mencilxys

May 9, 2023

0.1 Mercedes-Benz Greener Manufacturing using XGBOOST

[ ]:

DESCRIPTION

Reduce the time a Mercedes-Benz spends on the test bench.

Problem Statement Scenario: To ensure the safety and reliability of every unique car configuration before they hit the road, the company's engineers have developed a robust testing system. As one of the world's biggest manufacturers of premium cars, safety and efficiency are paramount on Mercedes-Benz's production lines. However, optimizing the speed of their testing system for many possible feature combinations is complex and time-consuming without a powerful algorithmic approach.

You are required to reduce the time that cars spend on the test bench. Others will work with a dataset representing different permutations of features in a Mercedes-Benz car to predict the time it takes to pass testing. Optimal algorithms will contribute to faster testing, resulting in lower carbon dioxide emissions without reducing Mercedes-Benz's standards.

Following actions should be performed:

If for any column(s), the variance is equal to zero, then you need to remove those variable(s). Check for null and unique values for test and train sets. Apply label encoder. Perform dimensionality reduction. Predict your test_df values using XGBoost

[1]:
```
#importing the libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,r2_score
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
from warnings import filterwarnings
filterwarnings('ignore')
%matplotlib inline
```

[ ]:

[2]:
```
#importing dataset

train=pd.read_csv('train.csv')
test=pd.read_csv('test.csv')
```

[ ]:
```

```

[3]:
```
## Reading first 5 records

train.head()
```

[3]:

	ID	y	X0	X1	X2	X3	X4	X5	X6	X8	X10	X11	X12	X13	X14	X15	X16	\
0	0	130.81	k	v	at	a	d	u	j	o	0	0	0	1	0	0	0	
1	6	88.53	k	t	av	e	d	y	l	o	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	0	
3	9	80.62	az	t	n	f	d	x	l	e	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	

	X17	X18	X19	X20	X21	X22	X23	X24	X26	X27	X28	X29	X30	X31	X32	\
0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	
1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	
2	1	0	0	0	0	0	0	0	0	1	1	1	0	1	0	
3	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	
4	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	

	X33	X34	X35	X36	X37	X38	X39	X40	X41	X42	X43	X44	X45	X46	X47	\
0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	
1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	
2	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	
3	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	
4	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	

	X48	X49	X50	X51	X52	X53	X54	X55	X56	X57	X58	X59	X60	X61	X62	\
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	
2	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	
3	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	
4	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	

	X63	X64	X65	X66	X67	X68	X69	X70	X71	X73	X74	X75	X76	X77	X78	\
0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	
1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	
2	0	0	0	0	0	1	0	1	1	0	1	1	1	0	0	
3	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	
4	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	

	X79	X80	X81	X82	X83	X84	X85	X86	X87	X88	X89	X90	X91	X92	X93	\

0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
2	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0
3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

	X94	X95	X96	X97	X98	X99	X100	X101	X102	X103	X104	X105	X106	\
0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	1	0	1	0	1	1	0	0	0	0	0	
2	0	0	1	0	1	0	0	1	0	0	0	0	0	
3	0	0	1	0	1	0	0	1	0	0	0	0	0	
4	0	0	1	0	1	0	0	1	0	0	0	0	0	

	X107	X108	X109	X110	X111	X112	X113	X114	X115	X116	X117	X118	\
0	0	0	0	0	1	0	0	1	0	1	0	1	
1	0	0	0	0	1	0	0	0	0	0	0	1	
2	0	0	0	0	1	0	0	0	0	0	0	0	
3	0	1	0	0	1	0	0	1	0	0	0	0	
4	0	1	0	0	1	0	0	1	0	0	0	0	

	X119	X120	X122	X123	X124	X125	X126	X127	X128	X129	X130	X131	\
0	1	1	0	0	0	0	0	0	1	0	0	1	
1	1	1	0	0	0	0	0	1	1	0	0	0	
2	0	1	0	0	0	0	0	0	1	0	0	0	
3	0	1	0	0	0	0	0	0	1	0	0	0	
4	0	1	0	0	0	0	0	0	1	0	0	0	

	X132	X133	X134	X135	X136	X137	X138	X139	X140	X141	X142	X143	\
0	0	0	0	0	1	1	0	0	0	0	1	0	
1	1	0	0	0	1	0	0	0	0	0	1	0	
2	1	0	0	0	0	1	0	0	0	0	0	0	
3	1	0	0	0	0	0	0	0	0	0	1	0	
4	1	0	0	0	0	0	0	0	0	0	0	0	

	X144	X145	X146	X147	X148	X150	X151	X152	X153	X154	X155	X156	\
0	1	0	0	0	0	1	0	0	0	0	0	1	
1	1	0	0	0	0	1	0	0	0	0	0	1	
2	1	0	0	0	1	1	0	0	0	0	0	0	
3	1	0	0	0	1	1	0	0	0	0	0	0	
4	1	0	0	0	1	1	0	0	0	0	0	0	

	X157	X158	X159	X160	X161	X162	X163	X164	X165	X166	X167	X168	\
0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	1	0	0	0	
2	1	1	0	0	0	1	1	0	0	0	0	0	
3	1	0	0	0	0	1	0	0	0	1	0	0	
4	1	1	0	0	0	1	0	0	0	1	0	0	

	X169	X170	X171	X172	X173	X174	X175	X176	X177	X178	X179	X180	\
0	0	1	0	0	0	0	0	0	0	0	1	0	
1	0	0	0	0	0	0	0	0	0	1	0	0	
2	1	0	0	0	0	0	0	0	0	0	1	0	
3	0	0	0	0	0	1	0	0	0	0	1	0	
4	0	0	0	0	0	0	0	0	0	0	1	0	

	X181	X182	X183	X184	X185	X186	X187	X189	X190	X191	X192	X194	\
0	0	0	0	1	0	0	1	1	0	0	0	1	
1	0	0	0	0	0	0	1	1	0	0	0	1	
2	0	0	0	0	0	0	0	0	0	0	0	1	
3	0	0	0	0	0	0	0	0	0	0	0	1	
4	0	0	0	0	0	0	0	0	0	0	0	1	

	X195	X196	X197	X198	X199	X200	X201	X202	X203	X204	X205	X206	\
0	0	0	0	0	0	0	0	0	0	1	0	0	
1	0	0	0	0	0	0	0	0	0	0	1	0	
2	0	0	0	0	0	0	0	0	0	0	1	1	
3	0	0	0	0	0	0	0	0	0	0	1	0	
4	0	0	0	0	0	0	0	0	0	0	1	0	

	X207	X208	X209	X210	X211	X212	X213	X214	X215	X216	X217	X218	\
0	0	0	1	0	0	0	0	0	0	0	0	0	
1	0	0	1	0	0	0	0	0	0	0	0	0	
2	0	0	1	0	0	0	0	0	0	0	0	1	
3	0	0	1	0	0	0	0	0	0	0	0	1	
4	0	0	1	0	0	0	0	0	0	0	0	1	

	X219	X220	X221	X222	X223	X224	X225	X226	X227	X228	X229	X230	\
0	0	1	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	1	0	
2	1	1	0	0	1	1	0	0	0	1	0	0	
3	0	1	0	0	1	0	0	0	0	0	1	0	
4	0	1	0	0	1	0	0	0	0	0	1	0	

	X231	X232	X233	X234	X235	X236	X237	X238	X239	X240	X241	X242	\
0	0	0	0	1	0	0	1	0	0	0	0	0	
1	0	0	0	0	0	0	0	1	0	0	0	0	
2	0	1	0	0	0	0	0	0	0	0	1	0	
3	0	1	0	0	0	0	0	0	0	0	1	0	
4	0	1	0	0	0	0	0	0	0	0	0	0	

