“KiP” Distribution system optimization

Problem Description

KiP (Knowledge is Power) is a wholesale company wanting to update their distribution system. They currently have six Distribution Centers, each supplying Retailers from its own Region. Changes in the demand, however, have caused some Distribution Centers to become insufficient, as Retailer’s from certain regions now require more goods. Seeing how the current capacity of Centers is enough in total, but not per region, KiP hopes to solve the issue in the short term by reallocating Retailers to Centers not necessarily in the same region.

Nevertheless, the company wants to be prepared for a more distant future too, because if demand trends continue, their Distribution Centers will soon become incapable of satisfying the total demand as well. Therefore, they also want a distribution plan that could satisfy the projected demand in three years.

Our objectives are to:
- Familiarize ourselves with the data, analyze it, and prepare it for further work
- Optimize distribution in the present moment, so that each Retailer’s demand is satisfied with minimal cost, by reallocating Retailers such that they do not have to be supplied by Distribution Center of the Region they belong to
- Predict demand over the next three years
- Optimize distribution in three years time, based on predicted demand, so that each Retailer’s demand is satisfied, by renting new Locations, and reallocating Retailers once again with the aim of minimizing the cost

Each of these objectives will have a dedicated chapter in this document, detailing methods and results.

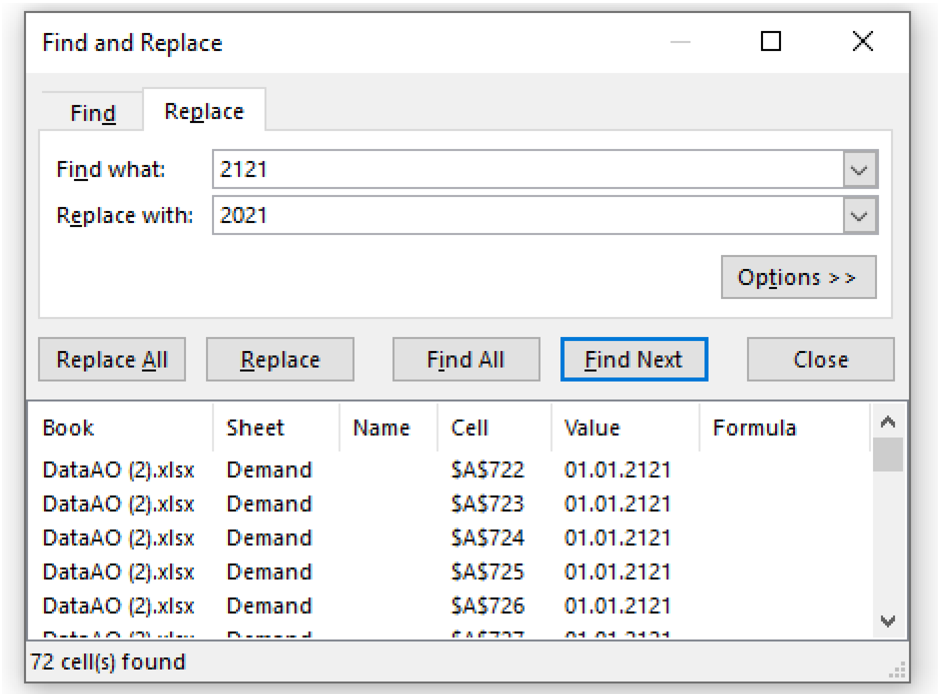
Data Analysis

The dataset provided for analysis consists of several spreadsheets containing valuable information about the demand in different regions of Serbia. It consists of:
- Map: A visual representation of the regions in Serbia.
- Retailers: This spreadsheet catalogs the retailers and their geographical coordinates and present month demand.
- Regions: A tab includes the region by its ID, name, and coordinates, coupled with the capacity details that indicate the potential volume of goods handled.
- Locations: An analysis of potential new Distribution Center (DC) locations with their coordinates, capacity, and monthly rental costs.
- Demand: The historical monthly demand data from January 2011 to October 2023 is presented on a per-region basis.

Focusing on the Demand tab, which is crucial for business insights, we need to first get this data ready for detailed analysis.

