About Dataset:

Context:

An international e-commerce company based wants to discover key insights from their customer database. They want to use some of the most advanced machine learning techniques to study their customers. The company sells electronic products.

Content:

The dataset used for model building contained 10999 observations of 12 variables.

The data contains the following information:

Columns:

ID: ID Number of Customers.

Warehouse block: The Company have big Warehouse which is divided in to block such as A, B, C, D, E.

Mode of shipment: The Company Ships the products in multiple way such as Ship, Flight and Road.

Customer care calls: The number of calls made for enquiry of the shipment.

Customer rating: The company has rated from every customer. 1 is the lowest (Worst), 5 is the highest (Best).

Cost of the product: Cost of the Product in US Dollars.

Prior purchases: The Number of Prior Purchase.

Product importance: The company has categorized the product in the various parameter such as low, medium, high.

Gender: Male and Female.

Discount offered: Discount offered to the customers.

Weight in grams: It is the weight in grams.

Reached on time: It is the target variable, where 1 Indicates that the product has NOT reached on time and 0 indicates it has reached on time.

HADOOP QUESTIONS:

1. Display average of Weight in grams according to Warehouse blocks? => ECargo1.
2. Which method of shipment receives the most inquiries from customers? => ECargo2.
3. Was the delivery made on time, and did it additionally include a breakdown of the proportion of on-time deliveries to those that weren't? => ECargo3.
4. Based on their prior purchases, do buyers get bigger discounts? => ECargo4.