	X243	X244	X245	X246	X247	X248	X249	X250	X251	X252	X253	X254	\
0	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	
2	0	1	0	1	0	0	0	1	0	1	0	0	

```
3     0     1     0     1     0     0     0     1     0     1     0     0
4     0     0     0     1     0     0     0     1     0     1     0     0

    X255  X256  X257  X258  X259  X260  X261  X262  X263  X264  X265  X266  \
0     0     0     0     0     0     0     0     1     1     0     0     1
1     0     0     0     0     0     0     0     0     1     0     1     0
2     0     1     0     0     0     0     0     0     0     0     0     0
3     0     1     0     0     0     0     0     0     0     0     0     0
4     0     1     0     0     0     0     0     0     0     0     0     0

    X267  X268  X269  X270  X271  X272  X273  X274  X275  X276  X277  X278  \
0     0     0     0     0     0     0     1     0     1     0     0     0
1     0     0     0     0     0     0     1     0     1     0     0     0
2     0     0     0     0     0     1     1     1     0     1     0     0
3     0     0     0     0     0     1     1     0     0     1     0     0
4     0     0     0     0     0     1     1     0     0     1     0     0

    X279  X280  X281  X282  X283  X284  X285  X286  X287  X288  X289  X290  \
0     0     0     0     0     0     0     1     0     0     0     0     0
1     0     0     0     0     0     0     1     0     0     0     0     0
2     1     0     0     0     0     0     0     1     0     0     0     0
3     1     0     0     0     0     0     0     1     0     0     0     0
4     1     0     0     0     0     0     0     1     0     0     0     0

    X291  X292  X293  X294  X295  X296  X297  X298  X299  X300  X301  X302  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
3     1     0     0     0     0     0     0     0     0     0     0     0
4     0     0     0     0     0     0     0     0     0     0     0     0

    X304  X305  X306  X307  X308  X309  X310  X311  X312  X313  X314  X315  \
0     0     0     1     0     0     0     0     0     0     0     0     0
1     1     0     0     0     0     0     0     1     0     0     0     0
2     1     0     0     0     0     0     0     0     0     0     0     0
3     1     0     0     0     0     0     0     0     0     0     0     0
4     1     0     0     0     0     0     0     0     0     0     0     0

    X316  X317  X318  X319  X320  X321  X322  X323  X324  X325  X326  X327  \
0     1     0     0     0     0     0     0     0     1     0     0     1
1     1     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     1     0     0     0
3     0     0     0     0     0     0     0     0     0     0     0     0
4     0     0     0     0     0     0     0     0     0     0     0     0

    X328  X329  X330  X331  X332  X333  X334  X335  X336  X337  X338  X339  \
0     0     1     0     0     0     0     1     0     0     0     0     0
```

```
1     0     1     0     0     0     0     0     0     1     1     0     0
2     1     0     0     0     0     0     1     0     0     0     0     0
3     1     0     0     0     0     0     0     0     0     0     0     0
4     1     0     0     0     0     0     1     0     0     0     0     0

   X340  X341  X342  X343  X344  X345  X346  X347  X348  X349  X350  X351  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     1     0     0     0
2     0     0     0     0     0     0     0     0     1     0     1     0
3     0     0     0     0     0     0     0     0     1     0     1     0
4     0     0     0     0     0     0     0     0     1     0     1     0

   X352  X353  X354  X355  X356  X357  X358  X359  X360  X361  X362  X363  \
0     0     0     1     0     0     0     0     0     0     1     0     0
1     0     0     0     0     0     0     0     0     0     1     0     1
2     0     0     1     0     0     0     1     0     0     1     0     1
3     0     0     0     0     0     0     1     0     0     1     0     1
4     0     0     0     0     0     0     1     0     0     1     0     1

   X364  X365  X366  X367  X368  X369  X370  X371  X372  X373  X374  X375  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     1
2     0     0     0     0     0     0     0     0     0     0     0     0
3     0     0     0     0     0     0     0     0     1     0     0     0
4     0     0     0     0     0     0     0     1     0     0     0     0

   X376  X377  X378  X379  X380  X382  X383  X384  X385
0     0     1     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     1     0     0     0
3     0     0     0     0     0     0     0     0     0
4     0     0     0     0     0     0     0     0     0
```

[ ]:

[4]:
```
#finding shape of the train dataset

train.shape
```

[4]:
```
(4209, 378)
```

[5]:
```
#Check for null and unique values for train sets.
```

[6]:
```
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
```

[ ]:

[7]:
```
#checking null values in train data

sum(train.isnull().sum())
```

[7]: 0

[8]:
```
#If for any column(s), the variance is equal to zero, then you need to remove␣
 ↪those variable(s).
```

[9]:
```
#checking variance

pd.options.display.float_format = '{:,.4f}'.format
var=train.var()
var=var.reset_index()
var.columns=['id','values']
var.sort_values(by='values',ascending=False)
```

[9]:

	id	values
0	ID	5,941,936.1180
1	y	160.7667
116	X127	0.2500
322	X337	0.2498
347	X362	0.2496
178	X191	0.2492
174	X186	0.2488
180	X194	0.2488
319	X334	0.2487
236	X250	0.2473
209	X223	0.2470
166	X178	0.2468
206	X220	0.2463
314	X329	0.2459
299	X314	0.2454
6	X14	0.2449
343	X358	0.2447
49	X58	0.2444
309	X324	0.2444
175	X187	0.2437
247	X261	0.2436
126	X137	0.2434
232	X246	0.2418

75	X85	0.2416
37	X46	0.2406
296	X311	0.2404
237	X251	0.2389
340	X355	0.2357
109	X119	0.2351
108	X118	0.2351
55	X64	0.2345
159	X171	0.2253
335	X350	0.2241
360	X375	0.2172
210	X224	0.2167
18	X27	0.2167
362	X377	0.2158
204	X218	0.2149
121	X132	0.2145
90	X100	0.2139
151	X163	0.2114
298	X313	0.2105
336	X351	0.2089
105	X115	0.2041
145	X157	0.2028
144	X156	0.2028
259	X273	0.2016
42	X51	0.2009
261	X275	0.1986
156	X168	0.1975
36	X45	0.1892
348	X363	0.1856
86	X96	0.1834
233	X247	0.1832
188	X202	0.1832
306	X321	0.1818
28	X37	0.1784
26	X35	0.1784
22	X31	0.1784
146	X158	0.1770
131	X142	0.1770
71	X81	0.1766
359	X374	0.1757
93	X103	0.1691
41	X50	0.1683
142	X154	0.1653
138	X150	0.1646
286	X300	0.1642
271	X285	0.1635
339	X354	0.1618

220	X234	0.1611
149	X161	0.1586
106	X116	0.1581
301	X316	0.1573
133	X144	0.1552
341	X356	0.1475
187	X201	0.1462
168	X180	0.1331
104	X114	0.1248
12	X20	0.1224
269	X283	0.1209
312	X327	0.1119
321	X336	0.1112
280	X294	0.1094
122	X133	0.1088
118	X129	0.1076
40	X49	0.1072
170	X182	0.0949
230	X244	0.0929
62	X71	0.0929
74	X84	0.0929
195	X209	0.0912
11	X19	0.0897
201	X215	0.0889
227	X241	0.0878
211	X225	0.0876
251	X265	0.0856
169	X181	0.0851
128	X139	0.0823
14	X22	0.0794
139	X151	0.0782
176	X189	0.0774
224	X238	0.0769
61	X70	0.0737
328	X343	0.0723
143	X155	0.0707
345	X360	0.0707
289	X304	0.0701
4	X12	0.0695
59	X68	0.0680
242	X256	0.0678
34	X43	0.0670
205	X219	0.0629
91	X101	0.0603
194	X208	0.0590
353	X368	0.0588
152	X164	0.0586

5	X13	0.0546
88	X98	0.0542
361	X376	0.0540
316	X331	0.0529
272	X286	0.0517
337	X352	0.0512
70	X80	0.0502
333	X348	0.0500
352	X367	0.0491
165	X177	0.0476
107	X117	0.0468
167	X179	0.0457
331	X346	0.0453
287	X301	0.0446
52	X61	0.0440
334	X349	0.0429
137	X148	0.0429
291	X306	0.0418
66	X76	0.0416
45	X54	0.0416
125	X136	0.0416
218	X232	0.0412
20	X29	0.0412
265	X279	0.0412
249	X263	0.0412
110	X120	0.0405
43	X52	0.0405
117	X128	0.0399
119	X130	0.0399
270	X284	0.0394
135	X146	0.0392
127	X138	0.0392
150	X162	0.0392
129	X140	0.0388
99	X109	0.0388
313	X328	0.0385
215	X229	0.0383
250	X264	0.0379
115	X126	0.0375
214	X228	0.0375
262	X276	0.0370
132	X143	0.0368
258	X272	0.0361
65	X75	0.0348
346	X361	0.0328
29	X38	0.0322
154	X166	0.0322

19	X28	0.0315
140	X152	0.0313
212	X226	0.0313
311	X326	0.0313
183	X197	0.0313
344	X359	0.0308
60	X69	0.0290
300	X315	0.0279
57	X66	0.0264
124	X135	0.0264
120	X131	0.0259
101	X111	0.0246
69	X79	0.0246
158	X170	0.0237
318	X333	0.0234
184	X198	0.0225
325	X340	0.0218
39	X48	0.0218
123	X134	0.0218
208	X222	0.0218
136	X147	0.0218
103	X113	0.0218
330	X345	0.0218
327	X342	0.0218
163	X175	0.0218
307	X322	0.0214
47	X56	0.0207
15	X23	0.0202
363	X378	0.0202
63	X73	0.0196
241	X255	0.0191
192	X206	0.0189
358	X373	0.0189
173	X185	0.0184
162	X174	0.0170
164	X176	0.0168
72	X82	0.0168
189	X203	0.0166
217	X231	0.0159
273	X287	0.0157
197	X211	0.0147
98	X108	0.0145
130	X141	0.0141
356	X371	0.0141
147	X159	0.0134
48	X57	0.0131
2	X10	0.0131

290	X305	0.0131
96	X106	0.0129
38	X47	0.0127
67	X77	0.0124
181	X195	0.0115
54	X63	0.0113
32	X41	0.0113
35	X44	0.0113
288	X302	0.0113
23	X32	0.0110
277	X291	0.0103
182	X196	0.0101
260	X274	0.0099
161	X173	0.0096
293	X308	0.0094
364	X379	0.0094
308	X323	0.0092
253	X267	0.0089
278	X292	0.0089
89	X99	0.0085
329	X344	0.0085
326	X341	0.0080
207	X221	0.0080
365	X380	0.0080
10	X18	0.0078
235	X249	0.0075
9	X17	0.0075
366	X382	0.0075
302	X317	0.0075
84	X94	0.0073
228	X242	0.0073
80	X90	0.0073
203	X217	0.0073
294	X309	0.0071
111	X122	0.0071
229	X243	0.0071
78	X88	0.0071
305	X320	0.0071
323	X338	0.0068
92	X102	0.0068
44	X53	0.0068
225	X239	0.0068
200	X214	0.0068
355	X370	0.0066
186	X200	0.0066
157	X169	0.0066
223	X237	0.0066

