**Data Science Coding Exercise: Risk Identification for Textile Dye Suppliers**

# Context:

You are a data scientist working with the **Textile and Apparel industry giant**. It relies on a global supply chain for its ingredients, including **Textile Dyes**, which are supplied by 10 major companies worldwide. To ensure a stable supply chain, the company needs to monitor potential risks associated with these suppliers, such as geopolitical issues, environmental challenges, financial instability, logistics disruptions, and market competition.

Your task is to **analyze news articles** related to these 10 suppliers for the **last 2 years (2023 and 2024)** and identify risks that could impact the company’s supply chain. The end goal is to provide actionable insights to the company’s supply chain management team, helping them mitigate risks and ensure uninterrupted operations.

The 10 suppliers are:

1. **Welspun Living Limited** – A global leader in home textiles, specializing in towels, bed linens, and flooring solutions.
2. **Teejay Lanka PLC** – A major textile manufacturer known for producing knitted fabrics for intimate wear, activewear, and casualwear.
3. **Arvind Limited** – A diversified textile and apparel company excelling in denim, advanced textiles, and branded apparel.
4. **Caleres, Inc.** – A leading footwear company with brands like Famous Footwear and Sam Edelman, specializing in stylish and comfortable shoes.
5. **Interloop Limited** – One of the world’s largest hosiery manufacturers, supplying socks, leggings, and apparel to global brands.
6. **Kitex Garments Limited** – A top manufacturer of infant and children’s apparel, exporting to major global retailers.
7. **ThredUp Inc.** – A pioneer in online secondhand fashion, offering a marketplace for buying and selling pre-owned clothing.
8. **G-III Apparel Group, Ltd.** – A global fashion company producing apparel and accessories under brands like DKNY, Calvin Klein, and Tommy Hilfiger.
9. **Mint Velvet** – A UK-based fashion brand known for its relaxed luxury and modern, effortless style in womenswear.
10. **White Stuff Limited** – A British fashion and lifestyle brand specializing in unique, casual, and colorful clothing.

You will find a JSON file attached containing links to the news articles. Review the data file, extract the necessary articles & data features, and utilize it for your analysis.

# Objective:

Using **NLP and machine learning**, analyze news article links given in the JSON file related to the top 10 textile dye suppliers and identify the risks in the following five areas:

1. **Geopolitical and Regulatory Risks** (e.g., trade wars, tariffs, regulations)
2. **Agricultural and Environmental Risks** (e.g., droughts, crop failures, climate change)
3. **Financial and Operational Risks** (e.g., bankruptcy, labor strikes, production issues)
4. **Supply Chain and Logistics Risks** (e.g., transportation bottlenecks, fuel price hikes)
5. **Market and Competitive Risks** (e.g., price fluctuations, competitor actions)

# Deliverables:

1. **A structured overview of the problem-solving approach**, outlining the key steps taken from data extraction to risk classification. ***(Compulsory to submit)***
2. A **Python script or Jupyter Notebook** or **VS Code zip** containing:
   * Code and functions for extracting data from links and classifying them with proper risk tags
   * Preprocessing and analysis of the news articles
   * One or more machine learning models to classify articles into the five risk categories and identify the direction of risk (+ve or -ve i.e. increased risk or decreased risk)
   * Visualizations summarizing the risks for each supplier
   * A working code file that solves this problem to the *best extent possible within the given time.*
3. A **brief PPT report** (1-2 pages) explaining:
   * Key findings and insights
   * Recommendations for the Textile company to mitigate the identified risks

# Evaluation Criteria:

Your submission will be evaluated on the following parameters:

1. **Problem-Solving:** Creativity and effectiveness in addressing the problem
2. **Technical Skills:** Proficiency in web scraping, NLP, and machine learning
3. **Analysis and Insights:** Depth of analysis and clarity of insights
4. **Visualization:** Quality and effectiveness of visualizations
5. **Communication:** Clarity and structure of the PPT report