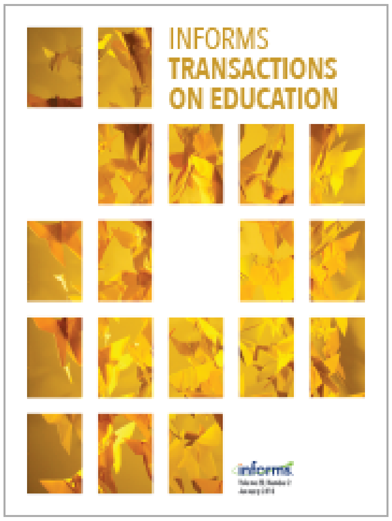


INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

An Exercise for Teaching Transportation Problem Using Spatial Data

Jyotirmoy Dalal

informs
https://pubsonline.informs.org/journal/ited

INFORMS TRANSACTIONS ON EDUCATION
Vol. 23, No. 1, September 2022, pp. 12–26
ISSN 1532-0545 (online)

An Exercise for Teaching Transportation Problem Using Spatial Data

Jyotirmoy Dalal[a]

[a] Decision Sciences Area, Indian Institute of Management Lucknow, Lucknow 226013, India

Contact: jyotirmoy.dalal@iiml.ac.in, https://orcid.org/0000-0001-5303-8043 (JD)

Received: May 30, 2021
Revised: October 9, 2021; November 24, 2021
Accepted: November 26, 2021
Published Online in Articles in Advance: February 9, 2022

https://doi.org/10.1287/ited.2021.0265

Abstract. We present an exercise for teaching the transportation problem using a mix of spatial and randomly generated data. It illustrates the potential of using qualitative and quantitative data and is suitable for undergraduate or introductory business school courses on operations research (OR), logistics, and supply chain management. It poses two challenges: (i) given the demand locations and volume, open a certain number of warehouses to ensure customer responsiveness and (ii) given those warehouses with capacity limits, determine an optimal distribution plan that minimizes the total distribution cost. This exercise is developed with the active participation of MBA students in an introductory OR course. The participants, attending the class online from different parts of India during the COVID-19 pandemic, helped generate realistic customer locations by sharing their location data. Visualizing this spatial data (after masking) in Google My Maps helps the students decide on suitable warehouse locations by considering the proximity to customers as well as diverse socioeconomic, political, and environmental factors. Then, using these warehouse and customer data, the optimal distribution plan is obtained by employing OpenSolver. Students appreciate the exposure—starting from data set generation to deriving an optimal solution—offered by this data-driven decision-making exercise.

Supplemental Material: The online supplement is available at https://doi.org/10.1287/ited.2021.0265.

Keywords: transportation problem • spatial data • data-driven decision making • teaching operations research • teaching distribution model

1. Introduction and Motivation

The transportation problem is taught at various levels of logistics and supply chain management/analytics, optimization, or operations research (OR) courses in undergraduate/graduate engineering and business school curricula as an introductory topic under distribution and network models. The applied mathematics or engineering departments generally focus on theoretical (e.g., integrality property) and methodological (e.g., network simplex algorithm, duality theory, etc.) aspects of this problem, considering it a stepping stone for teaching more advanced topics in the supply chain network design. The author's personal experience of teaching quantitative courses at business schools indicates that more than the theoretical and methodological nuances, the business students primarily appreciate the practical applicability of this exciting problem lying at the core of the transportation and logistics industry. Specifically, with the phenomenal growth in the online market demand (a significant proportion of which comprises first-time customers) in the wake of the COVID-19 pandemic situation (Baig et al. 2020), the importance of a time- and cost-effective logistical solution has never become more critical for the survival of businesses.

This paper introduces an exercise developed to encourage active learning of the transportation problem in an introductory OR course within an MBA curriculum by emphasizing data-driven optimization. It is somewhat surprising that even with the abundance of spatial data accessible through smartphone applications and Google's ubiquitous mapping tools, the opportunity to demonstrate a data-rich practical application of transportation/logistics problems in the OR classrooms is not explored enough. Although the relevant cases and pedagogical articles published in *INFORMS Transactions on Education* (Drake et al. 2011, Huggins 2019, Weltman and Tokar 2019, Wu 2020) illustrate several practical aspects for class discussion, the author needed some reasonably easy-to-implement exercise to use in an introductory OR class. The Ivey case titled "DHL Supply Chain" (Chu and

Ringrose 2012), despite being relevant to the topic, is not found suitable because the case problem involves multiple modes of transport and requires binary variables in its formulation. The author finds that the case problem in Huggins (2019) to be most relevant to the exercise discussed in this paper. Their case problem requires proximity-based characterization of inbound students at a higher education institute of the United States as "local," "regional," etc., using the students' zip code data. However, our exercise fundamentally differs from this work in two aspects. First, the case in Huggins (2019) takes the advantage of readily available latitude-longitude data for each U.S. zip code, which helps in approximating an inbound student's location. The nonavailability of such a ready-to-use database in some other countries would require building from scratch (as we do). Secondly, although the case in Huggins (2019) is helpful for analytics courses, this author wanted to conduct a suitable linear programming exercise that would provide the opportunity of using actual latitude-longitude data. The apparent unavailability of such an exercise or case motivated the author to develop this exercise.

Our exercise's unique feature is that its data set has been developed with an active engagement of the students who attended the course online from the Indian Institute of Management Lucknow, a premier public business school in India, in the academic years 2020 and 2021 because of the COVID-19 pandemic-related travel restrictions. The author's primary motivation behind developing this exercise came from recognizing the spatial dimension (the pan-India locations of the attendees), which can realistically imitate customer locations spread across the country. To this end, as a starting point of this exercise, the students are asked to capture their location data (i.e., latitude and longitude values) using some free mobile applications and share the same along with other information by filling in an online form. After some randomized masking of the inputs, four different customer demand data sets emerge corresponding to four product categories (electronics, books, apparel, and grocery).

The author breaks down the exercise into two phases. The first phase requires the students to construct a realistic customer (demand) data set. Then, given the demand points, students need to strategically locate three to five warehouses to ensure adequate customer responsiveness. To this end, students should consider several real-life factors instead of the quantitative aspects alone; this is how the exercise expands beyond a textbook topic coverage. Demand values for the four product categories are randomly generated for all customer locations and are provided as input data to the students. The warehouse capacities are also randomly generated while ensuring feasibility of the solution (see Algorithm 1 in Section 2.3). Then, with the warehouses and customer locations along with the associated supply capacity and demand data, students formulate and solve a transportation problem. Because our problem consists of more than 100 customers and three to five warehouse locations, the standard Excel Solver cannot be employed because of its 200-variable size limitation. Therefore, we use the OpenSolver to solve the problem in Excel. Moreover, to further strengthen the realistic aspect of the exercise, we use the actual driving distances instead of using the Euclidean distance between the supply demand node pairs. Because manually extracting hundreds of individual driving distances from Google Maps is cumbersome and does not add value to the exercise, the author developed a Python code routine to automate the process of populating the distance matrix using the "Open Source Routing Machine" (OSRM) application programming interface (API) (source: http://project-osrm.org/). A spreadsheet containing this distance matrix is provided to the students as the input data file for the second phase of the exercise. The execution of each step of this exercise is discussed in detail in the following section. Appendix A lists all the necessary files that we have shared in the online supplement.

