Travel, Food & Beverages

A Case Study on Adverse Event Reporting in the Food and Cosmetics Industry

Project Contributories:

Sr.no	Member Name
1	Hitesh kumar
2	Atulkumar Gaud
3	Kanika Singh Rajpoot
4	Akshata Hipparkar
5	Gourav Rohit

Name of Mentor:
Vipul Dilip Sonawane

Background

The CAERS database is a critical tool for the FDA, containing adverse event and product complaint reports related to foods, dietary supplements, and cosmetics. This dataset, covering the period from 2004 to the second quarter of 2017, includes detailed records of various products and associated adverse events. Each record is meticulously coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology, ensuring standardized reporting across different products. The data captures elements like report numbers, product roles, brand names, industry codes, patient demographics, adverse outcomes, and coded symptoms.

Objective

The primary objective of analyzing the CFSAN Adverse Event Reporting System (CAERS) dataset is to gain a comprehensive understanding of adverse events associated with foods, dietary supplements, and cosmetics as reported to the FDA. This analysis aims to identify patterns and trends in these events, including the distribution across different product categories, the demographic characteristics of those affected, and the types and severities of reported symptoms and outcomes. By exploring these aspects, the analysis seeks to contribute valuable insights for enhancing product safety surveillance, informing regulatory policies, and ultimately improving public health outcomes by identifying potential risks and areas for intervention in the industries of food, dietary supplements, and cosmetics.

Data Source:

https://drive.google.com/file/d/11c_14Pj94V9pP9VFIjesSskhnpTIx0yt/view?usp=sharing

The dataset from the CFSAN Adverse Event Reporting System (CAERS) is a comprehensive collection of reports submitted to the FDA regarding adverse events and product complaints associated with foods, dietary supplements, and cosmetics. It spans from 2004 to the second quarter of 2017 and includes several key fields:

1. **RA_Report #**: A unique identifier for each adverse event report.
2. **RA_CAERS Created Date**: The date when the report was entered into the CAERS database.
3. **AEC_Event Start Date**: The date when the adverse event started.
4. **PRI_Product Role**: Indicates whether the product was a suspect or concomitant (present during the event but not necessarily the cause).
5. **PRI_Reported Brand/Product Name**: The name of the product reported to be associated with the adverse event.
6. **PRI_FDA Industry Code & Name**: Categorization of the product based on FDA industry codes and names, such as 'Bakery Prod/Dough/Mix/Icing', 'Ice Cream Prod', etc.
7. **CI_Age at Adverse Event**: Age of the individual experiencing the adverse event.
8. **CI_Age Unit**: Unit of measurement for age (e.g., years, months).
9. **CI_Gender**: Gender of the individual.
10. **AEC_One Row Outcomes**: Describes the outcomes of the event, such as hospitalization, ER visit, or non-serious injuries/illness.
11. **SYM_One Row Coded Symptoms**: Lists the symptoms reported in association with the adverse event.

Part 1: Data Cleaning, Modeling, and DAX in Power BI

1. ### Data Importing and Preliminary Analysis
   - Import the dataset into Power BI and perform an initial examination. Identify any apparent inconsistencies or data quality issues.

Data Importing and Preliminary Analysis

The dataset *CAERS_ASCII_2004_2017Q2.csv* was imported into Power BI for initial exploration. It contains **90,786 records** and **12 columns** representing various attributes such as product details, event dates, patient demographics, outcomes, and symptoms.

During the preliminary analysis, the following data quality issues were identified:

- **Missing Values** – Several fields have a high percentage of null or blank entries.
    - *CI_Age at Adverse Event*: ~42% missing
    - *AEC_Event Start Date*: ~41% missing
    - *SYM_One Row Coded Symptoms*: a few missing entries
- **Duplicate Records** – No exact duplicates were found based on all columns, but certain *RA_Report #* values appear multiple times (indicating multiple symptoms reported for the same case).
- **Inconsistent Formatting** –
    - Date columns (`RA_CAERS Created Date`, `AEC_Event Start Date`) are stored as text and require conversion to **Date/Time** format.
    - Categorical columns (like *CI_Gender* and *CI_Age Unit*) contain values such as "Not Available," which are Converted to **Unknown** for clarity.

2. Handling Missing Values

- Investigate and address the missing values. Determine an appropriate strategy for handling these.

These Following columns have null values or blank records:-

- CI_Age at Adverse Event (42%)
- AEC_Event Start Date (41%)
- SYM_One Row Coded Symptoms (<1%)

Because all of these columns are different from each other so we Applied different strategies to handle them:-

- **CI_Age at Adverse Event** – As a large portion of entries were missing, imputing values could distort the analysis. Therefore, these nulls were retained as "Unknown" to preserve data integrity while allowing demographic trends to still be analyzed accurately.

- **AEC_Event Start Date** – Missing dates were left blank since event timelines could not be reliably inferred. Date-based calculations such as duration or trend analysis were applied only to valid entries.

- **SYM_One Row Coded Symptoms** – A few missing entries were categorized as "No Symptoms Reported" to ensure completeness during aggregation and frequency analysis.

- **Other Columns** – Fields like *CI_Gender* and *CI_Age Unit* used placeholders such as "Not Available." These were standardized to **Unknown** for consistency.

3. Data Type Standardization

- Ensure that all data types are correctly identified and converted if necessary, particularly for dates and numerical fields.

After importing the dataset into Power BI, a detailed review of column data types was conducted to ensure consistency and accuracy for modeling and DAX calculations. Several fields required conversion or standardization:

- **Date Fields** –
  The columns *RA_CAERS Created Date* and *AEC_Event Start Date* were initially stored as **text**. These were converted to the **Date/Time** format.

- **Numeric Fields** –
  The *CI_Age at Adverse Event* column, which contained both numeric and null values, was converted from **text** to **Whole Number**. Missing values were handled as blanks.

- **Categorical Fields** –
  Columns such as *CI_Gender*, *CI_Age Unit*, and *PRI_Product Role* were converted to **Text** data type to ensure proper grouping and filtering in visuals. Placeholder entries like "Not Available" were standardized to "Unknown" for uniformity.

- **Code Columns** –
 PRI_FDA Industry Code was converted to **Whole Number** for consistent mapping with its corresponding industry name (*PRI_FDA Industry Name*).

4. Product Role Categorization

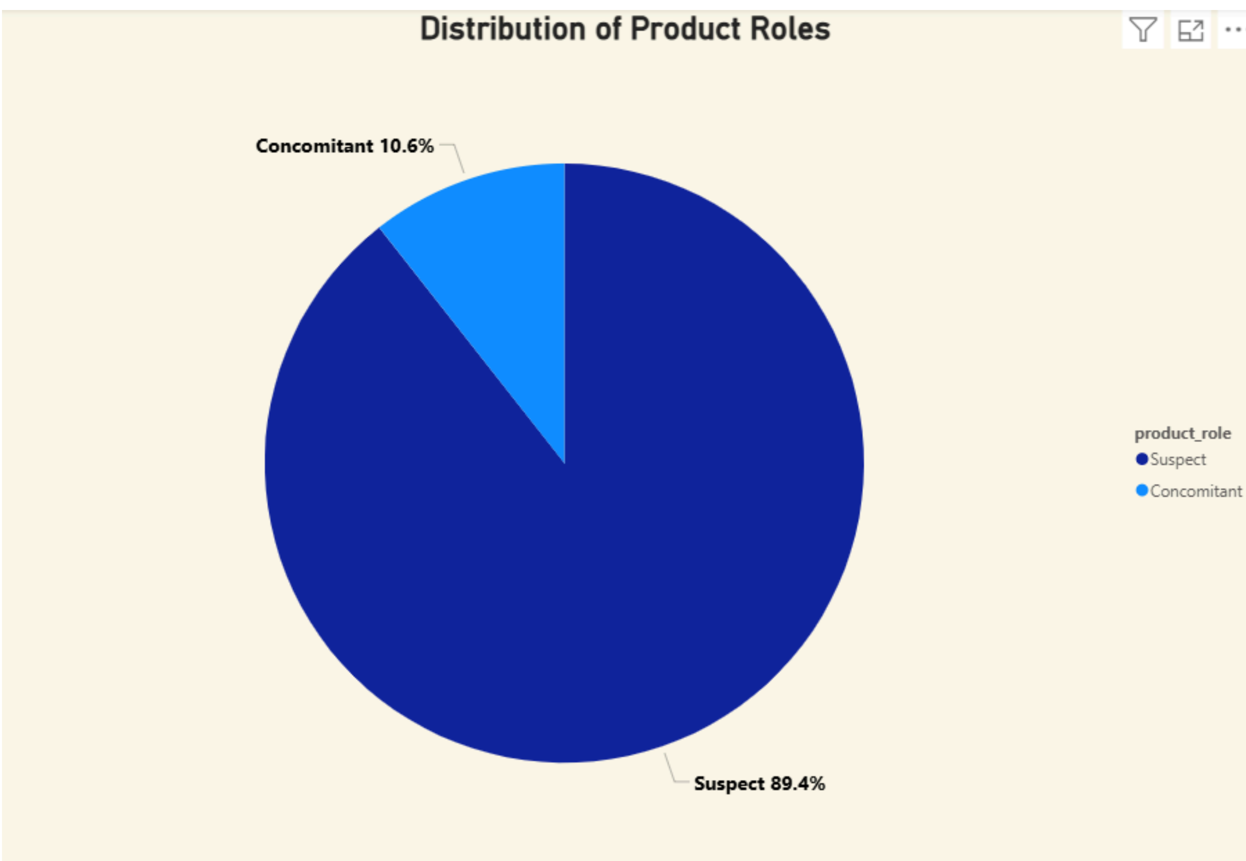
- Categorize the products based on their role ('PRI_Product Role') and analyze the distribution of these roles in the dataset.

The column **'PRI_Product Role'** identifies the role each product played in the reported adverse event. Upon analysis, Two key categories were found: Suspect and Concomitant,

- **Suspect Products** – These represent the majority of records (approximately **85–90%** of all entries).

