Data_Preprocessing1

March 2, 2022

[56]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import sys
import os
import csv
import warnings
warnings.filterwarnings("ignore")
```

[57]:
```
#Read the data
data= pd.read_csv("/Users/aminameghezzi/Downloads/Yocket-dataset.csv", sep=',')
data
```

[57]:

	Name	University	Course	Year	Status	GRE_SCORE	Eng_test
0	Nishigandha	Arizona State University	Management Information System	Fall 2019	Admite	N.A.	ENG TEST
1	pranalimore808atgmail.com	Arizona State University	Management Information System	Fall 2018	Admite	305	ENG TEST
2	bostonner	Arizona State University	Management Information System	Fall 2018	Admite	315	TOEFL
3	KrithikaCT	Arizona State University	Management Information System	Fall 2018	Admite	301	TOEFL
4	Giridhar312	Arizona State University	Management Information System	Fall 2018	Admite	311	IELTS
…	…	…	…	…	…	…	…
1277	keval	Worcester Polytechnic Institute	Management Information System	Fall 2016	Reject	311	TOEFL
1278	pranil23	Worcester Polytechnic Institute	Management Information System	Fall 2016	Reject	295	TOEFL
1279	sonaligc	Worcester Polytechnic Institute	Management Information System	Fall 2015	Reject	307	TOEFL
1280	pratz10	Worcester Polytechnic Institute					
1281	sandy2912	Worcester Polytechnic Institute					

```
1280  Management Information System    Fall 2015   Reject       301     TOEFL
1281  Management Information System  Spring 2016   Reject       300     TOEFL

     Test_score Undergrad_score  work_ex
0            NA           65.30%     18.0
1            NA              7.1      NaN
2           104              77%      3.0
3             0           80.12%     60.0
4             7            7.74     25.0
…           …              …        …
1277         95              61%      NaN
1278         90           60.60%     18.0
1279         88                7      NaN
1280         89           62.60%      NaN
1281         93                5      NaN

[1282 rows x 10 columns]
```

[58]:
```
print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1282 entries, 0 to 1281
Data columns (total 10 columns):
Name               1282 non-null object
University         1282 non-null object
Course             1282 non-null object
Year               1282 non-null object
Status             1282 non-null object
GRE_SCORE          1282 non-null object
Eng_test           1282 non-null object
Test_score         1282 non-null object
Undergrad_score    1282 non-null object
work_ex            1101 non-null float64
dtypes: float64(1), object(9)
memory usage: 100.3+ KB
None
```

[59]:
```
#normalize the GPA Score
data['Undergrad_score'] = data['Undergrad_score'].str.split('/').str[0]
data['GPA'] = data['Undergrad_score'].str.split('%').str[0]
data.drop(columns=['Undergrad_score'], inplace=True)
#convert GPA into integer
data['GPA'] = pd.to_numeric(data['GPA'], errors='coerce')
data['GPA'] = data['GPA'].apply(lambda x: x*7.1+12 if(x>4 and x<10) else x)
```

[60]:
```
#convert status to binary
data.loc[data['Status'] == "Admite", 'Status'] = 1
```

```
data.loc[data['Status'] == "Reject", 'Status'] = 0
```

[61]:
```
#convert features into float
data['Status'] = pd.to_numeric(data['Status'], errors='coerce')
data['GRE_SCORE'] = pd.to_numeric(data['GRE_SCORE'], errors='coerce')
#convert features string
data['University'].str
```

[61]:
```
<pandas.core.strings.StringMethods at 0x10b46af50>
```

[62]:
```
data['GRE Verbal'] = data['GRE_SCORE']/2
data['GRE Quant']=data['GRE_SCORE']-data['GRE Verbal']
data['GRE Verbal']=data['GRE Verbal'].apply(np.floor)
data['GRE Quant']=data['GRE Quant'].apply(np.ceil)
```

[63]:
```
data.rename(columns={"Test_score":"Language proficiency"}, inplace=True )
data['Language proficiency'] = pd.to_numeric(data['Language proficiency'],␣
↪errors='coerce')
data.rename(columns={"Status":"Decision"}, inplace=True )
data.rename(columns={"Course":"Major"}, inplace=True )
data.rename(columns={"Language proficiency":"Language Proficiency"},␣
↪inplace=True )
data = data[['University', 'Major', 'Decision', 'GRE_SCORE','GRE Quant','GRE␣
↪Verbal','work_ex','Name','GPA','Language Proficiency']]
data
```

[63]:

	University	Major	Decision	GRE_SCORE	GRE Quant	GRE Verbal	work_ex
0	Arizona State University	Management Information System	1	NaN	NaN	NaN	18.0
1	Arizona State University	Management Information System	1	305.0	153.0	152.0	NaN
2	Arizona State University	Management Information System	1	315.0	158.0	157.0	3.0
3	Arizona State University	Management Information System	1	301.0	151.0	150.0	60.0
4	Arizona State University	Management Information System	1	311.0	156.0	155.0	25.0
…	…	…	…	…	…	…	…
1277	Worcester Polytechnic Institute	Management Information System	0	311.0	156.0	155.0	NaN
1278	Worcester Polytechnic Institute	Management Information System					
1279	Worcester Polytechnic Institute	Management Information System					
1280	Worcester Polytechnic Institute	Management Information System					
1281	Worcester Polytechnic Institute	Management Information System					

```
1278         0      295.0      148.0       147.0     18.0
1279         0      307.0      154.0       153.0      NaN
1280         0      301.0      151.0       150.0      NaN
1281         0      300.0      150.0       150.0      NaN

                           Name     GPA  Language Proficiency
0                     Nishigandha  65.300                  NaN
1     pranalimore808atgmail.com    62.410                  NaN
2                       bostonner  77.000                104.0
3                     KrithikaCT   80.120                  0.0
4                     Giridhar312  66.954                  7.0
…                             …     …                     …
1277                        keval  61.000                 95.0
1278                    pranil23   60.600                 90.0
1279                    sonaligc   61.700                 88.0
1280                     pratz10   62.600                 89.0
1281                   sandy2912   47.500                 93.0

[1282 rows x 10 columns]
```

[64]:
```
#drop the nan values
data.dropna(inplace = True)
```

[65]:
```
#check if we have nan values
data.isnull().any().any()
```

[65]:
```
False
```

[66]:
```
data1 = pd.read_csv('/Users/aminameghezzi/Downloads/Fall14.csv',␣
↪encoding='latin1')
data1.head()
```

[66]:
```
  Timestamp             Name                          University Applied  \
0       NaN              NaN                            Acceptance Ratio
1     2-Mar              NaN   Western new england,1198 massachusettes
2    15-Mar            jigar                           Arizona State univ
3    18-Jan          Karthik                 Arizona State University
4    23-Jan  Gagandeep Singh                 Arizona State University

  Date of Application Accept / Reject Date of Decision  \
0                 NaN           71.74              NaN
1           27-Jan-14          Accept        19-Feb-14
2                 NaN          Accept              NaN
3                 NaN          Accept              NaN
4           30-Dec-13          Accept        17-Jan-14

                    Major  GRE  GRE (Quants)  AWA  TOEFL  IELTS Work-Ex  \
```

```
0                 Averages  309           160  3.3  101.0    6.5     NaN
1             Engg Mngmnt   289           147  3.5   84.0    NaN     NaN
2              mechanical   311           161  3.0  109.0    NaN     NaN
3                 CS(AME)   324           163  NaN  106.0    NaN    1 yr
4  Construction Management  312           163  3.0  102.0    NaN       0

   International Papers Under Graduate Aggregate  Scale Undergrad Univesity
0                   NaN                      NaN    NaN                  NaN
1                   NaN                       62  100.0               Mumbai
2                   NaN                    61.05  100.0          Mumbai Univ
3                   NaN                     7.51   10.0                  NaN
4                   1.0                     7.02   10.0  Manipal University
```

[67]:
```
#combine all TOEFL and IELTS scores

toefl=data1.TOEFL
ielts=data1.IELTS
#toefl = toefl.rename(columns=d)
#gre = gre.rename(columns=d)
#df_total = pd.concat([toefl, ielts], ignore_index=False)
df_total = data1['TOEFL'].fillna(data1['IELTS'])
print (df_total)
print(len(df_total))
print(df_total[3])

data1['English Language Proficiency']=df_total
data1.drop(columns=['TOEFL','IELTS'], inplace=True)
```

