

# Mercedes-Benz Greener Manufacturing

February 10, 2022

## 1 Name: Sunil Pradhan

### 1.1 Project: 1

### 1.2 Project Name: Mercedes-Benz Greener Manufacturing

```
[1]: #import libraries

import numpy as np
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import mean_squared_error
from math import sqrt
```

```
[2]: #read the test and train dataset
df_train=pd.read_csv("train.csv")
df_test=pd.read_csv("test.csv")
```

```
[3]: #printing the train dataset

df_train.head()
```

```
[3]:      ID      y  X0 X1  X2 X3 X4 X5 X6 X8 ... X375  X376  X377  X378  X379  \
0    0  130.81   k  v  at  a  d  u  j  o  ...    0    0    1    0    0
1    6   88.53   k  t  av  e  d  y  l  o  ...    1    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
```

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

```
[5 rows x 378 columns]
```

```
[4]: #printing the test dataset
```

```
df_test.head()
```

```
[4]:
```

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

  

	X382	X383	X384	X385
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

[5 rows x 377 columns]

```
[5]: df_train.columns
```

```
[5]: Index(['ID', 'y', 'X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X8',  
        ...,  
        'X375', 'X376', 'X377', 'X378', 'X379', 'X380', 'X382', 'X383', 'X384',  
        'X385'],  
        dtype='object', length=378)
```

```
[6]: df_test.columns
```

```
[6]: Index(['ID', 'X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X8', 'X10',  
        ...,  
        'X375', 'X376', 'X377', 'X378', 'X379', 'X380', 'X382', 'X383', 'X384',  
        'X385'],  
        dtype='object', length=377)
```

Dropping the column ID as it not required for model prediction

```
[7]: df_train=df_train.drop(['ID'], axis=1)  
df_test=df_test.drop(['ID'], axis=1)
```

```
[8]: df_train.head()
```

```
[8]:
```

	y	X0	X1	X2	X3	X4	X5	X6	X8	X10	...	X375	X376	X377	X378	X379	\
0	130.81	k	v	at	a	d	u	j	o	0	...	0	0	1	0	0	
1	88.53	k	t	av	e	d	y	l	o	0	...	1	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

```

```

      X380  X382  X383  X384  X385
0         0     0     0     0     0
1         0     0     0     0     0
2         0     1     0     0     0
3         0     0     0     0     0
4         0     0     0     0     0

```

[5 rows x 377 columns]

```
[9]: df_test.head()
```

```

[9]:   X0 X1  X2 X3 X4 X5 X6 X8  X10 X11 ... X375 X376 X377 X378 X379 \
0  az  v   n  f  d  t  a  w    0    0 ...    0    0    0    1    0
1   t  b  ai  a  d  b  g  y    0    0 ...    0    0    1    0    0
2  az  v  as  f  d  a  j  j    0    0 ...    0    0    0    1    0
3  az  l   n  f  d  z  l  n    0    0 ...    0    0    0    1    0
4   w  s  as  c  d  y  i  m    0    0 ...    1    0    0    0    0

```

```

      X380  X382  X383  X384  X385
0         0     0     0     0     0
1         0     0     0     0     0
2         0     0     0     0     0
3         0     0     0     0     0
4         0     0     0     0     0

```

[5 rows x 376 columns]

**1.2.1 If for any column(s), the variance is equal to zero, then you need to remove those variable(s)**

```
[10]: var_col_zero=df_train.var()[df_train.var()==0].index.values
```

<ipython-input-10-37cc858c5963>:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric\_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

```
var_col_zero=df_train.var()[df_train.var()==0].index.values
```

```
[11]: #column index having zero variance
var_col_zero
```

```
[11]: array(['X11', 'X93', 'X107', 'X233', 'X235', 'X268', 'X289', 'X290',
            'X293', 'X297', 'X330', 'X347'], dtype=object)
```

```
[12]: #dropping the column having zero variance
df_train1=df_train.drop(var_col_zero, axis=1)
df_test1=df_test.drop(var_col_zero, axis=1)
```

```
[13]: #checking the initial columns and final columns after dropping column having
      ↪variance=0
df_train.shape[1], df_train1.shape[1]
```

```
[13]: (377, 365)
```

```
[14]: df_test.shape[1], df_test1.shape[1]
```

```
[14]: (376, 364)
```

### 1.2.2 Check for null and unique values for test and train sets

```
[15]: #checking for null values
np.sum(df_train1.isnull().sum())
```

```
[15]: 0
```

```
[16]: np.sum(df_test1.isnull().sum())
```

```
[16]: 0
```

In both train and test data set there is no null values

```
[17]: #checking for unique value
np.sum(df_train1.nunique().sum())
```

```
[17]: 3452
```

```
[18]: np.sum(df_test1.nunique().sum())
```

```
[18]: 908
```

There is 3452 unique values are in the train dataset and 908 unique values are in the dataset

```
[19]: #printing all the unique values of each column from train dataset

for i in df_train1:
    print("\nUnique value for the column",i)
    print(df_train1[i].unique())
```