53	X62	0.0059
202	X216	0.0059
160	X172	0.0059
310	X325	0.0057
68	X78	0.0057
198	X212	0.0054
25	X34	0.0054
46	X55	0.0052
240	X254	0.0052
216	X230	0.0052
17	X26	0.0050
284	X298	0.0045
285	X299	0.0045
153	X165	0.0045
21	X30	0.0045
27	X36	0.0045
87	X97	0.0043
297	X312	0.0043
171	X183	0.0040
268	X282	0.0040
320	X335	0.0036
213	X227	0.0031
114	X125	0.0031
226	X240	0.0028
350	X365	0.0028
349	X364	0.0028
185	X199	0.0028
102	X112	0.0028
8	X16	0.0026
13	X21	0.0026
112	X123	0.0026
295	X310	0.0026
267	X281	0.0026
244	X258	0.0024
179	X192	0.0024
95	X105	0.0024
292	X307	0.0021
56	X65	0.0021
338	X353	0.0021
257	X271	0.0021
58	X67	0.0019
199	X213	0.0019
16	X24	0.0019
94	X104	0.0019
81	X91	0.0017
367	X383	0.0017
369	X385	0.0014

234	X248	0.0014
51	X60	0.0014
239	X253	0.0014
263	X277	0.0014
134	X145	0.0014
76	X86	0.0014
172	X184	0.0014
252	X266	0.0014
248	X262	0.0014
73	X83	0.0012
351	X366	0.0012
342	X357	0.0012
148	X160	0.0012
82	X92	0.0009
155	X167	0.0009
77	X87	0.0009
100	X110	0.0009
238	X252	0.0007
50	X59	0.0007
141	X153	0.0007
64	X74	0.0007
303	X318	0.0007
79	X89	0.0007
231	X245	0.0007
317	X332	0.0007
31	X40	0.0007
264	X278	0.0005
113	X124	0.0005
304	X319	0.0005
255	X269	0.0005
7	X15	0.0005
354	X369	0.0005
222	X236	0.0005
368	X384	0.0005
357	X372	0.0005
191	X205	0.0002
190	X204	0.0002
33	X42	0.0002
324	X339	0.0002
245	X259	0.0002
196	X210	0.0002
246	X260	0.0002
85	X95	0.0002
281	X295	0.0002
282	X296	0.0002
256	X270	0.0002
177	X190	0.0002

```
30    X39         0.0002
24    X33         0.0002
274   X288        0.0002
266   X280        0.0002
193   X207        0.0002
243   X257        0.0002
276   X290        0.0000
315   X330        0.0000
3     X11         0.0000
221   X235        0.0000
279   X293        0.0000
219   X233        0.0000
275   X289        0.0000
83    X93         0.0000
332   X347        0.0000
97    X107        0.0000
254   X268        0.0000
283   X297        0.0000
```

[10]:
```
# We will remove the varialbes with variance 0 and low variance (less than 0.2)
# we will also remove id column
```

[11]:
```
variance=var.loc[var['values'] < 0.2,'id']
train1=train.drop(variance,axis=1)
train1=train1.drop('ID',axis=1)
train1.head()
```

[11]:

	y	X0	X1	X2	X3	X4	X5	X6	X8	X14	X27	X46	X51	X58	X64	X85	X100	\
0	130.8100	k	v	at	a	d	u	j	o	0	0	1	0	1	0	1	0	
1	88.5300	k	t	av	e	d	y	l	o	0	1	0	1	0	0	1	1	
2	76.2600	az	w	n	c	d	x	j	x	0	1	1	1	1	0	1	0	
3	80.6200	az	t	n	f	d	x	l	e	0	1	1	0	0	0	0	0	
4	78.0200	az	v	n	f	d	h	d	n	0	1	1	1	0	0	0	0	

	X115	X118	X119	X127	X132	X137	X156	X157	X163	X171	X178	X186	\
0	0	1	1	0	0	1	1	0	0	0	0	0	
1	0	1	1	1	1	0	1	0	0	0	1	0	
2	0	0	0	0	1	1	0	1	1	0	0	0	
3	0	0	0	0	1	0	0	1	0	0	0	0	
4	0	0	0	0	1	0	0	1	0	0	0	0	

	X187	X191	X194	X218	X220	X223	X224	X246	X250	X251	X261	X273	\
0	1	0	1	0	1	0	0	0	0	0	0	1	
1	1	0	1	0	0	0	0	0	1	0	0	1	
2	0	0	1	1	1	1	1	1	1	0	0	1	
3	0	0	1	1	1	1	0	1	1	0	0	1	
4	0	0	1	1	1	1	0	1	1	0	0	1	

```
   X311  X313  X314  X324  X329  X334  X337  X350  X351  X355  X358  X362  \
0     0     0     0     1     1     1     0     0     0     0     0     0
1     1     0     0     0     1     0     1     0     0     0     0     0
2     0     0     0     1     0     1     0     1     0     0     1     0
3     0     0     0     0     0     0     0     1     0     0     1     0
4     0     0     0     0     0     1     0     1     0     0     1     0

   X375  X377
0     0     1
1     1     0
2     0     0
3     0     0
4     0     0
```

[ ]:

[12]:
```
#checking shape of the dataset after remving low var columns

train1.shape
```

[12]:
```
(4209, 55)
```

[13]:
```
#finding corr betwenn idvs

c=train1.corr().abs()
s=c.unstack()
```

[14]:
```
s=pd.DataFrame(s)
s.reset_index(inplace=True)
s.columns.values[2]='corr'
s["flag"] = np.where(s["level_0"] == s["level_1"],"same","not same")
s.head()
```

[14]:
```
  level_0 level_1   corr      flag
0       y       y 1.0000      same
1       y     X14 0.1936  not same
2       y     X27 0.0535  not same
3       y     X46 0.1360  not same
4       y     X51 0.2300  not same
```

[ ]:

[15]:
```
# Remove the variables with correlation more than .9
name = s.loc[(s["corr"] > .9) & (s["flag"] != "same") ,"level_1"]
corr_col = name.unique()
train2=train1.drop(corr_col,1)
```

[16]:
```
train2.head()
```

[16]:

	y	X0	X1	X2	X3	X4	X5	X6	X8	X27	X46	X51	X64	X85	X100	X115	\
0	130.8100	k	v	at	a	d	u	j	o	0	1	0	0	1	0	0	
1	88.5300	k	t	av	e	d	y	l	o	1	0	1	0	1	1	0	
2	76.2600	az	w	n	c	d	x	j	x	1	1	1	0	1	0	0	
3	80.6200	az	t	n	f	d	x	l	e	1	1	0	0	0	0	0	
4	78.0200	az	v	n	f	d	h	d	n	1	1	1	0	0	0	0	

	X127	X132	X163	X171	X191	X218	X220	X223	X224	X273	X313	X329	\
0	0	0	0	0	0	0	1	0	0	1	0	1	
1	1	1	0	0	0	0	0	0	0	1	0	1	
2	0	1	1	0	0	1	1	1	1	1	0	0	
3	0	1	0	0	0	1	1	1	0	1	0	0	
4	0	1	0	0	0	1	1	1	0	1	0	0	

	X350	X351	X355	X375	X377
0	0	0	0	0	1
1	0	0	0	1	0
2	1	0	0	0	0
3	1	0	0	0	0
4	1	0	0	0	0

[ ]:
```
```

[17]:
```
#checking shape after removing high corr between idvs to avoid multicolinearity
train2.shape
```

[17]: (4209, 33)

[18]:
```
train2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4209 entries, 0 to 4208
Data columns (total 33 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   y       4209 non-null   float64
 1   X0      4209 non-null   object
 2   X1      4209 non-null   object
 3   X2      4209 non-null   object
 4   X3      4209 non-null   object
 5   X4      4209 non-null   object
 6   X5      4209 non-null   object
 7   X6      4209 non-null   object
 8   X8      4209 non-null   object
 9   X27     4209 non-null   int64
```

```
 10  X46     4209 non-null   int64
 11  X51     4209 non-null   int64
 12  X64     4209 non-null   int64
 13  X85     4209 non-null   int64
 14  X100    4209 non-null   int64
 15  X115    4209 non-null   int64
 16  X127    4209 non-null   int64
 17  X132    4209 non-null   int64
 18  X163    4209 non-null   int64
 19  X171    4209 non-null   int64
 20  X191    4209 non-null   int64
 21  X218    4209 non-null   int64
 22  X220    4209 non-null   int64
 23  X223    4209 non-null   int64
 24  X224    4209 non-null   int64
 25  X273    4209 non-null   int64
 26  X313    4209 non-null   int64
 27  X329    4209 non-null   int64
 28  X350    4209 non-null   int64
 29  X351    4209 non-null   int64
 30  X355    4209 non-null   int64
 31  X375    4209 non-null   int64
 32  X377    4209 non-null   int64
dtypes: float64(1), int64(24), object(8)
memory usage: 1.1+ MB
```

[ ]:

[19]:
```
#final columns
train2.columns
```

[19]:
```
Index(['y', 'X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X8', 'X27', 'X46',
       'X51', 'X64', 'X85', 'X100', 'X115', 'X127', 'X132', 'X163', 'X171',
       'X191', 'X218', 'X220', 'X223', 'X224', 'X273', 'X313', 'X329', 'X350',
       'X351', 'X355', 'X375', 'X377'],
      dtype='object')
```

[ ]:

[20]:
```
#checking unique values
train2.nunique()
```

[20]:
```
y       2545
X0        47
X1        27
X2        44
X3         7
```

```
X4          4
X5         29
X6         12
X8         25
X27         2
X46         2
X51         2
X64         2
X85         2
X100        2
X115        2
X127        2
X132        2
X163        2
X171        2
X191        2
X218        2
X220        2
X223        2
X224        2
X273        2
X313        2
X329        2
X350        2
X351        2
X355        2
X375        2
X377        2
dtype: int64
```

[21]:
```
train2.describe()
```

[21]:

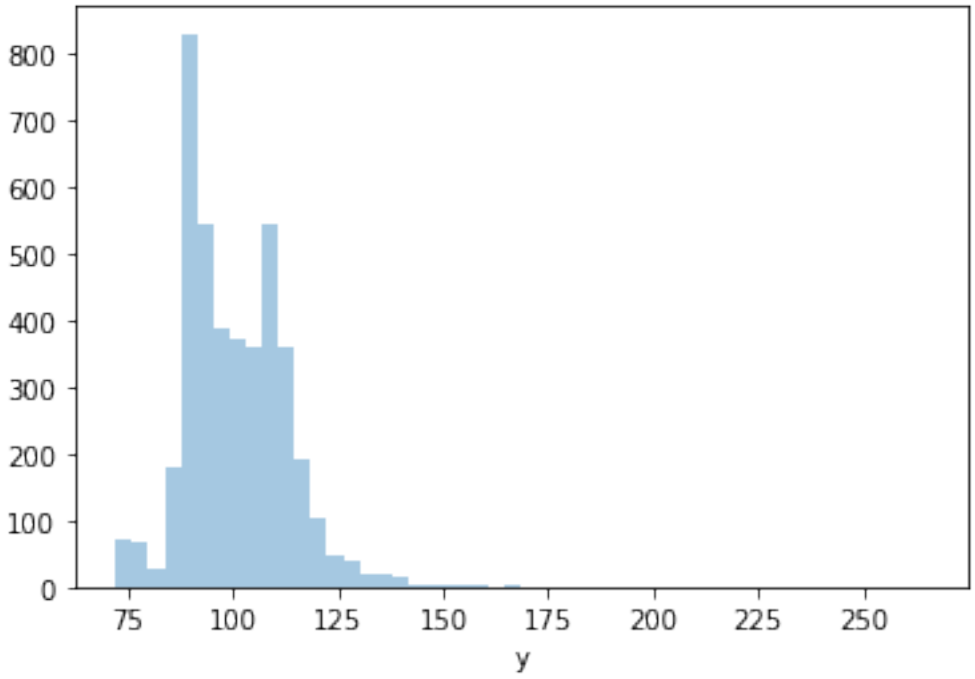
	y	X27	X46	X51	X64	X85	\
count	4,209.0000	4,209.0000	4,209.0000	4,209.0000	4,209.0000	4,209.0000	
mean	100.6693	0.6826	0.5973	0.7218	0.3751	0.4082	
std	12.6794	0.4655	0.4905	0.4482	0.4842	0.4916	
min	72.1100	0.0000	0.0000	0.0000	0.0000	0.0000	
25%	90.8200	0.0000	0.0000	0.0000	0.0000	0.0000	
50%	99.1500	1.0000	1.0000	1.0000	0.0000	0.0000	
75%	109.0100	1.0000	1.0000	1.0000	1.0000	1.0000	
max	265.3200	1.0000	1.0000	1.0000	1.0000	1.0000	

	X100	X115	X127	X132	X163	X171	\
count	4,209.0000	4,209.0000	4,209.0000	4,209.0000	4,209.0000	4,209.0000	
mean	0.6902	0.2856	0.4951	0.6885	0.3034	0.6574	
std	0.4625	0.4517	0.5000	0.4632	0.4598	0.4746	
min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

```
25%       0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
50%       1.0000     0.0000     0.0000     1.0000     0.0000     1.0000
75%       1.0000     1.0000     1.0000     1.0000     1.0000     1.0000
max       1.0000     1.0000     1.0000     1.0000     1.0000     1.0000

            X191       X218       X220       X223       X224       X273  \
count 4,209.0000 4,209.0000 4,209.0000 4,209.0000 4,209.0000 4,209.0000
mean      0.4709     0.3124     0.5612     0.5552     0.3174     0.7201
std       0.4992     0.4635     0.4963     0.4970     0.4655     0.4490
min       0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
25%       0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
50%       0.0000     0.0000     1.0000     1.0000     0.0000     1.0000
75%       1.0000     1.0000     1.0000     1.0000     1.0000     1.0000
max       1.0000     1.0000     1.0000     1.0000     1.0000     1.0000

            X313       X329       X350       X351       X355       X375  \
count 4,209.0000 4,209.0000 4,209.0000 4,209.0000 4,209.0000 4,209.0000
mean      0.3010     0.4353     0.3388     0.2972     0.3804     0.3188
std       0.4588     0.4958     0.4734     0.4571     0.4855     0.4661
min       0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
25%       0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
50%       0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
75%       1.0000     1.0000     1.0000     1.0000     1.0000     1.0000
max       1.0000     1.0000     1.0000     1.0000     1.0000     1.0000

            X377
count 4,209.0000
mean      0.3148
std       0.4645
min       0.0000
25%       0.0000
50%       0.0000
75%       1.0000
max       1.0000
```

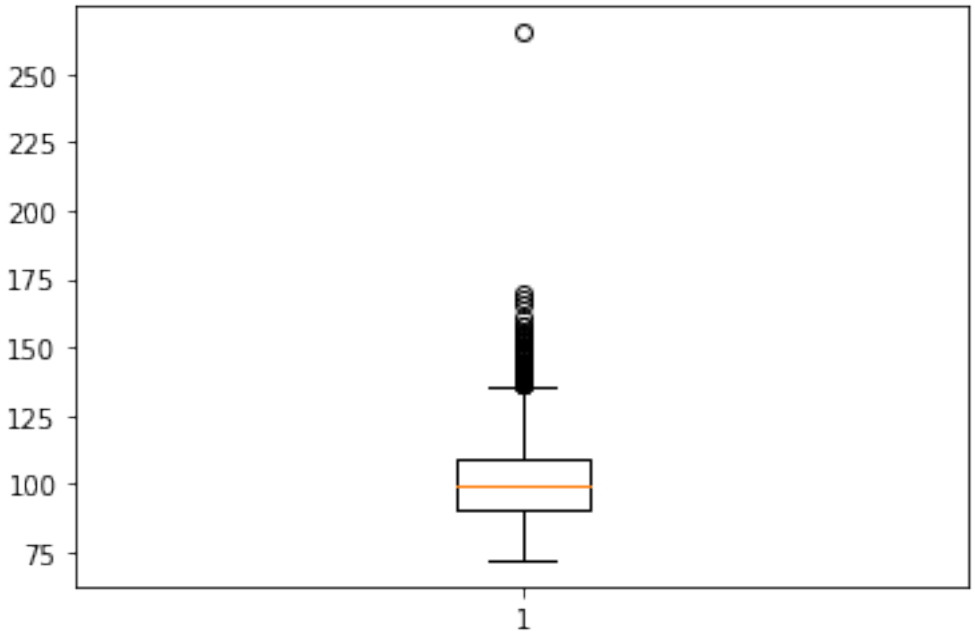
[22]:
```
# Distribution plot for output variable.
sns.distplot(train2['y'],kde = False)
plt.title('Target value distribution')
```

[22]:
```
<AxesSubplot:xlabel='y'>
```



[23]:
```
# Observation: We can see it clearly that datapoints has been distributed␣
↪mostly between around 65 to 140
```

[24]:
```
# Box plot for output variable.
plt.boxplot(train2['y'])
plt.show()
```



[25]:
```
# Observation: We can see it clearly that we have outliers at top side above 140
```

[ ]:
```

```

[26]:
```
#lets check train test data

test.head()
```

[26]:

	ID	X0	X1	X2	X3	X4	X5	X6	X8	X10	X11	X12	X13	X14	X15	X16	X17	X18	\
0	1	az	v	n	f	d	t	a	w	0	0	0	0	0	0	0	0	0	
1	2	t	b	ai	a	d	b	g	y	0	0	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	
3	4	az	l	n	f	d	z	l	n	0	0	0	0	0	0	0	0	0	
4	5	w	s	as	c	d	y	i	m	0	0	0	0	1	0	0	0	0	

	X19	X20	X21	X22	X23	X24	X26	X27	X28	X29	X30	X31	X32	X33	X34	\
0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	
1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	
2	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	
3	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	
4	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	

	X35	X36	X37	X38	X39	X40	X41	X42	X43	X44	X45	X46	X47	X48	X49	\
0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	
1	1	0	1	0	0	0	0	0	0	0	1	1	0	0	1	
2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
3	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

	X50	X51	X52	X53	X54	X55	X56	X57	X58	X59	X60	X61	X62	X63	X64	\
0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	
1	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	
2	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	
3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	
4	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	

	X65	X66	X67	X68	X69	X70	X71	X73	X74	X75	X76	X77	X78	X79	X80	\
0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	
1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	
2	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	
3	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	
4	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	

	X81	X82	X83	X84	X85	X86	X87	X88	X89	X90	X91	X92	X93	X94	X95	\
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	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	0
4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

	X96	X97	X98	X99	X100	X101	X102	X103	X104	X105	X106	X107	X108	\
0	1	0	1	0	0	1	0	0	0	0	0	0	0	
1	0	0	1	0	0	1	0	1	0	0	0	0	0	
2	1	0	1	0	1	1	0	1	0	0	0	0	0	
3	1	0	1	0	0	1	0	0	0	0	0	0	0	
4	1	0	1	0	1	1	0	1	0	0	1	0	0	

	X109	X110	X111	X112	X113	X114	X115	X116	X117	X118	X119	X120	\
0	0	0	1	0	0	1	0	0	0	0	0	0	
1	0	0	1	0	0	0	0	1	0	1	1	1	
2	0	0	1	0	0	0	0	0	0	0	0	1	
3	0	0	1	0	0	1	0	0	0	0	0	1	
4	0	0	1	0	0	0	0	0	0	1	1	1	

