Data cleaning - Students' data

Data cleaning steps that we are going to follow:

- 1. Deleting redundant columns
- 2. Renaming columns
- 3. Dropping duplicates
- 4. Remove NaN(null) values from the dataset
- 5. Check for some more transformations

```
In [1]: import pandas as pd
In [2]: pd. version
Out[2]:
                      '2.2.3'
                       data = pd.read \ csv(r"C:\Users\balus\OneDrive\Desktop\Projects\Data \ cleaning - Pandas-student \ info\Students \ info\Control \ info\Con
In [4]:
                      data
In [5]:
                                       Unnamed:
                                                                Gender EthnicGroup
                                                                                                                 ParentEduc
                                                                                                                                               LunchType
                                                                                                                                                                           TestPrep ParentMaritalStatus PracticeSport IsFirstChild
                                                         0
                                                                                                                      bachelor's
                                0
                                                         0
                                                                                                     NaN
                                                                                                                                                     standard
                                                                                                                                                                                                                             married
                                                                  female
                                                                                                                                                                                    none
                                                                                                                                                                                                                                                           regularly
                                                                                                                                                                                                                                                                                                 yes
                                                                                                                            degree
                                                                                                                               some
                                                                  female
                                                                                              group C
                                                                                                                                                      standard
                                                                                                                                                                                      NaN
                                                                                                                                                                                                                             married
                                                                                                                                                                                                                                                       sometimes
                                                                                                                                                                                                                                                                                                 yes
                                                                                                                            college
                                                                                                                         master's
                                2
                                                         2
                                                                  female
                                                                                              group B
                                                                                                                                                     standard
                                                                                                                                                                                    none
                                                                                                                                                                                                                                single
                                                                                                                                                                                                                                                       sometimes
                                                                                                                                                                                                                                                                                                 yes
                                                                                                                            degree
                                                                                                                    associate's
                                3
                                                         3
                                                                     male
                                                                                              group A
                                                                                                                                              free/reduced
                                                                                                                                                                                                                             married
                                                                                                                                                                                    none
                                                                                                                                                                                                                                                                never
                                                                                                                                                                                                                                                                                                  no
                                                                                                                            degree
                                                                                                                               some
                                4
                                                         4
                                                                                                                                                     standard
                                                                     male
