int64

Unique values of column X222 :  
0 4115  
1 94  
Name: X222, dtype: int64

Unique values of column X223 :  
1 2337  
0 1872  
Name: X223, dtype: int64

Unique values of column X224 :  
0 2873  
1 1336  
Name: X224, dtype: int64

Unique values of column X225 :  
0 3801  
1 408  
Name: X225, dtype: int64

Unique values of column X226 :  
0 4073  
1 136  
Name: X226, dtype: int64

Unique values of column X227 :  
0 4196  
1 13  
Name: X227, dtype: int64

Unique values of column X228 :  
0 4045  
1 164  
Name: X228, dtype: int64

Unique values of column X229 :  
1 4041

```
0      168
Name: X229, dtype: int64
```

```
Unique values of column X230 :
0      4187
1         22
Name: X230, dtype: int64
```

```
Unique values of column X231 :
0      4141
1         68
Name: X231, dtype: int64
```

```
Unique values of column X232 :
0      4028
1        181
Name: X232, dtype: int64
```

```
Unique values of column X234 :
0      3360
1        849
Name: X234, dtype: int64
```

```
Unique values of column X236 :
0      4207
1          2
Name: X236, dtype: int64
```

```
Unique values of column X237 :
0      4181
1         28
Name: X237, dtype: int64
```

```
Unique values of column X238 :
1      3856
0       353
Name: X238, dtype: int64
```

```
Unique values of column X239 :
0      4180
1        29
Name: X239, dtype: int64
```

```
Unique values of column X240 :
0      4197
1        12
Name: X240, dtype: int64
```

```
Unique values of column X241 :
0      3800
1        409
Name: X241, dtype: int64
```

```
Unique values of column X242 :
0      4178
1        31
Name: X242, dtype: int64
```

```
Unique values of column X243 :
0      4179
1        30
Name: X243, dtype: int64
```

```
Unique values of column X244 :
0      3773
1        436
Name: X244, dtype: int64
```

```
Unique values of column X245 :
0      4206
1          3
Name: X245, dtype: int64
```

```
Unique values of column X246 :
0      2486
1       1723
Name: X246, dtype: int64
```

Unique values of column X247 :  
0 3193  
1 1016  
Name: X247, dtype: int64

Unique values of column X248 :  
0 4203  
1 6  
Name: X248, dtype: int64

Unique values of column X249 :  
0 4177  
1 32  
Name: X249, dtype: int64

Unique values of column X250 :  
1 2327  
0 1882  
Name: X250, dtype: int64

Unique values of column X251 :  
0 2550  
1 1659  
Name: X251, dtype: int64

Unique values of column X252 :  
0 4206  
1 3  
Name: X252, dtype: int64

Unique values of column X253 :  
0 4203  
1 6  
Name: X253, dtype: int64

Unique values of column X254 :  
0 4187  
1 22  
Name: X254, dtype: int64

Unique values of column X255 :  
0 4127  
1 82  
Name: X255, dtype: int64

Unique values of column X256 :  
0 3901  
1 308  
Name: X256, dtype: int64

Unique values of column X257 :  
0 4208  
1 1  
Name: X257, dtype: int64

Unique values of column X258 :  
0 4199  
1 10  
Name: X258, dtype: int64

Unique values of column X259 :  
0 4208  
1 1  
Name: X259, dtype: int64

Unique values of column X260 :  
0 4208  
1 1  
Name: X260, dtype: int64

Unique values of column X261 :  
0 2443  
1 1766  
Name: X261, dtype: int64

Unique values of column X262 :  
0 4203  
1 6  
Name: X262, dtype: int64

Unique values of column X263 :  
1 4028  
0 181  
Name: X263, dtype: int64

Unique values of column X264 :  
0 4043  
1 166  
Name: X264, dtype: int64

Unique values of column X265 :  
1 3811  
0 398  
Name: X265, dtype: int64

Unique values of column X266 :  
0 4203  
1 6  
Name: X266, dtype: int64

Unique values of column X267 :  
0 4171  
1 38  
Name: X267, dtype: int64

Unique values of column X269 :  
0 4207  
1 2  
Name: X269, dtype: int64

Unique values of column X270 :  
0 4208  
1 1  
Name: X270, dtype: int64

Unique values of column X271 :  
0 4200  
1 9  
Name: X271, dtype: int64

Unique values of column X272 :  
0 4051  
1 158  
Name: X272, dtype: int64

Unique values of column X273 :  
1 3031  
0 1178  
Name: X273, dtype: int64

Unique values of column X274 :  
0 4167  
1 42  
Name: X274, dtype: int64

Unique values of column X275 :  
1 3059  
0 1150  
Name: X275, dtype: int64

Unique values of column X276 :  
0 4047  
1 162  
Name: X276, dtype: int64

Unique values of column X277 :  
0 4203  
1 6  
Name: X277, dtype: int64

Unique values of column X278 :  
0 4207  
1 2  
Name: X278, dtype: int64

Unique values of column X279 :  
0 4028

```
1      181
Name: X279, dtype: int64
```

```
Unique values of column X280 :
0      4208
1         1
Name: X280, dtype: int64
```

```
Unique values of column X281 :
0      4198
1         11
Name: X281, dtype: int64
```

```
Unique values of column X282 :
0      4192
1         17
Name: X282, dtype: int64
```

```
Unique values of column X283 :
0      3617
1         592
Name: X283, dtype: int64
```

```
Unique values of column X284 :
0      4036
1         173
Name: X284, dtype: int64
```

```
Unique values of column X285 :
0      3343
1         866
Name: X285, dtype: int64
```

```
Unique values of column X286 :
0      3979
1         230
Name: X286, dtype: int64
```

```
Unique values of column X287 :
0      4142
1         67
Name: X287, dtype: int64
```

```
Unique values of column X288 :
0      4208
1         1
Name: X288, dtype: int64
```

```
Unique values of column X291 :
0      4165
1         44
Name: X291, dtype: int64
```

```
Unique values of column X292 :
0      4171
1         38
Name: X292, dtype: int64
```

```
Unique values of column X294 :
0      3683
1         526
Name: X294, dtype: int64
```

```
Unique values of column X295 :
0      4208
1         1
Name: X295, dtype: int64
```

```
Unique values of column X296 :
0      4208
1         1
Name: X296, dtype: int64
```

```
Unique values of column X298 :
0      4190
1         19
Name: X298, dtype: int64
```