```
0       101.0
1        84.0
2       109.0
3       106.0
4       102.0
        …
1662      NaN
1663      6.5
1664      7.5
1665      6.5
1666    108.0
Name: TOEFL, Length: 1667, dtype: float64
1667
106.0
```

[68]:
```
# remove all letters from work ex
data1['work-Ex'] = data1['Work-Ex'].str.extract('(\d+)', expand=False)
```

[69]:
```
#Remove special characters in GPA

data1['new gpa temp'] = data1['Under Graduate Aggregate'].str.split('/').str[0]
data1['new gpa'] = data1['new gpa temp'].str.split('%').str[0]
#xyz.drop(columns=['new gpa'], inplace=True)
data1.drop(columns=['new gpa temp'], inplace=True)
data1['new gpa'] = pd.to_numeric(data1['new gpa'], errors='coerce')
data1['new gpa'] = data1['new gpa'].apply(lambda x: x*7.1+12 if(x>4 and x<10)␣
↪else x)

data1['work-ex'] = pd.to_numeric(data1['Work-Ex'], errors='coerce')
```

[70]:
```
#remove redundant column Acads

data1.drop(columns=['Under Graduate Aggregate'], inplace=True)
data1.drop(columns=['International Papers','Scale','Undergrad Univesity'],␣
↪inplace=True)
data1.drop(columns=['Timestamp','Date of Application','AWA','Date of Decision'␣
↪], inplace=True)
```

[71]:
```
print(data1.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1667 entries, 0 to 1666
Data columns (total 11 columns):
Name                            1570 non-null object
University Applied              1666 non-null object
Accept / Reject                 1667 non-null object
Major                           1667 non-null object
GRE                             1667 non-null int64
GRE (Quants)                    1667 non-null int64
Work-Ex                         1118 non-null object
English Language Proficiency    1626 non-null float64
work-Ex                         1009 non-null object
new gpa                         1661 non-null float64
work-ex                         436 non-null float64
dtypes: float64(3), int64(2), object(6)
memory usage: 143.4+ KB
None
```

[72]:
```
data1.rename(columns={"University Applied": "University", "GRE": "GRE_SCORE",␣
↪"GRE (Quants)":"GRE Quant", "work-Ex":"work_ex","new gpa":"GPA","English␣
↪Language Proficiency":"Language Proficiency"}, inplace=True)
data1.rename(columns={"Accept / Reject": "Decision"}, inplace=True)
```

[73]:
```
data1
```

[73]:
```
                 Name                                University Decision  \
0                 NaN                          Acceptance Ratio    71.74
1                 NaN   Western new england,1198 massachusettes   Accept
2               jigar                        Arizona State univ   Accept
3             Karthik                  Arizona State University   Accept
4     Gagandeep Singh                  Arizona State University   Accept
…                 …                                       …      …
1662            Mohan                  Wright State University    Reject
1663            Mohan                  Wright state University    Reject
1664     Suhaib Siraj                    Wyane State University   Accept
1665            Mohan                Youngstown state university   Reject
1666      Ankit Mitra                                     NaN   Accept

                            Major  GRE_SCORE  GRE Quant   Work-Ex  \
0                        Averages        309        160       NaN
1                     Engg Mngmnt        289        147       NaN
2                      mechanical        311        161       NaN
3                         CS(AME)        324        163      1 yr
4          Construction Management        312        163         0
…                              …          …          …       …
1662                          CS          295        147  6 Months
1663                          CS          296        147       NaN
1664                   MS in CS           306        157       NaN
1665                       CIS            295        147   6months
1666  MS in Electrical Engineering        313        161       NaN

      Language Proficiency work_ex     GPA  work-ex
0                    101.0     NaN     NaN      NaN
1                     84.0     NaN  62.000      NaN
2                    109.0     NaN  61.050      NaN
3                    106.0       1  65.321      NaN
4                    102.0       0  61.842      0.0
…                    …     …     …      …
1662                   NaN       6  65.000      NaN
1663                   6.5     NaN  65.000      NaN
1664                   7.5     NaN  63.000      NaN
1665                   6.5       6  64.000      NaN
1666                 108.0     NaN  74.835      NaN

[1667 rows x 11 columns]
```

[74]:
```
#drop the nan values
data1.dropna(inplace = True)
```

[75]:
```
data1['GRE Verbal'] = data1['GRE_SCORE'] - data1['GRE Quant']
data1 = data1[['University', 'Major', 'Decision', 'GRE_SCORE','GRE Quant','GRE␣
 ↪Verbal','work_ex','Name','GPA','Language Proficiency']]
```

```
data1
```

[75]:
```
                        University                    Major Decision  \
4         Arizona State University  Construction Management   Accept
5         Arizona State University  Industrial Engineering    Accept
9         Arizona State University          Computer Science   Accept
16        Arizona State University                  MS in CS   Accept
19        Arizona State University                       CS   Accept
…                              …                       …      …
1640  western michigan university               electricals   Accept
1645     Wichita State University                 MS in CSE   Accept
1648     wichita state university                       EE   Accept
1654      Wright State University                      eee   Accept
1658      Wright State University          Computer Science   Reject

      GRE_SCORE  GRE Quant  GRE Verbal work_ex             Name     GPA  \
4           312        163         149       0  Gagandeep Singh  61.842
5           319        163         156       0    Jineet Lilani  63.160
9           318        161         157       0           Sujata  69.865
16          322        163         159       0  Aamir Goriawala  69.120
19          311        164         147       1             Uday  70.910
…           …          …           …         …                …      …
1640        299        154         145       0  Sachin B Bethur   3.760
1645        294        154         140       0            Gulam  62.220
1648        291        150         141       0    DIVYA SAHITHI  63.000
1654        284        152         132       0            ashok  69.300
1658        280        150         130       1              sai  75.000

      Language Proficiency
4                    102.0
5                    114.0
9                    114.0
16                   117.0
19                    97.0
…                     …
1640                  96.0
1645                   7.0
1648                   7.0
1654                   6.0
1658                   5.5

[408 rows x 10 columns]
```

[76]:
```
data1.isnull().any().any()
```

[76]:
```
False
```

[77]:
```
#Read the data
data2= pd.read_csv("/Users/aminameghezzi/Downloads/Fall17.csv", sep=',')
data2
```

[77]:

```
                                                 1                   Timestamp  \
0    To add an entry to the sheet please go to:  http://bit.ly/fall17MSform
1                                                3             12/7/2016 19:43
2                                                4             12/7/2016 20:28
3                                                5             12/7/2016 20:40
4                                                6             12/7/2016 21:19
..                                             ...                         ...
193                                            195             1/30/2017 19:00
194                                            196             1/30/2017 19:06
195                                            197             1/30/2017 19:20
196                                            198             1/30/2017 20:23
197                                            199             1/30/2017 20:27

                              Applied University Name           Applied Branch  \
0    If you feel any entry is fake let me know at ->  srujanbarai75@gmail.com
1                                                UTD                     MS CS
2                                                UTD                      Mech
3                  Michigan technological University             Data science
4                           Arizona State University       Business Analytics
..                                               ...                       ...
193                 Stevens Institute of Technology   Engineering Management
194                                              UTD                  Telecom
195                 Stevens institute of technology                     MIS
196                      Carnegie Mellon University           MSIT ebusiness
197                      University of Pennsylvania                      CIS

    Decision    Acads     UG University  GRE Total  GRE Quant  GRE Verbal  \
0        NaN      NaN               NaN        NaN        NaN         NaN
1      Admit   7.6/10                MU      326.0      167.0       159.0
2      Admit   7.5/10   Anna University      318.0      158.0       160.0
3      Admit   6.6/10  GITAM University      303.0      154.0       149.0
4      Admit     7.85                MU      307.0      161.0       146.0
..       ...      ...               ...        ...        ...         ...
193    Admit     2.69             NMIMS      308.0      161.0       147.0
194    Admit      8.9               VTU      308.0      159.0       149.0
195    Admit  6.55/10                MU      301.0      159.0       142.0
196    Admit      7.6       ISM Dhanbad      325.0      170.0       155.0
197    Admit   76.80%               VTU      321.0      165.0       156.0

     ... IELTS    Work Ex   UG Branch Category (your assumption)  \
0    ...   NaN         NaN         NaN                        NaN
1    ...   NaN           0          CE                       Safe
2    ...   NaN           0  Mechanical                       Safe
```

```
3    …   NaN          0         ECE                        Mod
4    …   NaN          0          EE                        Mod
..   …   …        …          …                        …
193  …   NaN         7M        EXTC                      Safe
194  …   NaN          0     Telecom                       Mod
195  …     0  3M intern          CS                      Safe
196  …   NaN      2Y 7M         CSE                      Ambi
197  …   NaN          0          CS                      Ambi

    Application Date   Decision Date Your name (Optional)  \
0                NaN             NaN                   NaN
1         22/10/2016      12/11/2016                Nirmod
2           18/10/16        15/11/16               Akilesh
3         16/10/2016       6/11/2016               Abhinav
4            1st Oct        26th Oct                   NaN
..               …              …                   …
193    15th November   18th November                   NaN
194     31st October     3rd January                   NaN
195            6-Dec           5-Jan             Balaji K
196           12-Dec          26-Jan         Sakshi Gopal
197        28-Oct-16  January 2,2017                   NaN

      Your mail ID (Optional) Paper Published  \
0                         NaN             NaN
1                         NaN             NaN
2                         NaN             NaN
3      chabhinav.95@gmail.com             NaN
4                         NaN             NaN
..                        …             …
193                       NaN             NaN
194                       NaN               0
195  balaji.katakam@gmail.com               0
196    sakshigopal7@gmail.com               0
197                       NaN               0

                Outstanding achievements
0                                    NaN
1                                    NaN
2                                    NaN
3                                    NaN
4                                    NaN
..                                   …
193                                  NaN
194                                  NaN
195                                 None
196  1 patent applied, startup experience
197                                  NaN
```

```
[198 rows x 21 columns]
```

[78]:
```
#combine all TOEFL and IELTS scores

toefl=data2.TOEFL
ielts=data2.IELTS
#toefl = toefl.rename(columns=d)
#gre = gre.rename(columns=d)
#df_total = pd.concat([toefl, ielts], ignore_index=False)
df_total = data2['TOEFL'].fillna(data2['IELTS'])
print (df_total)
print(len(df_total))
print(df_total[3])

data2['English Language Proficiency']=df_total
data2.drop(columns=['TOEFL','IELTS'], inplace=True)
```