	X122	X123	X124	X125	X126	X127	X128	X129	X130	X131	X132	X133	\
0	0	0	0	0	0	0	1	0	0	0	1	0	
1	0	0	0	0	0	1	1	1	0	0	0	1	
2	0	0	0	0	0	0	1	0	0	0	1	0	
3	0	0	0	0	0	0	1	0	0	0	1	0	
4	0	0	0	0	0	0	1	0	0	0	1	0	

	X134	X135	X136	X137	X138	X139	X140	X141	X142	X143	X144	X145	\
0	0	0	0	0	0	0	0	0	0	0	1	0	
1	0	0	1	1	1	0	1	0	1	0	0	0	
2	0	0	0	1	0	0	0	0	1	0	1	0	
3	0	0	0	0	0	0	0	0	0	0	1	0	
4	0	0	1	1	0	0	0	0	1	0	1	0	

	X146	X147	X148	X150	X151	X152	X153	X154	X155	X156	X157	X158	\
0	0	0	1	1	0	0	0	0	0	0	1	1	
1	1	0	0	0	0	0	0	0	0	1	0	0	
2	0	0	0	1	0	0	0	0	0	0	1	0	
3	0	0	1	1	0	0	0	0	0	0	1	1	
4	0	0	0	1	1	0	0	0	0	1	0	0	

	X159	X160	X161	X162	X163	X164	X165	X166	X167	X168	X169	X170	\
0	0	0	0	1	0	0	0	1	0	0	0	0	
1	0	0	1	0	1	1	0	0	0	0	0	0	
2	1	0	0	1	0	0	0	1	0	0	0	0	
3	0	0	0	1	0	0	0	1	0	0	0	0	
4	0	0	0	0	1	0	0	0	0	0	0	0	

	X171	X172	X173	X174	X175	X176	X177	X178	X179	X180	X181	X182	\
0	0	0	0	0	0	0	0	0	1	0	0	0	
1	1	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	1	0	0	0	
4	1	0	0	0	0	0	0	0	0	0	0	0	

	X183	X184	X185	X186	X187	X189	X190	X191	X192	X194	X195	X196	\
0	0	0	1	0	0	0	0	0	0	1	0	0	
1	0	0	0	1	0	1	0	0	0	0	0	0	
2	0	0	1	0	0	0	0	1	0	1	0	0	
3	0	0	1	0	0	0	0	0	0	1	0	0	
4	0	0	0	0	1	1	0	1	0	1	0	0	

	X197	X198	X199	X200	X201	X202	X203	X204	X205	X206	X207	X208	\
0	0	0	0	0	0	0	0	1	0	0	0	0	
1	0	0	0	0	0	1	0	1	0	0	0	1	
2	0	0	0	0	0	0	0	1	0	0	0	0	
3	0	0	0	0	0	0	0	0	1	0	0	0	
4	0	0	0	0	0	0	0	0	1	0	0	0	

	X209	X210	X211	X212	X213	X214	X215	X216	X217	X218	X219	X220	\
0	1	0	0	0	0	0	0	0	0	1	0	1	
1	0	0	0	0	0	0	1	0	0	1	0	0	
2	1	0	0	0	0	0	0	0	0	1	0	0	
3	1	0	0	0	0	0	0	0	0	1	0	1	
4	1	0	0	0	0	0	0	0	0	0	0	1	

	X221	X222	X223	X224	X225	X226	X227	X228	X229	X230	X231	X232	\
0	0	0	1	0	0	0	0	0	0	0	0	1	
1	0	0	0	0	0	0	0	0	1	0	0	0	
2	0	0	1	1	0	0	0	0	0	0	0	1	
3	0	0	1	0	0	0	0	0	1	0	0	1	
4	0	0	1	1	0	0	0	0	1	0	0	0	

	X233	X234	X235	X236	X237	X238	X239	X240	X241	X242	X243	X244	\
0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	1	0	0	0	1	0	0	0	0	0	0	
2	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	
4	0	0	0	0	0	1	0	0	0	0	0	0	

	X245	X246	X247	X248	X249	X250	X251	X252	X253	X254	X255	X256	\
0	0	1	0	0	0	1	0	0	0	0	0	1	
1	0	0	1	0	0	1	0	0	0	0	0	0	
2	0	0	0	0	0	0	1	0	0	0	0	0	
3	0	1	0	0	0	1	0	1	0	0	0	1	

```
4     0     1     0     0     0     0     1     0     0     0     0     0

    X257  X258  X259  X260  X261  X262  X263  X264  X265  X266  X267  X268  \
0      0     0     0     0     0     0     0     0     0     0     0     0
1      0     0     0     0     0     0     1     0     1     0     0     0
2      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
4      0     0     0     0     1     0     1     0     1     0     0     0

    X269  X270  X271  X272  X273  X274  X275  X276  X277  X278  X279  X280  \
0      0     0     0     1     1     0     0     1     0     0     1     0
1      0     0     0     0     1     0     1     0     0     0     0     0
2      0     0     0     0     0     1     0     1     0     0     1     0
3      0     0     0     1     1     0     0     1     0     0     1     0
4      0     0     0     0     1     0     1     0     0     0     0     0

    X281  X282  X283  X284  X285  X286  X287  X288  X289  X290  X291  X292  \
0      0     0     0     0     0     1     0     0     0     0     0     0
1      0     0     0     0     0     0     0     0     0     0     0     0
2      0     0     0     0     1     0     0     0     0     0     0     0
3      0     0     0     0     0     1     0     0     0     0     0     1
4      0     0     0     0     1     0     0     0     0     0     0     0

    X293  X294  X295  X296  X297  X298  X299  X300  X301  X302  X304  X305  \
0      0     0     0     0     0     0     0     0     0     0     1     0
1      0     1     0     0     0     0     0     0     0     0     1     0
2      0     0     0     0     0     0     0     0     0     0     1     0
3      0     0     0     0     0     0     0     0     0     0     1     0
4      0     0     0     0     0     0     0     1     0     0     1     0

    X306  X307  X308  X309  X310  X311  X312  X313  X314  X315  X316  X317  \
0      0     0     0     0     0     0     0     0     0     0     0     0
1      0     0     0     0     0     1     0     0     0     0     1     0
2      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
4      0     0     0     0     0     1     0     0     1     0     0     0

    X318  X319  X320  X321  X322  X323  X324  X325  X326  X327  X328  X329  \
0      0     0     0     0     0     0     0     0     0     0     1     0
1      0     0     0     0     0     0     1     0     0     1     0     0
2      0     0     0     0     0     0     1     0     0     0     1     1
3      0     0     0     0     0     0     0     0     0     0     1     0
4      0     0     0     0     0     0     1     0     0     0     0     1

    X330  X331  X332  X333  X334  X335  X336  X337  X338  X339  X340  X341  \
0      0     0     0     0     1     0     0     0     0     0     0     0
1      0     0     0     0     1     0     0     0     0     0     0     0
```

```
2     0     0     0     0     1     0     0     0     0     0     0     0
3     0     0     0     0     1     0     0     0     0     0     0     0
4     0     0     0     0     0     0     1     1     0     0     0     0

   X342  X343  X344  X345  X346  X347  X348  X349  X350  X351  X352  X353  \
0     0     0     0     0     0     0     1     0     1     0     0     0
1     0     0     0     0     0     0     1     0     0     0     0     0
2     0     1     0     0     0     0     1     0     1     0     0     0
3     0     0     0     0     0     0     1     0     1     0     0     0
4     0     0     0     0     0     0     1     0     1     1     0     0

   X354  X355  X356  X357  X358  X359  X360  X361  X362  X363  X364  X365  \
0     0     0     0     0     1     0     0     1     0     1     0     0
1     0     0     1     0     0     0     0     1     1     0     0     0
2     0     0     0     0     0     0     0     1     0     1     0     0
3     0     0     0     0     1     0     0     1     0     1     0     0
4     0     0     0     0     1     0     0     1     0     1     0     0

   X366  X367  X368  X369  X370  X371  X372  X373  X374  X375  X376  X377  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     1     0     0     0     0     0     0     0     0     1
2     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
4     0     0     0     0     0     0     0     0     0     1     0     0

   X378  X379  X380  X382  X383  X384  X385
0     1     0     0     0     0     0     0
1     0     0     0     0     0     0     0
2     1     0     0     0     0     0     0
3     1     0     0     0     0     0     0
4     0     0     0     0     0     0     0
```

[ ]:

[27]:
```
#shape of the test data

test.shape
```

[27]: (4209, 377)

[28]:
```
#select columns with variance less than 0.2 which we preprocessed in train

test1=test[['X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X8', 'X27', 'X46',
       'X51', 'X64', 'X85', 'X100', 'X115', 'X127', 'X132', 'X163', 'X171',
       'X191', 'X218', 'X220', 'X223', 'X224', 'X273', 'X313', 'X329', 'X350',
       'X351', 'X355', 'X375', 'X377']]
```

[29]:
```
test1.head()
```

[29]:
```
   X0 X1  X2 X3 X4 X5 X6 X8  X27  X46  X51  X64  X85  X100  X115  X127  X132  \
0  az  v   n  f  d  t  a  w    1    1    0    0    0     0     0     0     1
1   t  b  ai  a  d  b  g  y    1    1    1    0    0     0     0     1     0
2  az  v  as  f  d  a  j  j    1    0    1    0    1     1     0     0     1
3  az  l   n  f  d  z  l  n    1    1    0    0    0     0     0     0     1
4   w  s  as  c  d  y  i  m    1    0    1    1    1     1     0     0     1

   X163  X171  X191  X218  X220  X223  X224  X273  X313  X329  X350  X351  \
0     0     0     0     1     1     1     0     1     0     0     1     0
1     1     1     0     1     0     0     0     1     0     0     0     0
2     0     0     1     1     0     1     1     0     0     1     1     0
3     0     0     0     1     1     1     0     1     0     0     1     0
4     1     1     1     0     1     1     1     1     0     1     1     1

   X355  X375  X377
0     0     0     0
1     0     0     1
2     0     0     0
3     0     0     0
4     0     1     0
```