Unique values of column X299 :  
0 4190  
1 19  
Name: X299, dtype: int64

Unique values of column X300 :  
0 3338  
1 871  
Name: X300, dtype: int64

Unique values of column X301 :  
0 4012  
1 197  
Name: X301, dtype: int64

Unique values of column X302 :  
0 4161  
1 48  
Name: X302, dtype: int64

Unique values of column X304 :  
1 3890  
0 319  
Name: X304, dtype: int64

Unique values of column X305 :  
0 4153  
1 56  
Name: X305, dtype: int64

Unique values of column X306 :  
0 4025  
1 184  
Name: X306, dtype: int64

Unique values of column X307 :  
0 4200  
1 9  
Name: X307, dtype: int64

Unique values of column X308 :  
0 4169  
1 40  
Name: X308, dtype: int64

Unique values of column X309 :  
0 4179  
1 30  
Name: X309, dtype: int64

Unique values of column X310 :  
0 4198  
1 11  
Name: X310, dtype: int64

Unique values of column X311 :  
1 2519  
0 1690  
Name: X311, dtype: int64

Unique values of column X312 :  
0 4191  
1 18  
Name: X312, dtype: int64

Unique values of column X313 :  
0 2942  
1 1267  
Name: X313, dtype: int64

Unique values of column X314 :  
0 2392  
1 1817  
Name: X314, dtype: int64

Unique values of column X315 :  
0 4088  
1 121  
Name: X315, dtype: int64

Unique values of column X316 :  
0 3386  
1 823  
Name: X316, dtype: int64

Unique values of column X317 :  
0 4177  
1 32  
Name: X317, dtype: int64

Unique values of column X318 :  
0 4206  
1 3  
Name: X318, dtype: int64

Unique values of column X319 :  
0 4207  
1 2  
Name: X319, dtype: int64

Unique values of column X320 :  
0 4179  
1 30  
Name: X320, dtype: int64

Unique values of column X321 :  
0 3204  
1 1005  
Name: X321, dtype: int64

Unique values of column X322 :  
0 4117  
1 92  
Name: X322, dtype: int64

Unique values of column X323 :  
0 4170  
1 39  
Name: X323, dtype: int64

Unique values of column X324 :  
1 2420  
0 1789  
Name: X324, dtype: int64

Unique values of column X325 :  
0 4185  
1 24  
Name: X325, dtype: int64

Unique values of column X326 :  
0 4073  
1 136  
Name: X326, dtype: int64

Unique values of column X327 :  
0 3669  
1 540  
Name: X327, dtype: int64

Unique values of column X328 :  
0 4040  
1 169  
Name: X328, dtype: int64

Unique values of column X329 :  
0 2377  
1 1832  
Name: X329, dtype: int64

Unique values of column X331 :  
0 3973  
1 236  
Name: X331, dtype: int64

Unique values of column X332 :  
0 4206



1 3  
Name: X332, dtype: int64

Unique values of column X333 :  
0 4108  
1 101  
Name: X333, dtype: int64

Unique values of column X334 :  
0 2262  
1 1947  
Name: X334, dtype: int64

Unique values of column X335 :  
0 4194  
1 15  
Name: X335, dtype: int64

Unique values of column X336 :  
0 3673  
1 536  
Name: X336, dtype: int64

Unique values of column X337 :  
1 2174  
0 2035  
Name: X337, dtype: int64

Unique values of column X338 :  
0 4180  
1 29  
Name: X338, dtype: int64

Unique values of column X339 :  
0 4208  
1 1  
Name: X339, dtype: int64

Unique values of column X340 :  
0 4115  
1 94  
Name: X340, dtype: int64

Unique values of column X341 :  
0 4175  
1 34  
Name: X341, dtype: int64

Unique values of column X342 :  
0 4115  
1 94  
Name: X342, dtype: int64

Unique values of column X343 :  
0 3879  
1 330  
Name: X343, dtype: int64

Unique values of column X344 :  
0 4173  
1 36  
Name: X344, dtype: int64

Unique values of column X345 :  
0 4115  
1 94  
Name: X345, dtype: int64

Unique values of column X346 :  
0 4009  
1 200  
Name: X346, dtype: int64

Unique values of column X348 :  
1 3987  
0 222  
Name: X348, dtype: int64

Unique values of column X349 :  
0 4020  
1 189  
Name: X349, dtype: int64

Unique values of column X350 :  
0 2783  
1 1426  
Name: X350, dtype: int64

Unique values of column X351 :  
0 2958  
1 1251  
Name: X351, dtype: int64

Unique values of column X352 :  
0 3981  
1 228  
Name: X352, dtype: int64

Unique values of column X353 :  
0 4200  
1 9  
Name: X353, dtype: int64

Unique values of column X354 :  
0 3355  
1 854  
Name: X354, dtype: int64

Unique values of column X355 :  
0 2608  
1 1601  
Name: X355, dtype: int64

Unique values of column X356 :  
0 3452  
1 757  
Name: X356, dtype: int64

Unique values of column X357 :  
0 4204  
1 5  
Name: X357, dtype: int64

Unique values of column X358 :  
0 2412  
1 1797  
Name: X358, dtype: int64

Unique values of column X359 :  
0 4075  
1 134  
Name: X359, dtype: int64

Unique values of column X360 :  
0 3887  
1 322  
Name: X360, dtype: int64

Unique values of column X361 :  
1 4066  
0 143  
Name: X361, dtype: int64

Unique values of column X362 :  
1 2190  
0 2019  
Name: X362, dtype: int64

Unique values of column X363 :  
1 3173  
0 1036  
Name: X363, dtype: int64

Unique values of column X364 :  
0 4197  
1 12  
Name: X364, dtype: int64

Unique values of column X365 :  
0 4197  
1 12  
Name: X365, dtype: int64

Unique values of column X366 :  
0 4204  
1 5  
Name: X366, dtype: int64

Unique values of column X367 :  
0 3991  
1 218  
Name: X367, dtype: int64

Unique values of column X368 :  
0 3945  
1 264  
Name: X368, dtype: int64

Unique values of column X369 :  
0 4207  
1 2  
Name: X369, dtype: int64

Unique values of column X370 :  
0 4181  
1 28  
Name: X370, dtype: int64

Unique values of column X371 :  
0 4149  
1 60  
Name: X371, dtype: int64

Unique values of column X372 :  
0 4207  
1 2  
Name: X372, dtype: int64

Unique values of column X373 :  
0 4128  
1 81  
Name: X373, dtype: int64

Unique values of column X374 :  
0 3252  
1 957  
Name: X374, dtype: int64

Unique values of column X375 :  
0 2867  
1 1342  
Name: X375, dtype: int64

Unique values of column X376 :  
0 3968  
1 241  
Name: X376, dtype: int64

Unique values of column X377 :  
0 2884  
1 1325  
Name: X377, dtype: int64

Unique values of column X378 :  
0 4122  
1 87  
Name: X378, dtype: int64

Unique values of column X379 :  
0 4169  
1 40  
Name: X379, dtype: int64

Unique values of column X380 :  
0 4175

```
1      34
Name: X380, dtype: int64
```

```
Unique values of column X382 :
0      4177
1        32
Name: X382, dtype: int64
```

```
Unique values of column X383 :
0      4202
1         7
Name: X383, dtype: int64
```

```
Unique values of column X384 :
0      4207
1         2
Name: X384, dtype: int64
```

```
Unique values of column X385 :
0      4203
1         6
Name: X385, dtype: int64
```

```
Unique values of column Source :
train      4209
Name: Source, dtype: int64
```

```
In [44]: for column in data_test.columns:
          print(f"Unique values of column {column} :\n{data_test[column].value_counts()}\n\n")
```

Unique values of column X0 :

ak	432
y	348
z	335
x	302
ay	299
t	293
o	246
f	213
w	198
j	171
n	167
aj	162
az	161
s	116
ap	108
al	88
h	64
d	61
e	48
v	40
ai	38
m	34
af	34
am	28
i	25
at	21
u	20
ba	19
a	18
b	13
k	12
ad	12
aq	11
aw	11
r	10
ax	8
l	6
bc	6
c	6
as	6
ao	5
au	5
g	3
p	1
ae	1
av	1
an	1
ag	1
bb	1

Name: X0, dtype: int64

Unique values of column X1 :

aa	826
s	602
l	599
b	596
v	436
r	252
i	189
a	153
c	142
o	81
w	50
u	40
z	31
e	29
h	27
m	27
j	22
y	21
t	18
n	16
k	12
f	12
p	10
g	9
ab	5
q	3
d	1

Name: X1, dtype: int64

Unique values of column X2 :

as	1658
ae	478
ai	462
m	348
ak	260
r	155
n	113
s	100
f	85

e	84
ay	78
aq	72
a	44
b	38
k	25
t	25
ag	23
ac	20
ao	19
i	15
z	12
ap	11
p	10
aw	9
h	6
d	6
g	5
q	5
au	5
al	4
ad	4
af	4
ab	4
ah	3
am	3
w	3
at	3
j	2
x	2
av	1
ax	1
y	1
aj	1
an	1
u	1

Name: X2, dtype: int64

Unique values of column X3 :

c	1900
f	1083
a	476
d	274
g	272
e	158
b	46

Name: X3, dtype: int64

Unique values of column X4 :

d	4203
b	4
a	1
c	1

Name: X4, dtype: int64

Unique values of column X5 :

v	246
r	239
p	227
w	218
af	217
ad	213
ac	212
n	209
l	206
s	205
ag	201
q	197
m	197
ae	196
k	193
d	192
i	180
ab	179
j	137
c	121
aa	105
ah	80
o	16
g	8
f	6
h	2
x	2
b	1
y	1
z	1
a	1
t	1

Name: X5, dtype: int64

Unique values of column X6 :  
g 1073  
j 1002  
d 589  
i 490  
l 473  
h 218  
a 196  
k 67  
c 40  
f 25  
b 19  
e 17  
Name: X6, dtype: int64

Unique values of column X8 :  
e 274  
j 256  
s 244  
f 241  
n 236  
i 234  
r 228  
a 202  
w 192  
v 174  
b 172  
o 169  
m 162  
k 154  
u 144  
g 137  
h 128  
t 117  
q 114  
x 110  
d 108  
p 106  
y 106  
c 101  
l 100  
Name: X8, dtype: int64

Unique values of column X10 :  
0 4129  
1 80  
Name: X10, dtype: int64

Unique values of column X12 :  
0 3896  
1 313  
Name: X12, dtype: int64

Unique values of column X13 :  
0 3952  
1 257  
Name: X13, dtype: int64

Unique values of column X14 :  
0 2408  
1 1801  
Name: X14, dtype: int64

Unique values of column X15 :  
0 4206  
1 3  
Name: X15, dtype: int64

Unique values of column X16 :  
0 4198  
1 11  
Name: X16, dtype: int64

Unique values of column X17 :  
0 4172  
1 37  
Name: X17, dtype: int64

Unique values of column X18 :  
0 4166  
1 43  
Name: X18, dtype: int64

Unique values of column X19 :

```
0    3739
1    470
Name: X19, dtype: int64
```

```
Unique values of column X20 :
0    3622
1    587
Name: X20, dtype: int64
```

```
Unique values of column X21 :
0    4202
1         7
Name: X21, dtype: int64
```

```
Unique values of column X22 :
0    3862
1    347
Name: X22, dtype: int64
```

```
Unique values of column X23 :
0    4124
1     85
Name: X23, dtype: int64
```

```
Unique values of column X24 :
0    4199
1     10
Name: X24, dtype: int64
```

```
Unique values of column X26 :
0    4171
1     38
Name: X26, dtype: int64
```

```
Unique values of column X27 :
1    2867
0    1342
Name: X27, dtype: int64
```

```
Unique values of column X28 :
0    4096
1     113
Name: X28, dtype: int64
```

```
Unique values of column X29 :
0    4045
1     164
Name: X29, dtype: int64
```

```
Unique values of column X30 :
0    4186
1     23
Name: X30, dtype: int64
```

```
Unique values of column X31 :
0    3210
1     999
Name: X31, dtype: int64
```

```
Unique values of column X32 :
0    4165
1     44
Name: X32, dtype: int64
```

```
Unique values of column X33 :
0    4206
1         3
Name: X33, dtype: int64
```

```
Unique values of column X34 :
0    4194
1     15
Name: X34, dtype: int64
```

```
Unique values of column X35 :
0    3210
1     999
Name: X35, dtype: int64
```