Throughout the exercise, students are encouraged to use Google My Maps, wherever needed, to visualize the spatial data. At the end of this exercise, students submit write-ups with analyses and observations. The overall learning experience and student perceptions are captured by collecting their anonymous, voluntary feedback at the end of the exercise. The encouraging comments and suggestions received by the author testify to the positive contribution of the exercise toward achieving the active learning goal.

The major contributions of this paper are as follows. First, the exercise encourages students to collect and visualize spatial data, thus enhancing classroom teaching of the OR problems by establishing a deeper connection with the real world. Second, the author developed a Python code routine to automate the actual driving distance matrix population as part of this exercise development. Other instructors may adopt this code for handling a large data set while conducting similar exercises. Third, the input data files containing latitude-longitude values (shared as one of the online supplement files) can be used in any other data-driven decision-making/analytics course having a geographic information system (GIS) component.

The rest of the paper is organized as follows. In Section 2, we describe the exercise development followed by the pedagogical goals in Section 3. We include some teaching suggestions in Section 4, present detailed discussions on the participating students' profile and classroom experience in Section 5, and conclude the paper in Section 6.

2. The Exercise Development

As the exercise constitutes two phases and involves 100-odd remotely located students, a clear plan of action is essential for its successful execution. The author used online tools to instruct, collect data, clarify doubts at any step, and exchange information with the class for this exercise spanning two weeks. Figure 1, illustrating the steps in a sequence, shows the students' and instructor's actions by black and white boxes, respectively. An in-depth discussion on these steps follows.

2.1. Preprocessing for Phase-I

This step includes the important tasks for generating input data for Phase-I of this exercise. To construct the customer demand points, the locations of students remotely attending the class becomes a critical input. To this end, the author shares a Google Form with the class, requesting the students to fill in the required information. Using any free mobile GPS application (the form suggests a few names) that captures the latitude and longitude data at the phone's location, students record and enter their locations in the form. A representative sample of this form (after masking the filled-in data) is shown in Figure 2.

2.1.1. Validation of Entered Data and Masking. This step is conducted by the instructor. Note that although for our exercise only latitude and longitude values suffice to capture a location, some redundancies are deliberately included in the data collection form to cross-check correctness of the student-entered coordinate values with the help of pin code and state name. Because some students make mistakes while entering the coordinates into the form, a visualization of the entire collected data by importing the comma-separated values (CSV) file (generated by Google Form) into the Google My Maps provides a quick data validation and helps to rectify any incorrect entry. Mostly the errors involve missing the decimal points and swapping between the latitude and longitude values.

As we use the student-entered data to construct the demand data, to ensure privacy, a masking of the actual coordinates is important before circulating the

Figure 1. Schematic of the Execution Plan: Students' Actions in Black Boxes; Instructor's Actions in White Boxes

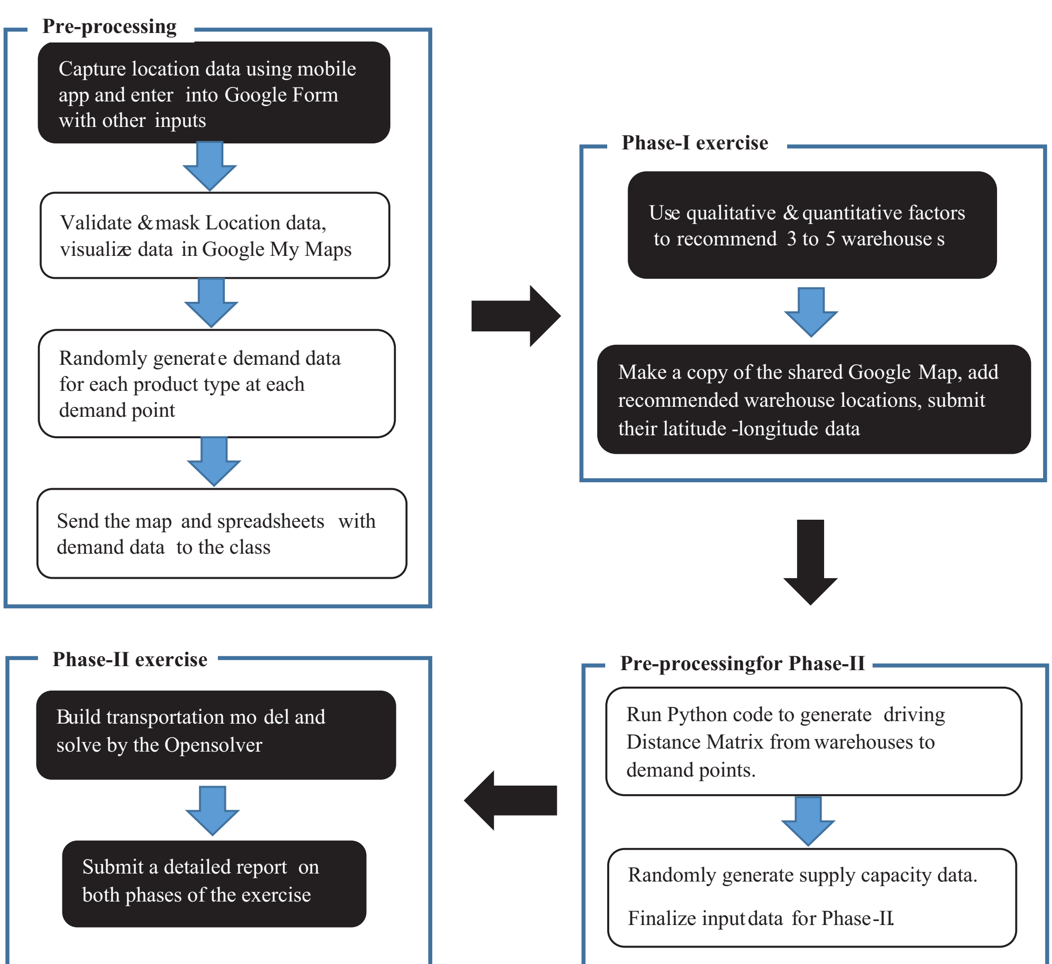


Figure 2. Sample Filled-In Form that Collects the Students' Location and Demand Data

Student location and demand data collection form

Instruction: *For a hands-on exercise development for our classroom use, I request you to fill in the form with the Latitude/Longitude of your current location, and a few other data. There are many free mobile Apps that would give you the Latitude/Longitude values of your current location. Some suggested names are: GPS Coordinates, Latitude Longitude, My GPS Coordinates, etc.*

Enter Name (optional)	XXXXX
Enter State Name	Andhra Pradesh
Enter Pincode (digits only)	517502
Enter Latitude of your current location (up to 8 decimal point precision)	13.619065
Enter Longitude of your current location (up to 8 decimal point precision)	79.324446

Rank the following product categories in your preferred order of shopping online:

	Electronics	Book	Apparel	Grocery
First Choice	X			
Second Choice			X	
Third Choice		X		
Fourth Choice				X

data set. To this end, the instructor perturbs all latitude and longitude values by adding a random fraction to the actual entries.

