

# DATA SCIENCE LAB-2

K.SASI KIRAN

MCA(R)

2019202049

## 1. SORTING

Jupyter open csv file Last Checkpoint: 02/25/2021 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

In [9]: `res.sort_values(['Restaurant'], ascending=0)`

Out[9]:

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising
17	18	Vitality Bowls	San Ramon, Calif.	37	30.10%	77	24.20%	535	Yes
6	7	The Simple Greek	Blue Bell, Pa.	24	52.50%	36	33.30%	775	Yes
29	30	The Pizza Press	Anaheim, Calif.	27	23.70%	27	22.70%	1130	Yes
30	31	The Lost Cajun	Frisco, Texas	20	23.00%	26	4.00%	785	Yes
22	23	The Little Beet	New York, N.Y.	23	26.50%	12	33.30%	2230	No
41	42	The Human Bean	Medford, Ore.	47	19.00%	97	19.80%	535	Yes
39	40	The Flying Biscuit Cafe	Atlanta, Ga.	29	19.50%	21	16.70%	1510	Yes
12	13	Teriyaki Madness	Denver, Colo.	41	34.10%	63	65.80%	890	Yes
43	44	Tacos 4 Life	Conway, Ark.	25	18.40%	16	6.70%	1620	Yes
35	36	Surcheros	Douglas, Ga.	22	21.30%	20	17.60%	1230	Yes
47	48	Super Duper Burgers	San Francisco, Calif.	39	16.90%	16	14.30%	2630	No
25	26	Sugarfire Smokehouse	Olivette, Mo.	39	25.40%	15	15.40%	2820	No
48	49	StoneFire Grill	Agoura Hills, Calif.	31	16.00%	13	8.30%	2550	No
42	43	Spitz Mediterranean Street Food	Los Angeles, Calif.	28	18.80%	11	10.00%	2700	No
2	3	Slapfish	Huntington Beach, Calif.	21	81.00%	21	90.90%	1370	Yes
15	16	Rapid Fired Pizza	Kettering, Ohio	24	32.20%	35	29.60%	780	Yes
34	35	Quickway Japanese Hibachi	Fairfax, Va.	22	21.40%	25	19.00%	985	Yes
4	5	Pokeworks	Irvine, Calif.	49	77.10%	50	56.30%	1210	Yes
5	6	Playa Bowls	Belmar, N.J.	39	62.90%	76	28.80%	580	Yes
44	45	Pita Mediterranean Street Food	Fairburn, Ga.	38	17.80%	32	10.30%	1260	Yes
14	15	Original ChopShop	Plano, Texas	21	32.50%	12	20.00%	1930	No

In [10]: res.sort\_values(['location'],ascending=1)

Out[10]:

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising
48	49	StoneFire Grill	Agoura Hills, Calif.	31	16.00%	13	8.30%	2550	No
29	30	The Pizza Press	Anaheim, Calif.	27	23.70%	27	22.70%	1130	Yes
39	40	The Flying Biscuit Cafe	Atlanta, Ga.	29	19.50%	21	16.70%	1510	Yes
5	6	Playa Bowls	Belmar, N.J.	39	62.90%	76	28.80%	580	Yes
6	7	The Simple Greek	Blue Bell, Pa.	24	52.50%	36	33.30%	775	Yes
1	2	Clean Juice	Charlotte, N.C.	44	121.90%	105	94.40%	560	Yes
36	37	Famous Toastery	Charlotte, N.C.	40	21.20%	27	8.00%	1540	Yes
32	33	Jen's Splendid Ice Creams	Columbus, Ohio	42	22.40%	44	22.20%	1050	No
38	39	Condado Tacos	Columbus, Ohio	38	19.90%	15	15.40%	2755	No
26	27	Bibibop Asian Grill	Columbus, Ohio	44	24.60%	37	23.30%	1330	No
43	44	Tacos 4 Life	Conway, Ark.	25	18.40%	16	6.70%	1620	Yes
12	13	Teriyaki Madness	Denver, Colo.	41	34.10%	63	65.80%	890	Yes
20	21	Bulla Gastrobar	Doral, Fla.	32	28.20%	8	14.30%	4300	No
35	36	Surcheros	Douglas, Ga.	22	21.30%	20	17.60%	1230	Yes
44	45	Pita Mediterranean Street Food	Fairburn, Ga.	38	17.80%	32	10.30%	1260	Yes
34	35	Quickway Japanese Hibachi	Fairfax, Va.	22	21.40%	25	19.00%	985	Yes
30	31	The Lost Cajun	Frisco, Texas	20	23.00%	26	4.00%	785	Yes
2	3	Slapfish	Huntington Beach, Calif.	21	81.00%	21	90.90%	1370	Yes
4	5	Pokeworks	Invine, Calif.	49	77.10%	50	56.30%	1210	Yes
15	16	Rapid Fired Pizza	Kettering, Ohio	24	32.20%	35	29.60%	780	Yes
42	43	Spitz Mediterranean Street Food	Los Angeles, Calif.	28	18.80%	11	10.00%	2700	No

