In [1]: import pandas as pd Reading CSV which is in our computer df=pd.read csv(r"C:\Users\USER\Downloads\aug train.csv") In [2]: df.head() enrolled\_university Out[2]: enrollee\_id city\_development\_index gender relevent\_experience education\_level major\_discipline experience Has relevent 0 0.920 8949 city\_103 Male no\_enrollment Graduate STEM >2 experience No relevent 29725 city\_40 0.776 Male no\_enrollment Graduate STEM 1 experience No relevent 2 0.624 11561 Full time course Graduate STEM city 21 NaN experience No relevent 0.789 NaN **Business Degree** 3 33241 city\_115 NaN Graduate experience Has relevent 4 666 city\_162 0.767 Male no\_enrollment Masters STEM >2 experience Reading CSV from URL - like directly from Github (LOADING FROM SERVER) dfl=pd.read csv("https://media.githubusercontent.com/media/datablist/sample-csv-files/main/files/customers/cust In [3]: df1.head() In [4]: Out[4]: First Last Index **Customer Id** City Phone 1 Phone 2 Country Emai Company Name Name First Last 0 Index Customer Id Company City Country Phone 1 Phone 2 Emai Name Name Rasmussen DD37Cf93aecA6Dc Shervl East Leonard Chile 229.077.5154 397.884.0519x718 zunigavanessa@smith.info Group

Vega-East 686-620-2 1Ef7b82A4CAAD10 Preston Djibouti 5153435776 vmata@colon.com Lozano Gentry Jimmychester 1820x944 Antigua +1-539-402-Murillo-(496)978beckycarr@hogan.com 3 6F94879bDAfE5a6 Roy Berry Isabelborough Perry 0259 3969x58947 Barbuda Dominguez, Mcmillan Dominican 001-808-617-5Cef8BFA16c5e3c +1-813-324-8756 stanleyblackwell@benson.org Linda Olsen Bensonview Republic 6467x12895 Donovan

Reading CSV file but separated with tab (Tab Separated file)

In [5]: df2=pd.read\_csv(r"C:\Users\USER\Downloads\movie\_titles\_metadata.tsv", sep="\t")

In [6]: df2.head(4)

62847 m0 10 things i hate about you 1999 6.90 ['comedy' 'romance'] 6.2 10421.0 0 m1 1492: conquest of paradise 1992 ['adventure' 'biography' 'drama' 'history'] 1 m2 15 minutes 2001 6.1 25854.0 ['action' 'crime' 'drama' 'thriller'] **2** m3 2001: a space odyssey 1968 8.4 163227.0 ['adventure' 'mystery' 'sci-fi'] 3 48 hrs. 1982 6.9 22289.0 ['action' 'comedy' 'crime' 'drama' 'thriller']

Notice something in tab separated values, look first row become columns because there is no column names, so pass list

In [7]: df3= pd.read\_csv(r"C:\Users\USER\Downloads\movie\_titles\_metadata.tsv", sep="\t", names=['symbol','movie','relea"]

In [8]: df3.head()

profit genere Out[8]: symbol movie release\_year rating 0 10 things i hate about you 62847.0 ['comedy' 'romance'] m01999 6.9 1 m1 1492: conquest of paradise 1992 6.2 10421.0 ['adventure' 'biography' 'drama' 'history'] 2 m2 2001 6.1 25854.0 ['action' 'crime' 'drama' 'thriller'] 15 minutes 3 m3 2001: a space odyssey 1968 8.4 163227.0 ['adventure' 'mystery' 'sci-fi'] m4 48 hrs 1982 6.9 22289.0 ['action' 'comedy' 'crime' 'drama' 'thriller']

Now lets look - index\_col Parameter

In [9]: #lets see first df df.head() enrollee\_id city city\_development\_index gender relevent\_experience enrolled\_university education\_level major\_discipline Out[9]: Has relevent 0 0.920 STEM 8949 city\_103 Male no\_enrollment Graduate >2 experience No relevent 1 29725 city 40 0.776 Male no enrollment Graduate STEM 1 experience No relevent 2 11561 city\_21 0.624 NaN Full time course Graduate STFM experience No relevent 3 33241 city\_115 0.789 NaN NaN Graduate **Business Degree** experience Has relevent 4 666 city\_162 0.767 Male no\_enrollment Masters STEM >2( experience

