



Week 3 :
Data Cleaning using Pandas

Assignment 1 : Data Cleaning (Pandas)

Uncleaned Data

Cleaned Data

	Index	Age	Salary	Rating	Location	Established	Easy Apply
0	0	44.0	\$44k-\$99k	5.4	India,In	1999	TRUE
1	1	66.0	\$55k-\$66k	3.5	New York,Ny	2002	TRUE
2	2	NaN	\$77k-\$89k	-1.0	New York,Ny	-1	-1
3	3	64.0	\$44k-\$99k	4.4	India,In	1988	-1
4	4	25.0	\$44k-\$99k	6.4	Australia,Aus	2002	-1
5	5	44.0	\$77k-\$89k	1.4	India,In	1999	TRUE
6	6	21.0	\$44k-\$99k	0.0	New York,Ny	-1	-1
7	7	44.0	\$44k-\$99k	-1.0	Australia,Aus	-1	-1
8	8	35.0	\$44k-\$99k	5.4	New York,Ny	-1	-1
9	9	22.0	\$44k-\$99k	7.7	India,In	-1	TRUE
10	10	55.0	\$10k-\$49k	5.4	India,In	2008	TRUE
11	11	44.0	\$10k-\$49k	6.7	India,In	2009	-1
12	12	NaN	\$44k-\$99k	0.0	India,In	1999	-1
13	13	25.0	\$44k-\$99k	-1.0	Australia,Aus	2019	TRUE
14	14	66.0	\$44k-\$99k	4.0	Australia,Aus	2020	TRUE
15	15	44.0	\$88k-\$101k	3.0	Australia,Aus	1999	-1
16	16	19.0	\$19k-\$40k	4.5	India,In	1984	-1
17	17	NaN	\$44k-\$99k	5.3	New York,Ny	1943	TRUE
18	18	35.0	\$44k-\$99k	6.7	New York,Ny	1954	TRUE
19	19	32.0	\$44k-\$99k	3.3	New York,Ny	1955	TRUE
20	20	NaN	\$44k-\$99k	5.7	New York,Ny	1944	TRUE
21	21	35.0	\$44k-\$99k	5.0	New York,Ny	1946	-1
22	22	19.0	\$55k-\$66k	7.8	New York,Ny	1988	TRUE
23	23	NaN	\$44k-\$99k	2.4	New York,Ny	1999	TRUE
24	24	13.0	\$44k-\$99k	-1.0	New York,Ny	1987	-1
25	25	55.0	\$44k-\$99k	0.0	Australia,Aus	1980	TRUE
26	26	NaN	\$55k-\$66k	NaN	India,In	1934	TRUE
27	27	52.0	\$44k-\$99k	5.4	India,In	1935	-1
28	28	NaN	\$39k-\$88k	3.4	Australia,Aus	1932	-1

	age	rating	established	easy_apply	starting_salary	ending_salary	country	initials
0	44	5.4	1999	True	44000	99000	India	In
1	66	3.5	2002	True	55000	66000	New York	Ny
2	39	4.0	Unknown	False	77000	89000	New York	Ny
3	64	4.4	1988	False	44000	99000	India	In
4	25	6.4	2002	False	44000	99000	Australia	Aus
5	44	1.4	1999	True	77000	89000	India	In
6	21	0.0	Unknown	False	44000	99000	New York	Ny
7	44	4.0	Unknown	False	44000	99000	Australia	Aus
8	35	5.4	Unknown	False	44000	99000	New York	Ny
9	22	7.7	Unknown	True	44000	99000	India	In
10	55	5.4	2008	True	10000	49000	India	In
11	44	6.7	2009	False	10000	49000	India	In
12	39	0.0	1999	False	44000	99000	India	In
13	25	4.0	2019	True	44000	99000	Australia	Aus
14	66	4.0	2020	True	44000	99000	Australia	Aus
15	44	3.0	1999	False	88000	101000	Australia	Aus
16	19	4.5	1984	False	19000	40000	India	In
17	39	5.3	1943	True	44000	99000	New York	Ny
18	35	6.7	1954	True	44000	99000	New York	Ny
19	32	3.3	1955	True	44000	99000	New York	Ny
20	39	5.7	1944	True	44000	99000	New York	Ny
21	35	5.0	1946	False	44000	99000	New York	Ny
22	19	7.8	1988	True	55000	66000	New York	Ny
23	39	2.4	1999	True	44000	99000	New York	Ny
24	13	4.0	1987	False	44000	99000	New York	Ny
25	55	0.0	1980	True	44000	99000	Australia	Aus
26	39	4.0	1934	True	55000	66000	India	In
27	52	5.4	1935	False	44000	99000	India	In
28	39	3.4	1932	False	39000	88000	Australia	Aus

Assignment 2 : Data Cleaning (chipotle.tsv):

Cleaned Version 1

Uncleaned Data

	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 Data

	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

Assignment 2 : Data Cleaning (chipotle.tsv):

Cleaned Version 2

Uncleaned Data

	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 Data

order_id	quantity	standardized_item_name	choice_1	choice_2	choice_3	choice_4	choice_5	choice_6	choice_7	choice_8	choice_9	choice_10	item_price
0	1	1	Chips and Fresh Tomato Salsa	None	None	None	None	None	None	None	None	None	2.75
1	1	1	Izze	Clementine	None	None	None	None	None	None	None	None	3.39
2	1	1	Nantucket Nectar	Apple	None	None	None	None	None	None	None	None	3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	None	None	None	None	None	None	None	None	None	2.39
4	2	2	Chicken Bowl	Tomatillo-Red Chili Salsa (Hot)	Black Beans	Rice	Cheese	Sour Cream	None	None	None	None	19.33
...
4617	1833	1	Steak Burrito	Fresh Tomato Salsa	Rice	Black Beans	Sour Cream	Cheese	Lettuce	Guacamole	None	None	9.99
4618	1833	1	Steak Burrito	Fresh Tomato Salsa	Rice	Sour Cream	Cheese	Lettuce	Guacamole	None	None	None	9.99
4619	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa	Fajita Vegetables	Pinto Beans	Guacamole	Lettuce	None	None	None	None	9.91
4620	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa	Fajita Vegetables	Lettuce	None	None	None	None	None	None	9.91
4621	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa	Fajita Vegetables	Pinto Beans	Lettuce	None	None	None	None	None	9.91