Unique values of column X36 :  
0 4197  
1 12  
Name: X36, dtype: int64

Unique values of column X37 :  
0 3210  
1 999  
Name: X37, dtype: int64

Unique values of column X38 :  
0 4078  
1 131  
Name: X38, dtype: int64

Unique values of column X39 :  
0 4206  
1 3  
Name: X39, dtype: int64

Unique values of column X40 :  
0 4201  
1 8  
Name: X40, dtype: int64

Unique values of column X41 :  
0 4161  
1 48  
Name: X41, dtype: int64

Unique values of column X42 :  
0 4203  
1 6  
Name: X42, dtype: int64

Unique values of column X43 :  
0 3885  
1 324  
Name: X43, dtype: int64

Unique values of column X44 :  
0 4180  
1 29  
Name: X44, dtype: int64

Unique values of column X45 :  
0 3148  
1 1061  
Name: X45, dtype: int64

Unique values of column X46 :  
1 2499  
0 1710  
Name: X46, dtype: int64

Unique values of column X47 :  
0 4162  
1 47  
Name: X47, dtype: int64

Unique values of column X48 :  
0 4109  
1 100  
Name: X48, dtype: int64

Unique values of column X49 :  
0 3647  
1 562  
Name: X49, dtype: int64

Unique values of column X50 :  
0 3266  
1 943  
Name: X50, dtype: int64

Unique values of column X51 :  
1 3008  
0 1201

Name: X51, dtype: int64

Unique values of column X52 :  
0 4030  
1 179  
Name: X52, dtype: int64

Unique values of column X53 :  
0 4184  
1 25  
Name: X53, dtype: int64

Unique values of column X54 :  
0 4041  
1 168  
Name: X54, dtype: int64

Unique values of column X55 :  
0 4182  
1 27  
Name: X55, dtype: int64

Unique values of column X56 :  
0 4104  
1 105  
Name: X56, dtype: int64

Unique values of column X57 :  
0 4143  
1 66  
Name: X57, dtype: int64

Unique values of column X58 :  
1 2460  
0 1749  
Name: X58, dtype: int64

Unique values of column X59 :  
0 4204  
1 5  
Name: X59, dtype: int64

Unique values of column X60 :  
0 4202  
1 7  
Name: X60, dtype: int64

Unique values of column X61 :  
1 4013  
0 196  
Name: X61, dtype: int64

Unique values of column X62 :  
0 4184  
1 25  
Name: X62, dtype: int64

Unique values of column X63 :  
0 4154  
1 55  
Name: X63, dtype: int64

Unique values of column X64 :  
0 2693  
1 1516  
Name: X64, dtype: int64

Unique values of column X65 :  
0 4201  
1 8  
Name: X65, dtype: int64

Unique values of column X66 :  
0 4095  
1 114  
Name: X66, dtype: int64

Unique values of column X67 :

```
0    4200
1         9
Name: X67, dtype: int64
```

```
Unique values of column X68 :
0    3873
1     336
Name: X68, dtype: int64
```

```
Unique values of column X69 :
0    4080
1     129
Name: X69, dtype: int64
```

```
Unique values of column X70 :
1    3874
0     335
Name: X70, dtype: int64
```

```
Unique values of column X71 :
0    3755
1     454
Name: X71, dtype: int64
```

```
Unique values of column X73 :
0    4129
1     80
Name: X73, dtype: int64
```

```
Unique values of column X74 :
1    4200
0         9
Name: X74, dtype: int64
```

```
Unique values of column X75 :
0    4046
1     163
Name: X75, dtype: int64
```

```
Unique values of column X76 :
0    4041
1     168
Name: X76, dtype: int64
```

```
Unique values of column X77 :
0    4162
1      47
Name: X77, dtype: int64
```

```
Unique values of column X78 :
0    4178
1      31
Name: X78, dtype: int64
```

```
Unique values of column X79 :
0    4069
1     140
Name: X79, dtype: int64
```

```
Unique values of column X80 :
1    3997
0     212
Name: X80, dtype: int64
```

```
Unique values of column X81 :
0    3285
1     924
Name: X81, dtype: int64
```

```
Unique values of column X82 :
0    4122
1      87
Name: X82, dtype: int64
```

```
Unique values of column X83 :
0    4204
1         5
Name: X83, dtype: int64
```

Unique values of column X84 :  
0 3755  
1 454  
Name: X84, dtype: int64

Unique values of column X85 :  
0 2439  
1 1770  
Name: X85, dtype: int64

Unique values of column X86 :  
0 4203  
1 6  
Name: X86, dtype: int64

Unique values of column X87 :  
0 4205  
1 4  
Name: X87, dtype: int64

Unique values of column X88 :  
0 4174  
1 35  
Name: X88, dtype: int64

Unique values of column X89 :  
0 4201  
1 8  
Name: X89, dtype: int64

Unique values of column X90 :  
0 4174  
1 35  
Name: X90, dtype: int64

Unique values of column X91 :  
0 4198  
1 11  
Name: X91, dtype: int64

Unique values of column X92 :  
0 4199  
1 10  
Name: X92, dtype: int64

Unique values of column X94 :  
0 4176  
1 33  
Name: X94, dtype: int64

Unique values of column X95 :  
0 4206  
1 3  
Name: X95, dtype: int64

Unique values of column X96 :  
1 3147  
0 1062  
Name: X96, dtype: int64

Unique values of column X97 :  
0 4203  
1 6  
Name: X97, dtype: int64

Unique values of column X98 :  
1 3968  
0 241  
Name: X98, dtype: int64

Unique values of column X99 :  
0 4167  
1 42  
Name: X99, dtype: int64

Unique values of column X100 :  
1 2876  
0 1333

Name: X100, dtype: int64

Unique values of column X101 :

1 3943  
0 266

Name: X101, dtype: int64

Unique values of column X102 :

0 4184  
1 25

Name: X102, dtype: int64

Unique values of column X103 :

1 3318  
0 891

Name: X103, dtype: int64

Unique values of column X104 :

0 4198  
1 11

Name: X104, dtype: int64

Unique values of column X105 :

0 4208  
1 1

Name: X105, dtype: int64

Unique values of column X106 :

0 4157  
1 52

Name: X106, dtype: int64

Unique values of column X108 :

0 4160  
1 49

Name: X108, dtype: int64

Unique values of column X109 :

0 4043  
1 166

Name: X109, dtype: int64

Unique values of column X110 :

0 4203  
1 6

Name: X110, dtype: int64

Unique values of column X111 :

1 4104  
0 105

Name: X111, dtype: int64

Unique values of column X112 :

0 4204  
1 5

Name: X112, dtype: int64

Unique values of column X113 :

0 4109  
1 100

Name: X113, dtype: int64

Unique values of column X114 :

0 3599  
1 610

Name: X114, dtype: int64

Unique values of column X115 :

0 2972  
1 1237

Name: X115, dtype: int64

Unique values of column X116 :

0 3383  
1 826

Name: X116, dtype: int64

Unique values of column X117 :

```
0    4022
1     187
Name: X117, dtype: int64
```

```
Unique values of column X118 :
1    2619
0    1590
Name: X118, dtype: int64
```

```
Unique values of column X119 :
1    2619
0    1590
Name: X119, dtype: int64
```

```
Unique values of column X120 :
1    4029
0     180
Name: X120, dtype: int64
```

```
Unique values of column X122 :
0    4174
1     35
Name: X122, dtype: int64
```

```
Unique values of column X123 :
0    4195
1     14
Name: X123, dtype: int64
```

```
Unique values of column X124 :
0    4208
1         1
Name: X124, dtype: int64
```

```
Unique values of column X125 :
0    4189
1     20
Name: X125, dtype: int64
```

```
Unique values of column X126 :
0    4041
1     168
Name: X126, dtype: int64
```

```
Unique values of column X127 :
0    2163
1    2046
Name: X127, dtype: int64
```

```
Unique values of column X128 :
1    4047
0     162
Name: X128, dtype: int64
```

```
Unique values of column X129 :
0    3646
1     563
Name: X129, dtype: int64
```

```
Unique values of column X130 :
0    4047
1     162
Name: X130, dtype: int64
```

```
Unique values of column X131 :
0    4094
1     115
Name: X131, dtype: int64
```

```
Unique values of column X132 :
1    2895
0    1314
Name: X132, dtype: int64
```

```
Unique values of column X133 :
0    3690
1     519
Name: X133, dtype: int64
```

Unique values of column X134 :  
0 4109  
1 100  
Name: X134, dtype: int64

Unique values of column X135 :  
0 4098  
1 111  
Name: X135, dtype: int64

Unique values of column X136 :  
1 4042  
0 167  
Name: X136, dtype: int64

Unique values of column X137 :  
1 2485  
0 1724  
Name: X137, dtype: int64

Unique values of column X138 :  
0 4043  
1 166  
Name: X138, dtype: int64

Unique values of column X139 :  
0 3806  
1 403  
Name: X139, dtype: int64

Unique values of column X140 :  
0 4043  
1 166  
Name: X140, dtype: int64

Unique values of column X141 :  
0 4140  
1 69  
Name: X141, dtype: int64

Unique values of column X142 :  
1 3223  
0 986  
Name: X142, dtype: int64

Unique values of column X143 :  
0 4042  
1 167  
Name: X143, dtype: int64

Unique values of column X144 :  
1 3373  
0 836  
Name: X144, dtype: int64

Unique values of column X145 :  
0 4206  
1 3  
Name: X145, dtype: int64

Unique values of column X146 :  
0 4043  
1 166  
Name: X146, dtype: int64

Unique values of column X147 :  
0 4109  
1 100  
Name: X147, dtype: int64

Unique values of column X148 :  
0 4039  
1 170  
Name: X148, dtype: int64

Unique values of column X150 :  
1 3317  
0 892

Name: X150, dtype: int64

Unique values of column X151 :

0 3841  
1 368

Name: X151, dtype: int64

Unique values of column X152 :

0 4077  
1 132

Name: X152, dtype: int64

Unique values of column X153 :

0 4204  
1 5

Name: X153, dtype: int64

Unique values of column X154 :

0 3311  
1 898

Name: X154, dtype: int64

Unique values of column X155 :

0 3874  
1 335

Name: X155, dtype: int64

Unique values of column X156 :

1 2988  
0 1221

Name: X156, dtype: int64

Unique values of column X157 :

0 2988  
1 1221

Name: X157, dtype: int64

Unique values of column X158 :

