National College of Ireland

National College of Ireland

MSc / PgDip in Data Analytics

Release Date: Friday 21st June 2024
Submission Date: Saturday 27th July 2024

Modelling, Simulation, and Optimisation

Project CA (60%)

Project Outline

'We-Doo', the start-up company the early stage of which we investigated last year, is growing rapidly. The fundamental idea of the business is to take over the delivery services in commuter townships everywhere. Essentially, they allow the normal delivery services to drop their packages in a secure location in the township and in the evening local drivers will distribute these packages on an electric cargo bike to the end customers.

For the global delivery organisations this brings substantial savings, as all deliveries can be dropped at once in the local delivery centre. For the customer this means guaranteed secure in person delivery. A number of trials have shown that customers are willing to pay a small monthly subscription fee, while the big logistics companies are willing to outsource the last mile.

The business idea became a big success, but this creates new problems. Due to a clever social media campaign, more customers are signing up as was anticipated. As a result, one driver cannot handle the number of packages. Because of the short delivery time window, the only option is the use of two drivers. Your task is to help develop a distribution strategy using two or more driver. This requires the selection of the optimal delivery centre location.

The Problem

You are given a map of the town and a list of customers that have already signed up, an example is shown in figure 1. Further you are given a set of candidate warehouse locations, as shown for example in figure 2.

To evaluate a candidate warehouse location, you have to run a simulation of the company operating out of this location. In class we will develop a basic simulation model for the day to day operation of the company. Based on a synthetic workload the simulation will provide data about the length of the delivery route, the number of parcels delivered, and the working time of the driver. Figure 3 gives an example for the different delivery routes used over a number of days out of one delivery centre. Based on the length of the delivery route and the working time of the driver you can compute the operating costs on a day per day basis.

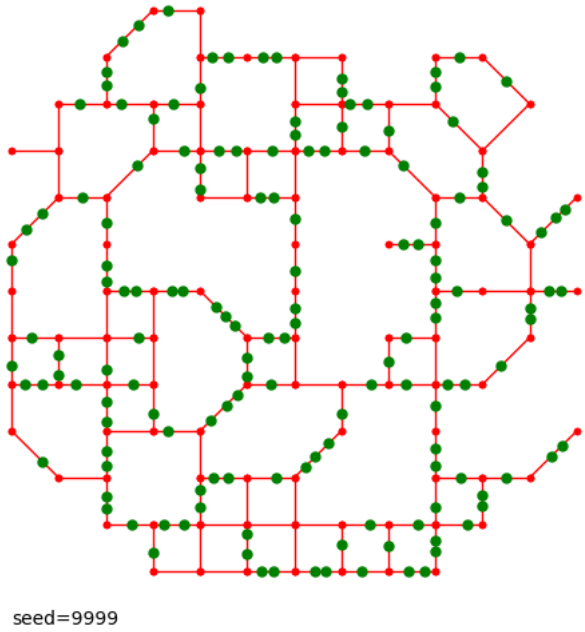

Figure 1 Sample street map with customer locations (green). The map and the customer locations will be create based on the last four digits of your student ID.

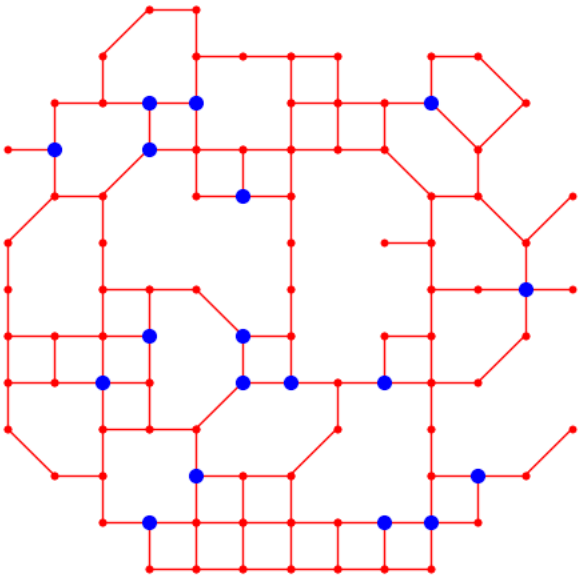

Figure 2 Candidate Warehouse locations (blue) in the town. The warehouse locations are randomly created on the map based on the last four digits of your student ID.

