In [ ]:

```
                              WORKING WITH MAP REDUCE
```

In [ ]:

```
                              K.SASI KIRAN
                              MCA(R)
                              2019202049
```

In [1]:

```python
import pandas as pd
import seaborn as sn
rest=pd.read_csv("Future50.csv")
rest.head()
```

Out[1]:

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Franchisir
0	1	Evergreens	Seattle, Wash.	24	130.50%	26	116.70%	1150	N
1	2	Clean Juice	Charlotte, N.C.	44	121.90%	105	94.40%	560	Ye
2	3	Slapfish	Huntington Beach, Calif.	21	81.00%	21	90.90%	1370	Ye
3	4	Clean Eatz	Wilmington, N.C.	25	79.70%	46	58.60%	685	Ye
4	5	Pokeworks	Irvine, Calif.	49	77.10%	50	56.30%	1210	Ye

In [2]:

```python
rest.columns
```

Out[2]:

```
Index(['Rank', 'Restaurant', 'Location', 'Sales', 'YOY_Sales', 'Units',
       'YOY_Units', 'Unit_Volume', 'Franchising'],
      dtype='object')
```

In [3]:

```python
rest['Location'].value_counts()
```

Out[3]:

```
New York, N.Y.            8
Columbus, Ohio            3
San Francisco, Calif.     2
Charlotte, N.C.           2
Pasadena, Calif.          1
Fairburn, Ga.             1
Frisco, Texas             1
Agoura Hills, Calif.      1
Yorba Linda,  Calif.      1
Doral, Fla.               1
Denver, Colo.             1
McAllen, Texas            1
Memphis, Tenn.            1
Scottsdale, Ariz.         1
Anaheim, Calif.           1
Irvine, Calif.            1
Fairfax, Va.              1
San Ramon, Calif.         1
Omaha, Neb.               1
Conway, Ark.              1
Washington, D.C.          1
Wilmington, N.C.          1
Olivette, Mo.             1
Plano, Texas              1
Blue Bell, Pa.            1
Orlando, Fla.             1
Douglas, Ga.              1
Wall Township, N.J.       1
Belmar,  N.J.             1
Atlanta, Ga.              1
Orange Park, Fla.         1
Medford, Ore.             1
Spartanburg, S.C.         1
Los Angeles, Calif.       1
Louisville, Ky.           1
Seattle, Wash.            1
Kettering, Ohio           1
Mechanicsburg, Pa.        1
Huntington Beach, Calif.  1
Name: Location, dtype: int64
```

In [4]:

```python
rest['Location'].unique()
```

Out[4]:

```
array(['Seattle, Wash.', 'Charlotte, N.C.', 'Huntington Beach, Calif.',
       'Wilmington, N.C.', 'Irvine, Calif.', 'Belmar,  N.J.',
       'Blue Bell, Pa.', 'New York, N.Y.', 'Yorba Linda,  Calif.',
       'Louisville, Ky.', 'Spartanburg, S.C.', 'Pasadena, Calif.',
       'Denver, Colo.', 'Plano, Texas', 'Kettering, Ohio',
       'San Francisco, Calif.', 'San Ramon, Calif.', 'Orlando, Fla.',
       'Orange Park, Fla.', 'Doral, Fla.', 'Mechanicsburg, Pa.',
       'Olivette, Mo.', 'Columbus, Ohio', 'Wall Township, N.J.',
       'Anaheim, Calif.', 'Frisco, Texas', 'Fairfax, Va.', 'Douglas, Ga.',
       'Scottsdale, Ariz.', 'Atlanta, Ga.', 'Omaha, Neb.',
       'Medford, Ore.', 'Los Angeles, Calif.', 'Conway, Ark.',
       'Fairburn, Ga.', 'McAllen, Texas', 'Washington, D.C.',
       'Agoura Hills, Calif.', 'Memphis, Tenn.'], dtype=object)
```

