Case Assignment:
Food and Beverage Market Analysis

The datasets below captures some details about new food and beverage (F&B) product launches across the world in various years.

Datasets:

- *Product Launch Dataset.csv*
- *Flavor Classification Dataset.csv*
- *Positioning Category Mapping Dataset.csv*

Use these datasets for the following tasks.

1. Data Exploration: Data Handling and Formatting:

Review the "flavor" column in the *Product Launch Dataset*. Each product (as specified by a product ID) could come in combination of multiple flavors. If you eyeball some of entries in the column you will notice that there are multiple terms strung together into a list to describe the flavor combination. Each unique flavor used in a product is separated from another by a semicolon (;).

For instance, Product ID #10 carries the following list under the "flavor" column:

Cherry, not specified; Vanilla, not specified

You can infer that this product (#10) comes in combination of two unique flavors. The first flavor "Cherry, not specified" means the flavor is cherry but it is not mentioned what particular type of cherry flavor is used (e.g red or black). The second entry stands for vanilla but it is not mentioned what type of vanilla it is (e.g. French).

Likewise, Product ID #8 lists the following in the "flavor" column:

Apple, Red; Pineapple; Guava

This represents a combination of three unique flavors: red apple, pineapple and guava.

If you look carefully at the data you will also notice a double pipe separator (||) in some entries. These indicate product variants i.e. the same product launched with a new flavor. Think about how you would like to deal with these product variants in your analysis for the client. Would you treat them as different products altogether or just

ignore them? Or is there anything else would you would do? Consider the consequences of each option and justify your decision.

Can you use this interpretation of the "flavor" column to process the data and reveal some insights on the distribution of unique flavors based on following notes:

- Create a list of unique flavors based on your analysis. What is the total number of unique flavors in your list?
- Plot histogram of market subcategory against eventdate (years). Do any categories show negative trend over years?
- Which market subcategory has highest unique flavors?

2. Data Exploration: Merging and Manipulating Datasets

Review the dataset below:
Flavor Classification Dataset. csv

It assigns each unique flavor to a particular "Flavor Group" and gives you the mapping.

Could you use this Flavor Group data together with the Product Launch data and reveal some insights on the number of product launches over different quarters for "fruit" Flavor?

3. Data Formatting, Merging and Manipulating Datasets

Review the "positioning" column in the *Product Launch Dataset*. Each product (as specified by a product ID) could come in combination of multiple positioning categories.

The dataset "*Positioning Category Mapping Dataset.csv"* assigns each unique positioning category to a particular "Positioning Group" and gives you the mapping.

Could you use this Positioning Group data together with the Product Launch data and reveal some insights on the distribution for "*Convenience*" & "*Ethical Positioning*" Groups?

4. Data Aggregation

The "*eventdate*" column in the data represents the date on which the particular product was launched. Using this date information, generate monthly, quarterly and yearly summary statistics of product launches.

How would you package all these summaries and insights into an interactive excel tool for the client? You may want to use Excel Pivots with Slicers.

5. Data Exploration: Filtering and Subsetting

If you review the *Product Launch Dataset* you will notice that it covers product launches across the world for all products.

- One of your clients is interested in having the data from the Canadian market for the year 2013 about Energy drinks with ethical packaging. Help extract this data for that client.
- Identify TOP 5 unique flavors across countries in 2013?

6. Hypothesis Testing: TOP 5 Positioning Groups

The client is interested in determining TOP 5 popular positioning categories (groups) across countries in 2013. Define the Top 5 based on total product launches.

To add a statistical significance to your results, can you test out if each of them are significantly different. Say, the 1st one is significantly different from the 2nd one and so on.