**Thuong Pham, Shrey Vachhani, Jildo Lumumba, Trent Korycinska**

**MSIS415 Group Project Deliverable 1**

* **What’s the URL of the website you would like to collect data from? Will you use web scraping techniques taught in class to gather the data or the data is directly available for download?**

URL :

<https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts> : This is the FDA's website that publishes information about food recalls, safety alerts, and related product recalls.

<https://api.fda.gov/food/enforcement.json>: This is the official FDA API, providing food recall data in JSON format.

We don’t use web scraping because the FDA data is provided directly through the API, we will use API requests to get the data instead of having to scrape to collect information from the website.

* **What data would you like to collect? List the data fields that you will save into your database.**

**Data collected from FDA API:**

| **Fields** | **Description** |
| --- | --- |
| "recall\_number" | Recall number (unique ID for each case) |
| "recall\_initiation\_date" | Date the recall began |
| "product\_description" | Description of the recalled product |
| "reason\_for\_recall" | Reason for the product recall |
| "recalling\_firm" | Name of the company that manufactured the product |
| "state" | State affected by the recall |
| "product\_quantity" | Number of products recalled |
| "classification" | Severity of the recall (Class I, II, III) |
| "report\_date" | Date the recall was reported |
| "distribution\_pattern" | Region or country where the product was distributed |

* **Describe what descriptive analysis (including data visualization) you would like to perform on the data you collected.**

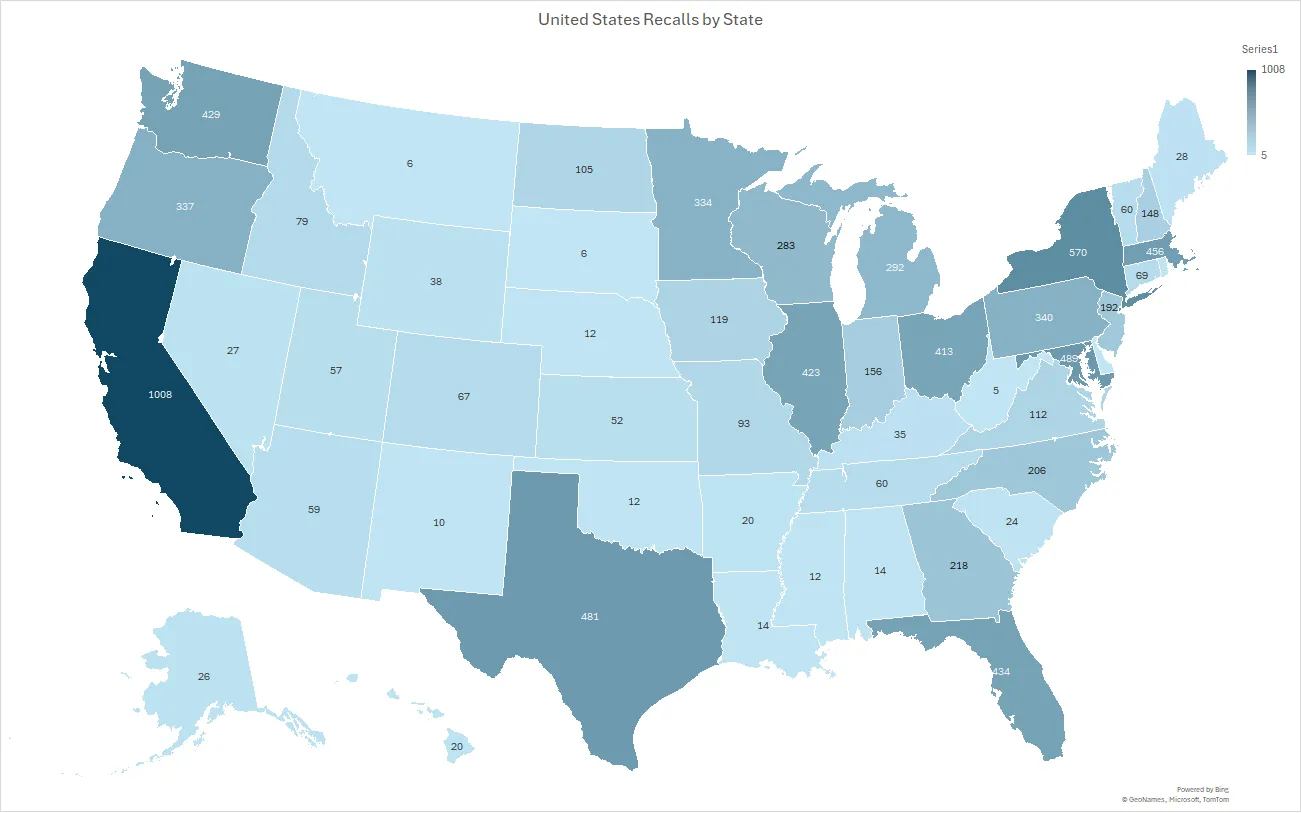
First, we’ll look at **how the number of recalls has changed over the years** by drawing a line graph. It helps us see if the number of recalls is increasing or decreasing over time.

Next, we’ll look at **which companies have the most recalls** by drawing a bar chart to compare them. We’ll also look at the most common reasons for recalls, by counting the frequency of each reason then drawing a pie chart to show them.

Finally, **which states in the US are most affected by recalls**, so we’ll use a heat map to show the number of recalls by state.

These charts will help us see the overall data in visual and understandable

* **What is the first analysis you would like to conduct on the data you collected?**

****

* 1. ***What questions you would like to answer with the analysis?***

We wanted to know the number of food recalls has been increasing or decreasing over time.

* 1. ***What answer do you expect?***

We expect the number of recalls to increase over the years, especially as food safety regulations become more close. In some years, there may be an increase in mutations associated with disease outbreaks or significant policy changes.

* 1. ***Why is this interesting?***

It helps us to understand whether the trend of recalls is increasing or decreasing, and also can help authorities evaluate the effectiveness of their regulatory efforts. Especially, help consumers gauge food safety levels over time

* 1. ***What data fields will you use to perform the analysis?***

recall\_initiation\_date

product\_quantity

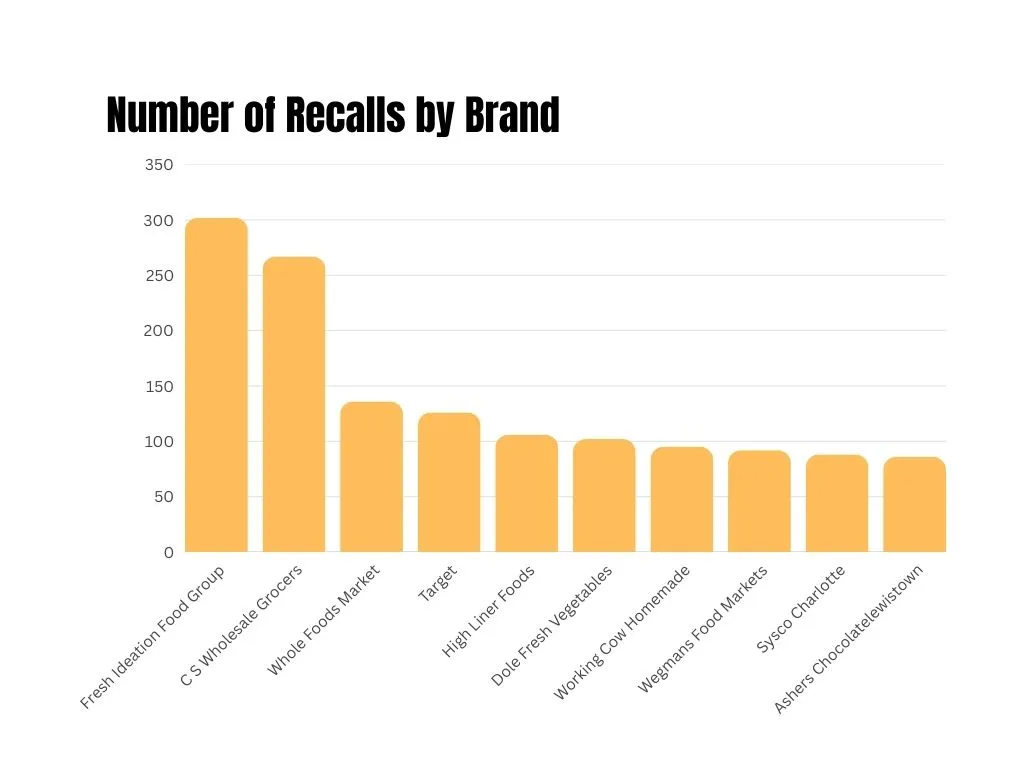
* 1. ***What analytic techniques will you use for this analysis?***

We will group the data by year based on recall\_initiation\_date, count the number of recalls per year, then draw a line chart to see the trend change over time

* 1. ***Does the result answer your questions satisfactorily?***

Yes. If recalled food increases, we can find the cause and make adjustments. If there is a sudden increase, we will consider whether it is due to new laws, disease or quality issues or pandemic ect.

* **What is the second analysis you would like to conduct on the data you collected?**



* 1. ***What questions would you like to answer with the analysis?***

Which companies have the highest number of recalls?

Are there specific companies that frequently recall products?

* 1. ***What answer do you expect?***

We expect to see a few companies having a significantly higher number of recalls than others. Large-scale manufacturers and distributors might be responsible for more recalls due to their large production volume.

* 1. ***Why is this interesting?***

This analysis helps identify whether certain companies have recurring safety issues. If specific companies are responsible for frequent recalls, regulators and consumers can pay closer attention to their practices and products.

* 1. ***What data fields will you use to perform the analysis?***

recalling\_firm (Company name)

recall\_number (Unique recall count per company)

* 1. ***What analytic techniques will you use for this analysis?***
     1. Group the data by recalling\_firm
     2. Count the number of recalls per company
     3. Visualize the results using a **bar chart** to compare recall frequencies across different companies.
  2. ***Does the result answer your questions satisfactorily?***

Yes, if certain companies consistently appear at the top of the recall list, it indicates potential quality control issues. If the recalls are evenly distributed, it suggests that food safety incidents are more widespread across different manufacturers.

* **What is the third analysis you would like to conduct on the data you collected?**
  1. ***What questions would you like to answer with the analysis?***
* Which states experience the highest number of food recalls?
* Are recalls concentrated in specific regions, or are they evenly distributed across the U.S.?
  1. ***What answer do you expect?***

We expect states with larger populations, such as California, Texas, and New York, to have more recalls due to their high food consumption and distribution networks.

* 1. ***Why is this interesting?***

Understanding regional patterns in food recalls can help pinpoint areas where food safety issues are more prevalent. It can also help regulators focus on enforcing stricter safety measures in high-risk areas.

* 1. ***What data fields will you use to perform the analysis?***
* state (Location of the recall)
* recall\_number (To count the number of recalls per state)
  1. ***What analytic techniques will you use for this analysis?***
* Group the data by state
* Count the number of recalls in each state
* Visualize the results using a **heat map** to show recall intensity by state
  1. ***Does the result answer your questions satisfactorily?***

Yes, if certain states consistently experience high numbers of recalls, we can investigate whether it is due to high food distribution volumes, specific regional factors, or regulatory differences. If recalls are evenly spread across states, it suggests that food safety issues are widespread rather than localized.