In [5]:

```python
rest['Units'].value_counts(normalize=True)
```

Out[5]:

```
7      0.06
26     0.04
14     0.04
27     0.04
16     0.04
25     0.04
35     0.04
21     0.04
15     0.04
50     0.04
12     0.04
44     0.02
46     0.02
32     0.02
105    0.02
38     0.02
37     0.02
36     0.02
48     0.02
97     0.02
60     0.02
41     0.02
33     0.02
13     0.02
31     0.02
28     0.02
22     0.02
20     0.02
19     0.02
17     0.02
90     0.02
77     0.02
76     0.02
11     0.02
8      0.02
71     0.02
69     0.02
63     0.02
Name: Units, dtype: float64
```

In [6]:

```python
rest.groupby(by='Location').size()
```

Out[6]:

```
Location
Agoura Hills, Calif.       1
Anaheim, Calif.            1
Atlanta, Ga.               1
Belmar,  N.J.              1
Blue Bell, Pa.             1
Charlotte, N.C.            2
Columbus, Ohio             3
Conway, Ark.               1
Denver, Colo.              1
Doral, Fla.                1
Douglas, Ga.               1
Fairburn, Ga.              1
Fairfax, Va.               1
Frisco, Texas              1
Huntington Beach, Calif.   1
Irvine, Calif.             1
Kettering, Ohio            1
Los Angeles, Calif.        1
Louisville, Ky.            1
McAllen, Texas             1
Mechanicsburg, Pa.         1
Medford, Ore.              1
Memphis, Tenn.             1
New York, N.Y.             8
Olivette, Mo.              1
Omaha, Neb.                1
Orange Park, Fla.          1
Orlando, Fla.              1
Pasadena, Calif.           1
Plano, Texas               1
San Francisco, Calif.      2
San Ramon, Calif.          1
Scottsdale, Ariz.          1
Seattle, Wash.             1
Spartanburg, S.C.          1
Wall Township, N.J.        1
Washington, D.C.           1
Wilmington, N.C.           1
Yorba Linda,  Calif.       1
dtype: int64
```

In [7]:

```python
rest.groupby('Location')['Units'].mean()
```

Out[7]:

```
Location
Agoura Hills, Calif.       13.000
Anaheim, Calif.            27.000
Atlanta, Ga.               21.000
Belmar,  N.J.              76.000
Blue Bell, Pa.             36.000
Charlotte, N.C.            66.000
Columbus, Ohio             32.000
Conway, Ark.               16.000
Denver, Colo.              63.000
Doral, Fla.                 8.000
Douglas, Ga.               20.000
Fairburn, Ga.              32.000
Fairfax, Va.               25.000
Frisco, Texas              26.000
Huntington Beach, Calif.   21.000
Irvine, Calif.             50.000
Kettering, Ohio            35.000
Los Angeles, Calif.        11.000
Louisville, Ky.            17.000
McAllen, Texas             25.000
Mechanicsburg, Pa.         90.000
Medford, Ore.              97.000
Memphis, Tenn.             28.000
New York, N.Y.             28.625
Olivette, Mo.              15.000
Omaha, Neb.                14.000
Orange Park, Fla.          33.000
Orlando, Fla.               7.000
Pasadena, Calif.           50.000
Plano, Texas               12.000
San Francisco, Calif.      43.500
San Ramon, Calif.          77.000
Scottsdale, Ariz.           7.000
Seattle, Wash.             26.000
Spartanburg, S.C.          41.000
Wall Township, N.J.        31.000
Washington, D.C.           35.000
Wilmington, N.C.           46.000
Yorba Linda,  Calif.       60.000
Name: Units, dtype: float64
```

3/12/2021   map_reduce - Jupyter Notebook

In [8]:

```python
rest.groupby('Location')['Units'].describe()
```

Out[8]:

Location	count	mean	std	min	25%	50%	75%	max
Agoura Hills, Calif.	1.0	13.000	NaN	13.0	13.00	13.0	13.00	13.0
Anaheim, Calif.	1.0	27.000	NaN	27.0	27.00	27.0	27.00	27.0
Atlanta, Ga.	1.0	21.000	NaN	21.0	21.00	21.0	21.00	21.0
Belmar, N.J.	1.0	76.000	NaN	76.0	76.00	76.0	76.00	76.0
Blue Bell, Pa.	1.0	36.000	NaN	36.0	36.00	36.0	36.00	36.0
Charlotte, N.C.	2.0	66.000	55.154329	27.0	46.50	66.0	85.50	105.0
Columbus, Ohio	3.0	32.000	15.132746	15.0	26.00	37.0	40.50	44.0
Conway, Ark.	1.0	16.000	NaN	16.0	16.00	16.0	16.00	16.0
Denver, Colo.	1.0	63.000	NaN	63.0	63.00	63.0	63.00	63.0
Doral, Fla.	1.0	8.000	NaN	8.0	8.00	8.0	8.00	8.0
Douglas, Ga.	1.0	20.000	NaN	20.0	20.00	20.0	20.00	20.0
Fairburn, Ga.	1.0	32.000	NaN	32.0	32.00	32.0	32.00	32.0
Fairfax, Va.	1.0	25.000	NaN	25.0	25.00	25.0	25.00	25.0
Frisco, Texas	1.0	26.000	NaN	26.0	26.00	26.0	26.00	26.0
Huntington Beach, Calif.	1.0	21.000	NaN	21.0	21.00	21.0	21.00	21.0
Irvine, Calif.	1.0	50.000	NaN	50.0	50.00	50.0	50.00	50.0
Kettering, Ohio	1.0	35.000	NaN	35.0	35.00	35.0	35.00	35.0
Los Angeles, Calif.	1.0	11.000	NaN	11.0	11.00	11.0	11.00	11.0
Louisville, Ky.	1.0	17.000	NaN	17.0	17.00	17.0	17.00	17.0
McAllen, Texas	1.0	25.000	NaN	25.0	25.00	25.0	25.00	25.0
Mechanicsburg, Pa.	1.0	90.000	NaN	90.0	90.00	90.0	90.00	90.0
Medford, Ore.	1.0	97.000	NaN	97.0	97.00	97.0	97.00	97.0
Memphis, Tenn.	1.0	28.000	NaN	28.0	28.00	28.0	28.00	28.0
New York, N.Y.	8.0	28.625	21.340354	7.0	13.50	20.5	40.50	69.0
Olivette, Mo.	1.0	15.000	NaN	15.0	15.00	15.0	15.00	15.0
Omaha, Neb.	1.0	14.000	NaN	14.0	14.00	14.0	14.00	14.0
Orange Park, Fla.	1.0	33.000	NaN	33.0	33.00	33.0	33.00	33.0
Orlando, Fla.	1.0	7.000	NaN	7.0	7.00	7.0	7.00	7.0
Pasadena, Calif.	1.0	50.000	NaN	50.0	50.00	50.0	50.00	50.0
Plano, Texas	1.0	12.000	NaN	12.0	12.00	12.0	12.00	12.0
San Francisco, Calif.	2.0	43.500	38.890873	16.0	29.75	43.5	57.25	71.0
San Ramon, Calif.	1.0	77.000	NaN	77.0	77.00	77.0	77.00	77.0
Scottsdale, Ariz.	1.0	7.000	NaN	7.0	7.00	7.0	7.00	7.0

localhost:8888/notebooks/map_reduce.ipynb#   7/13

Location	count	mean	std	min	25%	50%	75%	max
Seattle, Wash.	1.0	26.000	NaN	26.0	26.00	26.0	26.00	26.0
Spartanburg, S.C.	1.0	41.000	NaN	41.0	41.00	41.0	41.00	41.0
Wall Township, N.J.	1.0	31.000	NaN	31.0	31.00	31.0	31.00	31.0
Washington, D.C.	1.0	35.000	NaN	35.0	35.00	35.0	35.00	35.0
Wilmington, N.C.	1.0	46.000	NaN	46.0	46.00	46.0	46.00	46.0
Yorba Linda, Calif.	1.0	60.000	NaN	60.0	60.00	60.0	60.00	60.0