```
Unique value for the column y
[130.81  88.53  76.26 ...  85.71 108.77  87.48]
```

Unique value for the column X0

['k' 'az' 't' 'al' 'o' 'w' 'j' 'h' 's' 'n' 'ay' 'f' 'x' 'y' 'aj' 'ak' 'am'  
'z' 'q' 'at' 'ap' 'v' 'af' 'a' 'e' 'ai' 'd' 'aq' 'c' 'aa' 'ba' 'as' 'i'  
'r' 'b' 'ax' 'bc' 'u' 'ad' 'au' 'm' 'l' 'aw' 'ao' 'ac' 'g' 'ab']

Unique value for the column X1

['v' 't' 'w' 'b' 'r' 'l' 's' 'aa' 'c' 'a' 'e' 'h' 'z' 'j' 'o' 'u' 'p' 'n'  
'i' 'y' 'd' 'f' 'm' 'k' 'g' 'q' 'ab']

Unique value for the column X2

['at' 'av' 'n' 'e' 'as' 'aq' 'r' 'ai' 'ak' 'm' 'a' 'k' 'ae' 's' 'f' 'd'  
'ag' 'ay' 'ac' 'ap' 'g' 'i' 'aw' 'y' 'b' 'ao' 'al' 'h' 'x' 'au' 't' 'an'  
'z' 'ah' 'p' 'am' 'j' 'q' 'af' 'l' 'aa' 'c' 'o' 'ar']

Unique value for the column X3

['a' 'e' 'c' 'f' 'd' 'b' 'g']

Unique value for the column X4

['d' 'b' 'c' 'a']

Unique value for the column X5

['u' 'y' 'x' 'h' 'g' 'f' 'j' 'i' 'd' 'c' 'af' 'ag' 'ab' 'ac' 'ad' 'ae'  
'ah' 'l' 'k' 'n' 'm' 'p' 'q' 's' 'r' 'v' 'w' 'o' 'aa']

Unique value for the column X6

['j' 'l' 'd' 'h' 'i' 'a' 'g' 'c' 'k' 'e' 'f' 'b']

Unique value for the column X8

['o' 'x' 'e' 'n' 's' 'a' 'h' 'p' 'm' 'k' 'd' 'i' 'v' 'j' 'b' 'q' 'w' 'g'  
'y' 'l' 'f' 'u' 'r' 't' 'c']

Unique value for the column X10

[0 1]

Unique value for the column X12

[0 1]

Unique value for the column X13

[1 0]

Unique value for the column X14

[0 1]

Unique value for the column X15

[0 1]

Unique value for the column X16

[0 1]

Unique value for the column X17  
[0 1]

Unique value for the column X18  
[1 0]

Unique value for the column X19  
[0 1]

Unique value for the column X20  
[0 1]

Unique value for the column X21  
[1 0]

Unique value for the column X22  
[0 1]

Unique value for the column X23  
[0 1]

Unique value for the column X24  
[0 1]

Unique value for the column X26  
[0 1]

Unique value for the column X27  
[0 1]

Unique value for the column X28  
[0 1]

Unique value for the column X29  
[0 1]

Unique value for the column X30  
[0 1]

Unique value for the column X31  
[1 0]

Unique value for the column X32  
[0 1]

Unique value for the column X33  
[0 1]

Unique value for the column X34  
[0 1]

Unique value for the column X35  
[1 0]

Unique value for the column X36  
[0 1]

Unique value for the column X37  
[1 0]

Unique value for the column X38  
[0 1]

Unique value for the column X39  
[0 1]

Unique value for the column X40  
[0 1]

Unique value for the column X41  
[0 1]

Unique value for the column X42  
[0 1]

Unique value for the column X43  
[0 1]

Unique value for the column X44  
[0 1]

Unique value for the column X45  
[0 1]

Unique value for the column X46  
[1 0]

Unique value for the column X47  
[0 1]

Unique value for the column X48  
[0 1]

Unique value for the column X49  
[0 1]

Unique value for the column X50  
[0 1]

Unique value for the column X51  
[0 1]

Unique value for the column X52  
[0 1]

Unique value for the column X53  
[0 1]

Unique value for the column X54  
[0 1]

Unique value for the column X55  
[0 1]

Unique value for the column X56  
[0 1]

Unique value for the column X57  
[0 1]

Unique value for the column X58  
[1 0]

Unique value for the column X59  
[0 1]

Unique value for the column X60  
[0 1]

Unique value for the column X61  
[0 1]

Unique value for the column X62  
[0 1]

Unique value for the column X63  
[0 1]

Unique value for the column X64  
[0 1]

Unique value for the column X65  
[0 1]



Unique value for the column X66  
[0 1]

Unique value for the column X67  
[0 1]

Unique value for the column X68  
[1 0]

Unique value for the column X69  
[0 1]

Unique value for the column X70  
[1 0]

Unique value for the column X71  
[0 1]

Unique value for the column X73  
[0 1]

Unique value for the column X74  
[1 0]

Unique value for the column X75  
[0 1]

Unique value for the column X76  
[0 1]

Unique value for the column X77  
[0 1]

Unique value for the column X78  
[0 1]

Unique value for the column X79  
[0 1]

Unique value for the column X80  
[0 1]

Unique value for the column X81  
[0 1]

Unique value for the column X82  
[0 1]

Unique value for the column X83  
[0 1]

Unique value for the column X84  
[0 1]

Unique value for the column X85  
[1 0]

Unique value for the column X86  
[0 1]

Unique value for the column X87  
[0 1]

Unique value for the column X88  
[0 1]

Unique value for the column X89  
[0 1]

Unique value for the column X90  
[0 1]

Unique value for the column X91  
[0 1]

Unique value for the column X92  
[0 1]

Unique value for the column X94  
[0 1]

Unique value for the column X95  
[0 1]

Unique value for the column X96  
[0 1]

Unique value for the column X97  
[0 1]

Unique value for the column X98  
[0 1]

Unique value for the column X99  
[0 1]

Unique value for the column X100  
[0 1]

Unique value for the column X101  
[0 1]

Unique value for the column X102  
[0 1]

Unique value for the column X103  
[0 1]

Unique value for the column X104  
[0 1]

Unique value for the column X105  
[0 1]

Unique value for the column X106  
[0 1]

Unique value for the column X108  
[0 1]

Unique value for the column X109  
[0 1]

Unique value for the column X110  
[0 1]

Unique value for the column X111  
[1 0]

Unique value for the column X112  
[0 1]

Unique value for the column X113  
[0 1]

Unique value for the column X114  
[1 0]

Unique value for the column X115  
