

DataCo Delayed Shipments Reduction Proposal

Given the vast amounts of shipments globally and the growing issues with delays in the supply chain, this project will try to highlight factors that may result in delays to shipments. DataCo Global is a data collection company that acquired this data from multiple different types of shipments. The dataset was acquired from data science website Kaggle. This dataset from DataCo Smart Supply Chain for Big Data Analysis could be used with a company like FedEx that deals with shipping and logistics of moving products around the world. By identifying factors that can cause delays in shipping, FedEx could minimize delays in their shipments improving relationships with their customers. Maintaining consistent shipments would give greater confidence to other retailers who will ship their products through FedEx in a timely manner to ordering customers. By increasing confidence in their product and improving their reputation, FedEx could acquire new customers and improve revenues.

The goal of this project is to identify factors that might cause a greater chance for a shipment to be delayed. Once these factors have been identified, they can be minimized in order to decrease delays in shipments or even get shipments to arrive ahead of schedule. This dataset contains information on late shipments, on time shipments, and advance shipments. It also includes multiple factors including the product, whether the product was on discount as well as the locations for both the origin and destination for the product. One aspect of this problem I would like to consider is whether the distance the product has to travel is a large determining factor in the likelihood of delays occurring. One way I might try to solve this is by creating either one or two models that include and exclude the distance the product has to travel to reach its destination. By excluding distance from potential factors, the model could try to identify other factors that lead to shipment delays. Additionally, distance is not a factor that is easily controlled as it would be difficult to change the infrastructure surrounding the shipments of these products.

This analysis will be performed using Python. The process will include preprocessing and cleaning the data. Cleaning the data will help make it more usable for machine learning and linear modeling to train and test the dataset. After the data is cleaned and usable, it will be visualized in Python to help identify interesting correlations between data. Once the data has been preprocessed and trained, it can then be used to determine the best factors that can help decrease the number of late shipments. The model from this project would be available as a code set. Businesses such as FedEx could use the code set as a guiding tool in determining factors that increase shipment delays. The results from the code would then be placed on a slide deck. This slide deck would offer a clear and concise overview of the results from the project. In addition to the slide deck, this report would also contain a concluding project report that dives into further detail of the project results.

The dataset for this project can be found here:

www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis
is