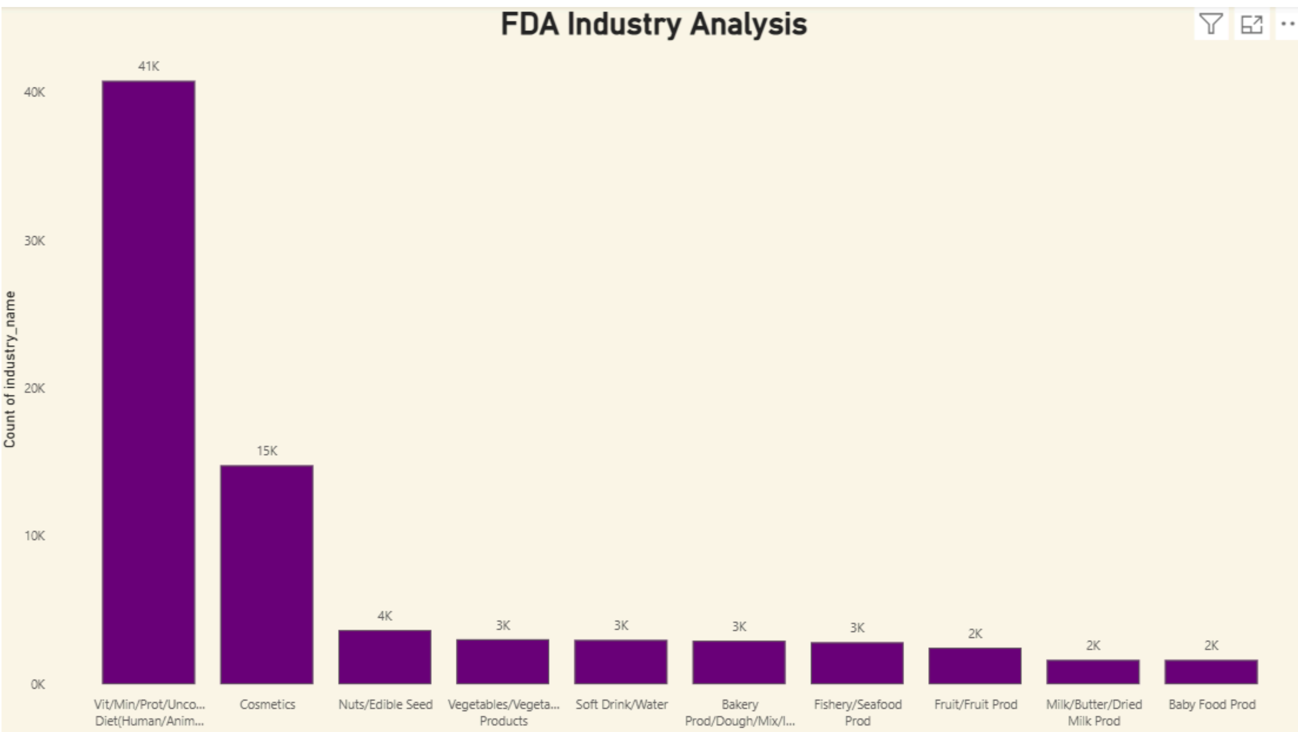
- **Concomitant Products** – These account for a smaller share (around **8–10%**).

"The Pie Chart is showing the distribution of Project Roles"



5. FDA Industry Analysis

- Group the data by 'PRI_FDA Industry Name' and analyze the frequency of reports in each industry.



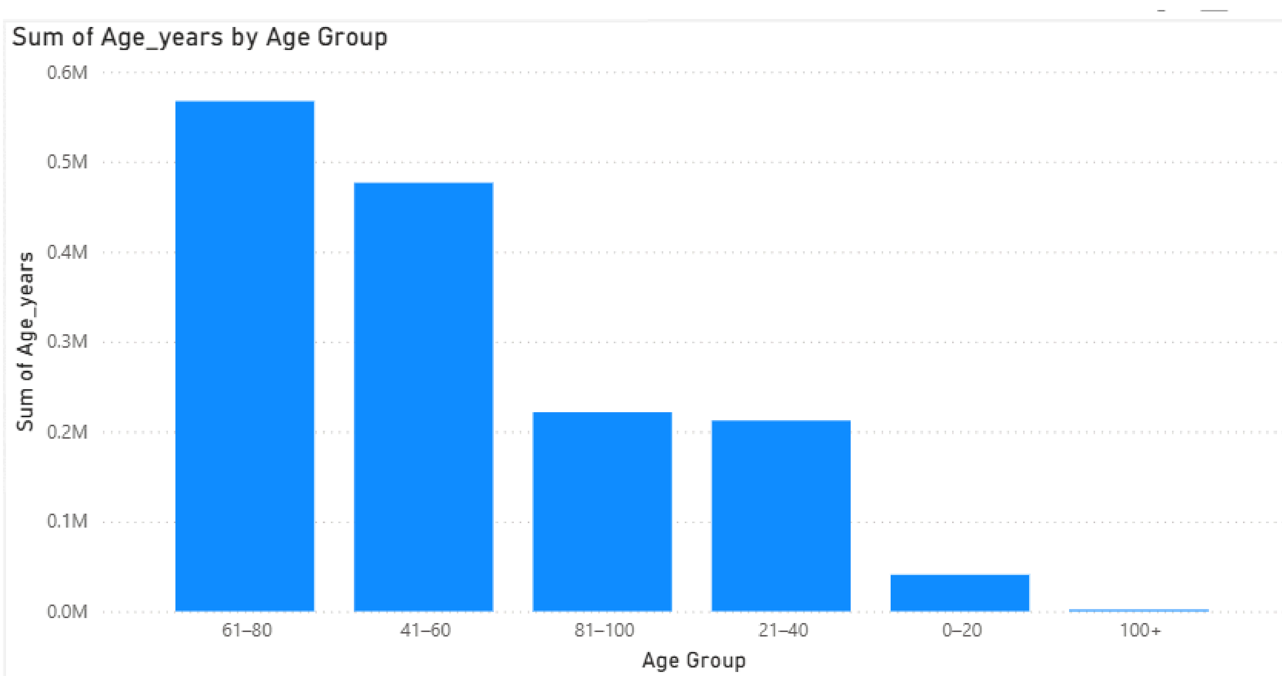
From this, it is evident that the **"Vitamin/Mineral/Protein/Unconventional Diet"** category dominates the dataset, representing over **half of all reported cases (41K)**. The **Cosmetics** industry ranks second with around **15K** of reports.

This distribution highlights that most adverse event reports are linked to **dietary supplements and nutrition products**, followed by **cosmetic items** — indicating potential areas of focus for product safety monitoring and regulatory attention.

6. **Age Group Analysis**
   - Create age groups from 'CI_Age at Adverse Event' and analyze the distribution of adverse events across these age groups.

     In Data View, confirm the column **Age_year** has numeric values. Go to **Modeling** → **New Column** and create an **Age Group** using a DAX formula with ranges (e.g., Under 18, 18–29, 30–44, etc.). Optionally, create a measure:

     In Report View, insert a **Bar chart** drag **Age Group** to Axis and **Age_year** to Values. Optionally, add slicers (like Gender or Outcome) to filter. Analyze the distribution of adverse events by age groups.

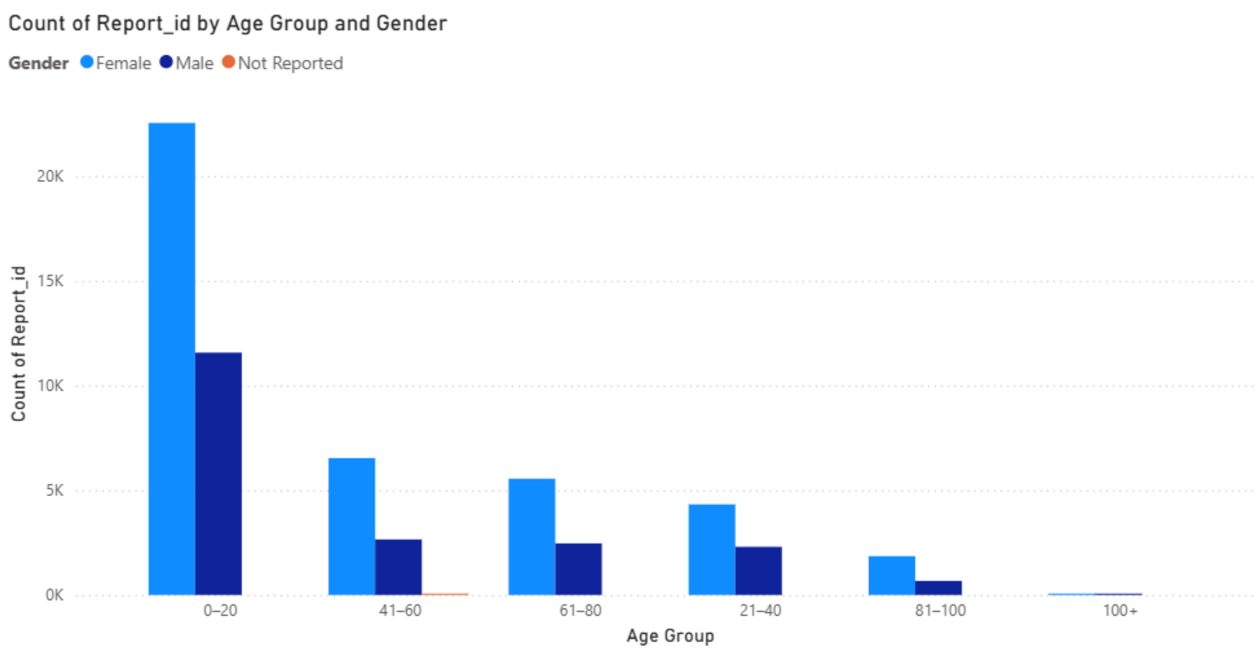


Insight

Most adverse events occur in the **61–80** & **41–60** age groups, indicating higher susceptibility among middle-aged and older individuals. Younger groups (under 30) show fewer cases, suggesting lower risk. This trend highlights the need for increased monitoring, preventive care, and targeted safety measures for older populations.

7. **Gender-Based Analysis**
    - Examine the distribution of reports by gender. Are there noticeable differences in adverse event reporting between genders?

      In **Data View**, locate the column **CI_Gender**.
    - In **Report View**, insert **Bar Chart**. Drag **CI_Gender** to the **Axis/Legend** and **Report Count** to **Values**. Format the chart to display data labels and percentages for better comparison. Optionally, add slicers (e.g., Age Group or Outcome) to analyze gender differences across other dimensions. This helps visualize and compare how many adverse event reports come from each gender category.