[30]:
```
#Splitting data into numerical and categorical for train and test
```

[31]:
```
#splitting train data

char_train=train2.select_dtypes(exclude='number')
num_train=train2.select_dtypes(include='number')
```

[32]:
```
char_train.nunique()
```

[32]:
```
X0    47
X1    27
X2    44
X3     7
X4     4
X5    29
X6    12
X8    25
dtype: int64
```

[33]:
```
#splitting test data

char_test=test1.select_dtypes(exclude='number')
num_test=test1.select_dtypes(include='number')
```

[34]:
```
char_test.nunique()
```

[34]:
```
X0    49
X1    27
X2    45
X3     7
X4     4
X5    32
X6    12
X8    25
dtype: int64
```

[35]:
```
#observation: we can clearly both data sets have different unique values in␣
 ↪categorical column
#this will impact our model while training the model
#so, we will merge both categorical column to get same number of columns while␣
 ↪creating dummy var
```

[36]:
```
#concatinating both train and test categorical columns
df_new=pd.concat([char_train,char_test],axis=0)
```

[37]:
```
df_new.head()
```

[37]:

	X0	X1	X2	X3	X4	X5	X6	X8
0	k	v	at	a	d	u	j	o
1	k	t	av	e	d	y	l	o
2	az	w	n	c	d	x	j	x
3	az	t	n	f	d	x	l	e
4	az	v	n	f	d	h	d	n

[38]:
```
df_new.shape
```

[38]:
```
(8418, 8)
```

[39]:
```
#creating dummy var for both datasets

dum_var=pd.get_dummies(df_new,drop_first=True)
```

[40]:
```
dum_var.head()
```

[40]:

	X0_aa	X0_ab	X0_ac	X0_ad	X0_ae	X0_af	X0_ag	X0_ai	X0_aj	X0_ak	\
0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	

```
   X0_al  X0_am  X0_an  X0_ao  X0_ap  X0_aq  X0_as  X0_at  X0_au  X0_av  \
0      0      0      0      0      0      0      0      0      0      0
1      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0

   X0_aw  X0_ax  X0_ay  X0_az  X0_b  X0_ba  X0_bb  X0_bc  X0_c  X0_d  X0_e  \
0      0      0      0      0     0      0      0      0     0     0     0
1      0      0      0      0     0      0      0      0     0     0     0
2      0      0      0      1     0      0      0      0     0     0     0
3      0      0      0      1     0      0      0      0     0     0     0
4      0      0      0      1     0      0      0      0     0     0     0

   X0_f  X0_g  X0_h  X0_i  X0_j  X0_k  X0_l  X0_m  X0_n  X0_o  X0_p  X0_q  \
0     0     0     0     0     0     1     0     0     0     0     0     0
1     0     0     0     0     0     1     0     0     0     0     0     0
2     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
4     0     0     0     0     0     0     0     0     0     0     0     0

   X0_r  X0_s  X0_t  X0_u  X0_v  X0_w  X0_x  X0_y  X0_z  X1_aa  X1_ab  X1_b  \
0     0     0     0     0     0     0     0     0     0      0      0     0
1     0     0     0     0     0     0     0     0     0      0      0     0
2     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
4     0     0     0     0     0     0     0     0     0      0      0     0

   X1_c  X1_d  X1_e  X1_f  X1_g  X1_h  X1_i  X1_j  X1_k  X1_l  X1_m  X1_n  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     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
4     0     0     0     0     0     0     0     0     0     0     0     0

   X1_o  X1_p  X1_q  X1_r  X1_s  X1_t  X1_u  X1_v  X1_w  X1_y  X1_z  X2_aa  \
0     0     0     0     0     0     0     0     1     0     0     0      0
1     0     0     0     0     0     1     0     0     0     0     0      0
2     0     0     0     0     0     0     0     0     1     0     0      0
3     0     0     0     0     0     1     0     0     0     0     0      0
4     0     0     0     0     0     0     0     1     0     0     0      0

   X2_ab  X2_ac  X2_ad  X2_ae  X2_af  X2_ag  X2_ah  X2_ai  X2_aj  X2_ak  \
0      0      0      0      0      0      0      0      0      0      0
1      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0
```

```
4      0      0      0      0      0      0      0      0      0      0

   X2_al  X2_am  X2_an  X2_ao  X2_ap  X2_aq  X2_ar  X2_as  X2_at  X2_au  \
0      0      0      0      0      0      0      0      0      1      0
1      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0

   X2_av  X2_aw  X2_ax  X2_ay  X2_b  X2_c  X2_d  X2_e  X2_f  X2_g  X2_h  X2_i  \
0      0      0      0      0     0     0     0     0     0     0     0     0
1      1      0      0      0     0     0     0     0     0     0     0     0
2      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
4      0      0      0      0     0     0     0     0     0     0     0     0

   X2_j  X2_k  X2_l  X2_m  X2_n  X2_o  X2_p  X2_q  X2_r  X2_s  X2_t  X2_u  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     1     0     0     0     0     0     0     0
3     0     0     0     0     1     0     0     0     0     0     0     0
4     0     0     0     0     1     0     0     0     0     0     0     0

   X2_w  X2_x  X2_y  X2_z  X3_b  X3_c  X3_d  X3_e  X3_f  X3_g  X4_b  X4_c  \
0     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
2     0     0     0     0     0     1     0     0     0     0     0     0
3     0     0     0     0     0     0     0     0     1     0     0     0
4     0     0     0     0     0     0     0     0     1     0     0     0

   X4_d  X5_aa  X5_ab  X5_ac  X5_ad  X5_ae  X5_af  X5_ag  X5_ah  X5_b  X5_c  \
0     1      0      0      0      0      0      0      0      0     0     0
1     1      0      0      0      0      0      0      0      0     0     0
2     1      0      0      0      0      0      0      0      0     0     0
3     1      0      0      0      0      0      0      0      0     0     0
4     1      0      0      0      0      0      0      0      0     0     0

   X5_d  X5_f  X5_g  X5_h  X5_i  X5_j  X5_k  X5_l  X5_m  X5_n  X5_o  X5_p  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     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
4     0     0     0     1     0     0     0     0     0     0     0     0

   X5_q  X5_r  X5_s  X5_t  X5_u  X5_v  X5_w  X5_x  X5_y  X5_z  X6_b  X6_c  \
0     0     0     0     0     1     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     1     0     0     0
```

```
2     0     0     0     0     0     0     0     1     0     0     0     0
3     0     0     0     0     0     0     0     1     0     0     0     0
4     0     0     0     0     0     0     0     0     0     0     0     0

   X6_d  X6_e  X6_f  X6_g  X6_h  X6_i  X6_j  X6_k  X6_l  X8_b  X8_c  X8_d  \
0     0     0     0     0     0     0     1     0     0     0     0     0
1     0     0     0     0     0     0     0     0     1     0     0     0
2     0     0     0     0     0     0     1     0     0     0     0     0
3     0     0     0     0     0     0     0     0     1     0     0     0
4     1     0     0     0     0     0     0     0     0     0     0     0

   X8_e  X8_f  X8_g  X8_h  X8_i  X8_j  X8_k  X8_l  X8_m  X8_n  X8_o  X8_p  \
0     0     0     0     0     0     0     0     0     0     0     1     0
1     0     0     0     0     0     0     0     0     0     0     1     0
2     0     0     0     0     0     0     0     0     0     0     0     0
3     1     0     0     0     0     0     0     0     0     0     0     0
4     0     0     0     0     0     0     0     0     0     1     0     0

   X8_q  X8_r  X8_s  X8_t  X8_u  X8_v  X8_w  X8_x  X8_y
0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     1     0
3     0     0     0     0     0     0     0     0     0
4     0     0     0     0     0     0     0     0     0
```

[ ]:

[41]:
```
#splitting both data

dum_train=dum_var.iloc[:char_train.shape[0],:]
dum_test=dum_var.iloc[char_train.shape[0]:,:]
```

[42]:
```
dum_train.shape,dum_test.shape
```

[42]:
```
((4209, 203), (4209, 203))
```

[43]:
```
#concat cat and numeric col for train data

train_final=pd.concat([num_train,dum_train],axis=1)
train_final.head()
```