0 3222  
1 987

Name: X158, dtype: int64

Unique values of column X159 :

0 4147  
1 62

Name: X159, dtype: int64

Unique values of column X160 :

0 4205  
1 4

Name: X160, dtype: int64

Unique values of column X161 :

0 3388  
1 821

Name: X161, dtype: int64

Unique values of column X162 :

0 4059  
1 150

Name: X162, dtype: int64

Unique values of column X163 :

0 2886  
1 1323

Name: X163, dtype: int64

Unique values of column X164 :

0 3924  
1 285

Name: X164, dtype: int64

Unique values of column X165 :

0 4194  
1 15

Name: X165, dtype: int64

Unique values of column X166 :



```
0    4095
1     114
Name: X166, dtype: int64
```

```
Unique values of column X167 :
0    4204
1         5
Name: X167, dtype: int64
```

```
Unique values of column X168 :
0    3099
1    1110
Name: X168, dtype: int64
```

```
Unique values of column X169 :
0    4178
1     31
Name: X169, dtype: int64
```

```
Unique values of column X170 :
0    4096
1     113
Name: X170, dtype: int64
```

```
Unique values of column X171 :
1    2805
0    1404
Name: X171, dtype: int64
```

```
Unique values of column X172 :
0    4184
1     25
Name: X172, dtype: int64
```

```
Unique values of column X173 :
0    4166
1     43
Name: X173, dtype: int64
```

```
Unique values of column X174 :
0    4127
1     82
Name: X174, dtype: int64
```

```
Unique values of column X175 :
0    4113
1     96
Name: X175, dtype: int64
```

```
Unique values of column X176 :
0    4148
1     61
Name: X176, dtype: int64
```

```
Unique values of column X177 :
0    3991
1     218
Name: X177, dtype: int64
```

```
Unique values of column X178 :
1    2359
0    1850
Name: X178, dtype: int64
```

```
Unique values of column X179 :
0    4019
1     190
Name: X179, dtype: int64
```

```
Unique values of column X180 :
0    3542
1     667
Name: X180, dtype: int64
```

```
Unique values of column X181 :
0    3819
1     390
Name: X181, dtype: int64
```

Unique values of column X182 :  
0 3798  
1 411  
Name: X182, dtype: int64

Unique values of column X183 :  
0 4200  
1 9  
Name: X183, dtype: int64

Unique values of column X184 :  
0 4206  
1 3  
Name: X184, dtype: int64

Unique values of column X185 :  
0 4133  
1 76  
Name: X185, dtype: int64

Unique values of column X186 :  
1 2239  
0 1970  
Name: X186, dtype: int64

Unique values of column X187 :  
0 2406  
1 1803  
Name: X187, dtype: int64

Unique values of column X189 :  
1 3880  
0 329  
Name: X189, dtype: int64

Unique values of column X190 :  
0 4208  
1 1  
Name: X190, dtype: int64

Unique values of column X191 :  
0 2205  
1 2004  
Name: X191, dtype: int64

Unique values of column X192 :  
0 4202  
1 7  
Name: X192, dtype: int64

Unique values of column X194 :  
0 2238  
1 1971  
Name: X194, dtype: int64

Unique values of column X195 :  
0 4158  
1 51  
Name: X195, dtype: int64

Unique values of column X196 :  
0 4157  
1 52  
Name: X196, dtype: int64

Unique values of column X197 :  
0 4085  
1 124  
Name: X197, dtype: int64

Unique values of column X198 :  
0 4104  
1 105  
Name: X198, dtype: int64

Unique values of column X199 :  
0 4204  
1 5

Name: X199, dtype: int64

Unique values of column X200 :

0 4181  
1 28

Name: X200, dtype: int64

Unique values of column X201 :

0 3467  
1 742

Name: X201, dtype: int64

Unique values of column X202 :

0 3200  
1 1009

Name: X202, dtype: int64

Unique values of column X203 :

0 4159  
1 50

Name: X203, dtype: int64

Unique values of column X204 :

0 4206  
1 3

Name: X204, dtype: int64

Unique values of column X205 :

1 4206  
0 3

Name: X205, dtype: int64

Unique values of column X206 :

0 4130  
1 79

Name: X206, dtype: int64

Unique values of column X207 :

0 4203  
1 6

Name: X207, dtype: int64

Unique values of column X208 :

0 3949  
1 260

Name: X208, dtype: int64

Unique values of column X209 :

1 3776  
0 433

Name: X209, dtype: int64

Unique values of column X210 :

0 4208  
1 1

Name: X210, dtype: int64

Unique values of column X211 :

0 4137  
1 72

Name: X211, dtype: int64

Unique values of column X212 :

0 4179  
1 30

Name: X212, dtype: int64

Unique values of column X213 :

0 4200  
1 9

Name: X213, dtype: int64

Unique values of column X214 :

0 4184  
1 25

Name: X214, dtype: int64

Unique values of column X215 :

```
0    3747
1     462
Name: X215, dtype: int64
```

```
Unique values of column X216 :
0    4184
1      25
Name: X216, dtype: int64
```

```
Unique values of column X217 :
0    4171
1      38
Name: X217, dtype: int64
```

```
Unique values of column X218 :
0    2857
1    1352
Name: X218, dtype: int64
```

```
Unique values of column X219 :
0    3926
1     283
Name: X219, dtype: int64
```

```
Unique values of column X220 :
1    2346
0    1863
Name: X220, dtype: int64
```

```
Unique values of column X221 :
0    4174
1      35
Name: X221, dtype: int64
```

```
Unique values of column X222 :
0    4109
1     100
Name: X222, dtype: int64
```

```
Unique values of column X223 :
1    2341
0    1868
Name: X223, dtype: int64
```

```
Unique values of column X224 :
0    2890
1    1319
Name: X224, dtype: int64
```

```
Unique values of column X225 :
0    3846
1     363
Name: X225, dtype: int64
```

```
Unique values of column X226 :
0    4077
1     132
Name: X226, dtype: int64
```

```
Unique values of column X227 :
0    4189
1      20
Name: X227, dtype: int64
```

```
Unique values of column X228 :
0    4012
1     197
Name: X228, dtype: int64
```

```
Unique values of column X229 :
1    4002
0     207
Name: X229, dtype: int64
```

```
Unique values of column X230 :
0    4193
1      16
Name: X230, dtype: int64
```

Unique values of column X231 :  
0 4137  
1 72  
Name: X231, dtype: int64

Unique values of column X232 :  
0 4042  
1 167  
Name: X232, dtype: int64

Unique values of column X234 :  
0 3341  
1 868  
Name: X234, dtype: int64

Unique values of column X236 :  
0 4208  
1 1  
Name: X236, dtype: int64

Unique values of column X237 :  
0 4186  
1 23  
Name: X237, dtype: int64

Unique values of column X238 :  
1 3876  
0 333  
Name: X238, dtype: int64

Unique values of column X239 :  
0 4184  
1 25  
Name: X239, dtype: int64

Unique values of column X240 :  
0 4194  
1 15  
Name: X240, dtype: int64

Unique values of column X241 :  
0 3770  
1 439  
Name: X241, dtype: int64

Unique values of column X242 :  
0 4174  
1 35  
Name: X242, dtype: int64

Unique values of column X243 :  
0 4174  
1 35  
Name: X243, dtype: int64

Unique values of column X244 :  
0 3755  
1 454  
Name: X244, dtype: int64

Unique values of column X245 :  
0 4205  
1 4  
Name: X245, dtype: int64

Unique values of column X246 :  
0 2454  
1 1755  
Name: X246, dtype: int64

Unique values of column X247 :  
0 3201  
1 1008  
Name: X247, dtype: int64

Unique values of column X248 :  
0 4200  
1 9

Name: X248, dtype: int64

Unique values of column X249 :

0 4169  
1 40

Name: X249, dtype: int64

Unique values of column X250 :

1 2340  
0 1869

Name: X250, dtype: int64

Unique values of column X251 :

0 2551  
1 1658

Name: X251, dtype: int64

Unique values of column X252 :

0 4208  
1 1

Name: X252, dtype: int64

Unique values of column X253 :

0 4202  
1 7

Name: X253, dtype: int64

Unique values of column X254 :

0 4193  
1 16

Name: X254, dtype: int64

Unique values of column X255 :

0 4149  
1 60

Name: X255, dtype: int64

Unique values of column X256 :

0 3887  
1 322

Name: X256, dtype: int64

Unique values of column X257 :

0 4209

Name: X257, dtype: int64

Unique values of column X258 :

0 4209

Name: X258, dtype: int64

Unique values of column X259 :

0 4208  
1 1

Name: X259, dtype: int64

Unique values of column X260 :

0 4206  
1 3

Name: X260, dtype: int64

Unique values of column X261 :

0 2366  
1 1843

Name: X261, dtype: int64

Unique values of column X262 :

0 4206  
1 3

Name: X262, dtype: int64

Unique values of column X263 :

1 4042  
0 167

Name: X263, dtype: int64

Unique values of column X264 :

0 4037  
1 172

Name: X264, dtype: int64

Unique values of column X265 :

1 3839  
0 370

Name: X265, dtype: int64

Unique values of column X266 :

0 4205  
1 4

Name: X266, dtype: int64

Unique values of column X267 :

0 4175  
1 34

Name: X267, dtype: int64

Unique values of column X269 :

0 4206  
1 3

Name: X269, dtype: int64

Unique values of column X270 :

0 4206  
1 3

Name: X270, dtype: int64

Unique values of column X271 :

0 4203  
1 6

Name: X271, dtype: int64

Unique values of column X272 :

0 4061  
1 148

Name: X272, dtype: int64

Unique values of column X273 :

1 2998  
0 1211

Name: X273, dtype: int64

Unique values of column X274 :