In [10]:

```python
rest['unit_bin']=pd.cut(x=rest['Units'],bins=[20,40,50,80])
```

In [11]:

```
rest
```

Out[11]:

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Fr
0	1	Evergreens	Seattle, Wash.	24	130.50%	26	116.70%	1150	
1	2	Clean Juice	Charlotte, N.C.	44	121.90%	105	94.40%	560	
2	3	Slapfish	Huntington Beach, Calif.	21	81.00%	21	90.90%	1370	
3	4	Clean Eatz	Wilmington, N.C.	25	79.70%	46	58.60%	685	
4	5	Pokeworks	Irvine, Calif.	49	77.10%	50	56.30%	1210	
5	6	Playa Bowls	Belmar, N.J.	39	62.90%	76	28.80%	580	
6	7	The Simple Greek	Blue Bell, Pa.	24	52.50%	36	33.30%	775	
7	8	Melt Shop	New York, N.Y.	20	39.60%	19	35.70%	1260	
8	9	Creamistry	Yorba Linda, Calif.	24	36.80%	60	27.70%	465	
9	10	Joella's Hot Chicken	Louisville, Ky.	29	35.50%	17	30.80%	1930	
10	11	Eggs Up Grill	Spartanburg, S.C.	30	35.40%	41	36.70%	860	
11	12	Dog Haus	Pasadena, Calif.	39	34.50%	50	42.90%	1200	
12	13	Teriyaki Madness	Denver, Colo.	41	34.10%	63	65.80%	890	
13	14	Bluestone Lane	New York, N.Y.	48	33.00%	48	37.10%	1175	
14	15	Original ChopShop	Plano, Texas	21	32.50%	12	20.00%	1930	
15	16	Rapid Fired Pizza	Kettering, Ohio	24	32.20%	35	29.60%	780	
16	17	Ike's Love & Sandwiches	San Francisco, Calif.	44	30.80%	71	29.10%	700	
17	18	Vitality Bowls	San Ramon, Calif.	37	30.10%	77	24.20%	535	
18	19	Hawkers Asian Street Fare	Orlando, Fla.	22	30.00%	7	40.00%	3800	
19	20	Maple Street Biscuit Co.	Orange Park, Fla.	39	28.90%	33	10.00%	1260	
20	21	Bulla Gastrobar	Doral, Fla.	32	28.20%	8	14.30%	4300	
21	22	Duck Donuts	Mechanicsburg, Pa.	44	28.00%	90	16.90%	530	
22	23	The Little Beet	New York, N.Y.	23	26.50%	12	33.30%	2230	

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Fr
23	24	Joe & The Juice	New York, N.Y.	47	25.90%	69	25.50%	760	
24	25	By Chloe	New York, N.Y.	37	25.60%	14	7.70%	2800	
25	26	Sugarfire Smokehouse	Olivette, Mo.	39	25.40%	15	15.40%	2820	
26	27	Bibibop Asian Grill	Columbus, Ohio	44	24.60%	37	23.30%	1330	
27	28	Bubbakoo's Burritos	Wall Township, N.J.	28	24.30%	31	14.80%	970	
28	29	Dos Toros Taqueria	New York, N.Y.	28	24.00%	22	10.00%	1375	
29	30	The Pizza Press	Anaheim, Calif.	27	23.70%	27	22.70%	1130	
30	31	The Lost Cajun	Frisco, Texas	20	23.00%	26	4.00%	785	
31	32	Just Salad	New York, N.Y.	42	22.70%	38	26.70%	1240	
32	33	Jeni's Splendid Ice Creams	Columbus, Ohio	42	22.40%	44	22.20%	1050	
33	34	Boqueria	New York, N.Y.	27	22.00%	7	16.70%	4260	
34	35	Quickway Japanese Hibachi	Fairfax, Va.	22	21.40%	25	19.00%	985	
35	36	Surcheros	Douglas, Ga.	22	21.30%	20	17.60%	1230	
36	37	Famous Toastery	Charlotte, N.C.	40	21.20%	27	8.00%	1540	
37	38	Culinary Dropout	Scottsdale, Ariz.	20	20.80%	7	16.70%	3120	
38	39	Condado Tacos	Columbus, Ohio	38	19.90%	15	15.40%	2755	
39	40	The Flying Biscuit Cafe	Atlanta, Ga.	29	19.50%	21	16.70%	1510	
40	41	Blue Sushi Sake Grill	Omaha, Neb.	49	19.50%	14	16.70%	3500	
41	42	The Human Bean	Medford, Ore.	47	19.00%	97	19.80%	535	
42	43	Spitz Mediterranean Street Food	Los Angeles, Calif.	28	18.80%	11	10.00%	2700	
43	44	Tacos 4 Life	Conway, Ark.	25	18.40%	16	6.70%	1620	
44	45	Pita Mediterranean Street Food	Fairburn, Ga.	38	17.80%	32	10.30%	1260	
45	46	LA Crawfish	McAllen, Texas	48	17.60%	25	13.60%	2050	
46	47	&pizza	Washington, D.C.	45	17.10%	35	9.40%	1350	
47	48	Super Duper Burgers	San Francisco, Calif.	39	16.90%	16	14.30%	2630	