[43]:
```
         y  X27  X46  X51  X64  X85  X100  X115  X127  X132  X163  X171  X191  \
0 130.8100    0    1    0    0    1     0     0     0     0     0     0     0
1  88.5300    1    0    1    0    1     1     0     1     1     0     0     0
2  76.2600    1    1    1    0    1     0     0     0     1     1     0     0
3  80.6200    1    1    0    0    0     0     0     0     1     0     0     0
4  78.0200    1    1    1    0    0     0     0     0     1     0     0     0
```

```
   X218  X220  X223  X224  X273  X313  X329  X350  X351  X355  X375  X377  \
0     0     1     0     0     1     0     1     0     0     0     0     1
1     0     0     0     0     1     0     1     0     0     0     1     0
2     1     1     1     1     1     0     0     1     0     0     0     0
3     1     1     1     0     1     0     0     1     0     0     0     0
4     1     1     1     0     1     0     0     1     0     0     0     0

   X0_aa  X0_ab  X0_ac  X0_ad  X0_ae  X0_af  X0_ag  X0_ai  X0_aj  X0_ak  \
0      0      0      0      0      0      0      0      0      0      0
1      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0

   X0_al  X0_am  X0_an  X0_ao  X0_ap  X0_aq  X0_as  X0_at  X0_au  X0_av  \
0      0      0      0      0      0      0      0      0      0      0
1      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0

   X0_aw  X0_ax  X0_ay  X0_az  X0_b  X0_ba  X0_bb  X0_bc  X0_c  X0_d  X0_e  \
0      0      0      0      0     0      0      0      0     0     0     0
1      0      0      0      0     0      0      0      0     0     0     0
2      0      0      0      1     0      0      0      0     0     0     0
3      0      0      0      1     0      0      0      0     0     0     0
4      0      0      0      1     0      0      0      0     0     0     0

   X0_f  X0_g  X0_h  X0_i  X0_j  X0_k  X0_l  X0_m  X0_n  X0_o  X0_p  X0_q  \
0     0     0     0     0     0     1     0     0     0     0     0     0
1     0     0     0     0     0     1     0     0     0     0     0     0
2     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
4     0     0     0     0     0     0     0     0     0     0     0     0

   X0_r  X0_s  X0_t  X0_u  X0_v  X0_w  X0_x  X0_y  X0_z  X1_aa  X1_ab  X1_b  \
0     0     0     0     0     0     0     0     0     0      0      0     0
1     0     0     0     0     0     0     0     0     0      0      0     0
2     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
4     0     0     0     0     0     0     0     0     0      0      0     0

   X1_c  X1_d  X1_e  X1_f  X1_g  X1_h  X1_i  X1_j  X1_k  X1_l  X1_m  X1_n  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     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
4     0     0     0     0     0     0     0     0     0     0     0     0

   X1_o  X1_p  X1_q  X1_r  X1_s  X1_t  X1_u  X1_v  X1_w  X1_y  X1_z  X2_aa  \
0     0     0     0     0     0     0     0     1     0     0     0      0
1     0     0     0     0     0     1     0     0     0     0     0      0
2     0     0     0     0     0     0     0     0     1     0     0      0
3     0     0     0     0     0     1     0     0     0     0     0      0
4     0     0     0     0     0     0     0     1     0     0     0      0

   X2_ab  X2_ac  X2_ad  X2_ae  X2_af  X2_ag  X2_ah  X2_ai  X2_aj  X2_ak  \
0      0      0      0      0      0      0      0      0      0      0
1      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0

   X2_al  X2_am  X2_an  X2_ao  X2_ap  X2_aq  X2_ar  X2_as  X2_at  X2_au  \
0      0      0      0      0      0      0      0      0      1      0
1      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0

   X2_av  X2_aw  X2_ax  X2_ay  X2_b  X2_c  X2_d  X2_e  X2_f  X2_g  X2_h  X2_i  \
0      0      0      0      0     0     0     0     0     0     0     0     0
1      1      0      0      0     0     0     0     0     0     0     0     0
2      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
4      0      0      0      0     0     0     0     0     0     0     0     0

   X2_j  X2_k  X2_l  X2_m  X2_n  X2_o  X2_p  X2_q  X2_r  X2_s  X2_t  X2_u  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     1     0     0     0     0     0     0     0
3     0     0     0     0     1     0     0     0     0     0     0     0
4     0     0     0     0     1     0     0     0     0     0     0     0

   X2_w  X2_x  X2_y  X2_z  X3_b  X3_c  X3_d  X3_e  X3_f  X3_g  X4_b  X4_c  \
0     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
2     0     0     0     0     0     1     0     0     0     0     0     0
3     0     0     0     0     0     0     0     0     1     0     0     0
4     0     0     0     0     0     0     0     0     1     0     0     0

   X4_d  X5_aa  X5_ab  X5_ac  X5_ad  X5_ae  X5_af  X5_ag  X5_ah  X5_b  X5_c  \
0     1      0      0      0      0      0      0      0      0     0     0
```

```
1     1      0      0      0      0      0      0      0      0      0      0
2     1      0      0      0      0      0      0      0      0      0      0
3     1      0      0      0      0      0      0      0      0      0      0
4     1      0      0      0      0      0      0      0      0      0      0

   X5_d  X5_f  X5_g  X5_h  X5_i  X5_j  X5_k  X5_l  X5_m  X5_n  X5_o  X5_p  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     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
4     0     0     0     1     0     0     0     0     0     0     0     0

   X5_q  X5_r  X5_s  X5_t  X5_u  X5_v  X5_w  X5_x  X5_y  X5_z  X6_b  X6_c  \
0     0     0     0     0     1     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     1     0     0     0
2     0     0     0     0     0     0     0     1     0     0     0     0
3     0     0     0     0     0     0     0     1     0     0     0     0
4     0     0     0     0     0     0     0     0     0     0     0     0

   X6_d  X6_e  X6_f  X6_g  X6_h  X6_i  X6_j  X6_k  X6_l  X8_b  X8_c  X8_d  \
0     0     0     0     0     0     0     1     0     0     0     0     0
1     0     0     0     0     0     0     0     0     1     0     0     0
2     0     0     0     0     0     0     1     0     0     0     0     0
3     0     0     0     0     0     0     0     0     1     0     0     0
4     1     0     0     0     0     0     0     0     0     0     0     0

   X8_e  X8_f  X8_g  X8_h  X8_i  X8_j  X8_k  X8_l  X8_m  X8_n  X8_o  X8_p  \
0     0     0     0     0     0     0     0     0     0     0     1     0
1     0     0     0     0     0     0     0     0     0     0     1     0
2     0     0     0     0     0     0     0     0     0     0     0     0
3     1     0     0     0     0     0     0     0     0     0     0     0
4     0     0     0     0     0     0     0     0     0     1     0     0

   X8_q  X8_r  X8_s  X8_t  X8_u  X8_v  X8_w  X8_x  X8_y
0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     1     0
3     0     0     0     0     0     0     0     0     0
4     0     0     0     0     0     0     0     0     0
```

[ ]:

[44]:
```
#final shape of the train data
train_final.shape
```

[44]: (4209, 228)

[45]:
```
#concat cat and numeric col for test data

test_final=pd.concat([num_test,dum_test],axis=1)
test_final.head()
```

[45]:

	X27	X46	X51	X64	X85	X100	X115	X127	X132	X163	X171	X191	X218	\
0	1	1	0	0	0	0	0	0	1	0	0	0	1	
1	1	1	1	0	0	0	0	1	0	1	1	0	1	
2	1	0	1	0	1	1	0	0	1	0	0	1	1	
3	1	1	0	0	0	0	0	0	1	0	0	0	1	
4	1	0	1	1	1	1	0	0	1	1	1	1	0	

	X220	X223	X224	X273	X313	X329	X350	X351	X355	X375	X377	X0_aa	\
0	1	1	0	1	0	0	1	0	0	0	0	0	
1	0	0	0	1	0	0	0	0	0	0	1	0	
2	0	1	1	0	0	1	1	0	0	0	0	0	
3	1	1	0	1	0	0	1	0	0	0	0	0	
4	1	1	1	1	0	1	1	1	0	1	0	0	

	X0_ab	X0_ac	X0_ad	X0_ae	X0_af	X0_ag	X0_ai	X0_aj	X0_ak	X0_al	\
0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	

	X0_am	X0_an	X0_ao	X0_ap	X0_aq	X0_as	X0_at	X0_au	X0_av	X0_aw	\
0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	

	X0_ax	X0_ay	X0_az	X0_b	X0_ba	X0_bb	X0_bc	X0_c	X0_d	X0_e	X0_f	\
0	0	0	1	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	1	0	0	0	0	0	0	0	0	
3	0	0	1	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	