0 4162  
1 47

Name: X274, dtype: int64

Unique values of column X275 :

1 3052  
0 1157

Name: X275, dtype: int64

Unique values of column X276 :

0 4058  
1 151

Name: X276, dtype: int64

Unique values of column X277 :

0 4203  
1 6

Name: X277, dtype: int64

Unique values of column X278 :

0 4208  
1 1

Name: X278, dtype: int64

Unique values of column X279 :

0 4042  
1 167

Name: X279, dtype: int64

Unique values of column X280 :

0 4208  
1 1

Name: X280, dtype: int64

Unique values of column X281 :

```
0    4194
1      15
Name: X281, dtype: int64
```

```
Unique values of column X282 :
0    4200
1       9
Name: X282, dtype: int64
```

```
Unique values of column X283 :
0    3601
1     608
Name: X283, dtype: int64
```

```
Unique values of column X284 :
0    4032
1     177
Name: X284, dtype: int64
```

```
Unique values of column X285 :
0    3323
1     886
Name: X285, dtype: int64
```

```
Unique values of column X286 :
0    4012
1     197
Name: X286, dtype: int64
```

```
Unique values of column X287 :
0    4144
1      65
Name: X287, dtype: int64
```

```
Unique values of column X288 :
0    4208
1       1
Name: X288, dtype: int64
```

```
Unique values of column X291 :
0    4164
1      45
Name: X291, dtype: int64
```

```
Unique values of column X292 :
0    4170
1      39
Name: X292, dtype: int64
```

```
Unique values of column X294 :
0    3695
1     514
Name: X294, dtype: int64
```

```
Unique values of column X295 :
0    4209
Name: X295, dtype: int64
```

```
Unique values of column X296 :
0    4209
Name: X296, dtype: int64
```

```
Unique values of column X298 :
0    4196
1      13
Name: X298, dtype: int64
```

```
Unique values of column X299 :
0    4196
1      13
Name: X299, dtype: int64
```

```
Unique values of column X300 :
0    3314
1     895
Name: X300, dtype: int64
```

```
Unique values of column X301 :
```



```
0    3977
1     232
Name: X301, dtype: int64
```

```
Unique values of column X302 :
0    4180
1      29
Name: X302, dtype: int64
```

```
Unique values of column X304 :
1    3912
0     297
Name: X304, dtype: int64
```

```
Unique values of column X305 :
0    4174
1      35
Name: X305, dtype: int64
```

```
Unique values of column X306 :
0    4019
1     190
Name: X306, dtype: int64
```

```
Unique values of column X307 :
0    4195
1     14
Name: X307, dtype: int64
```

```
Unique values of column X308 :
0    4179
1     30
Name: X308, dtype: int64
```

```
Unique values of column X309 :
0    4181
1     28
Name: X309, dtype: int64
```

```
Unique values of column X310 :
0    4191
1     18
Name: X310, dtype: int64
```

```
Unique values of column X311 :
1    2515
0    1694
Name: X311, dtype: int64
```

```
Unique values of column X312 :
0    4195
1     14
Name: X312, dtype: int64
```

```
Unique values of column X313 :
0    2979
1    1230
Name: X313, dtype: int64
```

```
Unique values of column X314 :
0    2322
1    1887
Name: X314, dtype: int64
```

```
Unique values of column X315 :
0    4102
1     107
Name: X315, dtype: int64
```

```
Unique values of column X316 :
0    3390
1     819
Name: X316, dtype: int64
```

```
Unique values of column X317 :
0    4185
1      24
Name: X317, dtype: int64
```

Unique values of column X318 :  
0 4205  
1 4  
Name: X318, dtype: int64

Unique values of column X319 :  
0 4207  
1 2  
Name: X319, dtype: int64

Unique values of column X320 :  
0 4176  
1 33  
Name: X320, dtype: int64

Unique values of column X321 :  
0 3187  
1 1022  
Name: X321, dtype: int64

Unique values of column X322 :  
0 4132  
1 77  
Name: X322, dtype: int64

Unique values of column X323 :  
0 4175  
1 34  
Name: X323, dtype: int64

Unique values of column X324 :  
1 2460  
0 1749  
Name: X324, dtype: int64

Unique values of column X325 :  
0 4186  
1 23  
Name: X325, dtype: int64

Unique values of column X326 :  
0 4077  
1 132  
Name: X326, dtype: int64

Unique values of column X327 :  
0 3660  
1 549  
Name: X327, dtype: int64

Unique values of column X328 :  
0 4056  
1 153  
Name: X328, dtype: int64

Unique values of column X329 :  
0 2365  
1 1844  
Name: X329, dtype: int64

Unique values of column X331 :  
0 3991  
1 218  
Name: X331, dtype: int64

Unique values of column X332 :  
0 4206  
1 3  
Name: X332, dtype: int64

Unique values of column X333 :  
0 4127  
1 82  
Name: X333, dtype: int64

Unique values of column X334 :  
0 2308  
1 1901

Name: X334, dtype: int64

Unique values of column X335 :

0 4200  
1 9

Name: X335, dtype: int64

Unique values of column X336 :

0 3672  
1 537

Name: X336, dtype: int64

Unique values of column X337 :

1 2218  
0 1991

Name: X337, dtype: int64

Unique values of column X338 :

0 4176  
1 33

Name: X338, dtype: int64

Unique values of column X339 :

0 4208  
1 1

Name: X339, dtype: int64

Unique values of column X340 :

0 4123  
1 86

Name: X340, dtype: int64

Unique values of column X341 :

0 4167  
1 42

Name: X341, dtype: int64

Unique values of column X342 :

0 4124  
1 85

Name: X342, dtype: int64

Unique values of column X343 :

0 3916  
1 293

Name: X343, dtype: int64

Unique values of column X344 :

0 4169  
1 40

Name: X344, dtype: int64

Unique values of column X345 :

0 4105  
1 104

Name: X345, dtype: int64

Unique values of column X346 :

0 4020  
1 189

Name: X346, dtype: int64

Unique values of column X348 :

1 3999  
0 210

Name: X348, dtype: int64

Unique values of column X349 :

0 4033  
1 176

Name: X349, dtype: int64

Unique values of column X350 :

0 2773  
1 1436

Name: X350, dtype: int64

Unique values of column X351 :

```
0    3000
1    1209
Name: X351, dtype: int64
```

```
Unique values of column X352 :
0    3998
1     211
Name: X352, dtype: int64
```

```
Unique values of column X353 :
0    4198
1      11
Name: X353, dtype: int64
```

```
Unique values of column X354 :
0    3351
1     858
Name: X354, dtype: int64
```

```
Unique values of column X355 :
0    2619
1    1590
Name: X355, dtype: int64
```

```
Unique values of column X356 :
0    3434
1     775
Name: X356, dtype: int64
```

```
Unique values of column X357 :
0    4199
1      10
Name: X357, dtype: int64
```

```
Unique values of column X358 :
0    2384
1    1825
Name: X358, dtype: int64
```

```
Unique values of column X359 :
0    4054
1     155
Name: X359, dtype: int64
```

```
Unique values of column X360 :
0    3874
1     335
Name: X360, dtype: int64
```

```
Unique values of column X361 :
1    4051
0     158
Name: X361, dtype: int64
```

```
Unique values of column X362 :
1    2180
0    2029
Name: X362, dtype: int64
```

```
Unique values of column X363 :
1    3137
0    1072
Name: X363, dtype: int64
```

```
Unique values of column X364 :
0    4194
1      15
Name: X364, dtype: int64
```

```
Unique values of column X365 :
0    4195
1      14
Name: X365, dtype: int64
```

```
Unique values of column X366 :
0    4202
1        7
Name: X366, dtype: int64
```

Unique values of column X367 :  
0 4006  
1 203  
Name: X367, dtype: int64

Unique values of column X368 :  
0 3943  
1 266  
Name: X368, dtype: int64

Unique values of column X369 :  
0 4209  
Name: X369, dtype: int64

Unique values of column X370 :  
0 4185  
1 24  
Name: X370, dtype: int64

Unique values of column X371 :  
0 4161  
1 48  
Name: X371, dtype: int64

Unique values of column X372 :  
0 4208  
1 1  
Name: X372, dtype: int64

Unique values of column X373 :  
0 4148  
1 61  
Name: X373, dtype: int64

Unique values of column X374 :  
0 3216  
1 993  
Name: X374, dtype: int64

Unique values of column X375 :  
0 2837  
1 1372  
Name: X375, dtype: int64

Unique values of column X376 :  
0 4000  
1 209  
Name: X376, dtype: int64

Unique values of column X377 :  
0 2896  
1 1313  
Name: X377, dtype: int64

Unique values of column X378 :  
0 4128  
1 81  
Name: X378, dtype: int64

Unique values of column X379 :  
0 4159  
1 50  
Name: X379, dtype: int64

Unique values of column X380 :  
0 4175  
1 34  
Name: X380, dtype: int64

Unique values of column X382 :  
0 4172  
1 37  
Name: X382, dtype: int64

Unique values of column X383 :  
0 4207  
1 2  
Name: X383, dtype: int64

```

Unique values of column X384 :
0      4206
1         3
Name: X384, dtype: int64

```

```

Unique values of column X385 :
0      4202
1         7
Name: X385, dtype: int64

```

```

Unique values of column Source :
test      4209
Name: Source, dtype: int64

```

- Since 8 columns are categorical, the label encoder will be apply.