[0 1]

Unique value for the column X116  
[1 0]

Unique value for the column X117  
[0 1]

Unique value for the column X118  
[1 0]

Unique value for the column X119  
[1 0]

Unique value for the column X120  
[1 0]

Unique value for the column X122  
[0 1]

Unique value for the column X123  
[0 1]

Unique value for the column X124  
[0 1]

Unique value for the column X125  
[0 1]

Unique value for the column X126  
[0 1]

Unique value for the column X127  
[0 1]

Unique value for the column X128  
[1 0]

Unique value for the column X129  
[0 1]

Unique value for the column X130  
[0 1]

Unique value for the column X131  
[1 0]

Unique value for the column X132  
[0 1]

Unique value for the column X133  
[0 1]

Unique value for the column X134  
[0 1]

Unique value for the column X135  
[0 1]

Unique value for the column X136  
[1 0]

Unique value for the column X137  
[1 0]

Unique value for the column X138  
[0 1]

Unique value for the column X139  
[0 1]

Unique value for the column X140  
[0 1]

Unique value for the column X141  
[0 1]

Unique value for the column X142  
[1 0]

Unique value for the column X143  
[0 1]

Unique value for the column X144  
[1 0]

Unique value for the column X145  
[0 1]

Unique value for the column X146  
[0 1]

Unique value for the column X147  
[0 1]

Unique value for the column X148  
[0 1]

Unique value for the column X150  
[1 0]

Unique value for the column X151  
[0 1]

Unique value for the column X152  
[0 1]

Unique value for the column X153  
[0 1]

Unique value for the column X154  
[0 1]

Unique value for the column X155  
[0 1]

Unique value for the column X156  
[1 0]

Unique value for the column X157  
[0 1]

Unique value for the column X158  
[0 1]

Unique value for the column X159  
[0 1]

Unique value for the column X160  
[0 1]

Unique value for the column X161  
[0 1]

Unique value for the column X162  
[0 1]

Unique value for the column X163  
[0 1]

Unique value for the column X164  
[0 1]

Unique value for the column X165  
[0 1]

Unique value for the column X166  
[0 1]

Unique value for the column X167  
[0 1]

Unique value for the column X168  
[0 1]

Unique value for the column X169  
[0 1]

Unique value for the column X170  
[1 0]

Unique value for the column X171  
[0 1]

Unique value for the column X172  
[0 1]

Unique value for the column X173  
[0 1]

Unique value for the column X174  
[0 1]

Unique value for the column X175  
[0 1]

Unique value for the column X176  
[0 1]

Unique value for the column X177  
[0 1]

Unique value for the column X178  
[0 1]

Unique value for the column X179  
[1 0]

Unique value for the column X180  
[0 1]

Unique value for the column X181  
[0 1]

Unique value for the column X182  
[0 1]

Unique value for the column X183  
[0 1]

Unique value for the column X184  
[1 0]

Unique value for the column X185  
[0 1]

Unique value for the column X186  
[0 1]

Unique value for the column X187  
[1 0]

Unique value for the column X189  
[1 0]

Unique value for the column X190  
[0 1]

Unique value for the column X191  
[0 1]

Unique value for the column X192  
[0 1]

Unique value for the column X194  
[1 0]

Unique value for the column X195  
[0 1]

Unique value for the column X196  
[0 1]

Unique value for the column X197  
[0 1]

Unique value for the column X198  
[0 1]

Unique value for the column X199  
[0 1]

Unique value for the column X200  
[0 1]



Unique value for the column X201  
[0 1]

Unique value for the column X202  
[0 1]

Unique value for the column X203  
[0 1]

Unique value for the column X204  
[1 0]

Unique value for the column X205  
[0 1]

Unique value for the column X206  
[0 1]

Unique value for the column X207  
[0 1]

Unique value for the column X208  
[0 1]

Unique value for the column X209  
[1 0]

Unique value for the column X210  
[0 1]

Unique value for the column X211  
[0 1]

Unique value for the column X212  
[0 1]

Unique value for the column X213  
[0 1]

Unique value for the column X214  
[0 1]

Unique value for the column X215  
[0 1]

Unique value for the column X216  
[0 1]

Unique value for the column X217  
[0 1]

Unique value for the column X218  
[0 1]

Unique value for the column X219  
[0 1]

Unique value for the column X220  
[1 0]

Unique value for the column X221  
[0 1]

Unique value for the column X222  
[0 1]

Unique value for the column X223  
[0 1]

Unique value for the column X224  
[0 1]

Unique value for the column X225  
[0 1]

Unique value for the column X226  
[0 1]

Unique value for the column X227  
[0 1]

Unique value for the column X228  
[0 1]

Unique value for the column X229  
[0 1]

Unique value for the column X230  
[0 1]

Unique value for the column X231  
[0 1]

Unique value for the column X232  
[0 1]

Unique value for the column X234  
[1 0]

Unique value for the column X236  
[0 1]

Unique value for the column X237  
[1 0]

Unique value for the column X238  
[0 1]

Unique value for the column X239  
[0 1]

Unique value for the column X240  
[0 1]

Unique value for the column X241  
[0 1]

Unique value for the column X242  
[0 1]

Unique value for the column X243  
[0 1]

Unique value for the column X244  
[0 1]

Unique value for the column X245  
[0 1]

Unique value for the column X246  
[0 1]

Unique value for the column X247  
[0 1]

Unique value for the column X248  
[0 1]

Unique value for the column X249  
