

Day 16

DIY

Q1. Problem Statement: Univariate EDA

Load the “*cereal.csv*” data into a DataFrame and perform the following tasks:

1. Explore the DataFrame using `info()` and `describe()` functions
2. Replace the manufacturer names with the dictionary of names given here –
{‘A’:‘American Home Food Products’, ‘G’:‘General Mills’,
‘K’:‘Kelloggs’, ‘N’:‘Nabisco’, ‘P’:‘Post’, ‘Q’:‘Quaker Oats’, ‘R’:‘Ralston Purina’}
3. Find out the top five cereal manufacturers based on customer ratings
4. With the help of a barplot, compare the manufacturers based on customer ratings
5. With the help of a pie chart, find out which of the manufacturers are the largest producers of cereals

Dataset:

	name	mfr	type	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	weight	cups	rating
0	100% Bran	N	C	70	4	1	130	10.0	5.0	6	280	25	3	1.0	0.33	68.402973
1	100% Natural Bran	Q	C	120	3	5	15	2.0	8.0	8	135	0	3	1.0	1.00	33.983679
2	All-Bran	K	C	70	4	1	260	9.0	7.0	5	320	25	3	1.0	0.33	59.425505
3	All-Bran with Extra Fiber	K	C	50	4	0	140	14.0	8.0	0	330	25	3	1.0	0.50	93.704912
4	Almond Delight	R	C	110	2	2	200	1.0	14.0	8	-1	25	3	1.0	0.75	34.384843

Sample Output:

1. Explore the DataFrame using `info()` and `describe()` functions

	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	weight
count	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000
mean	106.883117	2.545455	1.012987	159.675325	2.151948	14.597403	6.922078	96.077922	28.246753	2.207792	1.029610
std	19.484119	1.094790	1.006473	83.832295	2.383364	4.278956	4.444885	71.286813	22.342523	0.832524	0.150477
min	50.000000	1.000000	0.000000	0.000000	0.000000	-1.000000	-1.000000	-1.000000	0.000000	1.000000	0.500000
25%	100.000000	2.000000	0.000000	130.000000	1.000000	12.000000	3.000000	40.000000	25.000000	1.000000	1.000000
50%	110.000000	3.000000	1.000000	180.000000	2.000000	14.000000	7.000000	90.000000	25.000000	2.000000	1.000000
75%	110.000000	3.000000	2.000000	210.000000	3.000000	17.000000	11.000000	120.000000	25.000000	3.000000	1.000000
max	160.000000	6.000000	5.000000	320.000000	14.000000	23.000000	15.000000	330.000000	100.000000	3.000000	1.500000

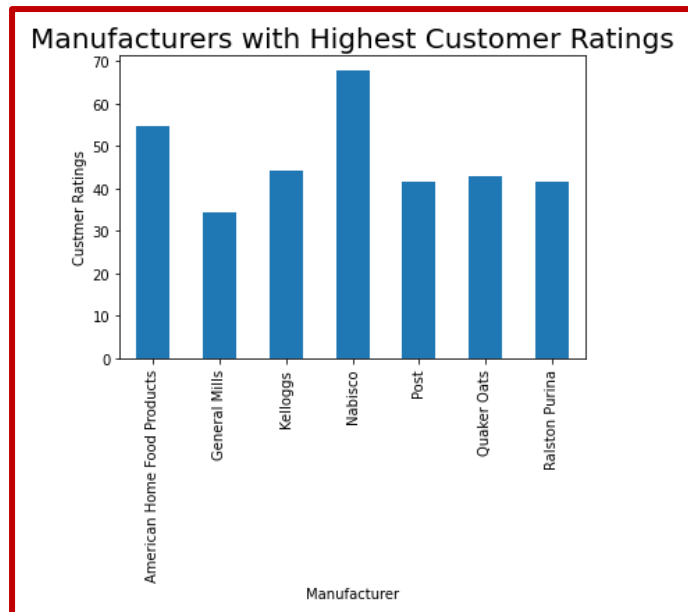
2. Replace the manufacturer names with the dictionary of names given here
 {'A':'American Home Food Products', 'G':'General Mills',
 'K':'Kelloggs', 'N':'Nabisco', 'P':'Post', 'Q':'Quaker Oats', 'R':'Ralston Purina'}

	index	mfr
0	Kelloggs	23
1	General Mills	22
2	Post	9
3	Quaker Oats	8
4	Ralston Purina	8
5	Nabisco	6
6	American Home Food Products	1

3. Find out the top five cereal manufacturers based on customer ratings

	name	mfr	type	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	weight	cups	rating
0	All-Bran with Extra Fiber	K	C	50	4	0	140	14.0	8.0	0	330	25	3	1.00	0.50	93.704912
1	Shredded Wheat 'n'Bran	N	C	90	3	0	0	4.0	19.0	0	140	0	1	1.00	0.67	74.472949
2	Shredded Wheat spoon size	N	C	90	3	0	0	3.0	20.0	0	120	0	1	1.00	0.67	72.801787
3	100% Bran	N	C	70	4	1	130	10.0	5.0	6	280	25	3	1.00	0.33	68.402973
4	Shredded Wheat	N	C	80	2	0	0	3.0	16.0	0	95	0	1	0.83	1.00	68.235885

4. With the help of a barplot, compare the manufacturers based on customer ratings



5. With the help of a pie chart, find out which of the manufacturers are the largest producers of cereals