Your task is to recommend the delivery centre location that promises the lowest cost base. Qualify your recommendation with an investigation if the cost base for the different warehouse locations is statistically significant. For this purpose use ANOVA and F-statistics or visual aids like Box plots. If there is a visual difference, but it is statistically not significant, check if you can increase the sample size by running a longer simulation.

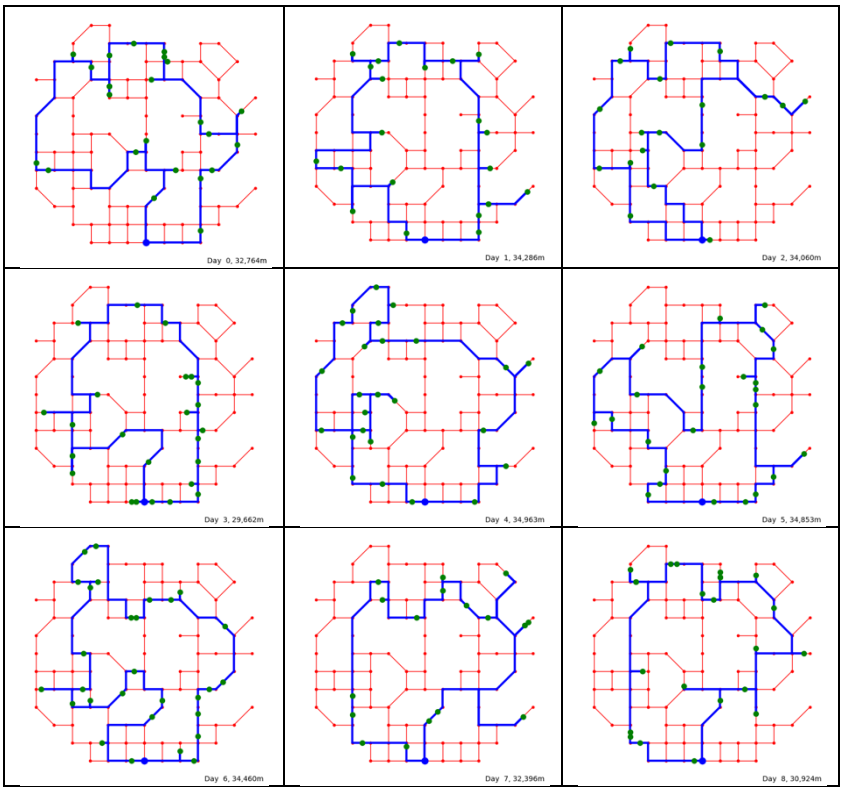

Figure 3 Optimal delivery tours out of warehouse (blue) to service the given customers (green) for a sequence of days. Needs to be updated.

Input Parameters

You are given the following information:

- A map M of the township encoded as a graph. A graph consists of a set V of vertices and a set E of edges. The vertices of the graph represent intersections, the edges represent road segments. Vertices are encoded as pairs of integers giving (x, y)-coordinates in a plane. Edges are encoded as pairs of Vertices.
- A set of customers C and their delivery addresses encoded as (x, y)-coordinates.
- A set of candidate warehouse locations $W \subseteq V$.
- The average number of parcels expected by a customer per day: $p = 0.15$.

M, C, and W are generated algorithmically. The data generation code is provided on Moodle in the Jupyter Notebook file **`Generate Map Data.ipynb`**, which you can use as the starting point for your project. To generate the parameters, call the **`generateData()`** method with an integer parameter consisting of the last four digits of your student ID.