```
0                    NaN
1       114(30,30,24,30)
2       110(29:29:24:28)
3        97(26:25:20:26)
4                     99
              …
193      98(22:22:20:28)
194                  108
195      88(22:19:24:23)
196     110(29:26:26:29)
197     114(27:30:17:30)
Name: TOEFL, Length: 198, dtype: object
198
97(26:25:20:26)
```

[79]:
```
# remove all letters from work ex
data2['Work Ex'] = data2['Work Ex'].str.extract('(\d+)', expand=False)
```

[80]:
```
#Remove special characters in GPA

data2['new gpa temp'] = data2['Acads'].str.split('/').str[0]
data2['new gpa'] = data2['new gpa temp'].str.split('%').str[0]
#xyz.drop(columns=['new gpa'], inplace=True)
data2.drop(columns=['new gpa temp'], inplace=True)
```

[81]:
```
#remove special characters in English language proficiency

#data2['English Language Proficiency']=df_total
```

```
data2['Language Proficiency'] = data2['English Language Proficiency'].str.
 ↪split('(').str[0]
data2.drop(columns=['English Language Proficiency'], inplace=True)
```

[82]:
```
#remove redundant column Acads

data2.drop(columns=['Acads'], inplace=True)
```

[83]:
```
# convert gpa and work ex columns to float types so that we can perform␣
 ↪calculations

data2['new gpa'] = pd.to_numeric(data2['new gpa'], errors='coerce')
data2['Work Ex'] = pd.to_numeric(data2['Work Ex'], errors='coerce')
print (data2.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 198 entries, 0 to 197
Data columns (total 20 columns):
1                              198 non-null object
Timestamp                      198 non-null object
Applied University Name        198 non-null object
Applied Branch                 198 non-null object
Decision                       196 non-null object
UG University                  197 non-null object
GRE Total                      197 non-null float64
GRE Quant                      197 non-null float64
GRE Verbal                     197 non-null float64
Work Ex                        175 non-null float64
UG Branch                      197 non-null object
Category (your assumption)     196 non-null object
Application Date               158 non-null object
Decision Date                  162 non-null object
Your name (Optional)           93 non-null object
Your mail ID (Optional)        49 non-null object
Paper Published                181 non-null object
Outstanding achievements       51 non-null object
new gpa                        192 non-null float64
Language Proficiency           191 non-null object
dtypes: float64(5), object(15)
memory usage: 31.1+ KB
None
```

[84]:
```
#Apply a function to convert the GPA to a 100 point scale if it is in a 10␣
 ↪point scale

data2['new gpa'] = data2['new gpa'].apply(lambda x: x*7.1+12 if(x>0 and x<10)␣
 ↪else x)
```

[85]:
```
data2.drop(columns=['1','Timestamp','UG University','Outstanding␣
↪achievements','Your mail ID (Optional)'], inplace=True)
data2.drop(columns=['Paper Published'], inplace=True)
```

[86]:
```
data2.rename(columns={"Applied University Name": "University", "Applied Branch":
↪ "Course", "Your name (Optional)":"Name"}, inplace=True)
data2.rename(columns={"GRE Total": "GRE_SCORE", "Work Ex": "work_ex", "new gpa":
↪"GPA"}, inplace=True)
data2.rename(columns={"Application Date": "Application_Date"}, inplace=True)
data2.drop(columns=['Category (your assumption)','Application_Date'],␣
↪inplace=True)
data2.rename(columns={"Course": "Major"}, inplace=True)
data2
```

[86]:

	University	Major	\
0	If you feel any entry is fake let me know at ->	srujanbarai75@gmail.com	
1	UTD	MS CS	
2	UTD	Mech	
3	Michigan technological University	Data science	
4	Arizona State University	Business Analytics	
..	…	…	
193	Stevens Institute of Technology	Engineering Management	
194	UTD	Telecom	
195	Stevens institute of technology	MIS	
196	Carnegie Mellon University	MSIT ebusiness	
197	University of Pennsylvania	CIS	

	Decision	GRE_SCORE	GRE Quant	GRE Verbal	work_ex	UG Branch	\
0	NaN	NaN	NaN	NaN	NaN	NaN	
1	Admit	326.0	167.0	159.0	0.0	CE	
2	Admit	318.0	158.0	160.0	0.0	Mechanical	
3	Admit	303.0	154.0	149.0	0.0	ECE	
4	Admit	307.0	161.0	146.0	0.0	EE	
..	…	…	…	…	…	…	
193	Admit	308.0	161.0	147.0	7.0	EXTC	
194	Admit	308.0	159.0	149.0	0.0	Telecom	
195	Admit	301.0	159.0	142.0	3.0	CS	
196	Admit	325.0	170.0	155.0	2.0	CSE	
197	Admit	321.0	165.0	156.0	0.0	CS	

	Decision Date	Name	GPA	Language Proficiency
0	NaN	NaN	NaN	NaN
1	12/11/2016	Nirmod	65.960	114
2	15/11/16	Akilesh	65.250	110
3	6/11/2016	Abhinav	58.860	97
4	26th Oct	NaN	67.735	99
..	…	…	…	…

```
193   18th November           NaN  31.099                  98
194     3rd January           NaN  75.190                 108
195           5-Jan      Balaji K  58.505                  88
196          26-Jan  Sakshi Gopal  65.960                 110
197  January 2,2017           NaN  76.800                 114

[198 rows x 12 columns]
```

[87]:
```
# fix the date formats for the application date

data2['Language Proficiency'] = pd.to_numeric(data2['Language Proficiency'],␣
↪errors='coerce')
```

[88]:
```
data2.drop([0],axis=0,inplace=True)
```

[89]:
```
#drop the nan values
data2.dropna(inplace = True)
#check if we have nan values
data2.isnull().any().any()
```

[89]:
```
False
```

[90]:
```
data2 = data2[['University', 'Major', 'Decision', 'GRE_SCORE','GRE Quant','GRE␣
↪Verbal','work_ex','Name','GPA','Language Proficiency']]
data2
```

[90]:
```
                          University           Major Decision  GRE_SCORE  \
1                                UTD           MS CS    Admit      326.0
2                                UTD            Mech    Admit      318.0
3    Michigan technological University    Data science    Admit      303.0
6                                UTD              CS    Admit      311.0
8              Santa Clara University             CSE    Admit      308.0
..                                ...             ...      ...        ...
187             Ohio State University              IE    Admit      320.0
188             Ohio State University              IE    Admit      320.0
192       University of Texas, Dallas              CS    Admit      329.0
195    Stevens institute of technology             MIS    Admit      301.0
196         Carnegie Mellon University  MSIT ebusiness    Admit      325.0

     GRE Quant  GRE Verbal  work_ex                 Name     GPA  \
1        167.0       159.0      0.0               Nirmod  65.960
2        158.0       160.0      0.0               Akilesh  65.250
3        154.0       149.0      0.0               Abhinav  58.860
6        161.0       150.0      1.0         Sunny Bangale  66.620
8        159.0       149.0      9.0          Parth Ladani  75.474
..         ...         ...      ...                  ...     ...
187      165.0       155.0      0.0                  AVS  77.320
```

```
188      165.0       155.0      0.0                   AVS  77.320
192      170.0       159.0      0.0  Yash Prakash Pradhan  76.894
195      159.0       142.0      3.0              Balaji K  58.505
196      170.0       155.0      2.0           Sakshi Gopal  65.960

     Language Proficiency
1                   114.0
2                   110.0
3                    97.0
6                    92.0
8                    92.0
..                    …
187                 104.0
188                 104.0
192                 111.0
195                  88.0
196                 110.0

[76 rows x 10 columns]
```

[91]:
```
frames = [data, data1, data2]
data4 = pd.concat(frames)
data4
```

[91]:
```
                         University                          Major Decision  \
2           Arizona State University  Management Information System        1
3           Arizona State University  Management Information System        1
4           Arizona State University  Management Information System        1
5           Arizona State University  Management Information System        1
6           Arizona State University  Management Information System        1
..                               …                                 …      …
187            Ohio State University                                IE    Admit
188            Ohio State University                                IE    Admit
192      University of Texas, Dallas                                CS    Admit
195  Stevens institute of technology                               MIS    Admit
196       Carnegie Mellon University                    MSIT ebusiness    Admit

     GRE_SCORE  GRE Quant  GRE Verbal work_ex                 Name     GPA  \
2        315.0      158.0       157.0       3             bostonner  77.000
3        301.0      151.0       150.0      60            KrithikaCT  80.120
4        311.0      156.0       155.0      25           Giridhar312  66.954
5        312.0      156.0       156.0      24            Shreepriya  73.060
6        307.0      154.0       153.0      34               muktika  68.800
..         …          …           …         …                   …      …
187      320.0      165.0       155.0       0                   AVS  77.320
188      320.0      165.0       155.0       0                   AVS  77.320
192      329.0      170.0       159.0       0  Yash Prakash Pradhan  76.894
```

```
195      301.0      159.0       142.0       3              Balaji K  58.505
196      325.0      170.0       155.0       2          Sakshi Gopal  65.960

