In [535]:

```
1 import pandas as pd
```

In [536]:

```
1 # file read and write operation
```

In [537]:

```
1 pd.read_excel("G:\My Drive\Data\data power supply.xlsx")
```

Out[537]:

	Date Time	Temmperature	humidity	Wind speed	General diffuse flows	power 1	power 2	power 3
0	2022-01-01	67	4	177	0.210125	40	31	38
1	2022-01-02	26	4	122	0.128509	27	21	24
2	2022-01-03	58	7	172	0.537862	22	25	16
3	2022-01-04	68	6	97	0.818062	23	45	41
4	2022-01-05	20	4	184	0.514938	24	46	41
5	2022-01-06	20	9	124	0.498937	41	43	49
6	2022-01-07	45	7	284	0.439151	38	15	14
7	2022-01-08	55	6	264	0.041091	29	10	10
8	2022-01-09	47	1	209	0.180481	19	40	47
9	2022-01-10	57	7	259	0.156179	14	31	24
10	2022-01-11	33	7	81	0.750898	14	23	24

In [539]:

```
1 #ls
2 #command is used to show the current working directory
3
```

In [540]:

```
# r is for unicode issue. in order to read the excel character

pd.read_excel(r"G:\My Drive\Data\data power supply.xlsx")
```

Out[540]:

	Date Time	Temmperature	humidity	Wind speed	General diffuse flows	power 1	power 2	power 3
0	2022-01-01	67	4	177	0.210125	40	31	38
1	2022-01-02	26	4	122	0.128509	27	21	24
2	2022-01-03	58	7	172	0.537862	22	25	16
3	2022-01-04	68	6	97	0.818062	23	45	41
4	2022-01-05	20	4	184	0.514938	24	46	41
5	2022-01-06	20	9	124	0.498937	41	43	49
6	2022-01-07	45	7	284	0.439151	38	15	14
7	2022-01-08	55	6	264	0.041091	29	10	10
8	2022-01-09	47	1	209	0.180481	19	40	47
9	2022-01-10	57	7	259	0.156179	14	31	24
10	2022-01-11	33	7	81	0.750898	14	23	24

In [541]:

```
# it gives rows and columns automatically
# first row is considered as a column
# pandas cant be used for unstructured data
# pandas is for structured data
# data type of entire table is dataframe (tabular structure)
#
```

In [542]:

```
df = pd.read_excel(r"G:\My Drive\Data\data power supply.xlsx")
```

In [543]:

```
type(df)
```

Out[543]:

```
pandas.core.frame.DataFrame
```

In [ ]:

```
# in case you dont want first row as a column
# header = 'infer' by default
# header = 'infer = None' then it gives a default header or column
# names = is used to change the name of the column name in the list
```

In [544]:

```
df1 = pd.read_csv("G:\\My Drive\\Data\\taxonomy.csv")
```

In [545]:

```
df1 # header = 'infer' by default
```

Out[545]:

	taxonomy_id	name	parent_id	parent_name
0	101	Emergency	NaN	NaN
1	101-01	Disaster Response	101	Emergency
2	101-02	Emergency Cash	101	Emergency
3	101-02-01	Help Pay for Food	101-02	Emergency Cash
4	101-02-02	Help Pay for Healthcare	101-02	Emergency Cash
...	...	...	...	...
285	111-01-07	Workplace Rights	111-01	Advocacy & Legal Aid
286	111-02	Mediation	111	Legal
287	111-03	Notary	111	Legal
288	111-04	Representation	111	Legal
289	111-05	Translation & Interpretation	111	Legal

290 rows × 4 columns

In [546]:

```
df1 = pd.read_csv("G:\\My Drive\\Data\\taxonomy.csv", header= None)
```

In [547]:

```
df1 # header = 'infer = None' then it gives a default header or column
```

Out[547]:

	0	1	2	3
0	taxonomy_id	name	parent_id	parent_name
1	101	Emergency	NaN	NaN
2	101-01	Disaster Response	101	Emergency
3	101-02	Emergency Cash	101	Emergency
4	101-02-01	Help Pay for Food	101-02	Emergency Cash
...	...	...	...	...
286	111-01-07	Workplace Rights	111-01	Advocacy & Legal Aid
287	111-02	Mediation	111	Legal
288	111-03	Notary	111	Legal
289	111-04	Representation	111	Legal
290	111-05	Translation & Interpretation	111	Legal

291 rows × 4 columns

In [548]:

```
df1 = pd.read_csv("G:\\My Drive\\Data\\taxonomy.csv", names= ['a','b','c','d'])
```

names = is used to change the name of the column name in the list

In [19]:

```
1 df1
```

Out[19]:

	a	b	c	d
0	taxonomy_id	name	parent_id	parent_name
1	101	Emergency	NaN	NaN
2	101-01	Disaster Response	101	Emergency
3	101-02	Emergency Cash	101	Emergency
4	101-02-01	Help Pay for Food	101-02	Emergency Cash
...	...	...	...	...
286	111-01-07	Workplace Rights	111-01	Advocacy & Legal Aid
287	111-02	Mediation	111	Legal
288	111-03	Notary	111	Legal
289	111-04	Representation	111	Legal
290	111-05	Translation & Interpretation	111	Legal

291 rows × 4 columns

In [22]:

```
1 df1 = pd.read_csv("G:\\My Drive\\Data\\taxonomy.csv", names= ['a','b','c'])
```

In [21]:

```
1 df1
```

Out[21]:

	a	b	c
taxonomy_id	name	parent_id	parent_name
101	Emergency	NaN	NaN
101-01	Disaster Response	101	Emergency
101-02	Emergency Cash	101	Emergency
101-02-01	Help Pay for Food	101-02	Emergency Cash
...	...	...	...
111-01-07	Workplace Rights	111-01	Advocacy & Legal Aid
111-02	Mediation	111	Legal
111-03	Notary	111	Legal
111-04	Representation	111	Legal
111-05	Translation & Interpretation	111	Legal

291 rows × 3 columns

In [26]:

```
1 pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv")
```

Out[26]:

	id	location_id	service_id	start_date	end_date	closed	opens_at	closes_at
0	1	22.0	NaN	January 01, 0001	January 01, 0001	True	NaN	NaN
1	2	22.0	NaN	12/24/2015	12/24/15	False	10:00	16:00
2	3	22.0	NaN	June 01, 0001	September 01, 0001	True	NaN	NaN
3	4	NaN	22.0	January 01, 0001	January 01, 0001	True	NaN	NaN

In [ ]:

```
1 # separated the ',' values with @
```

In [27]:

```
1 pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv")
```

Out[27]:

	id@location_id@service_id@start_date@end_date@closed@opens_at@closes_at
0	1@22@@"January 01@ 0001"@"January 01@ 0001"@tr...
1	2@22@@"12/24/2015"@"12/24/15"@false@10:00@16:00
2	3@22@@"June 01@ 0001"@"September 01@ 0001"@true@@
3	4@@22@"January 01@ 0001"@"January 01@ 0001"@tr...

In [29]:

```
1 pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv", sep= '@')
```

Out[29]:

	id	location_id	service_id	start_date	end_date	closed	opens_at	closes_at
0	1	22.0	NaN	January 01@ 0001	January 01@ 0001	True	NaN	NaN
1	2	22.0	NaN	12/24/2015	12/24/15	False	10:00	16:00
2	3	22.0	NaN	June 01@ 0001	September 01@ 0001	True	NaN	NaN
3	4	NaN	22.0	January 01@ 0001	January 01@ 0001	True	NaN	NaN

In [30]:

```
1 pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv", sep= '@')
```

Out[30]:

	id|location_id|service_id|start_date|end_date|closed|opens_at|closes_at
0	1|22||"January 01| 0001"|"January 01| 0001"|tr...
1	2|22||"12/24/2015"|"12/24/15"|false|10:00|16:00
2	3|22||"June 01| 0001"|"September 01| 0001"|true||
3	4||22|"January 01| 0001"|"January 01| 0001"|tr...

In [31]:

```
1 pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv", sep= '|')
```

Out[31]:

	id	location_id	service_id	start_date	end_date	closed	opens_at	closes_at
0	1	22.0	NaN	January 01| 0001	January 01| 0001	True	NaN	NaN
1	2	22.0	NaN	12/24/2015	12/24/15	False	10:00	16:00
2	3	22.0	NaN	June 01| 0001	September 01| 0001	True	NaN	NaN
3	4	NaN	22.0	January 01| 0001	January 01| 0001	True	NaN	NaN

In [ ]:

```
1 # we can only use on separator at a time
```

In [ ]:

```
1
```

In [ ]:

```
1
```

In [23]:

```
1 pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv")
```

Out[23]:

	id@location_id@service_id@start_date@end_date@closed@opens_at@closes_at
0	1@22@@"January 01@ 0001"@"January 01@ 0001"@tr...
1	2@22@@"12/24/2015"@"12/24/15"@false@10:00@16:00
2	3@22@@"June 01@ 0001"@"September 01@ 0001"@true@@
3	4@@22@"January 01@ 0001"@"January 01@ 0001"@tr...

In [24]:

```
1 pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv", sep= "|")
```

Out[24]:

	id@location_id@service_id@start_date@end_date@closed@opens_at@closes_at
0	1@22@@"January 01@ 0001"@"January 01@ 0001"@tr...
1	2@22@@"12/24/2015"@"12/24/15"@false@10:00@16:00
2	3@22@@"June 01@ 0001"@"September 01@ 0001"@true@@
3	4@@22@"January 01@ 0001"@"January 01@ 0001"@tr...

In [32]:

```
pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv")
```

Out[32]:

	id|location_id|service_id|start_date|end_date|closed|opens_at|closes_at
0	1|22||"January 01| 0001"|"January 01| 0001"|tr...
1	2|22||"12/24/2015"|"12/24/15"|false|10:00|16:00
2	3|22||"June 01| 0001"|"September 01| 0001"|true||
3	4||22|"January 01| 0001"|"January 01| 0001"|tr...

In [33]:

```
pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv", sep= "|")
```

Out[33]:

	id	location_id	service_id	start_date	end_date	closed	opens_at	closes_at
0	1	22.0	NaN	January 01| 0001	January 01| 0001	True	NaN	NaN
1	2	22.0	NaN	12/24/2015	12/24/15	False	10:00	16:00
2	3	22.0	NaN	June 01| 0001	September 01| 0001	True	NaN	NaN
3	4	NaN	22.0	January 01| 0001	January 01| 0001	True	NaN	NaN

In [34]:

```
pd.read_csv("G:\\My Drive\\Data\\holiday_schedules.csv", sep= "|", skiprows= [1,3])
```

Out[34]:

	id	location_id	service_id	start_date	end_date	closed	opens_at	closes_at
0	2	22.0	NaN	12/24/2015	12/24/15	False	10:00	16:00
1	4	NaN	22.0	January 01| 0001	January 01| 0001	True	NaN	NaN

In [35]:

```
1 pd.read_csv("G:\\My Drive\\Data\\services.csv")
```

Out[35]:

	id	location_id	program_id	accepted_payments	alternate_name	application_process	audience	
0	1	1	NaN	NaN	NaN	Walk in or apply by phone.	Older adults age 55 or over, ethnic minorities...	A fo
1	2	2	NaN	NaN	NaN	Apply by phone for an appointment.	Residents of San Mateo County age 55 or over	Pro
2	3	3	NaN	NaN	NaN	Phone for information (403-4300 Ext. 4322).	Older adults age 55 or over who can benefit fr...	Offe se

In [36]:

```
1 df1 = pd.read_csv("G:\\My Drive\\Data\\services.csv")
```

In [37]:

```
1 df1
```

Out[37]:

	id	location_id	program_id	accepted_payments	alternate_name	application_process	audience	
0	1	1	NaN	NaN	NaN	Walk in or apply by phone.	Older adults age 55 or over, ethnic minorities...	A fo
1	2	2	NaN	NaN	NaN	Apply by phone for an appointment.	Residents of San Mateo County age 55 or over	Pro
2	3	3	NaN	NaN	NaN	Phone for information (403-4300 Ext. 4322).	Older adults age 55 or over who can benefit fr...	Offe se