                                                                                              group C
                                                                                                                                                                                                                             married
                                                                                                                                                                                                                                                       sometimes
                                                                                                                                                                                                                                                                                                 yes
                                                                                                                                                                                    none
                                                                                                                            college
                      30636
                                                    816
                                                                  female
                                                                                                                   high school
                                                                                                                                                     standard
                                                                                              group D
                                                                                                                                                                                    none
                                                                                                                                                                                                                                single
                                                                                                                                                                                                                                                      sometimes
                                                                                                                                                                                                                                                                                                  no
                      30637
                                                    890
                                                                                                                                                     standard
                                                                     male
                                                                                              group E
                                                                                                                   high school
                                                                                                                                                                                    none
                                                                                                                                                                                                                                single
                                                                                                                                                                                                                                                          regularly
                                                                                                                                                                                                                                                                                                  no
                      30638
                                                    911
                                                                  female
                                                                                                    NaN
                                                                                                                   high school
                                                                                                                                              free/reduced completed
                                                                                                                                                                                                                             married
                                                                                                                                                                                                                                                       sometimes
                                                                                                                                                                                                                                                                                                  no
                                                                                                                    associate's
                      30639
                                                    934
                                                                                              group D
                                                                                                                                                     standard completed
                                                                                                                                                                                                                             married
                                                                                                                                                                                                                                                          regularly
                                                                  female
                                                                                                                                                                                                                                                                                                  no
                                                                                                                            degree
                                                                                                                               some
                      30640
                                                    960
                                                                                                                                                      standard
                                                                                                                                                                                                                             married
                                                                     male
                                                                                              group B
                                                                                                                                                                                    none
                                                                                                                                                                                                                                                                never