In [11]: res.sort\_values(['Restaurant','Location'],ascending=[1,0])

Out[11]:

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising
46	47	&pizza	Washington, D.C.	45	17.10%	35	9.40%	1350	No
26	27	Bibibop Asian Grill	Columbus, Ohio	44	24.60%	37	23.30%	1330	No
40	41	Blue Sushi Sake Grill	Omaha, Neb.	49	19.50%	14	16.70%	3500	No
13	14	Bluestone Lane	New York, N.Y.	48	33.00%	48	37.10%	1175	No
33	34	Boqueria	New York, N.Y.	27	22.00%	7	16.70%	4260	No
27	28	Bubbakoo's Burritos	Wall Township, N.J.	28	24.30%	31	14.80%	970	Yes
20	21	Bulla Gastrobar	Doral, Fla.	32	28.20%	8	14.30%	4300	No
24	25	By Chloe	New York, N.Y.	37	25.60%	14	7.70%	2800	No
3	4	Clean Eat	Wilmington, N.C.	25	79.70%	46	58.60%	685	Yes
1	2	Clean Juice	Charlotte, N.C.	44	121.90%	105	94.40%	560	Yes
38	39	Condado Tacos	Columbus, Ohio	38	19.90%	15	15.40%	2755	No
8	9	Creamistry	Yorba Linda, Calif.	24	36.80%	60	27.70%	465	Yes
37	38	Culinary Dropout	Scottsdale, Ariz.	20	20.80%	7	16.70%	3120	No
11	12	Dog Haus	Pasadena, Calif.	39	34.50%	50	42.90%	1200	Yes
28	29	Dos Toros Taqueria	New York, N.Y.	28	24.00%	22	10.00%	1375	No
21	22	Duck Donuts	Mechanicsburg, Pa.	44	28.00%	90	16.90%	530	Yes
10	11	Eggs Up Grill	Spartanburg, S.C.	30	35.40%	41	36.70%	860	Yes
0	1	Evergreens	Seattle, Wash.	24	130.50%	26	116.70%	1150	No
36	37	Famous Toastery	Charlotte, N.C.	40	21.20%	27	8.00%	1540	Yes
49	50	Gus's World Famous Fried Chicken	Memphis, Tenn.	44	14.40%	28	7.70%	1600	Yes
18	19	Hawkers Asian Street Fare	Orlando, Fla.	22	30.00%	7	40.00%	3800	No

## 2. INCLUDING COLUMN

Jupyter open csv file Last Checkpoint: 02/25/2021 (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python

17 18 Vitality Bowls San Ramon, Calif. 37 30.10% 77 24.20% 535 Yes

```
In [12]: res['Total']=res['Sales']+res['Units']+res['Unit_Volume']
```

```
In [13]: res['Total']=res['Sales']+res['Units']+res['Unit_Volume']
res.head(10)
```

Out[13]:

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising	Total
0	1	Evergreens	Seattle, Wash.	24	130.50%	26	116.70%	1150	No	1200
1	2	Clean Juice	Charlotte, N.C.	44	121.90%	105	94.40%	560	Yes	709
2	3	Slapfish	Huntington Beach, Calif.	21	81.00%	21	90.90%	1370	Yes	1412
3	4	Clean Eat	Wilmington, N.C.	25	79.70%	46	58.60%	685	Yes	756
4	5	Pokeworks	Irvine, Calif.	49	77.10%	50	56.30%	1210	Yes	1309
5	6	Playa Bowls	Belmar, N.J.	39	62.90%	76	28.80%	580	Yes	695
6	7	The Simple Greek	Blue Bell, Pa.	24	52.50%	36	33.30%	775	Yes	835
7	8	Melt Shop	New York, N.Y.	20	39.60%	19	35.70%	1260	Yes	1299
8	9	Creamistry	Yorba Linda, Calif.	24	36.80%	60	27.70%	465	Yes	549
9	10	Joella's Hot Chicken	Louisville, Ky.	29	35.50%	17	30.80%	1930	No	1976