The delivery data (i.e. the destinations of parcels to be delivered over a number of days) can be generated with **`generateDeliveryData`**() based on p, C, and the number of days n. The output of this method is a list of delivery targets per day. The delivery targets are given as indices for C. During the early stages of your work, use a small value for n, for example $n = 7$. It saves compute time. Once you get everything working increase it to $n = 20$ or larger, if required to determine statistical significance.

Simulation

The aim is to deliver all parcels arriving during day time at the distribution warehouse on the evening of the same day to the final customers. For each day you compute the required optimal delivery route for the parcels to be delivered. However this may not always be possible because the length of delivery tour may exceed the maximum range of the cargo bike. In this case parcels are left over for the following day. To avoid parcels staying for several days in the distribution centre, parcels should in general be prioritised in the sequence of their arrival, with the exception of multiple parcels for the same customer which will be delivered together. If it takes too long to compute the strict optimal shortest delivery route, you may choose a suitable heuristic (and hence a slightly suboptimal route) to speed up the simulation process. The following constraints apply to the simulation:

- The cargo bike has a maximum range of 40 km
- The time required for driving is based on the distance between way points at an average speed of 15km/h.
- The time for handing over the parcels to the customer consists of a call time (for the customer to come to the door) that follows an expovariate distribution with a mean time of 40s and an additional handover time per parcel, which again follows an expovariate distribution with a mean time of 10s per parcel.
- The cumulative preparation time (route planning and sorting of the parcels in the delivery order and packing the cargo-bike) is assumed to be 50 sec per parcel to be delivered.
- On average 10% of the customers are not at home. In this case, the parcels need to be returned to the delivery centre for next day delivery.
- The day-end procedure (reporting and setting up the cargo-bike for charging) is 10 minutes.
- The operational cost (electricity and maintenance) of the electric cargo bike is 8c/km
- The drivers are paid 30€/h with a minimum of 60€ per working day.

Deliverables

The project has two deliverables:

1. A final report in .pdf format describing your simulation study should have maximal 6 pages in IEEE conference format and follow the outline structure given below
2. A .zip file containing the code as (one or more) Jupyter Notebook files and any additional data, configuration files or documentation that may be required to run the code. The .zip file should further more contain earlier versions of your code files and the report, do prove the authenticity of your work, just in case your submission is flagged as created by AI tools.

The final report and the .zip will be uploaded through the Turn-it-in link provided on Moodle. Note that the upload limit on Moodle/Turn-it-in is 100MB uncompressed.

For evaluation criteria please check the Rubrics given at the end of this project outline. You may use components from the Simulation Study presented in class. When re-using code or design elements, refer to the version uploaded on Moodle.

Academic Integrity

- By submitting any work on Moodle you declare that this is your own work.
- Material created by others (human or AI) must be properly referenced. Verbatim text copies should be included in quotes.
- Figures included in the report should have their origin in one of the jupyter notebooks uploaded. Provide the file name and cell number as part of the figure caption. For figures not created by yourself the caption should include an acknowledgement detailing the name(s) of the creator(s) and proper references.
- Code copied from class material or other sources should be clearly marked as such and properly referenced. In particular it should not be (directly or implicitly) claimed as your own. Instead a comment should be included in the source code indicating where you obtained it from and what are your modifications, if any.
- Students are strongly advised to familiarise themselves with the Guide to Academic Integrity. All submissions will be electronically screened for evidence of academic misconduct, e.g. plagiarism, collusion and misrepresentation. Any submission showing evidence of such misconduct will be referred to the college's processes.