                                                                                                                                                                                                                                                                                                  no
                                                                                                                            college
                    30641 rows × 15 columns
                   data.shape ## Data contains 15 columns & 30641 rows
Out[]: (30641, 15)
In []: data.columns ## These are the names of the 15 columns
'NrSiblings', 'TransportMeans', 'WklyStudyHours', 'MathScore',
                                         'ReadingScore', 'WritingScore'],
                                      dtype='object')
In [ ]: data.head() ## This shows the first 5 rows along with column names
```

```
0
                                                bachelor's
          0
                      0
                          female
                                         NaN
                                                              standard
                                                                           none
                                                                                             married
                                                                                                          regularly
                                                                                                                          yes
                                                                                                                                      3.0
                                                   degree
                                                    some
                          female
                                      group C
                                                              standard
                                                                            NaN
                                                                                             married
                                                                                                        sometimes
                                                                                                                                      0.0
                                                                                                                          yes
                                                   college
                                                  master's
          2
                     2
                          female
                                      group B
                                                              standard
                                                                           none
                                                                                              single
                                                                                                        sometimes
                                                                                                                          yes
                                                                                                                                      4.0
                                                   dearee
                                                associate's
          3
                     3
                                                                                                                                      1.0
                           male
                                      group A
                                                           free/reduced
                                                                           none
                                                                                             married
                                                                                                            never
                                                                                                                           no
                                                   degree
                                                    some
          4
                      4
                                                              standard
                                                                                                                                      0.0
                                                                                             married
                                                                                                        sometimes
                           male
                                      aroup C
                                                                                                                          ves
                                                                           none
                                                   college
          4
 In [ ]: data.info() ## It is showing datatypes, no of non-null entries & various other info about each column
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 30641 entries, 0 to 30640
         Data columns (total 15 columns):
          #
              Column
                                      Non-Null Count
                                                        Dtype
          0
              Unnamed: 0
                                      30641 non-null
                                                         int64
                                      30641 non-null
               Gender
          1
                                                         object
          2
              EthnicGroup
                                      28801 non-null
                                                         obiect
          3
              ParentEduc
                                      28796 non-null
                                                         object
          4
              LunchType
                                      30641 non-null
                                                         obiect
          5
              TestPrep
                                       28811 non-null
                                                         object
              ParentMaritalStatus
          6
                                      29451 non-null
                                                         object
          7
                                      30010 non-null
              PracticeSport
                                                         object
          8
              IsFirstChild
                                      29737 non-null
                                                         object
          9
              NrSiblings
                                      29069 non-null
                                                         float64
                                      27507 non-null
          10
              TransportMeans
                                                         object
          11
              WklyStudyHours
                                      29686 non-null
              MathScore
                                      30641 non-null
          12
                                                         int64
          13
              ReadingScore
                                      30641 non-null
                                                         int64
          14
                                      30641 non-null
              WritingScore
                                                         int64
         dtypes: float64(1), int64(4), object(10)
         memory usage: 3.5+ MB
In [12]: '''1) Deleting redundant columns
           'Unnamed: 0','TestPrep', 'ParentMaritalStatus', 'PracticeSport', 'IsFirstChild' -
          These are the redundant columns (not so important columns) that we are deleting'''