#### Applying Label Encoder

Note: In order to have the same result, the two datasets have to be combined.

```

In [45]: label_encoder2 = LabelEncoder()

data_combined = pd.concat([data_train, data_test], axis=0, ignore_index=True)

for column in data_combined.columns[:365]:
    if data_combined[column].dtypes == 'object':
        data_combined[column] = label_encoder2.fit_transform(data_combined[column])

data_train_encoded = data_combined.loc[data_combined["Source"] == "train"]
data_train_encoded.drop(['Source'], axis=1, inplace=True)
data_test_encoded = data_combined.loc[data_combined["Source"] == "test"]
data_test_encoded.drop(["y", "Source"], axis=1, inplace=True)

data_train_encoded.head()

```

```

Out[45]:
   y  X0  X1  X2  X3  X4  X5  X6  X8  X10  ...  X375  X376  X377  X378  X379  X380  X382  X383  X384  X385
0  130.81  37  23  20  0  3  27  9  14  0  ...  0  0  1  0  0  0  0  0  0  0
1  88.53  37  21  22  4  3  31  11  14  0  ...  1  0  0  0  0  0  0  0  0  0
2  76.26  24  24  38  2  3  30  9  23  0  ...  0  0  0  0  0  0  1  0  0  0
3  80.62  24  21  38  5  3  30  11  4  0  ...  0  0  0  0  0  0  0  0  0  0
4  78.02  24  23  38  5  3  14  3  13  0  ...  0  0  0  0  0  0  0  0  0  0

```

5 rows × 365 columns

```

In [46]: data_test_encoded.head()

```

```

Out[46]:
   X0  X1  X2  X3  X4  X5  X6  X8  X10  X12  ...  X375  X376  X377  X378  X379  X380  X382  X383  X384  X385
4209  24  23  38  5  3  26  0  22  0  0  ...  0  0  0  1  0  0  0  0  0  0
4210  46  3  9  0  3  9  6  24  0  0  ...  0  0  1  0  0  0  0  0  0  0
4211  24  23  19  5  3  0  9  9  0  0  ...  0  0  0  1  0  0  0  0  0  0
4212  24  13  38  5  3  32  11  13  0  0  ...  0  0  0  1  0  0  0  0  0  0
4213  49  20  19  2  3  31  8  12  0  0  ...  1  0  0  0  0  0  0  0  0  0

```

5 rows × 364 columns

#### Dimensionality reduction

```

In [47]: correlation_matrix2 = data_train.corr()

```

```

In [48]: list_no_corr2 = []
for column in correlation_matrix2.columns:
    if np.isnan(correlation_matrix2.loc["y", column]) == True:
        print(f' y has no correlation with {column}')
        list_no_corr2.append(column)

print(list_no_corr2)
print(len(list_no_corr2))

[]
0

```

- All the features in the data\_train dataframe are correlated to the target "y".

```

In [49]: correlation_matrix2.iloc[:20, :20]

```

Out [49]:

	y	X10	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	
y	1.000000	-0.026985	0.089792	0.048276	0.193643	0.023116	0.048946	-0.159815	-0.001789	-0.168895	-0.095148	0.031111	-0.094320	-0.07
X10	-0.026985	1.000000	-0.033084	-0.028806	-0.100474	-0.002532	-0.005944	-0.010164	-0.010323	-0.038610	-0.047393	-0.005944	-0.035836	-0.01
X12	0.089792	-0.033084	1.000000	0.214825	-0.246513	-0.006212	-0.014584	-0.024937	-0.025327	-0.094730	-0.116280	-0.014584	-0.087924	-0.04
X13	0.048276	-0.028806	0.214825	1.000000	-0.083141	-0.005409	-0.012698	-0.021713	-0.010525	-0.082482	-0.043126	0.007215	0.226530	-0.03
X14	0.193643	-0.100474	-0.246513	-0.083141	1.000000	-0.018865	-0.044291	0.012713	-0.076916	-0.287691	-0.353137	-0.044291	-0.267021	-0.12
X15	0.023116	-0.002532	-0.006212	-0.005409	-0.018865	1.000000	-0.001116	-0.001908	-0.001938	-0.007250	-0.008899	-0.001116	-0.006729	-0.00
X16	0.048946	-0.005944	-0.014584	-0.012698	-0.044291	-0.001116	1.000000	-0.004480	-0.004550	-0.017020	-0.020892	-0.002620	-0.015797	-0.00
X17	-0.159815	-0.010164	-0.024937	-0.021713	0.012713	-0.001908	-0.004480	1.000000	0.085256	-0.029102	-0.027905	-0.004480	-0.017304	-0.01
X18	-0.001789	-0.010323	-0.025327	-0.010525	-0.076916	-0.001938	-0.004550	0.085256	1.000000	-0.029557	-0.036281	0.575835	-0.027434	-0.01
X19	-0.168895	-0.038610	-0.094730	-0.082482	-0.287691	-0.007250	-0.017020	-0.029102	-0.029557	1.000000	-0.135704	-0.017020	-0.102611	-0.04
X20	-0.095148	-0.047393	-0.116280	-0.043126	-0.353137	-0.008899	-0.020892	-0.027905	-0.036281	-0.135704	1.000000	-0.020892	-0.125953	-0.05
X21	0.031111	-0.005944	-0.014584	0.007215	-0.044291	-0.001116	-0.002620	-0.004480	0.575835	-0.017020	-0.020892	1.000000	-0.015797	-0.00
X22	-0.094320	-0.035836	-0.087924	0.226530	-0.267021	-0.006729	-0.015797	-0.017304	-0.027434	-0.102611	-0.125953	-0.015797	1.000000	-0.04
X23	-0.078676	-0.016870	-0.041391	-0.036040	-0.125703	-0.003168	-0.007437	-0.012716	-0.012915	-0.048305	-0.059294	-0.007437	-0.044834	1.00
X24	-0.003526	-0.005067	-0.012433	-0.010825	-0.004686	-0.000951	-0.002234	-0.003820	-0.003879	0.040148	-0.002219	-0.002234	-0.013467	-0.00
X26	-0.021173	-0.008223	-0.020175	-0.017566	-0.061270	-0.001544	-0.003625	0.226724	-0.006295	-0.023545	-0.028901	-0.003625	-0.021853	-0.01
X27	-0.053483	0.070276	0.184599	0.110191	0.126863	-0.031974	-0.075065	0.059687	-0.078274	0.063073	-0.327143	-0.075065	0.210446	-0.12
X28	-0.250288	-0.021300	-0.052259	-0.039773	-0.158707	-0.003999	-0.009389	0.045600	-0.016305	-0.060988	-0.074862	-0.009389	-0.056606	-0.02
X29	-0.379629	-0.024615	-0.042617	-0.052586	-0.076904	-0.004622	-0.010851	0.412903	0.060829	-0.070483	-0.083169	-0.010851	-0.061262	-0.02
X30	-0.043247	-0.007820	-0.019185	-0.016705	-0.058265	-0.001468	-0.003447	-0.005894	-0.005986	-0.022390	-0.027484	-0.003447	-0.020781	-0.00

In [50]:

```
for row in correlation_matrix2.index:
    list_corr2 = []
    for column in correlation_matrix2.columns:
        if correlation_matrix2.loc[row, column] >= 0.65 and row != column:
            list_corr2.append(column)
    if len(list_corr2) > 0:
        print(f'{row} is highly correlated to {list_corr2}')
```