2.1.2. Demand Data Generation. This step is also conducted by the instructor. From the order of preference entered by students in the Google Form regarding the four product types (i.e., electronics, books, apparel, and grocery), demand values are randomly generated for each product type at each customer location. To this end, we use two steps in Excel. At the first step, because each student's product choices are captured in the CSV file (a download option in Google Form) as strings, such as "Electronics," "Books," etc., we use the nested IF function of Excel to convert the texts to numbers, capturing the order of preference. For example, if a student mentions his or her decreasing order of preference as "Electronics > Apparel > Grocery > Books," the nested IF function assigns scores of 4, 3, 2, and 1 to those product types. A sample of the specific function used in our spreadsheet is mentioned below:

$$= IF(F2 = \text{"}Books\text{"}, 4, IF(G2 = \text{"}Books\text{"}, 3, IF(H2 = \text{"}Books\text{"}, 2, IF(I2 = \text{"}Books\text{"}, 1, 0))))$$

Note that columns F to I contain those strings and the transformed numeric scores appear at the J, K, L, and M columns in our spreadsheet. Thus, if "Books" is the first choice for a student (text entry in cell F2), its score is 4; if "Books" is the second choice (text entry in cell G2), then the score is 3, and so on.

Next, considering these 1–4 scores as inputs, we use the RANDBETWEEN function of Excel within a nested IF structure and generate demand values for each product type, depending on the preference order. Specifically, for any product type, the random demand value is translated as follows: first choice → (350, 450), second choice → (250, 350), third choice → (150, 250), and fourth choice → (50, 150). A sample of the specific function is given:

$$\begin{aligned} = IF(&J2 = 4, RANDBETWEEN(350, 450),\\ &IF(J2 = 3, RANDBETWEEN(250, 350),\\ &IF(J2 = 2, RANDBETWEEN(150, 250),\\ &RANDBETWEEN(50, 150)))) \end{aligned}$$

The final input data file, named "DataSet.xlsx" (shared in the online supplement), contains the following columns: Demand Point Name, State, Pin Code, Masked Latitude, Masked Longitude, Demand-Electronics, Demand-Apparel, Demand-Grocery, Demand-Books.

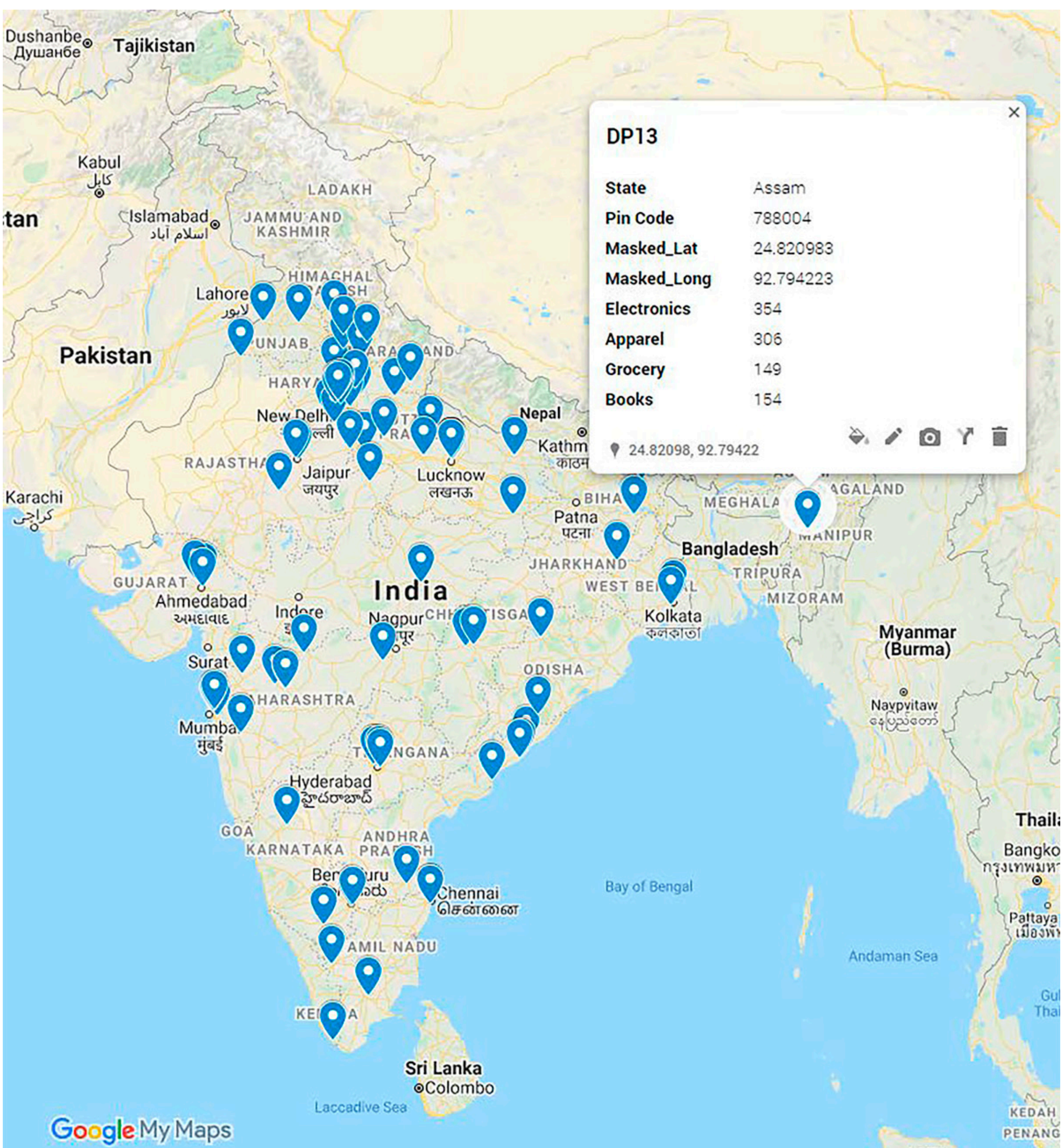
We finally obtain 101 demand points named DP1 to DP101. At each demand point, we specify the demand value for each product type. To conduct Phase-I of our exercise in two sections, each having around 64 students, we make 16 groups of 4 per section and assign one product type per group. Thus, for all groups, the customer network has identical structure; however, the underlying demand distributions may differ among the product types. The instructor circulates the “DataSet.xlsx” file and Google My Maps-based visualization of the customer network (see Figure 3) with the class as inputs for Phase-I.

2.2. Phase-I Exercise

All the steps under Phase-I are student led (see Figure 1). Each group needs to carefully study the spatial (location) and demand data. With the assumption that each customer location has a steady demand in the future, three to five warehouses should be opened to fulfill all customers’ demands.

To strategically open the warehouses, in addition to the geographic locations, students may consider various socioeconomic and political issues at a potential region, presence of transportation infrastructure, and any other relevant information. In a week’s time, each group needs to submit two deliverables: (i) a concise write-up explaining their problem-solving approach and (ii) a spreadsheet with latitude and longitude data (can be extracted from Google My Maps by putting a pin) for the suggested warehouse locations.

Figure 3. Visualization of All Customer Nodes—with Product Type-Wise Demand Value



Note that item (ii) would be used to generate the input data for Phase-II.