The Structure of the Final Report

Title

Abstract with Keywords

1. Introduction
   In this section you introduce the problem you intend to investigate and document the parameters you are using for your study as they have been generated from the last four digits of your student ID.
2. Literature Review
   To the extend you have referred to the literature for model parameters or in the evaluation section give a summary of your sources and the parameter values you have extracted. When you refer to data provided in class use as reference the material provided on Moodle.
3. Methodology
   Describe the sequence and possibly intermediate stages of your development and give references to the relevant section of code in the Jupyter Notebook files. Give details of the Model Validation you performed and the criteria you applied for the simulation runs, in particular the number of simulation runs, the load generator function used, and the statistical evaluation of the simulation results.
4. Results and Interpretation
   Report the results of your study and give an interpretation of the same. The statistical significance of the results should be discussed.
5. Reflections and Future Work
   Discuss how your research could be improved and suggest problems for future research. Check if your results are consistent with common sense. Should there be major deviations, discuss possible reasons for the same.
6. References

Rubrics

Grade Criterion	Solid H1 > 80%	H1 > 70%	H2.1 > 60%	H2.2 > 50%	PASS > 40%	FAIL < 40%
Methodology (20%)	All elements of project requirements have been thoroughly addressed. The logic of the simulation study is well presented.	All elements of the project requirements have been thoroughly addressed. Arguments have been given for the type and number of simulation runs.	Some minor requirements missing from project. No arguments have been given for the type and number of simulation runs.	Multiple omissions from the project. The proposed number of simulation runs is insufficient for statistical evaluation.	Major parts of the project are missing. The project may contain parts that are not relevant. The proposed simulation runs are insufficient for the project.	The solution bears no resemblance to the project requirements at all.
Simulation (20%)	An excellent, thorough simulation was carried out. Effort exceeds the requirements.	An excellent, fully complete simulation was carried out. The results go beyond the minimal requirement.	A very good and largely complete simulation of the required models was carried out.	A good and largely complete simulation of one model or an inadequate simulation of multiple models was carried out.	An inadequate simulation of one model was carried out. Some logical errors exist.	Little or no simulation carried out.
Code (20%)	The notebook executes without problems. Code is elegant and fully commented. The notebook is well presented. The implementation significantly exceeds the requirements	The notebook executes without problems. Code is fully commented. There is no excess code used. The notebook is well presented.	The notebook executes without problems. Code is partially commented. There is a minimal amount of excess code used.	The notebook works and allows the output to be reproduced. The notebook is poorly commented or contains a lot of excess code.	The output of the notebook file can't be reproduced. The notebook file contains minor errors, is not commented or contains a lot of unrelated code.	The notebook file doesn't contain output or contains errors that prevent it from being executed.
Evaluation & Results (30%)	Models are fully evaluated. Results are thoroughly discussed. Statistics is applied flawlessly. There is significant reflection on the challenges faced in this project and possible resolution to remaining problems	Models are fully evaluated. Results are presented and thoroughly discussed. Statistics is applied appropriately. There is significant reflection on the challenges faced in this project.	Models are evaluated based on a sufficient number of simulation runs. Results are presented and thoroughly discussed. There is very good reflection on the challenges faced in this project.	Models and results are presented and appropriately discussed. There is good reflection on the challenges faced in this project in particular related to the insufficient number of simulation runs, if this was the case.	Cursory evaluation of one model. Cursory discussion of the results. There is some reflection on the challenges faced in this project. There is no reflection on the insufficient number of simulation runs.	Little to no evaluation of model. Little to no discussion of results. There is no reflection on the challenges faced in this project
Quality of Writing (10%)	Very well written, with no language errors. All figures are well conceived and readable. The IEEE template is strictly adhered to. Report does not exceed the length limits. References are appropriately and correctly used.	Well written, with only minor language errors. All figures are well conceived and readable. The IEEE template is adhered to. Report does not exceed the length limits. References are appropriately and correctly used.	Main document has a few language and/or style errors. Figures are well presented. IEEE template and length limit are adhered to. References are complete, and correctly used.	Main document has a few language and/or style errors. Some figures are may be hard to read. IEEE template and length limit are largely adhered to. References are complete, and correctly used.	Main report is readable with some language and/or style errors. Figures may be hard to read or presented in a suboptimal manner. IEEE template may have been broken. References are mostly complete and correctly used.	Littered with typos, and/or poor use of English. IEEE template not used. Figures may be hard to read. References (if any) are incomplete or incorrect.