[0 1]

Unique value for the column X250  
[0 1]

Unique value for the column X251  
[0 1]

Unique value for the column X252  
[0 1]

Unique value for the column X253  
[0 1]

Unique value for the column X254  
[0 1]

Unique value for the column X255  
[0 1]

Unique value for the column X256  
[0 1]

Unique value for the column X257  
[0 1]

Unique value for the column X258  
[0 1]

Unique value for the column X259  
[0 1]

Unique value for the column X260  
[0 1]

Unique value for the column X261  
[0 1]

Unique value for the column X262  
[1 0]

Unique value for the column X263  
[1 0]

Unique value for the column X264  
[0 1]

Unique value for the column X265  
[0 1]

Unique value for the column X266  
[1 0]

Unique value for the column X267  
[0 1]

Unique value for the column X269  
[0 1]

Unique value for the column X270  
[0 1]

Unique value for the column X271  
[0 1]

Unique value for the column X272  
[0 1]

Unique value for the column X273  
[1 0]

Unique value for the column X274  
[0 1]

Unique value for the column X275  
[1 0]

Unique value for the column X276  
[0 1]

Unique value for the column X277  
[0 1]

Unique value for the column X278  
[0 1]

Unique value for the column X279  
[0 1]

Unique value for the column X280  
[0 1]

Unique value for the column X281  
[0 1]

Unique value for the column X282  
[0 1]

Unique value for the column X283  
[0 1]

Unique value for the column X284  
[0 1]

Unique value for the column X285  
[1 0]

Unique value for the column X286  
[0 1]

Unique value for the column X287  
[0 1]

Unique value for the column X288  
[0 1]

Unique value for the column X291  
[0 1]

Unique value for the column X292  
[0 1]

Unique value for the column X294  
[0 1]

Unique value for the column X295  
[0 1]

Unique value for the column X296  
[0 1]

Unique value for the column X298  
[0 1]

Unique value for the column X299  
[0 1]

Unique value for the column X300  
[0 1]

Unique value for the column X301  
[0 1]

Unique value for the column X302  
[0 1]

Unique value for the column X304  
[0 1]

Unique value for the column X305  
[0 1]

Unique value for the column X306  
[1 0]

Unique value for the column X307  
[0 1]

Unique value for the column X308  
[0 1]

Unique value for the column X309  
[0 1]

Unique value for the column X310  
[0 1]

Unique value for the column X311  
[0 1]

Unique value for the column X312  
[0 1]

Unique value for the column X313  
[0 1]

Unique value for the column X314  
[0 1]

Unique value for the column X315  
[0 1]

Unique value for the column X316  
[1 0]

Unique value for the column X317  
[0 1]

Unique value for the column X318  
[0 1]

Unique value for the column X319  
[0 1]

Unique value for the column X320  
[0 1]

Unique value for the column X321  
[0 1]

Unique value for the column X322  
[0 1]

Unique value for the column X323  
[0 1]

Unique value for the column X324  
[1 0]

Unique value for the column X325  
[0 1]

Unique value for the column X326  
[0 1]

Unique value for the column X327  
[1 0]

Unique value for the column X328  
[0 1]

Unique value for the column X329  
[1 0]

Unique value for the column X331  
[0 1]

Unique value for the column X332  
[0 1]

Unique value for the column X333  
[0 1]

Unique value for the column X334  
[1 0]

Unique value for the column X335  
[0 1]

Unique value for the column X336  
[0 1]

Unique value for the column X337  
[0 1]



Unique value for the column X338  
[0 1]

Unique value for the column X339  
[0 1]

Unique value for the column X340  
[0 1]

Unique value for the column X341  
[0 1]

Unique value for the column X342  
[0 1]

Unique value for the column X343  
[0 1]

Unique value for the column X344  
[0 1]

Unique value for the column X345  
[0 1]

Unique value for the column X346  
[0 1]

Unique value for the column X348  
[0 1]

Unique value for the column X349  
[0 1]

Unique value for the column X350  
[0 1]

Unique value for the column X351  
[0 1]

Unique value for the column X352  
[0 1]

Unique value for the column X353  
[0 1]

Unique value for the column X354  
[1 0]

Unique value for the column X355  
[0 1]

Unique value for the column X356  
[0 1]

Unique value for the column X357  
[0 1]

Unique value for the column X358  
[0 1]

Unique value for the column X359  
[0 1]

Unique value for the column X360  
[0 1]

Unique value for the column X361  
[1 0]

Unique value for the column X362  
[0 1]

Unique value for the column X363  
[0 1]

Unique value for the column X364  
[0 1]

Unique value for the column X365  
[0 1]

Unique value for the column X366  
[0 1]

Unique value for the column X367  
[0 1]

Unique value for the column X368  
[0 1]

Unique value for the column X369  
[0 1]

Unique value for the column X370  
[0 1]

Unique value for the column X371  
[0 1]

Unique value for the column X372  
[0 1]

Unique value for the column X373  
[0 1]

Unique value for the column X374  
[0 1]

Unique value for the column X375  
[0 1]

Unique value for the column X376  
[0 1]

Unique value for the column X377  
[1 0]

Unique value for the column X378  
[0 1]

Unique value for the column X379  
[0 1]

Unique value for the column X380  
[0 1]

Unique value for the column X382  
[0 1]

Unique value for the column X383  
[0 1]

Unique value for the column X384  
[0 1]

Unique value for the column X385  
[0 1]