	X0_g	X0_h	X0_i	X0_j	X0_k	X0_l	X0_m	X0_n	X0_o	X0_p	X0_q	X0_r	\
0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	
2	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	
4	0	0	0	0	0	0	0	0	0	0	0	0	

```
   X0_s  X0_t  X0_u  X0_v  X0_w  X0_x  X0_y  X0_z  X1_aa  X1_ab  X1_b  X1_c  \
0     0     0     0     0     0     0     0     0      0      0     0     0
1     0     1     0     0     0     0     0     0      0      0     1     0
2     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
4     0     0     0     0     1     0     0     0      0      0     0     0

   X1_d  X1_e  X1_f  X1_g  X1_h  X1_i  X1_j  X1_k  X1_l  X1_m  X1_n  X1_o  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
3     0     0     0     0     0     0     0     0     1     0     0     0
4     0     0     0     0     0     0     0     0     0     0     0     0

   X1_p  X1_q  X1_r  X1_s  X1_t  X1_u  X1_v  X1_w  X1_y  X1_z  X2_aa  X2_ab  \
0     0     0     0     0     0     0     1     0     0     0      0      0
1     0     0     0     0     0     0     0     0     0     0      0      0
2     0     0     0     0     0     0     1     0     0     0      0      0
3     0     0     0     0     0     0     0     0     0     0      0      0
4     0     0     0     1     0     0     0     0     0     0      0      0

   X2_ac  X2_ad  X2_ae  X2_af  X2_ag  X2_ah  X2_ai  X2_aj  X2_ak  X2_al  \
0      0      0      0      0      0      0      0      0      0      0
1      0      0      0      0      0      0      1      0      0      0
2      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0

   X2_am  X2_an  X2_ao  X2_ap  X2_aq  X2_ar  X2_as  X2_at  X2_au  X2_av  \
0      0      0      0      0      0      0      0      0      0      0
1      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      1      0      0      0
3      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      1      0      0      0

   X2_aw  X2_ax  X2_ay  X2_b  X2_c  X2_d  X2_e  X2_f  X2_g  X2_h  X2_i  X2_j  \
0      0      0      0     0     0     0     0     0     0     0     0     0
1      0      0      0     0     0     0     0     0     0     0     0     0
2      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
4      0      0      0     0     0     0     0     0     0     0     0     0

   X2_k  X2_l  X2_m  X2_n  X2_o  X2_p  X2_q  X2_r  X2_s  X2_t  X2_u  X2_w  \
0     0     0     0     1     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
3     0     0     0     1     0     0     0     0     0     0     0     0
```

```
4     0     0     0     0     0     0     0     0     0     0     0     0

   X2_x  X2_y  X2_z  X3_b  X3_c  X3_d  X3_e  X3_f  X3_g  X4_b  X4_c  X4_d  \
0     0     0     0     0     0     0     0     1     0     0     0     1
1     0     0     0     0     0     0     0     0     0     0     0     1
2     0     0     0     0     0     0     0     1     0     0     0     1
3     0     0     0     0     0     0     0     1     0     0     0     1
4     0     0     0     0     1     0     0     0     0     0     0     1

   X5_aa  X5_ab  X5_ac  X5_ad  X5_ae  X5_af  X5_ag  X5_ah  X5_b  X5_c  X5_d  \
0      0      0      0      0      0      0      0      0     0     0     0
1      0      0      0      0      0      0      0      0     1     0     0
2      0      0      0      0      0      0      0      0     0     0     0
3      0      0      0      0      0      0      0      0     0     0     0
4      0      0      0      0      0      0      0      0     0     0     0

   X5_f  X5_g  X5_h  X5_i  X5_j  X5_k  X5_l  X5_m  X5_n  X5_o  X5_p  X5_q  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     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
4     0     0     0     0     0     0     0     0     0     0     0     0

   X5_r  X5_s  X5_t  X5_u  X5_v  X5_w  X5_x  X5_y  X5_z  X6_b  X6_c  X6_d  \
0     0     0     1     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
3     0     0     0     0     0     0     0     0     1     0     0     0
4     0     0     0     0     0     0     0     1     0     0     0     0

   X6_e  X6_f  X6_g  X6_h  X6_i  X6_j  X6_k  X6_l  X8_b  X8_c  X8_d  X8_e  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     1     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     1     0     0     0     0     0     0
3     0     0     0     0     0     0     0     1     0     0     0     0
4     0     0     0     0     1     0     0     0     0     0     0     0

   X8_f  X8_g  X8_h  X8_i  X8_j  X8_k  X8_l  X8_m  X8_n  X8_o  X8_p  X8_q  \
0     0     0     0     0     0     0     0     0     0     0     0     0
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     1     0     0     0     0     0     0     0
3     0     0     0     0     0     0     0     0     1     0     0     0
4     0     0     0     0     0     0     0     1     0     0     0     0

   X8_r  X8_s  X8_t  X8_u  X8_v  X8_w  X8_x  X8_y
0     0     0     0     0     0     1     0     0
1     0     0     0     0     0     0     0     1
```

```
2     0     0     0     0     0     0     0     0
3     0     0     0     0     0     0     0     0
4     0     0     0     0     0     0     0     0
```

[ ]:

[46]:
```
#final shape of the test data
test_final.shape
```

[46]: (4209, 227)

[47]:
```
#segregating independent and dependent variable
X = train_final.drop("y",axis=1)
y = train_final.loc[:,"y"]
```

[48]:
```
# Splitting train and val data

from sklearn.model_selection import train_test_split
train_X, test_X, train_y, test_y = train_test_split(X, y,test_size = 0.3,␣
 ↪random_state = 123)
```

[80]:
```
#model building
import xgboost as xg
xgb_r = xg.XGBRegressor(objective ='reg:squarederror',n_estimators =30, seed␣
 ↪=123)
# Fitting the model
xgb_r.fit(train_X, train_y)
# Predict the model
y_pred = xgb_r.predict(test_X)
```

[81]:
```
from sklearn.metrics import mean_absolute_percentage_error,r2_score
```

[82]:
```
print(mean_absolute_percentage_error(test_y,y_pred))
```

```
0.05199845456795594
```

[83]:
```
#checking r2 score
score=r2_score(test_y,y_pred)
print(score)
```

```
0.5754236147467587
```

[ ]:
```
#lets tune the model by hyper parameters to get higher accuracy using␣
 ↪gridsearchCV
```

[ ]:

[75]:
```
# Define the parameter grid for the grid search
param_grid = {
    'max_depth': [3, 4],
    'learning_rate': [0.1, 0.01],
    'n_estimators': [50,100],
}
```

[76]:
```
# Create an XGBoost model
xgb_r= xg.XGBRegressor()

from sklearn.model_selection import GridSearchCV

# Perform a grid search with cross-validation
grid_search = GridSearchCV(xgb_r, param_grid=param_grid, cv=5,␣
 ↪scoring='neg_mean_squared_error')
grid_search.fit(train_X, train_y)
```

[76]:
```
GridSearchCV(cv=5,
             estimator=XGBRegressor(base_score=None, booster=None,
                                    colsample_bylevel=None,
                                    colsample_bynode=None,
                                    colsample_bytree=None, gamma=None,
                                    gpu_id=None, importance_type='gain',
                                    interaction_constraints=None,
                                    learning_rate=None, max_delta_step=None,
                                    max_depth=None, min_child_weight=None,
                                    missing=nan, monotone_constraints=None,
                                    n_estimators=100, n_jobs=None,
                                    num_parallel_tree=None, random_state=None,
                                    reg_alpha=None, reg_lambda=None,
                                    scale_pos_weight=None, subsample=None,
                                    tree_method=None, validate_parameters=False,
                                    verbosity=None),
             param_grid={'learning_rate': [0.1, 0.01], 'max_depth': [3, 4],
                         'n_estimators': [50, 100]},
             scoring='neg_mean_squared_error')
```

[78]:
```
# Print the best hyperparameters and the corresponding mean squared error
print('Best hyperparameters: {}'.format(grid_search.best_params_))
#print('Best mean squared error: {:.2f}'.format(-grid_search.best_score_))

# Evaluate the model on the test set using the best hyperparameters
y_predd = grid_search.predict(test_X)
mape = mean_absolute_percentage_error(test_y, y_predd)
print('Test mean squared error: {:.4f}'.format(mape))
```

```
Best hyperparameters: {'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50}
```

```
Test mean squared error: 0.0501
```

[79]:
```
#checking r2 score
score=r2_score(test_y,y_predd)
print(score)
```

```
0.5817154500983586
```

[ ]:
```

```

[84]:
```
#lets predict test data
test_pred = grid_search.predict(test_final)
```

[85]:
```
#creating dataframe to store predicted values
submission = pd.DataFrame({
    'ID' : test['ID'],
    'y' : test_pred
})
```

[86]:
```
submission.head(8)
```

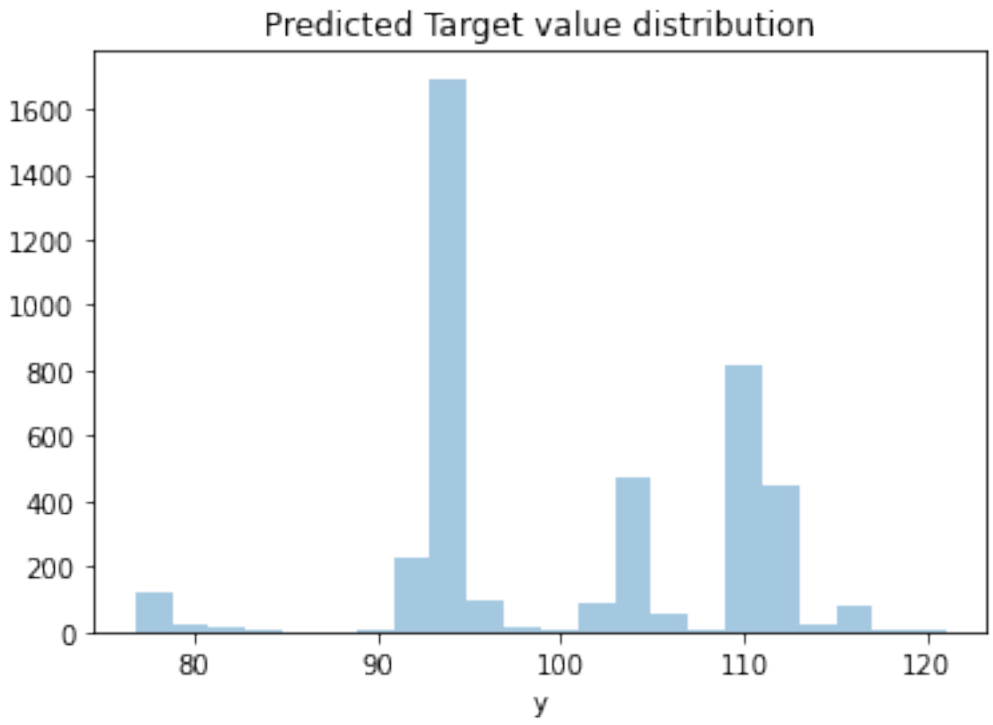
[86]:

	ID	y
0	1	78.2401
1	2	93.8482
2	3	78.7231
3	4	77.9545
4	5	111.3142
5	8	93.3457
6	10	110.6912
7	11	94.3795

[ ]:
```

```

[88]:
```
# Distribution plot for output variable.
sns.distplot(submission['y'],kde = False)
plt.title('Predicted Target value distribution')
```

[88]:
```
Text(0.5, 1.0, 'Predicted Target value distribution')
```



[ ]:
```
##conclusion:The XGBoost Regressor model, with a Mean Absolute Percentage Error␣
↪(MAPE) of 0.501 and a R-squared score of 0.581(58%).
#This model was able to capture the complex nonlinear relationship between the␣
↪features and the target variable,
#and therefore, it can be used for predicting the testing time with a␣
↪reasonable level of accuracy.
```