Problem statement:

Travel, Food & Beverages Industry

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

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:

CAERS_ASCII_2004_2017Q2.csv

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 Advanced Analysis in Tableau

1. **Data Preparation** ◦ Import the CAERS dataset into Tableau. Conduct an initial examination to identify data quality issues or inconsistencies, such as missing values or incorrect data types. ◦ Perform necessary data cleaning steps (e.g., handling missing values, standardizing symptom names).
2. **Adverse Event Analysis**: Explore the distribution of adverse events across different product categories. Which type of product is most associated with adverse events?
3. **Severity of Events Analysis**: Assess the severity of adverse events reported. What are the common severe outcomes, and how do they vary by product?
4. **Demographic Analysis**: Analyze how different demographics (age, gender) are affected by adverse events. Are there notable differences in event types or severity?
5. **Symptom Frequency Analysis**: Identify the most frequently reported symptoms. How do these vary across product categories?
6. **Geographical Distribution**: Examine the geographical spread of reports. Are certain regions more prone to reporting specific types of adverse events?
7. **Temporal Trends in Reporting**: Investigate the trends in adverse event reporting over time. Identify any seasonal or annual patterns.
8. **Duration between Event and Report Analysis**: Calculate the average time lag between the adverse event occurrence and report submission. Does this vary by product type or severity?
9. **Impact of Product Type on Event Severity**: Analyze the correlation between product types and the severity of reported events.
10. **Customer Complaints and Outcomes Analysis**: Study the relationship between the nature of customer complaints and the resulting health outcomes.
11. **Time-Series Analysis of Adverse Events**: Utilize Tableau's time-series capabilities to analyze changes in adverse event reporting over the years.
12. **Forecasting Future Trends**: Use Tableau's forecasting features to predict trends in adverse event reporting based on historical data.
13. **Product Risk Analysis**: Evaluate the risk associated with different products based on adverse event data.
14. **Impact of Product Ingredients on Event Severity**: Investigate if certain ingredients in products correlate with more severe adverse events.

15. **Correlation Analysis**: Explore correlations between age, gender, product type, and event severity.
16. **Clustering of Symptoms**: Use clustering techniques to group similar types of symptoms and analyze their associations with products.
17. **Analysis of Event Outcomes**: Study the different types of outcomes (hospitalizations, ER visits) and their frequencies.
18. **Competitive Product Analysis**: Compare similar products in terms of adverse event reports and outcomes.
19. **Advanced Calculations for Event Severity Index**: Develop a severity index for adverse events and compare it across product categories.
20. **Analysis of Reporting Delay**: Assess the impact of delay in reporting on the severity of events and outcomes.
21. **Network Analysis of Products and Symptoms**: Construct a network graph to visualize the relationships between products and reported symptoms.
22. **Predictive Model for Event Outcomes**: Create a model to predict the outcomes of events based on initial report details.
23. **Keyword Analysis**: By analyzing the 'SYM_One Row Coded Symptoms' column, identify the most frequently reported symptoms.(This can be done using text parsing techniques to split the symptoms into individual words or phrases and then counting their occurrences.)
24. **Product Analysis**: Identify the products most frequently associated with adverse events.
25. **Trend Analysis over Time**: By correlating the textual data with dates from 'RA_CAERS Created Date' or 'AEC_Event Start Date', analyze trends over time, like the emergence of new symptoms or changes in the frequency of reports related to specific products.

Part 2: Dashboard Building

1. **Interactive Dashboard Creation** ◦ Develop an interactive dashboard in Tableau that showcases visuals for adverse event frequency, severity, demographic distribution, and symptom analysis. Include filters for product type, event date, and demographic details.
2. **Dashboard Design and Aesthetics** ◦ Focus on creating a user-friendly, visually appealing dashboard that provides clear and concise insights at a glance.

3. Dashboard Interactivity: Slicers and Filters

- Implement interactive elements like slicers and filters to enable users to explore the data dynamically. Users should be able to filter views based on factors like product categories, symptom types, or severity of events.

1. **Summary and Insights**

- Include a section in the dashboard that summarizes key findings from the analysis. This could cover the most frequently reported products associated with adverse events, demographic insights, or notable trends in the data over time.