PROJECT: EXPLORATORY DATA ANALYSIS

Data: cereals_data

Abstract

This project is designed for Full Time students of Data Science/Analytics for sharpening their skills related to Data Exploration. Students will also learn as how to work as team, showcase their work through Power Point Presentation for attracting stake holders and recruiters.

Project

Topic: Exploratory Data Analysis

Data Sets:

(1) cereals_data [77 rows by 16 columns]

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	name	mfr	type	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	weight	cups	rating
2	100%_Bran	N	C	70	4	1	130	10	5	6	280	25	3	1	0.33	68.40297
3	100%_Natural_Bran	Q	C	120	3	5	15	2	8	8	135	0	3	1	1	33.98368
4	All-Bran	K	C	70	4	1	260	9	7	5	320	25	3	1	0.33	59.42551
5	All-Bran_with_Extra_Fiber	K	C	50	4	0	140	14	8	0	330	25	3	1	0.5	93.70491
6	Almond_Delight	R	C	110	2	2	200	1	14	8		25	3	1	0.75	34.38484
7	Apple_Cinnamon_Cheerios	G	C	110	2	2	180	1.5	10.5	10	70	25	1	1	0.75	29.50954
8	Apple_Jacks	K	C	110	2	0	125	1	11	14	30	25	2	1	1	33.17409
9	Basic_4	G	C	130	3	2	210	2	18	8	100	25	3	1.33	0.75	37.03856
10	Bran_Chex	R	C	90	2	1	200	4	15	6	125	25	1	1	0.67	49.12025

(2) cereals_description

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1. Name: Name of cereal															
2. mfr: Manufacturer of cereal where A = American Home Food Products; G = General Mills; K = Kelloggs; N = Nabisco; P = Post; Q = Quaker Oats; R = Ralston Purina															
3. type: cold or hot															
4. calories: calories per serving															
5. protein: grams of protein															
6. fat: grams of fat															
7. sodium: milligrams of sodium															
8. fiber: grams of dietary fiber															
9. carbo: grams of complex carbohydrates															
10. sugars: grams of sugars															
11. potass: milligrams of potassium															
12. vitamins: vitamins and minerals - 0, 25, or 100, indicating the typical percentage of FDA recommended															
13. shelf: display shelf (1, 2, or 3, counting from the floor)															
14. weight: weight in ounces of one serving															
15. cups: number of cups in one serving															
16. rating: a rating of the cereals calculated by Consumer Reports															

Pretext:

TuSigma has approached your institute for getting assistance in analyzing data given by their client **Jolly Foods Inc**. **JF** wants to build their understanding about calories in relation with other nutrients and ratings across different manufacturers. Trusting your capability your institute's administration has given this special task to you. Post your analysis you alongwith **TuSigma** representative will be required to give an impressive presentation to **Jolly Foods Inc** for 10 minutes.

Your mentor has advised you to perform following tasks. He advised you to form a team of 2 members. As Data Science work is a team effort, your mentor wants you to learn good team work at this stage.

Tasks:

1) Data is having missing values. Develop an appropriate approach for addressing missing values
2) Evaluate all 16 variables one by one through concepts of sample statistics and build your understanding and observe interesting findings
3) Do slicing and dicing of the data which you may call Data Manipulation and do deeper analysis and observe interesting hidden insights
4) Post above, prepare power point presentation with 10 slides maximum for giving presentation to **JF**
5) Before giving presentation to **JF** you are required to submit following files to your mentor/Team Lead for his approval
   a. Python Script
   b. Power Point Presentation
6) You need to give 10 mnts presentation in teams during classroom session/s

Best of Luck!

Stay Safe, Stay Focussed & Stay Blessed!