What we are noticing? we saw that, there is normal pandas default index too 0,1,2... and there is enrollment id as enrollee\_id, which can act as primary key

city\_development\_index gender relevent\_experience enrolled\_university education\_level major\_discipline experience

no\_enrollment

Masters

STEM

>20

In [10]: df=pd.read\_csv(r"C:\Users\USER\Downloads\aug\_train.csv",index\_col='enrollee\_id')
df.head()

enrollee\_id Has relevent 8949 city\_103 0.920 Male no\_enrollment Graduate STEM >20 experience No relevent 29725 0.776 Male no enrollment Graduate STEM 15 city 40 experience No relevent 11561 0.624 STFM 5 NaN Full time course Graduate city 21 experience No relevent 33241 city\_115 0.789 NaN NaN Graduate **Business Degree** <1 experience

Has relevent

experience

Now, lets look Header Parameter

666 city\_162

In [11]: #First of all lets look the problem
 df4=pd.read\_csv(r"C:\Users\USER\Downloads\test.csv")
 df4.head(2)

0.767

Male

Out[11]:		Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed:	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	ı
	0	0	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_discipline	•
	1	1	29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	STEM	

What we notice? - header is becoming our first row

In [13]: df4=pd.read\_csv(r"C:\Users\USER\Downloads\test.csv",header=1)
df4.head(2)

Out[13]: 0 enrollee\_id city city\_development\_index gender relevent\_experience enrolled\_university education\_level major\_discipline experier No relevent 0 1 29725 city\_40 0.776 no\_enrollment Graduate STEM Male experience No relevent 1 2 11561 city\_21 0.624 NaN Full time course Graduate STEM experience

Now, sometime, in ML we may not need all column, we may only need few columns - we can decrease and can only import needed column if we want directly during import of data rather than dropping later in preprocessing steps.

In [14]: #lets look df in this too
 df.head()

```
city_development_index gender relevent_experience enrolled_university education_level major_discipline experience
Out[14]:
           enrollee id
                                                                        Has relevent
                                                                                                                                 STEM
                                                                                                                                               >20
                 8949 city_103
                                                 0.920
                                                          Male
                                                                                          no enrollment
                                                                                                              Graduate
                                                                         experience
                                                                         No relevent
                                                                                                                                 STFM
                29725
                        city_40
                                                 0.776
                                                          Male
                                                                                          no_enrollment
                                                                                                              Graduate
                                                                                                                                                15
                                                                         experience
                                                                         No relevent
                11561
                        city_21
                                                 0.624
                                                           NaN
                                                                                        Full time course
                                                                                                              Graduate
                                                                                                                                 STEM
                                                                                                                                                  5
                                                                         experience
                                                                         No relevent
                33241 city_115
                                                 0.789
                                                           NaN
                                                                                                  NaN
                                                                                                              Graduate
                                                                                                                        Business Degree
                                                                                                                                                 <1
                                                                         experience
                                                                        Has relevent
                                                 0.767
                                                                                                                                 STEM
                                                                                                                                               >20
                  666 city_162
                                                          Male
                                                                                          no enrollment
                                                                                                               Masters
                                                                         experience
           #suppose I only want city, gender and education level
           df=pd.read csv(r"C:\Users\USER\Downloads\aug train.csv",usecols=['city','gender','education level'])
           df.head()
                  city
                       gender education_level