In [38]:

```
1 type(df1)
```

Out[38]:

```
pandas.core.frame.DataFrame
```

In [40]:

```
1 df1.dtypes # it returns the data types of each and every column
```

Out[40]:

```
id                          int64
location_id                 int64
program_id                float64
accepted_payments          object
alternate_name             object
application_process        object
audience                   object
description                object
eligibility                object
email                      object
fees                       object
funding_sources            object
interpretation_services    object
keywords                   object
languages                  object
name                       object
required_documents         object
service_areas              object
status                     object
wait_time                  object
website                    object
taxonomy_ids               object
dtype: object
```

In [41]:

```
1 df1.head()  # first five records or rows of a dataframe
```

Out[41]:

	id	location_id	program_id	accepted_payments	alternate_name	application_process	audie
0	1	1	NaN	NaN	NaN	Walk in or apply by phone.	C adults 55 or e minorit
1	2	2	NaN	NaN	NaN	Apply by phone for an appointment.	Resic of M County 55 or
2	3	3	NaN	NaN	NaN	Phone for information (403-4300 Ext. 4322).	C adults 55 or whc benefi
3	4	4	NaN	NaN	NaN	Apply by phone.	Par chil fan prob c
4	5	5	NaN	NaN	NaN	Phone for information.	inc wo fan chil tr

5 rows × 22 columns

In [42]:

```
1 df1.head(2)
```

Out[42]:

	id	location_id	program_id	accepted_payments	alternate_name	application_process	audie
0	1	1	NaN	NaN	NaN	Walk in or apply by phone.	C adults 55 or e minorit
1	2	2	NaN	NaN	NaN	Apply by phone for an appointment.	Resic of M County 55 or

2 rows × 22 columns

In [43]:

```
1 df1.tail()
```

Out[43]:

	id	location_id	program_id	accepted_payments	alternate_name	application_process	au
18	19	19	NaN	NaN	NaN	Call for screening appointment (650-347-3648).	
19	20	20	NaN	NaN	NaN	Walk in.	
20	21	21	NaN	NaN	NaN	By phone during business hours.	
21	22	22	NaN	Cash, Check, Credit Card	Fotos para pasaportes	Walk in or apply by phone or mail	Pr nc busir the
22	23	22	NaN	NaN	NaN	Walk in or apply by phone or mail	S servi nc busir

5 rows × 22 columns

In [44]:

```
1 df1.tail(2
2         )
```

Out[44]:

	id	location_id	program_id	accepted_payments	alternate_name	application_process	au
21	22	22	NaN	Cash, Check, Credit Card	Fotos para pasaportes	Walk in or apply by phone or mail	Pr nc busir the
22	23	22	NaN	NaN	NaN	Walk in or apply by phone or mail	S servi nc busir

2 rows × 22 columns

In [ ]:

```
# in jupyter notebook we dont see all the reocrds but it reads all the records
```

In [45]:

```
df.columns
```

Out[45]:

```
Index(['Date Time', 'Temmperature', 'humidity', 'Wind speed',
       'General diffuse flows', 'power 1', 'power 2', 'power 3'],
      dtype='object')
```

In [46]:

```
df1.columns
```

Out[46]:

```
Index(['id', 'location_id', 'program_id', 'accepted_payments',
       'alternate_name', 'application_process', 'audience', 'description',
       'eligibility', 'email', 'fees', 'funding_sources',
       'interpretation_services', 'keywords', 'languages', 'name',
       'required_documents', 'service_areas', 'status', 'wait_time', 'websit
e',
       'taxonomy_ids'],
      dtype='object')
```

In [ ]:

```
# data from some of the columns
```

In [47]:

```
1 df1['service_areas']
```

Out[47]:

```
0                                    Colma
1                          San Mateo County
2                          San Mateo County
3                          San Mateo County
4                          San Mateo County
5                          San Mateo County
6       Belmont, Burlingame, East Palo Alto
7                   Belmont, East Palo Alto
8                          San Mateo County
9                          San Mateo County
10                         San Mateo County
11                                Daly City
12                         San Mateo County
13      Belmont, Burlingame, East Palo Alto
14         Alameda County, San Mateo County
15                                       NaN
16    Colma, Daly City, South San Francisco
17                           East Palo Alto
18                     Belmont, Burlingame
19                                       NaN
20                         San Mateo County
21         Alameda County, San Mateo County
22         San Mateo County, Alameda County
Name: service_areas, dtype: object
```

In [48]:

```
1 df1['status']
```

Out[48]:

```
0       active
1       active
2       active
3       active
4       active
5       active
6       active
7       active
8       active
9       active
10      active
11      active
12      active
13      active
14      active
15      active
16      active
17      active
18      active
19     defunct
20    inactive
21      active
22      active
Name: status, dtype: object
```

In [ ]:

```
1 # in case multiple columns
```

In [51]:

```
1 a = df1['status']
```

In [52]:

```
1 a
```

Out[52]:

```
0       active
1       active
2       active
3       active
4       active
5       active
6       active
7       active
8       active
9       active
10      active
11      active
12      active
13      active
14      active
15      active
16      active
17      active
18      active
19     defunct
20    inactive
21      active
22      active
Name: status, dtype: object
```

In [53]:

```
1 type(a)
```

Out[53]:

```
pandas.core.series.Series
```

In [54]:

```
1 type(df1['service_areas'])
```

Out[54]:

```
pandas.core.series.Series
```

In [ ]:

```
1 # in data frame we have series and dataframe
2 # data frame is same as all the operations in SQL
3 # in case you are going to pull out one single row and single column then it is call as
4 # data frame is a combination of list
```

In [55]:

```
1 df1['languages', 'name']
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
~\anaconda3\lib\site-packages\pandas\core\indexes\base.py in get_loc(self, k
ey, method, tolerance)
   3079             try:
-> 3080                 return self._engine.get_loc(casted_key)
   3081             except KeyError as err:

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHa
shTable.get_item()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHa
shTable.get_item()

KeyError: ('languages', 'name')

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
<ipython-input-55-ed452e6fb418> in <module>
----> 1 df1['languages', 'name']

~\anaconda3\lib\site-packages\pandas\core\frame.py in __getitem__(self, key)
   3022             if self.columns.nlevels > 1:
   3023                 return self._getitem_multilevel(key)
-> 3024             indexer = self.columns.get_loc(key)
   3025             if is_integer(indexer):
   3026                 indexer = [indexer]

~\anaconda3\lib\site-packages\pandas\core\indexes\base.py in get_loc(self, k
ey, method, tolerance)
   3080                 return self._engine.get_loc(casted_key)
   3081             except KeyError as err:
-> 3082                 raise KeyError(key) from err
   3083
   3084         if tolerance is not None:

KeyError: ('languages', 'name')
```

In [ ]:

```
1 # in case you have multiple columns you have to enclosed into 2 brackets
2 # pass list as data frame
3
```

In [56]:

```
1 df1[['languages', 'name']]
```

Out[56]:

	languages	name
0	NaN	Fair Oaks Adult Activity Center
1	NaN	Second Career Employment Program
2	NaN	Senior Peer Counseling
3	NaN	Family Visitation Center
4	NaN	Economic Self-Sufficiency Program
5	NaN	Little House Recreational Activities
6	NaN	Rosener House Adult Day Services
7	NaN	Meals on Wheels - South County
8	NaN	Fair Oaks Branch
9	NaN	Main Library
10	NaN	Schaberg Branch
11	NaN	Project Read
12	NaN	Redwood Shores Branch
13	NaN	Redwood City Corps
14	NaN	Adult Rehabilitation Center
15	NaN	Sunnyvale Corps
16	NaN	South San Francisco Citadel Corps
17	NaN	Project Smile
18	NaN	San Mateo Free Medical Clinic
19	NaN	Service with blank fields
20	NaN	Service for Admin Test Location
21	Spanish	Passport Photos
22	NaN	Example Service Name

In [58]:

```
b= df1[['languages', 'name']]
b
```

Out[58]:

	languages	name
0	NaN	Fair Oaks Adult Activity Center
1	NaN	Second Career Employment Program
2	NaN	Senior Peer Counseling
3	NaN	Family Visitation Center
4	NaN	Economic Self-Sufficiency Program
5	NaN	Little House Recreational Activities
6	NaN	Rosener House Adult Day Services
7	NaN	Meals on Wheels - South County
8	NaN	Fair Oaks Branch
9	NaN	Main Library
10	NaN	Schaberg Branch
11	NaN	Project Read
12	NaN	Redwood Shores Branch
13	NaN	Redwood City Corps
14	NaN	Adult Rehabilitation Center
15	NaN	Sunnyvale Corps
16	NaN	South San Francisco Citadel Corps
17	NaN	Project Smile
18	NaN	San Mateo Free Medical Clinic
19	NaN	Service with blank fields
20	NaN	Service for Admin Test Location
21	Spanish	Passport Photos
22	NaN	Example Service Name

In [59]:

```
type(df1[['languages', 'name']])
```

Out[59]:

```
pandas.core.frame.DataFrame
```

In [60]:

```
1 type(df1[['name']])
```

Out[60]:

```
pandas.core.frame.DataFrame
```

In [ ]:

```
1 # in case of a single column name in double brackets it is going ot pass as dataframe
```

In [ ]:

```
1 # EXCEL FILE READING
```

In [65]:

```
1 pd.read_excel("LUSID Excel - Manage Orders.xlsx")
```

Out[65]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Un
0	NaN	This sheet allows you to manage executions:	NaN	NaN	NaN	NaN	NaN	
1	NaN	1. List executions	NaN	NaN	NaN	NaN	NaN	
2	NaN	2. Get execution	NaN	NaN	NaN	NaN	NaN	
3	NaN	3. Upsert executions	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	
65	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
66	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
67	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
68	NaN	NaN	Code	NaN	NaN	NaN	NaN	
69	NaN	NaN	MyExecutionCode01	NaN	NaN	NaN	NaN	

70 rows × 25 columns

In [66]:

```
1 pd.read_excel("LUSID Excel - Manage Orders.xlsx", sheet_name= "Orders")
```

Out[66]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unn
0	NaN	This sheet allows you to manage Orders:	NaN	NaN	NaN	NaN	NaN	
1	NaN	1. ListOrders	NaN	NaN	NaN	NaN	NaN	
2	NaN	2. Getorder	NaN	NaN	NaN	NaN	NaN	
3	NaN	3. UpsertOrders	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	
66	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
67	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
68	NaN	NaN	Code	NaN	NaN	NaN	NaN	
69	NaN	NaN	MyOrderCode02	NaN	NaN	NaN	NaN	
70	NaN	NaN	MyOrderCode01	NaN	NaN	NaN	NaN	

71 rows × 25 columns

In [68]:

```
df2 = pd.read_excel("LUSID Excel - Manage Orders.xlsx")
df2
```

Out[68]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Un
0	NaN	This sheet allows you to manage executions:	NaN	NaN	NaN	NaN	NaN	
1	NaN	1. List executions	NaN	NaN	NaN	NaN	NaN	
2	NaN	2. Get execution	NaN	NaN	NaN	NaN	NaN	
3	NaN	3. Upsert executions	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	
65	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
66	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
67	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
68	NaN	NaN	Code	NaN	NaN	NaN	NaN	
69	NaN	NaN	MyExecutionCode01	NaN	NaN	NaN	NaN	

70 rows × 25 columns

In [3]:

```
# in case you want to know the sheet in the excel file
import pandas as pd
```

In [4]:

```
df2 = pd.ExcelFile("LUSID Excel - Manage Orders.xlsx")
```

In [5]:

```
df2.sheet_names
```

Out[5]:

```
['Executions', 'Orders', 'Allocations', 'Placements', 'Blocks']
```

In [ ]:

```
# read all the sheets one by one
```

In [6]:

```
k= []
for i in df2.sheet_names:
    k = pd.read_excel("LUSID Excel - Manage Orders.xlsx", sheet_name=i)

