**CASE STUDY**: Food and Beverages

1. Unique Flavor Analysis
   1. **Data Exploration: Data Handling and Formatting**

* **Scenario**: Launched products statistics based on the features given in dataset
* **Dataset:** “Product launch Dataset.csv” , “Flavor classification Dataset.csv”
* **Objective:** Preparing the data and identifying the suitable approach for analysis
* **Assumptions:**

**Aprroach1**: each “ || ” separated entry is a new variant , So for a product, multiple variants can be launched on same day , Considering those variants as absolute new product .

**Consequences:** A new variant can have different effects on market, target audience and led to different marketing strategies.

**eg**. If two variants are launched one in “Fruit” flavor group and other in “Alcohol” flavor group, both can have different target audience, market size etc. to look forward. Hence it is better to analyse considering as new product.

**Approach2**: each “ || ” separated entry is a new variant , consider all variants as single product.

**Consequences:** If the variants share same features like flavor group , From market standpoint there is not much significant difference in analysis while considering as different products or same products. But it will limit our analysis for target market categories and difficult to analyse product variant performances.

**Approach3**: Consider those variants as different product which has significant difference in their nature .

Eg: If variants of a product is of different flavor group, consider then as new product, otherwise keep as new flavor only. Two variants are of fruit flavor and one variant is of alcohol , so consider alcohol as new product and fruit flavors as one product with different flavours.

* *Approach 2 is implemented in the case study.*
* **Data Cleaning**:

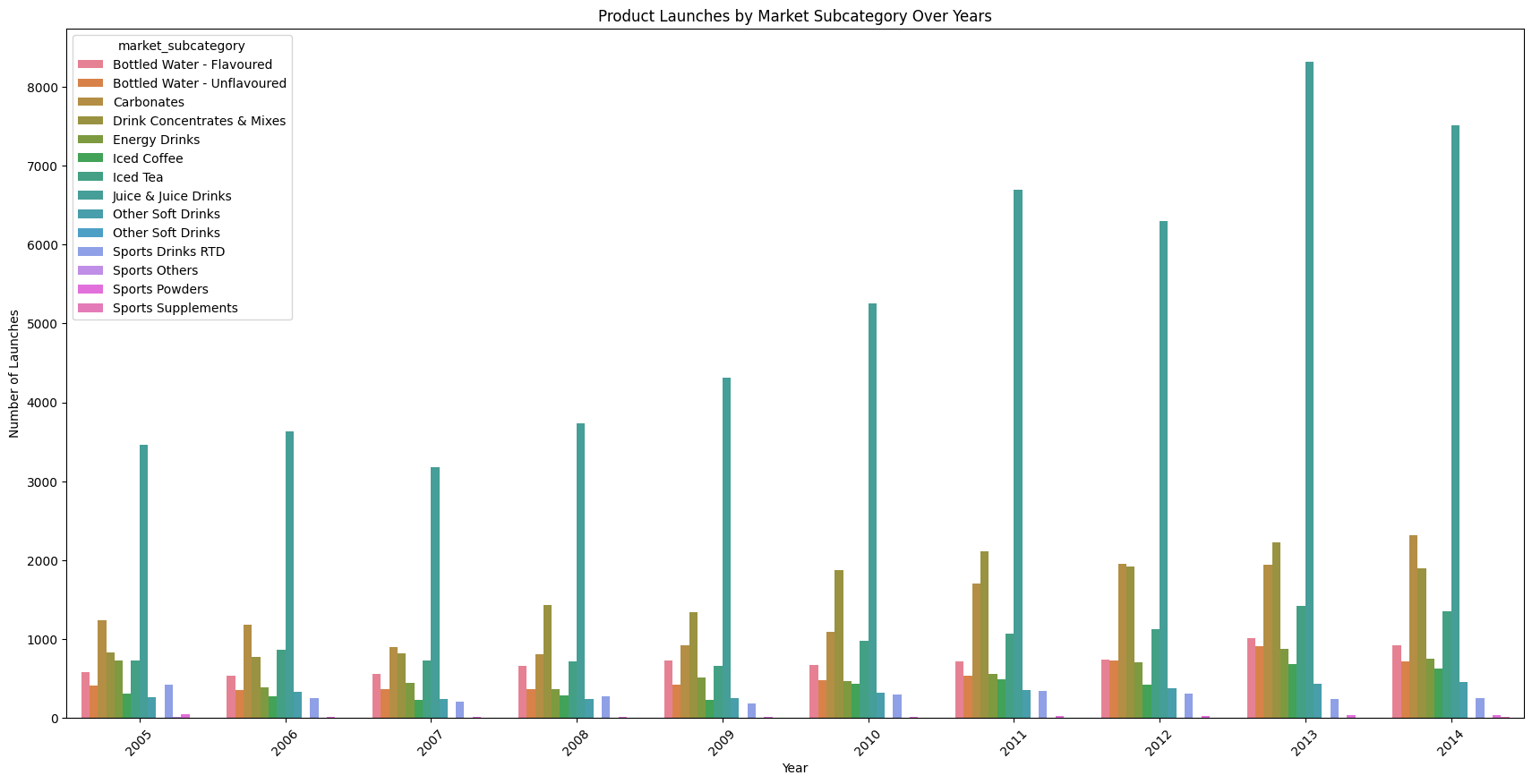
1. “Flavor” column has 1976 missing values , “Positioning” has 4321 missing values.
2. Question1 analysis is based on flavor ,hence if we remove positioning missing data it affects the count of unique flavors . So this step included removing of all missing flavor datapoints.