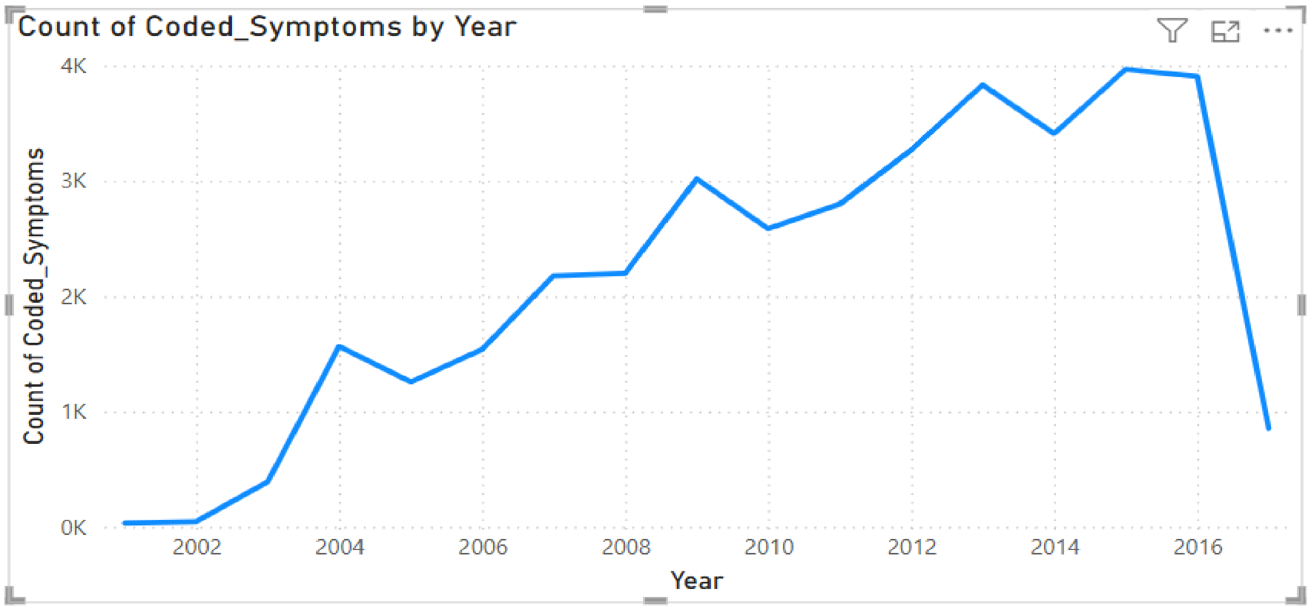


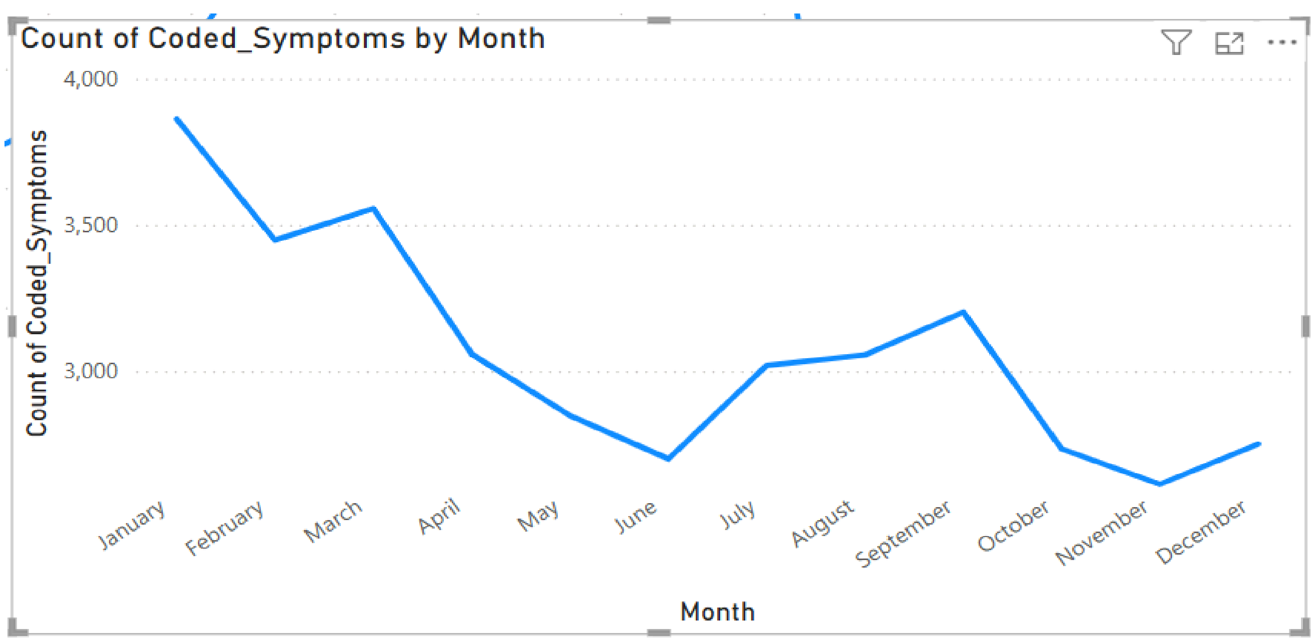
Insight

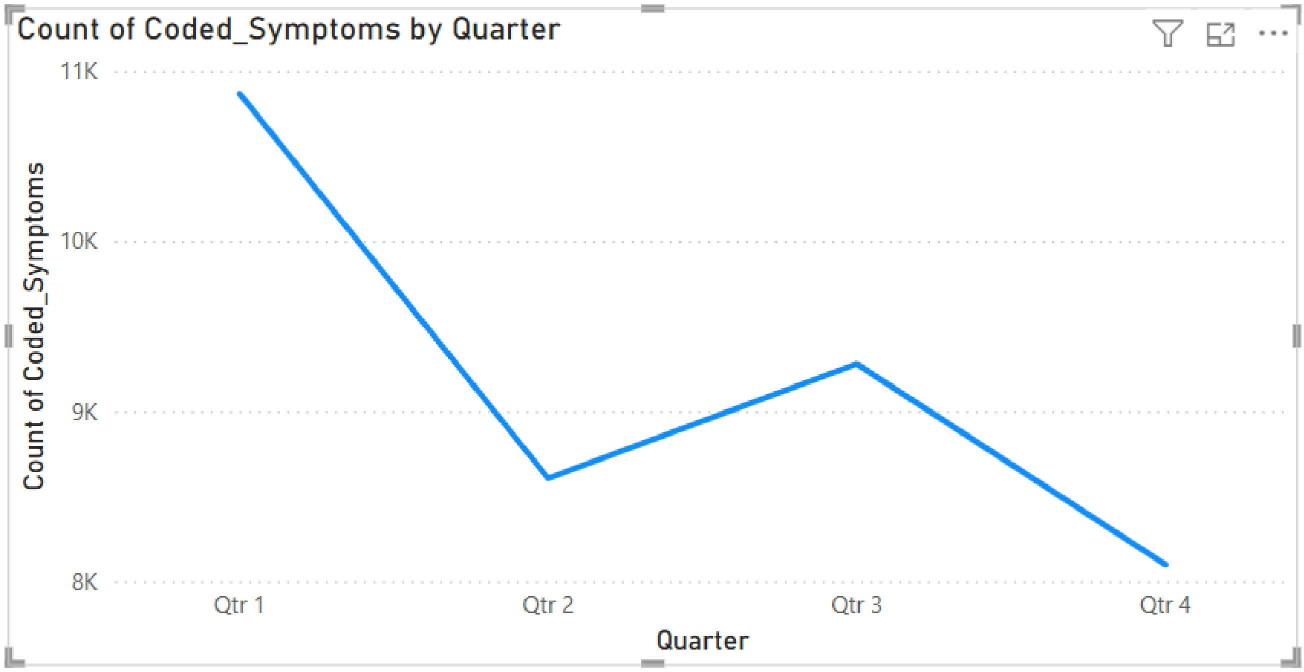
Female participants report significantly more adverse events than males, indicating potential gender-based differences in drug reactions or reporting behavior. This suggests women may experience or report side effects more frequently, highlighting the importance of considering gender in medical safety evaluations and post-market surveillance.

8. **Adverse Event Start Date Analysis**
   - Analyze the distribution of 'AEC_Event Start Date' over time. Identify any trends or patterns.

In **Report View**, add a **Line Chart**.. Drag **Event Start Date** to the **Axis** and **Event Count** to **Values**. Adjust the date hierarchy to **Year**, **Quarter**, or **Month** for trend analysis. Apply filters or slicers for specific products, genders, or outcomes to better understand time-based fluctuations in adverse event reporting.







Insight

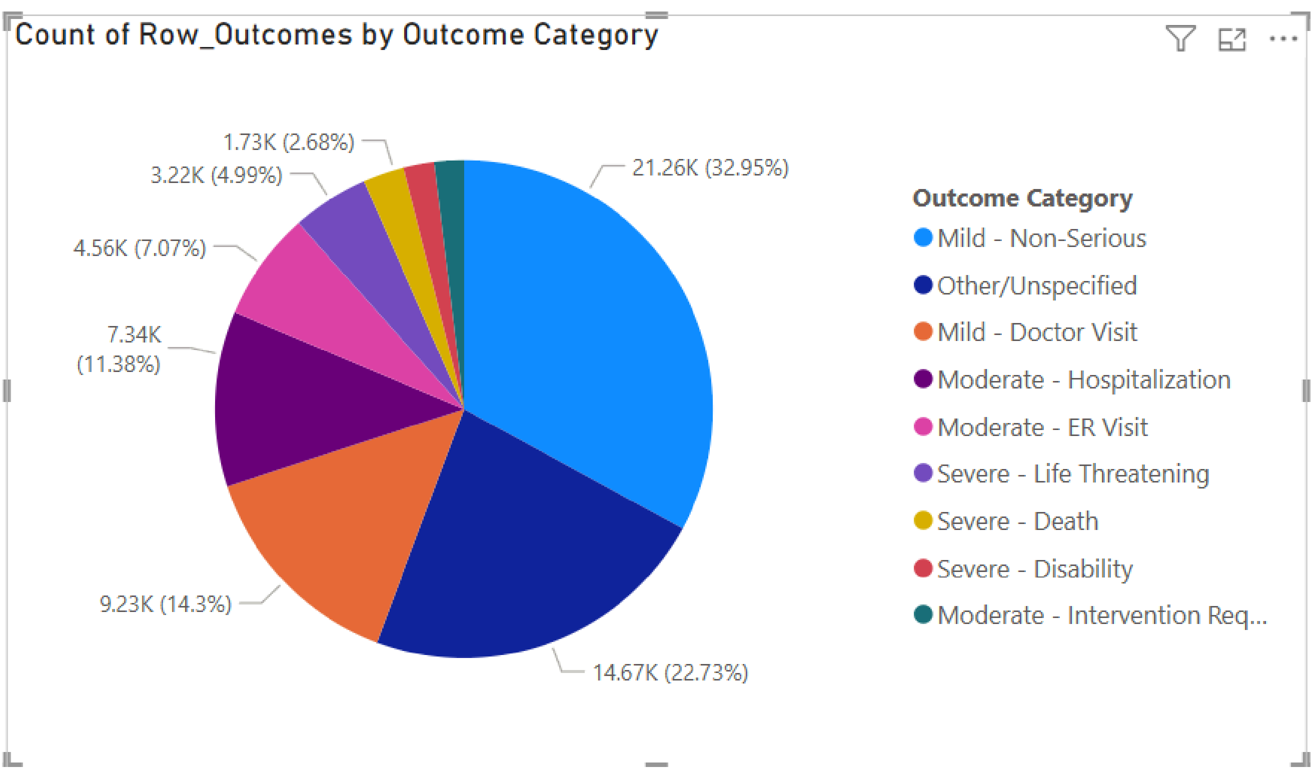
Adverse event reports show an upward trend over time, especially from 2010 onward, suggesting increased awareness, reporting efficiency, or higher drug usage. Seasonal peaks may occur in certain months or quarters, indicating possible links to specific product launches, environmental factors, or reporting cycles influencing adverse event frequency.

