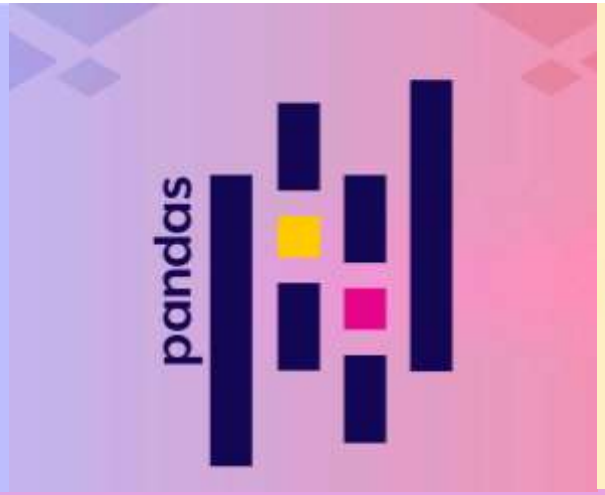
PrepInsta

Data cleaning using Python library

PANDAS

Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring and manipulating data.



PrepInsta

Data Cleaning using Pandas

Uncleaned 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

Cleaned Data

	age	rating	established	easy_apply	starting_salary	ending_salary	country	country_ISO_code
0	44	5.4	1999	TRUE	44000	99000	India	In
1	66	3.5	2002	TRUE	55000	66000	New York	Ny
2	39	1.0	1638	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	1638	FALSE	44000	99000	New York	Ny
7	44	1.0	1638	FALSE	44000	99000	Australia	Aus
8	35	5.4	1638	FALSE	44000	99000	New York	Ny
9	22	7.7	1638	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	1.0	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	1.0	1987	FALSE	44000	99000	New York	Ny
25	55	0.0	1980	TRUE	44000	99000	Australia	Aus
26	39	3.5	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

PrepInsta

Data Cleaning using chipotle.tsv

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	item_name	item_price	Choice_Description_1	Choice_Description_2	Choice_Description_3	Choice_Description_4	Choice_Description_5	Choice_Description_6	Choice_Description_7	Choice_Description_8	Choice_Description_9	Choice_Description_10
0	1	1	Chips and Fresh Tomato Salsa	3.29	None	None	None	None	None	None	None	None	None	None
1	1	1	Izze	3.39	Clementine	None	None	None	None	None	None	None	None	None
2	1	1	Nantucket Nectar	3.64	Apple	None	None	None	None	None	None	None	None	None
3	1	1	Chips and TomatilloGreen Chili Salsa	2.54	None	None	None	None	None	None	None	None	None	None
4	2	2	Chicken Bowl	20.27	TomatilloRed Chili Salsa Hot	Black Beans	Rice	Cheese	Sour Cream	None	None	None	None	None
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4617	1833	1	Steak Burrito	10.46	Fresh Tomato Salsa	Rice	Black Beans	Sour Cream	Cheese	Lettuce	Guacamole	None	None	None
4618	1833	1	Steak Burrito	10.46	Fresh Tomato Salsa	Rice	Sour Cream	Cheese	Lettuce	Guacamole	None	None	None	None
4619	1834	1	Chicken Salad Bowl	11.17	Fresh Tomato Salsa	Fajita Vegetables	Pinto Beans	Guacamole	Lettuce	None	None	None	None	None
4620	1834	1	Chicken Salad Bowl	11.17	Fresh Tomato Salsa	Fajita Vegetables	Lettuce	None	None	None	None	None	None	None
4621	1834	1	Chicken Salad Bowl	11.17	Fresh Tomato Salsa	Fajita Vegetables	Pinto Beans	Lettuce	None	None	None	None	None	None

PrepInsta

Top 5 items by Quantity and Price

	order_id	quantity	item_name	item_price	Choice_Description_1	Choice_Description_2	Choice_Description_3	Choice_Description_4	Choice_Description_5	Choice_Description_6	Choice_Description_7	Choice_Description_8	Choice_Description_9	Choice_Description_10
0	1	1	Chips and Fresh Tomato Salsa	3.29	None	None	None	None	None	None	None	None	None	None
1	1	1	Izze	3.39	Clementine	None	None	None	None	None	None	None	None	None
2	1	1	Nantucket Nectar	3.64	Apple	None	None	None	None	None	None	None	None	None
3	1	1	Chips and TomatilloGreen Chili Salsa	2.54	None	None	None	None	None	None	None	None	None	None
4	2	2	Chicken Bowl	20.27	TomatilloRed Chili Salsa Hot	Black Beans	Rice	Cheese	Sour Cream	None	None	None	None	None