Data Cleaning

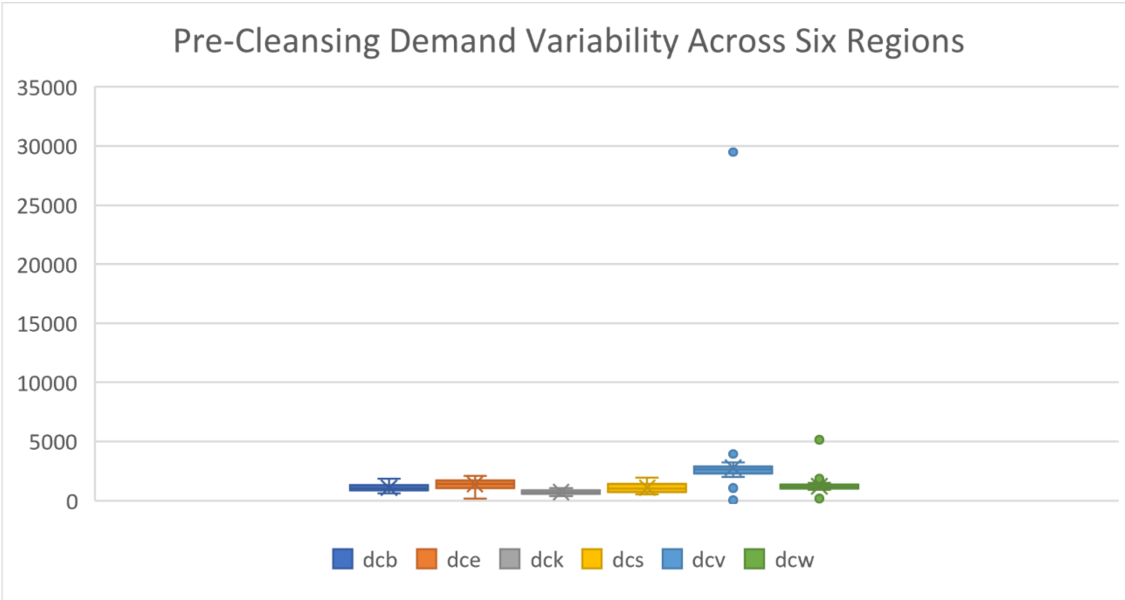
The first step is preparing the dataset for the future analysis involved converting columns to the proper data types—dates for the month-year column and numerical values for capacities. Anomalies in date formats were corrected, including resolving incorrect years such as "2121" being adjusted to "2021".



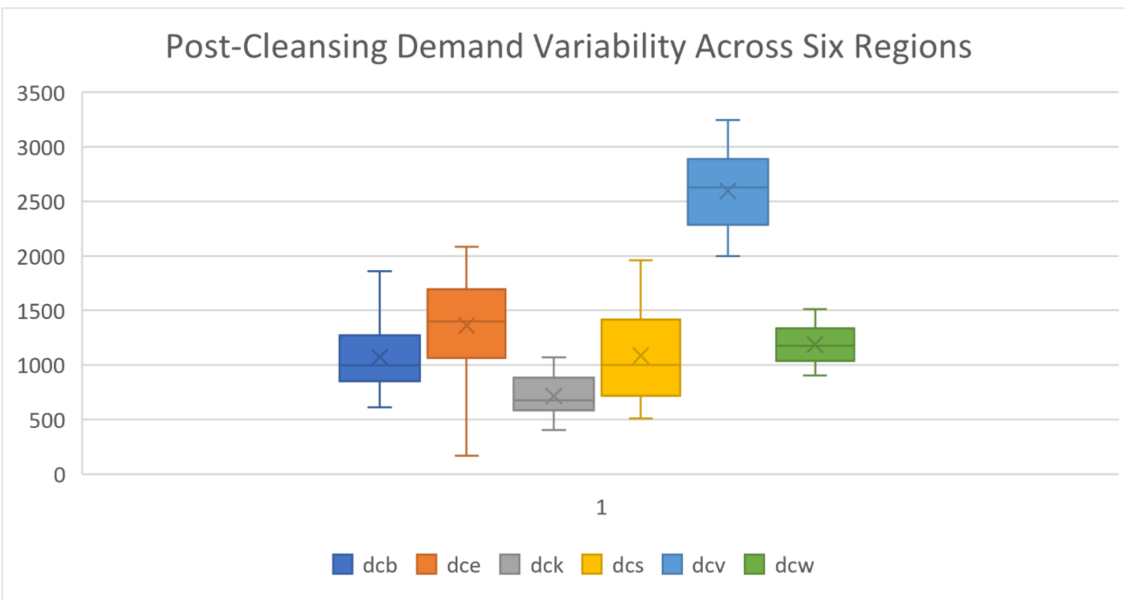
Handling Outliers

Before starting our analysis, we created a pivot table to organize the data by regions, which sets the stage for handling outliers. Box plots were constructed for each region's data to identify outliers. By manually checking and using our knowledge of the data, we corrected the obvious typos. For example, within a sequence of 1143, 1150, 1153, 5164, 1174, 1185, 1182, it was clear that 5164 was an outlier and was consequently adjusted to 1164. Similarly, the value of 29481 was deemed to be an outlier and thus altered to 2948, following the logical pattern observed in the data.