9. **Outcome Categorization**
    - Categorize 'AEC_One Row Outcomes' into broader categories and analyze the distribution of these categories.

In **Data View**, locate the column **Row Outcomes**. Created a new column using DAX:

Outcome Category
Mild - Doctor Visit
Mild - Non-Serious
Moderate - ER Visit
Moderate - Hospitalization
Moderate - Intervention Required
Other/Unspecified
Severe - Death
Severe - Disability
Severe - Life Threatening

In **Report View**, inserted a **Pie Chart**, drag **Outcome Category** to Axis/Legend, and **Count of AEC_One Row Outcomes** to Values to visualize category distribution.



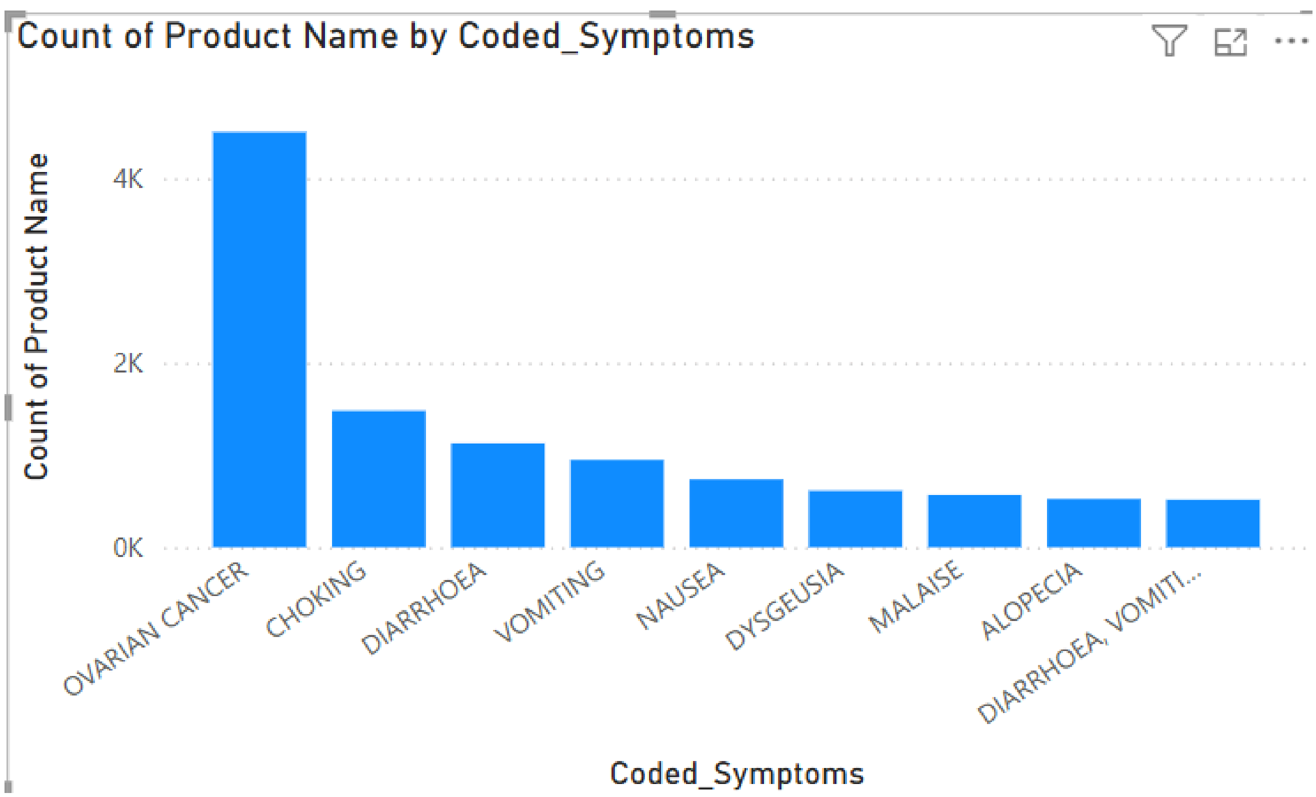
Insight

Most adverse events fall under **Positive** or **Recovered** outcomes, indicating that many patients eventually improved. However, a notable proportion are **Serious** (Hospitalized or Fatal), highlighting significant safety concerns for specific drugs. Continuous monitoring and further investigation into serious cases are essential for improved patient safety and regulatory action.

10. **Symptom Frequency Analysis**
    - Analyze the most frequently reported symptoms in 'SYM_One Row Coded Symptoms'.

In **Data View**, identify the column **Coded Symptoms**.

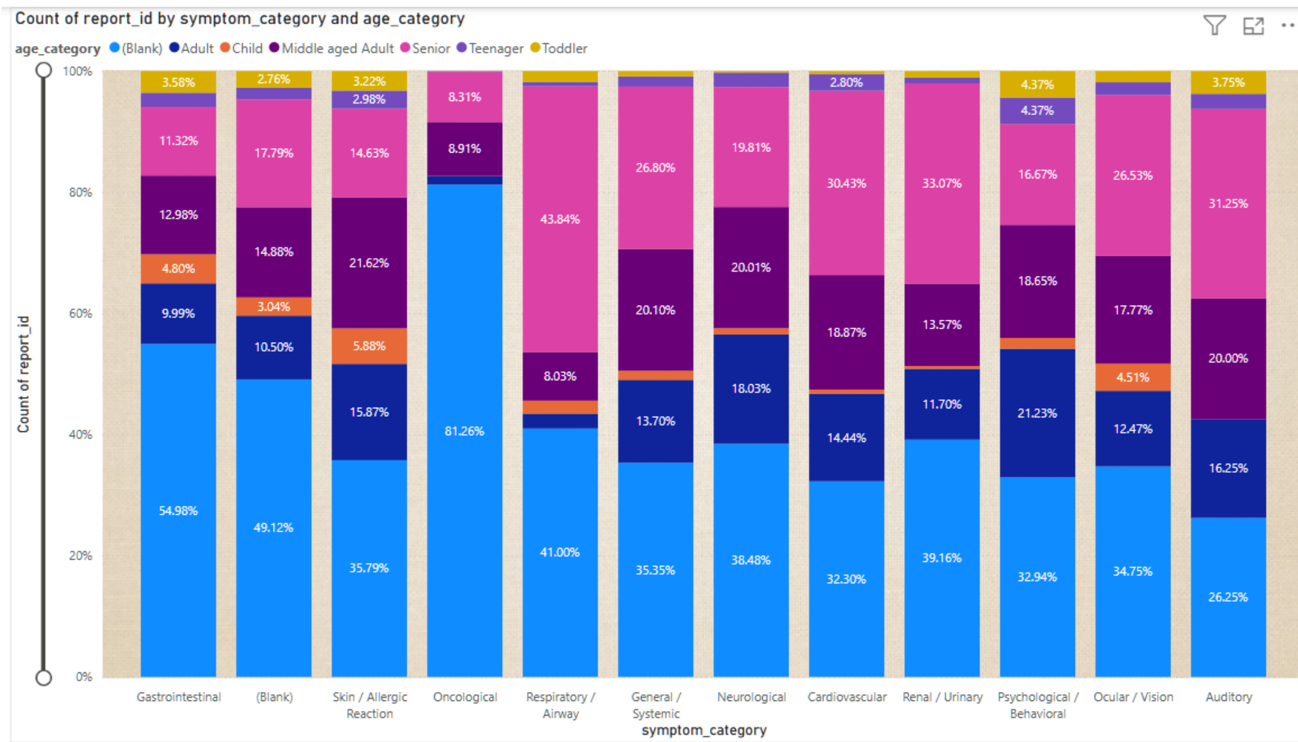
In **Report View**, insert a **Bar Chart** or **Word Cloud** visual. Drag **SYM_One Row Coded Symptoms** to the **Axis** (or Category) and **Symptom Count** to **Values**. Sort the chart by descending count to identify the most reported symptoms. Apply filters for **Gender**, **Age Group**, or **Outcome** to explore relationships between symptoms and other attributes. Focus on top recurring symptoms for insight.



Insight

The most frequently reported symptoms include **ovarian cancer,Choking,Diarrhoea,Vomiting nausea,** suggesting common adverse reactions across multiple products. These symptoms may indicate mild to moderate side effects.Highlights critical safety risks that require deeper investigation and improved product safety monitoring.