* **Data Preparation**:

1. “||” separated entries are exploded and assigned with new sub product ids.

Eg. Product ID #7 has 5 variants. Each variant is assigned new sub product id as 7.1, 7.2, 7.3 etc.

* 1. **Analysis**
* **Objective :** Create a list of unique flavors based on your analysis. What is the total number of unique flavors in your list?
* **Assumption:** Combination of two or more unique flavors considered as new unique flavor. New variant considered as new unique flavor.
* **Results:** Total unique flavors are **10881.**
  1. **Analysis**
* **Objective:** Plot histogram of market subcategory against eventdate (years). Do any categories show negative trend over years?
* **A graph with lines and numbers

  Description automatically generatedVisualization:**
* **Inference:**

**1**. Steady Decline in Certain Categories

* Categories like Carbonates and Iced Tea appear to show a general downward or stagnant trend over time, indicating a decrease or lack of growth in product launches.
* Over the years , product launches in "sports drink category" is less and decreasing or showing constant trend. From 2005 to 2009 - constant trend is shown. From 2009 to 2010 - number of launches has increased by only 100-200. From 2010 to 2014 - Negative trend is shown (Number of launches decreased or remain same)
* "carbonated drinks”: From 2005 to 2008 - lauches has decreased by 50-80 in comparison to each previous year, and from 2009 each year product launches increased.
* “Other Soft Drinks” also appears to be either stable or slightly declining, suggesting a lack of new launches in that area.

**2**.Saturation in Popular Categories

* Some subcategories, such as Bottled Water - Flavoured and Bottled Water - Unflavoured, show limited growth with some stagnation or minor declines over the years. This could indicate market saturation, where new product launches are not increasing as expected due to potentially high competition or a crowded market.

