An International E-Commerce company (Electronic goods) wants to use some of the most advanced machine learning techniques to analyse their customers with respect to their services and some important customer success matrix. They also have future expansion plans to India.

They have some specific key insights to be found out from their existing customer database.



As a Data Scientist, they want you to build a model to predict if the shipments are or will reach on time or not.

For this, they want you to use Logit/Probabilistic technique with the most accurate model.

| **Data Information:**   * ID – Id number of the customer |
| --- |
| * Warehouse\_block - The company has a big warehouse which is divided in various blocks such as A,B,C,D and so on. |
| * Mode\_of\_Shipment - The company ships the products by different modes of transport such as ship, air and road |
| * Customer\_care\_calls - This variable indicates the number of calls made for enquiry of the shipment.(Sometimes customer make too many calls, hence the company wants to know that are these customers unknowingly favoured. |
| * Customer\_rating - The company has rated every customer on various parameters, 1 being the lowest (Worst), 5 being highest (Best) |
| * Cost\_of\_the\_Product - It is the cost of the product in USD |
| * Prior\_purchases - This variable indicates the number of prior purchases |
| * Product\_importance - The company has categorised the products in the range of high, medium and low based on various parameters |
| * Gender - Male or female |
| * Discount\_offered - It is the percentage of discount offered on that specific product. |
| * Weight\_in\_gms - It is the weight in grams |
| * Reached.on.Time\_Y.N - It is the Y variable, where 1 Indicates that the product has **NOT** reached on time and **0 indicates it has reached on time**   The company operates in various states of USA, but the customer data is only specific to one state in the USA.  The warehouse is located on the eastern part whereas, the state to which shipments are delivered is at the western part of USA. |

As a Data Scientist, they want you to build model to predict if the shipments are or will reach on time or not.

For this, they want you to use various Logit/Probabilistic techniques with the most accurate model. The main models they want you to build and compare the accuracy are:

1. Logistic Regression

2. Support Vector Machines

3. Random Forest

4. XgBoost or any other boosting technique.