



PANDAS

By Aditya Rana

Week #3

Data Cleaning using Pandas & Numpy

Uncleaned dataset

```
|> data = pd.read_csv ("/content/Data-cleaning-for-beginners-using-pandas.csv")
|> print(data)
```

Index	Age	Salary	Rating	Location	Established	Easy	Apply
0	44.0	\$44k-\$99k	5.4	India,In	1999	TRUE	
1	66.0	\$55k-\$66k	3.5	New York,Ny	2002	TRUE	
2	NaN	\$77k-\$89k	-1.0	New York,Ny	-1	-1	
3	64.0	\$44k-\$99k	4.4	India In	1988	-1	
4	25.0	\$44k-\$99k	6.4	Australia Aus	2002	-1	
5	44.0	\$77k-\$89k	1.4	India,In	1999	TRUE	
6	21.0	\$44k-\$99k	0.0	New York,Ny	-1	-1	
7	44.0	\$44k-\$99k	-1.0	Australia Aus	-1	-1	
8	35.0	\$44k-\$99k	5.4	New York,Ny	-1	-1	
9	22.0	\$44k-\$99k	7.7	India,In	-1	TRUE	
10	55.0	\$10k-\$49k	5.4	India,In	2008	TRUE	
11	44.0	\$10k-\$49k	6.7	India,In	2009	-1	
12	NaN	\$44k-\$99k	0.0	India,In	1999	-1	
13	25.0	\$44k-\$99k	-1.0	Australia Aus	2019	TRUE	
14	66.0	\$44k-\$99k	4.0	Australia Aus	2020	TRUE	
15	44.0	\$88k-\$101k	3.0	Australia Aus	1999	-1	
16	19.0	\$19k-\$40k	4.5	India,In	1984	-1	
17	NaN	\$44k-\$99k	5.3	New York,Ny	1943	TRUE	
18	35.0	\$44k-\$99k	6.7	New York,Ny	1954	TRUE	
19	32.0	\$44k-\$99k	3.3	New York,Ny	1955	TRUE	
20	NaN	\$44k-\$99k	5.7	New York,Ny	1944	TRUE	
21	35.0	\$44k-\$99k	5.0	New York,Ny	1946	-1	
22	19.0	\$55k-\$66k	7.8	New York,Ny	1988	TRUE	
23	NaN	\$44k-\$99k	2.4	New York,Ny	1999	TRUE	
..							
5	25	55.0	\$44k-\$99k	0.0	Australia Aus	1980	TRUE
6	NaN	\$55k-\$66k	NaN	India,In	1934	TRUE	
7	27	52.0	\$44k-\$99k	5.4	India,In	1935	-1
8	28	NaN	\$39k-\$88k	3.4	Australia Aus	1932	-1

Cleaned dataset

Unique values in the 'Easy Apply' column:
[nan]

Data type of the 'Easy Apply' column: object

Dataset after transforming 'Easy Apply' column:

Index	Age	Salary	Rating	Location	Established	Easy	Apply
0	0	44.0	\$44k-\$99k	5.4	india,in	1999	NaN
1	1	66.0	\$55k-\$66k	3.5	new york,ny	2002	NaN
2	2	39.0	\$77k-\$89k	-1.0	new york,ny	-1	NaN
3	3	64.0	\$44k-\$99k	4.4	india,in	1988	NaN
4	4	25.0	\$44k-\$99k	6.4	australia aus	2002	NaN
5	5	44.0	\$77k-\$89k	1.4	india,in	1999	NaN
6	6	21.0	\$44k-\$99k	0.0	new york,ny	-1	NaN
7	7	44.0	\$44k-\$99k	-1.0	australia aus	-1	NaN
8	8	35.0	\$44k-\$99k	5.4	new york,ny	-1	NaN
9	9	22.0	\$44k-\$99k	7.7	india,in	-1	NaN
10	10	55.0	\$10k-\$49k	5.4	india,in	2008	NaN
11	11	44.0	\$10k-\$49k	6.7	india,in	2009	NaN
12	12	39.0	\$44k-\$99k	0.0	india,in	1999	NaN
13	13	25.0	\$44k-\$99k	-1.0	australia aus	2019	NaN
14	14	66.0	\$44k-\$99k	4.0	australia aus	2020	NaN
15	15	44.0	\$88k-\$101k	3.0	australia aus	1999	NaN
16	16	19.0	\$19k-\$40k	4.5	india,in	1984	NaN
17	17	39.0	\$44k-\$99k	5.3	new york,ny	1943	NaN
..							
25	25	55.0	\$44k-\$99k	0.0	australia aus	1980	NaN
26	26	39.0	\$55k-\$66k	NaN	india,in	1934	NaN
27	27	52.0	\$44k-\$99k	5.4	india,in	1935	NaN
28	28	39.0	\$39k-\$88k	3.4	australia aus	1932	NaN

Uncleaned dataset

	order_id	quantity	item_name	choice_description	item_price
0	1	1	Chips and Fresh Tomato Salsa	NaN	\$2.39
1	1	1	Izze	[Clementine]	\$3.39
2	1	1	Nantucket Nectar	[Apple]	\$3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	\$2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	\$16.98
...
4617	1833	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Sour ...	\$11.75
4618	1833	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Sour Cream, Cheese...	\$11.75
4619	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	\$11.25
4620	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Lettu...	\$8.75
4621	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	\$8.75

Cleaned dataset

	order_id	quantity	standardized_item_name	choice_description	item_price
0	1	1	Chips and Fresh Tomato Salsa	None	2.75
1	1	1	Izze	Clementine	3.39
2	1	1	Nantucket Nectar	Apple	3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	None	2.39
4	2	2	Chicken Bowl	Tomatillo-Red Chili Salsa (Hot), Black Beans, ...	19.33
...
4558	1833	1	Steak Burrito	Fresh Tomato Salsa, Rice, Black Beans, Sour Cr...	9.99
4559	1833	1	Steak Burrito	Fresh Tomato Salsa, Rice, Sour Cream, Cheese, ...	9.99
4560	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa, Fajita Vegetables, Pinto B...	9.91
4561	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa, Fajita Vegetables, Lettuce	9.91
4562	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa, Fajita Vegetables, Pinto B...	9.91