In [ ]:

Jupyter open csv file Last Checkpoint: 02/25/2021 (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python

9 10 Joella's Hot Chicken Louisville, Ky. 29 35.50% 17 30.80% 1930 No NaN

```
In [17]: res['Total']=res.iloc[:,4:10].sum(axis=1)
cols=list(res.columns)
res=res[cols[0:4]+[cols[-1]]+cols[4:12]]
res.head(10)
```

Out[17]:

	Rank	Restaurant	Location	Sales	Total	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising	Total
0	1	Evergreens	Seattle, Wash.	24	1176	130.50%	26	116.70%	1150	No	1176
1	2	Clean Juice	Charlotte, N.C.	44	665	121.90%	105	94.40%	560	Yes	665
2	3	Slapfish	Huntington Beach, Calif.	21	1391	81.00%	21	90.90%	1370	Yes	1391
3	4	Clean Eat	Wilmington, N.C.	25	731	79.70%	46	58.60%	685	Yes	731
4	5	Pokeworks	Irvine, Calif.	49	1260	77.10%	50	56.30%	1210	Yes	1260
5	6	Playa Bowls	Belmar, N.J.	39	656	62.90%	76	28.80%	580	Yes	656
6	7	The Simple Greek	Blue Bell, Pa.	24	811	52.50%	36	33.30%	775	Yes	811
7	8	Melt Shop	New York, N.Y.	20	1279	39.60%	19	35.70%	1260	Yes	1279
8	9	Creamistry	Yorba Linda, Calif.	24	525	36.80%	60	27.70%	465	Yes	525
9	10	Joella's Hot Chicken	Louisville, Ky.	29	1947	35.50%	17	30.80%	1930	No	1947

In [ ]:

### 3. REGULAR EXPRESSION

```
jupyter open csv file Last Checkpoint: 02/25/2021 (autosaved)
```

```
File Edit View Insert Cell Kernel Widgets Help
```

```
Not Trusted
```

```
Code
```

```
28 10.00% 1375 No
31 26.70% 1240 No
33 16.70% 4260 No
```

```
In [30]: import re
res.loc[res['Location'].str.contains('Columbus, Ohio|New York, N.Y.', regex=True)]
```

```
Out[30]:
```

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising
7	8	Melt Shop	New York, N.Y.	20	39.60%	19	35.70%	1260	Yes
13	14	Bluestone Lane	New York, N.Y.	48	33.00%	48	37.10%	1175	No
22	23	The Little Beet	New York, N.Y.	23	26.50%	12	33.30%	2230	No
23	24	Joe & The Juice	New York, N.Y.	47	25.90%	69	25.50%	760	Yes
24	25	By Chloe	New York, N.Y.	37	25.60%	14	7.70%	2800	No
26	27	Bibibop Asian Grill	Columbus, Ohio	44	24.60%	37	23.30%	1330	No
28	29	Dos Toros Taqueria	New York, N.Y.	28	24.00%	22	10.00%	1375	No
31	32	Just Salad	New York, N.Y.	42	22.70%	38	26.70%	1240	No
32	33	Jeni's Splendid Ice Creams	Columbus, Ohio	42	22.40%	44	22.20%	1050	No
33	34	Boqueria	New York, N.Y.	27	22.00%	7	16.70%	4260	No
38	39	Condado Tacos	Columbus, Ohio	38	19.90%	15	15.40%	2755	No

```
In [31]: res = res[res['Restaurant'].str.contains('^t[a-z]*t', regex = True)]
```

```
In [32]: res
```

```
Out[32]:
```