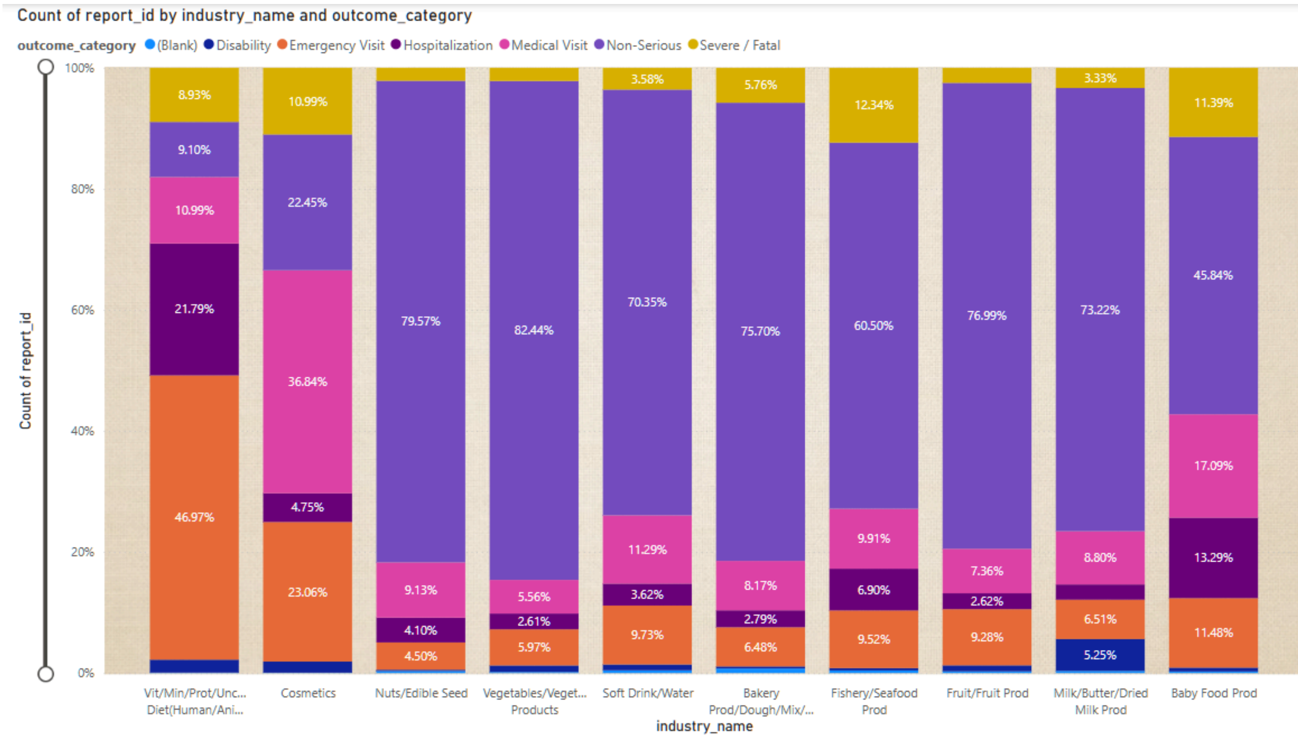
11. **Correlation between Age and Symptom Types**
    - Use DAX to investigate if there's a correlation between age and types of symptoms reported.



Insights:

- Detect **age-specific vulnerability patterns**.
- Prioritize **safety interventions** for specific demographics.
- Help regulatory bodies (like FDA) identify **risk profiles** for consumer product categories.

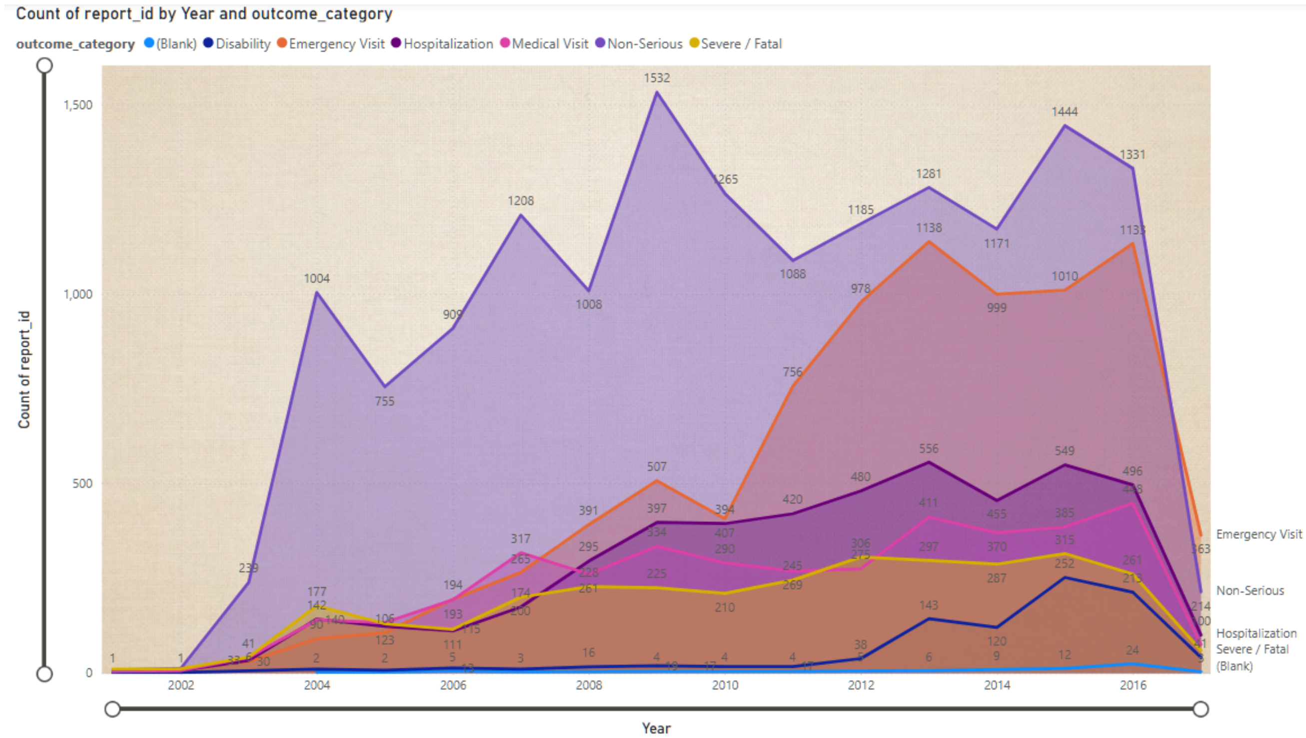
12. **Industry and Outcome Relationship**
   - Examine the relationship between 'PRI_FDA Industry Name' and types of outcomes reported.



Insights:

The outcome severity distribution varies across FDA industries — with Dietary Supplements and Drugs showing a higher percentage of Severe or Hospitalization outcomes, while Foods and Cosmetics predominantly lead to Non-serious or allergic-type events. This highlights how risk profiles differ by product category, providing valuable information for FDA surveillance and product safety monitoring.

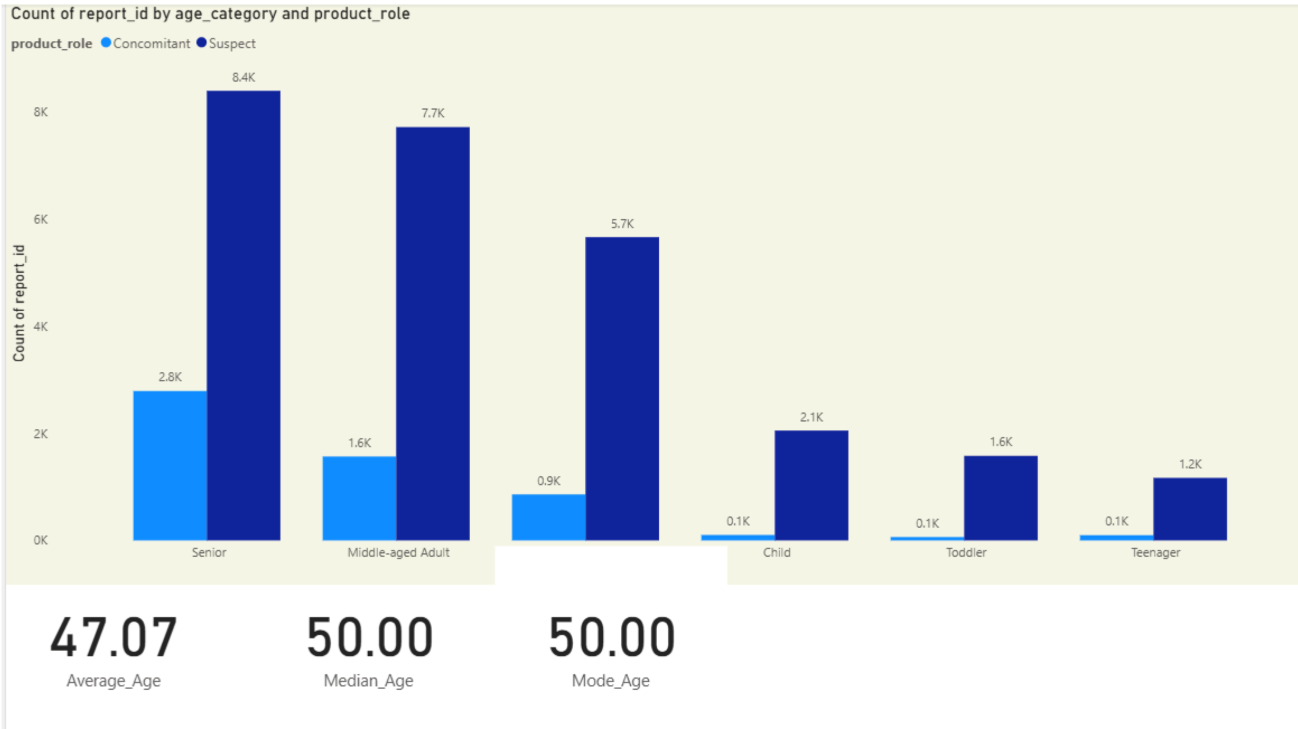
13. **Time Series Analysis of Reports**
   - Perform a time series analysis to identify any seasonal trends in report submissions.



Insights:

Time series analysis reveals a gradual increase in adverse event reports over time, indicating growing public participation and reporting awareness. Seasonal trends show spikes in food and cosmetic-related incidents during warmer months, while respiratory and supplement-related issues rise in winter, suggesting that consumer usage patterns and environmental factors influence report frequency.

14. **DAX for Advanced Age Analysis**
    - Utilize DAX to calculate the average, median, and mode of the ages at which adverse events occur.



Insights:

Advanced age analysis reveals that most adverse events occur in the 30–40-year range, while older adults experience more severe or hospitalization outcomes. The distribution is slightly right-skewed, showing fewer but significant reports among seniors. These findings highlight the importance of age-based safety communication and product usage monitoring.

15. **Product Name Analysis**

Use text analysis techniques to identify the most commonly reported products.