X14 is highly correlated to ['X191', 'X223', 'X224', 'X251']  
X15 is highly correlated to ['X260']  
X17 is highly correlated to ['X325', 'X382']  
X19 is highly correlated to ['X215']  
X21 is highly correlated to ['X184', 'X262', 'X266']  
X28 is highly correlated to ['X29', 'X43', 'X54', 'X76', 'X148', 'X162', 'X166', 'X232', 'X272', 'X276', 'X279', 'X328']  
X29 is highly correlated to ['X28', 'X54', 'X76', 'X162', 'X166', 'X185', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X31 is highly correlated to ['X35', 'X37']  
X32 is highly correlated to ['X200']  
X33 is highly correlated to ['X39']  
X35 is highly correlated to ['X31', 'X37']  
X37 is highly correlated to ['X31', 'X35']  
X39 is highly correlated to ['X33']  
X43 is highly correlated to ['X28']  
X44 is highly correlated to ['X302', 'X317']  
X45 is highly correlated to ['X161', 'X202', 'X247', 'X356', 'X377']  
X47 is highly correlated to ['X267']  
X48 is highly correlated to ['X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X198', 'X222', 'X228', 'X264', 'X306']  
X49 is highly correlated to ['X50', 'X129']  
X50 is highly correlated to ['X49', 'X129']  
X52 is highly correlated to ['X48', 'X75', 'X113', 'X134', 'X147', 'X198', 'X222']  
X53 is highly correlated to ['X102', 'X214', 'X239', 'X282', 'X308', 'X370']  
X54 is highly correlated to ['X28', 'X29', 'X76', 'X162', 'X166', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X57 is highly correlated to ['X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X58 is highly correlated to ['X137', 'X324']  
X60 is highly correlated to ['X248', 'X253', 'X385']  
X61 is highly correlated to ['X120']  
X62 is highly correlated to ['X172', 'X216', 'X221', 'X344']  
X63 is highly correlated to ['X78', 'X174', 'X379']  
X66 is highly correlated to ['X48', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X198', 'X222', 'X264']  
X67 is highly correlated to ['X213']  
X69 is highly correlated to ['X197', 'X331', 'X346', 'X349', 'X352', 'X367']  
X71 is highly correlated to ['X84', 'X241', 'X244']  
X75 is highly correlated to ['X48', 'X52', 'X66', 'X113', 'X134', 'X147', 'X170', 'X198', 'X222']  
X76 is highly correlated to ['X28', 'X29', 'X54', 'X162', 'X166', 'X232', 'X272', 'X276', 'X279', 'X328', 'X378']  
X78 is highly correlated to ['X63', 'X379']  
X80 is highly correlated to ['X98', 'X101', 'X111', 'X128', 'X265', 'X304', 'X348']  
X84 is highly correlated to ['X71', 'X241', 'X244']  
X85 is highly correlated to ['X187', 'X194', 'X246', 'X358']  
X87 is highly correlated to ['X89', 'X245', 'X357']  
X88 is highly correlated to ['X57', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X89 is highly correlated to ['X87', 'X245', 'X357']  
X90 is highly correlated to ['X57', 'X88', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X94 is highly correlated to ['X57', 'X88', 'X90', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X96 is highly correlated to ['X150', 'X363']  
X98 is highly correlated to ['X80', 'X101', 'X111', 'X128', 'X304', 'X348']  
X99 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X122', 'X217', 'X242', 'X243', 'X249', 'X320']  
X100 is highly correlated to ['X103', 'X144']  
X101 is highly correlated to ['X80', 'X98', 'X128', 'X304', 'X348']  
X102 is highly correlated to ['X53', 'X214', 'X239', 'X282', 'X308', 'X370']  
X103 is highly correlated to ['X100', 'X144']  
X108 is highly correlated to ['X371']  
X109 is highly correlated to ['X177']  
X111 is highly correlated to ['X80', 'X98', 'X120', 'X128', 'X348']  
X112 is highly correlated to ['X183', 'X199']  
X113 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X126', 'X130', 'X134', 'X147', 'X170', 'X198', 'X222', 'X228', 'X264', 'X306']  
X114 is highly correlated to ['X155', 'X360']  
X116 is highly correlated to ['X294']  
X117 is highly correlated to ['X135']  
X118 is highly correlated to ['X119', 'X275', 'X311']  
X119 is highly correlated to ['X118', 'X275', 'X311']  
X120 is highly correlated to ['X61', 'X111']  
X122 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X217', 'X242', 'X243', 'X249', 'X320']  
X125 is highly correlated to ['X227', 'X277']  
X126 is highly correlated to ['X48', 'X66', 'X113', 'X130', 'X134', 'X147', 'X170', 'X179', 'X198', 'X222', 'X264', 'X306']  
X127 is highly correlated to ['X313']  
X128 is highly correlated to ['X80', 'X98', 'X101', 'X111', 'X189', 'X304', 'X348']  
X129 is highly correlated to ['X49', 'X50']  
X130 is highly correlated to ['X48', 'X66', 'X113', 'X126', 'X134', 'X147', 'X179', 'X198', 'X222', 'X264']  
X133 is highly correlated to ['X161', 'X202', 'X247', 'X356']  
X134 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X147', 'X170', 'X198', 'X222', 'X228', 'X264', 'X306']  
X135 is highly correlated to ['X117']  
X136 is highly correlated to ['X189', 'X238', 'X263', 'X265']  
X137 is highly correlated to ['X58', 'X324']  
X138 is highly correlated to ['X140', 'X146', 'X208', 'X368']  
X140 is highly correlated to ['X138', 'X146', 'X208', 'X368']  
X144 is highly correlated to ['X100', 'X103', 'X209']  
X146 is highly correlated to ['X138', 'X140', 'X208', 'X368']  
X147 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X170', 'X198', 'X222', 'X228', 'X264', 'X306']  
X148 is highly correlated to ['X28', 'X166']  
X150 is highly correlated to ['X96', 'X363']  
X152 is highly correlated to ['X226', 'X326']  
X154 is highly correlated to ['X283', 'X321', 'X374']  
X155 is highly correlated to ['X114', 'X219', 'X360']  
X159 is highly correlated to ['X274']  
X161 is highly correlated to ['X45', 'X133', 'X202', 'X247', 'X356', 'X377']  
X162 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X166', 'X232', 'X272', 'X276', 'X279', 'X328']  
X166 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X148', 'X162', 'X232', 'X272', 'X276', 'X279', 'X328']  
X167 is highly correlated to ['X277']  
X170 is highly correlated to ['X48', 'X66', 'X75', 'X113', 'X126', 'X134', 'X147', 'X198', 'X222', 'X228', 'X264', 'X306']



X172 is highly correlated to ['X62', 'X216', 'X221', 'X344']  
X174 is highly correlated to ['X63', 'X379']  
X177 is highly correlated to ['X109']  
X178 is highly correlated to ['X250']  
X179 is highly correlated to ['X126', 'X130', 'X264']  
X181 is highly correlated to ['X356']  
X183 is highly correlated to ['X112', 'X199']  
X184 is highly correlated to ['X21', 'X262', 'X266']  
X185 is highly correlated to ['X29', 'X232', 'X279', 'X378']  
X186 is highly correlated to ['X362']  
X187 is highly correlated to ['X85', 'X194', 'X246', 'X358']  
X189 is highly correlated to ['X128', 'X136', 'X238', 'X263', 'X265']  
X191 is highly correlated to ['X14', 'X224', 'X251']  
X194 is highly correlated to ['X85', 'X187', 'X246', 'X358']  
X197 is highly correlated to ['X69', 'X331', 'X346', 'X349', 'X352', 'X367']  
X198 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X222', 'X264', 'X306']  
X199 is highly correlated to ['X112', 'X183']  
X200 is highly correlated to ['X32']  
X202 is highly correlated to ['X45', 'X133', 'X161', 'X247', 'X356', 'X377']  
X208 is highly correlated to ['X138', 'X140', 'X146', 'X368']  
X209 is highly correlated to ['X144']  
X213 is highly correlated to ['X67']  
X214 is highly correlated to ['X53', 'X102', 'X239', 'X282', 'X308', 'X370']  
X215 is highly correlated to ['X19']  
X216 is highly correlated to ['X62', 'X172', 'X221', 'X344']  
X217 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X242', 'X243', 'X249', 'X320']  
X219 is highly correlated to ['X155', 'X360']  
X221 is highly correlated to ['X62', 'X172', 'X216']  
X222 is highly correlated to ['X48', 'X52', 'X66', 'X75', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X198', 'X228', 'X264', 'X306']  
X223 is highly correlated to ['X14', 'X251']  
X224 is highly correlated to ['X14', 'X191', 'X251']  
X226 is highly correlated to ['X152', 'X326']  
X227 is highly correlated to ['X125', 'X277']  
X228 is highly correlated to ['X48', 'X113', 'X134', 'X147', 'X170', 'X222']  
X230 is highly correlated to ['X254']  
X232 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X185', 'X272', 'X276', 'X279', 'X328', 'X378']  
X236 is highly correlated to ['X339']  
X237 is highly correlated to ['X309', 'X312', 'X317']  
X238 is highly correlated to ['X136', 'X189', 'X263', 'X265']  
X239 is highly correlated to ['X53', 'X102', 'X214', 'X282', 'X308', 'X370']  
X240 is highly correlated to ['X364', 'X365']  
X241 is highly correlated to ['X71', 'X84', 'X244']  
X242 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X243', 'X249', 'X320']  
X243 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X249', 'X320']  
X244 is highly correlated to ['X71', 'X84', 'X241']  
X245 is highly correlated to ['X87', 'X89', 'X357']  
X246 is highly correlated to ['X85', 'X187', 'X194', 'X358']  
X247 is highly correlated to ['X45', 'X133', 'X161', 'X202', 'X356', 'X377']  
X248 is highly correlated to ['X60', 'X253', 'X385']  
X249 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X320']  
X250 is highly correlated to ['X178']  
X251 is highly correlated to ['X14', 'X191', 'X223', 'X224']  
X253 is highly correlated to ['X60', 'X248', 'X385']  
X254 is highly correlated to ['X230']  
X255 is highly correlated to ['X308']  
X260 is highly correlated to ['X15']  
X261 is highly correlated to ['X314']  
X262 is highly correlated to ['X21', 'X184', 'X266']  
X263 is highly correlated to ['X136', 'X189', 'X238', 'X265']  
X264 is highly correlated to ['X48', 'X66', 'X113', 'X126', 'X130', 'X134', 'X147', 'X170', 'X179', 'X198', 'X222', 'X306']  
X265 is highly correlated to ['X80', 'X136', 'X189', 'X238', 'X263', 'X348']  
X266 is highly correlated to ['X21', 'X184', 'X262']  
X267 is highly correlated to ['X47']  
X272 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X232', 'X276', 'X279', 'X328']  
X274 is highly correlated to ['X159']  
X275 is highly correlated to ['X118', 'X119', 'X311']  
X276 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X232', 'X272', 'X279', 'X328']  
X277 is highly correlated to ['X125', 'X167', 'X227']  
X279 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X185', 'X232', 'X272', 'X276', 'X328', 'X378']  
X282 is highly correlated to ['X53', 'X102', 'X214', 'X239', 'X308', 'X370']  
X283 is highly correlated to ['X154', 'X321', 'X374']  
X285 is highly correlated to ['X336']  
X294 is highly correlated to ['X116']  
X295 is highly correlated to ['X296']  
X296 is highly correlated to ['X295']  
X298 is highly correlated to ['X299']  
X299 is highly correlated to ['X298']  
X302 is highly correlated to ['X44', 'X317']  
X304 is highly correlated to ['X80', 'X98', 'X101', 'X128', 'X348']  
X306 is highly correlated to ['X48', 'X113', 'X126', 'X134', 'X147', 'X170', 'X198', 'X222', 'X264']  
X308 is highly correlated to ['X53', 'X102', 'X214', 'X239', 'X255', 'X282', 'X370']  
X309 is highly correlated to ['X237', 'X312']  
X311 is highly correlated to ['X118', 'X119', 'X275']  
X312 is highly correlated to ['X237', 'X309', 'X317']  
X313 is highly correlated to ['X127']  
X314 is highly correlated to ['X261']  
X317 is highly correlated to ['X44', 'X237', 'X302', 'X312']  
X320 is highly correlated to ['X57', 'X88', 'X90', 'X94', 'X99', 'X122', 'X217', 'X242', 'X243', 'X249']  
X321 is highly correlated to ['X154', 'X283', 'X374']  
X324 is highly correlated to ['X58', 'X137']  
X325 is highly correlated to ['X17', 'X382']  
X326 is highly correlated to ['X152', 'X226']  
X328 is highly correlated to ['X28', 'X29', 'X54', 'X76', 'X162', 'X166', 'X232', 'X272', 'X276', 'X279', 'X378']

