1. Importing pandas

In [1]:

```
import pandas as pd
```

2. Opening a local csv file

In [34]:

```
df = pd.read_csv('aug_train.csv')
df
```

Out[34]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_c
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	
1	29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	
2	11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	
3	33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	
4	666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	
...	...	...	...	...	...	...	...	
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollment	Graduate	Hu
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollment	High School	
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollment	Primary School	

19158 rows × 14 columns

3. Opening a csv file from an URL

In [35]:

```
import requests
from io import StringIO

url = "https://raw.githubusercontent.com/cs109/2014_data/master/countries.csv"
headers = {"User-Agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.14; rv:66.0) Gecko/2
0100101 Firefox/66.0"}
req = requests.get(url, headers=headers)
data = StringIO(req.text)

pd.read_csv(data)
```

Out[35]:

	Country	Region
0	Algeria	AFRICA
1	Angola	AFRICA
2	Benin	AFRICA
3	Botswana	AFRICA
4	Burkina	AFRICA
...	...	...
189	Paraguay	SOUTH AMERICA
190	Peru	SOUTH AMERICA
191	Suriname	SOUTH AMERICA
192	Uruguay	SOUTH AMERICA
193	Venezuela	SOUTH AMERICA

194 rows × 2 columns

4. Sep Parameter

In [41]:

```
pd.read_csv('movie_titles_metadata.tsv',sep='\t',names=['sno','name','release_year','rating','votes','genres'])
```

Out[41]:

	sno	name	release_year	rating	votes	genres
0	m0	10 things i hate about you	1999	6.9	62847.0	['comedy' 'romance']
1	m1	1492: conquest of paradise	1992	6.2	10421.0	['adventure' 'biography' 'drama' 'history']
2	m2	15 minutes	2001	6.1	25854.0	['action' 'crime' 'drama' 'thriller']
3	m3	2001: a space odyssey	1968	8.4	163227.0	['adventure' 'mystery' 'sci-fi']
4	m4	48 hrs.	1982	6.9	22289.0	['action' 'comedy' 'crime' 'drama' 'thriller']
...	...	...	...	...	...	...
612	m612	watchmen	2009	7.8	135229.0	['action' 'crime' 'fantasy' 'mystery' 'sci-fi'...
613	m613	xxx	2002	5.6	53505.0	['action' 'adventure' 'crime']
614	m614	x-men	2000	7.4	122149.0	['action' 'sci-fi']
615	m615	young frankenstein	1974	8.0	57618.0	['comedy' 'sci-fi']
616	m616	zulu dawn	1979	6.4	1911.0	['action' 'adventure' 'drama' 'history' 'war']

617 rows × 6 columns

5. Index_col parameter

In [43]:

```
pd.read_csv('aug_train.csv',index_col='enrollee_id')
```

Out[43]:

enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_disciplin
8949	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM
29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	STEM
11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	STEM
33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	Busines Degre
666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	STEM
...	...	...	...	...	...	...	.
7386	city_173	0.878	Male	No relevent experience	no_enrollment	Graduate	Humanitie
31398	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM
24576	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM
5756	city_65	0.802	Male	Has relevent experience	no_enrollment	High School	Nal
23834	city_67	0.855	NaN	No relevent experience	no_enrollment	Primary School	Nal

19158 rows × 13 columns

6. Header parameter

In [46]:

```
pd.read_csv('test.csv',header=1)
```

Out[46]:

	0	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_dis
0	1	29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	
1	2	11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	
2	3	33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	Bu
3	4	666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	

7. use_cols parameter

In [48]:

```
pd.read_csv('aug_train.csv',usecols=['enrollee_id','gender','education_level'])
```

Out[48]:

	enrollee_id	gender	education_level
0	8949	Male	Graduate
1	29725	Male	Graduate

	enrollee_id	gender	education_level
2	11561	NaN	Graduate
3	33241	NaN	Graduate
4	666	Male	Masters
...	...	...	...
19153	7386	Male	Graduate
19154	31398	Male	Graduate
19155	24576	Male	Graduate
19156	5756	Male	High School
19157	23834	NaN	Primary School

19158 rows × 3 columns

8. Squeeze parameters

In [50]:

```
pd.read_csv('aug_train.csv',usecols=['gender'],squeeze=True)
```

Out[50]:

```
0         Male
1         Male
2          NaN
3          NaN
4         Male
          ...
19153     Male
19154     Male
19155     Male
19156     Male
19157      NaN
Name: gender, Length: 19158, dtype: object
```

9. Skiprows/nrows Parameter

In [103]:

```
pd.read_csv('aug_train.csv',nrows=100)
```

Out[103]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_disci
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	S
1	29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	S
2	11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	S
3	33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	Bus D
4	666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	S
...	...	...	...	...	...	...	...	
95	12081	city_65	0.802	Male	Has relevent experience	Full time course	Graduate	S
96	7364	city_160	0.920	NaN	No relevent experience	Full time course	High School	

97	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_disc
98	7016	city_65	0.802	Male	Has relevent experience	no_enrollment	Graduate	
99	8695	city_11	0.550	Male	Has relevent experience	no_enrollment	Graduate	

100 rows × 14 columns

10. Encoding parameter

In [97]:

```
pd.read_csv('zomato.csv',encoding='latin-1')
```

Out[97]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cu
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565443	F Japa De
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553708	Jap
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...	Edsa Shangri-La, Ortigas, Mandaluyong City	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...	121.056831	14.581404	Sea Fi
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.056475	14.585318	Japa
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.057508	14.584450	Japa K
...	...	...	...	...	...	...	...	...	...	
9546	5915730	NamlÛ± Gurme	208	ÛÁstanbul	Kemankeô Karamustafa Paôa Mahallesi, RÛ±htÛ±...	Karakí_y	Karakí_y, ÛÁstanbul	28.977392	41.022793	T
9547	5908749	Ceviz AÛôacÛ±	208	ÛÁstanbul	Koôuyolu Mahallesi, Muhittin îstí_ndaÛô Cadd...	Koôuyolu	Koôuyolu, ÛÁstanbul	29.041297	41.009847	Cu Patis
9548	5915807	Huqqa	208	ÛÁstanbul	Kuruí_eôme Mahallesi, Muallim Naci Caddesi, N...	Kuruí_eôme	Kuruí_eôme, ÛÁstanbul	29.034640	41.055817	I C

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cu Resta
9549	5916112	Aşık Kahve	208	İstanbul	Kuruí_eôme Mahallesi, Muallim Naci Caddesi, N...	Kuruí_eôme	Kuruí_eôme, İstanbul	29.036019	41.057979	
9550	5927402	Walter's Coffee Roastery	208	İstanbul	CafeaÛôa Mahallesi, BademaltÛ± Sokak, No 21/B,...	Moda	Moda, İstanbul	29.026016	40.984776	

9551 rows × 21 columns

11. Skip bad lines

In [93]:

```
pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1",error_bad_lines=False)
```

```
b'Skipping line 6452: expected 8 fields, saw 9\nSkipping line 43667: expected 8 fields, s
aw 10\nSkipping line 51751: expected 8 fields, saw 9\n'
b'Skipping line 92038: expected 8 fields, saw 9\nSkipping line 104319: expected 8 fields,
saw 9\nSkipping line 121768: expected 8 fields, saw 9\n'
b'Skipping line 144058: expected 8 fields, saw 9\nSkipping line 150789: expected 8 fields
, saw 9\nSkipping line 157128: expected 8 fields, saw 9\nSkipping line 180189: expected 8
fields, saw 9\nSkipping line 185738: expected 8 fields, saw 9\n'
b'Skipping line 209388: expected 8 fields, saw 9\nSkipping line 220626: expected 8 fields
, saw 9\nSkipping line 227933: expected 8 fields, saw 11\nSkipping line 228957: expected
8 fields, saw 10\nSkipping line 245933: expected 8 fields, saw 9\nSkipping line 251296: e
xpected 8 fields, saw 9\nSkipping line 259941: expected 8 fields, saw 9\nSkipping line 26
1529: expected 8 fields, saw 9\n'
```

Out[93]:

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-U
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/01951534
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/00020050
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/00609731
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/03741570
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/03930452
...	...	...	...	...	...	
271355	0440400988	There's a Bat in Bunk Five	Paula Danziger	1988	Random House Childrens Pub (Mm)	http://images.amazon.com/images/P/04404009
271356	0525447644	From One to One Hundred	Teri Sloat	1991	Dutton Books	http://images.amazon.com/images/P/05254476
271357	006008667X	Lily Dale : The True Story of the Town that Ta...	Christine Wicker	2004	HarperSanFrancisco	http://images.amazon.com/images/P/00600866

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-L
271358	0192126040	Republic (World's Classics)	Plato	1998	Oxford University Press	http://images.amazon.com/images/P/0192126040.01.LZZZZZZZ.jpg
271359	0767409752	A Guided Tour of Rene Descartes' Meditations o...	Christopher Biffle	2000	McGraw-Hill Humanities/Social Sciences/Languages	http://images.amazon.com/images/P/0767409752.01.LZZZZZZZ.jpg

271360 rows × 8 columns

12. dtypes parameter

In [108]:

```
pd.read_csv('aug_train.csv',dtype={'target':int}).info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19158 entries, 0 to 19157
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   enrollee_id             19158 non-null  int64  
 1   city                    19158 non-null  object 
 2   city_development_index  19158 non-null  float64
 3   gender                  14650 non-null  object 
 4   relevent_experience     19158 non-null  object 
 5   enrolled_university     18772 non-null  object 
 6   education_level         18698 non-null  object 
 7   major_discipline        16345 non-null  object 
 8   experience              19093 non-null  object 
 9   company_size            13220 non-null  object 
 10  company_type            13018 non-null  object 
 11  last_new_job            18735 non-null  object 
 12  training_hours          19158 non-null  int64  
 13  target                  19158 non-null  int32  
dtypes: float64(1), int32(1), int64(2), object(10)
memory usage: 2.0+ MB
```

13. Handling Dates

In [112]:

```
pd.read_csv('IPL Matches 2008-2020.csv',parse_dates=['date']).info()
```

```
<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   date             816 non-null    datetime64[ns]
 3   player_of_match  812 non-null    object        
 4   venue            816 non-null    object        
 5   neutral_venue    816 non-null    int64         
 6   team1            816 non-null    object        
 7   team2            816 non-null    object        
 8   toss_winner      816 non-null    object        
 9   toss_decision    816 non-null    object        
 10  winner           812 non-null    object        
 11  result           812 non-null    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
```

In [142]:

```
def rename(name):
    if name == "Royal Challengers Bangalore":
        return "RCB"
    else:
        return name
```

In [143]:

```
rename("Royal Challengers Bangalore")
```

Out[143]:

```
'RCB'
```

14. Convertors

In [144]:

```
pd.read_csv('IPL Matches 2008-2020.csv',converters={'team1':rename})
```

Out[144]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_d
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	RCB	Kolkata Knight Riders	Royal Challengers Bangalore	
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	
...	...	...	...	...	...	...	...	...	...	...
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	RCB	Mumbai Indians	Mumbai Indians	
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mumbai Indians	Delhi Capitals	Delhi Capitals	
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	RCB	Sunrisers Hyderabad	Sunrisers Hyderabad	
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	Delhi Capitals	Sunrisers Hyderabad	Delhi Capitals	
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	Delhi Capitals	Mumbai Indians	Delhi Capitals	

816 rows × 17 columns

15. na_values parameter

In [147]:

```
pd.read_csv('aug_train.csv',na_values=['Male',])
```

Out[147]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_c
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	
1	29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	
2	11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	
3	33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	
4	666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	
...	...	...	...	...	...	...	...	
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollment	Graduate	Hu
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollment	High School	
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollment	Primary School	

19158 rows × 14 columns

16. Loading a huge dataset in chunks

In [151]:

```
dfs = pd.read_csv('aug_train.csv',chunksize=5000)
```

In [152]:

```
for chunks in dfs:
    print(chunk.shape)
```

```
(4158, 14)
(4158, 14)
(4158, 14)
(4158, 14)
```

In [ ]:

```

```