**Module 2**

**Week 2: R Practice**

**Logo

Description automatically generated**

**ALY6010**

**Instructor: Prof. Amin Karimpour**

**Date: 14 Nov 2021**

**Submitted By:-**

**Jeseeka Shah**

**Nuid - 002134289**

**Introduction**

The dataset is about the cereals that we consume everyday for breakfast. The dataset consist of 77 rows and 16 columns. I have taking this dataset as I was interested in knowning that what makes them and how the contribute to a healthy diet or not ?.

This analysis is done based on the contents that make cereals like protein, sugar,carbohyrates, fat, sodium along with some other factors such as ratings, brands, type of consumption of beverage along with it and manufacturers involved.

**Exploratory Data Analytics**

Initially, the data was read from the csv file and cleaned. The abbreviations are converted to its original manufacturer names as follows.

**Manufacturer of Cereals**

|  |  |
| --- | --- |
| Abbreviation of manufacturers | Name of the manufacturers |
| A | American Home Food Products |
| G | General Mills |
| K | Kellogg’s |
| N | Nabisco |
| P | Post |
| Q | Quaker Oats |
| R | Ralston Purina |

**Performing Descriptive Analytics**

Performing the descriptive statistics on the features such as fat, calories, sugar, fiber and protein to get an idea in general about the proportion of the content present in cereals.

**Table 1:**

Table

Description automatically generated

**Figure 1: Parameter that make the contents of the cereals**

The same is plotted using box plot to get idea clear visual understanding what makes the cereal. Looking at the values we can say cereals are not healthy are they are advertised to be. The cereals mainly contain calories and very little fiber and protein. The boxplot shows that carbohydrates are maximum excluding a single outlier followed by sugar. There are few exceptions in case of fiber and protein but in total the fiber and protein contain is also very less (shown in figure 2: Boxplot).

Chart, box and whisker chart

Description automatically generated

Figure 2: Cereal Composition

**Total product and Type of beverage used in cereals**

Chart, timeline

Description automatically generated

Figure 2: Total product and type of beverage required as manufactured in cereals

The figure 2 shows Kellogg’s and General Mills have maximum cereal product compared to other manufacturers and least cereals are produced by American Home Food Products out of the seven manufactures. Also, the second bar plot shows that maximum cereal manufactured are to be eaten with cold beverages than hot ones. Only American Home Food products produces cereals that are only to be eaten with hot beverages.

**Calories Content in Cereals**

The Figure 4 graph represents mean calories present in the cereals manufactured by the seven manufacturing companies. The maximum mean calories were found to be in **Kellogg’s** followed by **General Mills** and least in American home food products. Among all the **Quaker Oats** have done better as it has manufactured products almost same number of cereals just **like Ralston Purina** yet it has less mean calories than Ralston Purina and Post.

Chart, scatter chart

Description automatically generated

Figure 4: Calories present in cereals

**Rating Vs. Manufactures**

Chart, scatter chart

Description automatically generated

Figure 5: Rating Vs. Manufactures

The maximum rating is given to Kellogg’s that is above 90 percent but that is an outlier, and the entire performance can not be decided on it. But looking at the graph quaker oats have got good amount ratings compared to Post, Ralson Purina and American home food products.

**Comparing Sugar w.r.t rating and manufactures**

Chart, scatter chart

Description automatically generated Chart

Description automatically generated

Figure 6: Rating Vs. Sugar and Calories Figure 7: Sugar vs. Manufactures

Figure 6 clearly shows the relationship between fall in rating as the sugar and calories increase in the cereal. The blue line in figure 6 shows steep decline in the rating. And figure 7 shows that sugar rate in cereals manufactured by different manufacture. It can be clearly seen that Quaker Oats again has less sugar in their cereals compared to Kellogg’s, General Mills and Post. It’s clear that Quaker Oats is a better brand to purchase for cereals rather than other brands.

**Quaker Oat -Protein Analysis**

Chart

Description automatically generated

Figure 8: Comparing Protein Vs. Fiber & Carbo and Box Plot Protein

In figure 8, shows that there is positive linear relationship between protein and fiber whereas there is a negative relationship between the carbo and protein of Quaker Oats graph. The last graph shows are the Box plot which shows the mean protein to be around 2.5 in Quacker oats.

Our analysis has used scatter plot and jitter plot mainly as the to see the proper distribution and density which is not visible when you use boxplot. Looking at **Boxplot** in figure 8 does not give us the clarity on the protein distribution but when we check the boxplot in figure 5 represents the distribution and range both clearly with the help of jitter plot. **Jitter plot** is used to show proper **distribution** of the density or clearer view of overlapping points. Whereas **Scatterplot** is better way of showing **relationship** between two variables. These charts clearly show the relationship, distribution without blocking or hiding any point like boxplot which is seen in figure 8.

The **three-line table format** is when a table name on the **top**, category name on the **rows** side and column names on **columns** side with no borders except on the top and bottom of the table.

Table2: Showing the contents of Quaker Oats

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean  (n=8) | Median  (n=8) | Standard Deviation  (n=8) |
| Protein  Carbohydrates  Fiber | 2.620  10.000  1.3400 | 2.500  12.00  1.500 | 1.598  4.810  0.998 |

The table should be clear and show represent all the values properly along with units or sample set used. Above is the example of the three -line table format.

**Summary**

The analysis on cereals showed as that cereals in general have a lot of carbohydrates and sugar compared to protein and fiber. The Kellogg’s brand maximum products in cereals compared to other six. American Home Food Product is the only cereal brand that makes cereals for hot beverages. Also, mean calories intake is maximum in Kellogg’s and minimum in American Home Food Product. Furthermore, the Kellogg’s seems to have maximum rating, but the products are in variety, but Quaker Oats has lesser variety yet has been given a lot of votes. Quaker Oats cereal has less sugar content than other brand like Kellogg’s, Post, and General Mills. After protein analysis on the quaker oats it showed that with carbohydrate increase the protein content reduced and with protein increase the fiber content also increases.

**Reference**

1)Axis labels not showing up. (2016, March 8). Stack Overflow.

<https://stackoverflow.com/questions/35876938/axis-labels-not-showing-up>

2)Sharghi, S. (2018, September 14). US cereal dataset Analysis [Blog Post]. Wordpress.

<https://blogs.bgsu.edu/math6820simas/2018/09/14/us-cereal-data-set-analysis/>

3)J. (2021, June 15). 80 Cereals - R. Kaggle. https://www.kaggle.com/jcraggy/80-cerealsr#Best-Predictors-of-Good-Rating--:-Fiber,-Protein <https://thesoulfullproject.com/blogs/soulfull-living/hot-cereal>

4)Data Cleanup: Remove NA rows in R. (2020, December 17). ProgrammingR.

<https://www.programmingr.com/examples/remove-na-rows-in-r/>