# Spoonshot Data Science Internship Test 2022

Name: Abhiram S

Roll No.: CH19B037

## **Question 1**

The approach is as follows:

1. Use a pre-trained NLP model to create embeddings for each ingredient and the entire article.
2. Find cosine similarity scores of each embedded ingredient with the article.
3. The higher cosine similarity scores indicate the most relevant ingredients for the article.

The final ranking is was obtained to be:

['Munch Chocolate', 'Vegan Marshmallow', 'Vegan White Cake', 'Smokies Sausage', 'Vegan Puff Pastry Sheet', 'Vegan Chorizo', 'Vegan Carob Chips', 'Vegan Semisweet Chocolate Chips', 'Mulberry', 'Pork Chump Chops', 'Strawberry Gelatin', 'Vegan Chicken Strips', 'Vegetable Stock', 'Pork Lungs', 'Crunch Bars', 'Pickled Cauliflower', 'Muesli', 'Celery Soup', 'Snickers Spread', 'Cadbury Double Decker Chocolate Bar', 'Chia Meal', 'Ambrette Seed', 'Organic Maize', 'Hog Maw', 'Peanut Butter', 'Apple Cinnamon Granola', 'Cardamom', 'Mango', 'Organic Peruvian Groundcherry', 'Murukku Packet', 'Mccormick Montreal Steak Seasoning', 'Pork Tripe', 'Organic Tartar Cream', 'Americano Coffee', 'Campari Tomato', 'Orange Extract', 'SalmonTomato', 'Baby Abalone', 'Vinegar', 'Giardiniera', 'Tamarind', 'Arizona Seasoning']

We see that the vegan ingredients are most relevant, which makes sense since the article is about a vegan food delivery company.

## Question 2

Each element of the output array is a product of some of the elements of the input array. The elements of the input array are chosen according to combinations of its indices. The code can be generalised for longer arrays.

## Question 3

1. As an intern at Spoonshot, I wish to develop my skills in applied data science by working on real time projects at scale. I would also like to form connects with people in the company who would help me achieve my goals.
2. My motivation for pursuing a career in data science lies in the fact that there are so many problems that society faces today, which can be solved using a data driven approach. The amount of data generated every day is increasing exponentially. Since my interest also lies in the mathematical concepts of probability and statistics, along with being proficient at coding, it made sense to me to pursue data science.