          data.drop(columns=['Unnamed: 0','TestPrep', 'ParentMaritalStatus', 'PracticeSport', 'IsFirstChild'],inplace=Truc
In [13]: data
Out[13]:
                  Gender
                          EthnicGroup
                                       ParentEduc
                                                    LunchType NrSiblings TransportMeans WklyStudyHours
                                                                                                             MathScore
                                                                                                                        ReadingScore
                                         bachelor's
               0
                                  NaN
                                                       standard
                                                                       3.0
                                                                                                         < 5
                                                                                                                     71
                                                                                                                                    71
                  female
                                                                                 school_bus
                                            degree
                                             some
                   female
                               group C
                                                       standard
                                                                       0.0
                                                                                       NaN
                                                                                                      5 - 10
                                                                                                                     69
                                                                                                                                    90
                                            college
                                           master's
                                                                                                                     87
               2
                  female
                               group B
                                                       standard
                                                                       4.0
                                                                                 school_bus
                                                                                                         < 5
                                                                                                                                    93
                                            degree
                                         associate's
               3
                                                                       1.0
                                                                                       NaN
                                                                                                      5 - 10
                                                                                                                     45
                                                                                                                                    56
                    male
                               group A
                                                    free/reduced
                                            degree
                                             some
                                                                       0.0
                                                                                                      5 - 10
                                                                                                                     76
                                                                                                                                    78
                    male
                               group C
                                                       standard
                                                                                 school bus
                                            college
                                                                       20
                                                                                                      5 - 10
                                                                                                                     59
                                                                                                                                    61
          30636
                  female
                               group D
                                        high school
                                                       standard
                                                                                 school bus
          30637
                    male
                                        high school
                                                       standard
                                                                       1.0
                                                                                                      5 - 10
                                                                                                                     58
                                                                                                                                    53
                               group E
                                                                                     private
          30638
                                        high school
                                                    free/reduced
                                                                       1.0
                                                                                                                     61
                                                                                                                                    70
                  female
                                  NaN
                                                                                     private
                                                                                                      5 - 10
                                         associate's
          30639
                   female
                               group D
                                                       standard
                                                                       3.0
                                                                                 school_bus
                                                                                                      5 - 10
                                                                                                                     82
                                                                                                                                    90
                                            degree
                                             some
          30640
                                                       standard
                                                                       1.0
                                                                                 school_bus
                                                                                                      5 - 10
                                                                                                                     64
                                                                                                                                    60
                    male
                               group B
                                            college
          30641 rows × 10 columns
