# Final Project

**Data Exploration, Modification, and Cleaning using R Programming language.**

Created by

Sridhar Malladi (pza558)

for the course: Advanced R Programming for Data Science (6233)

# Table of contents



# Table of Figures

# 1. Project Overview

The COVID-19 pandemic has caused unstopping disruptions to industries around the world, making it difficult for companies to manage production, supply, and demand. Some industries are facing a big surge in demand, but are struggling to produce enough supply to meet it. Others have excess inventory on hand; therefore, this company is struggling through its supply chain for this purpose.

ABC, a famous technology company in North and Latin America, is also facing challenges due to the pandemic and its consequences. They have been one in this industry of production for the past 20 years. They have long-term experience in creating a low-cost, low-risk supply chain strategy over the past many years. However, ABC is now at risk because of its vast supply chain and its manufacturing dependence in Asia and Europe providing more flexibility. Many factories and Original Design Manufacturers (ODMs) were shut down due to the pandemic. Logistics and transportation from third-party logistics carriers are also challenging to track and handle, making it hard to estimate delivery times to customers.

Due to the increased demand for electronics, desktops, and laptops, ABC needs to cut down delivery times to stay competitive. However, the lead timing uncertainty makes it hard to set accurate customer expectation fulfilment. They are seeking creative solutions to minimize this risk and directly contribute to the profit.

# 2. Data and Analysis overview

In this project, I used R to explore, clean, and reshape the data in preparation for analysis. I read the data file into the R and open the file to look and further work on it. The data file seems to have some missing values which I have to later deal with

### 2.1 Loading of Data, Adding the required columns and calculating ILT and MLT

To start working with the data frame I first loaded it from the directory. Successfully read the data from the file directory and added new columns for the quarter and year. I then calculated the In Transit Lead Time (ILT) and Manufacturing Lead Time (MLT) based on the provided dates.

### 2.2 Cleaning Data for Analysis

During data exploration, I discovered that the dataset contained many missing values, which needed to be dealt with. I removed the missing values specifically from ILT and MLT columns, as they were crucial for the analysis. The original dataset contained 9124 rows, out of which 145 rows were removed during cleaning. The remaining 8979 rows were used for future modeling.

## 2.3 Finding a Correlation between ILT and MLT

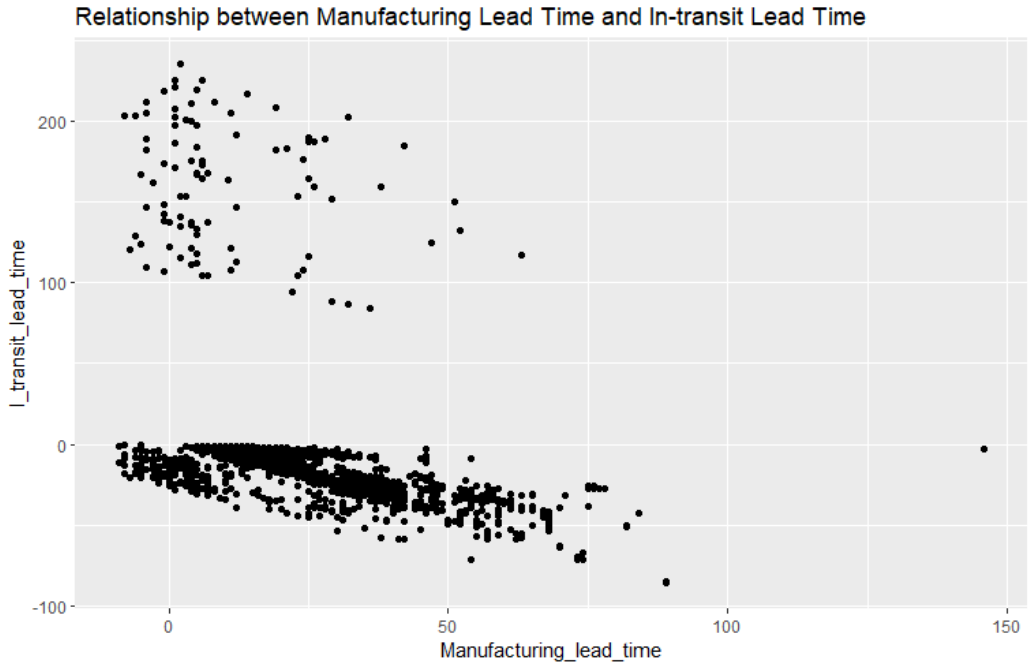


Figure A

Additionally, I reported the correlations between all variable pairs. One notable correlation I found was between ILT and MLT, which I converted into a numeric form for further calculations.

## 2.4 Finding the correlation between Variables

I successfully found the relationship between all variables and in the figure B below I have attached it for the reference.

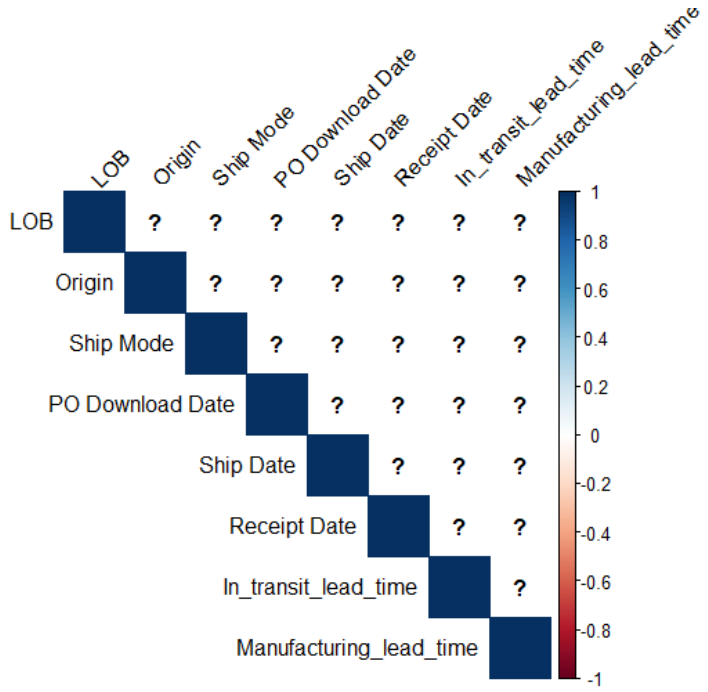


Figure B

**2.5 Function for calculating the percentage of missing values in a Data frame**

I created a function that calculated the percentage of missing elements in the data frame, which helped to identify any remaining issues with the data. Tables and figures were used throughout the report to present descriptive statistics and correlations. I thoroughly proofread the final report before submitting it, ensuring that it was professional and clear for the target audience.