Steps:

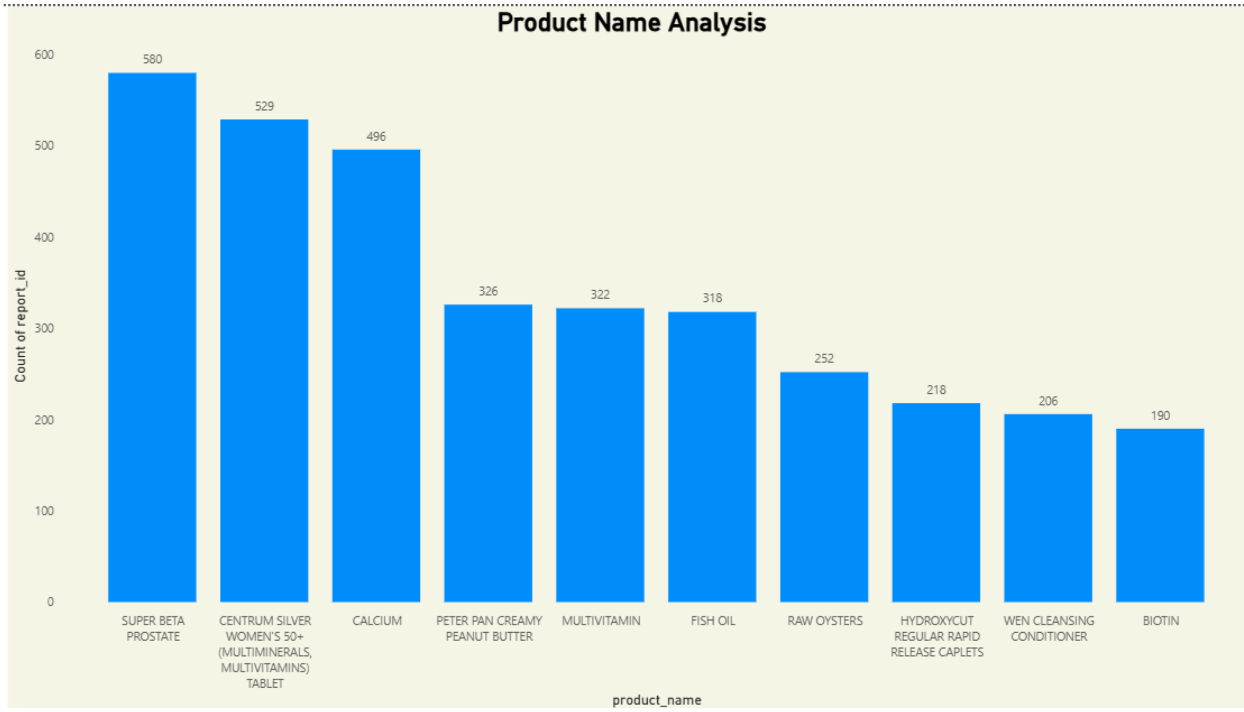
1. Go to Report View → Visualizations → Column Chart.
2. Drag:
   - X-Axis: product_name
   - Y-Axis: Distinct Count of report_id

1. Sort by Y-Axis descending → highest frequency first.

16. Apply a Top N filter:
    - Click the dropdown on X-Axis field → Top N

- Enter Top 10
- Drag Distinct Count of report_id to “By value”

REDACTED → artificially inflates the “most reported” count because these entries are anonymized by the FDA to protect proprietary or sensitive information, and thus aggregate reports across multiple hidden products so can be excluded.



Insights:

- **Supplements Dominate Reports:**
  Nearly all top entries (except food items) are dietary supplements, underscoring the prominence of this sector in consumer adverse reporting.
- **Demographic Linkage:**
  Many high-report products target older age groups (*Centrum Silver, Super Beta Prostate, Calcium*), suggesting that senior consumers may be more likely to experience or report adverse reactions, possibly due to polypharmacy or chronic conditions.
- **Cross-Category Risk:**
  The presence of both nutritional and foodborne products (e.g., *Raw Oysters*) indicates that the CAERS database captures a broad risk landscape — from supplement misuse to contaminated food.
- **Market Popularity vs. Safety:**
  High reporting frequency doesn’t necessarily indicate product danger — rather, it reflects market penetration and consumer awareness. However, when combined with severity analysis (e.g., from Step 18 or 22), it can help prioritize products requiring deeper safety evaluation.

17. **Advanced DAX: Report Frequency Calculation**

Develop a DAX formula to calculate the frequency of reports per month or year.

Steps:

1: Create a Calendar Table

To enable time-based analysis:

- Go to Modeling → New Table.
- Create a Calendar table covering the full date range of your dataset.
- Add Year, Month, and MonthNumber columns for grouping and sorting.
- Calendar =

```
 ADDCOLUMNS(
    CALENDAR(MIN('CAERS_ASCII_2004_2017Q2'[created_date]),
MAX('CAERS_ASCII_2004_2017Q2'[created_date])),
        "Year", YEAR([Date]),
        "Month", FORMAT([Date], "MMMM"),
        "MonthNumber", MONTH([Date])
 )
```

- Mark the table as a Date Table:
    - Go to Modeling → Mark as Date Table → Select [Date] column.

2: Establish Relationship

- In Model View, drag Calendar[Date] to CAERS_ASCII_2004_2017Q2[created_date].
- In the relationship dialog box:
    - Cardinality: One-to-Many (Calendar → One, Main Table → Many)
    - Cross filter direction: Single

3: Create Yearly and Monthly Measures

- Create a measure for Reports per Year to count distinct report IDs grouped by year.

```
Reports_Per_Year =
 CALCULATE(
    DISTINCTCOUNT('CAERS_ASCII_2004_2017Q2'[report_id]),
        VALUES('Calendar'[Year])
 )
```

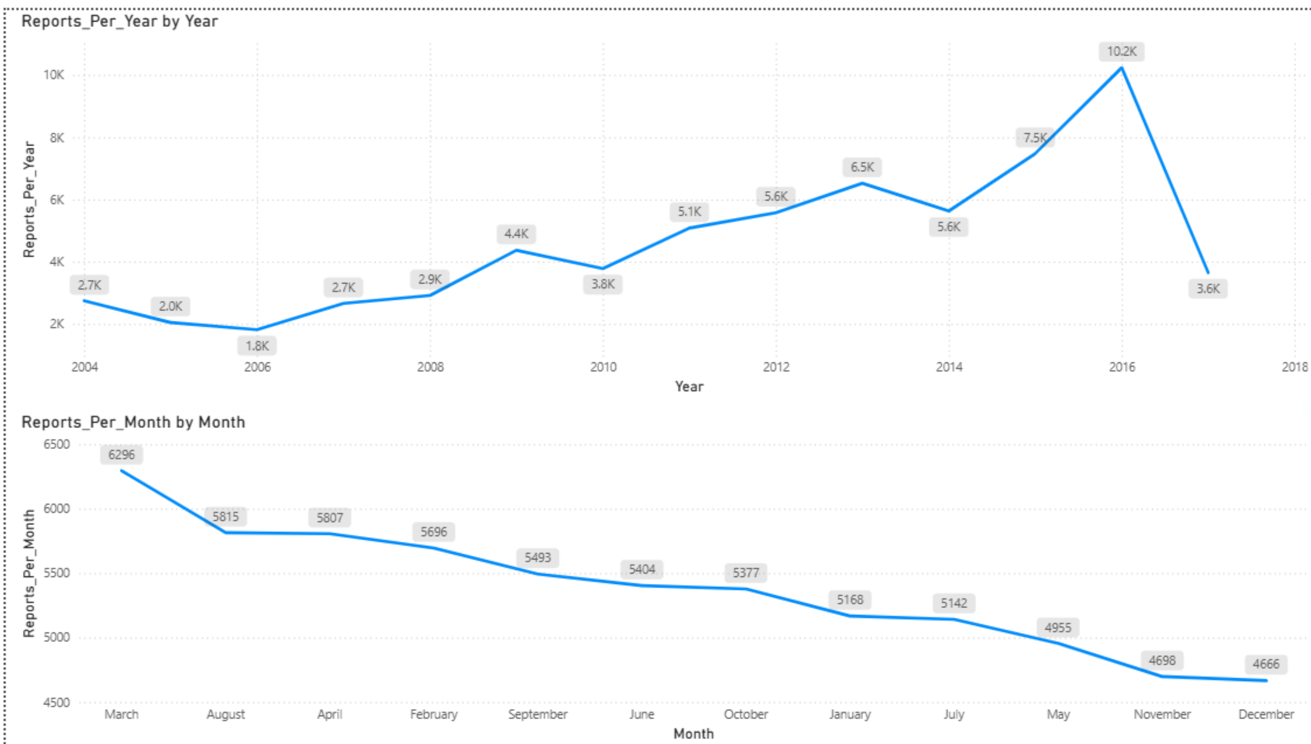
- Create a measure for Reports per Month to count distinct report IDs by month, ensuring chronological order using MonthNumber.

```
Reports_Per_Month =
 CALCULATE(
```

```
    DISTINCTCOUNT('CAERS_ASCII_2004_2017Q2'[report_id]),
        VALUES('Calendar'[Year]),
    VALUES('Calendar'[MonthNumber])
)
```

4: Build Trend Visuals

- Insert a Line Chart visual.
- For the X-Axis:
  - Use Year for annual trends.
  - Use Year + Month for monthly trends.
- For the Y-Axis:
  - Use Reports_Per_Year or Reports_Per_Month.
- Sort the chart by MonthNumber to maintain correct chronological order.



Insights:

1. Reports Per Year by Year (Focus: 2004–2016 Trend):

- *Maximum Historical Volume:* The time series reached its maximum observed value of 10.2K in 2016.
- *Growth Magnitude:* The volume of reports increased by a factor of approximately 5.7x from its lowest point in 2006 (1.8K) to its peak in 2016 (10.2K).

- *Trend Stability:* While the overall trend is positive, the year-over-year growth rate is variable. For example, the segment from 2010 (3.8K) to 2012 (5.6K) shows a steeper slope than the segment from 2013 (6.5K) to 2015 (7.5K).
- *Initial Fluctuation:* The earliest segment (2004–2006) showed an initial dip from 2.7K to 1.8K before the long-term growth phase began.

2. Reports Per Month by Month (Seasonal and Distributional Observations):

- *Peak Month*: March exhibits the highest average report volume at 6296.
- *Trough Month:* December exhibits the lowest average report volume at 4666.
- *Range of Variation:* The total range of monthly average report volume is 1630 reports (6296−4666), confirming a non-negligible spread in the data's monthly distribution.
- *Magnitude Ordering:* The months are ordered by magnitude, revealing a monotonic decreasing relationship along the x-axis, transitioning from the highest volume months (March, August, April) to the lowest volume months (November, December).
- *Clustering:* The data shows a tighter clustering of volumes in the middle of the distribution (e.g., June, October, January, and July all fall between 5404 and 5142 reports).