     Language Proficiency
2                   104.0
3                     0.0
4                     7.0
5                   110.0
6                   103.0
..                    …
187                 104.0
188                 104.0
192                 111.0
195                  88.0
196                 110.0

[1491 rows x 10 columns]
```

[92]:
```
data4.loc[data4['Decision'] == "Accept", 'Decision'] = 1
data4.loc[data4['Decision'] == "Admit", 'Decision'] = 1
data4.loc[data4['Decision'] == "Reject", 'Decision'] = 0
```

[93]:
```
#converting toefl score into ielts acoording to ETS scale
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 4.
↪5 if(x>32 and x<35) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 5.
↪0 if(x>34 and x<45) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 5.
↪5 if(x>44 and x<59) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 6.
↪0 if(x>58 and x<78) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 6.
↪5 if(x>77 and x<93) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 7.
↪0 if(x>92 and x<101) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 7.
↪5 if(x>100 and x<109) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 8.
↪0 if(x>108 and x<114) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 8.
↪5 if(x>113 and x<117) else x)
data4['Language Proficiency'] = data4['Language Proficiency'].apply(lambda x: 9.
↪0 if(x>116 and x<120) else x)
```

[98]:
```
data4.loc[data4['University'] == "Arizona State University", 'University'] = 50
data4.loc[data4['University'] == "University of Arizona", 'University'] = 50
```

```
data4.loc[data4['University'] == "Boston University", 'University'] = 29
data4.loc[data4['University'] == "Carnegie Mellon University", 'University'] =␣
↪19
data4.loc[data4['University'] == "University of Delaware", 'University'] = 89
data4.loc[data4['University'] == "Drexel University", 'University'] = 95
data4.loc[data4['University'] == "Illinois Institute of Technology",␣
↪'University'] = 78
data4.loc[data4['University'] == "Iowa State University", 'University'] = 91
data4.loc[data4['University'] == "Indiana University Bloomington",␣
↪'University'] = 60
data4.loc[data4['University'] == "Northeastern University", 'University'] = 65
data4.loc[data4['University'] == "New York University", 'University'] = 17
data4.loc[data4['University'] == "Pennsylvania State University", 'University']␣
↪= 28
data4.loc[data4['University'] == "Rochester Institute of Technology",␣
↪'University'] = 107
data4.loc[data4['University'] == "Rensselaer Polytechnic Institute",␣
↪'University'] = 39
data4.loc[data4['University'] == "Rutgers University-New Brunswick",␣
↪'University'] = 69
data4.loc[data4['University'] == "Santa Clara University", 'University'] = 54
data4.loc[data4['University'] == "Stevens Institute of Technology",␣
↪'University'] = 74
data4.loc[data4['University'] == "University at Buffalo SUNY", 'University'] =␣
↪79
data4.loc[data4['University'] == "Syracuse University", 'University'] = 54
data4.loc[data4['University'] == "Syracuse University Management Information␣
↪System", 'University'] = 54
data4.loc[data4['University'] == "Texas A&M; University, College Station",␣
↪'University'] = 58
data4.loc[data4['University'] == "University of Cincinnati", 'University'] = 139
data4.loc[data4['University'] == "University of California, Los Angeles",␣
↪'University'] = 20
data4.loc[data4['University'] == "University of Florida", "University"] = 34
data4.loc[data4['University'] == "University of Illinois at Chicago",␣
↪"University"] = 138
data4.loc[data4['University'] == "University of Maryland, College Park",␣
↪'University'] = 64
data4.loc[data4['University'] == "University of North Carolina at Charlotte",␣
↪'University'] = 228
data4.loc[data4['University'] == "University of Pennsylvania", 'University'] = 6
data4.loc[data4['University'] == "University of Utah", 'University'] = 147
data4.loc[data4['University'] == "University of Texas at Dallas", 'University']␣
↪= 104
data4.loc[data4['University'] == "University of Washington", 'University'] = 23
```

```
data4.loc[data4['University'] == "University of Minnesota, Twin Cities",␣
↪'University'] = 70
data4.loc[data4['University'] == "Northwestern University", 'University'] = 9
data4.loc[data4['University'] == "Rutgers University, Newark", 'University'] =␣
↪62
data4.loc[data4['University'] == "University of Texas at Arlington",␣
↪'University'] = 293
data4.loc[data4['University'] == "University of Washingto", 'University'] = 14
data4.loc[data4['University'] == "UTD", 'University'] = 147
data4.loc[data4['University'] == "Michigan technological University",␣
↪'University'] = 147
data4.loc[data4['University'] == "University of Texas Arlington", 'University']␣
↪= 293
data4.loc[data4['University'] == "Wayne state university", 'University'] = 246
data4.loc[data4['University'] == "RIT", 'University'] = 104
data4.loc[data4['University'] == "Stevens institute of technology",␣
↪'University'] = 74
data4.loc[data4['University'] == "Ohio State University", 'University'] = 54
data4.loc[data4['University'] == "University of Texas Dallas", 'University'] =␣
↪147
data4.loc[data4['University'] == "Michigan Technological University",␣
↪'University'] = 147
data4.loc[data4['University'] == "university of wyoming", 'University'] = 228
data4.loc[data4['University'] == "University at Buffalo, SUNY", 'University'] =␣
↪79
data4.loc[data4['University'] == "Purdue University", 'University'] = 57
data4.loc[data4['University'] == "TAMU, College station", 'University'] = 70
data4.loc[data4['University'] == "University of Maryland", 'University'] = 64
data4.loc[data4['University'] == "UPenn", 'University'] = 6
data4.loc[data4['University'] == "UIC", 'University'] = 132
data4.loc[data4['University'] == "Boston university", 'University'] = 40
data4.loc[data4['University'] == "university of houston", 'University'] = 185
data4.loc[data4['University'] == "University of Nebraska Omahao", 'University']␣
↪= 381
data4.loc[data4['University'] == "university of akron", 'University'] = 293
data4.loc[data4['University'] == "Northern Illinois Univ", 'University'] = 293

data4.loc[data4['University'] == "University of North Carolina at CHAPEL HILL",␣
↪'University'] = 80
data4.loc[data4['University'] == "George mason university", 'University'] = 153
data4.loc[data4['University'] == "University of Texas arlington", 'University']␣
↪= 293
data4.loc[data4['University'] == "NYU Tandon", 'University'] = 40
data4.loc[data4['University'] == "SUNY Binghamton", 'University'] = 79
data4.loc[data4['University'] == "Uic", 'University'] = 132
data4.loc[data4['University'] == "NCSU", 'University'] = 84
```

```
data4.loc[data4['University'] == "Texas A&M College Station", 'University'] = 70
data4.loc[data4['University'] == "UT Arlington", 'University'] = 293
data4.loc[data4['University'] == "UTD", 'University'] = 147
data4.loc[data4['University'] == "university of akron", 'University'] = 293
data4.loc[data4['University'] == "Northern Illinois Univ", 'University'] = 293


data4.loc[data4['University'] == "University of Southern California",␣
↪'University'] = 22
data4.loc[data4['University'] == "stevens", 'University'] = 74
data4.loc[data4['University'] == "Michigan Tech", 'University'] = 147
data4.loc[data4['University'] == "UT Dallas", 'University'] = 147
data4.loc[data4['University'] == "UPenn", 'University'] = 6
data4.loc[data4['University'] == "North Carolina State University",␣
↪'University'] = 84
data4.loc[data4['University'] == "CSU Fresno", 'University'] = 211
data4.loc[data4['University'] == "SYRACUSE", 'University'] = 54
data4.loc[data4['University'] == "Syracuse", 'University'] = 54
data4.loc[data4['University'] == "University of Texas, Dallas", 'University'] =␣
↪147
data4.loc[data4['University'] == "University of Connecticut", 'University'] = 64
data4.loc[data4['University'] == "umass dartmouth", 'University'] = 218

data4.loc[data4['University'] == "SUNY Buffalo", 'University'] = 79
data4.loc[data4['University'] == "IIT - Chicago", 'University'] = 117
data4.loc[data4['University'] == "ASu", 'University'] = 117
data4.loc[data4['University'] == "Southern methodist university", 'University']␣
↪= 64
data4.loc[data4['University'] == "University of North Carolina, Charlotte",␣
↪'University'] = 228
data4.loc[data4['University'] == "Bowling Green State Univeristy",␣
↪'University'] = 246
data4.loc[data4['University'] == "George mason", 'University'] = 153
data4.loc[data4['University'] == "OHIO STATE UNIVERSITY", 'University'] = 54
data4.loc[data4['University'] == "MIT", 'University'] = 3
data4.loc[data4['University'] == "Arizona", 'University'] = 117
data4.loc[data4['University'] == "RIT", 'University'] = 104
data4.loc[data4['University'] == "TAMU college station", 'University'] = 70

data4.loc[data4['University'] == "UT Arlington", 'University'] = 293
data4.loc[data4['University'] == "NCSU", 'University'] = 84
data4.loc[data4['University'] == "Northern Illinois University", 'University']␣
↪= 293
data4.loc[data4['University'] == "University of Texas - Austin", 'University']␣
↪= 48
data4.loc[data4['University'] == "TU DELFT", 'University'] = 17
```