          '''2) Renaming columns, we are going to rename column NrSiblings as Numberofsiblings '''
In [17]:
```

data.rename(columns={"NrSiblings":"Noofsiblings"},inplace=True)

Gender EthnicGroup ParentEduc LunchType TestPrep ParentMaritalStatus PracticeSport IsFirstChild NrSiblings

Unnamed:

	data									
8]:		Gender	EthnicGroup	ParentEduc	LunchType	Noofsiblings	TransportMeans	WklyStudyHours	MathScore	ReadingScore
	0	female	NaN	bachelor's degree	standard	3.0	school_bus	< 5	71	71
	1	female	group C	some college	standard	0.0	NaN	5 - 10	69	90
	2	female	group B	master's degree	standard	4.0	school_bus	< 5	87	93
	3	male	group A	associate's degree	free/reduced	1.0	NaN	5 - 10	45	56
	4	male	group C	some college	standard	0.0	school_bus	5 - 10	76	78
	30636	female	group D	high school	standard	2.0	school_bus	5 - 10	59	61
	30637	male	group E	high school	standard	1.0	private	5 - 10	58	53
	30638	female	NaN	high school	free/reduced	1.0	private	5 - 10	61	70
	30639	female	group D	associate's degree	standard	3.0	school_bus	5 - 10	82	90
	30640	male	group B	some college	standard	1.0	school_bus	5 - 10	64	60
3	30641 rd	ows × 10	columns							
	4									
		Droppin	α duplicates	- we are o	noina to che	ck how many	duplicate reco	rds are there &	delete th	em'''
	data.d	luplicat		we are g	joing to the	zek now marry	dupticate reco	rus are there o	t detete th	···
	np.in	luplicat		we are g	Joing to the	cer now many	dupered to reco	Tus are there o	. detec en	
	np.in	uplicat						rus ure enere o		-
	np.in	luplicat t64(9) luplicat	ed()					rus ure enere o		
	np.int	luplicat t64(9) luplicat t64(9)	ed()	# There are	e 9 duplica	te rows in t	he data	rus ure enere o		
	np.int	luplicat t64(9) luplicat t64(9) Irop_dup	ed() ed().sum() #	# There are ace= True) #	e 9 duplicat	te rows in to	he data	Tus are there o		
	np.int	luplicat t64(9) luplicat t64(9) Irop_dup	ed() ed().sum() # licates(inpl	# There are ace= True) #	e 9 duplicat	te rows in to	he data	Tus ute there o		
	np.intdata.ddata.ddata.ddata.ddata.dd	duplicat t64(9) duplicat t64(9) drop_dup duplicat t64(0) Remove	ed().sum() # licates(inpl ed().sum() # NaN(null) va	<pre># There are ace=True) # # now there lues from t</pre>	## deleting ## are no dup	those duplicate rows	he data cate rows ing to find Nul	l values & remo	ove them'''	
	np.intdata.ddata.ddata.ddata.ddata.dd	duplicat t64(9) duplicat t64(9) drop_dup duplicat t64(0) Remove sna() #	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr	# There are ace=True) # # now there lues from t ue wherever	## deleting e are no dup the dataset there is n	those duplicate rows - we are goonull values of	he data cate rows ing to find Nul & false whereve	l values & remo	ove them''' l values	
	np.intdata.ddata.ddata.ddata.ddata.dd	duplicat t64(9) duplicat t64(9) drop_dup duplicat t64(0) Remove sna() #	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr	# There are ace=True) # # now there lues from t ue wherever	## deleting e are no dup the dataset there is n	those duplicate rows - we are goonull values of	he data cate rows ing to find Nul & false whereve	l values & remo r there non-nul	ove them''' l values	
	np.intdata.ddata.ddata.ddata.ddata.dd	duplicat t64(9) duplicat t64(9) drop_dup duplicat t64(0) Remove sna() # Gender	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup	# There are ace=True) # # now there lues from t ue wherever ParentEduc	## deleting e are no dup the dataset there is n	those duplicate rows - we are good and a values of	he data cate rows ing to find Nul \$\false \text{ false whereve}\$ TransportMeans	l values & remo r there non-nul WklyStudyHours	ove them''' l values MathScore	ReadingScore
	np.infddata.dddata.dddata.ddnp.infddata.idddata.id	duplicat t64(9) duplicat t64(9) drop_dup duplicat t64(0) Remove sna() # Gender False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True	# There are ace=True) # # now there lues from t we wherever ParentEduc False	## deleting e are no dup the dataset there is n LunchType False	those duplicate rows - we are gooull values of Noofsiblings False	ing to find Nul & false whereve TransportMeans	l values & remo r there non-nul WklyStudyHours False	ove them''' ll values MathScore False	ReadingScore False
	np.intdata.ddata.ddata.ddata.ddata.ddata.idd	luplicat t64(9) luplicat t64(9) lrop_dup luplicat t64(0) Remove .sna() # Gender False False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False	# There are ace=True) # # now there lues from t ue wherever ParentEduc False False	## deleting e are no dup the dataset there is no LunchType False False	those duplicate rows - we are good ull values of the Noofsiblings False False	ting to find Nul false whereve TransportMeans False True	l values & remo r there non-nul WklyStudyHours False False	ove them''' l values MathScore False False	ReadingScore False False
	np.intdata.ddata.ddata.ddata.ddata.id	duplicat t64(9) duplicat t64(9) drop_dup duplicat t64(0) Remove sna() # Gender False False False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False False	# There are ace=True) # # now there lues from t we wherever ParentEduc False False False	## deleting e are no dup the dataset there is no LunchType False False False	those duplicate rows - we are good ull values of the valu	ing to find Nul S false whereve TransportMeans False True False	l values & remo r there non-nul WklyStudyHours False False	we them''' l values MathScore False False False	ReadingScore False False False
	np.infdata.ddnp.infdata.ddnp.infdata.ddnp.infdata.id	luplicat t64(9) luplicat t64(9) lrop_dup luplicat t64(0) Remove .sna() # Gender False False False False False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False False False False	# There are ace=True) # # now there lues from t ue wherever ParentEduc False False False False False False	the dataset there is n LunchType False False False False False False	those duplicate rows - we are goodull values of the value of the va	ing to find Nul false whereve TransportMeans False True False True False	l values & remor there non-nul WklyStudyHours False False False False False	MathScore False False False False False	ReadingScore False False False False
	np.infddata.ddnp.infddata.ddnp.infddata.idda	luplicat t64(9) luplicat t64(9) lrop_dup luplicat t64(0) Remove sna() # Gender False False False False False False False False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False False False False False False	# There are ace=True) # # now there lues from t we wherever ParentEduc False	the dataset there is r LunchType False	those duplicate rows - we are goodult values of the value	ing to find Nul S false whereve TransportMeans False True False True False	l values & remo r there non-nul WklyStudyHours False False False False False	we them''' l values MathScore False False False False False False False False False	ReadingScore False False False False False False
	np.intdata.ddata.ddata.ddata.ddata.id	luplicat t64(9) luplicat t64(9) lrop_dup luplicat t64(0) Remove sna() # Gender False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False False False False False False False False False	# There are ace=True) # # now there lues from t ue wherever ParentEduc False	the dataset there is r LunchType False	those duplicate rows - we are good and a values of the second se	ing to find Nul false whereve TransportMeans False True False True False False	l values & remor there non-nul WklyStudyHours False False False False False False False False	MathScore False False False False False False False False False	ReadingScore False False False False False False False
	np.infddata.ddnp.infddata.ddnp.infddata.idda	luplicat t64(9) luplicat t64(9) lrop_dup luplicat t64(0) Remove .sna() # Gender False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False	# There are ace=True) # # now there lues from t we wherever ParentEduc False	the dataset there is r LunchType False	those duplicate rows - we are gonull values of the values	ing to find Nul S false whereve TransportMeans False True False True False False False False False False	l values & remor there non-nul WklyStudyHours False	MathScore False	ReadingScore False False False False False False False False False