Out[17]:
           0 city 103
                         Male
                                      Graduate
               city_40
                         Male
                                      Graduate
               city_21
                          NaN
                                      Graduate
           3 city_115
                          NaN
                                      Graduate
           4 city_162
                         Male
                                       Masters
           Skiprows parameter
In [18]:
Out[18]:
                      city
                           gender
                                    education_level
                0 city_103
                              Male
                                          Graduate
                   city_40
                              Male
                1
                                          Graduate
                2
                   city_21
                              NaN
                                          Graduate
                3 city_115
                              NaN
                                          Graduate
                  city_162
                              Male
                                           Masters
           19153 city_173
                              Male
                                          Graduate
           19154 city_103
                              Male
                                          Graduate
           19155 city_103
                              Male
                                          Graduate
           19156
                                        High School
                   city 65
                              Male
           19157
                   city_67
                              NaN
                                     Primary School
           19158 rows × 3 columns
In [19]:
           #Now suppose I do not need 3 and 4th rows, look city 21 and 115, now i do not need that row
           df=df=pd.read_csv(r"C:\Users\USER\Downloads\aug_train.csv",skiprows=[3,4])
In [20]: df.head()
              enrollee_id
                                  city_development_index gender relevent_experience enrolled_university
                                                                                                          education_level major_discipline experience
Out[20]:
                              city
                                                                           Has relevent
           0
                                                                                                                                                  >2(
                    8949
                          city_103
                                                    0.920
                                                             Male
                                                                                             no_enrollment
                                                                                                                 Graduate
                                                                                                                                     STEM
                                                                            experience
                                                                            No relevent
                   29725
                                                    0.776
                                                                                                                 Graduate
                                                                                                                                     STEM
                                                                                                                                                    1
           1
                           city_40
                                                             Male
                                                                                             no enrollment
                                                                            experience
                                                                           Has relevent
           2
                                                    0.767
                                                                                             no enrollment
                                                                                                                                    STEM
                                                                                                                                                  >20
                     666 city 162
                                                             Male
                                                                                                                  Masters
                                                                            experience
                                                                           Has relevent
           3
                   21651 city_176
                                                    0.764
                                                              NaN
                                                                                           Part time course
                                                                                                                 Graduate
                                                                                                                                     STFM
                                                                            experience
                                                                           Has relevent
           4
                   28806 city_160
                                                    0.920
                                                             Male
                                                                                             no_enrollment
                                                                                                               High School
                                                                                                                                      NaN
                                                                            experience
```

dtypes parameter

In [21]: #it is useful when we want to convert datatypes of column. suppose:
 df=pd.read\_csv(r"C:\Users\USER\Downloads\aug\_train.csv",index\_col='enrollee\_id')
 df.head()

```
Has relevent
                8949 city_103
                                             0.920
                                                     Male
                                                                                                     Graduate
                                                                                                                       STEM
                                                                                                                                    >20
                                                                                   no_enrollment
                                                                   experience
                                                                   No relevent
                                                                                                     Graduate
                                                                                                                       STEM
                                                                                                                                     15
               29725
                      city 40
                                             0.776
                                                     Male
                                                                                  no enrollment
                                                                   experience
                                                                   No relevent
                                                                                                                       STEM
                                                                                                                                     5
               11561
                      city 21
                                             0.624
                                                      NaN
                                                                                 Full time course
                                                                                                     Graduate
                                                                   experience
                                                                   No relevent
               33241 city 115
                                             0.789
                                                      NaN
                                                                                          NaN
                                                                                                     Graduate
                                                                                                              Business Degree
                                                                                                                                     <1
                                                                   experience
                                                                  Has relevent
                666 city_162
                                             0.767
                                                     Male
                                                                                   no_enrollment
                                                                                                      Masters
                                                                                                                       STEM
                                                                                                                                    >20
                                                                   experience
In [22]: df.info()
          <class 'pandas.core.frame.DataFrame'>
          Int64Index: 19158 entries, 8949 to 23834
          Data columns (total 13 columns):
           #
                Column
                                           Non-Null Count
                                                             Dtype
           0
                citv
                                           19158 non-null
                                                             object
                city_development_index
           1
                                           19158 non-null
                                                             float64
           2
                gender
                                           14650 non-null
                                                             object
           3
                                           19158 non-null
                relevent experience
                                                             object
           4
                enrolled_university
                                           18772 non-null
                                                             object
           5
                education_level
                                           18698 non-null
                                                             object
           6
                major discipline
                                           16345 non-null
                                                             obiect
           7
                                           19093 non-null
                experience
                                                             object
           8
                                           13220 non-null
                company_size
                                                             object
           9
                company_type
                                           13018 non-null
                                                             object
           10
                last new job
                                           18735 non-null
                                                             object
                                           19158 non-null
           11
                training_hours
                                                             int64
           12
               target
                                           19158 non-null
                                                             float64
          dtypes: float64(2), int64(1), object(10)
          memory usage: 2.0+ MB
          #here target is float and suppose we want target as integer.
In [23]:
          df=pd.read_csv(r"C:\Users\USER\Downloads\aug_train.csv",index_col='enrollee_id',dtype={'target':int})
          df.head()
                         city_development_index gender relevent_experience enrolled_university education_level major_discipline experience
          enrollee id
                                                                  Has relevent
                                                                                                                       STFM
                                                                                                                                    >20
                8949 city_103
                                             0.920
                                                     Male
                                                                                  no enrollment
                                                                                                     Graduate
                                                                   experience
                                                                   No relevent
               29725
                      city 40
                                             0.776
                                                      Male
                                                                                   no_enrollment
                                                                                                     Graduate
                                                                                                                       STEM
                                                                                                                                     15
                                                                   experience
                                                                   No relevent
                                                                                 Full time course
               11561
                      city_21
                                             0.624
                                                      NaN
                                                                                                     Graduate
                                                                                                                       STEM
                                                                                                                                     5
                                                                   experience
                                                                   No relevent
               33241 city_115
                                             0.789
                                                      NaN
                                                                                          NaN
                                                                                                     Graduate
                                                                                                              Business Degree
                                                                                                                                     <1
                                                                   experience
                                                                  Has relevent
                666 city_162
                                             0.767
                                                                                                      Masters
                                                                                                                       STEM
                                                                                                                                    >20
                                                     Male
                                                                                  no enrollment
                                                                   experience
In [24]: df.info()
          <class 'pandas.core.frame.DataFrame'>
          Int64Index: 19158 entries, 8949 to 23834
          Data columns (total 13 columns):
                                           Non-Null Count
           #
                Column
                                                             Dtype
          - - -
           0
                city
                                           19158 non-null
                                                             object
           1
                city development index
                                           19158 non-null
                                                             float64
           2
                                           14650 non-null
                gender
                                                             object
           3
                relevent_experience
                                           19158 non-null
                                                             object
           4
                                           18772 non-null
                enrolled university
                                                             object
           5
                education level
                                           18698 non-null
                                                             obiect
           6
                major discipline
                                           16345 non-null
                                                             object
           7
                                           19093 non-null
                experience
                                                             object
           8
                                           13220 non-null
                company_size
                                                             object
           9
                                           13018 non-null
                company_type
                                                             object
           10
                last new job
                                           18735 non-null
                                                             object
                                                             int64
           11
                training_hours
                                           19158 non-null
                                           19158 non-null
           12
                                                             int32
                target
          dtypes: float64(1), int32(1), int64(1), object(10)
          memory usage: 2.0+ MB
```

city city\_development\_index gender relevent\_experience enrolled\_university education\_level major\_discipline experience

Out[21]:

enrollee\_id

Handling dates

```
df6=pd.read_csv(r"C:\Users\USER\Downloads\IPL Matches 2008-2020.csv")
In [25]:
          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 816 entries, 0 to 815
          Data columns (total 17 columns):
           #
                Column
                                   Non-Null Count
                                                    Dtype
          - - -
           0
                id
                                   816 non-null
                                                    int64
           1
                city
                                   803 non-null
                                                    object
           2
                                   816 non-null
                                                    object
                date
           3
                player_of_match
                                  812 non-null
                                                     object
                venue
                                   816 non-null
                                                    object
           5
                neutral venue
                                   816 non-null
                                                    int64
           6
                                   816 non-null
                                                    object
                team1
           7
                team2
                                   816 non-null
                                                     object
           8
                toss winner
                                   816 non-null
                                                    object
                toss_decision
           9
                                   816 non-null
                                                    obiect
           10
                winner
                                   812 non-null
                                                     object
           11
                                   812 non-null
                result
                                                    object
           12
                result margin
                                   799 non-null
                                                     float64
                                   812 non-null
           13
                eliminator
                                                    object
           14
                method
                                   19 non-null
                                                     object
           15
               umpire1
                                   816 non-null
                                                    object
           16 umpire2
                                   816 non-null
                                                    object
          dtypes: float64(1), int64(2), object(14)
          memory usage: 108.5+ KB
          #Its a cricket match data and I directly use info function to see datatypes, see the date, it is object - means
In [26]:
          #Convert it in date
          df6=pd.read csv(r"C:\Users\USER\Downloads\IPL Matches 2008-2020.csv", parse dates=['date'])
          df6.info()
          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 816 entries, 0 to 815
          Data columns (total 17 columns):
           #
                Column
                                   Non-Null Count
                                                    Dtype
           0
                                   816 non-null
                                                    int64
                id
           1
                city
                                   803 non-null
                                                     object
           2
                                   816 non-null
                                                     datetime64[ns]
                date
           3
                player_of_match 812 non-null
                                                    object
           4
                                   816 non-null
                venue
                                                     object
           5
                neutral_venue
                                   816 non-null
                                                     int64
           6
                team1
                                   816 non-null
                                                    object
           7
                                   816 non-null
                                                    object
                team2
           8
                toss_winner
                                   816 non-null
                                                    object
           9
                toss_decision
                                   816 non-null
                                                    object
           10
                                   812 non-null
               winner
                                                    object
                                   812 non-null
           11
                result
                                                    object
           12
                result margin
                                   799 non-null
                                                     float64
           13
                eliminator
                                   812 non-null
                                                    object
           14
               method
                                   19 non-null
                                                    object
           15
               umpire1
                                   816 non-null
                                                     object
           16
               umpire2
                                   816 non-null
                                                    object
          dtypes: datetime64[ns](1), float64(1), int64(2), object(13)
          memory usage: 108.5+ KB
          Converters
In [27]:
          df6.head()
                                     player_of_match
                                                                 neutral_venue
                                                                                  team1
                                                                                                   toss_winner toss_decision
                                                                                                                                winner
Out[27]:
                          city
                                date
                                                          venue
                                                                                             team2
                                                                                   Royal
                                                                                            Kolkata
                                                                                                         Royal
                                                                                                                                Kolkata
                               2008-
          0 335982
                     Bangalore
                                        BB McCullum Chinnaswamy
                                                                           0 Challengers
                                                                                                     Challengers
                                                                                             Kniaht
                                                                                                                        field
                                                                                                                                 Kniaht
                               04-18
                                                         Stadium
                                                                                Bangalore
                                                                                             Riders
                                                                                                      Bangalore
                                                                                                                                 Riders
                                                          Punjab
                                                                                            Chennai
                                                                                                                                Chennai
                                                          Cricket
                               2008-
                                                                                 Kings XI
                                                                                                       Chennai
          1 335983 Chandigarh
                                         MEK Hussey
                                                      Association
                                                                           0
                                                                                              Super
                                                                                                                         bat
                                                                                                                                 Super
                               04-19
                                                                                  Punjab
                                                                                                     Super Kings
                                                        Stadium,
                                                                                              Kinas
                                                                                                                                  Kinas
                                                          Mohali
                                                      Feroz Shah
                                                                                          Rajasthan
                                                                                                      Rajasthan
                               2008-
                                                                                   Delhi
                                                                                                                                  Delhi
          2 335984
                                         MF Maharoof
                                                                           0
                               04-19
                                                           Kotla
                                                                               Daredevils
                                                                                                         Royals
                                                                                                                              Daredevils
                                                                                             Royals
                                                                                              Royal
                                                                                                                                 Royal
                               2008
                                                       Wankhede
                                                                                 Mumbai
                                                                                                        Mumbai
                                         MV Boucher
                                                                                                                            Challengers
Bangalore
                                                                           0
                                                                                         Challengers
          3 335985
                       Mumbai
                                                                                                        Indians
                               04-20
                                                         Stadium
                                                                                  Indians
                                                                                          Bangalore
```

```
In [29]: #suppose in above dataframe instead of Kolkata Knight rider, I need KKR

def rename(a):
    if a=="Kolkata Knight Riders":
        return "KKR"
```