2.3. Preprocessing for Phase-II

This step, conducted by the instructor, requires construction of the distance matrix as an input to the transportation problem. With three to five warehouses and 100-odd demand nodes, a large number of individual pair-wise distance data are required. To emphasize on the realistic aspect of this exercise, we prefer using driving distances as opposed to the Euclidean distances between the warehouse-customer node pairs. Although it is possible to manually extract those individual driving distances from Google My Maps, the required effort undermines any value addition from the tedious task. To reduce this data preparation challenge, the author developed a Python code routine (included in the online supplement as “calcDM.py”) that reads input data from two sources: (i) “DataSet.xlsx” file containing all customer node coordinates and (ii) Excel file submitted by each team containing the proposed warehouse coordinates (naming convention: “Sec_[section name]_Group_[group number]_[product type].xlsx”). Then, using the OSRM API (source: http://project-osrm.org/), the Python code routine determines the driving distance between each warehouse-customer node pair and outputs the distance matrix in a new Excel file (naming convention: “Sec_[section name]_Group_[group number]_[product type]_DM.xlsx”). Figure 4 illustrates the overall scheme for populating this driving distance matrix.

The next step involves supply capacity generation for each warehouse. The transportation model taught in the class is presented in Appendix B. Note that we do not allow shortages, that is, all customer demands must be fulfilled (≥ constraints) by warehouses while obeying their capacity limits (≤ constraints). Balancing of the transportation model via dummy source or destination is not discussed in class; therefore, to keep this exercise straightforward, the instructor ensures feasibility of the model by making total supply capacity adequate to meet total demand. To this end, in one session, the instructor demonstrates the required steps for generating random capacity data, as explained next in Algorithm 1 (refer to Appendix B for notations). After generating capacity data following Algorithm 1, the groups start formulating the model.

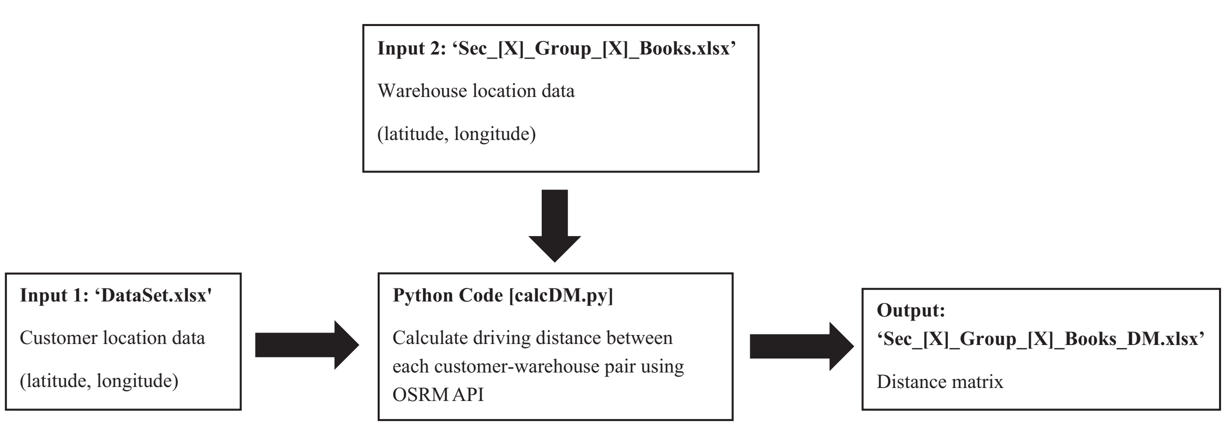
Algorithm 1 (Randomly Generating Supply Capacities)

1. Calculate total demand $D = \sum_{j \in J} D_j$.
2. Calculate base value of warehouse capacity, $\bar{S} = D/|I|$, where $|I|$ = number of warehouses.
3. Randomly set capacity of warehouse $i \in I$ as $S_i = \lceil Uniform[0.8, 1.2] \times \bar{S} \rceil$.
4. Calculate $S = \sum_{i \in I} S_i$.
5. If $S \geq D$, STOP and use current S_i values in the exercise. Else go to step 3.

2.4. Phase-II: Solving a Transportation Problem with Realistic Driving Distance Data

After the second preprocessing, each group gets one spreadsheet containing demand values (D_j) for one specific product type at the customer nodes, supply capacities (S_i) for warehouses, and driving distance matrix $[d_{ij}]_{|I| \times |J|}$. For simplicity, we assume the unit transportation cost $\alpha = 1$, thus, making transportation cost $c_{ij} = d_{ij}$. Noting the limitation of Excel Solver regarding the numbers of variables and constraints in an optimization model, we quickly demonstrate the use of OpenSolver in Excel and proceed to solve our transportation model that contains 303–505 x_{ij} variables (101 demand points and three to five warehouses). On completion of Phase-II, all groups submit a detailed report containing the optimal objective function value, flow values, and discussion on interesting observations from the strategic and tactical decisions made in Phase-I and Phase-II, respectively. Because warehouse supply capacities are not considered in Phase-I, it is expected that in Phase-II, some

Figure 4. Distance Matrix Generation Scheme—Preprocessing for Phase-II



demand points might receive supplies from distant warehouse(s).

3. Pedagogical Goals

The primary motivation behind developing this exercise is to provide students with an opportunity of hands-on OR problem implementation, where the students themselves would collect (or generate) the necessary model parameters by utilizing some readily available open source technologies. Instead of teaching the transportation model by standard textbook examples where the supply/demand nodes with their respective supply capacity/demand data, distance, or cost matrices are provided upfront, the author wanted to make the students directly involved in designing of a distribution network from scratch. Such *learning by doing* would boost the students' confidence in taking up some industry project later, where all input data would not be laid down in front of them in a textbook problem's fashion. Specifically, the pedagogical goals in this exercise are as follows:

1. Attempting a decision-making problem using freely available computing resources and self-generated realistic spatial data;
2. Experiencing hands-on OR problem solving and interpreting the optimal solution;
3. Thinking beyond the theoretical models in textbooks by considering different practical aspects;
4. Understanding the value of integrated decision making: students should recognize that locating the warehouses in Phase-I by merely focusing on their proximity to demand points may incur higher transportation cost at the end.

4. Teaching Suggestions

As a group assignment, the exercise can be used in an undergraduate engineering department's as well as in a business school's introductory OR course. Although the author used Google My Maps for the ease of sharing the visualizations with remotely located students, in a face-to-face session, use of professional GIS packages, such as QGIS (open source) or ArcGIS (by ESRI), can be more engaging. The class needs a spreadsheet program, such as Excel or LibreOffice Calc (open source), and the open source spreadsheet-based optimizer OpenSolver (Mason 2012) that can efficiently solve a linear programming model with a large number of variables (the standard Excel Solver's limit is 200 variables).

As explained in Section 2, the instructions for Phase-II of this exercise should be released after completion of Phase-I. Doing so would enable the instructor to lead a discussion at the end of Phase-II regarding the impact of supply node capacities. In Phase-I, students are encouraged to brainstorm on various qualitative factors to be considered while deciding the warehouse locations. Phase-II involves formulating and solving the transportation model in spreadsheet and interpreting of the solution obtained. Although the author does not introduce dummy source or destination node to create a balanced transportation model in this exercise, an instructor may consider adding a dummy warehouse (source) node to avoid steps 4 and 5 of Algorithm 1 (ensuring total supply $\geq$ total demand) during the preprocessing before Phase-II.