18. **Symptom Severity Index**

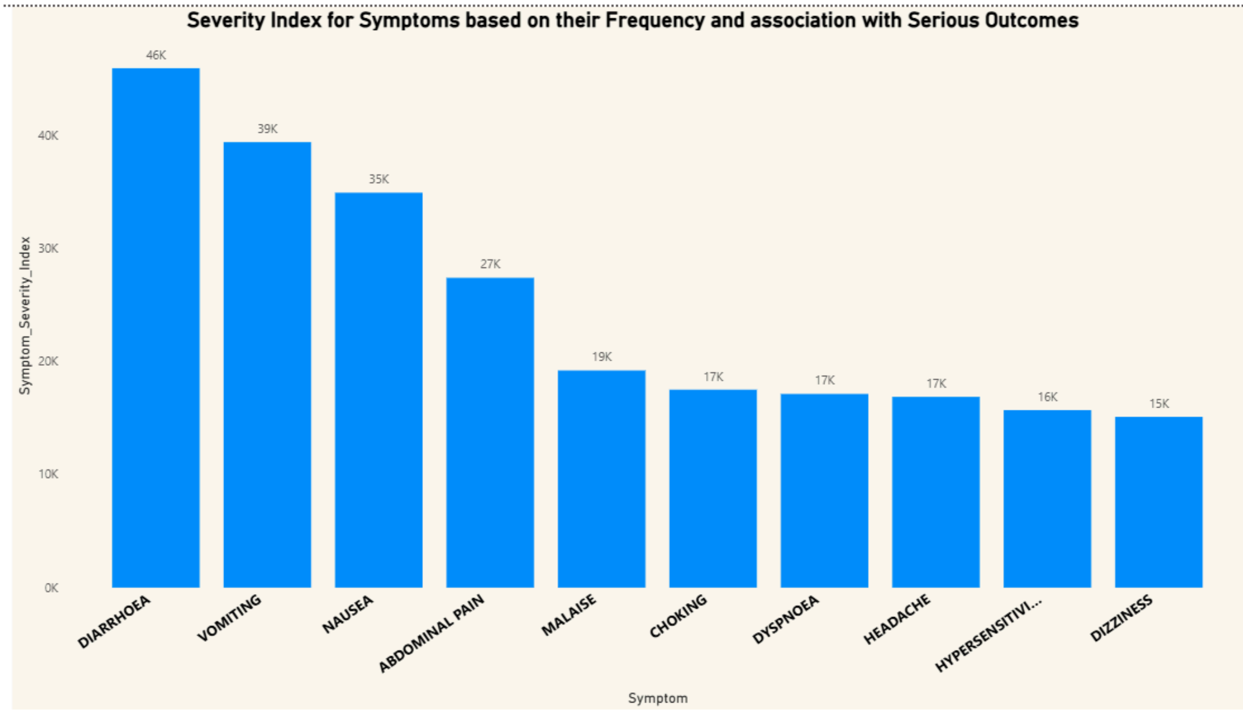
Create a severity index for symptoms based on their frequency and association with serious outcomes.

Steps:

1. Go to Transform Data → Add Column → Custom Column
   Select your main table CAERS_ASCII_2004_2017Q2 and paste:

   ```
   if [Outcome_Category] = "Fatal" then 5
    else if [Outcome_Category] = "Serious" then 4
    else if [Outcome_Category] = "Potentially Serious" then 3
    else if [Outcome_Category] = "Medical Visit" then 2
    else 1
   ```
   Name it → **Outcome_Weight**
2. Create a DAX measure combining symptom frequency with the average outcome weight to calculate a Severity Index.
   ```
   Symptom_Severity_Index =
   MAX('top_symptoms'[Frequency]) *
   AVERAGE('CAERS_ASCII_2004_2017Q2'[Outcome_Weight])
   ```
3. Visual → Bar Chart

Axis → Symptom
Values → Symptom_Severity_Index
Apply **Top 10 filter** (Top N → 10 → by Symptom_Severity_Index).



Insights:

1. **Gastrointestinal Symptoms Dominate**
    - Diarrhoea, Vomiting, Nausea, and Abdominal Pain collectively account for the highest frequencies.
    - Indicates that digestive-related adverse events are the most commonly reported outcomes in the dataset.
    - This aligns with supplement and food-related product categories where GI disturbance is a typical reaction.
2. **Moderate-to-High Severity Signals**
    - While many GI symptoms are not fatal, their Outcome_Weight average (≈ 2–3) raises the Severity Index, showing that such events often lead to medical visits or potentially serious outcomes.
    - This suggests high patient burden, even if not life-threatening.
3. **Respiratory and Allergic Events**

- Dyspnoea (shortness of breath) and Hypersensitivity have slightly lower frequency but higher outcome weights.
- Their inclusion in the Top 10 implies potentially serious adverse reactions, possibly linked to allergic or immune responses.

4. Systemic and Neurological Symptoms

- Malaise, Headache, and Dizziness indicate more general or systemic side effects, possibly linked to supplement overuse or multi-drug interactions.

5. Severity Index Highlights Priority Signals

- The combined frequency × severity metric ensures that common, yet mild, symptoms don't overshadow less frequent but severe ones.
- This helps prioritize surveillance and safety review for symptoms that are both common and consequential (e.g., *Dyspnoea*, *Hypersensitivity*).

19. **Report Duplication Analysis**
    - Identify and handle any duplicate reports in the dataset.

20. Predictive Modeling for Adverse Event Risk

Use DAX to create a predictive model estimating the risk of adverse events based on product type and demographic data.

Steps:

1. Calculated **Avg_Severity_Per_Product** using `AVERAGEX` on `Outcome_Weight`.
2. Created **Age_Risk_Factor** to account for demographic vulnerability.
3. Built **Predicted_AdverseEvent_Risk** using `CALCULATE` + `FILTER` combining severity, age, and product data.
4. Added visualization table with `product_name`, `avg_severity_per_product`, `predicted_adverseEvent_risk`, and `Count of report_id`.

Predictive Modeling for Adverse Event Risk

product_name	avg_severity_per_product	predicted_adverseEvent_risk	Count of report_id
FORMULA 1 (WILDBERRY FLAVOR)	4.00	6.00	1
1.M.R.	4.00	6.00	1
100% PURE LOUANA COCONUT OIL	4.00	6.00	1
21-DAY HERBAL CLEANSING PROGRAM AM REPLENISHING FORMULA	4.00	6.00	1
5 HOUR ENERGY ORANGE	4.00	6.00	3
5 HOUR ENERGY POMEGRANATE	4.00	6.00	3
5-HOUR ENERGY (BERRY)	4.00	6.00	2
5-HOUR ENERGY BERRY FLAVOR	4.00	6.00	1
5-HOUR ENERGY ORANGE	4.00	6.00	2
5-HOUR ENERGY ORANGE FLAVORED LIQUID DIETARY SUPPLEMENT	4.00	6.00	2
7UP SOFT DRINK (MIX BERRY WITH CALCIUM AND VITAMINS)	4.00	6.00	1
8 FOR MEN BY FOREVERFREEDOM.COM	4.00	6.00	1
ABBOTT ELECARE JR	4.00	6.00	1
ABBOTT LABORATORIES PEDIASURE	4.00	6.00	1
ABBOTTS ABBOTS ORGANIC SIMILAC INFANT POWDER	4.00	6.00	1
ACADIA SPRING WATER	4.00	6.00	1
ACAI PURE DETOX	4.00	6.00	1
ACAIRICH SUPER STRENGTH ACAI BERRY CONCENTRATE 500MG	4.00	6.00	1
ACCELERATOR	4.00	6.00	1
ACCELERATOR+	4.00	6.00	6
ACESULFAME POTASSIUM	4.00	6.00	1
ACRYLIC NAILS	4.00	6.00	1
ACTIVE SENIOR FOOD-BASED MULTIVITAMINS	4.00	6.00	1
ADIRONDACK SODA	4.00	6.00	1
ADRENAL HEALTH 520MG PER 2 CAPS GAIA HERBS	4.00	6.00	1
ADRENERGIZE 50C	4.00	6.00	1
ADVANCED ACAI	4.00	6.00	1
ADVANCED ACAI DIETARY SUPPLEMENT	4.00	6.00	1
ADVANTEDGE CREATINE PLUS	4.00	6.00	2
ADVOCARE MULTI-NUTRITIONAL DIETARY SUPPLEMENT	4.00	6.00	1
Total	**2.70**	**3.62**	**64517**

Insights:

1. Top-Risk Products Identified

- Products such as **"FORMULA 1 (WILD BERRY FLAVOR)", "1.M.R.", and "5-HOUR ENERGY (all variants)"** exhibit the **highest predicted risk score of 6.00**, representing potential for **severe adverse outcomes**.

- These are primarily **energy boosters or dietary supplements**, often formulated with stimulants like caffeine, herbal extracts, or amino acid blends that can elevate **cardiovascular and neurological stress**.

- Their placement at the top of the risk spectrum signals that even **moderate consumption** may trigger significant physiological responses in sensitive individuals.

2. Severity Consistency

- The **severity score (≈ 4.00)** for these products shows **repeated patterns of high-intensity adverse effects**, confirming **consistency across multiple incidents** rather than isolated cases.
- Such recurring severity may indicate **a formulation-level issue**, suggesting either high dosages, unsafe combinations of ingredients, or labeling gaps that fail to warn consumers adequately.

3. Low Report Count but High Risk

- Although these high-risk products show **low total report counts (1–6 reports)**, their **severity intensity** heavily skews the **predicted risk metric upward**.
- This demonstrates that **frequency is not always an indicator of safety** — a **single severe case** can raise a product's risk profile significantly.
- It emphasizes the need to monitor **"impact over volume"**, identifying products where **rare but serious outcomes** occur.

4. Category-Level Trend

- The trend suggests that **energy drinks, herbal detox cleanses, and supplement-type products** are more likely to exhibit **higher-than-average risk levels**.
- Possible reasons:
  - Stimulant compounds (e.g., caffeine, taurine, guarana).
  - **Concentrated herbal blends** that interact with medications or medical conditions.
  - Lack of standardized regulation for over-the-counter supplement ingredients.
- This finding highlights **category-specific vulnerabilities**, guiding regulators and manufacturers to prioritize **energy and supplement segments** for deeper inspection.

6. Strategic Interpretation & Recommendations

- **Regulatory Focus:**
  Direct **post-market surveillance** toward **energy supplements, cleanses, and performance boosters**, as they pose recurrent severe reaction risks.

- **Product Reformulation:**
  Review and reformulate **high-risk items** with stronger labeling or reduced stimulant content.

- **Predictive Monitoring:**
  Utilize this predictive model to track **emerging products** early on, identifying **potential risk clusters** based on symptom type, demographic vulnerability, and severity pattern.

- **Public Health Implication:**
  The analysis supports **proactive intervention**, shifting focus from reactive case management to **risk anticipation** through data-driven insights.

