

**PROJECT TITLE:-**

**CEREAL ANALYSIS**

Submitted By:

Y.S.N Praneeth

T. Ram Vardhan

T. Teja Srinivasu

S. Srujan

SK. Siddik

CEREAL ANALYSIS

**Introduction :**

Artificial Intelligence(AI) is the simulation of human intelligence possessed by machines . AI Includes many technologies like Automation,Machine Learning,Natural Language processing,Robotics,etc.

Machine Learning is a part of AI which is evolved from computer science that primarily studies the design of algorithms that can learn from experience. To learn, they need **data** that has certain attributes based on which the algorithms try to find some meaningful predictive patterns. The ultimate goal of ML algorithms is to be able to take decisions without any human intervention correctly. Predicting the stocks or weather are a couple of applications of machine learning algorithms.

Python is a general-purpose high level programming language that is being increasingly used in data science and in designing machine learning algorithms

Machine Learning is the Science of getting a machine to act without programming . Deep learning is a subset of ML that can be thought of as the automation of predictive analytics . There are three types of Machine Learning Algorithms :

* **Supervised learning :** Data sets are labeled so that patterns can be detected and used to label new data sets.
* **Unsupervised learning** **:** Data sets aren’t labeled and are sorted according to similarities or differences.
* **Reinforcement learning** **:** Data sets aren’t labeled but,after performing an action or several actions,the AI system is given feedback

**Objectives of Research :**

The project objective is to find the high dietary food that is predicted on the basis of rating of the food. The research should achieve the best food with high diet analytics.

The food is rated accordingly to get dietary benefits . Rating is given high for the food with high protein , fibre , vitamins and Rating is given low for the food with low fats , sugars , etc

The main objectives of the project are :

1. To find Which quantities are showing more impact on the rating of food
2. To show the food which is impacting less on the rating of food
3. Measure the inputs and outputs
4. Collect the data set
5. Find how the quantity is impacting rating
6. Use different ML algorithms to predict which food has high dietary benefits

**Problem Statement :**

A customer wants to buy some food items with high dietary benefits so that he wants to know which food item has high dietary benefits.It is so difficult to choose an item .

Usually a customer expects to consume dietary cereals with high proteins, fiber and low sugars, fats. Predicting a brand with high dietary cereals became a big issue.

Hence we came up with supervised machine learning where we can predict the brand ratings using Multiple linear regression.

**Review Of Literature :**

Given a data set with some ingredients and cereals that present in the food and food ratings have been taken and based on rating the food can be categorized.

The food with high ratings can be treated as high dietary benefit food and else are treated as low dietary benefit food.

Rating of food is given on the basis of content it consists. The food consists of proteins , carbohydrates , fats , sugar , vitamins , sodium , potassium ,

etc. These quantities determine the food rating and how it can be more or less beneficiary diet food. We use machine learning to predict the food with high beneficiary diet. Machine learning is used to determine which quantities are impacting rating and how they can change the rating of food.

Here we use multiple regression algorithm to predict the rating of the food.We take type , calories , vitamins , proteins , fats , sugars , sodium , potassium as variables that change the rating of food . In this two or more items can be treated as inputs and rating as an output. Data set is divided into training and testing sets through which data can be trained by the machine with inputs and rating can be predicted. In training set the new predicted outputs are matched with the old output then the accuracy of the machine can be found.

**Data Collection :**

The Data set is given and consists of the required items and quantities that can be used to predict the rating of the food.

A screenshot of a computer

Description automatically generated

**Methodology :**

Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables.

The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable). The variables we are using to predict the value of the dependent variable are called the independent variables (or sometimes, the predictor, explanatory or regressor variables).

Multiple regression also allows you to determine the overall fit (variance explained) of the model and the relative contribution of each of the predictors to the total variance explained.

There are 3 major uses for multiple linear regression analysis.

**1.** It might be used to identify the strength of the effect that the independent variables have on a dependent variable.

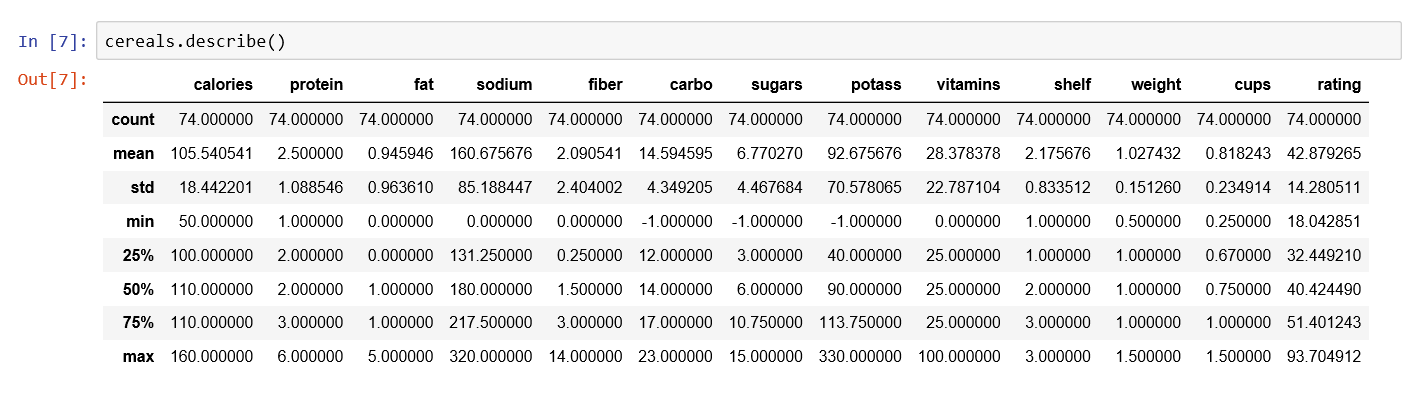
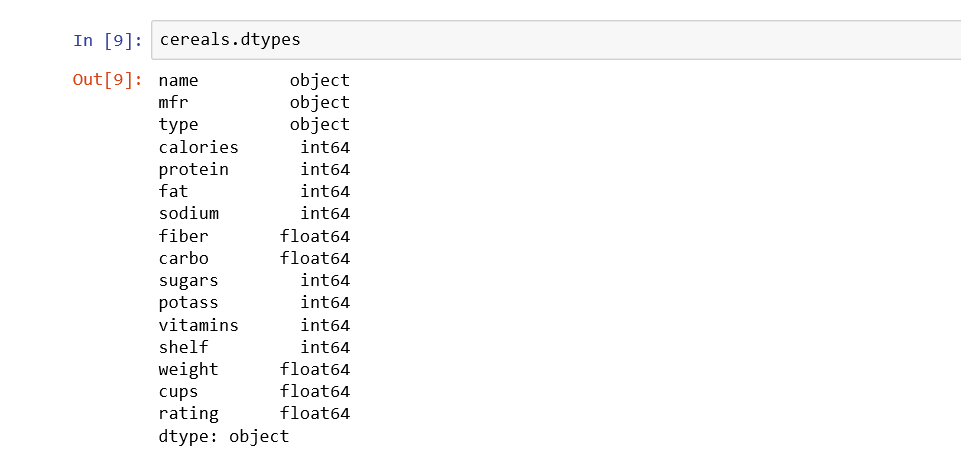
**2.** Second, it can be used to forecast effects or impacts of changes.  That is, multiple linear regression analysis helps us to understand how much will the dependent variable change when we change the independent variables.

**3.** Multiple linear regression analysis predicts trends and future values.  The multiple linear regression analysis can be used to get point estimates.

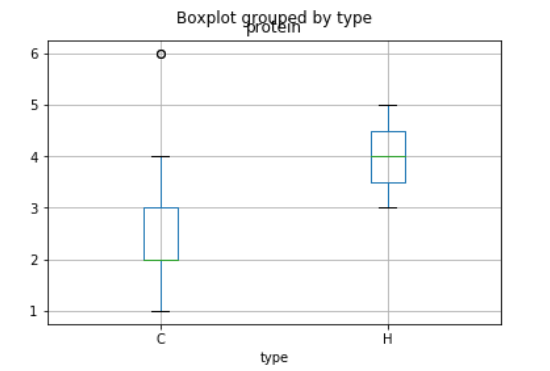
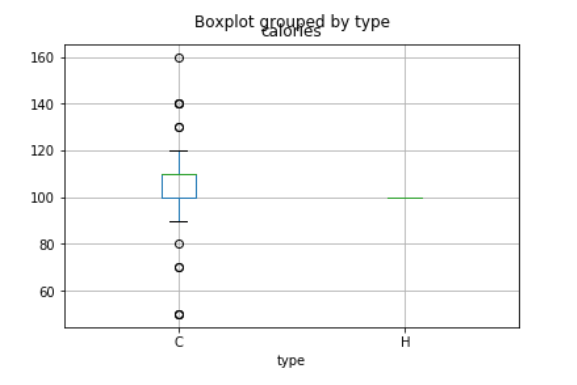
**Exploratory Data Analysis :**

**Figures and Tables :**

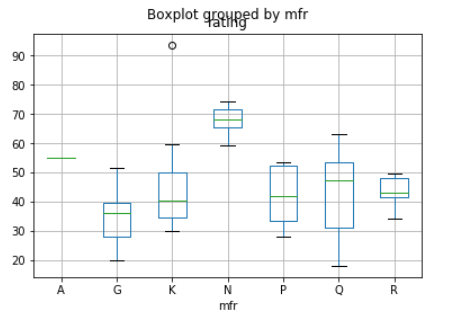
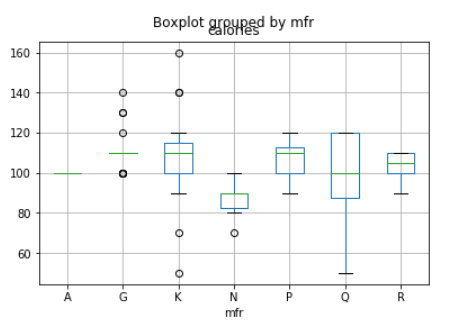
Description of Data:

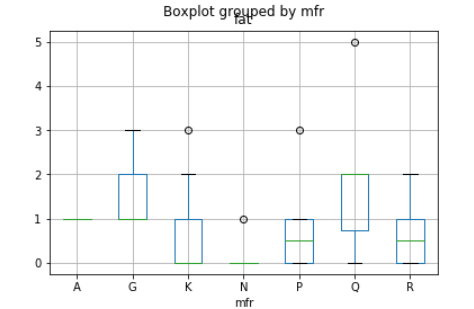
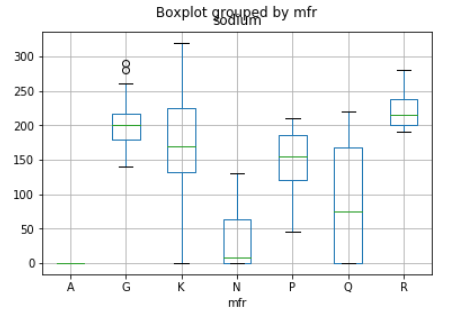
Data describeData Types 

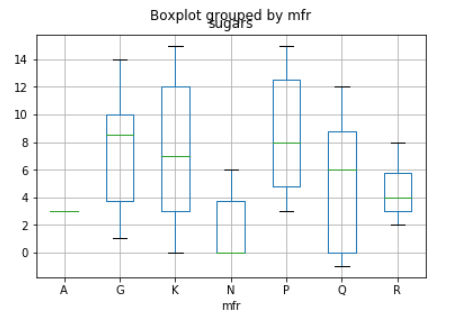
Box Plot



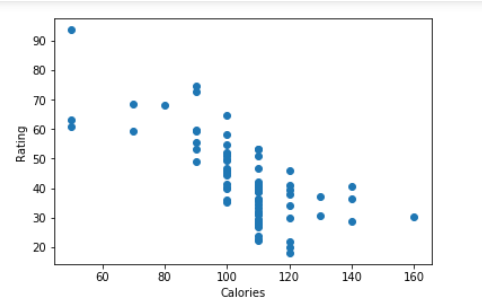
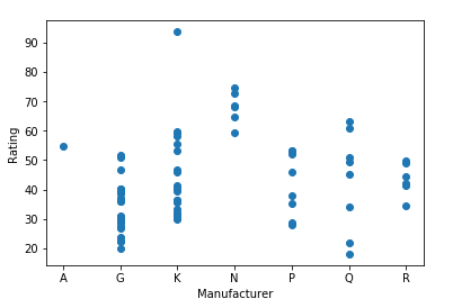
Box Plot by mfr



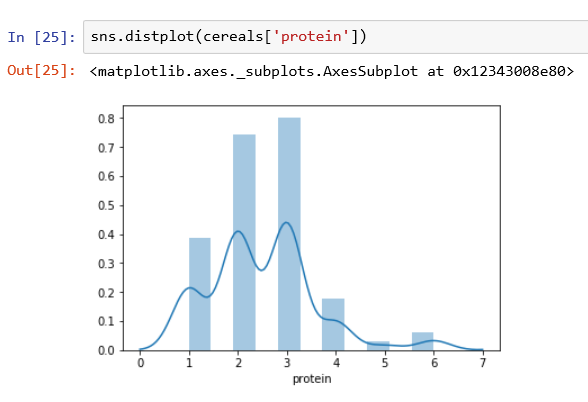
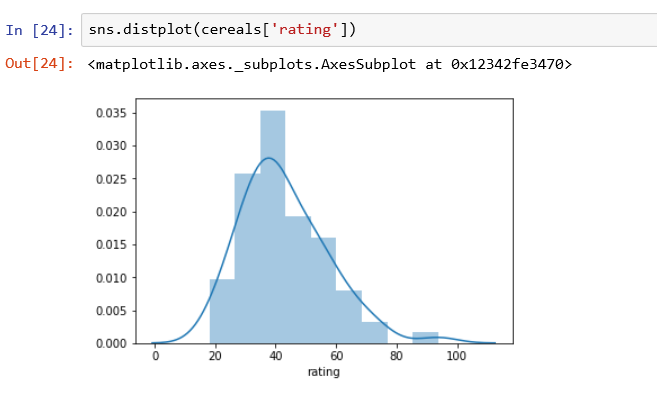


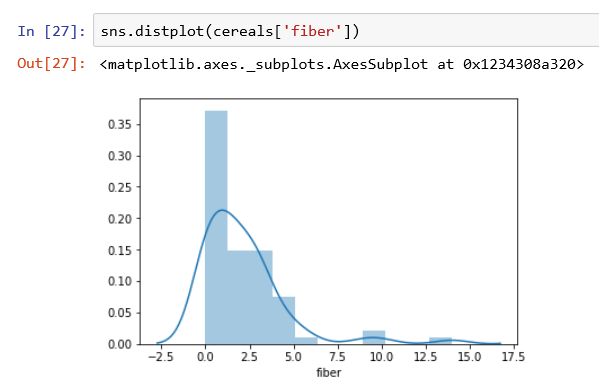
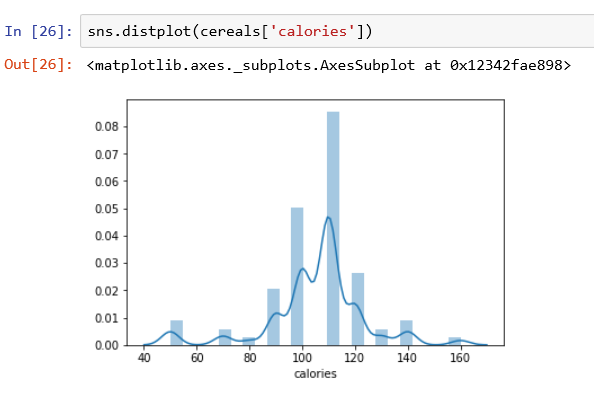


Scatter Plot

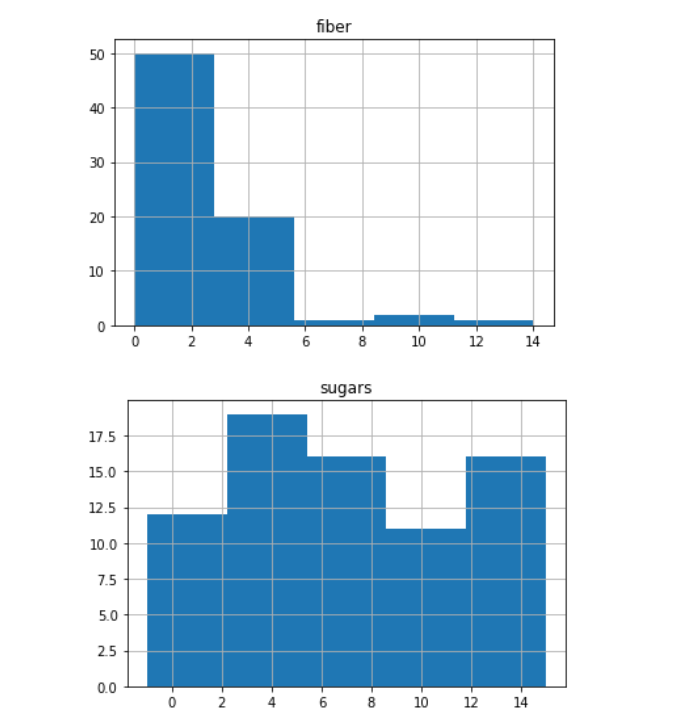
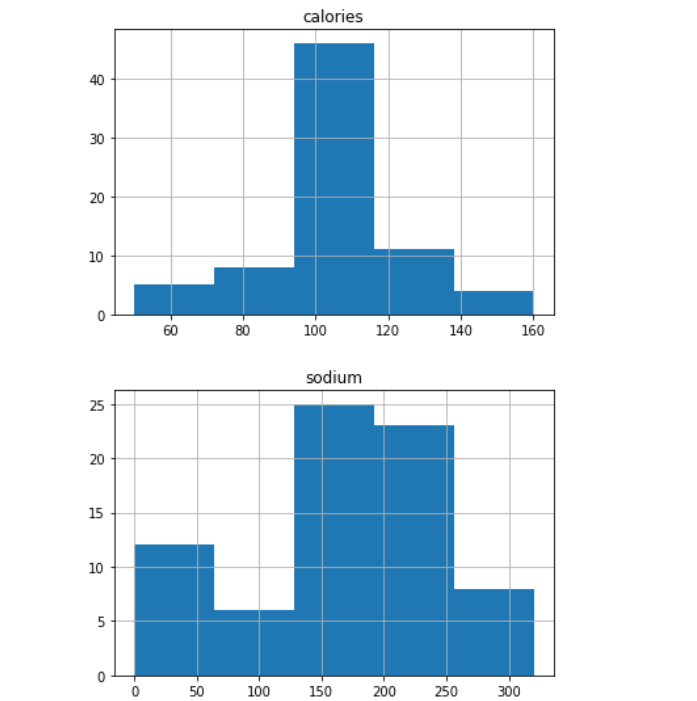


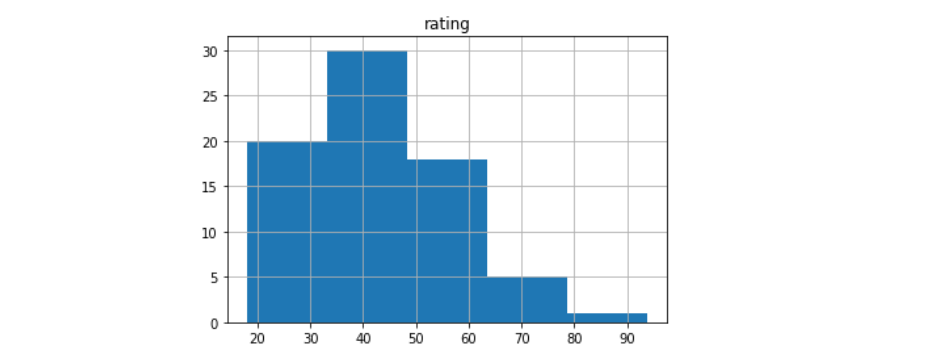
Bar graph





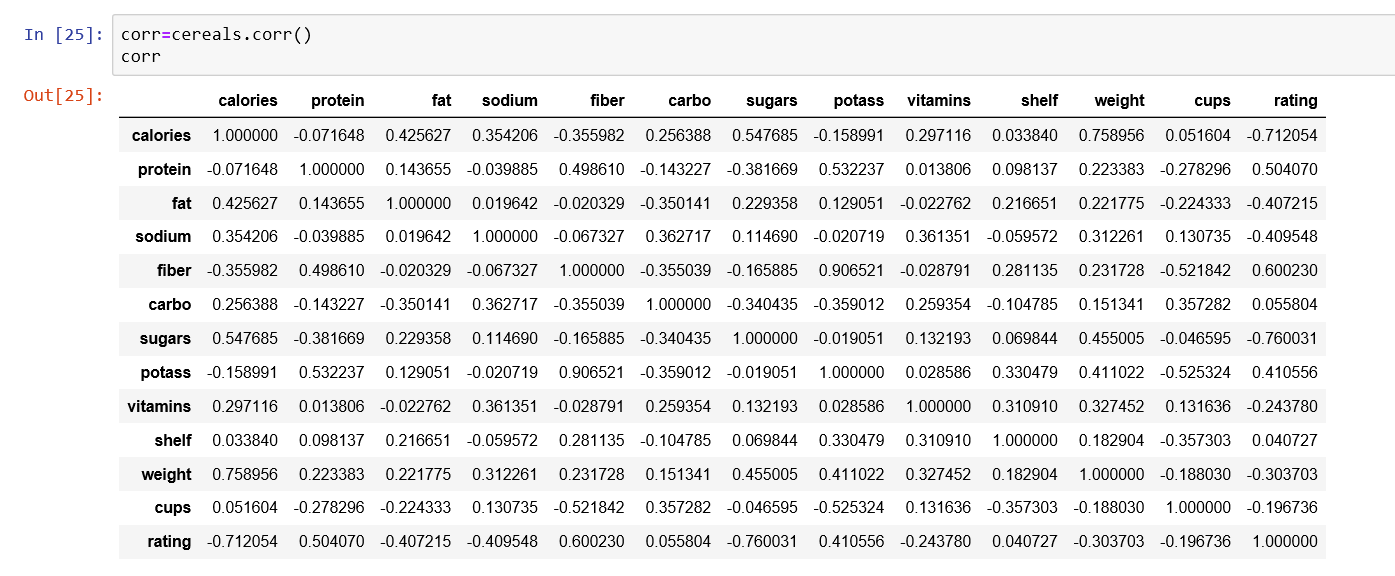
Data Visualization :

****

****

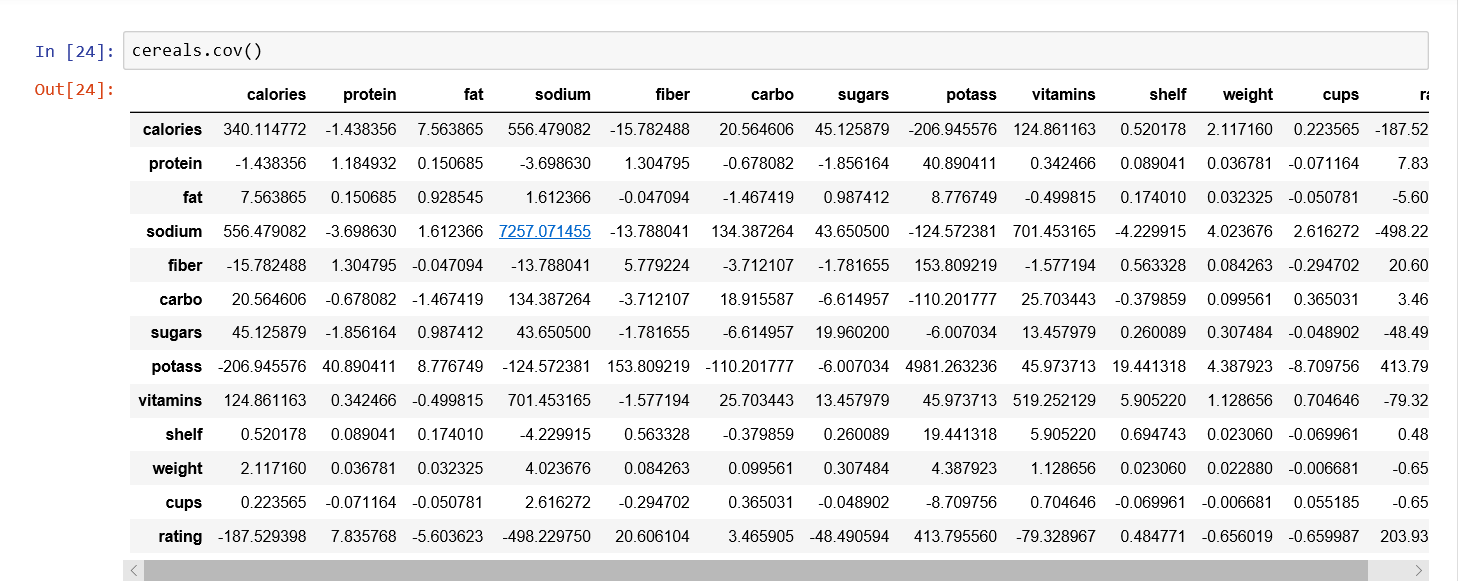
Data Correlation :

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

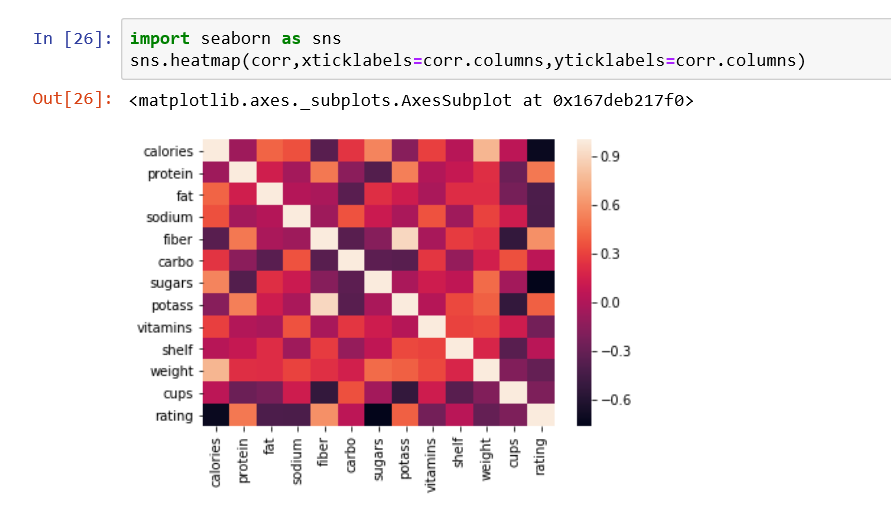


Data Covariance :

Covariance is the measure of relationship among different variables.



Heat map :



**Data Modelling :**

In this project Multiple Linear Regression is used.Given Data set consists of different items.We take multiple inputs for the prediction of output.In given data , 6 quantities are taken as inputs for the prediction of output.The inputs taken are

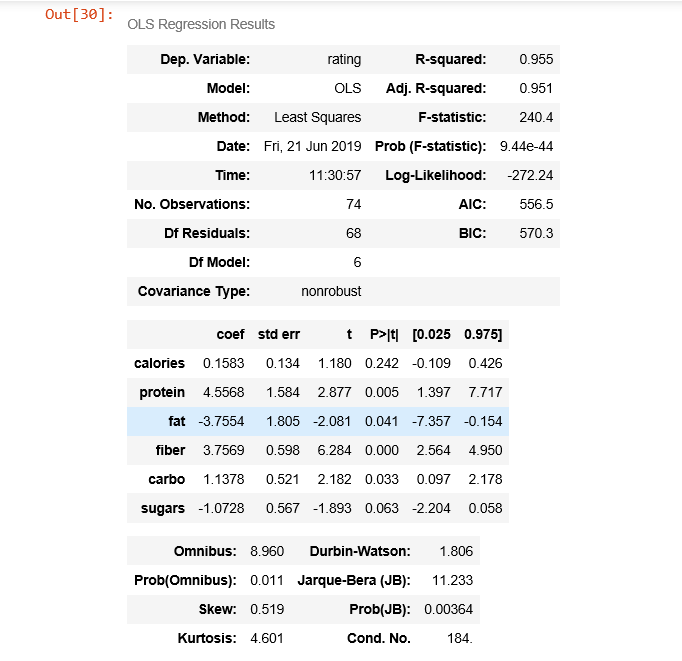
1. Calories
2. Protein
3. fat
4. carbo
5. Fiber
6. sugars

From these inputs given ,Rating can be predicted. We use matplot and boxplot functions to plot different graphs and maps to determine the dependency of each quantity with the prediction.

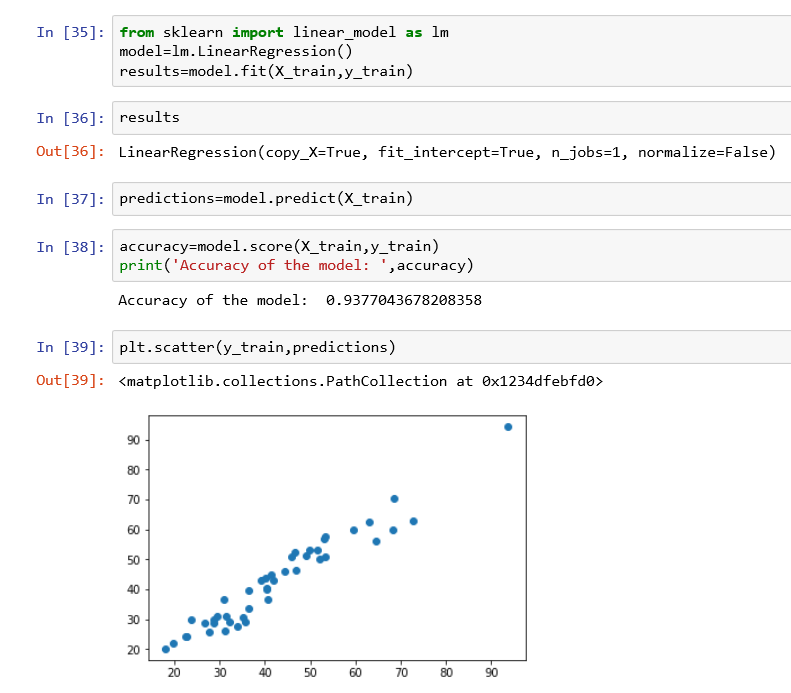
After this, OLS model can be implemented through which results prediction analysis can be identified.We divide the data set into two parts one is training set and other is testing set.

Machine learns how to predict the data from training and applies it on testing.In both training and testing,we take two variables X and Y , for input and output.With the new predictions by matching with the given predictions model checks and gives the percentage of accuracy of predicting the rating .In this project 60% is given for Training and 40% is given for Testing.

**OLS Model (Ordinary Least Square)**



Model Evaluation and Accuracy





**Conclusion :**

The model can predict the rating of the food more accurate by giving the inputs which are the cereals and ingredients present in the food. Thus,a customer can get high dietary food by the rating of the food given to it from the cereals and ingredients present. In the above model Training and Testing is given 60% and 40% where we get best model with accuracy of 93.9 percentage.