**Project Definition**

**Purpose**

The purpose of this project is to find the most suitable cereal having the right ingredients from the cereal data set and find any possible co-relation between ingredients of cereals that may lead to a better knowledge for healthy breakfast.

**Problems/Opportunities Addressed**

The problem can be addressed is the lack of proper quantity of all the ingredients in the cereals intended for healthy breakfast.

**Project Scope**

A Dendrogram with the clusters summarized and a knowledge discovery from the Dendrogram.

**Key Stakeholders**

The key stake holders are all individuals who consume cereal on regular basis.

**Literature Survey**

In this report, the use of data mining tools on a pre-compiled dataset cereals using the tool ‘Weka’. There have been a major improvement and increase of concern among individuals for a perfect diet and a healthy life. So, this report is to satisfy that need, Data mining is the process of selecting, exploring and modeling large amounts of data. This process has become an increasingly extensive activity in all areas. Specially in medical science research. Data mining has resulted in the discovery of useful hidden patterns from massive databases. Data mining problems are often solved using different approaches from both computer sciences, such as multi-dimensional databases, machine learning, soft computing and data visualization; and statistics, including hypothesis testing, clustering, classification, and regression techniques.

It is found through various medical researches that the breakfast has a lot impact on an individual’s regular life and a healthy breakfast can enhance one’s daily activities. So, as the starting meal of the day everyone should maintain consistency in both food intake timings and the types of food they choose. Dietary consistency helps peoples to prevent diabetics, high blood pressure and heart diseases. Meal planning includes choosing nutritious foods and eating the right amount of food at the right time.

**Methods**

Step 1: I made a .arff file of the given data set and I open it on weka (data mining application).

Step 2: With the filter option, I added an id attribute to the data set for better consideration in the later steps.

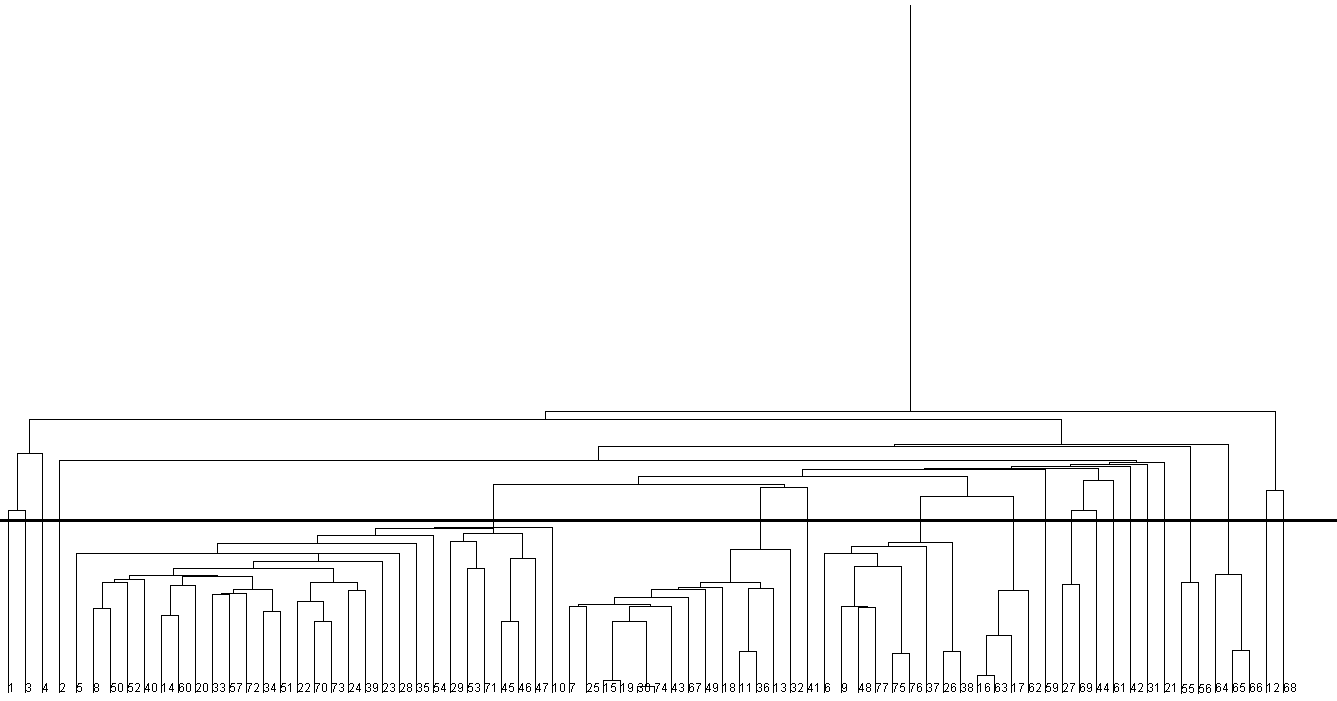
Step 3: From the Cluster tab I have chosen hierarchical clustering and applied it on the data set using training set.

Step 4: Right click on the result list and go to the ‘visualize tree’ option

Step 5: Took a screenshot and with that applied the cutting point on the tree

Step 6: Separate the clusters.

**Hierarchical cluster tree with cut points**



Data Set Structure:

@relation cereal

@attribute cerealname string

@attribute calories numeric

@attribute protein numeric

@attribute fat numeric

@attribute sodium numeric

@attribute dietaryfiber numeric

@attribute complexcarbohydrates numeric

@attribute sugars numeric

@attribute displayshelf numeric

@attribute potassium numeric

@data

**Cluster 1:**

(Low calorie, Medium protein, Low fat, Low sodium, High fiber, Low carbohydrate, Low sugar, High potassium)

100%\_Bran 70 4 1 130 10 5 6 280

**Cluster 2:**

(Low calorie, Medium protein, Low fat, High sodium, High fiber, Low carbohydrate, Low sugar, High potassium)

All-Bran 70 4 1 260 9 7 5 320

**Cluster 3:**

(Low calorie, Medium protein, Zero fat, Medium sodium, High fiber, Low carbohydrate, Zero sugar, High potassium)

All-Bran\_with\_Extra\_Fiber 50 4 0 140 14 8 0 330

**Cluster 4:**

(High calorie, Medium protein, High fat, Very Low sodium, Low fiber, Medium carbohydrate, Medium sugar, Medium potassium)

100%\_Natural\_Bran 120 3 5 15 2 8 8 135

**Cluster 5:**

(Medium calorie, Medium protein, Medium fat, Medium sodium, Medium fiber, Medium carbohydrate, Medium High sugar, Medium potassium)

Almond\_Delight 110 2 2 200 1 14 8 -1

Basic\_4 130 3 2 210 2 18 8 100

Nutri-Grain\_Almond-Raisin 140 3 2 220 3 21 7 130

Oatmeal\_Raisin\_Crisp 130 3 2 170 1.5 13.5 10 120

Just\_Right\_Fruit\_&\_Nut 140 3 1 170 2 20 9 95

Clusters 110 3 2 140 2 13 7 105

Raisin\_Nut\_Bran 100 3 2 140 2.5 10.5 8 140

Cracklin'\_Oat\_Bran 110 3 3 140 4 10 7 160

Grape\_Nuts\_Flakes 100 3 1 140 3 15 5 85

Quaker\_Oat\_Squares 100 4 1 135 2 14 6 110

Total\_Whole\_Grain 100 3 1 200 3 16 3 110

Grape-Nuts 110 3 0 170 3 17 3 90

Nutri-grain\_Wheat 90 3 0 170 3 18 2 90

Crispix 110 2 0 220 1 21 3 30

Total\_Corn\_Flakes 110 2 1 200 0 21 3 35

Triples 110 2 1 250 0 21 3 60

Double\_Chex 100 2 0 190 1 18 5 80

Just\_Right\_Crunchy\_\_Nuggets 110 2 1 170 1 17 6 60

Crispy\_Wheat\_&\_Raisins 100 2 1 140 2 11 10 120

"Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats" 120 3 2 160 5 12 10 200

Great\_Grains\_Pecan 120 3 3 75 3 13 4 100

Product\_ 19 100 3 0 320 1 20 3 45

Fruitful\_Bran 120 3 0 240 5 14 12 190

Post\_Nat.\_Raisin\_Bran 120 3 1 200 6 11 14 260

Total\_Raisin\_Bran 140 3 1 190 4 15 14 230

"Muesli\_Raisins,\_Dates,\_&\_Almonds" 150 4 3 95 3 16 11 170

"Muesli\_Raisins,\_Peaches,\_&\_Pecans" 150 4 3 150 3 16 11 170

Mueslix\_Crispy\_Blend 160 3 2 150 3 17 13 160

Bran\_Flakes 90 3 0 210 5 13 5 190

**Cluster 6:**

(Medium calorie, Low protein, Low fat, Medium sodium, Low fiber, Low Medium carbohydrate, High sugar, Very Low potassium)

Apple\_Jacks 110 2 0 125 1 11 14 30

Froot\_Loops 110 2 1 125 1 11 13 30

Cocoa\_Puffs 110 1 1 180 0 12 13 55

Count\_Chocula 110 1 1 180 0 12 13 65

Fruity\_Pebbles 110 1 1 135 0 13 12 25

Trix 110 1 1 140 0 13 12 25

Lucky\_Charms 110 2 1 180 0 12 12 55

Smacks 110 2 1 70 1 9 15 40

Nut&Honey\_Crunch 120 2 1 190 0 15 9 40

Corn\_Pops 110 1 0 90 1 13 12 20

Cap'n'Crunch 120 1 2 220 0 12 12 35

Honey\_Graham\_Ohs 120 1 2 220 1 12 11 45

Cinnamon\_Toast\_Crunch 120 1 3 210 0 13 9 45

Golden\_Grahams 110 1 1 280 0 15 9 45

**Cluster 7:**

(Medium calorie, Low protein, Low fat, High sodium, No fiber, High carbohydrate, Low sugar, Low potassium)

Kix 110 2 1 260 0 21 3 40

**Cluster 8:**

(Medium calorie, Low protein, Low fat, High sodium, Medium fiber, High carbohydrate, Medium sugar, Low potassium)

Apple\_Cinnamon\_Cheerios 110 2 2 180 1.5 10.5 10 70

Multi-Grain\_Cheerios 100 2 1 220 2 15 6 90

Wheaties\_Honey\_Gold 110 2 1 200 1 16 8 60

Wheat\_Chex 100 3 1 230 3 17 3 115

Wheaties 100 3 1 200 3 17 3 110

Honey\_Nut\_Cheerios 110 3 1 250 1.5 11.5 10 90

Frosted\_Flakes 110 1 0 200 1 14 11 25

Honey-comb 110 1 0 180 0 14 11 35

**Cluster 9:**

(Medium calorie, Low protein, Zero fat, High sodium, Very Low fiber, High carbohydrate, Low sugar, Very Low potassium)

Corn\_Chex 110 2 0 280 0 22 3 25

Rice\_Krispies 110 2 0 290 0 22 3 35

Corn\_Flakes 100 2 0 290 1 21 2 35

Rice\_Chex 110 1 0 240 0 23 2 30

**Cluster 10:**

(Medium calorie, Medium protein, Low fat, High sodium, Medium fiber, Medium carbohydrate, High sugar, High potassium)

Raisin\_Bran 120 3 1 210 5 14 12 240

**Cluster 11:**

(Low calorie, Low protein, Zero fat, Low sodium, Medium fiber, Medium carbohydrate, Medium sugar, Low potassium)

Frosted\_Mini-Wheats 100 3 0 0 3 14 7 100

Strawberry\_Fruit\_Wheats 90 2 0 15 3 15 5 90

**Cluster 12:**

(Low calorie, Medium protein, Low fat, Zero sodium, Zero fiber, Medium carbohydrate, Low sugar, Low potassium)

Maypo 100 4 1 0 0 16 3 95

**Cluster 13:**

(Low calorie, Low protein, Zero fat, Zero sodium, Low fiber, Medium carbohydrate, Medium sugar, Medium potassium)

Raisin\_Squares 90 2 0 0 2 15 6 110

**Cluster 14:**

(Low calorie, Medium protein, Medium fat, Medium sodium, Low fiber, Medium carbohydrate, Medium sugar, Low potassium)

Life 100 4 2 150 2 12 6 95

**Cluster 15:**

(Low calorie, Low protein, Zero fat, Low sodium, Zero fiber, Low carbohydrate, High sugar, Low potassium)

Golden\_Crisp 100 2 0 45 0 11 15 40

**Cluster 16:**

(Low calorie, Medium protein, Zero fat, Low sodium, Low fiber, Low carbohydrate, Zero sugar, Negative Data on potassium)

Cream\_of\_Wheat\_(Quick) 100 3 0 80 1 21 0 -1

**Cluster 17:**

(Low calorie, Low protein, Zero fat, Zero sodium, Low fiber, Medium carbohydrate, Zero sugar, Low potassium)

Puffed\_Rice 50 1 0 0 0 13 0 15

Puffed\_Wheat 50 2 0 0 1 10 0 50

**Cluster 18:**

(Low calorie, Medium protein, Zero fat, Zero sodium, Medium fiber, High carbohydrate, Zero sugar, Medium potassium)

Shredded\_Wheat 80 2 0 0 3 16 0 95

Shredded\_Wheat\_'n'Bran 90 3 0 0 4 19 0 140

Shredded\_Wheat\_spoon\_size 90 3 0 0 3 20 0 120

**Cluster 19:**

(Medium calorie, High protein, Medium fat, High sodium, Low fiber, High carbohydrate, Low sugar, Medium potassium)

Cheerios 110 6 2 290 2 17 1 105

**Cluster 20:**

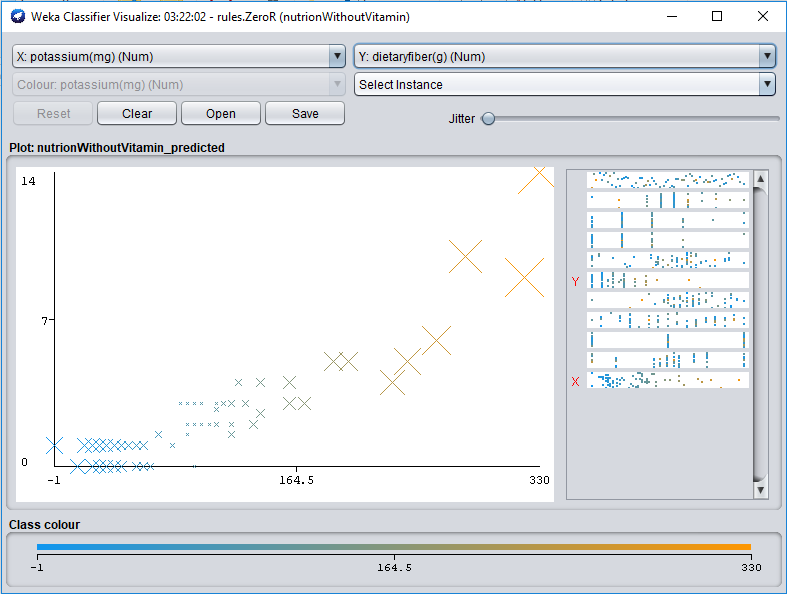
(Medium calorie, High protein, Zero fat, High sodium, Low fiber, High carbohydrate, Low sugar, Low potassium)

Special\_K 110 6 0 230 1 16 3 55

**Question answer:**

**Question: Is a strong correlation between dietary fiber and potassium?**

**Answer:** Yes, there is a strong correlation between dietary fiber and potassium. That is because the correlation of the fiber is proportional to potassium. We can also see from the plot area in the graph which is given below that the position of dietary fiber and potassium increased linearly or proportionally.



**Question: Are groups of cereals from which we can choose according to our preferences?**

**Answer:** Yes, from the 23 clusters anyone can get any combination of he/she likes such as-

1) People who want low calorie and no fat and no sugar can choose cluster 3 or 16.

2) People can take high potassium and high fiber and can have better cereals on cluster 1,2 & 3.

3) People who want high carbohydrates one can choose cereal from cluster 7,8 & 9.

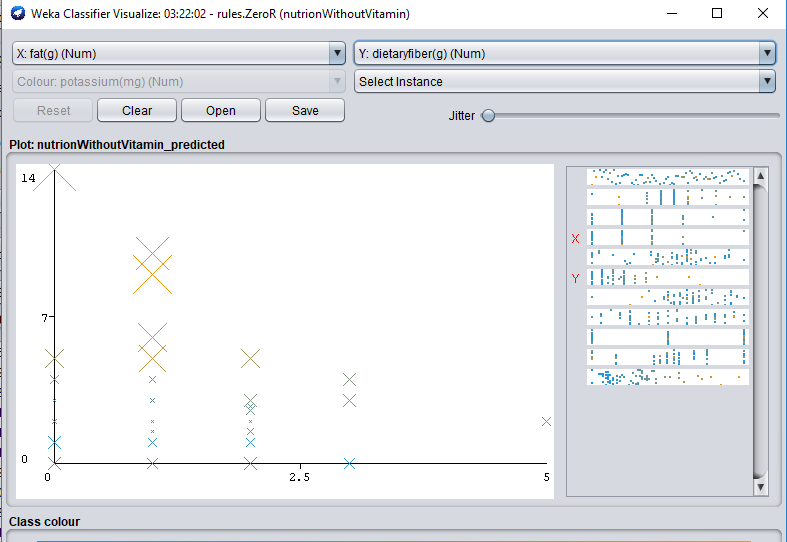
4) People who want high calories and fat with medium sugar can pick a cereal from cluster 4.

Etc.

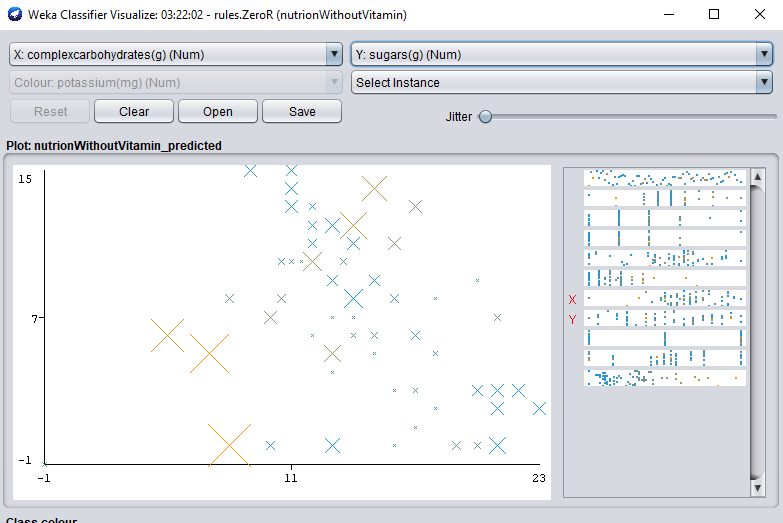
**Question: See other correlation between the data given in the files.**

**Answer:**

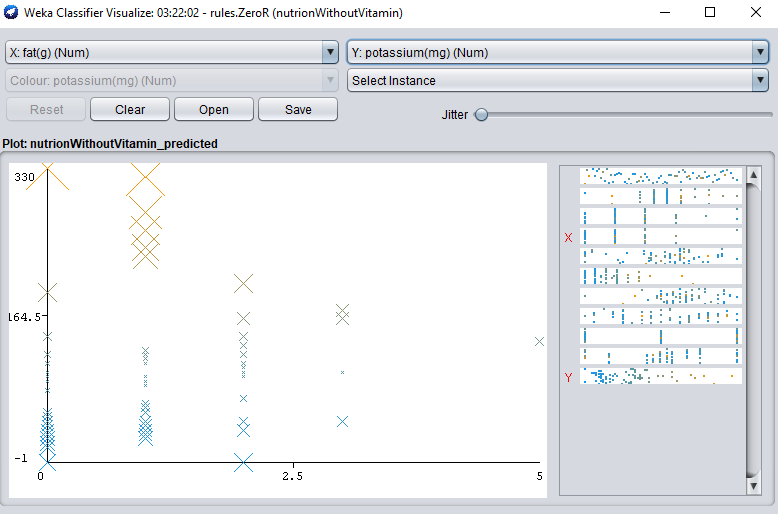
🡪Fat and Fiber corelation



🡪Carbohydrate and Sugar co-relation



🡪Fat and Potassium co-relation



Question 4:

Answer:

In our data set, we can notice that, where the fat is low, the sugar is low too.

I found that, wherever the calorie is low, the protein is low or on medium amount.

Also, this could be found that, the status of fiber does not affect calorie.