# Shipping Logistic Analysis

This project involves an in-depth analysis of a dataset related to the delivery of orders. The goal is to identify the factors influencing whether an order reaches its destination on time.

# Dataset

### 1. ID

Unique identifier for each order.

### 2. Warehouse\_block

The block of the warehouse where the order is processed.

### 3. Mode\_of\_Shipment

The mode of shipment used for the order (e.g., air, land, sea).

### 4. Customer\_care\_calls

The number of customer care calls made for the order.

### 5. Customer\_rating

The customer's rating for the order (on a scale of 1-5).

### 6. Cost\_of\_the\_Product

The cost of the product being shipped.

### 7. Prior\_purchases

The number of prior purchases made by the customer.

### 8. Product\_importance

The importance of the product being shipped (e.g., high, medium, low).

### 9. Gender

The customer's gender.

### 10. Discount\_offered

The discount offered on the order.

### 11. Weight\_in\_gms

The weight of the product in grams.

### 12. Reached\_on\_Time\_Y\_N

Whether the order reached its destination on time (1 = on time, 0 = not on time).

## Objectives

* To analyze the distribution of various categorical and numerical variables in the dataset
* To identify the relationship between these variables and the target variable (Reached.on.Time\_Y.N)
* To develop a predictive model using logistic regression to forecast whether an order will reach its destination on time based on the given input variables

## Methodology

1. **Data Preprocessing**: The dataset is cleaned, and categorical variables are encoded using LabelEncoder.
2. **Exploratory Data Analysis (EDA)**: Various data visualization techniques (pie charts, histograms, box plots, and heatmaps) are employed to understand the distribution of variables and their relationships.
3. **Univariate Analysis**: Attribute vs count plots (Histogram and pie chart).
4. **Bivariate Analysis**: Plotting two attributes using box plots.
5. **Multivariate Analysis**: Heatmap, pair plots and Confusion matrix.
6. **Correlation Analysis**: The correlation between variables and the target variable is analyzed to identify the most influential factors.
7. **Predictive Modeling**: A logistic regression model is developed to predict whether an order will reach its destination on time based on the given input variables.
8. **Model Evaluation**: The performance of the model is evaluated using classification reports and confusion matrices.

## Conclusions:

* The analysis reveals that several factors, including customer care calls, customer rating, product cost, prior purchases, discount offered, and weight in grams, influence whether an order reaches its destination on time.
* 59.7% of the total orders are reaching on time.
* Most of the orders were delivered via ships.
* Warehouse block F has the highest count.
* The logistic regression model developed in this project can accurately predict whether an order will reach its destination on time based on the given input variables. The accuracy of the model is slightly more than 70%.
* The model's performance is evaluated using classification reports and confusion matrices, which indicate its reliability and accuracy.