```
data4.loc[data4['University'] == "IIT, Chicago", 'University'] = 117
data4.loc[data4['University'] == "Ncsu", 'University'] = 84
data4.loc[data4['University'] == "IIT CHICAGO", 'University'] = 117
data4.loc[data4['University'] == "University Of Texas, Arlington",␣
↪'University'] = 293
data4.loc[data4['University'] == "Syracus University", 'University'] = 54
data4.loc[data4['University'] == "Wayne State", 'University'] = 246
data4.loc[data4['University'] == "Portland State University", 'University'] =␣
↪293
data4.loc[data4['University'] == "Columbia University", 'University'] = 3
data4.loc[data4['University'] == "IUPUI", 'University'] = 228
data4.loc[data4['University'] == "Rbradley", 'University'] = 4
data4.loc[data4['University'] == "Bradley Univ,Wright State University,Wichita␣
↪state university,California state university fullerton,Texas A&M␣
↪kingsville", 'University'] = 100
data4.loc[data4['University'] == "california inst tech(4034)", 'University'] =␣
↪12
data4.loc[data4['University'] == "UCalifornia State University - East bay",␣
↪'University'] = 80
data4.loc[data4['University'] == "California State University EastBay",␣
↪'University'] = 80
data4.loc[data4['University'] == "California State University- Long Beach",␣
↪'University'] = 20
data4.loc[data4['University'] == "California State University, San Bernardino",␣
↪'University'] = 41
data4.loc[data4['University'] == "Carnegie Mellon Univsesity- Robotics␣
↪Institue", 'University'] = 25
data4.loc[data4['University'] == "Clemson University", 'University'] = 70
data4.loc[data4['University'] == "Cleveland State University", "University"] =␣
↪293
data4.loc[data4['University'] == "Colorado state univers", "University"] = 166
data4.loc[data4['University'] == "Columbia University, College Park",␣
↪'University'] = 3
data4.loc[data4['University'] == "Cornell University", 'University'] = 17
data4.loc[data4['University'] == "Florida institute of technology",␣
↪'University'] = 202
data4.loc[data4['University'] == "George Mason University", 'University'] = 153
data4.loc[data4['University'] == "Houston clear lake", 'University'] = 43
data4.loc[data4['University'] == "IOWA state university", 'University'] = 121

data4.loc[data4['University'] == "Lamar University", 'University'] = 293
data4.loc[data4['University'] == "louisiana state university", 'University'] =␣
↪72
data4.loc[data4['University'] == "Michigan Technological University",␣
↪'University'] = 147
```

```
data4.loc[data4['University'] == "Midwestern state university", 'University'] =␣
↪70
data4.loc[data4['University'] == "Mississippi State University", 'University']␣
↪= 211
data4.loc[data4['University'] == "MIT", 'University'] = 3
data4.loc[data4['University'] == "MTU", 'University'] = 147
data4.loc[data4['University'] == "NCSU", 'University'] = 84
data4.loc[data4['University'] == "New Mexico State University", 'University'] =␣
↪263
data4.loc[data4['University'] == "North Eastern University", 'University'] = 40
data4.loc[data4['University'] == "NPU", 'University'] = 27
data4.loc[data4['University'] == "NUS", 'University'] = 4
data4.loc[data4['University'] == "Oklahama State University", 'University'] =␣
↪192
data4.loc[data4['University'] == "old dominion university", 'University'] = 263
data4.loc[data4['University'] == "Rbradley", 'University'] = 241
data4.loc[data4['University'] == "Penn State", 'University'] = 50
data4.loc[data4['University'] == "Portland State University", 'University'] =␣
↪293
data4.loc[data4['University'] == "Princeton University", 'University'] = 1
data4.loc[data4['University'] == "Purdue University", 'University'] = 57
data4.loc[data4['University'] == "Rutgers", 'University'] = 62
data4.loc[data4['University'] == "San Diego State University", 'University'] =␣
↪147
data4.loc[data4['University'] == "Souther illionois carbondale", 'University']␣
↪= 254
data4.loc[data4['University'] == "Stanford University", 'University'] = 6
data4.loc[data4['University'] == "Stony Brook University (SUNY Stony Brook)",␣
↪"University"] = 91
data4.loc[data4['University'] == "SUNY Buffalo", "University"] = 79
data4.loc[data4['University'] == "Towson University, College Park",␣
↪'University'] = 197
data4.loc[data4['University'] == "University of North Carolina", 'University']␣
↪= 80
data4.loc[data4['University'] == "University of Colorado, Boulder",␣
↪'University'] = 104
data4.loc[data4['University'] == "University of Illinois Chicago",␣
↪'University'] = 132
data4.loc[data4['University'] == "University Of Kentucky", 'University'] = 132
data4.loc[data4['University'] == "University of Maryland", 'University'] = 64

data4.loc[data4['University'] == "university of Massachusetts Amherst",␣
↪'University'] = 64
data4.loc[data4['University'] == "University of Missouri", 'University'] = 139
data4.loc[data4['University'] == "University of south carolina", 'University']␣
↪= 104
```

```
data4.loc[data4['University'] == "University of Texas", 'University'] = 48
data4.loc[data4['University'] == "University of Wisconsin - Madison",␣
↪'University'] = 46
data4.loc[data4['University'] == "Virginia Commonwealth University",␣
↪"University"] = 162
data4.loc[data4['University'] == "Wright State University", "University"] = 293
data4.loc[data4['University'] == "Worcester Polytechnic Institute",␣
↪"University"] = 64
data4.loc[data4['University'] == "University of Nebraska Omaha", 'University']␣
↪= 293
data4.loc[data4['University'] == "Kansas State University", 'University'] = 162
data4.loc[data4['University'] == "TAMU, College Station", 'University'] = 70
data4.loc[data4['University'] == "University of Wisconsin Madison",␣
↪'University'] = 149
data4.loc[data4['University'] == "Georgia Institute of Technology",␣
↪'University'] = 29
data4.loc[data4['University'] == "University of California, Los Angeles",␣
↪"University"] = 17
data4.loc[data4['University'] == "University of Texas at Austin", "University"]␣
↪= 67
data4.loc[data4['University'] == "Oklahoma State University", "University"] =␣
↪192


data4.loc[data4['University'] == "San Jose State University", 'University'] = 24
data4.loc[data4['University'] == "University of South Florida", 'University'] =␣
↪104
data4.loc[data4['University'] == "University of Texas, Arlington",␣
↪'University'] = 293

data4.loc[data4['University'] == "San Jose State University", 'University'] = 24
data4.loc[data4['University'] == "University of South Florida", 'University'] =␣
↪104
data4.loc[data4['University'] == "University of Texas, Arlington",␣
↪'University'] = 293
data4.loc[data4['University'] == "University of California, Irvine",␣
↪'University'] = 36
data4.loc[data4['University'] == "Michigan technological University",␣
↪'University'] = 147
data4.loc[data4['University'] == "University Of South Florida", "University"] =␣
↪104
data4.loc[data4['University'] == "University of Texas at Austin", "University"]␣
↪= 67


data4.loc[data4['University'] == "Michigan Tech University", 'University'] = 24
```

```
data4.loc[data4['University'] == "Upenn", 'University'] = 104
data4.loc[data4['University'] == "Arizona state university", 'University'] = 293
data4.loc[data4['University'] == "Ohio state university", 'University'] = 36
data4.loc[data4['University'] == "Michigan technological university",␣
↪'University'] = 147
data4.loc[data4['University'] == "arizona state university", "University"] = 104
data4.loc[data4['University'] == "Michigan Tech University", "University"] = 67


data4.loc[data4['University'] == "SUNY,Binghamton", 'University'] = 24
data4.loc[data4['University'] == "Suny Buffalo", 'University'] = 104
data4.loc[data4['University'] == "Carnegie Mellon university", 'University'] =␣
↪293
data4.loc[data4['University'] == "Mich tech", 'University'] = 36
data4.loc[data4['University'] == "Missouri Science Tech", 'University'] = 147
data4.loc[data4['University'] == "Arizona State University ", "University"] =␣
↪104
data4.loc[data4['University'] == "bradley", "University"] = 67

data4.loc[data4['University'] == "Bradley Univ,Wright State University,Wichita␣
↪state university,California state university fullerton,Texas A&M␣
↪kingsville,", 'University'] = 24
data4.loc[data4['University'] == "Buffalo", 'University'] = 104
data4.loc[data4['University'] == "California State University - East bay",␣
↪'University'] = 293
data4.loc[data4['University'] == "Carnegie Melon University", 'University'] = 36
data4.loc[data4['University'] == "Clemson", 'University'] = 147
data4.loc[data4['University'] == "cleveland state university ", "University"] =␣
↪104
data4.loc[data4['University'] == "coloado school of mines", "University"] = 67
data4.loc[data4['University'] == "Colorado State University", 'University'] = 24
data4.loc[data4['University'] == "Colorado State University, Fort Collins",␣
↪'University'] = 104
data4.loc[data4['University'] == "Florida International University",␣
↪'University'] = 293
data4.loc[data4['University'] == "Florida state university", 'University'] = 36
data4.loc[data4['University'] == "George Washington University", 'University']␣
↪= 147
data4.loc[data4['University'] == "Georgia Tech", "University"] = 104
data4.loc[data4['University'] == "Illinios Institute of Technology, Chicago",␣
↪"University"] = 67



data4.loc[data4['University'] == "Indiana University, Bloomington",␣
↪'University'] = 79
```