	np.intdata.ddata.ddata.ddata.ddata.ddata.ddata.ddata.iddat	luplicat t64(9) luplicat t64(9) lrop_dup luplicat t64(0) Remove .sna() # Gender False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False	# There are ace=True) # # now there lues from t ue wherever ParentEduc False	the dataset there is n LunchType False	those duplicate rows - we are goodult values of the value	ing to find Nul false whereve TransportMeans False True False True False False False False False False False False	l values & remor there non-nul WklyStudyHours False	MathScore False	ReadingScore False
	np.intdata.ddata.ddata.ddata.ddata.id	luplicat t64(9) luplicat t64(9) lrop_dup luplicat t64(0) Remove sna() # Gender False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False True	# There are ace=True) # # now there lues from t we wherever ParentEduc False	the dataset there is no large the false	those duplicate rows - we are gonull values of the value of the val	ing to find Nul false whereve TransportMeans False True False False	l values & remo r there non-nul WklyStudyHours False	MathScore False	ReadingScore False
	np.intdata.ddata.ddata.ddata.ddata.ddata.ddata.ddata.iddat	luplicat t64(9) luplicat t64(9) lrop_dup luplicat t64(0) Remove .sna() # Gender False	ed().sum() # licates(inpl ed().sum() # NaN(null) va # Showing Tr EthnicGroup True False	# There are ace=True) # # now there lues from t ue wherever ParentEduc False	the dataset there is n LunchType False	those duplicate rows - we are goodult values of the value	ing to find Nul false whereve TransportMeans False True False True False False False False False False False False	l values & remor there non-nul WklyStudyHours False	MathScore False	ReadingScore False

In [29]: data.isna().sum() ## Shows No of null values column wise

```
Out[29]: Gender
                                  0
          {\sf EthnicGroup}
                              1840
          ParentEduc
                              1845
          LunchType
                                  0
                              1571
          Noofsiblings
           TransportMeans
                              3134
          WklyStudyHours
                               955
          MathScore
                                  0
          ReadingScore
                                  0
          WritingScore
                                  0
          dtype: int64
In [33]: data.isna().sum().sum() ##Shows total No of null values across all columns
Out[33]: np.int64(9345)
 In [ ]: data.dropna(inplace=True) ## removing null records & saving it
 In [ ]: data.isna().sum().sum() ## now there are no null values
 Out[]: np.int64(0)
 In []: ## 5) Check for some more transformations
          data["Gender"].unique() ## checking unique entries in Gender column
 Out[]: array(['female', 'male'], dtype=object)
 In [ ]: ## Changing the gender values to 1 for female & 0 for male
          data["Gender"] = data["Gender"].apply(lambda x:1 if x == "female" else 0)
 In [ ]: data ## we can see Index is not right order
                 Gender EthnicGroup ParentEduc LunchType Noofsiblings TransportMeans WklyStudyHours MathScore ReadingScore
 Out[]:
                                          master's
              2
                       1
                              group B
                                                      standard
                                                                       4.0
                                                                                 school bus
                                                                                                        < 5
                                                                                                                    87
                                                                                                                                  93
                                           dearee
                                            some
              4
                       0
                                                      standard
                                                                       0.0
                                                                                                      5 - 10
                                                                                                                    76
                                                                                                                                  78
                              group C
                                                                                 school bus
                                           college
                                        associate's
              5
                       1
                              group B
                                                      standard
                                                                       1.0
                                                                                 school bus
                                                                                                      5 - 10
                                                                                                                    73
                                                                                                                                  84
                                           degree
                                            some
                                                      standard
                                                                       1.0
                                                                                                      5 - 10
                                                                                                                    85
                                                                                                                                  93
                              group B
                                                                                    private
                                           college
                                            some
              7
                       0
                              group B
                                                  free/reduced
                                                                       1.0
                                                                                    private
                                                                                                       > 10
                                                                                                                    41
                                                                                                                                  43
                                           college
                                        associate's
          30634
                       0
                                                                                                                    65
                              group A
                                                  free/reduced
                                                                       2.0
                                                                                 school_bus
                                                                                                      5 - 10
                                                                                                                                  60
                                           degree
                                            some
          30635
                       0
                              group C
                                                      standard
                                                                       2.0
                                                                                 school_bus
                                                                                                      5 - 10
                                                                                                                    58
                                                                                                                                  53
                                           college
          30636
                                                                       2.0
                                                                                                      5 - 10
                                                                                                                    59
                                                                                                                                  61
                       1
                              group D
                                       high school
                                                      standard
                                                                                 school_bus
          30637
                       0
                                       high school
                                                      standard
                                                                                    private
                                                                                                      5 - 10
                                                                                                                    58
                              group E
                                                                       10
                                                                                                                                  53
                                        associate's
          30639
                       1
                                                      standard
                                                                       3.0
                                                                                 school_bus
                                                                                                      5 - 10
                                                                                                                    82
                                                                                                                                  90
                              group D
                                           degree
         22343 rows × 10 columns
In [43]: data.reset_index(inplace=True)
In [42]: data
```

		Gender	EthnicGroup	ParentEduc	LunchType	Noofsiblings	TransportMeans	WklyStudyHours	MathScore	ReadingScore
	2	1	group B	master's degree	standard	4.0	school_bus	< 5	87	93
	4	0	group C	some college	standard	0.0	school_bus	5 - 10	76	78
	5	1	group B	associate's degree	standard	1.0	school_bus	5 - 10	73	84
	6	1	group B	some college	standard	1.0	private	5 - 10	85	93
	7	0	group B	some college	free/reduced	1.0	private	> 10	41	43
	30634	0	group A	associate's degree	free/reduced	2.0	school_bus	5 - 10	65	60
	30635	0	group C	some college	standard	2.0	school_bus	5 - 10	58	53
	30636	1	group D	high school	standard	2.0	school_bus	5 - 10	59	61
	30637	0	group E	high school	standard	1.0	private	5 - 10	58	53
	30639	1	group D	associate's degree	standard	3.0	school_bus	5 - 10	82	90
2	22343 rd	ows × 10	columns							
	4									þ.
:	data.d	lrop(col	umns="index"	,inplace=Tr	rue)					
]:	data									
]:		Gender	EthnicGroup	ParentEduc	LunchType	Noofsiblings	TransportMeans	WklyStudyHours	MathScore	ReadingScore
	0	1	group B	master's degree	standard	4.0	school_bus	< 5	87	93
	1	0	group C	some college	standard	0.0	school_bus	5 - 10	76	78
	2	0	group C		standard	0.0	school_bus	5 - 10 5 - 10	76 73	78 84
				college associate's			_			
	2	1	group B	college associate's degree some	standard	1.0	school_bus	5 - 10	73	84
	2	1	group B	college associate's degree some college some	standard	1.0	school_bus	5 - 10 5 - 10	73 85	84 93
	2 3 4	1 1 0	group B group B	college associate's degree some college some college	standard	1.0 1.0 1.0	school_bus private private	5 - 10 5 - 10 > 10	73 85 41	93 43
	2 3 4 	1 1 0	group B group B group B	college associate's degree some college some college associate's	standard standard free/reduced	1.0 1.0 1.0	school_bus private private	5 - 10 5 - 10 > 10 	73 85 41 	93 43
	2 3 4 22338	1 1 0 	group B group B group B group A	college associate's degree some college some college associate's degree some	standard standard free/reduced free/reduced	1.0 1.0 1.0 2.0	school_bus private private school_bus	5 - 10 5 - 10 > 10 5 - 10	73 85 41 65	84 93 43 60
	2 3 4 22338 22339	1 1 0 0	group B group B group B group A group C	college associate's degree some college some college associate's degree some college	standard standard free/reduced free/reduced standard	1.0 1.0 1.0 2.0	school_bus private private school_bus school_bus	5 - 10 5 - 10 > 10 5 - 10	73 85 41 65	84 93 43 60
	2 3 4 22338 22339 22340	1 1 0 0	group B group B group B group A group C group D	college associate's degree some college some college associate's degree some college	standard standard free/reduced free/reduced standard standard	1.0 1.0 1.0 2.0 2.0	school_bus private private school_bus school_bus school_bus	5 - 10 5 - 10 > 10 5 - 10 5 - 10	73 85 41 65 58 59	84 93 43 60 53 61
	2 3 4 22338 22339 22340 22341 22342	1 0 0	group B group B group B group A group C group D group E group D	college associate's degree some college some college associate's degree some college high school high school associate's	standard standard free/reduced free/reduced standard standard standard	1.0 1.0 1.0 2.0 2.0 2.0	school_bus private private school_bus school_bus school_bus private	5 - 10 5 - 10 > 10 5 - 10 5 - 10 5 - 10 5 - 10 5 - 10	73 85 41 65 58 59 58	84 93 43 60 53 61 53
2	2 3 4 22338 22339 22340 22341 22342	1 0 0 1 0	group B group B group B group A group C group D group E group D	college associate's degree some college some college associate's degree some college high school high school associate's	standard standard free/reduced free/reduced standard standard standard	1.0 1.0 1.0 2.0 2.0 2.0	school_bus private private school_bus school_bus school_bus private	5 - 10 5 - 10 > 10 5 - 10 5 - 10 5 - 10 5 - 10 5 - 10	73 85 41 65 58 59 58	84 93 43 60 53 61 53
2	2 3 4 22338 22339 22340 22341 22342	1 0 0 1 0 1 ows × 10	group B group B group B group A group C group D group E group D	college associate's degree some college some college associate's degree some college high school high school associate's degree	standard standard free/reduced free/reduced standard standard standard standard	1.0 1.0 1.0 2.0 2.0 2.0 1.0 3.0	school_bus private private school_bus school_bus school_bus private	5 - 10 5 - 10 > 10 5 - 10 5 - 10 5 - 10 5 - 10 5 - 10 5 - 10	73 85 41 65 58 59 58	84 93 43 60 53 61 53
2	2 3 4 22338 22339 22340 22341 22342 22343 rd	1 0 0 1 0 1 ows × 10	group B group B group B group A group C group D group E group D columns	college associate's degree some college some college associate's degree some college high school high school associate's degree	standard standard free/reduced free/reduced standard standard standard standard	1.0 1.0 1.0 2.0 2.0 2.0 1.0 3.0	school_bus private private school_bus school_bus private school_bus private	5 - 10 5 - 10 > 10 5 - 10 5 - 10 5 - 10 5 - 10 5 - 10 5 - 10	73 85 41 65 58 59 58	84 93 43 60 53 61 53

In [51]: data

0	1	group B	master's degree	standard	4	school_bus	< 5	87	93
1	0	group C	some college	standard	0	school_bus	5 - 10	76	78
2	1	group B	associate's degree	standard	1	school_bus	5 - 10	73	84
3	1	group B	some college	standard	1	private	5 - 10	85	93
4	0	group B	some college	free/reduced	1	private	> 10	41	43
22338	0	group A	associate's degree	free/reduced	2	school_bus	5 - 10	65	60
22339	0	group C	some college	standard	2	school_bus	5 - 10	58	53
22340	1	group D	high school	standard	2	school_bus	5 - 10	59	61
22341	0	group E	high school	standard	1	private	5 - 10	58	53
22342	1	group D	associate's degree	standard	3	school_bus	5 - 10	82	90

Now our Data cleaning is done. It is time to export the cleaned data

```
In [52]: data.to_excel("Cleaned_data.xlsx")
In [53]: data.to_csv("Cleaned_data.csv")
```

Thank you