

Big Data Project: Deliverable 1

Team :

1. Devu Satya Sai Ashish Chandra (801273533)
2. Aarsh Ghotra (801275191)
3. Spandana Pasupuleti (801275952)
4. Mounish Reddy Chintaparthi (801272611)
5. Govinda Satyanarayana Bandaru (801275228)

Communication :

1. The platform we are using to communicate the project related work is [Slack](#).
2. Together we came up with the idea of working with a shipping dataset that contains file SCMS_Delivery_History_Dataset_20150929.csv
3. Project Repository : <https://github.com/govind09051999/BigDataProject>

Business Problem or Opportunity, Domain Knowledge (link to information on domain relative to data, problem or opportunity :

With the growing usage of ecommerce websites, The demand to get the product delivered is also increasing. So to meet today's society's expectation to deliver the product as early as possible we need predictive analysis of shipments and why certain shipments are being delayed.

This dataset contains complete shipping data for all products delivered including mainly Country,Managed By,Fulfill Via,Vendor INCO Term,Shipment Mode,PQ First Sent to Client Date,PO Sent to Vendor Date,Scheduled Delivery Date,Delivered to Client Date,Delivery Recorded Date,Product Group,Sub Classification,Vendor,Item Description,Molecule/Test Type,Brand,Dosage,Dosage Form,Unit of Measure (Per Pack),Line Item Quantity,Line Item Value,Pack Price,Unit Price,Manufacturing Site,First Line Designation,Weight (Kilograms),Freight Cost (USD)

Data Structure :

Column	Data Type
ID	Int
Project Code	String
PQ #	String
PO / SO #	String
ASN/DN #	String
Country	String
Managed By	String
Fulfill Via	String
Vendor INCO Term	String
Shipment Mode	String
PQ First Sent to Client Date	String
PO Sent to Vendor Date	String
Scheduled Delivery Date	Date
Delivered to Client Date	Date
Delivery Recorded Date	Date
Product Group	String
Sub Classification	String
Vendor	String
Item Description	String
Molecule/Test Type	String
Brand	String
Dosage	String
Dosage Form	String
Unit of Measure (Per Pack)	Int

Line Item Quantity	Int
Line Item Value	Int
Pack Price	Double
Unit Price	Double
Manufacturing Site	String
First Line Designation	String
Weight (Kilograms)	String
Freight Cost (USD)	String
Line Item Insurance (USD)	Double

Data Set :

https://www.usaid.gov/opensgov/developer/datasets/SCMS_Delivery_History_Dataset_20150929.csv

Research Objectives and Question(s) (what you are trying to describe or predict with the data) :

1. To predict the estimated time of arrival of the shipment in number of days
2. Understand the correlation between the factors given in the data set and find out which ones affect the deliverable date of the package.
3. To modify the data with standard values wherever required to produce better results.
4. To build a Machine Learning model to predict the actual delivery date of the product.