Number of items sold

```
# Calculating number of different items sold
num_different_items = df['item_name'].nunique()

print("Number of Different Items Sold:", num_different_items)
```

```
Number of Different Items Sold: 50
```

Total number of orders

```
# Calculating total order count

total_orders = df['order_id'].nunique()

print("Total Order Count:", total_orders)
```

```
Total Order Count: 1827
```

Revenue

```
# Calculating revenue

df['total_price'] = df['quantity'] * df['item_price']
total_revenue = df['total_price'].sum()

print("Total Revenue:", total_revenue)
```

```
Total Revenue: 40975.32
```

PrepInsta

Average quantity and cost per unit of items

Average Quantity and Average Cost per Unit of Every Unique Item:

	item_name	quantity	item_price
0	6 Pack Soft Drink	1.018519	6.732407
1	Barbacoa Bowl	1.000000	10.220000
2	Barbacoa Burrito	1.000000	9.850000
3	Barbacoa Crispy Tacos	1.090909	11.923636
4	Barbacoa Salad Bowl	1.000000	10.950000
5	Barbacoa Soft Tacos	1.000000	10.020000
6	Bottled Water	1.287671	2.367466
7	Bowl	2.000000	29.600000
8	Burrito	1.000000	7.400000
9	Canned Soda	1.220000	1.622600
10	Canned Soft Drink	1.177936	1.731922
11	Carnitas Bowl	1.044118	11.308088
12	Carnitas Burrito	1.016949	10.301864
13	Carnitas Crispy Tacos	1.142857	12.730000
14	Carnitas Salad	1.000000	8.990000
15	Carnitas Salad Bowl	1.000000	11.060000
16	Carnitas Soft Tacos	1.000000	9.400000
17	Chicken Bowl	1.049365	10.630536
18	Chicken Burrito	1.070501	10.812059
19	Chicken Crispy Tacos	1.063830	10.690851
20	Chicken Salad	1.000000	9.010000
21	Chicken Salad Bowl	1.118182	12.490091
22	Chicken Soft Tacos	1.046729	10.111402
23	Chips	1.092683	2.567805
24	Chips and Fresh Tomato Salsa	1.181818	3.887091
25	Chips and Guacamole	1.057569	4.864776
26	Chips and Mild Fresh Tomato Salsa	1.000000	3.000000
27	Chips and Roasted Chili Corn Salsa	1.045455	3.220455
28	Chips and Roasted ChiliCorn Salsa	1.000000	2.390000
29	Chips and Tomatillo Green Chili Salsa	1.046512	3.233256
30	Chips and Tomatillo Red Chili Salsa	1.045455	3.220455
31	Chips and TomatilloGreen Chili Salsa	1.064516	2.704516
32	Chips and TomatilloRed Chili Salsa	1.277778	3.898889
33	Crispy Tacos	1.000000	7.400000
34	Izze	1.000000	3.390000
35	Nantucket Nectar	1.074074	3.909630
36	Salad	1.000000	7.400000
37	Side of Chips	1.089109	2.003960
38	Steak Bowl	1.047847	11.242967
39	Steak Burrito	1.049724	10.980110
40	Steak Crispy Tacos	1.028571	10.501714
41	Steak Salad	1.000000	8.910000
42	Steak Salad Bowl	1.068966	12.667241
43	Steak Soft Tacos	1.018182	9.927091
44	Veggie Bowl	1.023529	10.450235
45	Veggie Burrito	1.021053	10.047158
46	Veggie Crispy Tacos	1.000000	8.490000
47	Veggie Salad	1.000000	8.490000
48	Veggie Salad Bowl	1.000000	10.140000
49	Veggie Soft Tacos	1.142857	12.078571