```

In [7]:

```
k
```

Out[7]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3
0	NaN	This sheet allows you to manage Blocks:	NaN	NaN
1	NaN	1. ListBlocks	NaN	NaN
2	NaN	2. GetBlock	NaN	NaN
3	NaN	3. UpsertBlocks	NaN	NaN
4	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN
6	NaN	NaN	NaN	NaN

In [8]:

```
1 df3 = pd.read_csv("https://github.com/datasciencedojo/datasets/blob/master/titanic.csv"
```

```
---------------------------------------------------------------------------
ParserError                               Traceback (most recent call last)
<ipython-input-8-998cadcc3eb5> in <module>
----> 1 df3 = pd.read_csv("https://github.com/datasciencedojo/datasets/blob/master/titanic.csv")

~\anaconda3\lib\site-packages\pandas\io\parsers.py in read_csv(filepath_or_buffer, sep, delimiter, header, names, index_col, usecols, squeeze, prefix, mangle_dupe_cols, dtype, engine, converters, true_values, false_values, skipinitialspace, skiprows, skipfooter, nrows, na_values, keep_default_na, na_filter, verbose, skip_blank_lines, parse_dates, infer_datetime_format, keep_date_col, date_parser, dayfirst, cache_dates, iterator, chunksize, compression, thousands, decimal, lineterminator, quotechar, quoting, doublequote, escapechar, comment, encoding, dialect, error_bad_lines, warn_bad_lines, delim_whitespace, low_memory, memory_map, float_precision, storage_options)
    608     kwds.update(kwds_defaults)
    609
--> 610     return _read(filepath_or_buffer, kwds)
    611
    612

~\anaconda3\lib\site-packages\pandas\io\parsers.py in _read(filepath_or_buffer, kwds)
    466
    467     with parser:
--> 468         return parser.read(nrows)
    469
    470

~\anaconda3\lib\site-packages\pandas\io\parsers.py in read(self, nrows)
   1055     def read(self, nrows=None):
   1056         nrows = validate_integer("nrows", nrows)
-> 1057         index, columns, col_dict = self._engine.read(nrows)
   1058
   1059         if index is None:

~\anaconda3\lib\site-packages\pandas\io\parsers.py in read(self, nrows)
   2059     def read(self, nrows=None):
   2060         try:
-> 2061             data = self._reader.read(nrows)
   2062         except StopIteration:
   2063             if self._first_chunk:

pandas\_libs\parsers.pyx in pandas._libs.parsers.TextReader.read()

pandas\_libs\parsers.pyx in pandas._libs.parsers.TextReader._read_low_memory()

pandas\_libs\parsers.pyx in pandas._libs.parsers.TextReader._read_rows()

pandas\_libs\parsers.pyx in pandas._libs.parsers.TextReader._tokenize_rows()

pandas\_libs\parsers.pyx in pandas._libs.parsers.raise_parser_error()
```

```
ParserError: Error tokenizing data. C error: Expected 1 fields in line 26,
saw 376
```

In [10]:

```
1 df3 = pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/master/ti
```

In [15]:

```
1 df3
```

Out[15]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

891 rows × 12 columns

In [14]:

```
1 df3 = pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/master/ti
```

PANDAS DAY 2

In [16]:

```
pd.read_html("https://www.basketball-reference.com/leagues/NBA_2015_totals.html")
# it will try to scrap  the data only in the tabular format
# it will only scrap the data only in the tabular format from the html webpage
# html dr or
```

Out[16]:

```
[      Rk           Player Pos Age   Tm   G  GS    MP   FG  FGA  ...   FT%  O
RB  \
 0      1       Quincy Acy  PF  24  NYK  68  22  1287  152  331  ...  .784
79
 1      2     Jordan Adams  SG  20  MEM  30   0   248   35   86  ...  .609
9
 2      3     Steven Adams   C  21  OKC  70  67  1771  217  399  ...  .502  1
99
 3      4      Jeff Adrien  PF  28  MIN  17   0   215   19   44  ...  .579
23
 4      5   Arron Afflalo  SG  29  TOT  78  72  2502  375  884  ...  .843
27
 ..   ...             ...  ..  ..  ...  ..  ..   ...  ...  ...  ...   ...
...
 670  490  Thaddeus Young  PF  26  TOT  76  68  2434  451  968  ...  .655  1
27
 671  490  Thaddeus Young  PF  26  MIN  48  48  1605  289  641  ...  .682
75
 672  490  Thaddeus Young  PF  26  BRK  28  20   829  162  327  ...  .606
52
 673  491     Cody Zeller   C  22  CHO  62  45  1487  172  373  ...  .774
97
 674  492    Tyler Zeller   C  25  BOS  82  59  1731  340  619  ...  .823  1
46

      DRB  TRB  AST  STL BLK  TOV   PF   PTS
 0    222  301   68   27  22   60  147   398
 1     19   28   16   16   7   14   24    94
 2    324  523   66   38  86   99  222   537
 3     54   77   15    4   9    9   30    60
 4    220  247  129   41   7  116  167  1035
 ..   ...  ...  ...  ...  ..  ...  ...   ...
 670  284  411  173  124  25  117  171  1071
 671  170  245  135   86  17   75  115   685
 672  114  166   38   38   8   42   56   386
 673  265  362  100   34  49   62  156   472
 674  319  465  113   18  52   76  205   833

 [675 rows x 30 columns]]
```

In [ ]:

```
# currently data set is not in a readable format
# We have to make the data in the readable format
```

In [17]:

```
df4= pd.read_html("https://www.basketball-reference.com/leagues/NBA_2015_totals.html")
```

In [18]:

```
1 type(df4)
```

Out[18]:

```
list
```

In [ ]:

```
1 # how many elements are there in the list
```

In [19]:

```
1 len(df4)
```

Out[19]:

```
1
```

In [20]:

```
1 df4[0]
```

Out[20]:

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	FT%	ORB	DRB	TRB	A
0	1	Quincy Acy	PF	24	NYK	68	22	1287	152	331	...	.784	79	222	301	
1	2	Jordan Adams	SG	20	MEM	30	0	248	35	86	...	.609	9	19	28	
2	3	Steven Adams	C	21	OKC	70	67	1771	217	399	...	.502	199	324	523	
3	4	Jeff Adrien	PF	28	MIN	17	0	215	19	44	...	.579	23	54	77	
4	5	Arron Afflalo	SG	29	TOT	78	72	2502	375	884	...	.843	27	220	247	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
670	490	Thaddeus Young	PF	26	TOT	76	68	2434	451	968	...	.655	127	284	411	1
671	490	Thaddeus Young	PF	26	MIN	48	48	1605	289	641	...	.682	75	170	245	1
672	490	Thaddeus Young	PF	26	BRK	28	20	829	162	327	...	.606	52	114	166	
673	491	Cody Zeller	C	22	CHO	62	45	1487	172	373	...	.774	97	265	362	1
674	492	Tyler Zeller	C	25	BOS	82	59	1731	340	619	...	.823	146	319	465	1

675 rows × 30 columns

In [21]:

```
1 type(df4[0])
```

Out[21]:

```
pandas.core.frame.DataFrame
```

In [23]:

```
1 df5 = df4[0]
```

In [24]:

```
1 df5.columns
```

Out[24]:

```
Index(['Rk', 'Player', 'Pos', 'Age', 'Tm', 'G', 'GS', 'MP', 'FG', 'FGA', 'F
G%',
       '3P', '3PA', '3P%', '2P', '2PA', '2P%', 'eFG%', 'FT', 'FTA', 'FT%',
       'ORB', 'DRB', 'TRB', 'AST', 'STL', 'BLK', 'TOV', 'PF', 'PTS'],
      dtype='object')
```

In [26]:

```
1 df5.head()
```

Out[26]:

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	FT%	ORB	DRB	TRB	AST	S
0	1	Quincy Acy	PF	24	NYK	68	22	1287	152	331	...	.784	79	222	301	68	
1	2	Jordan Adams	SG	20	MEM	30	0	248	35	86	...	.609	9	19	28	16	
2	3	Steven Adams	C	21	OKC	70	67	1771	217	399	...	.502	199	324	523	66	
3	4	Jeff Adrien	PF	28	MIN	17	0	215	19	44	...	.579	23	54	77	15	
4	5	Arron Afflalo	SG	29	TOT	78	72	2502	375	884	...	.843	27	220	247	129	

5 rows × 30 columns

In [27]:

```
1 df5.head(3)
```

Out[27]:

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	FT%	ORB	DRB	TRB	AST	S
0	1	Quincy Acy	PF	24	NYK	68	22	1287	152	331	...	.784	79	222	301	68	
1	2	Jordan Adams	SG	20	MEM	30	0	248	35	86	...	.609	9	19	28	16	
2	3	Steven Adams	C	21	OKC	70	67	1771	217	399	...	.502	199	324	523	66	

3 rows × 30 columns

In [29]:

```
1 df5.dtypes
```

Out[29]:

```
Rk        object
Player    object
Pos       object
Age       object
Tm        object
G         object
GS        object
MP        object
FG        object
FGA       object
FG%       object
3P        object
3PA       object
3P%       object
2P        object
2PA       object
2P%       object
eFG%      object
FT        object
FTA       object
FT%       object
ORB       object
DRB       object
TRB       object
AST       object
STL       object
BLK       object
TOV       object
PF        object
PTS       object
dtype: object
```

In [30]:

```
1 df5[['Age', 'DRB']]
```

Out[30]:

	Age	DRB
0	24	222
1	20	19
2	21	324
3	28	54
4	29	220
...	...	...
670	26	284
671	26	170
672	26	114
673	22	265
674	25	319

675 rows × 2 columns

In [31]:

```
1 df5.to_csv("NBA.csv")
```

In [32]:

```
1 ls
```

```
 Volume in drive C is Windows
 Volume Serial Number is B8C0-2582

 Directory of C:\Users\shrav\Documents\iNuron\pandas

07-03-2022  22:24    <DIR>          .
07-03-2022  21:52    <DIR>          ..
07-03-2022  16:32    <DIR>          .ipynb_checkpoints
07-03-2022  22:24           388,199 iNuron.ipynb
04-03-2022  08:28           161,651 LUSID Excel - Business Agility - Making
Simple Changes Quickly & Easily.xlsx
04-03-2022  08:28           161,410 LUSID Excel - Maintain a product in mult
iple currencies and share classes.xlsx
04-03-2022  08:28            24,675 LUSID Excel - Manage instruments with ec
onomic definitions.xlsx
04-03-2022  08:28            84,319 LUSID Excel - Manage Orders.xlsx
04-03-2022  08:28           137,753 LUSID Excel - Manage your investment str
ategies.xlsx
04-03-2022  08:28           184,013 LUSID Excel - Setting up your IBOR.xlsx
04-03-2022  08:28           136,632 LUSID Excel - Setting up your market dat
a.xlsx
07-03-2022  22:24            81,551 NBA.csv
01-03-2022  02:18           170,352 pandas.ipynb
              10 File(s)      1,530,555 bytes
               3 Dir(s)  381,250,396,160 bytes free
```

In [ ]:

```
1 # how to remove the row indexes
```

In [33]:

```
1 df5.to_csv("NBA_noindex.csv" , index= False)
```

In [35]:

```
1 l=pd.read_html('https://www.espncricinfo.com/series/world-cup-league-2-2019-2021-22-119
```

In [36]:

```
1 type(l)
```

Out[36]:

```
list
```

In [39]:

```
1 len(l)
```

Out[39]:

```
9
```

In [40]:

```
1 l[0]
```

Out[40]:

	BATTING	Unnamed: 1	R	B	4s	6s	
0	Kashyap Prajapati	c †Aravind b Junaid Siddique	8	15	1	0	
1	NaN	NaN	NaN	NaN	NaN	NaN	
2	Jatinder Singh	c †Aravind b Zahoor Khan	19	37	2	0	
3	NaN	NaN	NaN	NaN	NaN	NaN	
4	Shoaib Khan	c †Aravind b Zahoor Khan	61	89	3	0	
5	NaN	NaN	NaN	NaN	NaN	NaN	
6	Ayaan Khan	b Rohan Mustafa	39	66	1	0	
7	NaN	NaN	NaN	NaN	NaN	NaN	
8	Zeeshan Maqsood (c)	c Basil Hameed b Zahoor Khan	35	43	2	0	
9	NaN	NaN	NaN	NaN	NaN	NaN	
10	Suraj Kumar †	lbw b Basil Hameed	9	19	0	0	
11	NaN	NaN	NaN	NaN	NaN	NaN	
12	Sandeep Goud	st †Aravind b Rohan Mustafa	22	21	3	0	
13	NaN	NaN	NaN	NaN	NaN	NaN	
14	Mohammad Nadeem	not out	4	8	0	0	
15	NaN	NaN	NaN	NaN	NaN	NaN	
16	Kaleemullah	not out	2	2	0	0	
17	NaN	NaN	NaN	NaN	NaN	NaN	
18	Extras	(b 1, lb 8, w 17)	26	NaN	NaN	NaN	
19	TOTAL	(50 Ov, RR: 4.50)	225/7	NaN	NaN	NaN	
20	Did not bat: Bilal Khan, Khawar Ali	Did not bat: Bilal Khan, Khawar Ali	Did not bat: Bilal Khan, Khawar Ali	Did not bat: Bilal Khan, Khawar Ali	Did not bat: Bilal Khan, Khawar Ali	Did not bat: Bilal Khan, Khawar Ali	Did Khan,
21	Fall of wickets: 1-14 (Kashyap Prajapati, 4.2 ...	Fall of wickets: 1-14 (Kashyap Prajapati, 4.2 ...	Fall of wickets: 1-14 (Kashyap Prajapati, 4.2 ...	Fall of wickets: 1-14 (Kashyap Prajapati, 4.2 ...	Fall of wickets: 1-14 (Kashyap Prajapati, 4.2 ...	Fall of wickets: 1-14 (Kashyap Prajapati, 4.2 ...	wicket (K Prajap

In [49]:

```
for i in range(len(l)):
    l[i]
    print(l[i])
```

```
                                          BATTING  \
0                                Kashyap Prajapati
1                                              NaN
2                                   Jatinder Singh
3                                              NaN
4                                     Shoaib Khan
5                                              NaN
6                                      Ayaan Khan
7                                              NaN
8                             Zeeshan Maqsood (c)
9                                              NaN
10                                  Suraj Kumar †
11                                             NaN
12                                   Sandeep Goud
13                                             NaN
14                                Mohammad Nadeem
15                                             NaN
16                                    Kaleemullah
17                                             NaN
```

In [59]:

```
d ="""{

"name": "Adviteeya Shrav",
"email": "shrav@gmail.com",
"tech": ["ML", "DL","CV"]

}
"""

```

In [ ]:

```
# convert data into a tabular format
```

In [60]:

```
import json
```

In [61]:

```
json.loads(d)
```

Out[61]:

```
{'name': 'Adviteeya Shrav',
 'email': 'shrav@gmail.com',
 'tech': ['ML', 'DL', 'CV']}
```

In [ ]:

```
1
```

In [ ]:

```
1
```

In [ ]:

```
1
```

In [ ]:

```
1
```

In [62]:

```
1 type(d)
```

Out[62]:

```
str
```

In [63]:

```
1 m= json.loads(d)
```

In [64]:

```
1 m
```

Out[64]:

```
{'name': 'Adviteeya Shrav',
 'email': 'shrav@gmail.com',
 'tech': ['ML', 'DL', 'CV']}
```

In [65]:

```
1 type(m)
```

Out[65]:

```
dict
```

In [66]:

```
1 pd.read_json(m)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-66-6d3395c85d55> in <module>
----> 1 pd.read_json(m)

~\anaconda3\lib\site-packages\pandas\util\_decorators.py in wrapper(*args, *
*kwargs)
    197                 else:
    198                     kwargs[new_arg_name] = new_arg_value
--> 199             return func(*args, **kwargs)
    200
    201         return cast(F, wrapper)

~\anaconda3\lib\site-packages\pandas\util\_decorators.py in wrapper(*args, *
*kwargs)
    297                 )
    298                 warnings.warn(msg, FutureWarning, stacklevel=stackle
vel)
--> 299             return func(*args, **kwargs)
    300
    301         return wrapper

~\anaconda3\lib\site-packages\pandas\io\json\_json.py in read_json(path_or_b
uf, orient, typ, dtype, convert_axes, convert_dates, keep_default_dates, num
py, precise_float, date_unit, encoding, lines, chunksize, compression, nrow
s, storage_options)
    538         convert_axes = True
    539
--> 540     json_reader = JsonReader(
    541         path_or_buf,
    542         orient=orient,

~\anaconda3\lib\site-packages\pandas\io\json\_json.py in __init__(self, file
path_or_buffer, orient, typ, dtype, convert_axes, convert_dates, keep_defaul
t_dates, numpy, precise_float, date_unit, encoding, lines, chunksize, compre
ssion, nrows, storage_options)
    620                 raise ValueError("nrows can only be passed if lines=
True")
    621
--> 622         data = self._get_data_from_filepath(filepath_or_buffer)
    623         self.data = self._preprocess_data(data)
    624

~\anaconda3\lib\site-packages\pandas\io\json\_json.py in _get_data_from_file
path(self, filepath_or_buffer)
    657             or file_exists(filepath_or_buffer)
    658         ):
--> 659             self.handles = get_handle(
    660                 filepath_or_buffer,
    661                 "r",

~\anaconda3\lib\site-packages\pandas\io\common.py in get_handle(path_or_buf,
mode, encoding, compression, memory_map, is_text, errors, storage_options)
    556
    557     # open URLs
--> 558     ioargs = _get_filepath_or_buffer(
    559         path_or_buf,
```

```
    560         encoding=encoding,

~\anaconda3\lib\site-packages\pandas\io\common.py in _get_filepath_or_buffer
(filepath_or_buffer, encoding, compression, mode, storage_options)
    369     if not is_file_like(filepath_or_buffer):
    370         msg = f"Invalid file path or buffer object type: {type(filep
ath_or_buffer)}"
--> 371         raise ValueError(msg)
    372
    373     return IOArgs(

ValueError: Invalid file path or buffer object type: <class 'dict'>
```

In [67]:

```
pd.DataFrame(m)
```

Out[67]:

	name	email	tech
0	Adviteeya Shrav	shrav@gmail.com	ML
1	Adviteeya Shrav	shrav@gmail.com	DL
2	Adviteeya Shrav	shrav@gmail.com	CV

In [ ]:

```
############## list to DataFrame is possible###########
#############
```

In [68]:

```
d = {"packetType":"D","data":{"checkEngineLightFlag":"F","batteryVoltageStableTime":0,'
```

In [176]:

```
type(d)
```

Out[176]:

```
pandas.core.frame.DataFrame
```

In [70]:

```
d1 = d['data']
```

In [71]:

```
1 d1
```

Out[71]:


```
{'checkEngineLightFlag': 'F',
 'batteryVoltageStableTime': 0,
 'batteryVoltageStable': '0',
 'batteryVoltageOff': '12.42',
 'batteryCrankParamTN': '-0.08',
 'batteryCrankParamVN': '0.00',
 'batteryCrankParamTP': '-0.08',
 'batteryCrankParamVP': '0.00',
 'batteryCrankParamTT': '-0.00008',
 'batteryCrankParamV0': '0.00',
 'batteryVoltageMaxOn': '13.05',
 'batteryVoltageMinOn': '12.97',
 'batteryVoltageMaxOff': '12.46',
 'batteryVoltageMinOff': '12.36',
 'batteryVoltageOnAverage': '13.02',
 'engineLoadMax': '84',
 'engineLoadAverage': '39.98',
 'rpmMax': '3487',
 'rpmAverage': '1431.29',
 'gpsSpeedAverage': '21.99',
 'vssMax': '53.44',
 'vssAverage': '23.06',
 'tcuTemperatureMin': '82.40',
 'tcuTemperatureMax': '109.40',
 'tcuTemperatureAverage': '104.87',
 'coolantMin': '158.00',
 'coolantMax': '188.60',
 'coolantAverage': '180.20',
 'packetStartLocal': 1508143346000,
 'tripStartLocal': 1508143346000,
 'milIndicator': 'F',
 'monitorsNotReady': 0,
 'imei': '60DF5417',
 'gatewayTs': 1515613306592,
 'diagnosticTroubleCodeData': [],
 'diagnosticPidData': [[64768, 47, 100],
  [64768, 1, 517376],
  [64800, 1, 262144],
  [64768, 5, 125]]}
```

In [72]:

```
1 pd.DataFrame(d1)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-72-17ddab1ee376> in <module>
----> 1 pd.DataFrame(d1)

~\anaconda3\lib\site-packages\pandas\core\frame.py in __init__(self, data, i
ndex, columns, dtype, copy)
    527
    528         elif isinstance(data, dict):
--> 529             mgr = init_dict(data, index, columns, dtype=dtype)
    530         elif isinstance(data, ma.MaskedArray):
    531             import numpy.ma.mrecords as mrecords

~\anaconda3\lib\site-packages\pandas\core\internals\construction.py in init_
dict(data, index, columns, dtype)
    285             arr if not is_datetime64tz_dtype(arr) else arr.copy() fo
r arr in arrays
    286         ]
--> 287     return arrays_to_mgr(arrays, data_names, index, columns, dtype=d
type)
    288
    289

~\anaconda3\lib\site-packages\pandas\core\internals\construction.py in array
s_to_mgr(arrays, arr_names, index, columns, dtype, verify_integrity)
     78         # figure out the index, if necessary
     79         if index is None:
---> 80             index = extract_index(arrays)
     81         else:
     82             index = ensure_index(index)

~\anaconda3\lib\site-packages\pandas\core\internals\construction.py in extra
ct_index(data)
    399             lengths = list(set(raw_lengths))
    400             if len(lengths) > 1:
--> 401                 raise ValueError("arrays must all be same length")
    402
    403             if have_dicts:

ValueError: arrays must all be same length
```

In [73]:

```
1 del d1['diagnosticTroubleCodeData']
```

In [74]:

```
1 d1
```

Out[74]:


```
{'checkEngineLightFlag': 'F',
 'batteryVoltageStableTime': 0,
 'batteryVoltageStable': '0',
 'batteryVoltageOff': '12.42',
 'batteryCrankParamTN': '-0.08',
 'batteryCrankParamVN': '0.00',
 'batteryCrankParamTP': '-0.08',
 'batteryCrankParamVP': '0.00',
 'batteryCrankParamTT': '-0.00008',
 'batteryCrankParamV0': '0.00',
 'batteryVoltageMaxOn': '13.05',
 'batteryVoltageMinOn': '12.97',
 'batteryVoltageMaxOff': '12.46',
 'batteryVoltageMinOff': '12.36',
 'batteryVoltageOnAverage': '13.02',
 'engineLoadMax': '84',
 'engineLoadAverage': '39.98',
 'rpmMax': '3487',
 'rpmAverage': '1431.29',
 'gpsSpeedAverage': '21.99',
 'vssMax': '53.44',
 'vssAverage': '23.06',
 'tcuTemperatureMin': '82.40',
 'tcuTemperatureMax': '109.40',
 'tcuTemperatureAverage': '104.87',
 'coolantMin': '158.00',
 'coolantMax': '188.60',
 'coolantAverage': '180.20',
 'packetStartLocal': 1508143346000,
 'tripStartLocal': 1508143346000,
 'milIndicator': 'F',
 'monitorsNotReady': 0,
 'imei': '60DF5417',
 'gatewayTs': 1515613306592,
 'diagnosticPidData': [[64768, 47, 100],
  [64768, 1, 517376],
  [64800, 1, 262144],
  [64768, 5, 125]]}
```

In [75]:

```
1 pd.DataFrame(d1)
```

Out[75]:

	checkEngineLightFlag	batteryVoltageStableTime	batteryVoltageStable	batteryVoltageOff	batte
0	F	0	0	12.42	
1	F	0	0	12.42	
2	F	0	0	12.42	
3	F	0	0	12.42	

4 rows × 35 columns

In [ ]:

```
1 ############url = "https://api.github.com/repos/pandas-dev/pandas/issues"
```