```
data4.loc[data4['University'] == "kansas state university", 'University'] = 162
data4.loc[data4['University'] == "lamar university", 'University'] = 293
data4.loc[data4['University'] == "Louisiana tech university", 'University'] =␣
 ↪272
data4.loc[data4['University'] == "MIR ", "University"] = 141
data4.loc[data4['University'] == "North Carolina State", 'University'] = 84
data4.loc[data4['University'] == "Northeastern", 'University'] = 40
data4.loc[data4['University'] == "Northeastern University, Boston",␣
 ↪'University'] = 40
data4.loc[data4['University'] == "northwest missouri state university",␣
 ↪'University'] = 95
data4.loc[data4['University'] == "Oklahoma City University", 'University'] = 240
data4.loc[data4['University'] == "Oklahoma state university still water",␣
 ↪'University'] = 192
data4.loc[data4['University'] == "Purdue", "University"] = 57
data4.loc[data4['University'] == "Purdue University, West Lafayette",␣
 ↪"University"] = 57

data4.loc[data4['University'] == "Rochester Institute of Technology ",␣
 ↪'University'] = 104
data4.loc[data4['University'] == "Rutgers State University, New Brunswick",␣
 ↪'University'] = 62
data4.loc[data4['University'] == "Southern Illinois University Edwardsville",␣
 ↪'University'] = 293
data4.loc[data4['University'] == "SOUTHERN ILLINOIS UNIVERSITY, EDWARDSVILLE",␣
 ↪'University'] = 293
data4.loc[data4['University'] == "SUNY STONYBROOK ", "University"] = 91
data4.loc[data4['University'] == "SUNY, Binghamton", "University"] = 79
data4.loc[data4['University'] == "SUNY, Buffalo", 'University'] = 79
data4.loc[data4['University'] == "syracuse", 'University'] = 54
data4.loc[data4['University'] == "northwest missouri state university",␣
 ↪'University'] = 95
data4.loc[data4['University'] == "Oklahoma City University", 'University'] = 240
data4.loc[data4['University'] == "Oklahoma state university still water",␣
 ↪'University'] = 192
data4.loc[data4['University'] == "Purdue", "University"] = 57
data4.loc[data4['University'] == "Purdue University, West Lafayette",␣
 ↪"University"] = 57


data4.loc[data4['University'] == "University of California, San Diego ",␣
 ↪'University'] = 104
data4.loc[data4['University'] == "University of California, San diego",␣
 ↪'University'] = 293
data4.loc[data4['University'] == "University of California, Santa Barbara",␣
 ↪'University'] = 36
```

```
data4.loc[data4['University'] == "University of Chicago", 'University'] = 147
data4.loc[data4['University'] == "University of Cinncinati ", "University"] =␣
↪104
data4.loc[data4['University'] == "University of Colorado, Denver",␣
↪"University"] = 67
data4.loc[data4['University'] == "University Of Florida", 'University'] = 24
data4.loc[data4['University'] == "UUniversity of Illinois - Urbana Champagne",␣
↪'University'] = 104
data4.loc[data4['University'] == "University of Illinois Urbana Chqmpaign",␣
↪'University'] = 293
data4.loc[data4['University'] == "University of Illinois, Chicago",␣
↪'University'] = 36
data4.loc[data4['University'] == "University of Illinois,Urbana Champaign",␣
↪'University'] = 147
data4.loc[data4['University'] == "University of kentucky, lexington",␣
↪"University"] = 104
data4.loc[data4['University'] == "university of Louisville ", "University"] = 67


data4.loc[data4['University'] == "University of Mary Hardin-Baylor ",␣
↪'University'] = 54
data4.loc[data4['University'] == "University of Maryland Baltimore County",␣
↪'University'] = 166
data4.loc[data4['University'] == "University of Maryland College Park",␣
↪'University'] = 64
data4.loc[data4['University'] == "University of Maryland, Baltimore County",␣
↪'University'] = 166
data4.loc[data4['University'] == "University of Massacheusetts ", "University"]␣
↪= 74
data4.loc[data4['University'] == "University of Massachusetts, Amherst",␣
↪"University"] = 77
data4.loc[data4['University'] == "University of minnesota Duluth",␣
↪'University'] = 45
data4.loc[data4['University'] == "University of missouri, kansas city",␣
↪'University'] = 263
data4.loc[data4['University'] == "university of New Mexico", 'University'] = 218
data4.loc[data4['University'] == "University of North Carolina Charlotte",␣
↪'University'] = 228
data4.loc[data4['University'] == "University of Texas, Austin", 'University'] =␣
↪48
data4.loc[data4['University'] == "University of Texas, Tyler", "University"] =␣
↪293
data4.loc[data4['University'] == "University of toledo ", "University"] = 230
```

```
data4.loc[data4['University'] == "university of virginia ", 'University'] = 25
data4.loc[data4['University'] == "University of Washington, Seattle",␣
↪'University'] = 62
data4.loc[data4['University'] == "University of Wisconsin, Madison",␣
↪'University'] = 46
data4.loc[data4['University'] == "Verginiya Tech", 'University'] = 74
data4.loc[data4['University'] == "villanova ", "University"] = 50
data4.loc[data4['University'] == "Villanova University", "University"] = 50
data4.loc[data4['University'] == "Virginia Tech", 'University'] = 50
data4.loc[data4['University'] == "Washington State University", 'University'] =␣
↪143
data4.loc[data4['University'] == "wright state", 'University'] = 293
data4.loc[data4['University'] == "Wright state university", 'University'] = 293
data4.loc[data4['University'] == "Youngstown state university", 'University'] =␣
↪290

data4.loc[data4['University'] == "Rochester Institute of Technology ",␣
↪'University'] = 104
data4.loc[data4['University'] == "Rutgers State University, New Brunswick",␣
↪'University'] = 62
data4.loc[data4['University'] == "Southern Illinois University Edwardsville",␣
↪'University'] = 293
data4.loc[data4['University'] == "SOUTHERN ILLINOIS UNIVERSITY, EDWARDSVILLE",␣
↪'University'] = 147
data4.loc[data4['University'] == "SUNY STONYBROOK ", "University"] = 104
data4.loc[data4['University'] == "SUNY, Binghamton", "University"] = 79
data4.loc[data4['University'] == "SUNY, Buffalo", 'University'] = 79
data4.loc[data4['University'] == "syracuse", 'University'] = 54
data4.loc[data4['University'] == "TAMUCC", 'University'] = 293
data4.loc[data4['University'] == "Texas A&M University", 'University'] = 70
data4.loc[data4['University'] == "Texas A&M University, College Station",␣
↪'University'] = 70
data4.loc[data4['University'] == "Texas A&M University, Commerce",␣
↪"University"] = 293
data4.loc[data4['University'] == "Texas A&M University,College Station",␣
↪"University"] = 70


data4.loc[data4['University'] == "Texas Tech University ", 'University'] = 218
data4.loc[data4['University'] == "UC ( CINCINNATI)", 'University'] = 139
data4.loc[data4['University'] == "ucla", 'University'] = 17
data4.loc[data4['University'] == "UHMC", 'University'] = 185
data4.loc[data4['University'] == "UIUC ", "University"] = 69
data4.loc[data4['University'] == "University of MissouriÐKansas City",␣
↪"University"] = 100
data4.loc[data4['University'] == "University of Akron", 'University'] = 293
```

```
data4.loc[data4['University'] == "University of Alabama, Huntsville",␣
↪'University'] = 263
data4.loc[data4['University'] == "University of Californinia, Irvine",␣
↪'University'] = 36
data4.loc[data4['University'] == "University of Californinia, Riverside",␣
↪'University'] = 91
data4.loc[data4['University'] == "University of California at Berkeley",␣
↪'University'] = 22
data4.loc[data4['University'] == "University of California Berkeley",␣
↪"University"] = 22
data4.loc[data4['University'] == "University of California Santa Barbara",␣
↪"University"] = 34



data4.loc[data4['University'] == "University of California, San Diego ",␣
↪'University'] = 37
data4.loc[data4['University'] == "University of California, San diego",␣
↪'University'] = 37
data4.loc[data4['University'] == "University of California, Santa Barbara",␣
↪'University'] = 34
data4.loc[data4['University'] == "University of Chicago", 'University'] = 6
data4.loc[data4['University'] == "University of Cinncinati ", "University"] =␣
↪139
data4.loc[data4['University'] == "University of Colorado, Denver",␣
↪"University"] = 254
data4.loc[data4['University'] == "University Of Florida", 'University'] = 8
data4.loc[data4['University'] == "UUniversity of Illinois - Urbana Champagne",␣
↪'University'] = 69
data4.loc[data4['University'] == "University of Illinois Urbana Chqmpaign",␣
↪'University'] = 69
data4.loc[data4['University'] == "University of Illinois, Chicago",␣
↪'University'] = 132
data4.loc[data4['University'] == "University of Illinois,Urbana Champaign",␣
↪'University'] = 69
data4.loc[data4['University'] == "University of kentucky, lexington",␣
↪"University"] = 132
data4.loc[data4['University'] == "university of Louisville ", "University"] =␣
↪192

data4.loc[data4['University'] == "Illinois Institute of Technology, Chicago",␣
↪'University'] = 6
data4.loc[data4['University'] == "illinois institute of technology,chicago",␣
↪'University'] = 6
data4.loc[data4['University'] == "MIR", "University"] = 141
```

```
data4.loc[data4['University'] == "North Carolina State University, Raleigh",␣
↪'University'] = 6
data4.loc[data4['University'] == "SUNY STONYBROOK", 'University'] = 6
data4.loc[data4['University'] == "SUNY,Buffalo", 'University'] = 6
data4.loc[data4['University'] == "Texas Tech University", 'University'] = 6
data4.loc[data4['University'] == "Towson University", 'University'] = 6
data4.loc[data4['University'] == "UIUC", 'University'] = 6


data4.loc[data4['University'] == "University of California, Berkeley",␣
↪'University'] = 22
data4.loc[data4['University'] == "University of California, Riverside",␣
↪'University'] = 91
data4.loc[data4['University'] == "University of California, San Diego ",␣
↪"University"] = 37
data4.loc[data4['University'] == "University of Cinncinati", 'University'] = 139
data4.loc[data4['University'] == "University of Illinois - Urbana Champagne",␣
↪'University'] = 69
data4.loc[data4['University'] == "University of Mary Hardin-Baylor",␣
↪'University'] = 43
data4.loc[data4['University'] == "University of Massacheusetts", 'University']␣
↪= 64
data4.loc[data4['University'] == "University of Massachusetts, Amherst ",␣
↪'University'] = 64
data4.loc[data4['University'] == "University of Massachusetts, Lowell",␣
↪'University'] = 179

data4.loc[data4['University'] == "University of Massachusetts, worchester",␣
↪'University'] = 45
data4.loc[data4['University'] == "University of North Carolina, Charlette",␣
↪'University'] = 194
data4.loc[data4['University'] == "university of north carolina,charlotte ",␣
↪"University"] = 194
data4.loc[data4['University'] == "University of toledo", 'University'] = 230
data4.loc[data4['University'] == "university of virginia", 'University'] = 25
data4.loc[data4['University'] == "villanova", 'University'] = 111
data4.loc[data4['University'] == "Wright State University ", 'University'] = 293
data4.loc[data4['University'] == "Wright state University , Amherst ",␣
↪'University'] = 293


data4.loc[data4['University'] == "university of north carolina,charlotte",␣
↪'University'] = 25
data4.loc[data4['University'] == "University of California, San Diego",␣
↪'University'] = 111
data4.loc[data4['University'] == "Wright state University ", 'University'] = 293
```

```
data4.loc[data4['University'] == "Wright state University , Amherst ",␣
↪'University'] = 293
data4.loc[data4['University'] == "cleveland state university", 'University'] =␣
↪294
```

[95]:
```
data4.loc[data4['Major'] == "Management Information System", 'Major'] = 0
data4.loc[data4['Major'] == "Information Management and Systems", 'Major'] = 0
data4.loc[data4['Major'] == "Information Systems", 'Major'] = 0
data4.loc[data4['Major'] == "Information Science", 'Major'] = 0
data4.loc[data4['Major'] == "ICS with concentration in Informatics", 'Major'] =␣
↪0
data4.loc[data4['Major'] == "Information Technology and Management", 'Major'] =␣
↪0
data4.loc[data4['Major'] == "n major", 'Major'] = 0
data4.loc[data4['Major'] == "Information Management", 'Major'] = 0
data4.loc[data4['Major'] == "MS CS", 'Major'] = 1
data4.loc[data4['Major'] == "Mech", 'Major'] = 0
data4.loc[data4['Major'] == "Data science", 'Major'] = 1
data4.loc[data4['Major'] == "Business Analytics", 'Major'] = 0
data4.loc[data4['Major'] == "CS", 'Major'] = 1
data4.loc[data4['Major'] == "IE", 'Major'] = 0
data4.loc[data4['Major'] == "CSE", 'Major'] = 1
data4.loc[data4['Major'] == "SE", 'Major'] = 1
data4.loc[data4['Major'] == "CE", 'Major'] = 1
data4.loc[data4['Major'] == "Civil", 'Major'] = 0
data4.loc[data4['Major'] == "MIS", 'Major'] = 0
data4.loc[data4['Major'] == "ITM", 'Major'] = 0
data4.loc[data4['Major'] == "Industrial", 'Major'] = 0
data4.loc[data4['Major'] == "EE", 'Major'] = 1
data4.loc[data4['Major'] == "anthropology", 'Major'] = 0
data4.loc[data4['Major'] == "MSIS", 'Major'] = 0
data4.loc[data4['Major'] == "ECE", 'Major'] = 0
data4.loc[data4['Major'] == "Telecom", 'Major'] = 0
data4.loc[data4['Major'] == "Network and communication management", 'Major'] = 0
data4.loc[data4['Major'] == "petroleum engineering", 'Major'] = 0
data4.loc[data4['Major'] == "DA", 'Major'] = 0
data4.loc[data4['Major'] == "Cs", 'Major'] = 1
data4.loc[data4['Major'] == "Mechanical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "Mis", 'Major'] = 0
data4.loc[data4['Major'] == "MECH", 'Major'] = 0
data4.loc[data4['Major'] == "mis", 'Major'] = 0
data4.loc[data4['Major'] == "Mecahnical", 'Major'] = 0
data4.loc[data4['Major'] == "Data Analytics", 'Major'] = 1
data4.loc[data4['Major'] == "MCIT (CS)", 'Major'] = 1
data4.loc[data4['Major'] == "Business analytics", 'Major'] = 0
data4.loc[data4['Major'] == "Mechanical", 'Major'] = 0
data4.loc[data4['Major'] == "MSBAPM", 'Major'] = 0
```

```
data4.loc[data4['Major'] == "Data Science", 'Major'] = 1
data4.loc[data4['Major'] == "Construction engineering & management", 'Major'] =␣
↪0
data4.loc[data4['Major'] == "msba", 'Major'] = 0
data4.loc[data4['Major'] == "BIOM", 'Major'] = 0
data4.loc[data4['Major'] == "Engineering Management", 'Major'] = 0
data4.loc[data4['Major'] == "ME", 'Major'] = 0
data4.loc[data4['Major'] == "Materials Science", 'Major'] = 0
data4.loc[data4['Major'] == "MSIM", 'Major'] = 0
data4.loc[data4['Major'] == "Industrial Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "Data management and analytics - ITM", 'Major'] = 0
data4.loc[data4['Major'] == "MECHANICAL ENGINEERING", 'Major'] = 0
data4.loc[data4['Major'] == "Civil Engineering (Construction)", 'Major'] = 0
data4.loc[data4['Major'] == "Civil Environmental", 'Major'] = 0
data4.loc[data4['Major'] == "Mathematical Finance", 'Major'] = 0
data4.loc[data4['Major'] == "MSIT ebusiness", 'Major'] = 1
data4.loc[data4['Major'] == "CIS", 'Major'] = 1
data4.loc[data4['Major'] == "Computer Science", 'Major'] = 1
data4.loc[data4['Major'] == "Industrial ", 'Major'] = 0
data4.loc[data4['Major'] == "MS Mech", 'Major'] = 0
data4.loc[data4['Major'] == "Aerospace Engineering ", 'Major'] = 0
data4.loc[data4['Major'] == "ee", 'Major'] = 1
data4.loc[data4['Major'] == "mechanical", 'Major'] = 0
data4.loc[data4['Major'] == "masters", 'Major'] = 1
data4.loc[data4['Major'] == "PhD", 'Major'] = 1
data4.loc[data4['Major'] == "MS in CS", 'Major'] = 1
data4.loc[data4['Major'] == "MS in Computer Science", 'Major'] = 1
data4.loc[data4['Major'] == "Ms CEE", 'Major'] = 1
data4.loc[data4['Major'] == "PhD Robotics", 'Major'] = 1
data4.loc[data4['Major'] == "MS in Robotics", 'Major'] = 1
data4.loc[data4['Major'] == "MS in Biomedical", 'Major'] = 0
data4.loc[data4['Major'] == "Mechanical Engg.", 'Major'] = 0
data4.loc[data4['Major'] == "industrial system", 'Major'] = 0
data4.loc[data4['Major'] == "Mechanical ", 'Major'] = 0
data4.loc[data4['Major'] == "Computer Engineering", 'Major'] = 1
data4.loc[data4['Major'] == "MS Industrial Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "MS EE", 'Major'] = 1
data4.loc[data4['Major'] == "Electrical Enginnering", 'Major'] = 1
data4.loc[data4['Major'] == "MS Mechanical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "MS in Civil", 'Major'] = 0
data4.loc[data4['Major'] == "metallurgy, phd", 'Major'] = 0
data4.loc[data4['Major'] == "Microbiology Immunology and Pathology", 'Major'] =␣
↪0
data4.loc[data4['Major'] == "PhD in Computational Biology", 'Major'] = 0
data4.loc[data4['Major'] == "MS Earth and Environmental Engineering", 'Major']␣
↪= 0
data4.loc[data4['Major'] == "PhD (Mechanical)", 'Major'] = 0
```

```
data4.loc[data4['Major'] == "MS-IT", 'Major'] = 1
data4.loc[data4['Major'] == "industrial", 'Major'] = 0
data4.loc[data4['Major'] == "MS in ECE/Computer Engnn", 'Major'] = 1
data4.loc[data4['Major'] == "ECE/Computer Engg", 'Major'] = 1
data4.loc[data4['Major'] == "M.S Aeronautics", 'Major'] = 0
data4.loc[data4['Major'] == "M S in Mechanical", 'Major'] = 0
data4.loc[data4['Major'] == "MS in Electrical and Computer Engineering",␣
 ↪'Major'] = 1
data4.loc[data4['Major'] == "MS", 'Major'] = 1
data4.loc[data4['Major'] == "MS in Manufacturing Engineering", 'Major'] = 1
data4.loc[data4['Major'] == "MS In Computer Science", 'Major'] = 1
data4.loc[data4['Major'] == "MS MEch", 'Major'] = 0
data4.loc[data4['Major'] == "cs", 'Major'] = 1
data4.loc[data4['Major'] == "MS ITM", 'Major'] = 0
data4.loc[data4['Major'] == "Industrial Enginnering", 'Major'] = 0
data4.loc[data4['Major'] == "ms (system science)", 'Major'] = 0
data4.loc[data4['Major'] == "fv", 'Major'] = 0
data4.loc[data4['Major'] == "Mechanical Engg", 'Major'] = 0
data4.loc[data4['Major'] == "MS in Mechanical", 'Major'] = 0
data4.loc[data4['Major'] == "Electrical", 'Major'] = 0
data4.loc[data4['Major'] == "Electrical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "Mechnical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "MS-EE", 'Major'] = 0
data4.loc[data4['Major'] == "MSCS", 'Major'] = 1
data4.loc[data4['Major'] == "MS in Chemistry", 'Major'] = 0
data4.loc[data4['Major'] == "MS-Mechanical", 'Major'] = 0
data4.loc[data4['Major'] == "civil engineering structural", 'Major'] = 0
data4.loc[data4['Major'] == "CS CN", 'Major'] = 0
data4.loc[data4['Major'] == "MS Industrial and Systems Engg", 'Major'] = 0
data4.loc[data4['Major'] == "MS in Mech Engg", 'Major'] = 0
data4.loc[data4['Major'] == "Masters in Analytics", 'Major'] = 0
data4.loc[data4['Major'] == "computer science", 'Major'] = 1
data4.loc[data4['Major'] == "MS Computer Science", 'Major'] = 1
data4.loc[data4['Major'] == "Mechanical MS", 'Major'] = 0
data4.loc[data4['Major'] == "ce", 'Major'] = 1
data4.loc[data4['Major'] == "M.S", 'Major'] = 1
data4.loc[data4['Major'] == "electrical(VLSI)", 'Major'] = 0
data4.loc[data4['Major'] == "MS in Mechanical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "Aerospace", 'Major'] = 0
data4.loc[data4['Major'] == "ITWS", 'Major'] = 0
data4.loc[data4['Major'] == "Software Engineering", 'Major'] = 1
data4.loc[data4['Major'] == "Information Sciences and Technology", 'Major'] = 0
data4.loc[data4['Major'] == "Industrial Eng.", 'Major'] = 0
data4.loc[data4['Major'] == "MS in EE", 'Major'] = 1
data4.loc[data4['Major'] == "MS ECE", 'Major'] = 1
data4.loc[data4['Major'] == "Bach in IT", 'Major'] = 1
data4.loc[data4['Major'] == "Civil Engineering", 'Major'] = 0
```

```
data4.loc[data4['Major'] == "Operations Research", 'Major'] = 0
data4.loc[data4['Major'] == "Electrical and Electronics Engineering", 'Major']␣
↪= 0
data4.loc[data4['Major'] == "MS IS", 'Major'] = 0
data4.loc[data4['Major'] == "MS in SUpply chain and Operations management",␣
↪'Major'] = 0
data4.loc[data4['Major'] == "ELE", 'Major'] = 1
data4.loc[data4['Major'] == "Engineering management", 'Major'] = 0
data4.loc[data4['Major'] == "MS in Information Management", 'Major'] = 0
data4.loc[data4['Major'] == "MS in Biomedical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "MSEE", 'Major'] = 1
data4.loc[data4['Major'] == "Elecreical", 'Major'] = 1
data4.loc[data4['Major'] == "MECHANICAL ENGINEERING( MSME)", 'Major'] = 0
data4.loc[data4['Major'] == "ms in ce", 'Major'] = 1
data4.loc[data4['Major'] == "MS in Mech", 'Major'] = 0
data4.loc[data4['Major'] == "Ms Cs", 'Major'] = 1
data4.loc[data4['Major'] == "Ms is", 'Major'] = 0
data4.loc[data4['Major'] == "Construction Management", 'Major'] = 0
data4.loc[data4['Major'] == "Construction management", 'Major'] = 0
data4.loc[data4['Major'] == "MS - Electrical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "PhD-Mechanical", 'Major'] = 0
data4.loc[data4['Major'] == "Mechanical M.S.", 'Major'] = 0
data4.loc[data4['Major'] == "MS Electrical and Computer Engineering", 'Major']␣
↪= 1
data4.loc[data4['Major'] == "industrial engg", 'Major'] = 0
data4.loc[data4['Major'] == "MIM", 'Major'] = 0
data4.loc[data4['Major'] == "computer enginneering.VLSI", 'Major'] = 1
data4.loc[data4['Major'] == "Embedded Systems", 'Major'] = 0
data4.loc[data4['Major'] == "MS - EE", 'Major'] = 1
data4.loc[data4['Major'] == "Microbiology", 'Major'] = 0
data4.loc[data4['Major'] == "Networking", 'Major'] = 0
data4.loc[data4['Major'] == "Computer science", 'Major'] = 1
data4.loc[data4['Major'] == "MS IT", 'Major'] = 1
data4.loc[data4['Major'] == "Computer Science ", 'Major'] = 1
data4.loc[data4['Major'] == "EMBS", 'Major'] = 0
data4.loc[data4['Major'] == "cggt/cs", 'Major'] = 0
data4.loc[data4['Major'] == "Chemical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "Electronics and Computer Engineering", 'Major'] = 1
data4.loc[data4['Major'] == "CS(MS)", 'Major'] = 1
data4.loc[data4['Major'] == "mech", 'Major'] = 0
data4.loc[data4['Major'] == "MS CS Game Development", 'Major'] = 1
data4.loc[data4['Major'] == "Information System", 'Major'] = 0
data4.loc[data4['Major'] == "Telecom Engg", 'Major'] = 0
data4.loc[data4['Major'] == "MSTE", 'Major'] = 0
data4.loc[data4['Major'] == "Telecommunication", 'Major'] = 0
data4.loc[data4['Major'] == "EEE", 'Major'] = 1
data4.loc[data4['Major'] == "MS Electrical Engineering", 'Major'] = 1
```

```
data4.loc[data4['Major'] == "masters in cs ", 'Major'] = 1
data4.loc[data4['Major'] == "cs masters ", 'Major'] = 1
data4.loc[data4['Major'] == "MS in Computing", 'Major'] = 1
data4.loc[data4['Major'] == "Chemical eng.", 'Major'] = 0
data4.loc[data4['Major'] == "PhD, Computer Science", 'Major'] = 1
data4.loc[data4['Major'] == "Engg. Mechanics (PhD)", 'Major'] = 0
data4.loc[data4['Major'] == "MS-ECE (VLSI)", 'Major'] = 0
data4.loc[data4['Major'] == "MS electrical - VLSI", 'Major'] = 0
data4.loc[data4['Major'] == "MS in Computer engineering", 'Major'] = 1
data4.loc[data4['Major'] == "Ph.D. in Chemical Engineering", 'Major'] = 0
data4.loc[data4['Major'] == "MS IN MECH ENGG", 'Major'] = 0
data4.loc[data4['Major'] == "MS Computer Engineering", 'Major'] = 1
data4.loc[data4['Major'] == "CS", 'Major'] = 1
data4.loc[data4['Major'] == "Telecommunication Engineering", 'Major'] = 1
data4.loc[data4['Major'] == "CS ", 'Major'] = 1
data4.loc[data4['Major'] == "n Management Information System", 'Major'] = 0
data4.loc[data4['Major'] == "MS-IS", 'Major'] = 0
```

[99]:
```
data4
```

[99]:
```
     University Major  Decision  GRE_SCORE  GRE Quant  GRE Verbal work_ex  \
2            50     0         1      315.0      158.0       157.0       3
3            50     0         1      301.0      151.0       150.0      60
4            50     0         1      311.0      156.0       155.0      25
5            50     0         1      312.0      156.0       156.0      24
6            50     0         1      307.0      154.0       153.0      34
..          …   …       …         …          …          …      …
187          54     0         1      320.0      165.0       155.0       0
188          54     0         1      320.0      165.0       155.0       0
192         147     1         1      329.0      170.0       159.0       0
195          74     0         1      301.0      159.0       142.0       3
196          19     1         1      325.0      170.0       155.0       2

                     Name     GPA  Language Proficiency
2               bostonner  77.000                   7.5
3              KrithikaCT  80.120                   0.0
4             Giridhar312  66.954                   7.0
5              Shreepriya  73.060                   8.0
6                 muktika  68.800                   7.5
..                    …     …                     …
187                   AVS  77.320                   7.5
188                   AVS  77.320                   7.5
192  Yash Prakash Pradhan  76.894                   8.0
195             Balaji K   58.505                   6.5
196          Sakshi Gopal  65.960                   8.0

[1491 rows x 10 columns]
```

[97]:
```
data4.to_csv(r'/Users/aminameghezzi/Downloads/Clean_data.csv')
```

[ ]:
```

```