Eden

Gardens

2008-

DJ Hussey

Kolkata

4 335986

Kolkata

Knight

Riders

0

Deccan

Chargers

Deccan

Chargers

Kolkata

Knight

Riders

bat

else: return a

In [31]: df6=pd.read\_csv(r"C:\Users\USER\Downloads\IPL Matches 2008-2020.csv",converters={'team1':rename,'team2':rename}

In [32]: df6.head()

Out[32]: id city date player\_of\_match team1 team2 toss winner toss decision winner venue neutral venue M Royal Royal Kolkata 2008-0 Challengers **0** 335982 Bangalore BB McCullum Chinnaswamy KKR Challengers field Knight Bangalore Riders Stadium Bangalore Punjab Cricket Chennai Chennai 2008-Kings XI Chennai MEK Hussey 0 1 335983 Chandigarh Association Super bat Super 04-19 Super Kings Punjab Stadium. Kings Kings Mohali Feroz Shah Rajasthan Delhi 2008-Delhi Rajasthan 2 335984 Delhi MF Maharoof 0 bat Daredevils Daredevils 04-19 Kotla Royals Royals Royal Royal 2008-Wankhede Mumbai Mumbai 0 **3** 335985 Mumbai MV Boucher Challengers bat Challengers 04-20 Stadium Indians Indians Bangalore Bangalore Kolkata 2008-Eden Deccan Deccan 4 335986 Kolkata DJ Hussey 0 KKR bat Knight 04-20 Gardens Chargers Chargers Riders

Loading a huge dataset (Need to work in chunk)

In [33]: #This below is just an example, suppose we will have 19lakh dateset, where our laptop/Machine #could not load and handle operation at a time so at that time we should divide data in chunk and need to work

In [36]: df6

36]:		id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	wini
	<b>o</b> 335	5982	Bangalore	2008- 04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	KKR	Royal Challengers Bangalore	field	Kolk Kni Rid
	1 335	5983	Chandigarh	2008- 04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chen Su <sub>l</sub> Kir
2	2 335	5984	Delhi	2008- 04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	D∉ Darede <sup>,</sup>
;	<b>3</b> 335	5985	Mumbai	2008- 04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Ro Challeng Bangal
4	<b>4</b> 335	5986	Kolkata	2008- 04-20	DJ Hussey	Eden Gardens	0	KKR	Deccan Chargers	Deccan Chargers	bat	Kolk Kni Rid
81	<b>1</b> 1216	6547	Dubai	2020- 09-28	AB de Villiers	Dubai International Cricket Stadium	0	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	Ro Challeng Bangalı
812	<b>2</b> 1237	7177	Dubai	2020- 11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mumbai Indians	Delhi Capitals	Delhi Capitals	field	Mum India
813	<b>3</b> 1237	7178	Abu Dhabi	2020- 11-06	KS Williamson	Sheikh Zayed Stadium	0	Royal Challengers Bangalore	Sunrisers Hyderabad	Sunrisers Hyderabad	field	Sunris Hyderab
814	<b>4</b> 1237	7180	Abu Dhabi	2020- 11-08	MP Stoinis	Sheikh Zayed Stadium	0	Delhi Capitals	Sunrisers Hyderabad	Delhi Capitals	bat	De Capit
81	<b>5</b> 1237	7181	Dubai	2020- 11-10	TA Boult	Dubai International Cricket Stadium	0	Delhi Capitals	Mumbai Indians	Delhi Capitals	bat	Mum India

In [37]: #Just an example lets divide df6 in chunks
dfs=pd.read\_csv(r"C:\Users\USER\Downloads\IPL Matches 2008-2020.csv", chunksize=200)

In [38]: for chunks in dfs:
 print(chunks.shape)

816 rows × 17 columns

```
(200, 17)
(200, 17)
(200, 17)
(200, 17)
(16, 17)
```

In [39]: #see it divided our data set 5 sets, 4 set of equal row and column and last is leftover data set.

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js