In [76]:

```
1 pd.read_json("https://api.github.com/repos/pandas-dev/pandas/issues")
```

Out[76]:

	url	repository_url	lal
0	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
1	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
2	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
3	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
4	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
5	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
6	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
7	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
8	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
9	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
10	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
11	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
12	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
13	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
14	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
15	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p

	url	repository_url	lal
16	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
17	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
18	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
19	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
20	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
21	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
22	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
23	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
24	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
25	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
26	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
27	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
28	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
29	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p

30 rows × 29 columns

In [ ]:

```
# how to load the data into a raw json
# request a raw data from the url
```

In [77]:

```
import requests
url = "https://api.github.com/repos/pandas-dev/pandas/issues"
```

In [81]:

```
data = requests.get(url)
```

In [82]:

```
data
```

Out[82]:

```
<Response [200]>
```

In [83]:

```
data
```

Out[83]:

```
<Response [200]>
```

In [87]:

```
data1 = data.json()
data1
```

Out[87]:

```
[{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/46252',
  'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/462
52/labels{/name}',
  'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/4
6252/comments',
  'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/462
52/events',
  'html_url': 'https://github.com/pandas-dev/pandas/issues/46252',
  'id': 1161597938,
  'node_id': 'I_kwDOAA0YD85FPJPy',
  'number': 46252,
  'title': 'BUG: Period formatting directive %l (millisecond) and %u (micr
oseconds) provide wrong results',
  'user': {'login': 'smarie',
   'id': 3236794,
   'node_id': 'MDQ6VXNlcjMyMzY3OTQ=',
   'avatar_url': 'https://avatars.githubusercontent.com/u/3236794?v=4',
```

In [ ]:

```
# how many unique dictionaries available
```

In [88]:

```
1 len(data1)
```

Out[88]:

```
30
```

In [ ]:

```
1 ############ give me userid from one dictionary
```

In [92]:

```
1 data1[0]['user']
```

Out[92]:

```
{'login': 'smarie',
 'id': 3236794,
 'node_id': 'MDQ6VXNlcjMyMzY3OTQ=',
 'avatar_url': 'https://avatars.githubusercontent.com/u/3236794?v=4',
 'gravatar_id': '',
 'url': 'https://api.github.com/users/smarie',
 'html_url': 'https://github.com/smarie',
 'followers_url': 'https://api.github.com/users/smarie/followers',
 'following_url': 'https://api.github.com/users/smarie/following{/other_use
r}',
 'gists_url': 'https://api.github.com/users/smarie/gists{/gist_id}',
 'starred_url': 'https://api.github.com/users/smarie/starred{/owner}{/rep
o}',
 'subscriptions_url': 'https://api.github.com/users/smarie/subscriptions',
 'organizations_url': 'https://api.github.com/users/smarie/orgs',
 'repos_url': 'https://api.github.com/users/smarie/repos',
 'events_url': 'https://api.github.com/users/smarie/events{/privacy}',
 'received_events_url': 'https://api.github.com/users/smarie/received_event
s',
 'type': 'User',
 'site_admin': False}
```

In [93]:

```
1 data1[0]['user']['id']
```

Out[93]:

```
3236794
```

In [103]:

```
for i in range(len(data1)):
    print(data1[i]['user']['id'])

```

```
3236794
97615200
25685004
3385756
20880292
8078968
1461273
36619387
56136443
20362314
30196
54762742
1020496
26364415
13765633
6618166
10712109
25179211
18547401
45562402
10820686
414652
10820686
46936498
51996504
25440862
5902745
4156237
100502242
10297014
```

In [104]:

```
1 pd.DataFrame(data1)
```

Out[104]:

	url	repository_url	lal
0	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
1	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
2	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
3	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
4	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
5	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
6	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
7	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
8	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
9	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
10	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
11	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
12	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
13	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
14	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
15	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p

	url	repository_url	lal
16	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
17	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
18	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
19	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
20	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
21	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
22	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
23	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
24	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
25	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
26	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
27	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
28	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p
29	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.github.com/repos/dev/p

30 rows × 29 columns

In [108]:

```
1 pd.DataFrame(data1 , columns= ['id','url','repository_url','labels_url','comments_url']
```

Out[108]:

	id	url	repository_url	
0	1161597938	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
1	1161253063	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
2	1161193811	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
3	1161008604	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
4	1160722098	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
5	1160677456	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
6	1160431471	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
7	1160389512	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
8	1160064471	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
9	1159934807	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
10	1159874789	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
11	1159791857	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
12	1159657724	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
13	1159602954	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
14	1159567404	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
15	1159216249	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
16	1159140752	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
17	1159078523	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
18	1158374398	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
19	1157875795	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
20	1157819010	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
21	1157652834	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith

	id	url	repository_url	
22	1157500304	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
23	1155867508	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
24	1154543878	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
25	1154522327	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
26	1154193534	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
27	1153888271	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
28	1153565414	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
29	1153437571	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith

In [109]:

```
1 pd.DataFrame(data1 , columns= ['id','url','repository_url','labels_url','comments_url',
```

Out[109]:

	id	url	repository_url	
0	1161597938	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
1	1161253063	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
2	1161193811	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
3	1161008604	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
4	1160722098	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
5	1160677456	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
6	1160431471	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
7	1160389512	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
8	1160064471	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
9	1159934807	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
10	1159874789	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
11	1159791857	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
12	1159657724	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
13	1159602954	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
14	1159567404	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
15	1159216249	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
16	1159140752	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith

	id	url	repository_url	
17	1159078523	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
18	1158374398	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
19	1157875795	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
20	1157819010	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
21	1157652834	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
22	1157500304	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
23	1155867508	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
24	1154543878	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
25	1154522327	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
26	1154193534	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
27	1153888271	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
28	1153565414	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith
29	1153437571	https://api.github.com/repos/pandas-dev/pandas...	https://api.github.com/repos/pandas-dev/pandas	https://api.gith

In [117]:

```
k= pd.DataFrame(data1 , columns= ['id','url','repository_url','labels_url','comments_ur
```

In [118]:

```
k.to_csv('request.csv')
```

In [119]:

```
1 ls
```

```
 Volume in drive C is Windows
 Volume Serial Number is B8C0-2582

 Directory of C:\Users\shrav\Documents\iNuron\pandas

08-03-2022  00:30    <DIR>          .
07-03-2022  21:52    <DIR>          ..
07-03-2022  16:32    <DIR>          .ipynb_checkpoints
08-03-2022  00:30           904,932 iNuron.ipynb
04-03-2022  08:28           161,651 LUSID Excel - Business Agility - Making
Simple Changes Quickly & Easily.xlsx
04-03-2022  08:28           161,410 LUSID Excel - Maintain a product in mult
iple currencies and share classes.xlsx
04-03-2022  08:28            24,675 LUSID Excel - Manage instruments with ec
onomic definitions.xlsx
04-03-2022  08:28            84,319 LUSID Excel - Manage Orders.xlsx
04-03-2022  08:28           137,753 LUSID Excel - Manage your investment str
ategies.xlsx
04-03-2022  08:28           184,013 LUSID Excel - Setting up your IBOR.xlsx
04-03-2022  08:28           136,632 LUSID Excel - Setting up your market dat
a.xlsx
07-03-2022  22:24            81,551 NBA.csv
07-03-2022  22:29            78,960 NBA_noindex.csv
01-03-2022  02:18           170,352 pandas.ipynb
08-03-2022  00:30            36,911 request.csv
              12 File(s)      2,163,159 bytes
               3 Dir(s)  381,058,576,384 bytes free
```

In [142]:

```
1 u = pd.DataFrame(data1 , columns=['user'])
```

In [116]:

```


"""
{
'login': 'jbrockmendel',
'id': 8078968,
'node_id': 'MDQ6VXNlcjgwNzg5Njg=',
'avatar_url': 'https://avatars.githubusercontent.com/u/8078968?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/jbrockmendel',
'html_url': 'https://github.com/jbrockmendel',
'followers_url': 'https://api.github.com/users/jbrockmendel/followers',
'following_url': 'https://api.github.com/users/jbrockmendel/following{/other_user}',
'gists_url': 'https://api.github.com/users/jbrockmendel/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/jbrockmendel/starred{/owner}{/repo}',
'subscriptions_url': 'https://api.github.com/users/jbrockmendel/subscriptions',
'organizations_url': 'https://api.github.com/users/jbrockmendel/orgs',
'repos_url': 'https://api.github.com/users/jbrockmendel/repos',
'events_url': 'https://api.github.com/users/jbrockmendel/events{/privacy}',
'received_events_url': 'https://api.github.com/users/jbrockmendel/received_events',
'type': 'User', 'site_admin': False}
"""
```

In [164]:

```
k=pd.DataFrame(u, columns=['user'])

```

In [155]:

```
len(u)
```

Out[155]:

```
30
```

In [156]:

```
type(u)
```

Out[156]:

```
pandas.core.frame.DataFrame
```

In [166]:

```
type(k)
```

Out[166]:

```
pandas.core.frame.DataFrame
```

In [174]:

```
1 k
```

Out[174]:

	user
0	{'login': 'smarie', 'id': 3236794, 'node_id': ...
1	{'login': 'haritha1022', 'id': 97615200, 'node...
2	{'login': 'franklucky001', 'id': 25685004, 'no...
3	{'login': 'eltenedor', 'id': 3385756, 'node_id...
4	{'login': 'ryansdowning', 'id': 20880292, 'nod...
5	{'login': 'jbrockmendel', 'id': 8078968, 'node...
6	{'login': 'sterlinm', 'id': 1461273, 'node_id'...
7	{'login': 'weikhor', 'id': 36619387, 'node_id'...
8	{'login': 'LSturtew', 'id': 56136443, 'node_id...
9	{'login': 'ehallam', 'id': 20362314, 'node_id'...
10	{'login': 'ianliu', 'id': 30196, 'node_id': 'M...
11	{'login': 'raffaem', 'id': 54762742, 'node_id'...
12	{'login': 'jorisvandenbossche', 'id': 1020496,...
13	{'login': 'topper-123', 'id': 26364415, 'node_...
14	{'login': 'hitvoice', 'id': 13765633, 'node_id...
15	{'login': 'twoertwein', 'id': 6618166, 'node_i...
16	{'login': 'mvashishtha', 'id': 10712109, 'node...
17	{'login': 'sergiykhan', 'id': 25179211, 'node_...
18	{'login': 'moojen', 'id': 18547401, 'node_id':...
19	{'login': 'rhshadrach', 'id': 45562402, 'node_...
20	{'login': 'NickCrews', 'id': 10820686, 'node_i...
21	{'login': 'rendner', 'id': 414652, 'node_id': ...
22	{'login': 'NickCrews', 'id': 10820686, 'node_i...
23	{'login': 'NumberPiOso', 'id': 46936498, 'node...
24	{'login': 'theinexorable', 'id': 51996504, 'no...
25	{'login': 'RagBlufThim', 'id': 25440862, 'node...
26	{'login': 'johannes-mueller', 'id': 5902745, '...
27	{'login': 'tritemio', 'id': 4156237, 'node_id'...
28	{'login': 'mach881040', 'id': 100502242, 'node...
29	{'login': 'macsakow', 'id': 10297014, 'node_id...

In [181]:

```
1 k['user']
```

Out[181]:

```
0     {'login': 'smarie', 'id': 3236794, 'node_id': ...
1     {'login': 'haritha1022', 'id': 97615200, 'node...
2     {'login': 'franklucky001', 'id': 25685004, 'no...
3     {'login': 'eltenedor', 'id': 3385756, 'node_id...
4     {'login': 'ryansdowning', 'id': 20880292, 'nod...
5     {'login': 'jbrockmendel', 'id': 8078968, 'node...
6     {'login': 'sterlinm', 'id': 1461273, 'node_id'...
7     {'login': 'weikhor', 'id': 36619387, 'node_id'...
8     {'login': 'LSturtew', 'id': 56136443, 'node_id...
9     {'login': 'ehallam', 'id': 20362314, 'node_id'...
10    {'login': 'ianliu', 'id': 30196, 'node_id': 'M...
11    {'login': 'raffaem', 'id': 54762742, 'node_id'...
12    {'login': 'jorisvandenbossche', 'id': 1020496,...
13    {'login': 'topper-123', 'id': 26364415, 'node_...
14    {'login': 'hitvoice', 'id': 13765633, 'node_id...
15    {'login': 'twoertwein', 'id': 6618166, 'node_i...
16    {'login': 'mvashishtha', 'id': 10712109, 'node...
17    {'login': 'sergiykhan', 'id': 25179211, 'node_...
18    {'login': 'moojen', 'id': 18547401, 'node_id':...
19    {'login': 'rhshadrach', 'id': 45562402, 'node_...
20    {'login': 'NickCrews', 'id': 10820686, 'node_i...
21    {'login': 'rendner', 'id': 414652, 'node_id': ...
22    {'login': 'NickCrews', 'id': 10820686, 'node_i...
23    {'login': 'NumberPiOso', 'id': 46936498, 'node...
24    {'login': 'theinexorable', 'id': 51996504, 'no...
25    {'login': 'RagBlufThim', 'id': 25440862, 'node...
26    {'login': 'johannes-mueller', 'id': 5902745, '...
27    {'login': 'tritemio', 'id': 4156237, 'node_id'...
28    {'login': 'mach881040', 'id': 100502242, 'node...
29    {'login': 'macsakow', 'id': 10297014, 'node_id...
Name: user, dtype: object
```

In [201]:

```
1 len(k)
```

Out[201]:

```
30
```

In [185]:

```
1 n=k['user'][0]
```

In [186]:

```
1 n
```

Out[186]:


```
{'login': 'smarie',
 'id': 3236794,
 'node_id': 'MDQ6VXNlcjMyMzY3OTQ=',
 'avatar_url': 'https://avatars.githubusercontent.com/u/3236794?v=4',
 'gravatar_id': '',
 'url': 'https://api.github.com/users/smarie',
 'html_url': 'https://github.com/smarie',
 'followers_url': 'https://api.github.com/users/smarie/followers',
 'following_url': 'https://api.github.com/users/smarie/following{/other_use
r}',
 'gists_url': 'https://api.github.com/users/smarie/gists{/gist_id}',
 'starred_url': 'https://api.github.com/users/smarie/starred{/owner}{/rep
o}',
 'subscriptions_url': 'https://api.github.com/users/smarie/subscriptions',
 'organizations_url': 'https://api.github.com/users/smarie/orgs',
 'repos_url': 'https://api.github.com/users/smarie/repos',
 'events_url': 'https://api.github.com/users/smarie/events{/privacy}',
 'received_events_url': 'https://api.github.com/users/smarie/received_event
s',
 'type': 'User',
 'site_admin': False}
```


In [202]:

```
1 len(n)
```

Out[202]:

```
18
```

In [187]:

```
1 type(n)
```

Out[187]:

```
dict
```

In [198]:

```
1
2 o = pd.DataFrame(n ,  index= [0])
3 o
```

Out[198]:

	login	id	node_id	avatar_url	grav
0	smarie	3236794	MDQ6VXNlcjMyMzY3OTQ=	https://avatars.githubusercontent.com/u/323679...	

In [231]:

```
for i in range(len(k)):
    n=k['user'][i]
    #print(n)
    for j in range(len(n)):
        o = pd.DataFrame(n,index= [j])
        print(o)
```

```
    login       id                node_id  \
0  smarie  3236794  MDQ6VXNlcjMyMzY3OTQ=

                                          avatar_url gravatar_id  \
0  https://avatars.githubusercontent.com/u/323679... (https://avatars.gith
ubusercontent.com/u/323679...)

                                     url                   html_url  \
0  https://api.github.com/users/smarie (https://api.github.com/users/smari
e)  https://github.com/smarie (https://github.com/smarie)

                                     followers_url  \
0  https://api.github.com/users/smarie/followers (https://api.github.com/u
sers/smarie/followers)

                                       following_url  \
0  https://api.github.com/users/smarie/following{... (https://api.github.c
om/users/smarie/following{...)
```

In [232]:

```
1
```

In [ ]:

```
1
```

In [ ]:

```
1
```

In [ ]:

```
1
```

In [ ]:

```
1
```

In [208]:

```
1 type(o)
```

Out[208]:

```
pandas.core.frame.DataFrame
```

In [214]:

```
1 o.to_csv('usertest')
2
```

In [215]:

```
1 ls
```

```
 Volume in drive C is Windows
 Volume Serial Number is B8C0-2582

 Directory of C:\Users\shrav\Documents\iNuron\pandas

08-03-2022  01:37    <DIR>          .
07-03-2022  21:52    <DIR>          ..
07-03-2022  16:32    <DIR>          .ipynb_checkpoints
08-03-2022  01:36         1,908,295 iNuron.ipynb
04-03-2022  08:28           161,651 LUSID Excel - Business Agility - Making
Simple Changes Quickly & Easily.xlsx
04-03-2022  08:28           161,410 LUSID Excel - Maintain a product in mult
iple currencies and share classes.xlsx
04-03-2022  08:28            24,675 LUSID Excel - Manage instruments with ec
onomic definitions.xlsx
04-03-2022  08:28            84,319 LUSID Excel - Manage Orders.xlsx
04-03-2022  08:28           137,753 LUSID Excel - Manage your investment str
ategies.xlsx
04-03-2022  08:28           184,013 LUSID Excel - Setting up your IBOR.xlsx
04-03-2022  08:28           136,632 LUSID Excel - Setting up your market dat
a.xlsx
07-03-2022  22:24            81,551 NBA.csv
07-03-2022  22:29            78,960 NBA_noindex.csv
01-03-2022  02:18           170,352 pandas.ipynb
08-03-2022  00:30            36,911 request.csv
08-03-2022  01:36               844 user_test
08-03-2022  01:37               844 usertest
              14 File(s)      3,168,210 bytes
               3 Dir(s)  380,895,551,488 bytes free
```

```
1 #    # PANDAS DAY 3 DATA MANIPULATION
```

In [1]:

```
1 import pandas as pd
```

In [2]:

```
1 pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7

891 rows × 12 columns

In [3]:

```
1 df10 = pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/master/t
```

In [4]:

```
1 type(df10)
```

Out[4]:

```
pandas.core.frame.DataFrame
```

In [5]:

```
1 df10.head(10)
```

Out[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708

In [6]:

```
1 df10.tail(10)
```

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.895
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.516
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.500
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.050
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.125
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.450
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.750

In [7]:

```
1 df10.dtypes
```

Out[7]:

```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```

In [10]:

```
1 df10.describe() # considered a columns having datatypes numeric non categorical data
```

Out[10]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [11]:

```
1 df10.columns
```

Out[11]:

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [ ]:

```
1 # How to find the description of the categorical data
```

In [12]:

```
df10.dtypes == 'object' # filter out all the columns with categorical data or object da
```

Out[12]:

```
PassengerId    False
Survived       False
Pclass         False
Name            True
Sex             True
Age            False
SibSp          False
Parch          False
Ticket          True
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

In [13]:

```
type(df10.dtypes)
```

Out[13]:

```
pandas.core.series.Series
```

In [ ]:

```
# in case of series we have indices
```

In [14]:

```
df10.dtypes[df10.dtypes == 'object']
```

Out[14]:

```
Name        object
Sex         object
Ticket      object
Cabin       object
Embarked    object
dtype: object
```

In [ ]:

```
# my are of interest find the list of column name
```

In [ ]:

```
# list of the column
```

In [15]:

```
df10.dtypes[df10.dtypes == 'object'].index
```

Out[15]:

```
Index(['Name', 'Sex', 'Ticket', 'Cabin', 'Embarked'], dtype='object')
```

In [16]:

```
col_name = df10.dtypes[df10.dtypes == 'object'].index
```

In [17]:

```
col_name
```

Out[17]:

```
Index(['Name', 'Sex', 'Ticket', 'Cabin', 'Embarked'], dtype='object')
```

In [18]:

```
df10[col_name].describe()
```

Out[18]:

	Name	Sex	Ticket	Cabin	Embarked
count	891	891	891	204	889
unique	891	2	681	147	3
top	Taussig, Mr. Emil	male	CA. 2343	B96 B98	S
freq	1	577	7	4	644

In [ ]:

```
# Add a new column
```

In [19]:

```
df10['iNuron'] = 'Adviteeya'
```

In [20]:

```
1 df10
```

Out[20]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

891 rows × 13 columns

In [21]:

```
1 df10['Name']
```

Out[21]:

```
0                                Braund, Mr. Owen Harris
1      Cumings, Mrs. John Bradley (Florence Briggs Th...
2                                Heikkinen, Miss. Laina
3           Futrelle, Mrs. Jacques Heath (Lily May Peel)
4                              Allen, Mr. William Henry
                              ...
886                              Montvila, Rev. Juozas
887                       Graham, Miss. Margaret Edith
888           Johnston, Miss. Catherine Helen "Carrie"
889                              Behr, Mr. Karl Howell
890                                Dooley, Mr. Patrick
Name: Name, Length: 891, dtype: object
```

In [22]:

```
1 df10['Name'][0:15] # first 15 names
```

Out[22]:

```
0                                Braund, Mr. Owen Harris
1      Cumings, Mrs. John Bradley (Florence Briggs Th...
2                                Heikkinen, Miss. Laina
3           Futrelle, Mrs. Jacques Heath (Lily May Peel)
4                              Allen, Mr. William Henry
5                                      Moran, Mr. James
6                              McCarthy, Mr. Timothy J
7                       Palsson, Master. Gosta Leonard
8      Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)
9                    Nasser, Mrs. Nicholas (Adele Achem)
10                       Sandstrom, Miss. Marguerite Rut
11                             Bonnell, Miss. Elizabeth
12                      Saundercock, Mr. William Henry
13                         Andersson, Mr. Anders Johan
14                 Vestrom, Miss. Hulda Amanda Adolfina
Name: Name, dtype: object
```

In [23]:

```
1 df10['Name'][0:15:2]
```

Out[23]:

```
0                                Braund, Mr. Owen Harris
2                                Heikkinen, Miss. Laina
4                              Allen, Mr. William Henry
6                              McCarthy, Mr. Timothy J
8      Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)
10                       Sandstrom, Miss. Marguerite Rut
12                      Saundercock, Mr. William Henry
14                 Vestrom, Miss. Hulda Amanda Adolfina
Name: Name, dtype: object
```

In [24]:

```
1 df10['Name'][: : -1]
```

Out[24]:

```
890                                  Dooley, Mr. Patrick
889                                Behr, Mr. Karl Howell
888            Johnston, Miss. Catherine Helen "Carrie"
887                         Graham, Miss. Margaret Edith
886                                Montvila, Rev. Juozas
                              ...
4                                 Allen, Mr. William Henry
3           Futrelle, Mrs. Jacques Heath (Lily May Peel)
2                                  Heikkinen, Miss. Laina
1      Cumings, Mrs. John Bradley (Florence Briggs Th...
0                                Braund, Mr. Owen Harris
Name: Name, Length: 891, dtype: object
```

In [25]:

```
1 df10
```

Out[25]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

891 rows × 13 columns

In [ ]:

```
# On what indexes we have a null value
```

In [26]:

```
df10['Age'].isnull()
```

Out[26]:

```
0      False
1      False
2      False
3      False
4      False
       ...
886    False
887    False
888     True
889    False
890    False
Name: Age, Length: 891, dtype: bool
```

In [27]:

```
df10['Age'].isnull()== True
```

Out[27]:

```
0      False
1      False
2      False
3      False
4      False
       ...
886    False
887    False
888     True
889    False
890    False
Name: Age, Length: 891, dtype: bool
```

In [28]:

```
df10[df10['Age'].isnull()== True]
```

Out[28]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2250
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959	7.8792
...	...	...	...	...	...	...	...	...	...	...
859	860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2292
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500

177 rows × 13 columns

In [29]:

```
df10[df10['Age'].isnull()== True].index  # How to get the index with the null value
```

Out[29]:

```
Int64Index([  5,  17,  19,  26,  28,  29,  31,  32,  36,  42,
            ...
            832, 837, 839, 846, 849, 859, 863, 868, 878, 888],
           dtype='int64', length=177)
```

In [30]:

```
1 len(df10[df10['Age'].isnull()== True].index)
```

Out[30]:

177

In [ ]:

```
1 # How to go for the row wise selection
```

In [31]:

```
1 k = df10[df10['Age'].isnull()== True].index
```

In [33]:

```
1 df10.iloc[k]
```

Out[33]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2250
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959	7.8792
...	...	...	...	...	...	...	...	...	...	...
859	860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2292
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500

177 rows × 13 columns

In [196]:

```
1 df10.loc[k]
```

Out[196]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2250
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959	7.8792
...	...	...	...	...	...	...	...	...	...	...
859	860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2292
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500

177 rows × 14 columns

In [34]:

```
df10[df10['Age'].isnull()]
```

Out[34]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2250
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959	7.8792
...	...	...	...	...	...	...	...	...	...	...
859	860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2292
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500

177 rows × 13 columns

In [ ]:

```
# Find the person name who has paid the highest fare
```

In [40]:

```
df10[df10['Fare']==df10["Fare"].max()]
```

Out[40]:

```
258                      Ward, Miss. Anna
679    Cardeza, Mr. Thomas Drake Martinez
737                Lesurer, Mr. Gustave J
Name: Name, dtype: object
```

In [41]:

```
df10[df10['Fare']==df10["Fare"].max()]['Name']
```

Out[41]:

```
258                      Ward, Miss. Anna
679    Cardeza, Mr. Thomas Drake Martinez
737                Lesurer, Mr. Gustave J
Name: Name, dtype: object
```

In [ ]:

```
#Q1 How many male and female passenger was onboarded
#Q2 How many survived we have
#Q3 How many casualty we have
#Q4 What is the name of a person who is the eldest one
#Q5 How many passenger do we have in first , second and third class
#Q6 How many person we have whose name starts with letter "s"
#Q7 Try to create a new column which is a summation of "Sibsp" and "Parch"
#Q8 How many person do we have below age 25
#Q9 How many person died whose age was less than 40
#Q10 From a cabin column separate test and numeric value.

```

In [42]:

```
1 df10.head()
```

Out[42]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	

In [47]:

```
df10[df10['Sex'] == 'male']
```

Out[47]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750
...	...	...	...	...	...	...	...	...	...	...
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

577 rows × 13 columns

In [48]:

```
1 df10[df10['Sex'] == 'female']
```

Out[48]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708
...	...	...	...	...	...	...	...	...	...	...
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500

314 rows × 13 columns

In [51]:

```
1 df10[df10['Survived']== 1]
```

Out[51]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0
...	...	...	...	...	...	...	...	...	...	
875	876	1	3	Najib, Miss. Adele Kiamie "Jane"	female	15.0	0	0	2667	7.2
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0

342 rows × 13 columns

In [54]:

```
df10[df10['Survived']== 1]['Survived'].count()
```

Out[54]:

342

In [55]:

```
df10[df10['Survived']== 0]['Survived'].count()
```

Out[55]:

549

In [ ]:

```
#Q4 What is the name of a person who is the eldest one
```

In [70]:

```
df10[df10['Age']== df10['Age'].max()]
```

Out[70]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
630	631	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0	0	0	27042	30.0	A23

In [ ]:

```
#Q5 How many passenger do we have in first , second and third class
```

In [74]:

```
df10[df10['Pclass'] == 1]['Pclass'].count()
```

Out[74]:

216

In [75]:

```
df10[df10['Pclass'] == 2]['Pclass'].count()
```

Out[75]:

184

In [76]:

```
df10[df10['Pclass'] == 3]['Pclass'].count()
```

Out[76]:

491

In [ ]:

```
#Q6 How many person we have whose name starts with letter "s"
```

In [189]:

```
c= 0
for i in range(len(df10['Name'])):
    if(df10['Name'][i][0] == 'S'):
        print(df10['Name'][i])
        c= c+1
print("Count of names starting with S = ",c)

```

```
Sandstrom, Miss. Marguerite Rut
Saundercock, Mr. William Henry
Sloper, Mr. William Thompson
Spencer, Mrs. William Augustus (Marie Eugenie)
Samaan, Mr. Youssef
Sirayanian, Mr. Orsen
Skoog, Master. Harald
Stewart, Mr. Albert A
Staneff, Mr. Ivan
Sheerlinck, Mr. Jan Baptist
Slocovski, Mr. Selman Francis
Shorney, Mr. Charles Joseph
Salkjelsvik, Miss. Anna Kristine
Sobey, Mr. Samuel James Hayden
Smiljanic, Mr. Mile
Sage, Master. Thomas Henry
Skoog, Mrs. William (Anna Bernhardina Karlsson)
Sivola, Mr. Antti Wilhelm
Smith, Mr. James Clinch
Sage, Miss. Constance Gladys
Sage, Mr. Frederick
Strom, Miss. Telma Matilda
Sunderland, Mr. Victor Francis
Sjostedt, Mr. Ernst Adolf
Strom, Mrs. Wilhelm (Elna Matilda Persson)
Stead, Mr. William Thomas
Smith, Mr. Thomas
Smith, Mr. Richard William
Stankovic, Mr. Ivan
Saalfeld, Mr. Adolphe
Spedden, Mrs. Frederic Oakley (Margaretta Corning Stone)
Slayter, Miss. Hilda Mary
Sage, Mr. George John Jr
Sedgwick, Mr. Charles Frederick Waddington
Smith, Miss. Marion Elsie
Skoog, Mr. Wilhelm
Sadlier, Mr. Matthew
Sandstrom, Mrs. Hjalmar (Agnes Charlotta Bengtsson)
Sdycoff, Mr. Todor
Sundman, Mr. Johan Julian
Silven, Miss. Lyyli Karoliina
Silvey, Mr. William Baird
Seward, Mr. Frederic Kimber
Smart, Mr. John Montgomery
Scanlan, Mr. James
Strandberg, Miss. Ida Sofia
Somerton, Mr. Francis William
Stanley, Mr. Edward Roland
Shellard, Mr. Frederick William
Svensson, Mr. Olof
Salonen, Mr. Johan Werner
```

```
Sharp, Mr. Percival James R
Sivic, Mr. Husein
Simmons, Mr. John
Stoytcheff, Mr. Ilia
Silvey, Mrs. William Baird (Alice Munger)
Stephenson, Mrs. Walter Bertram (Martha Eustis)
Slabenoff, Mr. Petco
Shutes, Miss. Elizabeth W
Sutton, Mr. Frederick
Stahelin-Maeglin, Dr. Max
Skoog, Miss. Mabel
Sagesser, Mlle. Emma
Skoog, Miss. Margit Elizabeth
Simonius-Blumer, Col. Oberst Alfons
Stanley, Miss. Amy Zillah Elsie
Sawyer, Mr. Frederick Charles
Saad, Mr. Khalil
Silverthorne, Mr. Spencer Victor
Soholt, Mr. Peter Andreas Lauritz Andersen
Stranden, Mr. Juho
Sinkkonen, Miss. Anna
Sjoblom, Miss. Anna Sofia
Sage, Miss. Stella Anna
Slemen, Mr. Richard James
Skoog, Master. Karl Thorsten
Stone, Mrs. George Nelson (Martha Evelyn)
Saad, Mr. Amin
Sirota, Mr. Maurice
Serepeca, Miss. Augusta
Sage, Mr. Douglas Bullen
Svensson, Mr. Johan
Swift, Mrs. Frederick Joel (Margaret Welles Barron)
Sage, Miss. Dorothy Edith "Dolly"
Shelley, Mrs. William (Imanita Parrish Hall)
Sutehall, Mr. Henry Jr
Count of names starting with S =  86
```

In [ ]:

```
1  #Q7 Try to create a new column which is a summation of "Sibsp" and "Parch"
```

In [83]:

```
1  df10['New_Column'] = df10['SibSp'] + df10['Parch']
```

In [84]:

```
1 df10
```

Out[84]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

891 rows × 14 columns

In [ ]:

```
#Q9 How many person died whose age was less than 40
#Q10 From a cabin column separate test and numeric value.
```

In [ ]:

```

```

In [100]:

```
age_40=df10[df10['Age']< 40]
```

In [101]:

```
1 age_40
```

Out[101]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500
...	...	...	...	...	...	...	...	...	...	...
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

551 rows × 14 columns

In [104]:

```
1 age_40[age_40['Survived']==0]
```

Out[104]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909
12	13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151
13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082
...	...	...	...	...	...	...	...	...	...
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376

322 rows × 14 columns

In [ ]:

```
1 #Q10 From a cabin column separate text and numeric value.
```

In [105]:

```
1 df10['Cabin']
```

Out[105]:

```
0       NaN
1       C85
2       NaN
3      C123
4       NaN
       ...
886     NaN
887     B42
888     NaN
889    C148
890     NaN
Name: Cabin, Length: 891, dtype: object
```

In [127]:

```
1 df10[df10['Cabin'].isnull()]
```

Out[127]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500
...	...	...	...	...	...	...	...	...	...	...
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000

204 rows × 14 columns

In [128]:

```
1 import pandas as pd
```

In [138]:

```
1
2 df10[df10['Cabin'].notnull()]
```

Out[138]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.458
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.075
...	...	...	...	...	...	...	...	...	...	
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.050
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.125
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.450
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.750

687 rows × 14 columns

In [140]:

```
1 type(df10[df10['Cabin'].notnull()])
```

Out[140]:

```
pandas.core.frame.DataFrame
```

In [170]:

```
(df10[df10['Cabin'].notnull()]['Cabin']== int).count()
```

Out[170]:

687

In [172]:

```
(df10[df10['Cabin'].notnull()]['Cabin']== str).count()
```

Out[172]:

687

In [175]:

```
(df10[df10['Cabin'].isnull()])['Cabin'].count()
```

Out[175]:

0

In [176]:

```
df10['Cabin']
```

Out[176]:

```
0        1.0
1        NaN
2        3.0
3        NaN
4        5.0
       ...
886    887.0
887      NaN
888    889.0
889      NaN
890    891.0
Name: Cabin, Length: 891, dtype: object
```

```
# COMPLEX BANK DATA SET
```

In [190]:

```
ls

```

```
 Volume in drive C is Windows
 Volume Serial Number is B8C0-2582

 Directory of C:\Users\shrav\Documents\iNuron\pandas

10-03-2022  13:06    <DIR>          .
10-03-2022  00:07    <DIR>          ..
09-03-2022  20:28    <DIR>          .ipynb_checkpoints
14-02-2012  14:38           461,474 bank.csv
14-02-2012  14:40         4,610,348 bank-full.csv
14-02-2012  15:10             3,864 bank-names.txt
10-03-2022  12:11         2,106,019 iNuron.ipynb
04-03-2022  08:28           161,651 LUSID Excel - Business Agility - Making
Simple Changes Quickly & Easily.xlsx
04-03-2022  08:28           161,410 LUSID Excel - Maintain a product in mult
iple currencies and share classes.xlsx
04-03-2022  08:28            24,675 LUSID Excel - Manage instruments with ec
onomic definitions.xlsx
04-03-2022  08:28            84,319 LUSID Excel - Manage Orders.xlsx
04-03-2022  08:28           137,753 LUSID Excel - Manage your investment str
ategies.xlsx
04-03-2022  08:28           184,013 LUSID Excel - Setting up your IBOR.xlsx
04-03-2022  08:28           136,632 LUSID Excel - Setting up your market dat
a.xlsx
07-03-2022  22:24            81,551 NBA.csv
07-03-2022  22:29            78,960 NBA_noindex.csv
01-03-2022  02:18           170,352 pandas.ipynb
08-03-2022  23:08            29,048 request.csv
08-03-2022  01:45               844 user
08-03-2022  01:36               844 user_test
08-03-2022  01:37               844 usertest
              18 File(s)      8,434,601 bytes
               3 Dir(s)  380,632,764,416 bytes free
```

In [191]:

```
1 pd.read_csv("bank.csv" )
```

Out[191]:

	age;"job";"marital";"education";"default";"balance";"housing";"loan";"contact";"day";"month";"
0	
1	
2	
3	
4	
...	
4516	
4517	
4518	
4519	
4520	

4521 rows × 1 columns

In [194]:

```
1 pd.read_csv("bank.csv" , delimiter=';')
```

Out[194]:

	age	job	marital	education	default	balance	housing	loan	contact	day	mc
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	
1	33	services	married	secondary	no	4789	yes	yes	cellular	11	
2	35	management	single	tertiary	no	1350	yes	no	cellular	16	
3	30	management	married	tertiary	no	1476	yes	yes	unknown	3	
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	
...	...	...	...	...	...	...	...	...	...	...	
4516	33	services	married	secondary	no	-333	yes	no	cellular	30	
4517	57	self-employed	married	tertiary	yes	-3313	yes	yes	unknown	9	
4518	57	technician	married	secondary	no	295	no	no	cellular	19	
4519	28	blue-collar	married	secondary	no	1137	no	no	cellular	6	
4520	44	entrepreneur	single	tertiary	no	1136	yes	yes	cellular	3	

4521 rows × 17 columns

In [195]:

```
1 pd.read_csv("bank-full.csv" , delimiter=';')
```

Out[195]:

	age	job	marital	education	default	balance	housing	loan	contact	day
0	58	management	married	tertiary	no	2143	yes	no	unknown	5
1	44	technician	single	secondary	no	29	yes	no	unknown	5
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5
4	33	unknown	single	unknown	no	1	no	no	unknown	5
...	...	...	...	...	...	...	...	...	...	...
45206	51	technician	married	tertiary	no	825	no	no	cellular	17
45207	71	retired	divorced	primary	no	1729	no	no	cellular	17
45208	72	retired	married	secondary	no	5715	no	no	cellular	17
45209	57	blue-collar	married	secondary	no	668	no	no	telephone	17
45210	37	entrepreneur	married	secondary	no	2971	no	no	cellular	17

45211 rows × 17 columns

In [ ]:

```
"""1 . how many campaign available in this dataset .
2 . how many users do we have with housing and personal loan .
3. how many person do we have whose age is 60+ .
4 . in which month we have trageted most of the customer .
5 . which mode of call is giving you more result
6 . how many enterpeures do we have in this list
7 . how many customers do we have with negative balance
8. prepare a group of data based on education level .
"""
```

In [197]:

```
df11 = pd.read_csv("bank.csv" , delimiter=';')
```

In [198]:

```
df11
```

Out[198]:

	age	job	marital	education	default	balance	housing	loan	contact	day	mo
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	
1	33	services	married	secondary	no	4789	yes	yes	cellular	11	
2	35	management	single	tertiary	no	1350	yes	no	cellular	16	
3	30	management	married	tertiary	no	1476	yes	yes	unknown	3	
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	
...	...	...	...	...	...	...	...	...	...	...	
4516	33	services	married	secondary	no	-333	yes	no	cellular	30	
4517	57	self-employed	married	tertiary	yes	-3313	yes	yes	unknown	9	
4518	57	technician	married	secondary	no	295	no	no	cellular	19	
4519	28	blue-collar	married	secondary	no	1137	no	no	cellular	6	
4520	44	entrepreneur	single	tertiary	no	1136	yes	yes	cellular	3	

4521 rows × 17 columns

In [ ]:

```

```

In [207]:

```
#1 . how many campaign available in this dataset .
c=0
for i in range(len(df11['campaign'])):
    c= c+df11['campaign'][i]
print(c)
```

```
12630
```

In [221]:

```
#2 . how many users do we have with housing and personal loan .
housing=df11[df11['housing']=='yes']['housing'].count()
print("housing loan count :", housing)
personal=df11[df11['loan']=='yes']['loan'].count()
print("personal loan count :", personal)
```

```
housing loan count : 2559
personal loan count : 691
```

In [228]:

```
# 3. how many person do we have whose age is 60+ .
age60=df11[df11['age']>60]['age'].count()
print("Users above age  60 in the dataset is ", age60)
```

```
Users above age  60 in the dataset is  127
```

In [390]:

```
# 4 . in which month we have targeted most of the customer .
month=df11['month'].unique()
print(month)
len(month)
month[0]
```

```
['oct' 'may' 'apr' 'jun' 'feb' 'aug' 'jan' 'jul' 'nov' 'sep' 'mar' 'dec']
```

Out[390]:

```
'oct'
```

In [385]:

```
targeted_customer=df11[df11['y']== "yes"]
targeted_customer
```

Out[385]:

	age	job	marital	education	default	balance	housing	loan	contact	day	n
13	20	student	single	secondary	no	502	no	no	cellular	30	
30	68	retired	divorced	secondary	no	4189	no	no	telephone	14	
33	32	management	single	tertiary	no	2536	yes	no	cellular	26	
34	49	technician	married	tertiary	no	1235	no	no	cellular	13	
36	78	retired	divorced	primary	no	229	no	no	telephone	22	
...	...	...	...	...	...	...	...	...	...	...	
4494	26	technician	single	secondary	no	668	yes	no	unknown	28	
4503	60	self-employed	married	primary	no	362	no	yes	cellular	29	
4504	42	blue-collar	single	secondary	no	1080	yes	yes	cellular	13	
4505	32	admin.	single	secondary	no	620	yes	no	unknown	26	
4511	46	blue-collar	married	secondary	no	668	yes	no	unknown	15	

521 rows × 17 columns

In [394]:

```
l=targeted_users['month']
l
l.iloc[0]
```

Out[394]:

```
'apr'
```

In [411]:

```
j=0
a=0
ju=0
au=0
s=0
o=0
d=0
n=0
ma=0
jn=0
f=0
m=0
l=targeted_users['month']
for i in range(len(l)):
    if(l.iloc[i]== "jan"):
        j=j+1
    if(l.iloc[i]== "feb"):
        f=f+1
    if(l.iloc[i]== "mar"):
        m=m+1
    if(l.iloc[i]== "apr"):
        a=a+1
    if(l.iloc[i]== "may"):
        ma=ma+1
    if(l.iloc[i]== "jun"):
        jn=jn+1
    if(l.iloc[i]== "jul"):
        ju=ju+1
    if(l.iloc[i]== "aug"):
        au=au+1
    if(l.iloc[i]== "sep"):
        s=s+1
    if(l.iloc[i]== "oct"):
        o=o+1
    if(l.iloc[i]== "nov"):
        n=n+1
    if(l.iloc[i]== "dec"):
        d=d+1
print("Count in JAN month is :",j)
print("Count in FEB month is :",f)
print("Count in MAR month is :",m)
print("Count in APR month is :",a)
print("Count in MAY month is :",ma)
print("Count in JUN month is :",jn)
print("Count in JUL month is :",ju)
print("Count in AUG month is :",au)
print("Count in SEP month is :",s)
print("Count in OCT month is :",o)
print("Count in NOV month is :",n)
print("Count in DEC month is :",d)

```

```
Count in JAN month is : 16
Count in FEB month is : 38
Count in MAR month is : 21
Count in APR month is : 56
Count in MAY month is : 93
Count in JUN month is : 55
```

```
Count in JUL month is : 61
Count in AUG month is : 79
Count in SEP month is : 17
Count in OCT month is : 37
Count in NOV month is : 39
Count in DEC month is : 9
```

In [413]:

```
"""c=0
max=0
l=targeted_users['month']
for i in range(len(month)):
    for j in range(len(l)):
        if(month[i]==l.loc[j]):
            c=c+1
            pass
    if(c>max):
        max=c
        mon = l[i]
    else:
        max=max
print("Month that have targeted most of the customers is ", mon)
"""
```

Out[413]:

```
'c=0\nmax=0\nl=targeted_users[\'month\']\nfor i in range(len(month)):\n    f
or j in range(len(l)):\n        if(month[i]==l.loc[j]):\n            c=c+1\n
pass\n    if(c>max):\n        max=c\n        mon = l[i]\n    else:\n
max=max\nprint("Month that have targeted most of the customers is ", mon)\n'
```

In [ ]:

```

```

In [ ]:

```

```

In [414]:

```
#c=0
k=df11[df11['month']=="jan"]['y'] == "yes"
print(k)
l=len(k)

#for i in range(len(k)):
#    if(k== True):
#        c=c+1

print("User in JAN month is :", l)
```

```
14      False
24      False
49       True
66      False
113     False
        ...
4410    False
4434    False
4479    False
4482    False
4488    False
Name: y, Length: 148, dtype: bool
User in JAN month is : 148
```

In [415]:

```
df11
```

Out[415]:

	age	job	marital	education	default	balance	housing	loan	contact	day	mo
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	
1	33	services	married	secondary	no	4789	yes	yes	cellular	11	
2	35	management	single	tertiary	no	1350	yes	no	cellular	16	
3	30	management	married	tertiary	no	1476	yes	yes	unknown	3	
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	
...	...	...	...	...	...	...	...	...	...	...	
4516	33	services	married	secondary	no	-333	yes	no	cellular	30	
4517	57	self-employed	married	tertiary	yes	-3313	yes	yes	unknown	9	
4518	57	technician	married	secondary	no	295	no	no	cellular	19	
4519	28	blue-collar	married	secondary	no	1137	no	no	cellular	6	
4520	44	entrepreneur	single	tertiary	no	1136	yes	yes	cellular	3	

4521 rows × 17 columns

In [ ]:

```
#5 . which mode of call is giving you more result
```

In [424]:

```
call=df11['contact'].unique()
call
```

Out[424]:

```
array(['cellular', 'unknown', 'telephone'], dtype=object)
```

In [436]:

```
df11['contact'][df11['contact']== "cellular"].count()
```

Out[436]:

```
2896
```

In [437]:

```
df11['contact'][df11['contact']== "unknown"].count()
```

Out[437]:

```
1324
```

In [438]:

```
df11['contact'][df11['contact']== "telephone"].count()
```

Out[438]:

```
301
```

In [ ]:

```
"""6 . how many enterpeures do we have in this list

```

In [445]:

```
df11['job'][df11['job']== "entrepreneur"].count()

```

Out[445]:

```
168
```

In [ ]:

```
"""7 . how many customers do we have with negative balance

```

In [448]:

```
df11['balance'][df11['balance']<0].count()
```

Out[448]:

366

In [ ]:

```
"""8. prepare a group of data based on education level .

```

In [453]:

```
edu=df11['education'].unique()
edu
```

Out[453]:

array(['primary', 'secondary', 'tertiary', 'unknown'], dtype=object)

In [472]:

```
for i in range(len(edu)):
    c=df11['education'][df11['education']== edu[i]].count()
    print("Count of education of users in",edu[i],"=",c)



```

```
Count of education of users in primary = 678
Count of education of users in secondary = 2306
Count of education of users in tertiary = 1350
Count of education of users in unknown = 187
```

In [533]:

```
for i in range(len(edu)):
    d=df11[df11['education']== edu[i]]
    e= d['education'][d['y']== "yes"].count()
    print("Count of education of users who subscribed in",edu[i],"=",e)



```

```
Count of education of users who subscribed in primary = 64
Count of education of users who subscribed in secondary = 245
Count of education of users who subscribed in tertiary = 193
Count of education of users who subscribed in unknown = 19
```

In [534]:

```
1 df11
```

Out[534]:

	age	job	marital	education	default	balance	housing	loan	contact	day	mo
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	
1	33	services	married	secondary	no	4789	yes	yes	cellular	11	
2	35	management	single	tertiary	no	1350	yes	no	cellular	16	
3	30	management	married	tertiary	no	1476	yes	yes	unknown	3	
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	
...	...	...	...	...	...	...	...	...	...	...	
4516	33	services	married	secondary	no	-333	yes	no	cellular	30	
4517	57	self-employed	married	tertiary	yes	-3313	yes	yes	unknown	9	
4518	57	technician	married	secondary	no	295	no	no	cellular	19	
4519	28	blue-collar	married	secondary	no	1137	no	no	cellular	6	
4520	44	entrepreneur	single	tertiary	no	1136	yes	yes	cellular	3	

4521 rows × 17 columns

In [ ]:

```
1
```