```
[20]: #printing all the unique values of each column from test dataset
```

```
for i in df_test1:  
    print("\nUnique value for the column",i)
```

```
print(df_train1[i].unique())
```

Unique value for the column X0

```
['k' 'az' 't' 'al' 'o' 'w' 'j' 'h' 's' 'n' 'ay' 'f' 'x' 'y' 'aj' 'ak' 'am'
 'z' 'q' 'at' 'ap' 'v' 'af' 'a' 'e' 'ai' 'd' 'aq' 'c' 'aa' 'ba' 'as' 'i'
 'r' 'b' 'ax' 'bc' 'u' 'ad' 'au' 'm' 'l' 'aw' 'ao' 'ac' 'g' 'ab']
```

Unique value for the column X1

```
['v' 't' 'w' 'b' 'r' 'l' 's' 'aa' 'c' 'a' 'e' 'h' 'z' 'j' 'o' 'u' 'p' 'n'
 'i' 'y' 'd' 'f' 'm' 'k' 'g' 'q' 'ab']
```

Unique value for the column X2

```
['at' 'av' 'n' 'e' 'as' 'aq' 'r' 'ai' 'ak' 'm' 'a' 'k' 'ae' 's' 'f' 'd'
 'ag' 'ay' 'ac' 'ap' 'g' 'i' 'aw' 'y' 'b' 'ao' 'al' 'h' 'x' 'au' 't' 'an'
 'z' 'ah' 'p' 'am' 'j' 'q' 'af' 'l' 'aa' 'c' 'o' 'ar']
```

Unique value for the column X3

```
['a' 'e' 'c' 'f' 'd' 'b' 'g']
```

Unique value for the column X4

```
['d' 'b' 'c' 'a']
```

Unique value for the column X5

```
['u' 'y' 'x' 'h' 'g' 'f' 'j' 'i' 'd' 'c' 'af' 'ag' 'ab' 'ac' 'ad' 'ae'
 'ah' 'l' 'k' 'n' 'm' 'p' 'q' 's' 'r' 'v' 'w' 'o' 'aa']
```

Unique value for the column X6

```
['j' 'l' 'd' 'h' 'i' 'a' 'g' 'c' 'k' 'e' 'f' 'b']
```

Unique value for the column X8

```
['o' 'x' 'e' 'n' 's' 'a' 'h' 'p' 'm' 'k' 'd' 'i' 'v' 'j' 'b' 'q' 'w' 'g'
 'y' 'l' 'f' 'u' 'r' 't' 'c']
```

Unique value for the column X10

