

```
In [1]: #1,2,... read csv file.....for pandas#
import pandas as pd
data=pd.read_csv("eda_data.csv")
```

```
In [9]: print(data)

      Unnamed: 0      Job Title \
0      0      Data Scientist
1      1      Healthcare Data Scientist
2      2      Data Scientist
3      3      Data Scientist
4      4      Data Scientist
...
737      737      Sr Scientist, Immuno-Oncology - Oncology
738      738      Senior Data Engineer
739      739      Project Scientist - Auton Lab, Robotics Institute
740      740      Data Science Manager
741      741      Research Scientist - Security and Privacy

      Salary Estimate \
0      $53K-$91K (Glassdoor est.)
1      $63K-$112K (Glassdoor est.)
2      $80K-$90K (Glassdoor est.)
3      $56K-$97K (Glassdoor est.)
4      $86K-$143K (Glassdoor est.)
...
737      $58K-$111K (Glassdoor est.)
738      $72K-$133K (Glassdoor est.)
739      $56K-$91K (Glassdoor est.)
740      $95K-$160K (Glassdoor est.)
741      $61K-$126K (Glassdoor est.)

      Job Description      Rating \
0      Data Scientist\nLocation: Albuquerque, NM\nEdu...      3.8
1      What You Will Do:\n\nI. General Summary\n\nThe...      3.4
2      KnowBe4, Inc. is a high growth information sec...      4.8
3      *Organization and Job ID*\nJob ID: 310709\n\n...      3.8
4      Data Scientist\nAffinity Solutions / Marketing...      2.9
...
737      Site Name: USA - Massachusetts - Cambridge\nPo...      3.9
738      THE CHALLENGE\nEventbrite has a world-class da...      4.4
739      The Auton Lab at Carnegie Mellon University is...      2.6
740      Data Science ManagerResponsibilities:\n\nOvers...      3.2
741      Returning Candidate? Log back in to the Career...      3.6

      Company Name      Location \
0      Tecolote Research\n3.8      Albuquerque, NM
1      University of Maryland Medical System\n3.4      Linthicum, MD
2      KnowBe4\n4.8      Clearwater, FL
3      PNNL\n3.8      Richland, WA
4      Affinity Solutions\n2.9      New York, NY
...
737      GSK\n3.9      Cambridge, MA
738      Eventbrite\n4.4      Nashville, TN
739      Software Engineering Institute\n2.6      Pittsburgh, PA
740      Numeric, LLC\n3.2      Allentown, PA
741      Riverside Research Institute\n3.6      Beavercreek, OH

      Headquarters      Size      Founded      ...      age \
0      Goleta, CA      501 to 1000 employees      1973      ...      47
1      Baltimore, MD      10000+ employees      1984      ...      36
2      Clearwater, FL      501 to 1000 employees      2010      ...      10
3      Richland, WA      1001 to 5000 employees      1965      ...      55
4      New York, NY      51 to 200 employees      1998      ...      22
...
737      Brentford, United Kingdom      10000+ employees      1830      ...      190
738      San Francisco, CA      1001 to 5000 employees      2006      ...      14
739      Pittsburgh, PA      501 to 1000 employees      1984      ...      36
740      Chadds Ford, PA      1 to 50 employees      -1      ...      -1
741      Arlington, VA      501 to 1000 employees      1967      ...      53

      python_yn      R_yn      spark      aws      excel      job_simp      seniority      desc_len \
0      1      0      0      0      1      data scientist      na      2536
1      1      0      0      0      0      data scientist      na      4783
2      1      0      1      0      1      data scientist      na      3461
3      1      0      0      0      0      data scientist      na      3883
4      1      0      0      0      1      data scientist      na      2728
...
737      0      0      0      1      0      na      senior      6162
738      1      0      1      1      0      data engineer      senior      6130
739      0      0      0      0      1      na      na      3078
740      0      0      0      0      1      manager      na      1642
741      1      0      0      0      0      na      na      3673

      num_comp
0      0
1      0
2      0
3      3
4      3
...
737      3
738      3
739      0
740      0
741      0
```

Out[9]:

Unnamed: 0	Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	...	age	python_yn	R_yn	spark	aws	excel	job_simp	seniority	
0	0	Data Scientist	53K–91K (Glassdoor est.)	Data Scientist\nLocation: Albuquerque, NM\nEdu...	3.8	Tecolote Research\n3.8	Albuquerque, NM	Goleta, CA	501 to 1000 employees	1973	...	47	1	0	0	0	1	data scientist	na
1	1	Healthcare Data Scientist	63K–112K (Glassdoor est.)	What You Will Do*\n\nI. General Summary\n\nThe...	3.4	University of Maryland Medical System\n3.4	Linthicum, MD	Baltimore, MD	10000+ employees	1984	...	36	1	0	0	0	0	data scientist	na
2	2	Data Scientist	80K–90K (Glassdoor est.)	KnowBe4, Inc. is a high growth information sec...	4.8	KnowBe4\n4.8	Clearwater, FL	Clearwater, FL	501 to 1000 employees	2010	...	10	1	0	1	0	1	data scientist	na
3	3	Data Scientist	56K–97K (Glassdoor est.)	*Organization and Job ID*\nJob ID: 310709\n\n...	3.8	PNNL\n3.8	Richland, WA	Richland, WA	1001 to 5000 employees	1965	...	55	1	0	0	0	0	data scientist	na
4	4	Data Scientist	86K–143K (Glassdoor est.)	Data Scientist\nAffinity Solutions / Marketing...	2.9	Affinity Solutions\n2.9	New York, NY	New York, NY	51 to 200 employees	1998	...	22	1	0	0	0	1	data scientist	na
5	5	Data Scientist	71K–119K (Glassdoor est.)	CyrusOne is seeking a talented Data Scientist ...	3.4	CyrusOne\n3.4	Dallas, TX	Dallas, TX	201 to 500 employees	2000	...	20	1	0	0	1	1	data scientist	na
6	6	Data Scientist	54K–93K (Glassdoor est.)	Description\n\n**Please only local candida...	4.1	ClearOne Advantage\n4.1	Baltimore, MD	Baltimore, MD	501 to 1000 employees	2008	...	12	0	0	0	0	1	data scientist	na
7	7	Data Scientist	86K–142K (Glassdoor est.)	Advanced Analytics – Lead Data Scientist\nOver...	3.8	Logic20/20\n3.8	San Jose, CA	Seattle, WA	201 to 500 employees	2005	...	15	1	0	1	1	1	data scientist	na
8	8	Research Scientist	38K–84K (Glassdoor est.)	SUMMARY\n\nThe Research Scientist I will be ta...	3.3	Rochester Regional Health\n3.3	Rochester, NY	Rochester, NY	10000+ employees	2014	...	6	0	0	0	0	0	na	na
9	9	Data Scientist	120K–160K (Glassdoor est.)	isn't your usual company. Our work is powered ...	4.6	<intent>\n4.6	New York, NY	New York, NY	51 to 200 employees	2009	...	11	1	0	1	0	0	data scientist	na

10 rows × 33 columns

```
In [10]: data.head(10)
```

[10 rows x 33 columns]

Unnamed: 0	Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	...	age	python_yn	R_yn	spark	aws	excel	job_simp	seniority	
0	0	Data Scientist	53K – 91K (Glassdoor est.)	Data Scientist\nLocation: Albuquerque, NM\nEdu...	3.8	Tecolote Research\n3.8	Albuquerque, NM	Goleta, CA	501 to 1000 employees	1973	...	47	1	0	0	0	1	data scientist	na
1	1	Healthcare Data Scientist	63K – 112K (Glassdoor est.)	What You Will Do\n\nI. General Summary\n\nThe...	3.4	University of Maryland Medical System\n3.4	Linthicum, MD	Baltimore, MD	10000+ employees	1984	...	36	1	0	0	0	0	data scientist	na
2	2	Data Scientist	80K – 90K (Glassdoor est.)	KnowBe4, Inc. is a high growth information sec...	4.8	KnowBe4\n4.8	Clearwater, FL	Clearwater, FL	501 to 1000 employees	2010	...	10	1	0	1	0	1	data scientist	na
3	3	Data Scientist	56K – 97K (Glassdoor est.)	*Organization and Job ID*\nJob ID: 310709\n\n...	3.8	PNNL\n3.8	Richland, WA	Richland, WA	1001 to 5000 employees	1965	...	55	1	0	0	0	0	data scientist	na
4	4	Data Scientist	86K – 143K (Glassdoor est.)	Data Scientist\nAffinity Solutions / Marketing...	2.9	Affinity Solutions\n2.9	New York, NY	New York, NY	51 to 200 employees	1998	...	22	1	0	0	0	1	data scientist	na
5	5	Data Scientist	71K – 119K (Glassdoor est.)	CyrusOne is seeking a talented Data Scientist ...	3.4	CyrusOne\n3.4	Dallas, TX	Dallas, TX	201 to 500 employees	2000	...	20	1	0	0	1	1	data scientist	na
6	6	Data Scientist	54K – 93K (Glassdoor est.)	Job Description\n\n**Please only local candida...	4.1	ClearOne Advantage\n4.1	Baltimore, MD	Baltimore, MD	501 to 1000 employees	2008	...	12	0	0	0	0	1	data scientist	na
7	7	Data Scientist	86K – 142K (Glassdoor est.)	Advanced Analytics – Lead Data Scientist\nOver...	3.8	Logic20/20\n3.8	San Jose, CA	Seattle, WA	201 to 500 employees	2005	...	15	1	0	1	1	1	data scientist	na
8	8	Research Scientist	38K – 84K (Glassdoor est.)	SUMMARY\n\nThe Research Scientist I will be ta...	3.3	Rochester Regional Health\n3.3	Rochester, NY	Rochester, NY	10000+ employees	2014	...	6	0	0	0	0	0	na	na
9	9	Data Scientist	120K – 160K (Glassdoor est.)	isn't your usual company. Our work is powered ...	4.6	<intent>\n4.6	New York, NY	New York, NY	51 to 200 employees	2009	...	11	1	0	1	0	0	data scientist	na

10 rows x 33 columns

```
In [11]: # 3....clean empty cell.....#
data.dropna()
```

data.dropna()

Out [11]:

Unnamed: 0	Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	...	age	python_yn	R_yn	spark	aws	excel	job_simp	seniority		
0	0	Data Scientist	53K–91K (Glassdoor est.)	Data Scientist\nLocation: Albuquerque, NM\nEdu...	3.8	Tecolote Research\n3.8	Albuquerque, NM	Goleta, CA	501 to 1000 employees	1973	...	47	1	0	0	0	1	data scientist	na	
1	1	Healthcare Data Scientist	63K–112K (Glassdoor est.)	What You Will Do:\n\nI. General Summary\n\nThe...	3.4	University of Maryland Medical System\n3.4	Linthicum, MD	Baltimore, MD	10000+ employees	1984	...	36	1	0	0	0	0	data scientist	na	
2	2	Data Scientist	80K–90K (Glassdoor est.)	KnowBe4, Inc. is a high growth information sec...	4.8	KnowBe4\n4.8	Clearwater, FL	Clearwater, FL	501 to 1000 employees	2010	...	10	1	0	1	0	1	data scientist	na	
3	3	Data Scientist	56K–97K (Glassdoor est.)	*Organization and Job ID*\nJob ID: 310709\n\n...	3.8	PNNL\n3.8	Richland, WA	Richland, WA	1001 to 5000 employees	1965	...	55	1	0	0	0	0	data scientist	na	
4	4	Data Scientist	86K–143K (Glassdoor est.)	Data Scientist\nAffinity Solutions / Marketing...	2.9	Affinity Solutions\n2.9	New York, NY	New York, NY	51 to 200 employees	1998	...	22	1	0	0	0	1	data scientist	na	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
737	737	Sr Scientist, Immuno-Oncology - Oncology	58K–111K (Glassdoor est.)	Site Name: USA - Massachusetts - Cambridge\nPo...	3.9	GSK\n3.9	Cambridge, MA	Brentford, United Kingdom	10000+ employees	1830	...	190	0	0	0	0	1	0	data scientist	na
738	738	Senior Data Engineer	72K–133K (Glassdoor est.)	THE CHALLENGE\nEventbrite has a world-class da...	4.4	Eventbrite\n4.4	Nashville, TN	San Francisco, CA	1001 to 5000 employees	2006	...	14	1	0	1	1	0	0	data engineer	na
739	739	Project Scientist - Auton Lab, Robotics Institute	56K–91K (Glassdoor est.)	The Auton Lab at Carnegie Mellon University is...	2.6	Software Engineering Institute\n2.6	Pittsburgh, PA	Pittsburgh, PA	501 to 1000 employees	1984	...	36	0	0	0	0	0	1	data scientist	na
740	740	Data Science Manager	95K–160K (Glassdoor est.)	Data Science ManagerResponsibilities:\n\nOvers...	3.2	Numeric, LLC\n3.2	Allentown, PA	Chadds Ford, PA	1 to 50 employees	-1	...	-1	0	0	0	0	1	1	manager	na
741	741	Research Scientist – Security and Privacy	61K–126K (Glassdoor est.)	Returning Candidate? Log back in to the Career...	3.6	Riverside Research Institute\n3.6	Beavercreek, OH	Arlington, VA	501 to 1000 employees	1967	...	53	1	0	0	0	0	0	data scientist	na

742 rows × 33 columns

```
In [14]: #.....remove duplicates.....#
data.drop_duplicates()
```

data.drop\_duplicates()

Out [41]:

Unnamed: 0	Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	...	age	python_yn	R_yn	spark	aws	excel	job_simp	seniority	
0	0	Data Scientist	53K–91K (Glassdoor est.)	Data Scientist\nLocation: Albuquerque, NM\nEdu...	3.8	Tecolote Research\n3.8	Albuquerque, NM	Goleta, CA	501 to 1000 employees	1973	...	47	1	0	0	0	1	data scientist	na
1	1	Healthcare Data Scientist	63K–112K (Glassdoor est.)	What You Will Do:\n\nI. General Summary\n\nThe...	3.4	University of Maryland Medical System\n3.4	Linthicum, MD	Baltimore, MD	10000+ employees	1984	...	36	1	0	0	0	0	data scientist	na
2	2	Data Scientist	80K–90K (Glassdoor est.)	KnowBe4, Inc. is a high growth information sec...	4.8	KnowBe4\n4.8	Clearwater, FL	Clearwater, FL	501 to 1000 employees	2010	...	10	1	0	1	0	1	data scientist	na
3	3	Data Scientist	56K–97K (Glassdoor est.)	*Organization and Job ID*\nJob ID: 310709\n\n...	3.8	PNNL\n3.8	Richland, WA	Richland, WA	1001 to 5000 employees	1965	...	55	1	0	0	0	0	data scientist	na
4	4	Data Scientist	86K–143K (Glassdoor est.)	Data Scientist\nAffinity Solutions / Marketing...	2.9	Affinity Solutions\n2.9	New York, NY	New York, NY	51 to 200 employees	1998	...	22	1	0	0	0	1	data scientist	na
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
737	737	Sr Scientist, Immuno-Oncology - Oncology	58K–111K (Glassdoor est.)	Site Name: USA - Massachusetts - Cambridge\nPo...	3.9	GSK\n3.9	Cambridge, MA	Brentford, United Kingdom	10000+ employees	1830	...	190	0	0	0	1	0	data scientist	na
738	738	Senior Data Engineer	72K–133K (Glassdoor est.)	THE CHALLENGE\nEventbrite has a world-class da...	4.4	Eventbrite\n4.4	Nashville, TN	San Francisco, CA	1001 to 5000 employees	2006	...	14	1	0	1	1	0	data engineer	na
739	739	Project Scientist - Auton Lab, Robotics Institute	56K–91K (Glassdoor est.)	The Auton Lab at Carnegie Mellon University is...	2.6	Software Engineering Institute\n2.6	Pittsburgh, PA	Pittsburgh, PA	501 to 1000 employees	1984	...	36	0	0	0	0	1	data scientist	na
740	740	Data Science Manager	95K–160K (Glassdoor est.)	Data Science ManagerResponsibilities:\n\nOvers...	3.2	Numeric, LLC\n3.2	Allentown, PA	Chadds Ford, PA	1 to 50 employees	-1	...	-1	0	0	0	0	1	manager	na
741	741	Research Scientist – Security and Privacy	61K–126K (Glassdoor est.)	Returning Candidate? Log back in to the Career...	3.6	Riverside Research Institute\n3.6	Beavercreek, OH	Arlington, VA	501 to 1000 employees	1967	...	53	1	0	0	0	0	data scientist	na

742 rows × 33 columns

```
In [16]: data.duplicated().sum()
```

```
Out[16]: 0
```

```
In [2]: #for numpys.....
import numpy as np
from array import *
ar= np.array([1,25,45,85,245,10,2,20,21,43])

#.....print(np.where(a==))......
print(ar[1:6])
print(ar[-5:-2])

ar1= np.array([[12,14,15,16,18],[19,45,23,7,6],[45,12,32,45,74]])
print(ar1.shape)

for i in range(0,3):
    for j in range(0,5):
        print(ar1[i][j], end =" ")
    print()

#.....split the array.....
print(np.array_split(ar, 4))

[ 25  45  85 245  10]
[10  2 20]
(3, 5)
12  14  15  16  18
19  45  23  7  6
45  12  32  45  74
[array([ 1, 25, 45]), array([ 85, 245, 10]), array([ 2, 20]), array([21, 43])]
```

```
In [ ]:
```