**Introduction**

Hello, nice to meet you!

We are **"snack for a friend"** - A non-profit organization that donates candy and snacks to children from needy families every week. There is no doubt that a donation of nutritious food is important, but there is nothing like something sweet to make a human heart happy.

We believe that snacks and sweets often boost the mood and are a comforting place on a difficult day. But at the same time it is important for us to keep the children healthy. Therefore, we hired Aviv and Elad to help us find the ultimate snacks for our children!

We provided them tables with a lot of data on the various snacks available to us. We would like Aviv and Elad to recommend us on the 3 top snacks in each category (Candy, Cookies & Biscuits, Cakes, Chips & Pretzels, Chocolate, Popcorn & Peanuts) by considering the following requirements:

1. A snack that is the least harmful to health (in terms of Calories, Carbs, Sugars and Fats)

2. A candy that contains relatively many vitamins.

3. A snack with a wrapper as colorful and eye-catching as possible for children.

4. A recommendation for one candy that is suitable for the sensitivity of peanuts/nuts

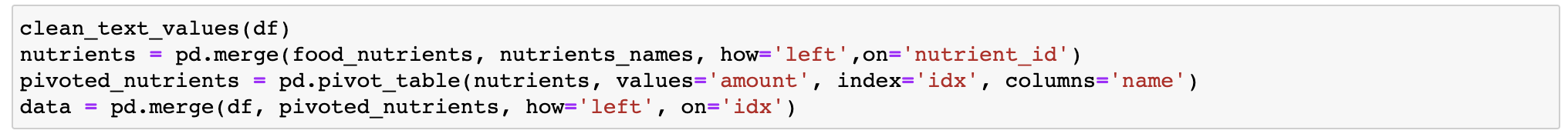
5. Nice to have - recommendation of products from the same brand to reduce purchase costs.

It is important to note that we put the health consideration at the top of our priorities!

Therefore, in our choices, it will take precedence over the colors and beauty of the snack wrapper.

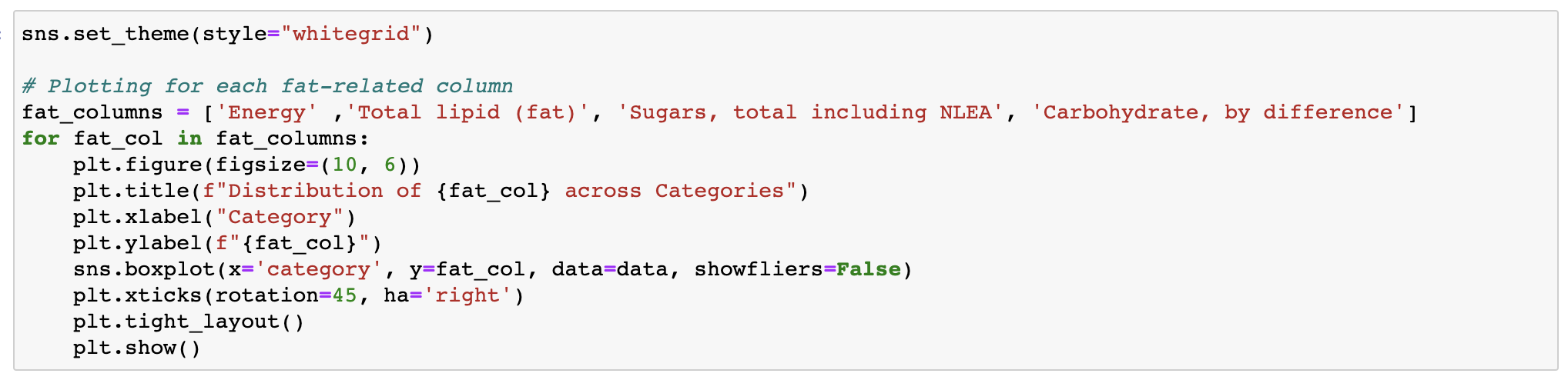
**Read the data**

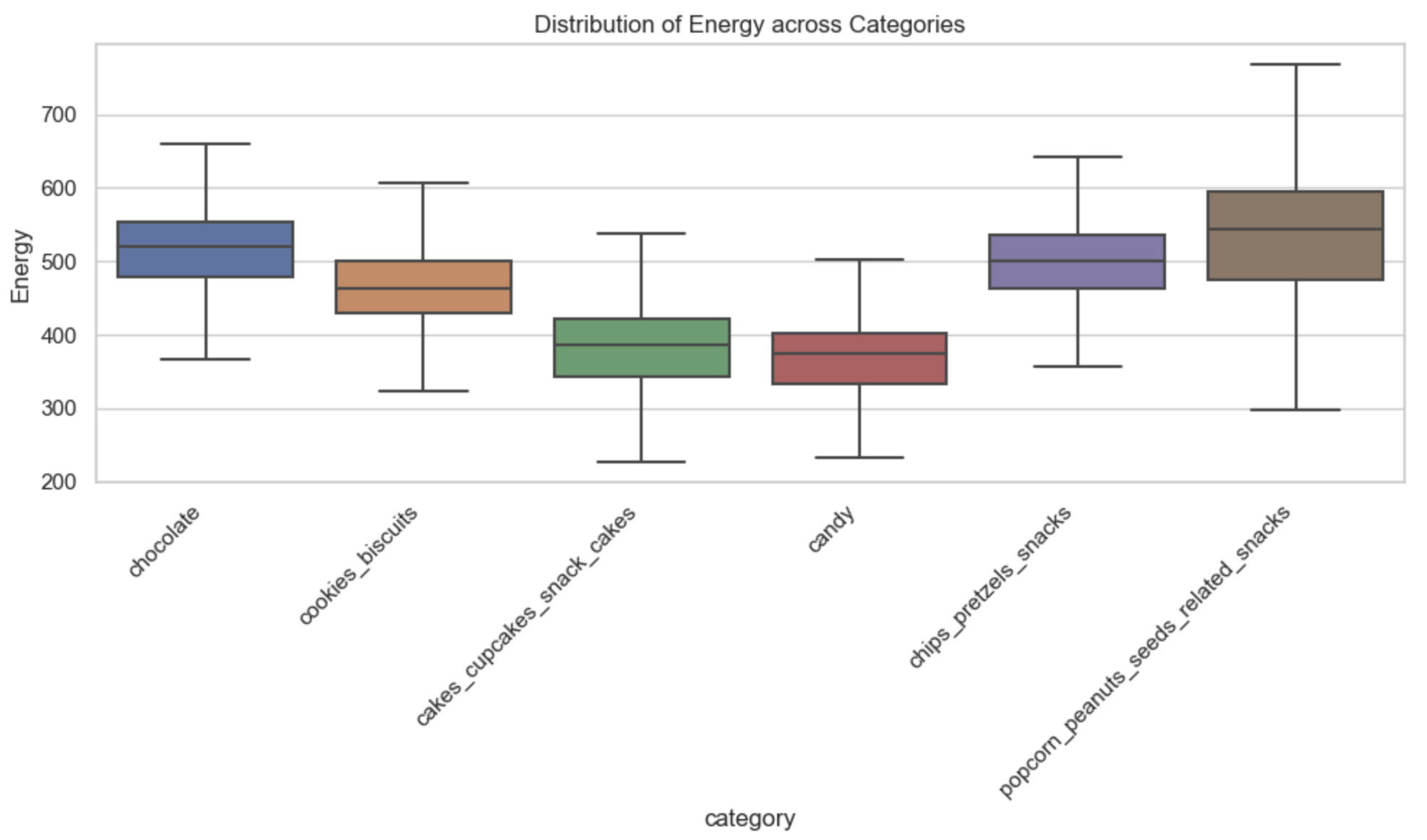


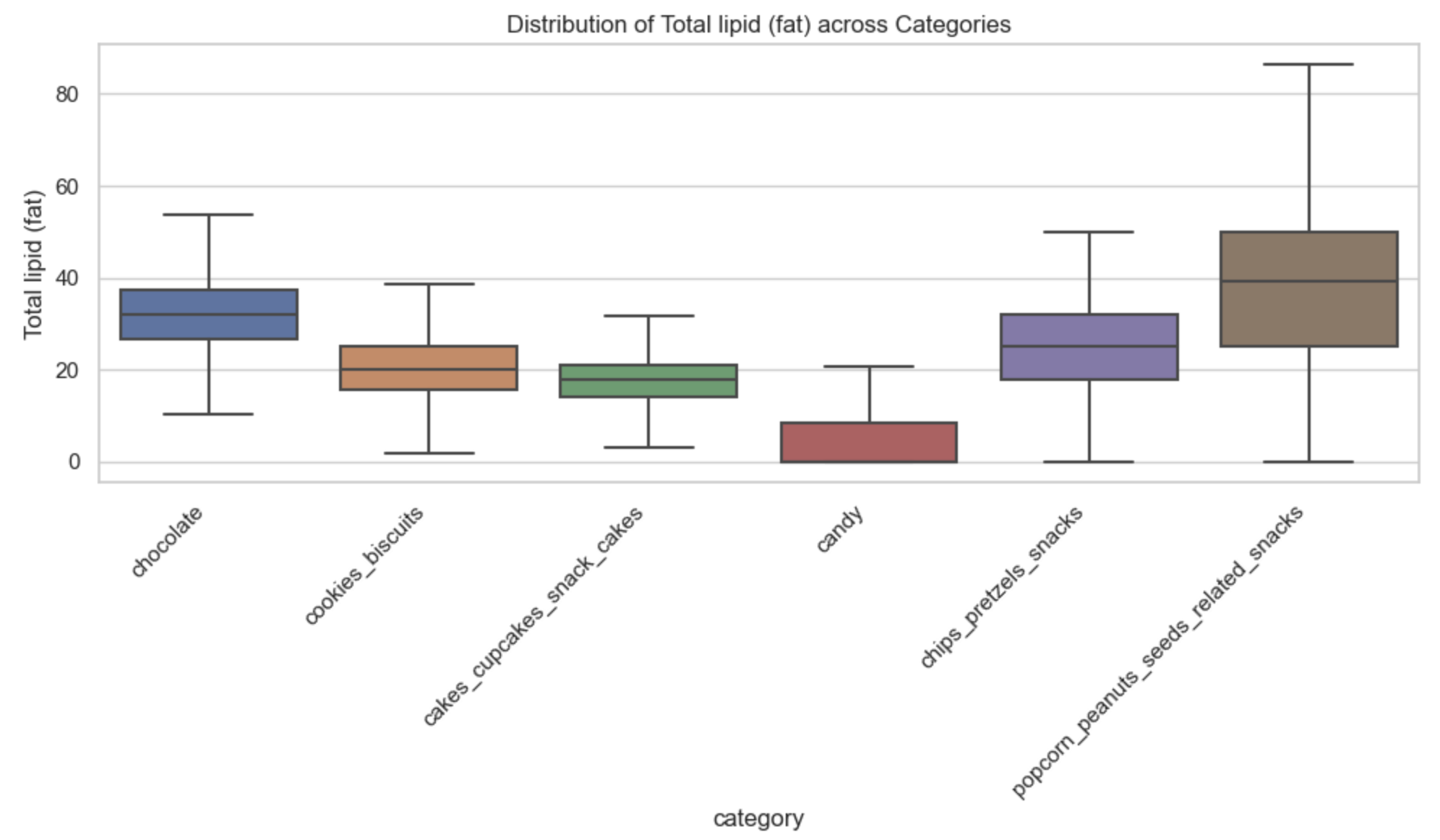
****

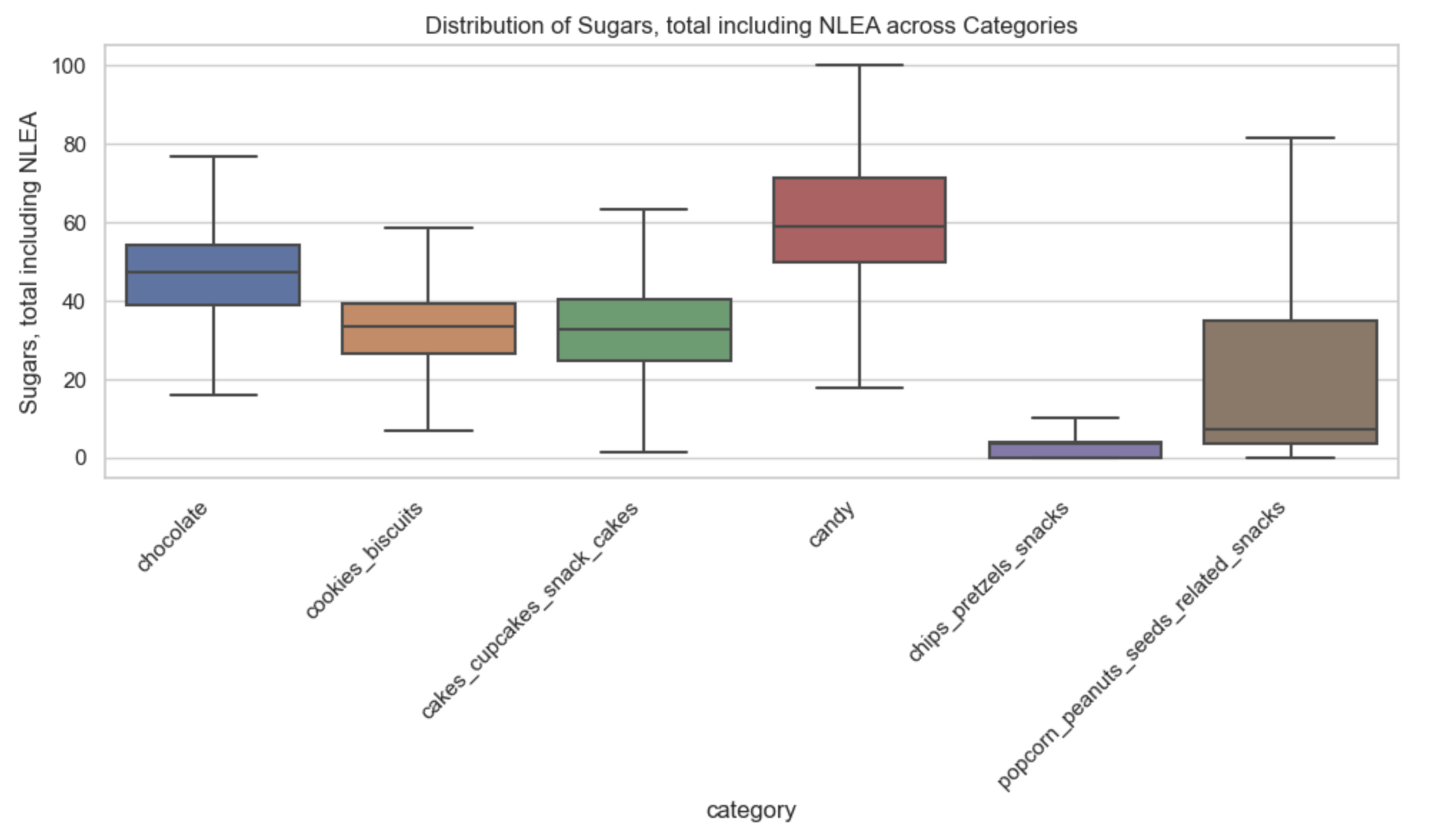
**Part 1 & 2 - Vitamins VS Calories, Carbs, Sugars and Fats**

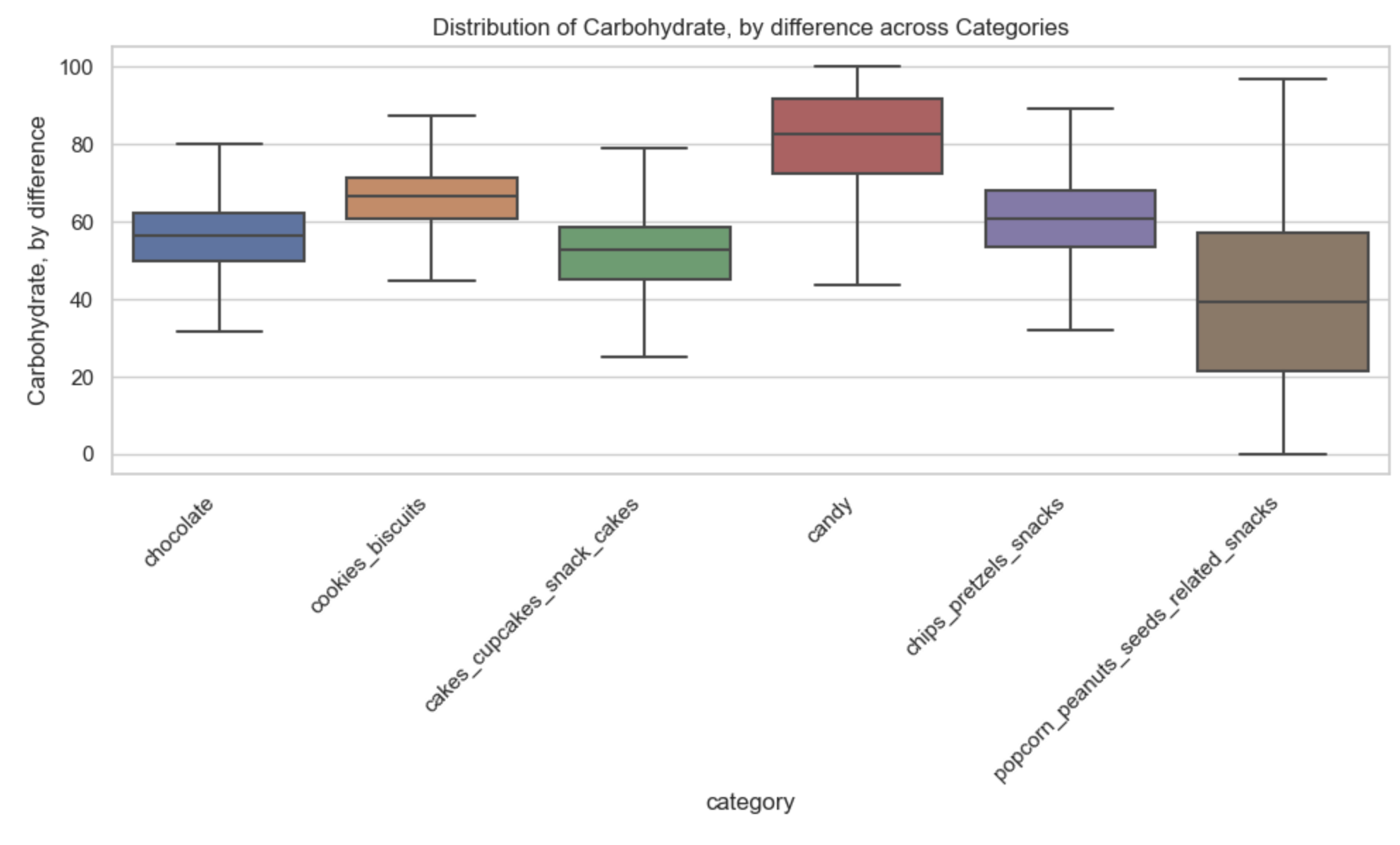
**Distribution of calories & fats & sugars & carbohydrates across Categories**











As you can see, snacks from Candy category have the lowest average sum of Energy and Fat but the highest average sum of Sugars and Carbohydrate. Moreover, snacks from Popcorn & Peanuts category have the highest average sum of Energe and Fat, but the lowest average sum of Sugars and Carbohydrates. In addition, Chips' snacks have the lowest average sum of Sugars.

**Vitamins score VS Unhealthy score**

We create 2 new columns:

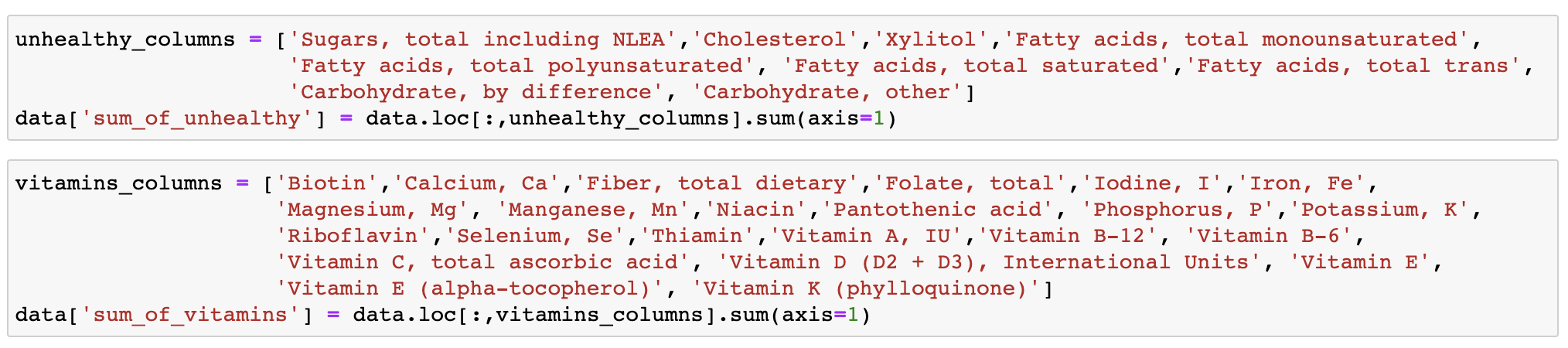
* *Vitamins score* - Contains the sum of all the amounts of vitamins found in the data for each snack.
* *Unhealthy score* - Contains the sum of all the amounts of Calories, Carbs, Sugars, Fats and unhealthy ingredients found in the data for each snack.

We are aware that there is a difference between the units of measure, but this doesn't affect the calculation of the score.

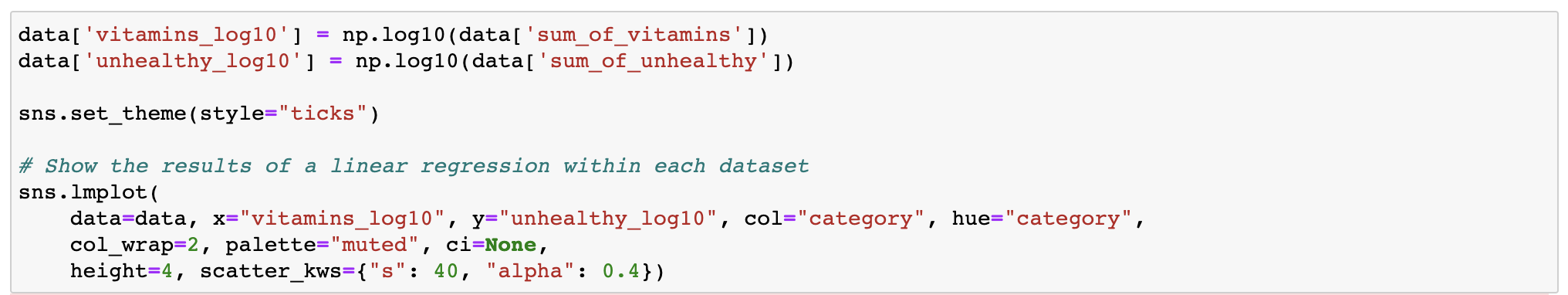
We will look for the top 5 snacks in each category that will have the highest vitamins score and the lowest unhealthy score in the same time. After that we wil focus on those top 5 snackes in each category and look over the color of the snack's wrapper.

We would like to choose the snackes with the most colorful wrapper, but we emphasize that the issue of health is of the utmost importance for us, therefore the color of the wrapper is only for the purpose of recommendation. Finally, we will make an informed choice of the 3 best snacks from each category.

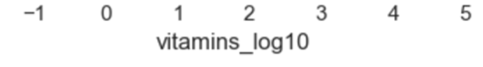
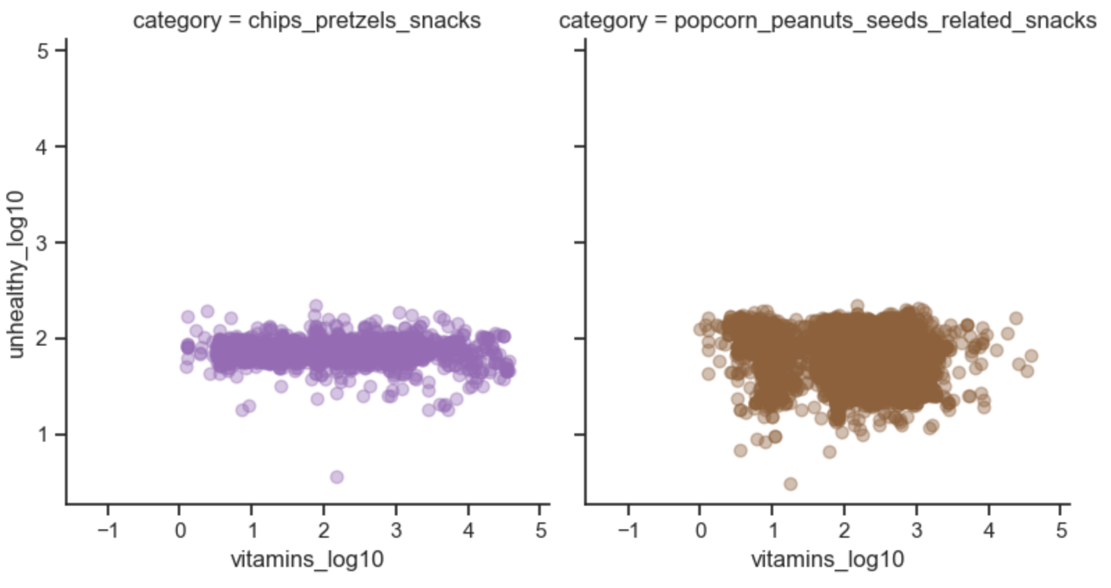
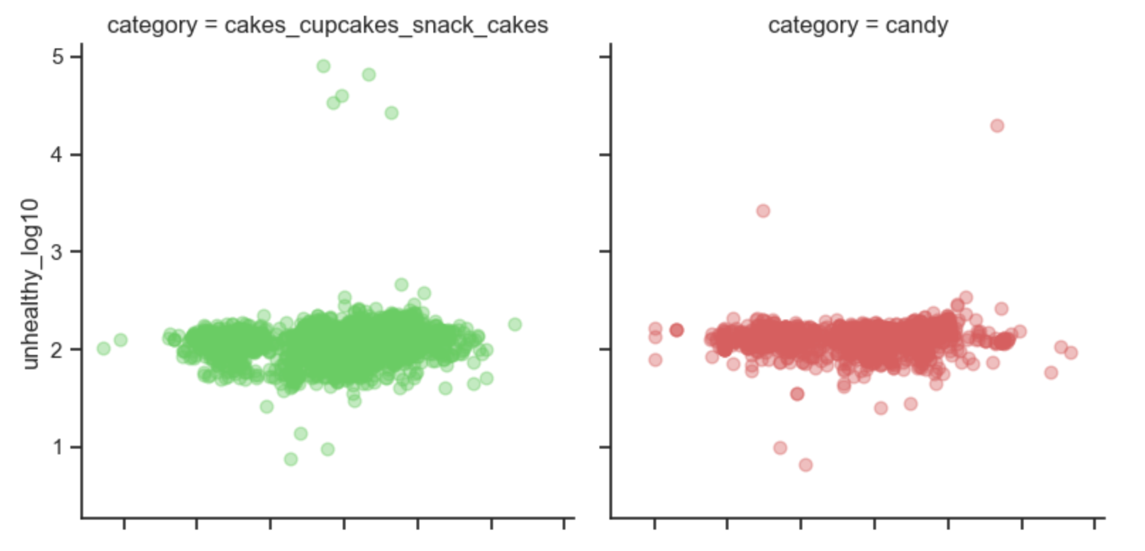
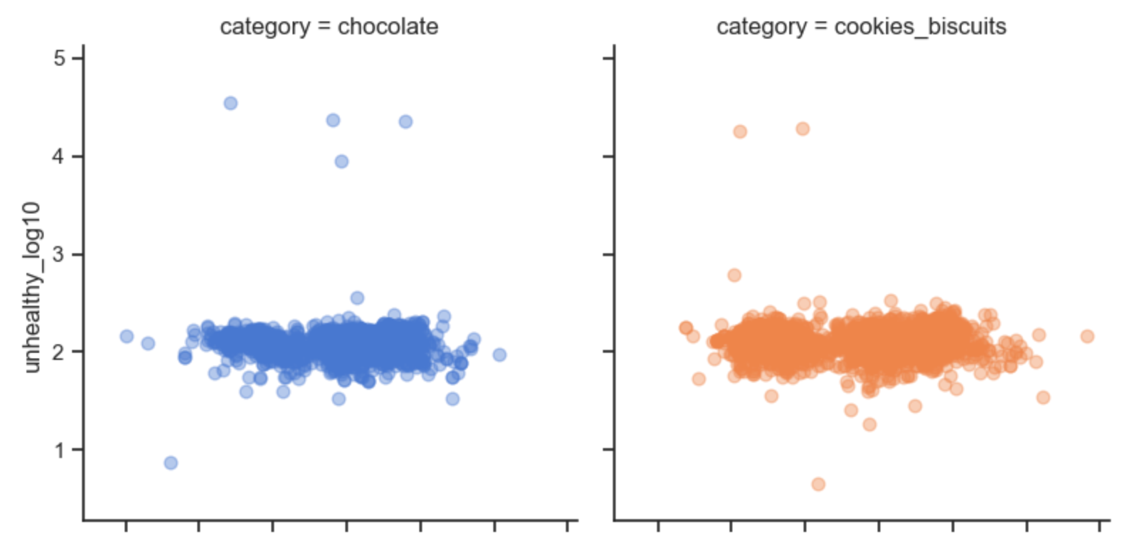
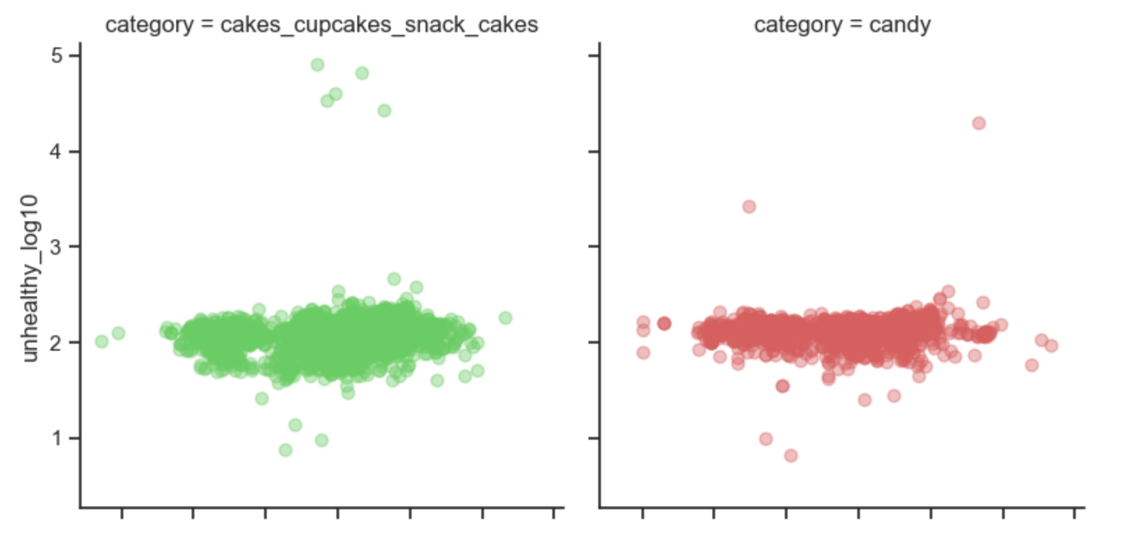
In the paper work we choose to show *one chosen category* for example. The information extraction work of the other categories is attached in the jupyter notebook "Exploratory Data Analysis".



**Scatter plot per category -** Using log transformation.

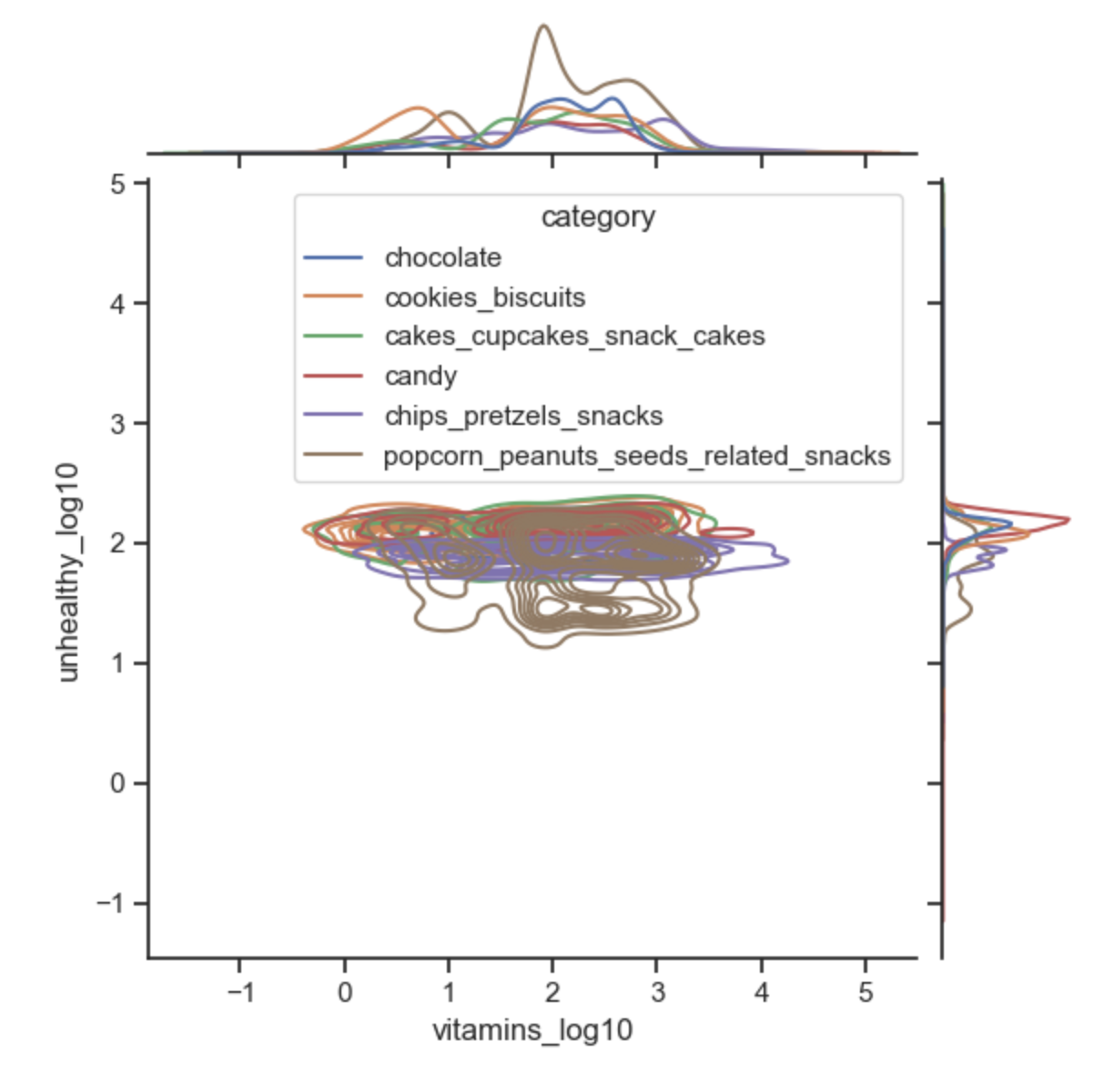


Add explain



## **Joint & Marginal** **Probability Distribution**

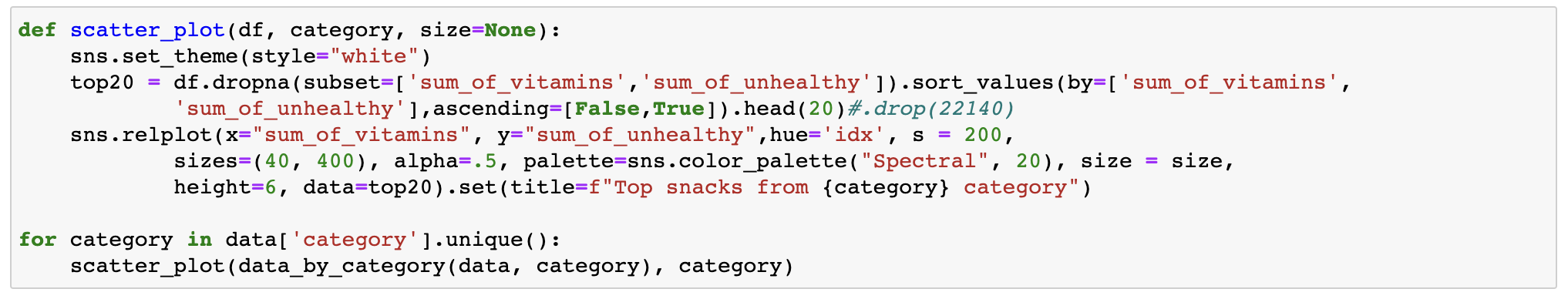


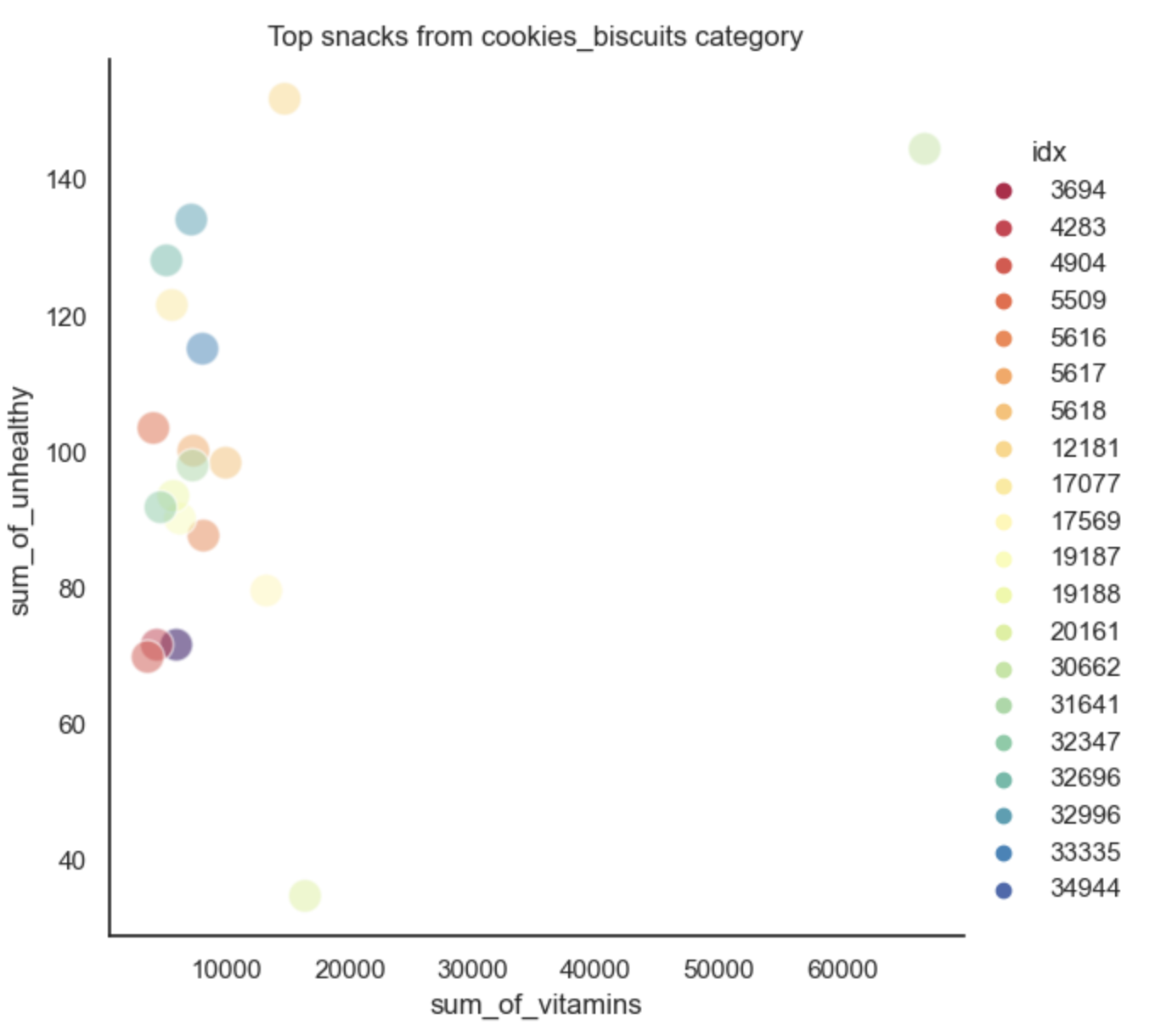


Add explain

**Scatter plot - Vitamins score VS Calories, Carbs, Sugars and Fats score**

The method:

* At firt, we removed rows with at least 1 Nan in 'sum\_of\_vitamins', 'sum\_of\_unhealthy' columns. (We prefer to look at snacks with all available information. We are aware that as a result our choice is reduced but we must be sure that the top product we recommend absolutely reflects the demands of the consumers).
* Sorting the values in each selected column and choose the top 10 indexes (min/max in accordance).
* Find the vitamins score and unhealthy score for each index (20 pairs in total).
* Create Scatter plot for the top 20 in each category. The vitamins score is in x axis, the unhealthy score is in y axis.
* We will focus on “Cookies” category. The results of the other categories are attached in "Exploratory Data Analysis" jupyter notebook.



### Selected Top 5 snacks

We prefer snacks that have more vitamins and are less harmful. Therefore, snacks that show in the right bottom area are the most recommended and snacks that show in the left top area are unrecommended.



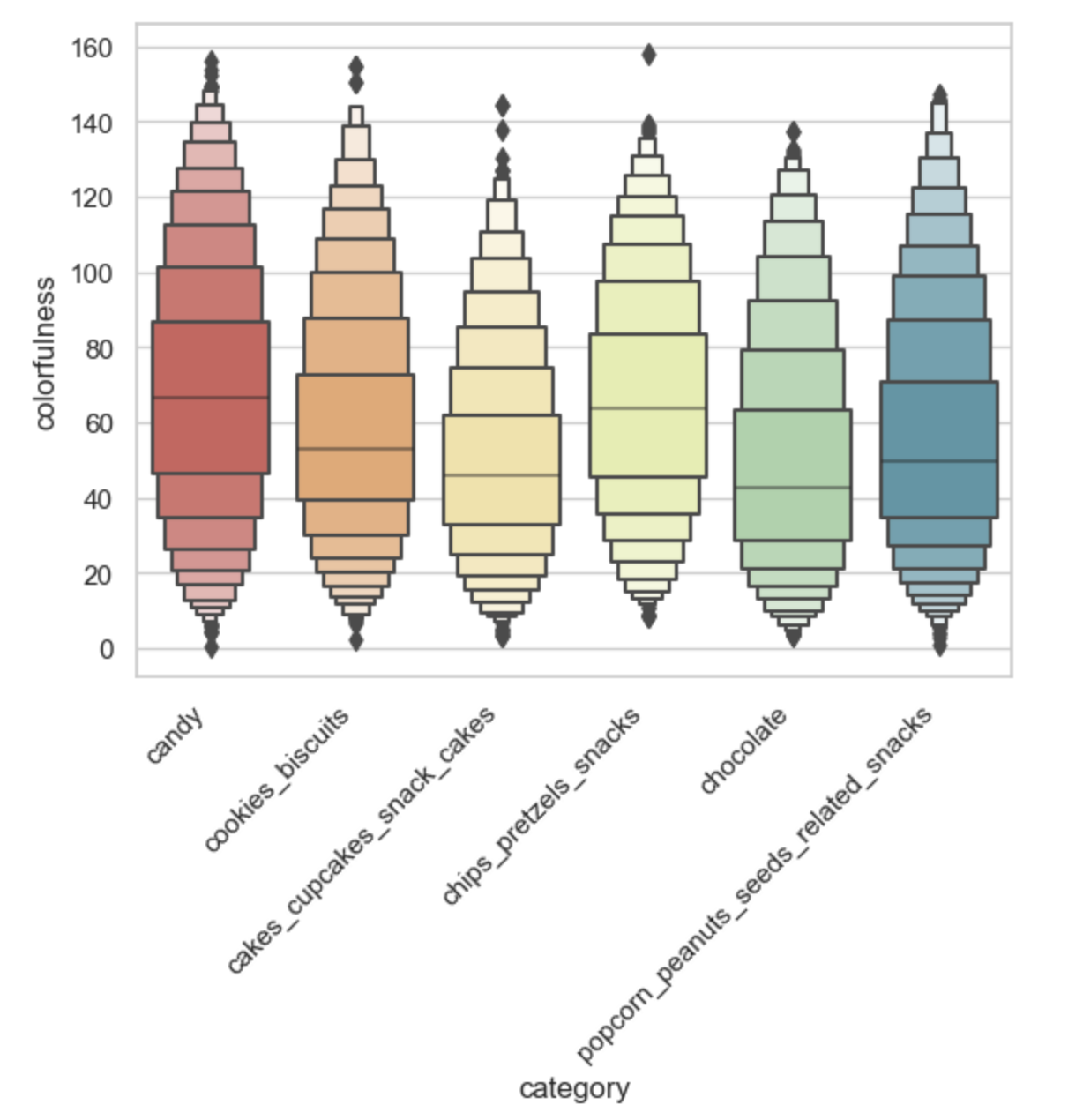
## **Part 3 - Colorful Wrapper**

As we said in the intreduction, besides supplying the least toxic food we shall also consider the colorfulnees of the wrap in order to makes the kids happy :)

There are many ways to define a "score" of colorfulness, we chose 'Hasler and Süsstrunk, 2003' way.

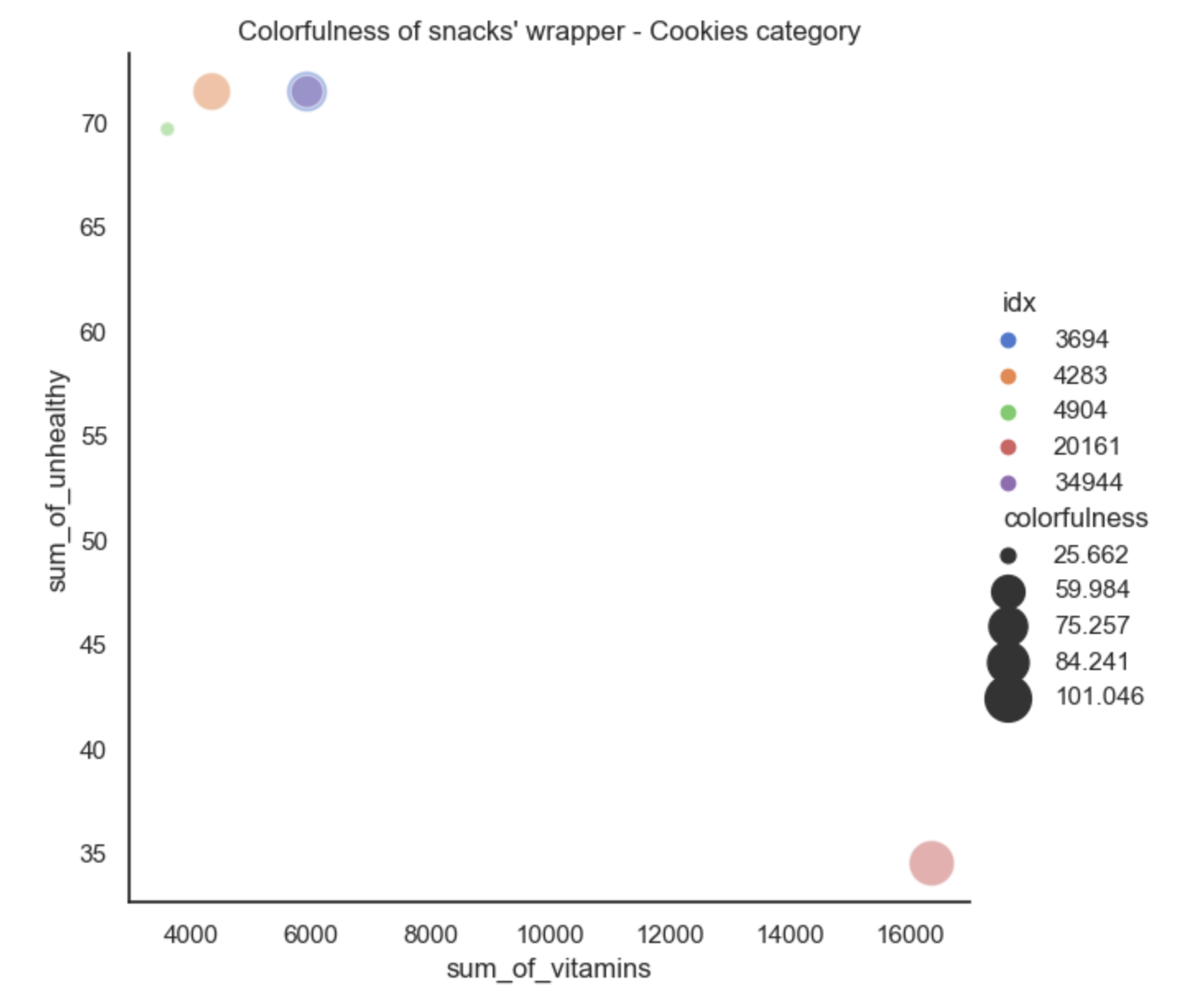






Its semms like canddies have the most colorful wrappers.  
To be honest, that's also what we guessed was happening:).  
  
**Scatter plot - Vitamins score VS Unhealthy score - The bigger the circle the more colorful snack wrapper**





After looking at the color of the wrapper, we used it to make an informed choice for the 3 most recommended snacks from each category, taking health considerations into account.

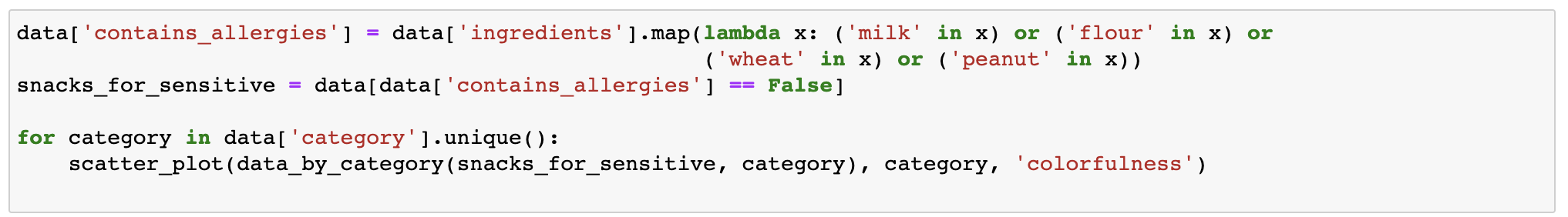


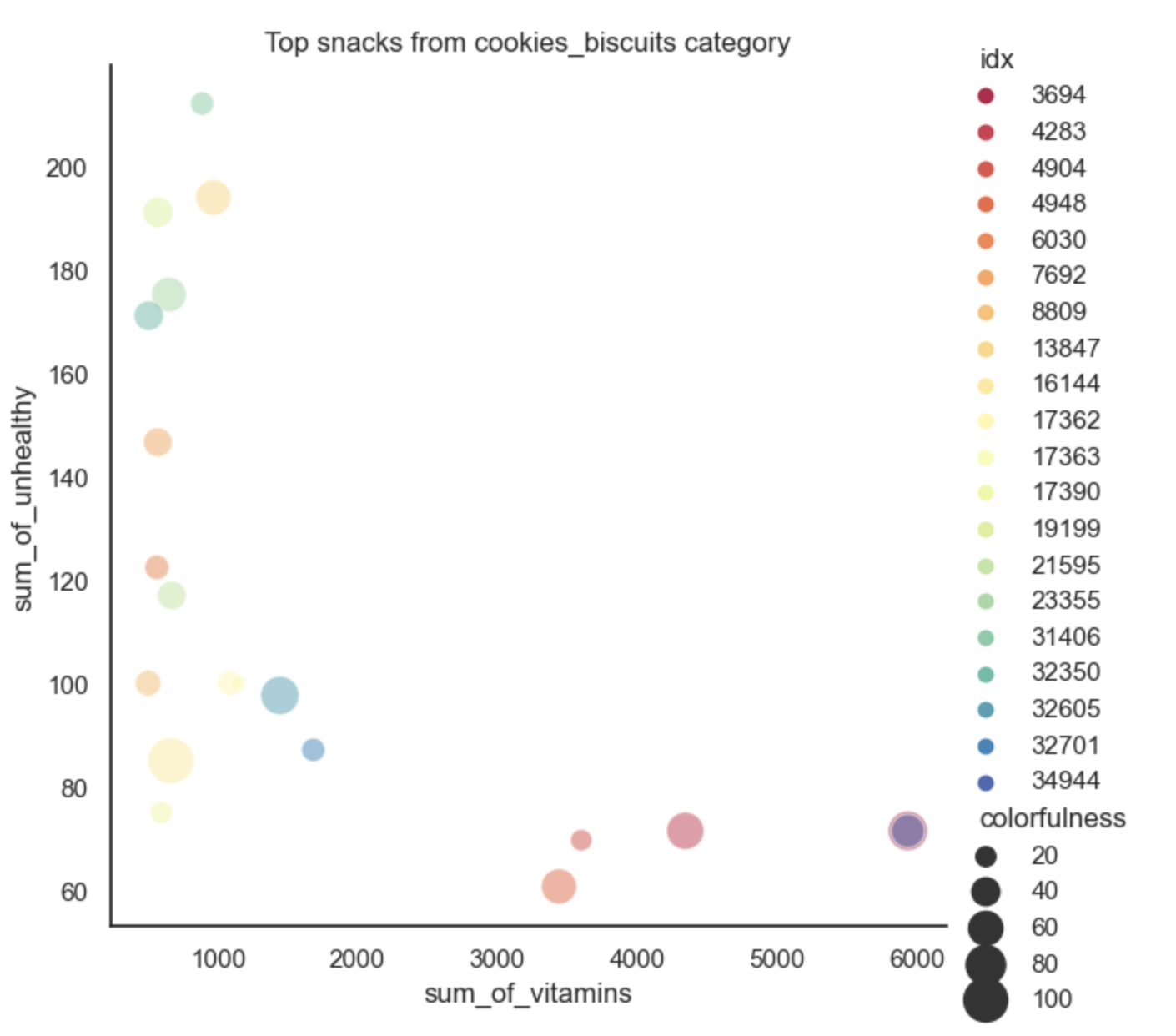
A box of oatmeal

Description automatically generated

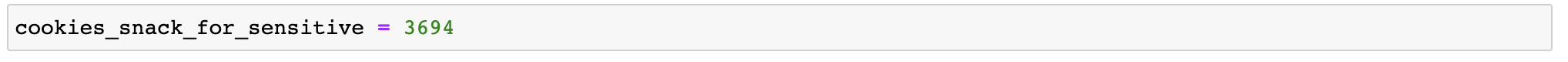


## **Part 4 - Sensitivity of Milk/Gluten/Peanut**





We have chosen from each category the most recommended snack taking sensitivity & health & colorfulness considerations into account.

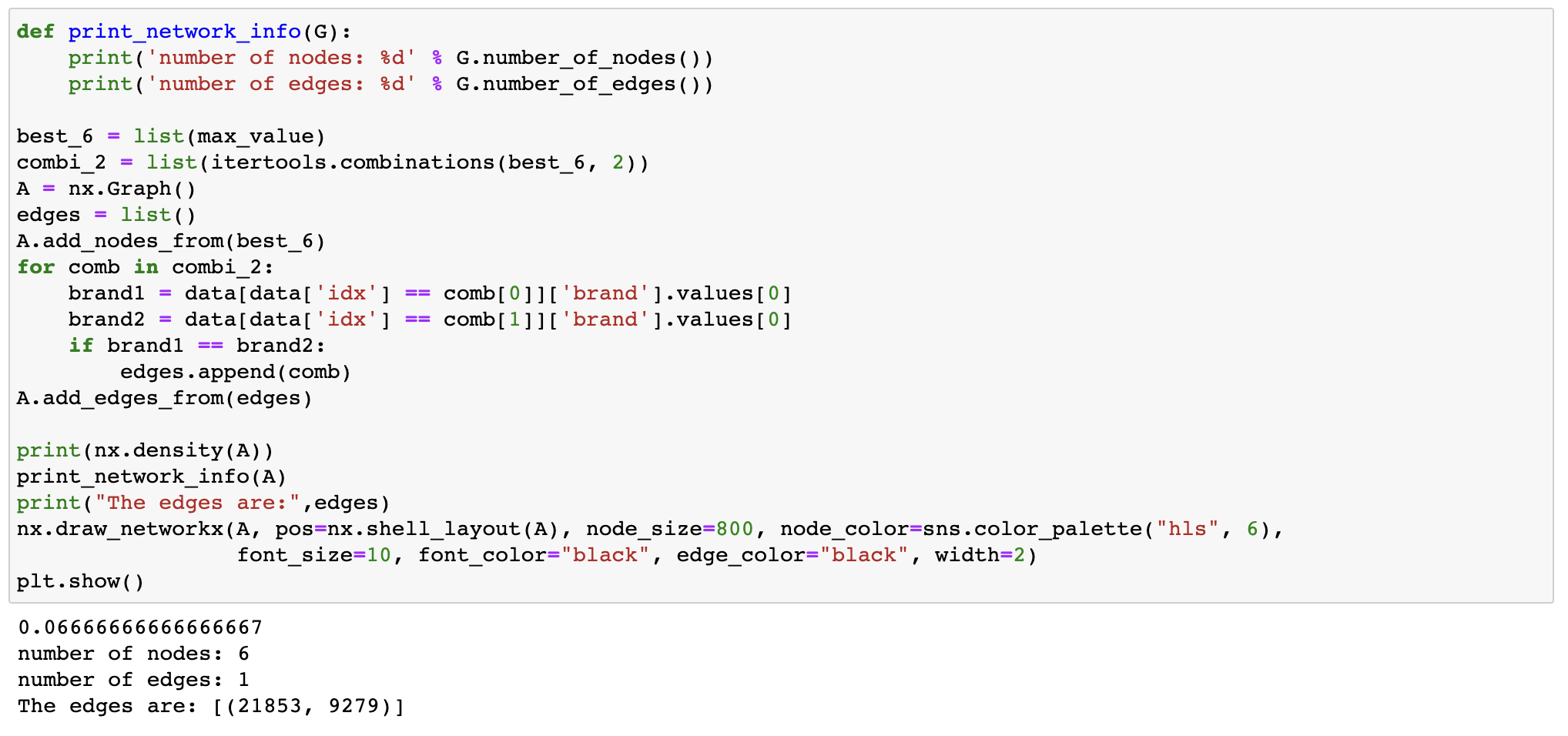




## **Part 5 - Products from the Same Brand**

#### Each idx is a node, Nodes are connected with edges if they are sharing the same brand , We are looking for the most connected(dense) graph

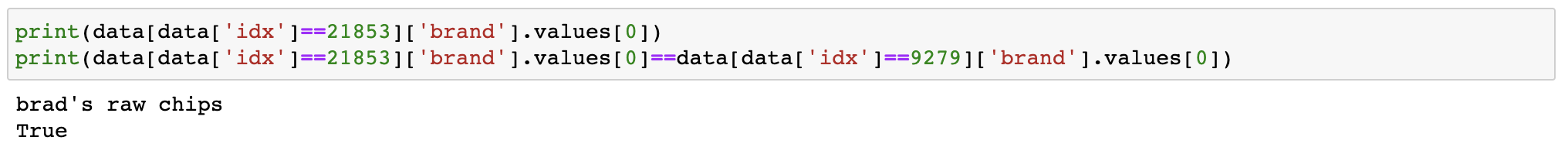


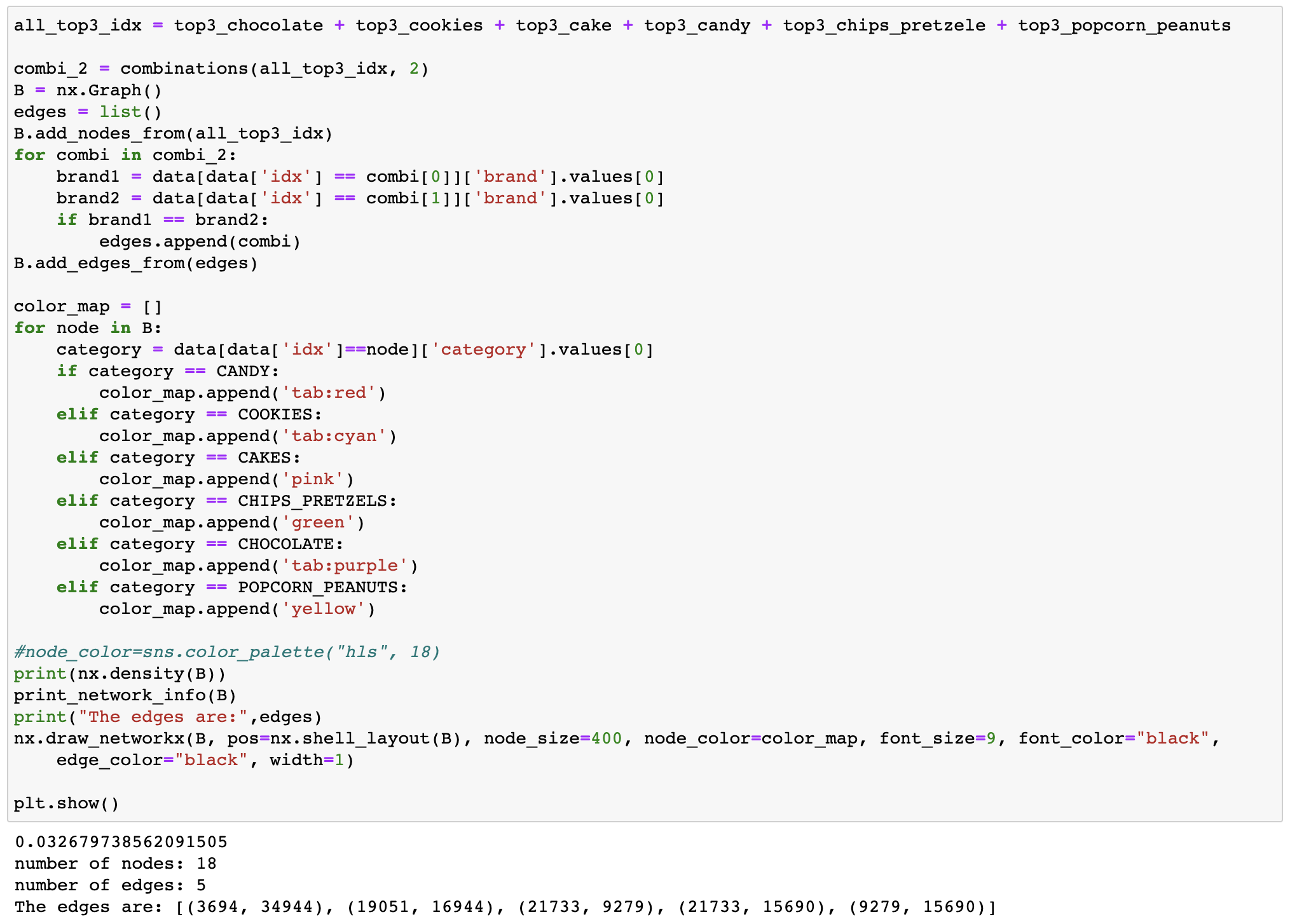


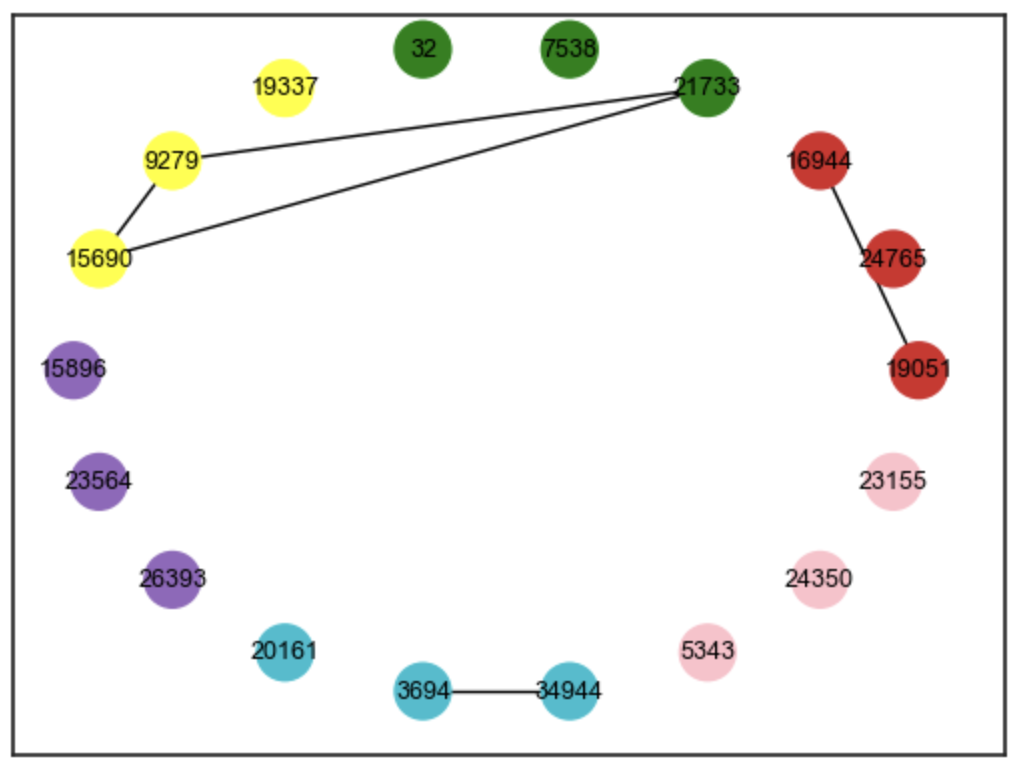




It seems that snack 21853 from chips & pretzels category and snack 9279 from popcorn & peanuts have the same brand!









# **Conclusions & Recommendations**

There are our final recommendations:  
**Chocolate** category: Top 3 recommended - **15896, 23564, 26393**, A snack for sensitive - **26393**  
**Cookies** category: Top 3 recommended - **20161, 3694, 34944**, A snack for sensitive - **3694**  
**Cake** category: Top 3 recommended - **5343, 24350, 23155**, A snack for sensitive - **17615**  
**Candy** category: Top 3 recommended - **19051, 24765, 16944**, A snack for sensitive - **24765**  
**Chips & Pretzele** ctegory: Top 3 recommended - **21733, 7538, 32**, A snack for sensitive - **7538**  
**Popcorn & Peanuts** ctegory: Top 3 recommended - **19337, 9279, 15690**, A snack for sensitive - **24319**

[Add explain on the connection graph]