ISYE6501 Course Project

https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Optimizing-Delivery-Routes

Background

ORION, the UPS On-Road Integrated Optimization and Navigation system, has transformed UPS's pickup and delivery operations. It plays a crucial role in UPS Small Package Operations. Each morning, it provides drivers with an optimized delivery sequence for their assigned packages. By integrating diverse data sources and cutting-edge analytical tools, it enhances UPS's flexibility and efficiency in providing customized services. It has contributed significantly to fuel savings of 8.5 million gallons annually and an 85,000 metric ton reduction in CO_2 emissions.

Introduction

The Capacitated Vehicle Routing Problem(CVRP) is a very well-researched problem and has been implemented in various contexts in the past. It is a generalization of the Traveling Salesman problem(TSP) and is classified as a NP-hard problem. I found the above implementation of CVRP by UPS very interesting and have explored how it would have been solved. At its core these types of problems can be solved using combinatorial optimizations, heuristics are often used in practice. So, I know the main model would be an optimization model, my goal is to also find other data analytics models(both descriptive and predictive) that would ensure the success of this prescriptive data analysis.

Approach

Below, I have reported a step-by-step approach that would be one of the solutions to this problem. I recognize that there would be a lot of challenges in a real-world implementation of this solution, so my approach is a simplistic one.

Data Collection and Cleaning:

Start by collecting detailed data like -

- **Shipment data**: Historical shipment details like origin, destination, package size, package priority, delivery window etc.
- **Geographical data**: Coordinates of delivery/pickup locations, coordinates of office/depot locations, traffic patterns, road networks etc.
- **Historical delivery/pickup data**: Previous delivery/pickup routes taken, delivery and pickup times, shipment preferences
- **Traffic Data:** Historical traffic patterns, congestion areas, and real-time traffic information.

After consolidating this data, ensure that it is accurate and consistent. Then perform data cleaning tasks like -

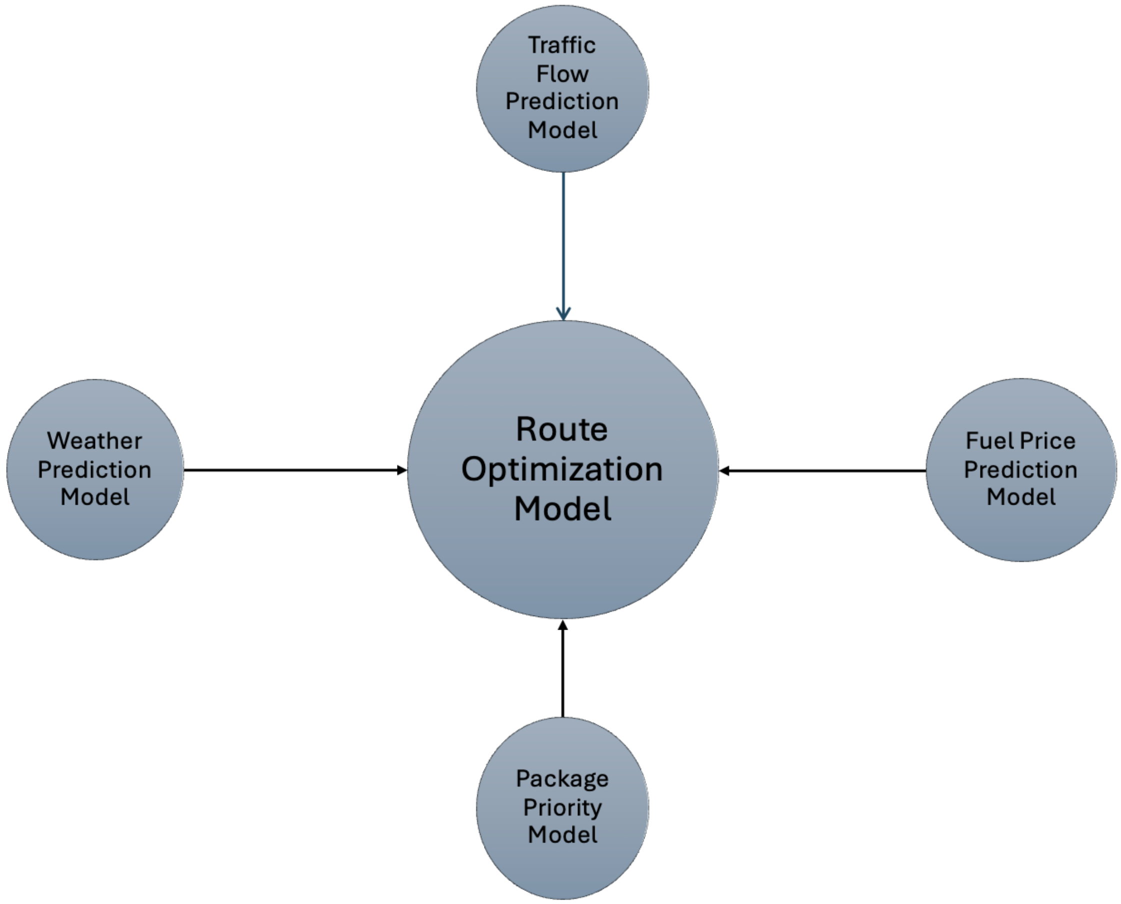
- Removing duplicates
- Handle missing data by using omission or imputation(as required and appropriate)
- Check data formats, for example date formats, measurement units etc.
- Validate data accuracy against reliable sources.
- Outlier detection and handling them after investigation as needed.

Steps:

Steps to solve the optimal route problem.

- Problem formulation
- Define the cost function
- Use predictive and prescriptive analytics to find inputs to calculate the cost of the route. Since cost of the route will change dynamically, the models will keep running in real-time and update its value.
- Use optimization to find the optimal routes for drivers to take at the beginning of the day and update the optimal routes for the drivers in real time based on cost of the route.
- Use simulation to analyze the optimal routes presented by the optimization model and make updates as needed.
- Iterate from step 3 till a good solution has been achieved.

Here is a diagram showing how different analytics models will work to support the optimization model.



Problem formulation

The goal of this analysis is to minimize the total delivery cost, which will include factors such as distance traveled, time spent, and operational expenses while ensuring that each customer is visited exactly once and the total demand on any route does not exceed the capacity of the assigned vehicle.

The objective is to -

- Minimize delivery time/cost
- Maximize delivery efficiency
- Balance workload among drivers

There will be several constraints like -

- Delivery time windows
- Vehicle capacity
- Traffic conditions
- Legal restrictions like weight limit etc.

I will use optimization to find optimal routes and other predictive and descriptive models to support the optimization model. Here are some of the questions I will try to answer.

- What would be the best route for a driver to take every day when they start their delivery route?
- How would the route change, based on real-world situations like accidents, traffic congestion etc?
- How to prioritize some deliveries over others, for example, some shipments might need to be delivered before a certain time frame.
- How to calculate the cost of a given route in real-time based on external conditions?

The cost of a route will depend on external factors like weather, traffic conditions, the number of shipments that need to be delivered on the day, delivery priority etc. Some of these factors will be known before a driver starts their route like number of packages that need to be delivered, priority of the packages etc., while some will change dynamically like traffic conditions, weather etc. We will need to incorporate these factors in the optimization model in real-time so that after an initial optimal route has been selected, we can also update the routes when any of these external factors change the cost of the route significantly.

Cost of Route

The cost of a route is dependent on some fixed features and some dynamically changing features. It can be modeled as a function of distance, time, traffic conditions and weather conditions and formulated mathematically as follows:

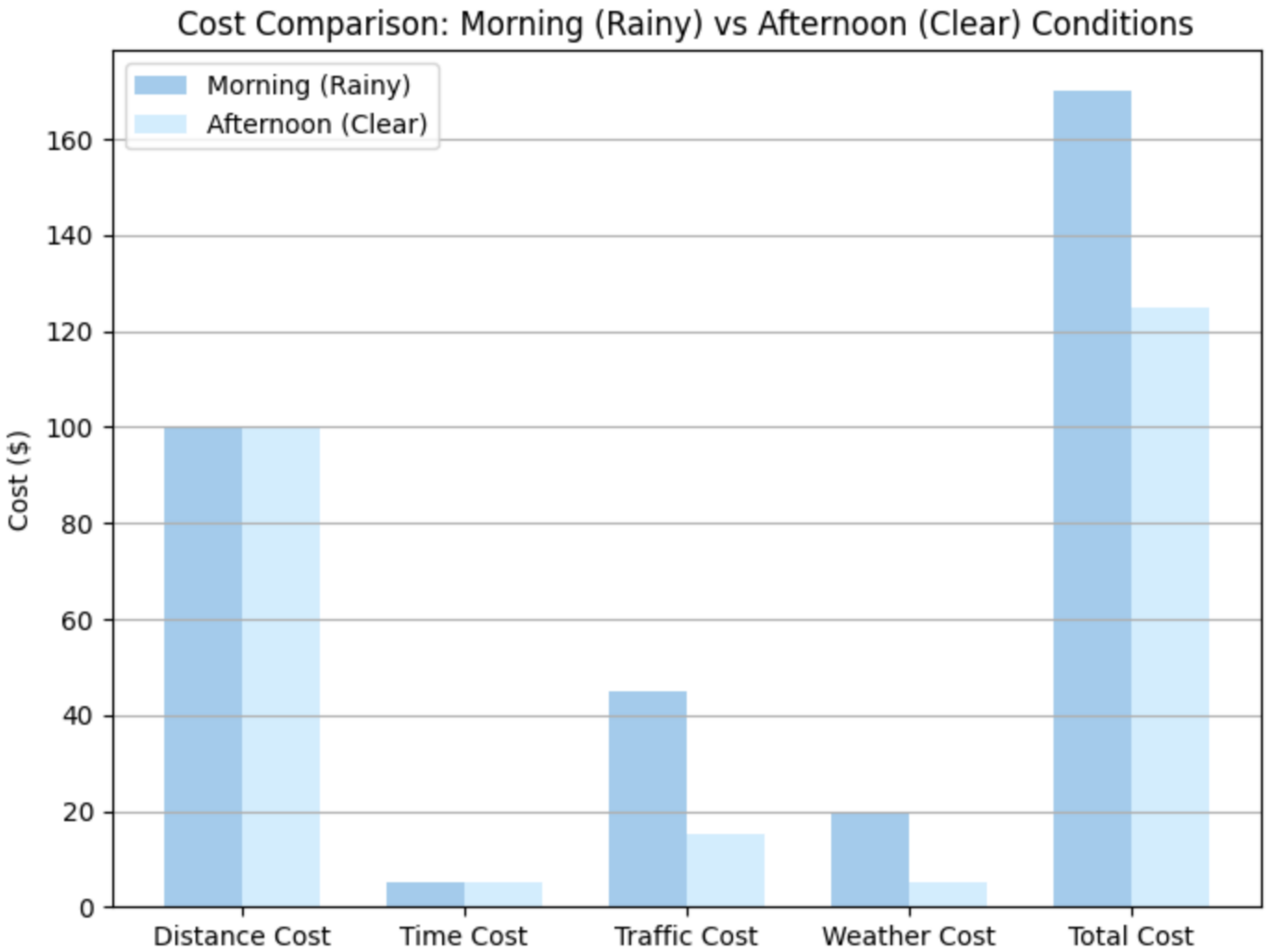
$\mathbf{C} = f(d, t, \text{traffic}, \text{weather})$

where:
$C_d = c_d * distance$ (c_d is cost per unit distance)
$C_t = c_t * time$ (c_t is cost per unit time)
$C_{traffic} = C_d * traffic_f$ ($traffic_f$ adjusts the base cost, $traffic_f > 1$ indicates increased cost due to heavy traffic and $traffic_f$ < 1 indicates decreased cost)
$C_{weather} = C_d * weather_f$ ($weather_f$ adjusts the cost based on weather conditions ex- rain or snow might increase the cost)

Combining these factors, we will get a cost function like this.

$C = c_d \cdot distance + c_t \cdot time + (c_d \cdot distance \cdot \text{traffic}_f) + (c_d \cdot distance \cdot \text{weather}_f)$

Here is a graph showing how the cost of the route can change with changes in weather and traffic conditions. For example- if it rains in the morning and becomes clear in the afternoon.



These dynamic changes in the cost of the route need to be incorporated into the route optimization model. To determine the value of the cost function, some predictive models are required to predict real-time traffic and weather conditions.

Predicting Traffic Flow

Since I will need real-time traffic flow prediction in the optimization model, time series models like **ARIMA(p,d,q)** to get real-time traffic predictions. For better results deep learning models like **RNN** and **LSTM** can be used as they provide better results, however for simplicity I am using ARIMA.

Usually, traffic data exhibits trends(long-term increase or decrease) and seasonality, however, we can check that using some statistical tests and visualization. In case the historical traffic data is not stationary we can use **differencing**(first order difference then second order difference till our data becomes stationary) in our ARIMA model. Since traffic is also seasonal, we can apply **seasonal differencing** in our ARIMA model as well.

Given:
- Traffic Counts: Hourly counts of vehicles on the road over a period of time(historical)

- Time stamps: Date and times corresponding to the above traffic counts
- External Factors: Weather data, holidays, events etc
Use: **ARIMA** (AutoRegressive Integrated Moving Average) model
To: Forecast the traffic volume on a route for the next hour

As we keep getting real data on traffic, it will be given to the model to use and predict future traffic flow. This prediction will help us calculate the real-time cost(c_{ij}) of the route in the optimization model and update the cost in the optimization model.

Real-time local weather prediction

Local weather will play an important part in the optimization model as it will affect the cost associated with each route. Practically, I would suggest using models like **regional atmospheric modeling system(RAMS)** that are good at high-resolution forecasting of local weather. For this project, I will be using a statistical model, **polynomial regression** to predict weather conditions based on historical data.

Given:
- Historical weather data(Temperature, humidity, precipitation, wind speed etc recorded hourly)
- Time variable(time of day, day of week, month and seasonal indicators that impact weather)
- Geographical factors like altitude, proximity to water bodies
- Historical weather events like storms etc
Use: **Polynomial regression**
To: Predict real-time weather conditions for the next hour

The result of this model will feed continuously into the optimization model as well so that the model can adjust the suggested routes based on real-time weather conditions.

Package priority

In real-world scenarios, there might be instances when some packages might not be delivered on time due to unforeseen events. To tackle such cases, we need to prioritize packages so we focus on delivering high-priority packages first. A classification algorithm would be beneficial to give us package priorities(for example high, medium, low) at the start of each day.

Given:
- Package features(size, weight, type(perishable, fragile, standard etc.), monetary value etc.)
- Customer features (customer profile, customer frequency of orders, customer delivery preference and urgency)
- Order date, time, delivery window
- Historical delivery times, and historical package priority

Use: **Logistic regression with decision thresholds**
To: Classify packages into high, medium and low priority based on predicted probabilities from the logistic regression model.

The output of this model will be integrated into the route optimization model which will use the package priorities to ensure high-priority packages are delivered first(by adding constraints) and adjust the optimal routes based on package priority classification. This will help enhance operational efficiency.

Finding optimal routes

At the beginning of each day for a given shipment depot, the below information would be available.

- Number of packages that need to be delivered on that day
- Address of the delivery locations
- Priority of delivery for these packages.
- Size and weight of the packages
- Capacity of the vehicles
- Number of drivers
- Weather and traffic conditions at the start of the route

First, we need to calculate the cost of each of the routes that our drivers will take that day **based on initial conditions**(weather, traffic, delivery timing constraints) and provide this information as input to the optimization model and run it to find the optimal routes by **minimizing the cost of delivery** on each route. As the driver continues their route delivering packages, predictive models running in real-time will keep feeding their predictions to the optimization model, and the model will run each time to keep updating the driver's route if required based on the predictions from the forecasting models that predict the external conditions like weather and traffic conditions.

Here is a description of the CVRP optimization model that I will use to determine the most efficient routes for the fleet of vehicles to deliver packages to a set of customers, subject to specific conditions and making sure that the **delivery time constraints** are factored in the optimization model.

Given:
- Deliveries locations along with starting location
- Cost associated with each route
- Delivery priority
- Number of drivers available

- Capacity of each vehicle

Use: **Optimization** with objective function to minimize the total cost for each route subject to below constraints.

- Each vehicle has limited package-carrying capacity
- Delivery timing constraints
- The driver route starts and ends at the same location
- Maximum duration allowed for a route
- Service time at each delivery location

To: Find a set of efficient routes(minimum cost) for each driver to deliver all the packages for that day.

Here is a mathematical representation of the optimization problem.

V = Number of nodes or stops including the starting depot.
K = Number of vehicles
Q = Capacity of each vehicle

Variables:
$\mathbf{x}_{i,j,k}$ = binary variable indicating active edge from node i to node j performed by vehicle k
$\mathbf{y}_{i,k}$ = binary variable indicating that node i is serviced by vehicle k

Constraints:

- $\sum_{k \in K} \sum_{j \in V} x_{ijk} = \sum_{k \in K} \sum_{j \in V} x_{jik} \qquad \forall i \in V \backslash \{0\}$
  (Each customer *i* is visited exactly once.)
- $\sum_{j \in V} x_{ijk} = \sum_{j \in V} x_{jik} \qquad \forall i \in V \backslash \{0\} k \in K$
  If any edge variable indexed by vehicle *k* goes to a node *i* or comes out of it, the demand *q* of this node is assigned to vehicle *k*.
- $\sum_{i \in V} q_i y_{ik} <= Q \qquad \forall k \in K$
  The total demand assigned to a vehicle must not exceed its capacity *Q*.
- $\sum_{k \in K} \sum_{j \in V} x_{0jk} = \sum_{k \in K} \sum_{i \in V} x_{i0k} = K$
  Exactly K vehicles start and arrive at the depot.

Objective function:
The objective is to minimize the total cost associated with each route traveled by each driver. Here x_{ijk} is a binary decision variable indicating if vehicle k traveled from node i and j. It is 1 if the route was taken and 0 otherwise.
$\mathbf{c}_{ij}$ represents the cost associated with the route between node i and j. The nested summation ensures that the objective function accounts for every possible route taken by every vehicle.

$$minimize \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} c_{ij} x_{ijk}$$

The objective function is aggregating the total cost incurred by all the vehicles over all potential routes and then minimizing it to find the most efficient set of routes to reduce the cost while satisfying all the constraints mentioned above.

Simulation

After the optimization model has been developed, testing and quantifying the results is important. Asking the drivers to follow the routes given by the model for testing could be very costly and impractical approach, So simulation will be used to model these routes and run the simulation many times to analyze and validate the results. **Real-world complexities** like traffic variations, road conditions, road closures, weather conditions which can impact routing decisions in the simulation model will be incorporated. Simulation will also help model scenarios like **delivery priority** or delivery times constraints, increase in demand, changes in delivery zones, or introduction of new customers and understand their potential impacts on routing and logistics operations.

Essentially this approach helps evaluate **performance of our route optimization under varying conditions** without the risk of real-world implications. This will allow us to better understand and navigate the complexities of logistics operations.

Given :
- **Historical delivery data**(past delivery routes and times, delivery volumes and schedules, customer location and delivery time window)
- **Traffic data**(historical traffic patterns and congestion data, real time-traffic conditions from "Traffic flow prediction model")
- **Weather data**(Weather conditions historical and forecasted from "real-time weather local weather prediction model")
- **Vehicle data**(Vehicle capacity and specifications, fuel consumption rates etc.)
- **External Factors**(local events, roadworks etc.)
- **Output from Optimization model**(Suggested optimized routes from the optimization model, estimated delivery times based on those routes)

Use : **Discrete Event Simulation(DES)**

To:
- Test the robustness of the optimized routes against real-world scenarios
- Feasibility and practicality of optimized routes
- Assess key performance indicators like total travel time, fuel consumption, reliability and driver workload.
- Comparing simulation results against benchmarks
- Scenario testing

By using simulation to test the results of an optimization model for real-time routing, we can gain valuable insights into the practical implications of routing strategies produced by optimization. Not only would we be able to validate the effectiveness of the optimization outputs but also prepare the company to handle real-world complexities, leading to improved efficiency and customer satisfaction.

Quantifying performance

To quantify the performance of the optimization, here are some key performance indicators that can be measured.

- **Total distance traveled** by all the vehicles during deliveries. Lower total distance indicates efficient routing resulting in reduced fuel consumption and improved time efficiency.
- **Total delivery time** to complete all deliveries. Shorter times result in improved SLAs and high customer satisfaction.
- **On-Time delivery rate** or the percentage of packages delivered within the specified time window. High on-time delivery rate reflects the model's effectiveness in meeting customer satisfaction.

$$On - Time\ delivery\ rate = \frac{Number\ of\ On - time\ delivery}{Total\ delivery} * 100$$

- **Vehicle Utilization** or how efficiently vehicle capacity was used. Higher utilization suggested better load management and cost-effectiveness of the model.

$$Vehicle\ Utilization = \frac{Total\ Load\ delivered}{Total\ Vehicle\ Capacity} * 100$$

- **Fuel Consumption** or total fuel used during the deliveries.
- **Customer satisfaction score** or feedback from customers regarding on-time delivery and service quality
- **Cost efficiency** or operational cost associated with deliveries including fuel, labor and maintenance.

$$Cost\ Efficiency\ = \frac{Total\ Delivery\ Cost}{Total\ Deliveries}$$

- **Service level agreement compliance** or adherence to predefined service levels or company specified rules on deliveries.

All these metrics from the past process(without optimization) will be collected and then compared against the same metrics on deliveries made using the optimization model. These comparison results will then be used to improve the optimization model. Not only will this help improve the optimization model but also help quantify the performance of the model.

We can monitor these metrics and assess the effectiveness of the optimization model in real-time and help make data-driven adjustments while continuously improving the delivery routing process.

Additional Insights

While the data analytics models and techniques described above focus on optimizing delivery routes, additional analytics can assist a shipping company like UPS with strategic decision-making. Here are some critical questions we can answer.

- How can we effectively select shipping depot(starting point offices) locations for effective customer coverage, in cost and time-effective ways?
- How to predict some operational costs like fuel, to efficiently manage operational resources?
- What is the optimal number of drivers needed at each location to balance operational costs and customer satisfaction?

I have explored these questions using analytics to provide insights to the company for long-term strategic planning and operational efficiency.

Shipping depot selection

To choose optimal depot locations for maximum customer coverage in a timely and cost-effective manner, clustering algorithms can be used to identify strategic areas for office locations. This model will group customers based on geographical proximity and demand characteristics that can be served by one office location.
Given:
- Customer locations (for proximity analysis)
- Customer demand data(historical)
- Distance metrics
Use: **K-means clustering**
To: Find optimal customer clusters based on geographical proximity and customer demand (we can use the elbow method to identify the optimal number of clusters while balancing maximum customer coverage and investment constraints)

Visualizations like **Voronoi Diagrams** can be presented to the stakeholders for a better understanding of customer demand and help them make data-driven decisions.

Predicting fuel prices

Fuel prices depend on a lot of factors like geopolitical events, supply and demand dynamics, economic indicators and seasonal trends. Practically this would be a hybrid model which will include a separate model for each of the below factors. For simplicity, I have used a linear regression model here.

Given:
- Historical fuel prices
- Historical Crude oil prices
- Economic indicators like inflation, GDP growth etc.
- Supply-demand data
- Exchange rates
- Seasonal factors
- Geopolitical events that have impacted fuel prices in the past

Use: **Linear Regression Model**

$$(FuelPrice_t = \beta_0 + \beta_1 . CrudeOilPrice + \beta_2 . EconomicIndicator + ... + \epsilon_t)$$

To: Predict fuel prices

Although the fuel price for the day will be known and the company might also have its own fuel reserve, predicting future fuel prices will help refine the constraints of the optimization model. Additionally, it will help in budgeting and give a competitive edge to the company.

Queuing Model for optimal number of drivers

To find optimal number of drivers needed at office location while balancing operational costs and customer satisfaction, a **queuing model** can be utilized. This model will help understand customer arrival patterns, service times, and the number of drivers required to meet the demand(deliveries).

Given :
- **Arrival rate(λ)**: Average number of delivery requests arriving per day(historical)
- **Service rate(μ)**: Average number of deliveries a driver can complete in a day(historical)
- **Number of drivers(c)**: Number of drivers available to deliver packages
- **Queuing discipline**: Higher priority packages should be delivered first

Use: A **queuing model** like (M/M/c queue)

To: Calculate metrics like average wait time, average time in the system, and utilization

Using analysis from the queuing model and calculated metrics, potential bottlenecks or service level breaches can be identified. The number of drivers can be adjusted, and the queuing model could be re-evaluated to find optimal number of drivers that are needed. Questions regarding variations in demand and service times can also be answered with this model and help the company reach an optimal staffing level.

Conclusion

Using the above approach to determine the optimal routes for drivers using data-driven decisions, the route optimization model minimizes transportation costs while ensuring timely deliveries across a diverse customer base. Through the integration of historical shipment data, demand patterns, and vehicle capacities, UPS can effectively streamline its logistics operations. Implementing this optimization not only reduces fuel consumption and operational expenses but also significantly improves service reliability.

To conclude, the route optimization model along with other supporting models will have to be validated against real data when possible and when not possible expert opinions and testing using experiment design techniques will be used. I remember from the lectures that data analytics is an iterative process, and it is true in this analysis as well. We will collect data, fit them to the models, validate, test, get **feedback** and then **make improvements to process**. Then we **iterate** all these steps again till a satisfactory solution has been achieved.

Citations:

Case Study - https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Optimizing-Delivery-Routes
For CVRP : https://en.wikipedia.org/wiki/Vehicle_routing_problem
For formulas using Latex - https://obsidian.md/plugins?search=latex, https://latex-tutorial.com