Freight Optimization through Analytical Model

By-


Aditya Jain


Anurag Sharma


Debabrata Pati


Richa Mishra

Our association with ISB for CBA 4 has been enriching and inspiring. It has helped us push ourselves harder, realize our potential and harness our ambitions. We would like to thank our Professors, TAs participants and the facilities for an experience and learning that has been nothing short of spectacular. A special thanks to our mentors at Deloitte for giving us the opportunity that helped us live through the nitty-gritty of a data science project end to end.

Background

Client has large freight logistics network operating across various modes (i.e. rail, road, air, ship etc.). Client has its operations in around 2000 locations worldwide. One of the key for their operational success is movement of freight from one point to other in optimum time and cost. To achieve this they have multiple vendors operating across all the routes. Client's requirement is to choose a vendor who minimizes their cost of transportation for that route, given the constraint of time available and mode of transfer required.

Any delay in reaching of material has impact on production. Whereas sending everything by fastest mode of transport has huge cost implications and hence an optimum solution is required to address both issues.

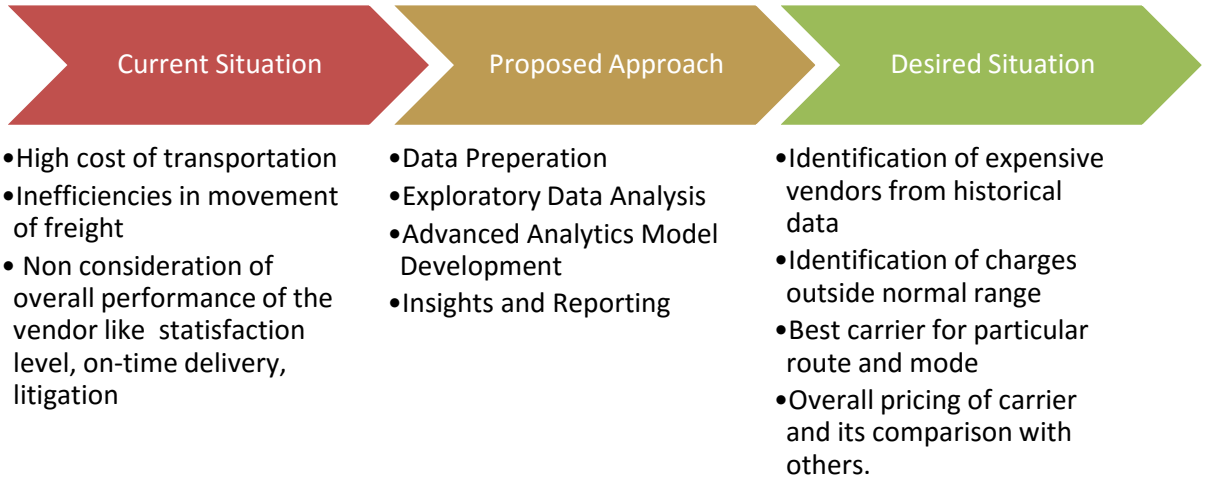
Project Objective

After various rounds of discussions and brainstorming with concerned stakeholders, the problems were refined to following objectives-

1) Identify inefficiencies from past data where they may have overspent on freight, and how that could be improved upon in future
2) Develop an advanced analytical model which, given a route and mode would find a best carrier taking into consideration parameters like cost, timeliness, safety, compliance, green rating, transit damage and corruption index rating.

Approach:

The solution utilized historical data to create model which would predict the cost of a transporting the goods between two cities based on distance, mode, product type, weight, oil prices and month of the year. In addition, looking at the historical data it would identify past inefficiencies.



Methodology

Provided below is the diagram explaining the various stages of our project

	Data Preparation	Exploratory Data Analysis	Model Development	Business Insights
ACTIVITIES	• Identify right data columns to support model • Clean data and make required transformations • Create derived KPIs • Source external data such as vendor performance rating to extend the model • Create distance function from given cities	• Analyze and visualize data • Explore relationships between variables • Understand key patterns and trends • Identification of outlier records and treat them for modeling dataset • **FEATURE ENGINEERING**	• Define Model • Identify relevant factors • Test What-If scenarios • Tweak the model for various scenarios • Find the optimum solution	• Understand and explain business significant drivers • Identification of inefficient routes • Identify the risk profile of vendor as a function of cost and rating of carrier • Reporting and Presentation in Tableau

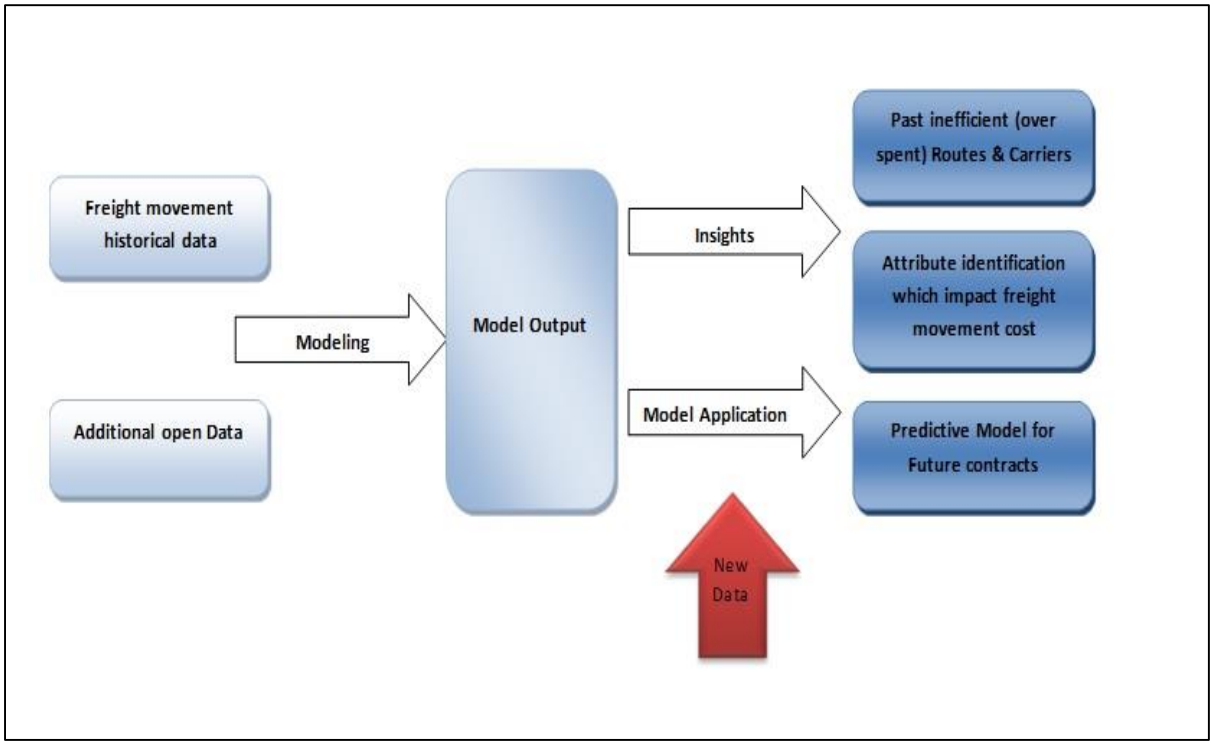
TECHNIQUES	**Data Cleaning** • Parsing • Transformation • Duplicate elimination **Missing value treatment** • Deletion • Distribution based imputation	• Scatter plots • Correlation • Bar graphs • Line graphs • Map analysis • Network Analysis	• Linear Regression • Clustering • Descriptive statistical analysis	• KPI reports • Dashboards • Trend Charts

As a final output we had to predict the best suited vendor and the ideal cost they should charge. We started with data preparation by treating outliers and missing values. We derived few variables like given the two cities, what is the distance between them. We asked client and included additional informative data like vendor ratings into the complete data set. We then did various exploratory analysis on our data to find most accurate variables and also derive a few new variables. Exploratory data analysis helped us to gain insights which helped us segregate outliers (or costlier) routes and identify inefficiencies in the historical data. These two were iterative process and took most of our time for completion.

For model building we used a logistic regression model to find the cost associated with any specific route given various parameters like product group, mode of transport, weight of item, city and distance. The output of logistic regression model was combined with other scoring parameters such as timeliness rating, safety rating, compliance and green rating etc. to come up with a final score. **The biggest contributory variables to our model were the feature engineering feats we did.** For e.g. from the dates provided we included the oil prices for that specific day into our model, since we though that they would have the biggest influence on the cost. We also identified and created a dummy variable for the months of winter as energy consumption and cost are higher during those periods. From various white papers and industry insights we got to understand that routes with higher traffic would have lower cost compared to others. To include this we feature engineered a variable which signified whether a route is high, medium or low traffic route. All these steps improved our model efficiency by leaps.

Once the model was created we utilized advanced data visualization techniques to provide interactive reports to the company to gain insights through our freight optimization model.

Provided below is the diagram depicting our solution architecture



End Result

Not only was our project able to predict near accurate cost for a specific route and mode of transport but also we were able to identify past inefficiency in the operations. For e.g. exploratory data analysis revealed to us that many LTL (Less than Truckload) were being transported by the customer between two cities on the same date. We suggested to combine them into one TL (Truck Load) which would save them a lot of cost.

Learnings:

The Capstone project was as challenging as it was enjoyable. It helped to undergo all the stages of an analytics project life cycle.

- **Invest time in data preparation**: Data preparation and exploratory data analysis would take almost 70-80% of the time, but these are the most important steps and proper execution of these would decide your model quality
- **Subject Matter Expertise:** Subject expertise is very important to identify and derive right set of variables. A poor understanding of subject may lead to irrelevant outputs. We were fortunate to have our mentor and faculty guide us at each step
- **Feature Engineering:** Thinking out of the box and doing some feature engineering to the data increases the model accuracy by leaps. In our case deriving the oil prices from the date of transportation provided helped us build a far better model