The topic can be brought to a closure by holding one session to discuss various clustering approaches adopted by the groups, to compare optimal solutions of the networks with different number of warehouses, and highlight any insightful observation. Specifically, leveraging on the flexibility offered to the students in terms of allowing three to five warehouses, a meaningful discussion can evolve by comparing the optimal solutions of groups that choose to open a different number of warehouses (see Appendix D). A group choosing three warehouses can be asked to explain their rationale behind opening less than the permitted number. If they assume that warehouse opening and operating incur significant expenses, which needs to be kept low, a brief discussion on trade-offs between fixed and transportation costs can brought up. In this context, we quote here a group's precise observation hinting at this underlying trade-off: "Some people [groups] had set up three warehouses, some did four or five. In order to take the final decision we need the cost of maintaining and establishing that warehouse."

Some discussion on the impacts of adding diverse components of the real-world logistics problems in terms of increased modeling and computational complexities can help students realize the possibilities of extending the base model in some advanced, hands-on OR/analytics courses (e.g., summer project, directed studies) that would require use of some state-of-the-art professional optimization packages. Few straightforward extensions for thinking beyond the classic model are suggested below:

1. Do not allow flow between certain warehouse-customer node pairs (introduce large penalty cost or fix corresponding $x_{ij} = 0$).
2. Consider the multicommodity setting, with different capacities for the commodities (which may also be zero) at different warehouses (although we have four product types, we have not mixed them in a single model).
3. Include multiple modes of transport with different unit transportation costs and lead times.
4. Include a minimum flow quantity constraint, that is, a flow, if happens, must be at least a threshold volume.

Figure 5. Student Background Data Collection Form

Name	...
Roll Number	...
Do you have prior work experience in industry?	Yes/No
If you have chosen "Yes" above, mention the duration (in months)	...
Is your bachelors degree in STEM (Science, technology, engineering, and mathematics)?	Yes/No
In a FIVE-point scale, express your prior exposure in quantitative subjects (1 = Very Low, 5 = Very High)	1 () 2 () 3 () 4 () 5 ()
What is your Major (bachelors degree stream)?	

Although extensions 1–3 are not difficult for the students to implement, including 4 would require auxiliary binary variables. Thus, this extension can be used as a logical transition from linear- to integer programming if both topics are covered in the same course.

5. Classroom Experience

In this section, we discuss the overall experience of using this exercise in a first-year MBA class. We highlight the students' profile, responses to the exercise questions, and feedback collected at the end of the exercise.

5.1. Student Profile

The author used this exercise as a take-home group assignment over a two-week span in two consecutive academic years. In the 2021 academic year, before assigning this exercise in two sections (approximately 64 students in each section), the author had floated a Google Form-based short survey. The student background data collection form is reproduced in Figure 5.

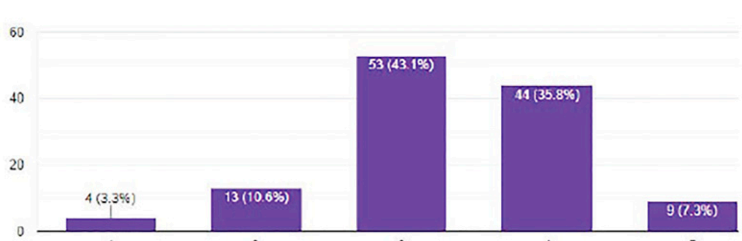
The compilation of 123 responses helps in understanding the students' prior quantitative background (Figure 6 (left)) and classroom diversity in terms of their prior academic background (Figure 6 (right)). A strong presence of STEM (almost 79%) students is observed, indicating a reasonably sound quantitative and computing skills in the class. Although 70% of the respondents indicated prior industry experience (average 32 months, maximum five years), the class had no exposure to hands-on optimization. Some STEM-background students had prior exposure to linear programming, but no student had used any optimization software before.

This exercise is assigned to the class after completing about 10 hours of classroom instruction on linear programming through textbook material, a couple of Harvard Business School cases, and hands-on practice sessions on the Excel Solver. The student groups are asked to submit two separate write-ups for Phase-I and Phase-II of the exercise. Three hours of instructions are delivered between these two phases to provide the required theoretical exposure to the transportation problem. After the final submission, the instructor summarizes main learnings and highlights interesting observations in a wrap-up session.

5.2. Students' Responses to the Exercise and Some Observations

Generally, all students find Phase-I to be challenging because they need to research and brainstorm with team members to identify various qualitative and quantitative factors that might influence the warehouse location decision. Despite not mentioning the objective function explicitly, the students can identify minimization of the demand-weighted distance/transportation cost/transportation time as the performance measure. To this end, the spatial demand data shared via Google My Maps (see Figure 3) helps them to visually identify the demand node clusters and locate the approximate centroids of those clusters as the potential warehouse locations. Although several

Figure 6. Academic Diversity: Prior Quantitative Exposure (1 = Very Low, 5 = Very High) (Left); Breakup (Right)



Stream	Percentage
Engineering	66.7
Commerce	9.8
Economics	6.5
Management	6.5
Agriculture	4.8
Science	4.1
Biotechnology	1.6
Total	**100.0**

groups cluster the customer nodes into the east, west, north, and south zones just by visual inspection, few groups leverage prior experience of some members and run the *k-means clustering algorithm* in some software with a specific k value. Figure 7 shows two representative samples from students' responses.

After clustering, some more insightful decision-making attempts are observed in terms of using the geographic information from Google Maps to identify any large city and/or transportation hub at the vicinity of the cluster centroids. If such cities or hubs exist, those are preferred by some groups over the cluster centroids for establishing warehouses to ensure better connectivity. Interestingly, a few groups prefer just the opposite and choose some remote locations close to the cluster centroids for locating warehouses, citing lower land acquisition cost, favorable state government policies for promoting remote area development, etc.

Although the above-discussed proximity-based approaches focus on the quantitative aspect, several groups include diverse real-world aspects, such as socioeconomic backgrounds, workforce availability, labor skills and cost, land, rent, electricity and other utility costs, government regulations, policies, and tax structures in the region. Obtaining actual data related to these factors and incorporating the same in a quantitative model is beyond the scope of this introductory OR course; however, the pedagogical goal regarding acknowledgment of relevant dimensions while recommending potential warehouse locations, as opposed to focusing only on the quantitative aspect of the problem, is satisfactorily addressed.

Note here that in Phase-I, students are asked to specifically focus on locating warehouses. An underlying assumption at this stage is as follows: any demand point *might be* served from a warehouse located at (or very close to) its corresponding cluster centroid. However, because of random setting of the warehouse capacities before Phase-II (while ensuring total capacity ≥ total demand), students observe that some distant warehouse(s) have to send items to some demand points because the nearest warehouse's capacity becomes inadequate. Some groups could clearly recognize that such allocations might trigger a tension between cost-effectiveness and customer dissatisfaction due to a longer lead time and articulate the same