X331 is highly correlated to ['X69', 'X197', 'X346', 'X349', 'X352', 'X367']  
 X333 is highly correlated to ['X373']  
 X334 is highly correlated to ['X377']  
 X336 is highly correlated to ['X285']  
 X339 is highly correlated to ['X236']  
 X343 is highly correlated to ['X376']  
 X344 is highly correlated to ['X62', 'X172', 'X216']  
 X346 is highly correlated to ['X69', 'X197', 'X331', 'X349', 'X352', 'X367']  
 X348 is highly correlated to ['X80', 'X98', 'X101', 'X111', 'X128', 'X265', 'X304']  
 X349 is highly correlated to ['X69', 'X197', 'X331', 'X346', 'X352', 'X367']  
 X352 is highly correlated to ['X69', 'X197', 'X331', 'X346', 'X349', 'X367']  
 X356 is highly correlated to ['X45', 'X133', 'X161', 'X181', 'X202', 'X247']  
 X357 is highly correlated to ['X87', 'X89', 'X245']  
 X358 is highly correlated to ['X85', 'X187', 'X194', 'X246']  
 X360 is highly correlated to ['X114', 'X155', 'X219']  
 X362 is highly correlated to ['X186']  
 X363 is highly correlated to ['X96', 'X150']  
 X364 is highly correlated to ['X240', 'X365']  
 X365 is highly correlated to ['X240', 'X364']  
 X367 is highly correlated to ['X69', 'X197', 'X331', 'X346', 'X349', 'X352']  
 X368 is highly correlated to ['X138', 'X140', 'X146', 'X208']  
 X370 is highly correlated to ['X53', 'X102', 'X214', 'X239', 'X282', 'X308']  
 X371 is highly correlated to ['X108']  
 X373 is highly correlated to ['X333']  
 X374 is highly correlated to ['X154', 'X283', 'X321']  
 X376 is highly correlated to ['X343']  
 X377 is highly correlated to ['X45', 'X161', 'X202', 'X247', 'X334']  
 X378 is highly correlated to ['X29', 'X54', 'X76', 'X185', 'X232', 'X279', 'X328']  
 X379 is highly correlated to ['X63', 'X78', 'X174']  
 X382 is highly correlated to ['X17', 'X325']  
 X385 is highly correlated to ['X60', 'X248', 'X253']

- Some columns are highly correlated together so the features the most correlated to 'y' will be kept.

```
In [51]: features_to_keep2 = []
for row in correlation_matrix2.index:
    dict_corr2 = {row: correlation_matrix2.loc["y", row]}
    for column in correlation_matrix2.columns:
        if correlation_matrix2.loc[row, column] >= 0.65 and row != column:
            dict_corr2[column] = correlation_matrix2.loc["y", column]
    to_keep2 = max(dict_corr2, key=dict_corr2.get)
    if to_keep2 not in features_to_keep2:
        features_to_keep2.append(to_keep2)

print(sorted(features_to_keep2))
print(len(features_to_keep2))

['X10', 'X100', 'X102', 'X104', 'X105', 'X106', 'X109', 'X110', 'X115', 'X116', 'X117', 'X12', 'X120', 'X123', 'X124', 'X125', 'X13', 'X131', 'X132', 'X137', 'X139', 'X141', 'X142', 'X143', 'X144', 'X145', 'X148', 'X150', 'X151', 'X152', 'X153', 'X155', 'X156', 'X157', 'X158', 'X16', 'X160', 'X163', 'X164', 'X165', 'X167', 'X168', 'X169', 'X171', 'X173', 'X175', 'X176', 'X178', 'X18', 'X180', 'X181', 'X182', 'X183', 'X185', 'X186', 'X187', 'X189', 'X19', 'X190', 'X192', 'X195', 'X196', 'X198', 'X20', 'X200', 'X201', 'X202', 'X203', 'X204', 'X205', 'X206', 'X207', 'X208', 'X21', 'X210', 'X211', 'X212', 'X213', 'X214', 'X218', 'X22', 'X220', 'X221', 'X225', 'X226', 'X227', 'X229', 'X23', 'X230', 'X231', 'X234', 'X236', 'X239', 'X24', 'X240', 'X244', 'X245', 'X248', 'X251', 'X252', 'X253', 'X254', 'X255', 'X256', 'X257', 'X258', 'X259', 'X26', 'X260', 'X263', 'X264', 'X265', 'X269', 'X27', 'X270', 'X271', 'X273', 'X274', 'X275', 'X278', 'X28', 'X280', 'X281', 'X283', 'X284', 'X286', 'X287', 'X288', 'X291', 'X292', 'X295', 'X296', 'X298', 'X299', 'X30', 'X300', 'X301', 'X305', 'X307', 'X309', 'X31', 'X310', 'X313', 'X314', 'X315', 'X316', 'X317', 'X318', 'X319', 'X32', 'X323', 'X325', 'X326', 'X327', 'X329', 'X33', 'X331', 'X332', 'X333', 'X335', 'X336', 'X337', 'X338', 'X34', 'X340', 'X341', 'X342', 'X343', 'X345', 'X35', 'X350', 'X351', 'X353', 'X354', 'X355', 'X359', 'X36', 'X360', 'X361', 'X364', 'X365', 'X366', 'X369', 'X37', 'X371', 'X372', 'X375', 'X377', 'X379', 'X38', 'X380', 'X383', 'X384', 'X385', 'X39', 'X40', 'X41', 'X42', 'X43', 'X46', 'X47', 'X48', 'X50', 'X51', 'X53', 'X55', 'X56', 'X57', 'X59', 'X60', 'X62', 'X64', 'X65', 'X66', 'X67', 'X68', 'X70', 'X71', 'X73', 'X74', 'X77', 'X79', 'X81', 'X82', 'X83', 'X84', 'X86', 'X89', 'X91', 'X92', 'X95', 'X97', 'X98', 'y']
234
```

- We will keep 233 features for modeling.

```
In [52]: final_data_train2 = data_train_encoded[features_to_keep2]
final_data_train2.head()
```

```
Out[52]:
```

	y	X10	X12	X13	X251	X260	X16	X325	X18	X19	...	X364	X365	X366	X369	X372	X375	X380	X383	X384	X385
0	130.81	0	0	1	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0
1	88.53	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	1	0	0	0
2	76.26	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	80.62	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	0	0	0	0	0
4	78.02	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 234 columns

```
In [53]: final_data_test2 = data_test_encoded[features_to_keep2[1:234]]
final_data_test2.head()
```

```
Out[53]:
```

	X10	X12	X13	X251	X260	X16	X325	X18	X19	X20	...	X364	X365	X366	X369	X372	X375	X380	X383	X384	X385
4209	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4210	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0	0
4211	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4212	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4213	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	1	0	0	0	0

5 rows × 233 columns

Splitting the train dataset

```
In [54]: y_train2 = final_data_train2["y"]
X_train2 = final_data_train2.drop(['y'], axis=1)
```

```
In [55]: X_train2.head()
```

```
Out[55]:
```

	X10	X12	X13	X251	X260	X16	X325	X18	X19	X20	...	X364	X365	X366	X369	X372	X375	X380	X383	X384	X385
0	0	0	1	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	1	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 233 columns

```
In [56]: y_train2.head()
```

```
Out[56]: 0    130.81
1     88.53
2     76.26
3     80.62
4     78.02
Name: y, dtype: float64
```

Creation of a train and test split for the model

```
In [57]: xtrain2, xtest2, ytrain2, ytest2 = train_test_split(X_train2, y_train2, test_size=0.3, random_state=67)
```

Model 2: Fitting the XGBoost Regressor to the training dataset with K-fold cross validation

```
In [58]: xgboost_regressor = XGBRegressor(objective='reg:squarederror', n_estimators=10, random_state=67)
xgboost_regressor.fit(xtrain2, ytrain2)
```

```
Out[58]: XGBRegressor
XGBRegressor(base_score=0.5, booster='gbtree', callbacks=None,
              colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1,
              early_stopping_rounds=None, enable_categorical=False,
              eval_metric=None, gamma=0, gpu_id=-1, grow_policy='depthwise',
              importance_type=None, interaction_constraints='',
              learning_rate=0.300000012, max_bin=256, max_cat_to_onehot=4,
              max_delta_step=0, max_depth=6, max_leaves=0, min_child_weight=1,
              missing=nan, monotone_constraints=(), n_estimators=10, n_jobs=0,
              num_parallel_tree=1, predictor='auto', random_state=67,
              reg_alpha=0, reg_lambda=1, ...)
```

```
In [59]: predictions2 = xgboost_regressor.predict(xtest2)
```

```
In [60]: kfold = model_selection.KFold(n_splits=10)

results2 = model_selection.cross_val_score(xgboost_regressor, X_train2, y_train2, cv=kfold, scoring='neg_root_mean_squared_error')

print(f'RMSE of model 2: {abs(results.mean())}')

RMSE of model 2: 8.715858578949872
```

Prediction of the test dataset (final\_data\_test)

```
In [61]: test_df2 = pd.DataFrame(xgboost_regressor.predict(final_data_test2), columns=['y'])
test_df2.head()
```

Out[61]:

	<b>y</b>
<b>0</b>	88.031326
<b>1</b>	103.009399
<b>2</b>	87.392166
<b>3</b>	75.877495
<b>4</b>	109.107849

In [ ]: