# Module 12 Challenge - Activities and Report

## Introduction

The UK Food Standards Agency evaluates various establishments across the United Kingdom and gives them a food hygiene rating. You've been contracted by the editors of a food magazine, *Eat Safe, Love*, to evaluate some of the ratings data to help their journalists and food critics decide where to focus future articles.

## Part 1: Database and Jupyter Notebook Set Up

- I used NoSQL_setup_starter.ipynb for this section of the challenge.
- I Imported the data provided in the establishments.json file from my Terminal into the database uk_food and the collection establishments.
- I imported the libraries and dependencies needed for the work.
- I created an instance of Mongo Client.
- I confirmed that I had created the database and loaded the data correctly.
- I confirmed that I had set up my Jupyter notebook correctly.

## Part 2: Update the Database

- Prepared and loaded the data for the Penang Flavours Halal Restaurant.
- I found the Restaurant/Café/Canteen business ID = 1. Added this to the Penang Flavours Halal Restaurant document.
- Found and removed all the documents where the LocalAuthorityName was Dover.
- ***My final Jupyter Notebook is NoSQL_Starter_Working-Final.ipynb***

## Part 3: Exploratory Analysis

1. **Which establishments have a hygiene score equal to 20?**

   - For this part of the challenge, I used NoSQL_analysis_starter.ipynb
   - RatingValue refers to the overall rating the Food Authority decides and ranges from 1-5. The higher the value, the better the rating.
   - The scores for Hygiene, Structural, and ConfidenceInManagement work in reverse. This means, the higher the value, the worse the establishment is in these areas.
   - I counted the number of establishments and found 41 establishments with a hygiene score of 20.
   - I displayed the first document with a hygiene score of 20. It is a Care Business called the Chase Rest Home in Eastbourne. I converted the MongoDB data to a Pandas

DataFrame. I limited the display to 5 relevant columns. Note that the “scores” is a nested dictionary, and all three scores, hygiene, structural, and confidential management, are all displayed.

2. **Which establishments in London have a RatingValue greater than or equal to 4?**

- Used ‘$regex to find the full name of the London LocalAuthorityName.
- Found that 33 establishments with London as the Local Authority had a RatingValue greater than or equal to 4.
- Converted MongoDB to Pandas DataFrame. Output 10 rows. Limited the output to 5 columns.

3. **Finding Restaurants/Cafes/Canteens close to Penang Flavours Halal Restaurant with a Rating Value of 5, and low Hygiene scores.**

- I decided to only search for Restaurants/Cafes/Canteens in the search area. I assume it is a method of discovering the competition near the Penang Flavours Halal Restaurant.
- I built the query, constraining the search area, only allowing enterprises with a RatingValue of 5, and limiting results to those with low Hygiene scores.
- I created a projection showing which values to display.
- I displayed the results before creating the DataFrame. This displayed the projected values of 5 enterprises.
- I created DataFrame and displayed 10 rows. I limited the display to 5 relevant columns. Note that geocode and scores are both nested dictionaries. In each case, they hold multiple values.

4. **How many establishments in each Local Authority area have a hygiene score 0?**

- I found a list of all the establishments with a Hygiene score of 0 and created a list of the top ten counts. I displayed the list.
- I created a DataFrame from this list. I counted the total number of businesses with this score at 16,827. I displayed a list of 10 of the enterprises with scores of 0.