```
[0 1]
```

Unique value for the column X12

```
[0 1]
```

Unique value for the column X13

```
[1 0]
```

Unique value for the column X14

```
[0 1]
```

Unique value for the column X15

[0 1]

Unique value for the column X16

[0 1]

Unique value for the column X17

[0 1]

Unique value for the column X18

[1 0]

Unique value for the column X19

[0 1]

Unique value for the column X20

[0 1]

Unique value for the column X21

[1 0]

Unique value for the column X22

[0 1]

Unique value for the column X23

[0 1]

Unique value for the column X24

[0 1]

Unique value for the column X26

[0 1]

Unique value for the column X27

[0 1]

Unique value for the column X28

[0 1]

Unique value for the column X29

[0 1]

Unique value for the column X30

[0 1]

Unique value for the column X31

[1 0]

Unique value for the column X32

[0 1]

Unique value for the column X33

[0 1]

Unique value for the column X34

[0 1]

Unique value for the column X35

[1 0]

Unique value for the column X36

[0 1]

Unique value for the column X37

[1 0]

Unique value for the column X38

[0 1]

Unique value for the column X39

[0 1]

Unique value for the column X40

[0 1]

Unique value for the column X41

[0 1]

Unique value for the column X42

[0 1]

Unique value for the column X43

[0 1]

Unique value for the column X44

[0 1]

Unique value for the column X45

[0 1]

Unique value for the column X46

[1 0]

Unique value for the column X47

[0 1]

Unique value for the column X48

[0 1]

Unique value for the column X49

[0 1]

Unique value for the column X50

[0 1]

Unique value for the column X51

[0 1]

Unique value for the column X52

[0 1]

Unique value for the column X53

[0 1]

Unique value for the column X54

[0 1]

Unique value for the column X55

[0 1]

Unique value for the column X56

[0 1]

Unique value for the column X57

[0 1]

Unique value for the column X58

[1 0]

Unique value for the column X59

[0 1]

Unique value for the column X60

[0 1]

Unique value for the column X61

[0 1]

Unique value for the column X62

[0 1]

Unique value for the column X63

[0 1]

Unique value for the column X64

[0 1]

Unique value for the column X65

[0 1]

Unique value for the column X66

[0 1]

Unique value for the column X67

[0 1]

Unique value for the column X68

[1 0]

Unique value for the column X69

[0 1]

Unique value for the column X70

[1 0]

Unique value for the column X71

[0 1]

Unique value for the column X73

[0 1]

Unique value for the column X74

[1 0]

Unique value for the column X75

[0 1]

Unique value for the column X76

[0 1]

Unique value for the column X77

[0 1]

Unique value for the column X78

[0 1]

Unique value for the column X79

[0 1]

Unique value for the column X80

[0 1]

Unique value for the column X81



[0 1]

Unique value for the column X82

[0 1]

Unique value for the column X83

[0 1]

Unique value for the column X84

[0 1]

Unique value for the column X85

[1 0]

Unique value for the column X86

[0 1]

Unique value for the column X87

[0 1]

Unique value for the column X88

[0 1]

Unique value for the column X89

[0 1]

Unique value for the column X90

[0 1]

Unique value for the column X91

[0 1]

Unique value for the column X92

[0 1]

Unique value for the column X94

[0 1]

Unique value for the column X95

[0 1]

Unique value for the column X96

[0 1]

Unique value for the column X97

[0 1]

Unique value for the column X98

[0 1]

Unique value for the column X99

[0 1]

Unique value for the column X100

[0 1]

Unique value for the column X101

[0 1]

Unique value for the column X102

[0 1]

Unique value for the column X103

[0 1]

Unique value for the column X104

[0 1]

Unique value for the column X105

[0 1]

Unique value for the column X106

[0 1]

Unique value for the column X108

[0 1]

Unique value for the column X109

[0 1]

Unique value for the column X110

[0 1]

Unique value for the column X111

[1 0]

Unique value for the column X112

[0 1]

Unique value for the column X113

[0 1]

Unique value for the column X114

[1 0]

Unique value for the column X115

[0 1]

Unique value for the column X116

[1 0]

Unique value for the column X117

[0 1]

Unique value for the column X118

[1 0]

Unique value for the column X119

[1 0]

Unique value for the column X120

[1 0]

Unique value for the column X122

[0 1]

Unique value for the column X123

[0 1]

Unique value for the column X124

[0 1]

Unique value for the column X125

[0 1]

Unique value for the column X126

[0 1]

Unique value for the column X127

[0 1]

Unique value for the column X128

[1 0]

Unique value for the column X129

[0 1]

Unique value for the column X130

[0 1]

Unique value for the column X131

[1 0]

Unique value for the column X132

[0 1]

Unique value for the column X133

[0 1]

Unique value for the column X134

[0 1]

Unique value for the column X135

[0 1]

Unique value for the column X136

[1 0]

Unique value for the column X137

[1 0]

Unique value for the column X138

[0 1]

Unique value for the column X139

[0 1]

Unique value for the column X140

[0 1]

Unique value for the column X141

[0 1]

Unique value for the column X142

[1 0]

Unique value for the column X143

[0 1]

Unique value for the column X144

[1 0]

Unique value for the column X145

[0 1]

Unique value for the column X146

[0 1]

Unique value for the column X147

[0 1]

Unique value for the column X148

[0 1]

Unique value for the column X150

[1 0]

Unique value for the column X151

[0 1]

Unique value for the column X152

[0 1]

Unique value for the column X153

[0 1]

Unique value for the column X154

[0 1]

Unique value for the column X155

[0 1]

Unique value for the column X156

[1 0]

Unique value for the column X157

[0 1]

Unique value for the column X158

[0 1]

Unique value for the column X159

[0 1]

Unique value for the column X160

[0 1]

Unique value for the column X161

[0 1]

Unique value for the column X162

[0 1]

Unique value for the column X163

[0 1]

Unique value for the column X164

[0 1]

Unique value for the column X165

[0 1]

Unique value for the column X166

[0 1]

Unique value for the column X167

[0 1]

Unique value for the column X168

[0 1]

Unique value for the column X169

[0 1]

Unique value for the column X170

[1 0]

Unique value for the column X171

[0 1]

Unique value for the column X172

[0 1]

Unique value for the column X173

[0 1]

Unique value for the column X174

[0 1]

Unique value for the column X175

[0 1]

Unique value for the column X176

[0 1]

Unique value for the column X177

[0 1]

Unique value for the column X178

[0 1]

Unique value for the column X179

[1 0]

Unique value for the column X180

[0 1]

Unique value for the column X181

[0 1]

Unique value for the column X182

[0 1]

Unique value for the column X183

[0 1]

Unique value for the column X184

[1 0]

Unique value for the column X185

[0 1]

Unique value for the column X186

[0 1]

Unique value for the column X187

[1 0]

Unique value for the column X189

[1 0]

Unique value for the column X190

[0 1]

Unique value for the column X191

[0 1]

Unique value for the column X192

[0 1]

Unique value for the column X194

[1 0]

Unique value for the column X195

[0 1]

Unique value for the column X196

[0 1]

Unique value for the column X197

[0 1]

Unique value for the column X198

