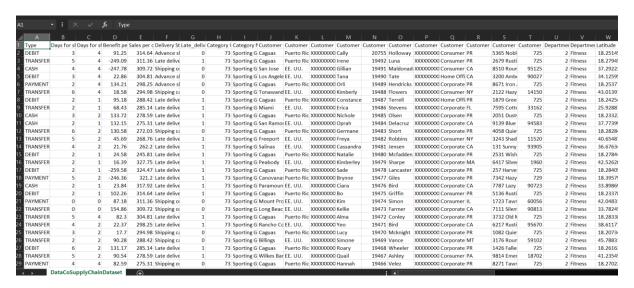
# Data-Driven Innovations in Supply Chain Management with Qlik Insights

## **Data Collection and Extraction**

<u>Collection:</u> The dataset (<u>link</u>) used in this project, titled "DataCo Smart Supply Chain for Big Data Analysis," is sourced from Kaggle. It encompasses various aspects of supply chain operations, making it suitable for comprehensive analysis and machine learning applications.

This dataset allows for the analysis of structured data across provisioning, production, sales, and commercial distribution activities. By centralizing this diverse information, it supports the application of advanced analytics to optimize various components of the supply chain.



## **Understanding the Data:**

#### **Data Structure**

#### The dataset includes:

- 1. **Customer Information**: Customer demographics, IDs, and purchase history.
- 2. **Product Details**: Product IDs, categories, and pricing.
- 3. **Order Information**: Order IDs, dates, and delivery status.
- 4. Sales and Revenue Data: Sales figures and revenue metrics.
- 5. **Logistics Data**: Shipment routes, transportation methods, and delivery times.

## **Data Quality and Preprocessing**

### To ensure data quality:

- Address missing values through imputation or exclusion.
- Normalize data for consistency.
- Detect and manage outliers.

## **Key Metrics and Analytical Applications**

Key metrics derived from the data include:

- Order fulfillment rate
- Inventory turnover ratio
- Customer satisfaction index
- Transportation cost efficiency

These metrics support descriptive, analytics, enabling businesses to optimize supply chain through Qlik's advanced analytics capabilities.

## **Data Preparation:**

- Cleaning the data: This involves removing or correcting any errors or inconsistencies in the data. This can include handling missing values, removing duplicate entries, and correcting formatting issues.
- Transforming the data: Data often needs to be transformed into a format that is suitable for visualization. This can include aggregating data, converting data types, and creating new variables or features.

- Exploring the data: Before visualizing the data, it's important to explore it to understand its structure and identify any patterns or trends. This can involve calculating summary statistics, creating visualizations such as histograms or scatter plots, and identifying outliers.
- **Filtering the data**: Sometimes it's necessary to focus on specific subsets of data for visualization. This can involve filtering the data based on certain criteria or selecting specific variables of interest.
- Preparing the data for visualization software: Depending on the visualization software being used, the data may need to be formatted in a specific way. This can include restructuring the data into tables or matrices, and ensuring that the data is in a format that the software can interpret.
- Ensuring data accuracy and completeness: It's important to verify that the data is accurate and complete before visualizing it. This can involve checking for errors or inconsistencies in the data, and ensuring that all necessary data is available for analysis.