Box & whisker plot of demand data for six regions before data cleansing to identify outliers:



Box & whisker plot of demand data for six regions after data cleansing to identify changes:



Handling Missing Data

The dataset contained minimal missing data, which was addressed by substituting the missing values with median values from the respective regions. This approach was selected for its simplicity and the minimal impact it would have on the overall dataset due to the small number of missing entries.

Short-term Distribution Optimization

At the moment, the Distribution Centers are capable of satisfying the demand in total, but not per Region. To solve the distribution issues in the shorter term, KiP decided to shift Retailers from Distribution Centers in Regions with higher demand to Centers in Regions where the demand is lower, therefore balancing out the distribution system. The goal of this reorganization is to keep the distribution costs minimal.

Assumptions

In order to successfully model the required system, we made the following assumptions:

- A Retailer can be supplied from one and only one Distribution Center. For example, a Retailer cannot receive 20 pallets from Vojvodina and 13 from Belgrade, for a total of 33; they must all come from the same place.
- Truck takes up to 10 pallets to just one Retailer per trip, so we round their demand divided by 10 up to the nearest integer to get the number of trips. For example, if a Retailer demands 33 pallets, truck will go there 4 times.
- Consequently, our formula for cost per Retailer is the unit cost of a trip times distance times 2 (going there and back), times weight, which represents how many trips a truck needs to take to said Retailer to fully satisfy their demand.

Mathematical model

We are aiming to minimize distribution costs, while satisfying each Retailer's demand and staying within each Distribution Center's capacity. Therefore, our mathematical model is as follows:

$$\min \sum_{i \in R} \sum_{j \in DC} l_{ij} * x_{ij}$$

Subject to:

1. $C_j - \sum_{i \in R} x_{ij} * D_i \geq 0, \forall j \in DC$
2. $\sum_{j \in DC} x_{ij} = 1, \forall i \in R$

Where:

1. $\boldsymbol{R}$ – a set of Retailers
2. $\boldsymbol{D_i}$ – demand of i-th Retailer
3. $\boldsymbol{DC}$ – a set of Distribution Centers
4. $\boldsymbol{C_j}$ – capacity of j-th Distribution Center
5. $\boldsymbol{x_{ij}}$ - the binary variable such that $x_{ij} = 1$ iff i-th Retailer is supplied by j-th Distribution Center, and 0 otherwise (decision)
6. $\boldsymbol{l_{ij}}$ – the cost of supplying i-th Retailer from j-th Distribution Center, calculated as:

$$l_{ij} = cost * ceiling(\frac{D_i}{tC}) * 2 * \sqrt[k]{|xDC_j - xR_i|^k + |yDC_j - yR_i|^k}$$

Where:

1. ***cost*** **–** the unit cost of trip, given in the problem description
2. ***tC*** – truck capacity, given in the problem description, therefore
3. ***ceiling(D_i / tC)*** **–** is the formula for calculating the aforementioned number of trips a truck has to make to satisfy i-th Retailer's demand,
4. ***2*** **–** serves to account for the truck going to a Retailer and back
5. ***k*** – the weight for the Minkowski distance function, given in the problem description
6. ***xDC_j*** and ***yDC_j*** – coordinates of j-th DC, given in the problem description
7. ***xR_i*** and ***yR_i*** – coordinates of i-th Retailer, given in the problem description

AMPL model/data/results

In our data file, project_shortterm.dat, we define values for all the necessary parameters.

The model file, project_shortterm.mod, contains AMPL representation of the mathematical model given in the previous section.

The run file, project_shortterm.run, is simple, to allow us to extrapolate the information we want to emphasize. It prints the value of objective function, and the list of Retailers supplied by each Distribution Center.

Results

Lacking almost 250 pallets, Vojvodina was bound to lose some Retailers to other Regions. With Belgrade having a surplus of more than 600, it is no wonder many are to receive goods from it in the new system, namely rv04, rv05, rv08, rv10, rv17, rv24, rv49, rv55, rv71, rv73, rv77, rv80. Vojvodina, however, also got two from Belgrade, rb08 and rb46, most likely as a result of cost minimization.

Belgrade will also take over Retailers from other regions, rs13, rs25, rs31, rs49 from Šumadia, which was just expected as it lacks almost 150 pallets, but also many (rw10, rw15, rw19, rw29, rw31, rw33, rw34, rw38, rw39, rw43, rw45, rw47, rw50) from the West, despite their DC having a surplus of 88 pallets.

Šumadia is, however, rerouting some of its goods to Vojvodina and Belgrade, to rv02, rv66, rb25, rb29, rb39, rb41, rb48 and rb50 in particular, as well as to the Eastern region (re11, re14, re19, re20, re23, re27, re28, re30, re31, re33, re36, re39, re47, re50, re51, re56). It's central location and irregular shaping of other Regions is most likely to blame for so many shifts - minimization of costs is likely to result in bordering Retailers shifting to a Center which happens to be closer despite belonging to a different Region.

Belgrade Retailers rb24 and rb35 are moving to West Distribution Center to make use of its surplus, as well many from the Šumadia - rs03, rs05, rs06, rs11, rs12, rs14, rs18, rs20, rs22, rs26, rs30, rs32, rs34, rs37, rs38, rs39, rs42, rs43, rs50.

With the greatest shortage, Distribution Center East logically only gets a few Retailers from other Regions: rs01, rs24, rs33 from the Šumadia, rk29 from Kosovo, and, surprisingly considering they are geographical opposites, rw17 from the West.

Finally, rw08, rw11, rw32, rw35, rw48 from the West and re04, re09, re38, re41, re59 from the East will be supplied from Kosovo in the new system, along with a vast majority of Retailers from the Region itself.

Total cost of distribution with the new system in place will be 71382.4 euros. Currently, with each retailer getting supplies from its own region's Center (assuming that was possible in terms of capacity), as we have shown in Excel tab "Current cost", KiP's distribution costs would be 82621.3, so we have achieved not only full retailer demand satisfaction, but also reduction in costs.

Predicting Future Demand

The data preparation phase had already been completed in the earlier stages of our research, allowing us to proceed to analysis and forecasting without the need for additional data cleansing actions. During our data analysis, we identified seasonality in the product demand data, which became a critical factor in selecting forecasting methods.

Many conventional forecasting techniques, such as forecasting based on the last period, arithmetic mean, and moving averages for 4, 5, and 6 months, did not produce satisfactory outcomes. This issue arose because most of these techniques do not take into account seasonality, which is critically important for our case.

We also explored the Holt-Winters method, which takes into account seasonal variations in demand. While this method allowed us to develop a quite adequate model, we encountered difficulties in making accurate predictions due to challenges in determining initial values for level, trend, and seasonality.

In the end, the most effective results were obtained using the FORECAST.ETS function in Excel, which automatically considers seasonality and provided reliable forecasts of demand for the next three years. This approach enabled us to accurately capture the seasonal fluctuations in demand and achieve a high degree of accuracy in our forecasts. Thanks to this, we successfully predicted future demand, which is crucial for the next step of our research: planning for long-term distribution.

Long-term Distribution Optimization

Per the results of future demand prediction, in three year's time KiP won't have the capacity to satisfy all Retailers with just the Distribution Centers currently owned. The longer-term solution is, therefore, to rent more Centers to increase capacity. This solution should, just like the shortterm one, aim to minimize the cost of distribution, including the cost of renting the new Distribution Centers.

Assumptions

In addition to assumptions already made for the short-term solution, we are adding another one here:

- Our goal is to create a plan for KiP's distribution system in three years, namely October 2026; we are not looking to gradually add Centers as demand increases, we are not interested at what precise moment the current Centers will no longer be enough, we are, in short, not concerned with the system between now and then.

Mathematical model

We are once again aiming to minimize the cost, however this time renting new Centers must also be taken into account. The mathematical model for long term optimization is hence as follows:

$$min \sum_{i \in R} \sum_{j \in DC} l_{ij} * x_{ij} + \sum_{j \in pDC} y_j * r_j$$

Subject to:

1. $C_j - \sum_{i \in R} x_{ij} * D_i \geq 0, \forall j \in DC$
2. $\sum_{j \in DC} x_{ij} = 1, \forall i \in R$
3. $y_j * n \geq \sum_{i \in R} x_{ij}, \forall j \in pDC$

Where:

1. ***R*** – a set of Retailers
2. ***D_i*** – demand of i-th Retailer
3. ***cDC*** – a set of current Distribution Centers
4. ***pDC*** – a set of potential Distribution Centers
5. ***DC*** – a set of all Distribution Centers, union of ***cDC*** and ***pDC***
6. ***C_j*** – capacity of j-th Distribution Center
7. ***r_j*** – rent of j-th potential Distribution Center
8. ***x_{ij}*** - the binary variable such that x_{ij} = 1 iff i-th Retailer is supplied by j-th Distribution Center, and 0 otherwise
9. ***y_j*** – the binary variable such that y_j = 1 if j-th potential Distribution Center is to be rented
10. ***l_{ij}*** – the cost of supplying i-th Retailer from j-th Distribution Center, calculated in the same way as for short term optimization.

AMPL model/data/results

In our data file, project_longterm.dat, we define values for all the necessary parameters.

The model file, project_longterm.mod, contains AMPL representation of the mathematical model given in the previous section.

The run file, project_longterm.run, is simple, to allow us to extrapolate the information we want to emphasize. It prints the value of objective function, and the list of Retailers supplied by each Distribution Center that is part of the solution. We ignore the locations that will not be rented.

Results

Vojvodina, being over 500 pallets short according to the prediction, now supplies even fewer Retailers, almost exclusively its own but it did get to keep rb08 from Belgrade.

On the other hand, Belgrade will still have a significant surplus, 450 pallets, so rv04, rv05, rv08, rv10, rv17, rv24, rv49, rv55, rv71, rv73, rv77, rv79, rv80 from Vojvodina, rs13, rs47, rs49 from Šumadia, and rw10, rw15, rw19, rw31, rw33, rw38, rw39, rw43, rw45, rw47, rw50 from the West, along with many from Belgrade itself, will continue to get goods from the capital.

Šumadia gets three of Vojvodina's Retailers: rv02, rv25, rv66; four from Belgrade: rb25, rb29, rb41, rb50; quite a few from the East: re11, re14, re15, re19, re20, re23, re27, re28, re30, re33, re36, re39, re47, re50, re51, re56; along with keeping some of its own.

With shortage of almost 750 pallets, it is no wonder a new distribution will be rented in this Region. Location ndcs1 will be taking over the supply of Šumadia's rs03, rs05, rs06, rs11, rs15, rs16, rs19, rs22, rs23, rs26, rs29, rs30, rs32, rs35, rs38, rs43, and a couple from the West as well: rw14 and rw37.

Western Distribution Center, with a small shortage of around 60 pallets, will be supplying many Retailers from Belgrade and Šumadia in three years: rb02, rb03, rb06, rb18, rb19, rb31, rb35, rb39, rb49, rs12, rs14, rs18, rs20, rs25, rs27, rs31, rs34, rs37, rs39, rs42, rs50, and some of its own.

Two new Locations are to be rented in Western region too - ndcw1 which will supply exclusively Western Retailers rw04, rw07, rw09, rw12, rw18, rw21, rw22, rw25 and rw41, and ndcw2, with a mix of Retailers from Šumadia, West and Kosovo: rs10, rs24, rw02, rw03, rw05, rw08, rw11, rw17, rw32, rw35, rw48, rw49, rk03, rk05, rk28. Considering West doesn't lack too many pallets, renting these two Locations is probably a result of cost minimization, due to either locations or rent - ndcw2 for example, has the lowest cost per pallet (therefore the best cost to capacity ratio) of all new Locations.

A significant shortage of 500 pallets in Eastern Center results in it covering rs01 and rs33 from Šumadia, rk29 from Kosovo, and some of its own Retailers in three years.

With a nearly insignificant surplus of 12 pallets, Kosovo Distribution Center will keep many of its own Retailers, but also add re04, re09, re10, re38, re41, re54 from East.

Total cost of distribution will be 125114.15 euros, which is a 75% increase in comparison with short term solution. The demand, however, went up almost 14%, resulting in increase in the number of trips necessary to supply each Retailer, and transport itself is estimated to be 56% more expensive, so the resulting increase in distribution cost was to be expected, even without the cost of renting new distribution centers.

Conclusion

The optimal solution for the long-term optimization consists of Centers with 13250 total capacity, while total predicted demand is 13151 - a surplus of only 100 pallets. If demand trends continue in the same vein, another distribution optimization might become necessary quite quickly, but provided the same locations are still available, it should not be too much of an issue, since 7 remain a possibility.

However, it is difficult to say what the future might hold. Perhaps KiP will replace some of its fleet with trucks of larger capacity, rendering our solution useless. Maybe they will gain new customers, or maybe they will decide to stop supplying those with demand as low as 1, 2, or 3 pallets per month, as rising cost of transport turns them unprofitable. Whatever happens, we will be there to support their business with data analytics and optimize their distribution system for ultimate customer satisfaction and cost reduction.