[0 1]

Unique value for the column X199

[0 1]

Unique value for the column X200

[0 1]

Unique value for the column X201

[0 1]

Unique value for the column X202

[0 1]

Unique value for the column X203

[0 1]

Unique value for the column X204

[1 0]

Unique value for the column X205

[0 1]

Unique value for the column X206

[0 1]

Unique value for the column X207

[0 1]

Unique value for the column X208

[0 1]

Unique value for the column X209

[1 0]

Unique value for the column X210

[0 1]

Unique value for the column X211

[0 1]

Unique value for the column X212

[0 1]

Unique value for the column X213

[0 1]

Unique value for the column X214

[0 1]

Unique value for the column X215



[0 1]

Unique value for the column X216

[0 1]

Unique value for the column X217

[0 1]

Unique value for the column X218

[0 1]

Unique value for the column X219

[0 1]

Unique value for the column X220

[1 0]

Unique value for the column X221

[0 1]

Unique value for the column X222

[0 1]

Unique value for the column X223

[0 1]

Unique value for the column X224

[0 1]

Unique value for the column X225

[0 1]

Unique value for the column X226

[0 1]

Unique value for the column X227

[0 1]

Unique value for the column X228

[0 1]

Unique value for the column X229

[0 1]

Unique value for the column X230

[0 1]

Unique value for the column X231

[0 1]

Unique value for the column X232

[0 1]

Unique value for the column X234

[1 0]

Unique value for the column X236

[0 1]

Unique value for the column X237

[1 0]

Unique value for the column X238

[0 1]

Unique value for the column X239

[0 1]

Unique value for the column X240

[0 1]

Unique value for the column X241

[0 1]

Unique value for the column X242

[0 1]

Unique value for the column X243

[0 1]

Unique value for the column X244

[0 1]

Unique value for the column X245

[0 1]

Unique value for the column X246

[0 1]

Unique value for the column X247

[0 1]

Unique value for the column X248

[0 1]

Unique value for the column X249

[0 1]

Unique value for the column X250

[0 1]

Unique value for the column X251

[0 1]

Unique value for the column X252

[0 1]

Unique value for the column X253

[0 1]

Unique value for the column X254

[0 1]

Unique value for the column X255

[0 1]

Unique value for the column X256

[0 1]

Unique value for the column X257

[0 1]

Unique value for the column X258

[0 1]

Unique value for the column X259

[0 1]

Unique value for the column X260

[0 1]

Unique value for the column X261

[0 1]

Unique value for the column X262

[1 0]

Unique value for the column X263

[1 0]

Unique value for the column X264

[0 1]

Unique value for the column X265

[0 1]

Unique value for the column X266

[1 0]

Unique value for the column X267

[0 1]

Unique value for the column X269

[0 1]

Unique value for the column X270

[0 1]

Unique value for the column X271

[0 1]

Unique value for the column X272

[0 1]

Unique value for the column X273

[1 0]

Unique value for the column X274

[0 1]

Unique value for the column X275

[1 0]

Unique value for the column X276

[0 1]

Unique value for the column X277

[0 1]

Unique value for the column X278

[0 1]

Unique value for the column X279

[0 1]

Unique value for the column X280

[0 1]

Unique value for the column X281

[0 1]

Unique value for the column X282

[0 1]

Unique value for the column X283

[0 1]

Unique value for the column X284

[0 1]

Unique value for the column X285

[1 0]

Unique value for the column X286

[0 1]

Unique value for the column X287

[0 1]

Unique value for the column X288

[0 1]

Unique value for the column X291

[0 1]

Unique value for the column X292

[0 1]

Unique value for the column X294

[0 1]

Unique value for the column X295

[0 1]

Unique value for the column X296

[0 1]

Unique value for the column X298

[0 1]

Unique value for the column X299

[0 1]

Unique value for the column X300

[0 1]

Unique value for the column X301

[0 1]

Unique value for the column X302

[0 1]

Unique value for the column X304

[0 1]

Unique value for the column X305

[0 1]

Unique value for the column X306

[1 0]

Unique value for the column X307

[0 1]

Unique value for the column X308

[0 1]

Unique value for the column X309

[0 1]

Unique value for the column X310

[0 1]

Unique value for the column X311

[0 1]

Unique value for the column X312

[0 1]

Unique value for the column X313

[0 1]

Unique value for the column X314

[0 1]

Unique value for the column X315

[0 1]

Unique value for the column X316

[1 0]

Unique value for the column X317

[0 1]

Unique value for the column X318

[0 1]

Unique value for the column X319

[0 1]

Unique value for the column X320

[0 1]

Unique value for the column X321

[0 1]

Unique value for the column X322

[0 1]

Unique value for the column X323

[0 1]

Unique value for the column X324

[1 0]

Unique value for the column X325

[0 1]

Unique value for the column X326

[0 1]

Unique value for the column X327

[1 0]

Unique value for the column X328

[0 1]

Unique value for the column X329

[1 0]

Unique value for the column X331

[0 1]

Unique value for the column X332

[0 1]

Unique value for the column X333

[0 1]

Unique value for the column X334

[1 0]

Unique value for the column X335

[0 1]

Unique value for the column X336

[0 1]

Unique value for the column X337

[0 1]

Unique value for the column X338

[0 1]

Unique value for the column X339

[0 1]

Unique value for the column X340

[0 1]

Unique value for the column X341

[0 1]

Unique value for the column X342

[0 1]

Unique value for the column X343

[0 1]

Unique value for the column X344

[0 1]

Unique value for the column X345

[0 1]

Unique value for the column X346

[0 1]

Unique value for the column X348

[0 1]

Unique value for the column X349

[0 1]

Unique value for the column X350

[0 1]

Unique value for the column X351

[0 1]

Unique value for the column X352

[0 1]

Unique value for the column X353



[0 1]

Unique value for the column X354

[1 0]

Unique value for the column X355

[0 1]

Unique value for the column X356

[0 1]

Unique value for the column X357

[0 1]

Unique value for the column X358

[0 1]

Unique value for the column X359

[0 1]

Unique value for the column X360

[0 1]

Unique value for the column X361

[1 0]

Unique value for the column X362

[0 1]

Unique value for the column X363

[0 1]

Unique value for the column X364

[0 1]

Unique value for the column X365

[0 1]

Unique value for the column X366

[0 1]

Unique value for the column X367

[0 1]

Unique value for the column X368

[0 1]

Unique value for the column X369

[0 1]

Unique value for the column X370

[0 1]

Unique value for the column X371

[0 1]

Unique value for the column X372

[0 1]

Unique value for the column X373

[0 1]

Unique value for the column X374

[0 1]

Unique value for the column X375

[0 1]

Unique value for the column X376

[0 1]

Unique value for the column X377

[1 0]

Unique value for the column X378

[0 1]

Unique value for the column X379

[0 1]

Unique value for the column X380

[0 1]

Unique value for the column X382

[0 1]

Unique value for the column X383

[0 1]

Unique value for the column X384

[0 1]

Unique value for the column X385

[0 1]

### 1.2.3 Applying Label Encoder