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising
--	------	------------	----------	-------	-----------	-------	-----------	-------------	-------------



jupyter open csv file Last Checkpoint: 02/25/2021 (unsaved changes) Python 3

File Edit View Insert Cell Kernel Widgets Help

Not Trusted

In [38]: `import re  
res.loc[res['Location'].str.contains('Columbus, Ohio|New York, N.Y.', regex=True)]`

Out[38]:

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising
7	8	Melt Shop	New York, N.Y.	20	39.60%	19	35.70%	1260	Yes
13	14	Bluestone Lane	New York, N.Y.	48	33.00%	48	37.10%	1175	No
22	23	The Little Beet	New York, N.Y.	23	26.50%	12	33.30%	2230	No
23	24	Joe & The Juice	New York, N.Y.	47	25.90%	69	25.50%	760	Yes
24	25	By Chloe	New York, N.Y.	37	25.60%	14	7.70%	2800	No
26	27	Bibibop Asian Grill	Columbus, Ohio	44	24.60%	37	23.30%	1330	No
28	29	Dos Toros Taqueria	New York, N.Y.	28	24.00%	22	10.00%	1375	No
31	32	Just Salad	New York, N.Y.	42	22.70%	38	26.70%	1240	No
32	33	Jeni's Splendid Ice Creams	Columbus, Ohio	42	22.40%	44	22.20%	1050	No
33	34	Boqueria	New York, N.Y.	27	22.00%	7	16.70%	4260	No
38	39	Condado Tacos	Columbus, Ohio	38	19.90%	15	15.40%	2755	No

In [ ]:

jupyter open csv file Last Checkpoint: 02/25/2021 (autosaved) Python 3

File Edit View Insert Cell Kernel Widgets Help

Not Trusted


In [39]: `import re  
res.loc[res['Restaurant'].str.contains('^The[a-z]*', flags=re.I, regex=True)]`


Out[39]:











	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising
6	7	The Simple Greek	Blue Bell, Pa.	24	52.50%	36	33.30%	775	Yes
22	23	The Little Beet	New York, N.Y.	23	26.50%	12	33.30%	2230	No
29	30	The Pizza Press	Anaheim, Calif.	27	23.70%	27	22.70%	1130	Yes
30	31	The Lost Cajun	Frisco, Texas	20	23.00%	26	4.00%	785	Yes
39	40	The Flying Biscuit Cafe	Atlanta, Ga.	29	19.50%	21	16.70%	1510	Yes
41	42	The Human Bean	Medford, Ore.	47	19.00%	97	19.80%	535	Yes

In [ ]:

## 4. DROPPING

jupyter open csv file Last Checkpoint: 02/25/2021 (unsaved changes)  [Log out](#)

File Edit View Insert Cell Kernel Widgets Help Not Trusted  Python 3

         Code 

Out[17]:

	Rank	Restaurant	Location	Sales	Total	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising	Total
0	1	Evergreens	Seattle, Wash.	24	1176	130.50%	26	116.70%	1150	No	1176
1	2	Clean Juice	Charlotte, N.C.	44	665	121.90%	105	94.40%	560	Yes	665
2	3	Slapfish	Huntington Beach, Calif.	21	1391	81.00%	21	90.90%	1370	Yes	1391
3	4	Clean Eat	Wilmington, N.C.	25	731	79.70%	46	58.60%	685	Yes	731
4	5	Pokeworks	Irvine, Calif.	49	1260	77.10%	50	56.30%	1210	Yes	1260
5	6	Playa Bowls	Belmar, N.J.	39	656	62.90%	76	28.80%	580	Yes	656
6	7	The Simple Greek	Blue Bell, Pa.	24	811	52.50%	36	33.30%	775	Yes	811
7	8	Melt Shop	New York, N.Y.	20	1279	39.60%	19	35.70%	1260	Yes	1279
8	9	Creamistry	Yorba Linda, Calif.	24	525	36.80%	60	27.70%	465	Yes	525
9	10	Joella's Hot Chicken	Louisville, Ky.	29	1947	35.50%	17	30.80%	1930	No	1947

In [18]: `res=res.drop(columns=['Total'])`  
`res.head(10)`

Out[18]:

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchising
0	1	Evergreens	Seattle, Wash.	24	130.50%	26	116.70%	1150	No
1	2	Clean Juice	Charlotte, N.C.	44	121.90%	105	94.40%	560	Yes
2	3	Slapfish	Huntington Beach, Calif.	21	81.00%	21	90.90%	1370	Yes
3	4	Clean Eat	Wilmington, N.C.	25	79.70%	46	58.60%	685	Yes
4	5	Pokeworks	Irvine, Calif.	49	77.10%	50	56.30%	1210	Yes
5	6	Playa Bowls	Belmar, N.J.	39	62.90%	76	28.80%	580	Yes
6	7	The Simple Greek	Blue Bell, Pa.	24	52.50%	36	33.30%	775	Yes
7	8	Melt Shop	New York, N.Y.	20	39.60%	19	35.70%	1260	Yes
8	9	Creamistry	Yorba Linda, Calif.	24	36.80%	60	27.70%	465	Yes

## 5. STORE IN NEW FILE

```
jupyter open csv file Last Checkpoint: 02/25/2021 (unsaved changes)
File Edit View Insert Cell Kernel Widgets Help
Not Trusted Python 3

In [27]: res.loc[(res['Units']==105)]
res.to_excel('future_new.xlsx', index=False)

In [28]: new_res=res.loc[(res['Location']=='New York, N.Y.') & (res['Franchising']=='No')]
new_res.to_excel('new.xlsx', index=False)

In [29]: print(new_res)

   Rank  Restaurant  Location  Sales  YOY_Sales  Units  \
13   14  Bluestone Lane  New York, N.Y.    48    33.00%    48
22   23   The Little Beet  New York, N.Y.    23    26.50%    12
24   25         By Chloe  New York, N.Y.    37    25.60%    14
28   29  Dos Toros Taqueria  New York, N.Y.    28    24.00%    22
31   32        Just Salad  New York, N.Y.    42    22.70%    38
33   34        Boqueria  New York, N.Y.    27    22.00%     7

   YOY_Units  Unit_Volume  Franchising
13    37.10%         1175           No
22    33.30%         2230           No
24     7.70%         2800           No
28    10.00%         1375           No
31    26.70%         1240           No
33    16.70%         4260           No
```

[illegible]

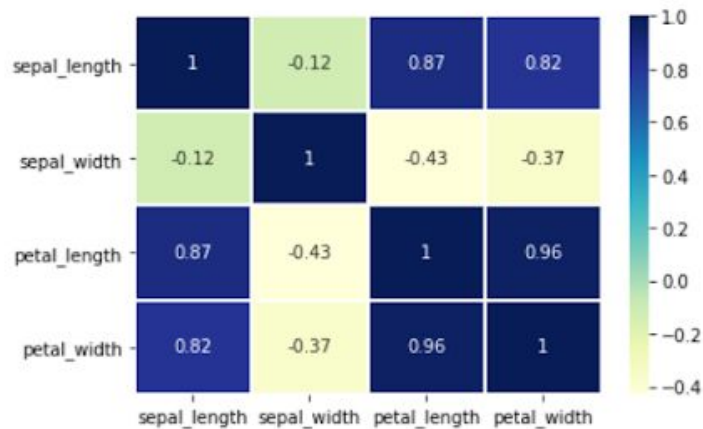
## 6. VISUALIZATION OF DATA

```
In [51]: g=sns.pairplot(data,hue="class")
```

```
-----  
NameError                                Traceback (most recent call last)  
<ipython-input-51-2d103d94d620> in <module>  
----> 1 g=sns.pairplot(data,hue="class")  
  
NameError: name 'data' is not defined
```

```
In [52]: sn.heatmap(iris.corr(),cmap="YlGnBu",linecolor='white',linewidths=1,annot=True)
```

```
Out[52]: <AxesSubplot:>
```



```
In [73]: import seaborn as sns  
sns.set()
```

```
In [74]: plt.plot(x, y)  
plt.legend('ABCDEF', ncol=2, loc='upper left');
```



```
In [75]: data = np.random.multivariate_normal([0, 0], [[5, 2], [2, 2]], size=2000)  
data = pd.DataFrame(data, columns=['x', 'y'])  
  
for col in 'xy':  
    plt.hist(data[col], normed=True, alpha=0.5)
```

```
-----  
AttributeError                            Traceback (most recent call last)  
<ipython-input-75-7c25cafad82f> in <module>  
3
```