	Rank	Restaurant	Location	Sales	YOY_Sales	Units	YOY_Units	Unit_Volume	Fr:
48	49	StoneFire Grill	Agoura Hills, Calif.	31	16.00%	13	8.30%	2550	
49	50	Gus's World Famous Fried Chicken	Memphis, Tenn.	44	14.40%	28	7.70%	1600	

In [12]:

```
rest['unit_bin'].unique()
```

Out[12]:

```
[(20.0, 40.0], NaN, (40.0, 50.0], (50.0, 80.0]]
Categories (3, interval[int64]): [(20, 40] < (40, 50] < (50, 80]]
```

In [15]:

```
rest.groupby(by='Location').agg({'Units':['count','mean','sum']})
```

Out[15]:

Location	Units count	mean	sum
Agoura Hills, Calif.	1	13.000	13
Anaheim, Calif.	1	27.000	27
Atlanta, Ga.	1	21.000	21
Belmar, N.J.	1	76.000	76
Blue Bell, Pa.	1	36.000	36
Charlotte, N.C.	2	66.000	132
Columbus, Ohio	3	32.000	96
Conway, Ark.	1	16.000	16
Denver, Colo.	1	63.000	63
Doral, Fla.	1	8.000	8
Douglas, Ga.	1	20.000	20
Fairburn, Ga.	1	32.000	32
Fairfax, Va.	1	25.000	25
Frisco, Texas	1	26.000	26
Huntington Beach, Calif.	1	21.000	21
Irvine, Calif.	1	50.000	50
Kettering, Ohio	1	35.000	35
Los Angeles, Calif.	1	11.000	11
Louisville, Ky.	1	17.000	17
McAllen, Texas	1	25.000	25
Mechanicsburg, Pa.	1	90.000	90
Medford, Ore.	1	97.000	97
Memphis, Tenn.	1	28.000	28
New York, N.Y.	8	28.625	229
Olivette, Mo.	1	15.000	15
Omaha, Neb.	1	14.000	14
Orange Park, Fla.	1	33.000	33
Orlando, Fla.	1	7.000	7
Pasadena, Calif.	1	50.000	50
Plano, Texas	1	12.000	12
San Francisco, Calif.	2	43.500	87
San Ramon, Calif.	1	77.000	77

	Units		
	count	**mean**	**sum**
Location			
Scottsdale, Ariz.	1	7.000	7
Seattle, Wash.	1	26.000	26
Spartanburg, S.C.	1	41.000	41
Wall Township, N.J.	1	31.000	31
Washington, D.C.	1	35.000	35
Wilmington, N.C.	1	46.000	46
Yorba Linda, Calif.	1	60.000	60

In [20]:

```python
rest.query("Location=='New York, N.Y.'")['Unit_Volume'].mean()
```

Out[20]:

1887.5

In [ ]: