Day 17

DIY

Q1. Problem Statement: Multivariate EDA

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

1. Explore the DataFrame using `info()` and `describe()` functions
2. Find out the top five cereal manufacturers based on customer ratings
3. 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'}`
4. Find out cereal manufacturers whose products are rich in protein but contain minimal calories, with the help of a scatter plot
5. Using a heatmap, plot the correlation and covariance between all the nutritional values present in the 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:

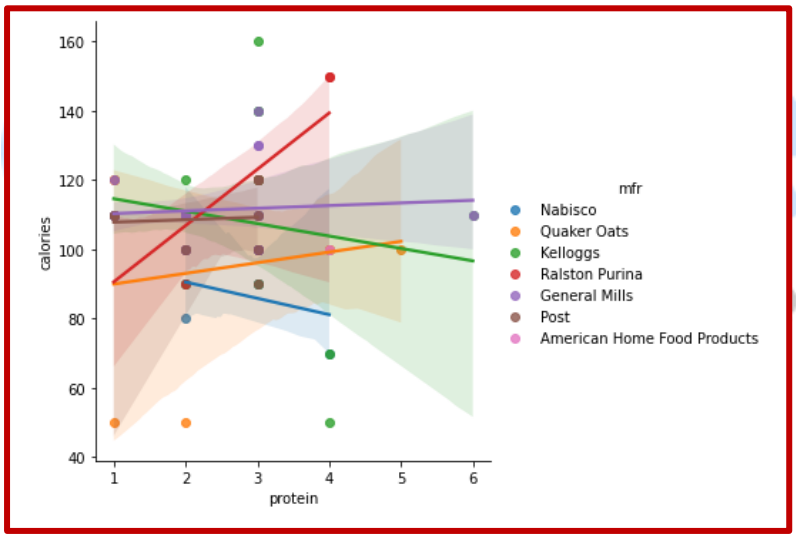
1. Replace the manufacturer names with the dictionary of names given here –

```
{'A':'American Home Food Products',
```

```
'G':'General Mills', 'K':'Kelloggs', 'N':'Nabisco',
'P':'Post', 'Q':'Quaker Oats',
'R':'Ralston Purina'}
```

```
0            Nabisco
1        Quaker Oats
2           Kelloggs
3           Kelloggs
4     Ralston Purina
           ...
72     General Mills
73     General Mills
74    Ralston Purina
75     General Mills
76     General Mills
Name: mfr, Length: 77, dtype: object
```

2. Find out cereal manufacturers whose products are rich in protein but contain minimal calories, with the help of a scatter plot



3. Using a heatmap, plot the correlation and covariance between all the nutritional values present in the cereals