**3**. Low Activity and Decline in Niche Categories

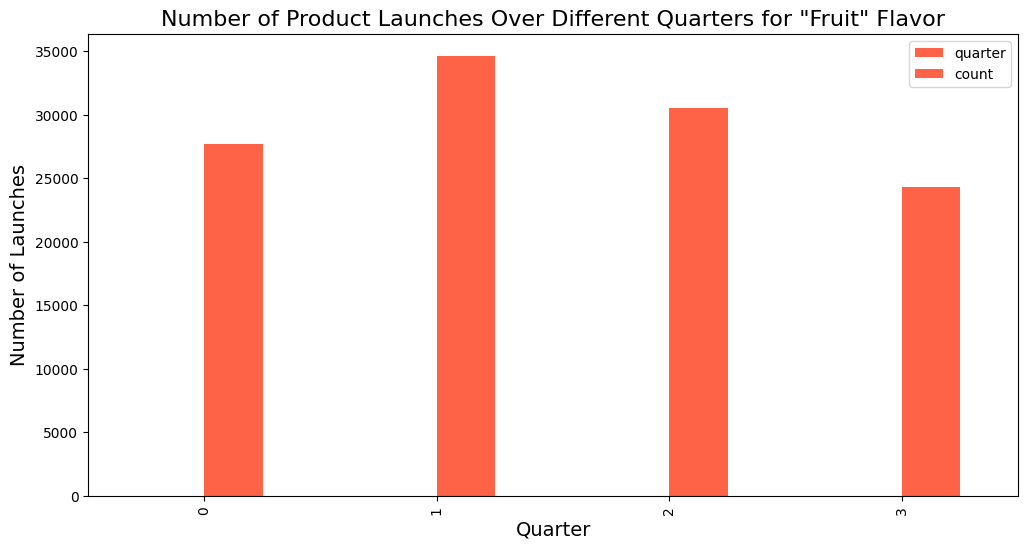
* Subcategories like Sports Powders and Sports Supplements show very low launch counts overall, with little to no growth. This could imply that these are niche markets with limited product innovation or launches.
  1. **Analysis:**
* **Objective:** Which market subcategory has highest unique flavors?
* **Result: “**Juice & Juice Drinks” with 5271 unique flavors.
* **Inference:**

1. High Consumer Demand for Variety and Customization

* The large number of unique flavors indicates that there is significant demand for variety in juice and juice drinks. Consumers may be seeking unique and tailored taste experiences, driving brands to diversify their flavor offerings.
* This trend aligns with a broader movement towards personalization in food and beverages, where consumers enjoy exploring new flavors or mixing multiple flavors to suit individual tastes.

2. Intense Competition Among Brands

* A high diversity in flavors often indicates intense competition in the market. Brands are likely trying to differentiate themselves from competitors by offering a wide range of flavors to attract various consumer segments.
* This could mean that new entrants or existing brands may need to continuously innovate with new flavors or flavor combinations to maintain market share.

1. Fruit Flavors Analysis
   1. **Data Exploration: Merging and Manipulating Datasets**
      * **Scenario**: Flavor statistics based on grouping all flavor categories together.
      * **Dataset:** “Product launch Dataset.csv” , “Flavor classification Dataset.csv”
      * **Objective:** 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.
      * **Data Preparation:**
2. Each product can be combination of multiple flavors so splitted and exploded each unique flavors for each product, based on “ ; ” delimeter.
3. For each launched product, identifying the “Flavor Group” mentioned in “flavor classification dataset” . Mapping and Merging the table using “flavor” feature.
4. Filtering the merged data for “fruit” flavor category specifically.
   * + **Visualization:**

1

4

3

2

* + - **Inference:**
* Q2 (second quarter) has the highest number of launches, followed closely by Q3. This suggests that a significant number of product launches for fruit-flavored products occur in the middle of the year.
* The peak in Q2 and Q3 likely aligns with spring and summer seasons in many regions, when consumers have a higher demand for refreshing, fruit-flavored products, particularly in beverages, snacks, and similar categories. This seasonal demand could drive companies to time their product launches accordingly.
* The lower launch numbers in Q4 (fall/winter) could reflect reduced consumer demand for fruit flavors in colder seasons or a shift towards flavors that are more popular in winter, such as spiced or warm flavors.

1. Positioning Categories and Subcategories Analysis
   1. **Data Formatting, Merging and Manipulating Datasets**

* **Scenario: “**Positioning Category” statistics of launched product where each product is combined with multiple position Subcategories.
* **Dataset:** “Positioning Category Mapping.csv”, “Product Launch Dataset.csv”
* **Objective:** use this Positioning Group data together with the Product Launch data and reveal some insights on the distribution for “Convenience” & “Ethical Positioning” Groups.
* **Data Preparation:**

1. Positioning of each launched product is combination of multiple positioning subcategories, so split each category by “,” delimeter and exploded the entries to different rows.
2. Merge the existing product launch data with positioning mapping data using the feature “positioning” which has one category only per product after doing step1.
3. Approach: There are overlapes and multiple counts of an product because each product may contain more than one positioning group so , we are counting how many products coming under each Positioning Group , does not mean each and every count of product implies a unique product.

**eg:** prodct id = 3 has three positioning Group (Juice , Convenience, Health(Passive)) so each group is having count of product 3, does not mean 3 different products we have.