Product Revenue

Revenue from Each Product:

item_name	
6 Pack Soft Drink	376.77
Barbacoa Bowl	654.08
Barbacoa Burrito	876.65
Barbacoa Crispy Tacos	153.02
Barbacoa Salad Bowl	87.60
Barbacoa Soft Tacos	250.50
Bottled Water	698.14
Bowl	148.00
Burrito	44.40
Canned Soda	228.76
Canned Soft Drink	680.94
Carnitas Bowl	855.62
Carnitas Burrito	628.08
Carnitas Crispy Tacos	111.38
Carnitas Salad	8.99
Carnitas Salad Bowl	66.36
Carnitas Soft Tacos	376.00
Chicken Bowl	8287.02
Chicken Burrito	6756.90
Chicken Crispy Tacos	562.74
Chicken Salad	81.09
Chicken Salad Bowl	1664.33
Chicken Soft Tacos	1178.52
Chips	620.40
Chips and Fresh Tomato Salsa	1176.63
Chips and Guacamole	2566.73
Chips and Mild Fresh Tomato Salsa	3.00
Chips and Roasted Chili Corn Salsa	77.02
Chips and Roasted ChiliCorn Salsa	43.02
Chips and Tomatillo Green Chili Salsa	151.37
Chips and Tomatillo Red Chili Salsa	154.04
Chips and TomatilloGreen Chili Salsa	94.02
Chips and TomatilloRed Chili Salsa	112.93
Crispy Tacos	14.80
Izze	61.02
Nantucket Nectar	120.12
Salad	14.80
Side of Chips	316.48
Steak Bowl	2585.74
Steak Burrito	4393.20
Steak Crispy Tacos	387.98
Steak Salad	35.64
Steak Salad Bowl	414.75
Steak Soft Tacos	565.48
Veggie Bowl	929.11
Veggie Burrito	1013.52
Veggie Crispy Tacos	8.49
Veggie Salad	50.94
Veggie Salad Bowl	182.52
Veggie Soft Tacos	105.68

Name: total_price, dtype: float64

PrepInsta

Total Cost and Quantity of Every Unique Item:

	item_name	quantity	total_cost
0	6 Pack Soft Drink	55	376.77
1	Barbacoa Bowl	64	654.08
2	Barbacoa Burrito	89	876.65
3	Barbacoa Crispy Tacos	12	153.02
4	Barbacoa Salad Bowl	8	87.60
5	Barbacoa Soft Tacos	25	250.50
6	Bottled Water	188	698.14
7	Bowl	4	148.00
8	Burrito	6	44.40
9	Canned Soda	122	228.76
10	Canned Soft Drink	331	680.94
11	Carnitas Bowl	71	855.62
12	Carnitas Burrito	60	628.08
13	Carnitas Crispy Tacos	8	111.38
14	Carnitas Salad	1	8.99
15	Carnitas Salad Bowl	6	66.36
16	Carnitas Soft Tacos	40	376.00
17	Chicken Bowl	744	8287.02
18	Chicken Burrito	577	6756.90
19	Chicken Crispy Tacos	50	562.74
20	Chicken Salad	9	81.09
21	Chicken Salad Bowl	123	1664.33
22	Chicken Soft Tacos	112	1178.52
23	Chips	224	620.40
24	Chips and Fresh Tomato Salsa	130	1176.63
25	Chips and Guacamole	496	2566.73
26	Chips and Mild Fresh Tomato Salsa	1	3.00
27	Chips and Roasted Chili Corn Salsa	23	77.02
28	Chips and Roasted ChiliCorn Salsa	18	43.02
29	Chips and Tomatillo Green Chili Salsa	45	151.37
30	Chips and Tomatillo Red Chili Salsa	46	154.04
31	Chips and TomatilloGreen Chili Salsa	33	94.02
32	Chips and TomatilloRed Chili Salsa	23	112.93
33	Crispy Tacos	2	14.80
34	Izze	18	61.02
35	Nantucket Nectar	29	120.12
36	Salad	2	14.80
37	Side of Chips	110	316.48
38	Steak Bowl	219	2585.74
39	Steak Burrito	380	4393.20
40	Steak Crispy Tacos	36	387.98
41	Steak Salad	4	35.64
42	Steak Salad Bowl	31	414.75
43	Steak Soft Tacos	56	565.48
44	Veggie Bowl	87	929.11
45	Veggie Burrito	97	1013.52
46	Veggie Crispy Tacos	1	8.49
47	Veggie Salad	6	50.94
48	Veggie Salad Bowl	18	182.52
49	Veggie Soft Tacos	8	105.68

Total cost and Quantity of every item