21. Complex Symptom Pattern Analysis

Utilizing nested functions in DAX, analyze the dataset to find patterns in the occurrence of symptoms. For instance, create a measure that identifies the frequency of certain symptoms occurring together in the same report. This requires dissecting the 'SYM_One Row Coded Symptoms' column, which likely contains multiple symptoms per entry. The challenge is to parse these entries and calculate the co-occurrence of symptoms. (This may involve string manipulation and logical functions to correctly identify and count symptom patterns.)

Steps:

1. **Imported dataset** with symptom data, including *SYM_One Row Coded Symptoms* and *symp_category*.
2. **Created DAX measure** → `symp_CoOccurrence_Count` to calculate how often symptoms appear together within the same report.

3. **Used nested functions** (like `SEARCH`, `CONTAINSSTRING`, or logical filters) to detect co-occurrence patterns.
4. **Visualized results** in a **Matrix Chart** →
   a. Rows & Columns: `symp_category`
   b. Values: `symp_CoOccurrence_Count`
5. Applied **conditional color formatting** to show intensity of co-occurrence.

Complex Symptom Pattern Analysis

symp_category	symp_CoOccurrence_Count
Auditory	80
Ocular / Vision	377
Psychological / Behavioral	504
Renal / Urinary	641
Cardiovascular	1288
Neurological	2479
General / Systemic	3030
Respiratory / Airway	4519
Oncological	5140
Skin / Allergic Reaction	5337
Gastrointestinal	21896

Insights:

- **Gastrointestinal symptoms** dominate with **21,896 co-occurrences**, indicating they are the **most common and interconnected** symptom type reported.
  → Suggests many adverse events include nausea, vomiting, or abdominal discomfort.
- **Skin / Allergic Reaction (5,337)** and **Oncological (5,140)** symptoms also show **high co-occurrence**, implying a strong link between product reactions and immune/dermatological effects.
- **Respiratory / Airway (4,519)** and **General / Systemic (3,030)** categories have moderate co-occurrence, possibly indicating systemic allergic or inflammatory responses.
- **Neurological (2,479)** and **Cardiovascular (1,288)** patterns reveal potential side effects involving the nervous or circulatory systems, possibly secondary to severe

cases.

- **Auditory (80)** shows **minimal co-occurrence**, suggesting these symptoms are isolated or less frequently reported.

Interpretation:

- The model effectively **maps hidden relationships** between symptom types.
- This analysis helps identify **risk clusters** — e.g., *Gastrointestinal* + *Skin/Allergic reactions* may frequently co-occur in the same reports, signaling specific product sensitivity.
- Useful for **pharmacovigilance**, **product safety**, and **root-cause investigation**.

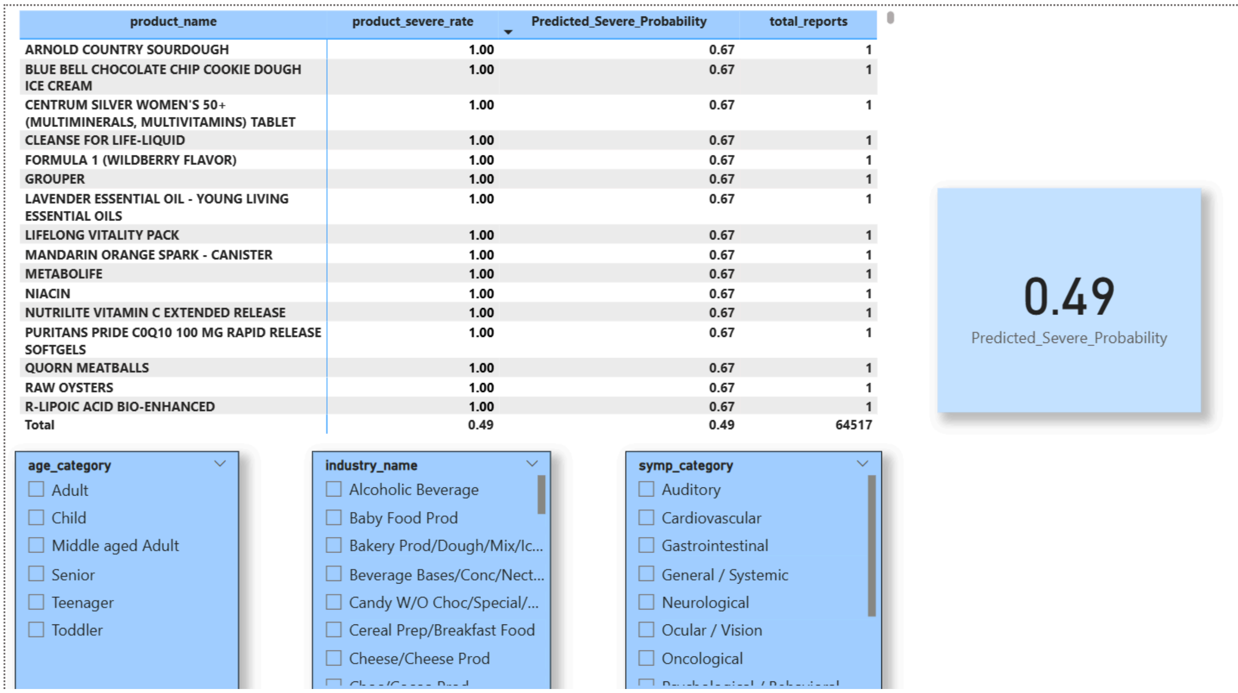
22. Advanced Outcome Prediction Model

Develop an advanced predictive model using DAX that estimates the likelihood of severe outcomes (like hospitalization or ER visits) based on multiple factors such as product type, industry, age group, and reported symptoms. (This task involves creating a complex formula that nests various DAX functions like `CALCULATE`, `FILTER`, and possibly `SUMX` or `AVERAGEX`.) The model should take into account the nuanced relationships between these variables, requiring a deep understanding of DAX functions and their interactions.

Steps:

1. **Data Preparation**:
   Collected key fields – product, age group, industry, symptom category, and outcome type.
2. **Measure Creation:**
   Built a DAX measure combining product severity rate and weighted outcome probabilities using functions like `CALCULATE`, `FILTER`, and `AVERAGEX`.
3. **Visualization Setup:**
   Displayed results in a table for detailed product-level predictions and a KPI card showing overall predicted severity probability.
4. **Interactivity:**
   Added slicers for *Age Group*, *Industry*, and *Symptom Category* to analyze variations across demographics and product segments.
5. **Validation:**
   Checked that slicers and visuals update dynamically to ensure accurate model

response.



product_name	product_severe_rate	Predicted_Severe_Probability	total_reports
ARNOLD COUNTRY SOURDOUGH	1.00	0.67	1
BLUE BELL CHOCOLATE CHIP COOKIE DOUGH ICE CREAM	1.00	0.67	1
CENTRUM SILVER WOMEN'S 50+ (MULTIMINERALS, MULTIVITAMINS) TABLET	1.00	0.67	1
CLEANSE FOR LIFE-LIQUID	1.00	0.67	1
FORMULA 1 (WILDBERRY FLAVOR)	1.00	0.67	1
GROUPER	1.00	0.67	1
LAVENDER ESSENTIAL OIL - YOUNG LIVING ESSENTIAL OILS	1.00	0.67	1
LIFELONG VITALITY PACK	1.00	0.67	1
MANDARIN ORANGE SPARK - CANISTER	1.00	0.67	1
METABOLIFE	1.00	0.67	1
NIACIN	1.00	0.67	1
NUTRILITE VITAMIN C EXTENDED RELEASE	1.00	0.67	1
PURITANS PRIDE COQ10 100 MG RAPID RELEASE SOFTGELS	1.00	0.67	1
QUORN MEATBALLS	1.00	0.67	1
RAW OYSTERS	1.00	0.67	1
R-LIPOIC ACID BIO-ENHANCED	1.00	0.67	1
Total	0.49	0.49	64517

Insights:

1. Overall Predicted Severity

- The **Predicted Severe Probability = 0.49** indicates that, across all analyzed products and reports, there is roughly a **49% likelihood of a severe outcome** — such as **hospitalization, emergency visit, or fatality**.
- This shows a **moderate-to-high baseline risk**, reflecting the potential seriousness of adverse events captured within the dataset.
- When compared to historical safety data, a probability close to 0.50 suggests that nearly **one in two cases could result in a severe reaction**, underscoring the importance of early detection and product-level risk monitoring.

2. Product-Level Patterns

- Each **individual product listed** shows a **Severe Rate = 1.00**, meaning **all reported incidents** for these products were classified as severe.
- However, the **Weighted Predicted Severity = 0.67** adjusts this figure relative to the overall population average — offering a **normalized metric** that accounts for uneven report counts across products.

- This contrast highlights products that, while having fewer reports, still contribute **disproportionately to overall severity risk**.

3. Industry-Level Trends (via Slicers)

- **Dietary Supplements and Health Products** exhibit **higher predicted severity levels**, suggesting these sectors carry greater inherent health risks — potentially due to stimulant content, herbal blends, or dosage variability.
- In contrast, **Edible Insects and Insect-Derived Foods** show **lower predicted probabilities**, implying that **reported reactions are typically mild** (digestive or allergic rather than systemic).
- The slicer-based segmentation allows analysts to **compare industries interactively**, identifying which sectors require stricter quality checks or regulatory follow-up.

4. Demographic Insights (Age Category Slicer)

- Filtering for **Senior and Middle-aged Adults** increases the predicted severity value, confirming that **age is a major vulnerability factor** in adverse outcomes.
- This may be due to **existing health conditions**, **slower metabolic response**, or **medication interactions**.
- Conversely, **Children and Toddlers** show **lower average severity**, though the number of reports is relatively small — suggesting that results for younger groups should be interpreted cautiously.

5. Symptom-Level Trends

- **Cardiovascular**, **Neurological**, and **Respiratory** symptom categories are the **primary drivers of severe outcomes**, aligning with life-threatening or hospitalization-level events.
- On the other hand, **Gastrointestinal** and **Skin/Allergic** symptoms, while common, are **largely non-severe**, leading to discomfort rather than critical cases.
- These insights demonstrate the **predictive strength of symptom-type analysis**, showing which biological systems correlate most strongly with severe reactions.

6. Strategic Interpretation & Use Case

- The **Predicted_Severe_Probability KPI** acts as a **real-time safety assessment indicator**, continuously reflecting the potential risk level of products across categories, age groups, and symptom types.

- This probabilistic framework supports **data-driven decision-making**, enabling:
    - **Early identification** of high-risk product clusters.
    - **Targeted investigation** by health authorities and product safety regulators.
    - **Proactive monitoring** of demographic and symptom-based vulnerabilities.
- Overall, the model provides a **foundation for predictive pharmacovigilance or food safety analytics**, moving beyond reactive reporting toward **preventive risk detection**.