* **Visualization:**
* **Inference:**
* ” Convenience” occupies the largest portion of the pie chart with 34% of total product launches. This high share suggests that a substantial portion of products are positioned to offer ease of use, quick preparation, or other convenience-related benefits. This trend might reflect consumer demand for time-saving products, such as ready-to-eat meals, portable snacks, or easy-to-prepare foods
* “Ethical” positioning constitutes **8%** of total product launches. This smaller, yet significant, segment indicates a focus on ethical considerations, such as sustainability, fair trade, or environmentally friendly packaging. It reflects a growing consumer awareness and preference for products that align with values such as social responsibility and environmental conservation.

1. Products launched over year, month, quarters.
   1. **Data Aggregation**

* **Objective:** Visualize the total products launched over year, months, quarters.
* **Dataset:** Product Launch Dataset.csv
* **Data Preparation:**

1. Pivot Tables, charts and slicers are used and count of each product id is taken as Y axis. To know how many products are launched over time.

* **Visualization:**

<product_launch_summary_final.xlsx>

* **Inference:**

1. From 2008 to 2014 , number of product launches are increased showing that market is growing and demands are increasing so company brands should consider launching new products for increasing revenue.
2. For every year , Quarter 2 and quarter 3 time is the peak time when most products get launched , hence brands should consider this insight for their next launching plan.
3. **Filtered Data As per client request**
   1. **Data Exploration: Filtering and Subsetting**

* **Objective:** Client is interested in having the data from the Canadian market for the year 2013 about Energy drinks with ethical packaging.
* **Dataset:** “Product launch Datset.csv”
* **A screenshot of a computer

  Description automatically generatedResults:**
* **Inference:** Only 4 products are launched in this filtering **.** As per Client’s request further analysis can be done in these data points.
  1. **Data Exploration: Filtering and Subsetting**
* **Objective:** Identify TOP 5 unique flavors across countries in 2013.
* **Dataset:** Product Launch Dataset.csv
* **Result:** The top 5 unique flavors across countries in 2013: ['orange, not specified', 'unflavored', 'apple, red', 'fruit, not specified', 'lemon'].
* **Inference:** In 2013, top 5 flavors launched are all in fruit category or no flavor , means fruit demand is increasing , and by time series analysis if the same trend is shown for next years also , we can suggest that lauching product in fruit category can increase the revenue as demand is high.

1. Top 5 Positioning Categories and Their significance
   1. **Hypothesis Testing: TOP 5 Positioning Groups**

* **Objective:** 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
* **Dataset:**
* **Result:** Positioning Group, Convenience-12986, Health (Passive)-9858, Ethical-5981, Choice-2958, Health (Active)- 2648
* **Inference:**

1. we can suggest that lauching product in these positioning category can increase the revenue as demand is high.
2. We can infer that these categories already have the high product supply chain, hence for starting, trying to launch in different niche categories can led to big success for business.
   1. **Hypothesis Testing: TOP 5 Positioning Groups**
      * **Objective:** 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.
      * **Hypothesis:** H0 = Significantly all categories are same. H1 = Significantly different categories.
      * **Testing Results:**
3. Chi square goodness-of-fit test is performed on top 5 categories where expected count per category = 6886.2.
4. P value ~ 0 , suggest that strongly reject null hypothesis.

* **Inference:**

1. **Significance of Differences:** Since the p-value is 0, this suggests that the differences in product launches among these five categories are not due to random chance. At least one category has a significantly higher or lower number of launches than would be expected if the launches were evenly distributed.
2. **Insight for Each Category:**

* Convenience has the highest number of launches (12,986), which is significantly above the expected count (6,886.2).
* Health (Passive), with 9,858 launches, also has a higher count than expected.
* Ethical and Choice categories have fewer launches than expected, with counts of 5,981 and 2,958, respectively.
* Health (Active) has the lowest count at 2,648, far below the expected count.
* Convenience and Health (Passive) are likely high-priority areas in terms of consumer demand or market focus, as evidenced by their large number of launches.
* The Ethical and Health (Active) categories might represent niche markets with less frequent product launches, suggesting limited consumer focus or strategic allocation in these areas.