```
[21]: #finding the column having datatype as object
obj_col=df_train1.describe(include=[object]).columns.values
obj_col
```

```
[21]: array(['X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X8'], dtype=object)
```

```
[22]: #label encoder
le=LabelEncoder()
for i in obj_col:
    le.fit(df_train1[i].append(df_test1[i]).values)
    df_train1[i]=le.transform(df_train1[i])
    df_test1[i]=le.transform(df_test1[i])
```

```
[23]: #create X and y
X=df_train1.drop(['y'], axis=1)
y=df_train1['y']
```

```
[24]: # create test and train split
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(X,y,test_size=0.3,random_state=7)
```

### 1.2.4 Performing dimensionality reduction

```
[25]: from sklearn.decomposition import PCA
```

```
[26]: #using PCA decomposition of the data from 364 components with 98% variance to
↳ lower dimension of 12 components
pca=PCA(0.98, svd_solver='full')
sklearn_pca=pca.fit(X)
```

```
[27]: sklearn_pca.n_components_
```

```
[27]: 12
```

```
[28]: sklearn_pca.explained_variance_ratio_
```

```
[28]: array([0.40868988, 0.21758508, 0.13120081, 0.10783522, 0.08165248,
        0.0140934 , 0.00660951, 0.00384659, 0.00260289, 0.00214378,
        0.00209857, 0.00180388])
```

### 1.2.5 predicting using XGBoost

```
[29]: pca_xtrain=pd.DataFrame(pca.transform(xtrain))
      pca_xtest=pd.DataFrame(pca.transform(xtest))
      pca_df_test1=pd.DataFrame(pca.transform(df_test1))
```

```
[30]: import xgboost as xgb
```

```
[31]: xgb_reg=xgb.XGBRegressor(objective='reg:squarederror',learning_rate=0.1)
      xgb_reg.fit(pca_xtrain,ytrain)
```

```
[31]: XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
                  colsample_bynode=1, colsample_bytree=1, enable_categorical=False,
                  gamma=0, gpu_id=-1, importance_type=None,
                  interaction_constraints='', learning_rate=0.1, max_delta_step=0,
                  max_depth=6, min_child_weight=1, missing=nan,
                  monotone_constraints='()', n_estimators=100, n_jobs=8,
                  num_parallel_tree=1, predictor='auto', random_state=0, reg_alpha=0,
                  reg_lambda=1, scale_pos_weight=1, subsample=1, tree_method='exact',
                  validate_parameters=1, verbosity=None)
```

```
[32]: xgb_reg_pred_ytest=xgb_reg.predict(pca_xtest)
```

```
[33]: print(sqrt(mean_squared_error(ytest,xgb_reg_pred_ytest)))
```

10.47229744101521

```
[34]: xgb_RFreg=xgb.XGBRFRegressor(objective='reg:squarederror',learning_rate=1)
      xgb_RFreg.fit(pca_xtrain,ytrain)
```

```
[34]: XGBRFRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
                  colsample_bytree=1, enable_categorical=False, gamma=0, gpu_id=-1,
                  importance_type=None, interaction_constraints='',
                  learning_rate=1, max_delta_step=0, max_depth=6,
                  min_child_weight=1, missing=nan, monotone_constraints='()',
                  n_estimators=100, n_jobs=8, num_parallel_tree=100,
                  objective='reg:squarederror', predictor='auto', random_state=0,
                  reg_alpha=0, scale_pos_weight=1, tree_method='exact',
                  validate_parameters=1, verbosity=None)
```

```
[35]: xgb_RFreg_pred_ytest=xgb_RFreg.predict(pca_xtest)
```

```
[36]: print(sqrt(mean_squared_error(ytest,xgb_RFreg_pred_ytest)))
```

10.64404619286274

```
[ ]:
```