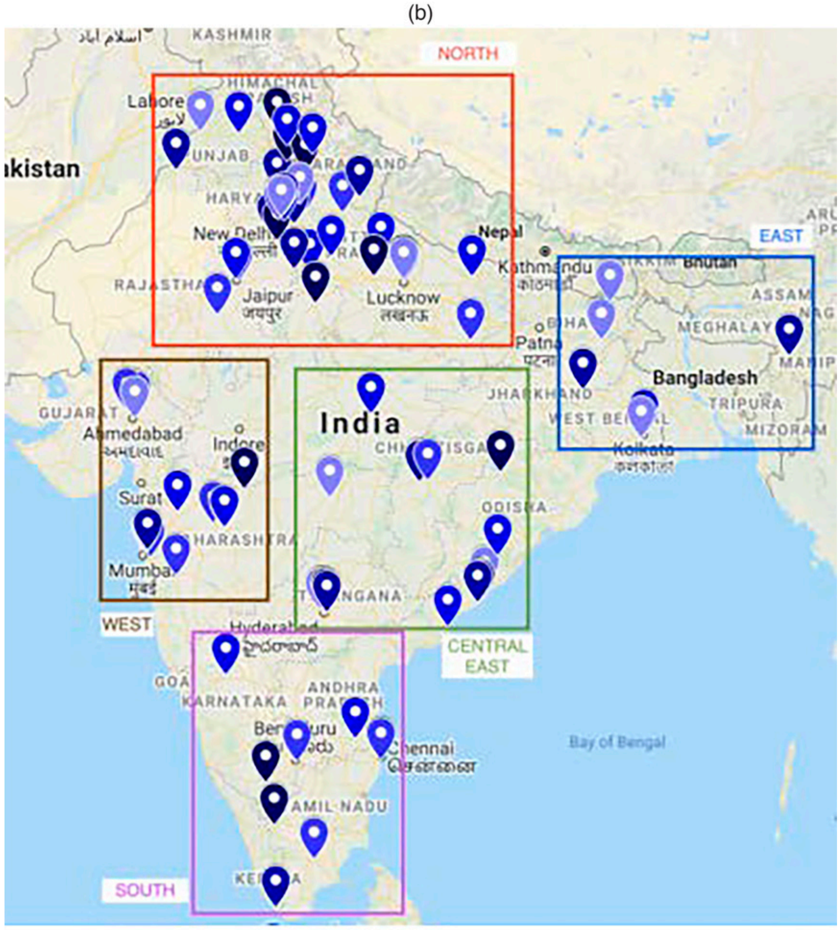
Figure 7. Clusters of Demand Points Identified by Different Groups

(a)

4 clusters

Figure 7. (Continued)

(b)



5 clusters

as follows: "*In quest of minimizing total cost, the company might end up catering a customer from a comparatively far-off warehouse against the closely viable options. Thus, every demand point is strictly catered but the individual customers might end up being frustrated due to extra time taken on the part of delivering from a farther warehouse.*"

We present a sample spreadsheet with data and the mathematical model in Figures C.1 and C.2 in Appendix C. Figure D.1 in Appendix D shows breakups of the optimal solutions obtained by four groups.

In summary, although the instructor does not explicitly discuss the nuances of location-logistics problems in the introductory OR course, through this exercise, students are able to recognize various real-world complexities that arise in the logistics network design that can significantly influence the distribution costs of goods.

5.3. Student Feedback and Suggestions for Improvement

Anonymous feedback is collected to capture the students' overall experience, major takeaways, as well as positive and negative aspects of this exercise. It is encouraging to observe that in two consecutive years, the students have given more than four on average in a five-point scale for their overall experience from this exercise. The students enthusiastically participate starting from the demand data generation process. Also, the visualization of spatial data via Google My Maps, in addition to other relevant geographical data, further strengthens the students' engagement level. The hands-on exercise gives the students an early exposure to a realistic data-driven decision-making assignment. Breakdown of the exercise into two phases and the shared learning experience from peers are also appreciated. Some representative comments in response to the question *"What are your key takeaways from this assignment?"* are quoted below:

1. "The fact that we could see the real locations of the demand points and facility locations on the map has helped us in realizing how important and effective OR methods can be in solving real issues."

2. "This is the very first continuous-assignment [probably the student means two phases] that I have done during my educational years. Breaking out an assignment in parts and building over the previous learnings. Had it been just an assignment on Transportation

problem, I am not sure the team would have analyzed the problem the way we did during part 1."

3. "The opportunity to think about a problem not only mathematically but also through a business approach considering various real parameters."

4. "The assignment simulated real life problems and compelled us to think of varied solutions. It also helped in understanding perspectives of team members from different fields of work and study."

The author also asks the participants to share the major challenge(s) faced and any suggestions for improvement. We mention here that the exercise was tested on a smaller network in the first year, when the students were required to manually extract the driving distances between warehouses and customer locations using Google My Maps. In this regard, several students of the first batch remarked the process of manually obtaining all the distance values as cumbersome, which can become impractical for a larger data set. Also, it was suggested that "a script can be written to accommodate the distance data from Google and provide to students if this assignment is repeated next year." The author incorporated this suggestion while assigning the exercise to the second batch. Currently, the distance matrix is populated by running a Python code routine and students receive a spreadsheet with preloaded distance matrix corresponding to their chosen warehouse locations. A valuable suggestion has been received regarding introduction of *storage compatibility of commodities* at warehouses; for example, electronics and furniture may be kept together but not with perishable food items. Also, inclusion of cost components for fuel, maintenance, storage, etc. have been suggested to make the problem more realistic.

6. Conclusion

This paper presents a hands-on implementation exercise that can be used for active learning in OR/analytics courses in an undergraduate engineering or introductory MBA level. Instead of formulating and solving a transportation model in a two-tier network comprising warehouse and customer nodes, we push the students to think critically and identify some relevant qualitative and quantitative issues influencing the strategic warehouse location decision. Also, by considering the students' real locations as proxies for the customer demand points, and realistic driving distances, we demonstrate data-driven decision making with (partially) real data. The modeling and solving of the transportation problem using OpenSolver while organizing data from different sources is a learning experience in itself. Moreover, visualization via Google My Maps helps in developing intuitions and validating insights of the future managers.

Acknowledgments

I thank the editor-in-chief, associate editor, and anonymous referees for comments and suggestions that helped improve the exercise design and overall organization of the paper.

Appendix A. Online Supplements

As part of this paper, the following resources are shared as online supplements.

1. An input Excel file containing customer locations and demand values ("DataSet.xlsx")

2. A sample Excel file containing proposed warehouse coordinates ("Section_1_Group_1_Books.xlsx")

3. Python code file to generate driving distance matrix ("calcDM.py")

4. An Excel file with the complete LP model ("Section_1_Group_1_Books_DM.xlsx")

Appendix B. Math Model

We present the mathematical model of the transportation problem.

Then, we formulate our mathematical model as follows:

Table B.1. List of Notations

	Description
Sets & indices	
I	Set of warehouses, $i \in I$
J	Set of demand points, $j \in J$
Parameters	
α	Unit transportation cost
d_{ij}	Distance between warehouse i and demand point j
c_{ij}	$= \alpha \times d_{ij}$, transportation cost between i and j
S_i	Supply capacity of the warehouse i
D_j	Demand value at the demand point j
Decision variables	
x_{ij}	Flow from warehouse i to demand point j

$$\text{Minimize } Z = \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij}, \tag{B.1}$$

subject to

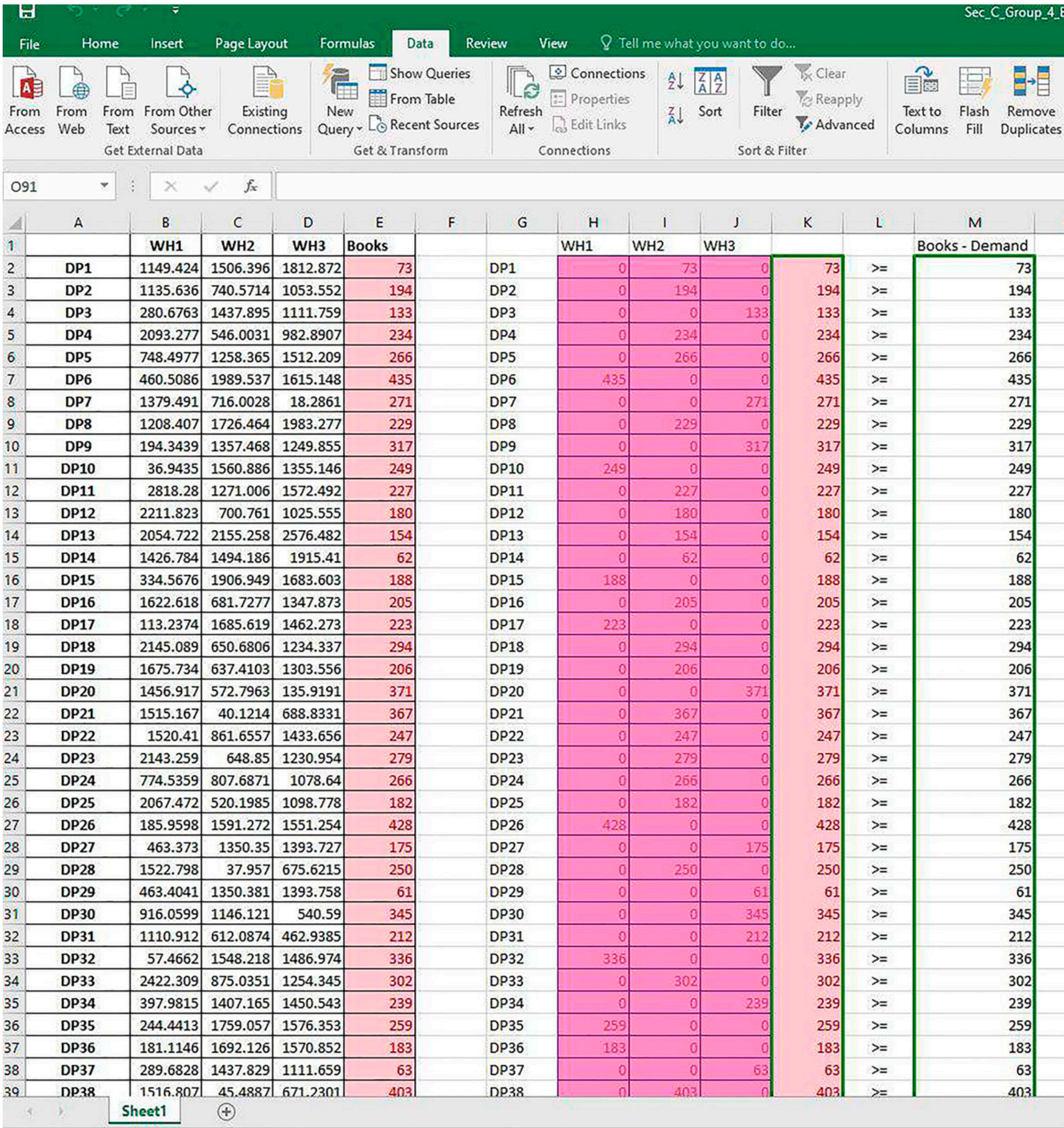
$$\sum_{j \in J} x_{ij} \le S_i \qquad \forall\, i \in I, \tag{B.2}$$

$$\sum_{i \in I} x_{ij} \ge D_j \qquad \forall\, j \in J, \tag{B.3}$$

$$x_{ij} \ge 0 \qquad \forall\, i \in I, j \in J. \tag{B.4}$$

Appendix C. Mathematical Model in Spreadsheet

Figure C.1. Part of Data and Mathematical Model in Spreadsheet



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		WH1	WH2	WH3	Books			WH1	WH2	WH3			Books - Demand
2	DP1	1149.424	1506.396	1812.872	73		DP1	0	73	0	73	>=	73
3	DP2	1135.636	740.5714	1053.552	194		DP2	0	194	0	194	>=	194
4	DP3	280.6763	1437.895	1111.759	133		DP3	0	0	133	133	>=	133
5	DP4	2093.277	546.0031	982.8907	234		DP4	0	234	0	234	>=	234
6	DP5	748.4977	1258.365	1512.209	266		DP5	0	266	0	266	>=	266
7	DP6	460.5086	1989.537	1615.148	435		DP6	435	0	0	435	>=	435
8	DP7	1379.491	716.0028	18.2861	271		DP7	0	0	271	271	>=	271
9	DP8	1208.407	1726.464	1983.277	229		DP8	0	229	0	229	>=	229
10	DP9	194.3439	1357.468	1249.855	317		DP9	0	0	317	317	>=	317
11	DP10	36.9435	1560.886	1355.146	249		DP10	249	0	0	249	>=	249
12	DP11	2818.28	1271.006	1572.492	227		DP11	0	227	0	227	>=	227
13	DP12	2211.823	700.761	1025.555	180		DP12	0	180	0	180	>=	180
14	DP13	2054.722	2155.258	2576.482	154		DP13	0	154	0	154	>=	154
15	DP14	1426.784	1494.186	1915.41	62		DP14	0	62	0	62	>=	62
16	DP15	334.5676	1906.949	1683.603	188		DP15	188	0	0	188	>=	188
17	DP16	1622.618	681.7277	1347.873	205		DP16	0	205	0	205	>=	205
18	DP17	113.2374	1685.619	1462.273	223		DP17	223	0	0	223	>=	223
19	DP18	2145.089	650.6806	1234.337	294		DP18	0	294	0	294	>=	294
20	DP19	1675.734	637.4103	1303.556	206		DP19	0	206	0	206	>=	206
21	DP20	1456.917	572.7963	135.9191	371		DP20	0	0	371	371	>=	371
22	DP21	1515.167	40.1214	688.8331	367		DP21	0	367	0	367	>=	367
23	DP22	1520.41	861.6557	1433.656	247		DP22	0	247	0	247	>=	247
24	DP23	2143.259	648.85	1230.954	279		DP23	0	279	0	279	>=	279
25	DP24	774.5359	807.6871	1078.64	266		DP24	0	266	0	266	>=	266
26	DP25	2067.472	520.1985	1098.778	182		DP25	0	182	0	182	>=	182
27	DP26	185.9598	1591.272	1551.254	428		DP26	428	0	0	428	>=	428
28	DP27	463.373	1350.35	1393.727	175		DP27	0	0	175	175	>=	175
29	DP28	1522.798	37.957	675.6215	250		DP28	0	250	0	250	>=	250
30	DP29	463.4041	1350.381	1393.758	61		DP29	0	0	61	61	>=	61
31	DP30	916.0599	1146.121	540.59	345		DP30	0	0	345	345	>=	345
32	DP31	1110.912	612.0874	462.9385	212		DP31	0	0	212	212	>=	212
33	DP32	57.4662	1548.218	1486.974	336		DP32	336	0	0	336	>=	336
34	DP33	2422.309	875.0351	1254.345	302		DP33	0	302	0	302	>=	302
35	DP34	397.9815	1407.165	1450.543	239		DP34	0	0	239	239	>=	239
36	DP35	244.4413	1759.057	1576.353	259		DP35	259	0	0	259	>=	259
37	DP36	181.1146	1692.126	1570.852	183		DP36	183	0	0	183	>=	183
38	DP37	289.6828	1437.829	1111.659	63		DP37	0	0	63	63	>=	63
39	DP38	1516.807	45.4887	671.2301	403		DP38	0	403	0	403	>=	403

Figure C.2. The Objective Function, Supply and Demand Constraints in the Spreadsheet

Get External Data | Get & Transform | Connections | Sort & Filter

X105 f_x

	A	B	C	D	E	F	G	H	I	J	K	L	M
73	DP72	36.8679	1574.003	1398.198	288		DP72	288	0	0	288	>=	288
74	DP73	38.6185	1574.671	1365.267	341		DP73	341	0	0	341	>=	341
75	DP74	275.6074	1847.989	1624.643	314		DP74	314	0	0	314	>=	314
76	DP75	275.7803	1833.447	1625.407	105		DP75	105	0	0	105	>=	105
77	DP76	1435.717	1503.12	1924.343	342		DP76	0	342	0	342	>=	342
78	DP77	463.3243	1350.301	1393.678	183		DP77	0	0	183	183	>=	183
79	DP78	463.14	1350.116	1393.494	54		DP78	0	54	0	54	>=	54
80	DP79	356.4394	1333.096	1376.474	172		DP79	0	172	0	172	>=	172
81	DP80	460.9082	1341.715	1385.092	128		DP80	0	128	0	128	>=	128
82	DP81	1505.904	45.6377	680.5111	118		DP81	0	118	0	118	>=	118
83	DP82	1524.995	42.7076	703.6285	57		DP82	0	57	0	57	>=	57
84	DP83	258.9687	1610.862	1654.239	215		DP83	215	0	0	215	>=	215
85	DP84	15.4126	1564.504	1393	435		DP84	435	0	0	435	>=	435
86	DP85	1128.749	792.8806	1082.765	346		DP85	0	346	0	346	>=	346
87	DP86	1262.452	547.4565	335.3317	189		DP86	0	0	189	189	>=	189
88	DP87	23.3577	1538.46	1429.403	386		DP87	386	0	0	386	>=	386
89	DP88	71.8274	1578.65	1498.806	375		DP88	375	0	0	375	>=	375
90	DP89	17.7974	1550.145	1372.041	70		DP89	70	0	0	70	>=	70
91	DP90	463.974	1350.95	1394.328	384		DP90	0	0	384	384	>=	384
92	DP91	31.8532	1570.617	1410.708	235		DP91	235	0	0	235	>=	235
93	DP92	412.1643	1312.178	972.5279	312		DP92	0	0	312	312	>=	312
94	DP93	706.214	1435.078	1578.488	273		DP93	0	273	0	273	>=	273
95	DP94	1241.428	739.8728	216.0887	428		DP94	0	0	428	428	>=	428
96	DP95	1196.836	1771.049	2029.187	380		DP95	0	380	0	380	>=	380
97	DP96	342.1742	1914.556	1615.027	157		DP96	157	0	0	157	>=	157
98	DP97	1263.005	550.7932	335.8843	62		DP97	0	0	62	62	>=	62
99	DP98	1798.115	417.7644	611.8468	218		DP98	0	218	0	218	>=	218
100	DP99	315.9095	1224.376	1190.925	55		DP99	0	0	55	55	>=	55
101	DP100	32.957	1569.046	1372.478	366		DP100	366	0	0	366	>=	366
102	DP101	1376.764	716.2305	19.3116	95		DP101	0	0	95	95	>=	95
103	Capacity	8548	9610	9568				8548	9610	6932			25090
104								<=	<=	<=			
105	Total Capacity	27726				Minimize		≤ 8548	9610	9568			
106	Total Demand	25090				min 14546511		binding	binding	nonbinding			
107	Excess capacity	2636											
108										Slack =			
109										2636			

Appendix D. Exhibit: Comparison of Optimal Solutions Among Some Groups

Figure D.1. Comparison of Cost Components Among Four Groups

1st Group 4:	**Objective Function Value**	**# Of warehouses**
	INR 14,889,335.00	**3**
Outflow volume from	WH-1	8455
	WH-2	8210
	WH-3	8425
Transportation cost incurred	WH-1	INR 911,515.00
	WH-2	INR 6,973,214.00
	WH-3	INR 7,004,606.00
2nd Group 8:	**Objective Function Value**	**# Of warehouses**
	INR 12,111,353.00	**4**
Outflow volume from	WH-1	7339
	WH-2	6085
	WH-3	5458
	WH-4	6208
Transportation cost incurred at	WH-1	INR 617,119.00
	WH-2	INR 4,602,186.00
	WH-3	INR 3,328,787.00
	WH-4	INR 3,563,261.00
3rd Group 12:	**Objective Function Value**	**# Of warehouses**
	INR 12,737,260.00	**4**
Outflow volume from	WH-1	6649
	WH-2	6775
	WH-3	5458
	WH-4	6733
Transportation cost incurred at	WH-1	INR 396289.01
	WH-2	INR 5680094.21
	WH-3	INR 3292994.12
	WH-4	INR 3367882.50
4th Group 16:	**Objective Function Value**	**# Of warehouses**
	INR 12,256,156.00	**4**
Outflow volume from	WH-1	5834
	WH-2	6649
	WH-3	5456
	WH-4	7151
Transportation cost incurred at	WH-1	INR 35,96,260.00
	WH-2	INR 50,34,960.00
	WH-3	INR 30,38,532.00
	WH-4	INR 5,86,404.00

All groups have assumed cost of transportation to be INR 1/km.

Note. INR, Indian rupees.

References

Baig A, Hall B, Jenkins P, Lamarre E, McCarthy B (2020) The COVID-19 recovery will be digital: A plan for the first 90 days (McKinsey Digital). Accessed February 3, 2022, https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-covid-19-recovery-will-be-digital-a-plan-for-the-first-90-days.

Chu S, Ringrose D (2012) DHL supply chain (Ivey Case). Accessed February 3, 2022, https://www.thecasecentre.org/products/view?id=108326 (requires instructor/subscriber login) and https://hbsp.harvard.edu/product/W12888-PDF-ENG (requires instructor/subscriber login).

Drake MJ, Griffin PM, Swann JL (2011) Case article–Keeping logistics under wraps. *INFORMS Trans. Ed.* 11(2):57–62.

Huggins E (2019) Case article–Converting zip code data into distances: A case study for teaching business analytics. *INFORMS Trans. Ed.* 19(2):105–107.

Mason AJ (2012) OpenSolver–An open source add-in to solve linear and integer progammes in Excel. Klatte D, Lüthi HJ, Schmedders K, eds. *Operations Research Proceedings 2011* (Springer, Berlin), 401–406.

Weltman D, Tokar T (2019) Using a Monte Carlo simulation exercise to teach principles of distribution: An enhanced version of the classic transportation problem. *INFORMS Trans. Ed.* 19(3):111–120.

Wu Y (2020) Case article–Integrating network design models for a global supply network. *INFORMS Trans. Ed.*, ePub ahead of print October 30, https://doi.org/10.1287/ited.2019.0237ca.