MGSC 404
Final Group Project
Smart City Analytics—Bike Sharing Operations

- Oral presentation: Tuesday and Thursday April 8 and 10 (in class)
- Written report: due April 11 at 11:59pm
- (Optional) Peer feedback and self-assessment form: due April 11 at 11:59pm

1 Project Goal

The goal of this final project is to apply decision analytics to a real-world problem. In particular, I am interested in your ability to generate ideas, process data, formulate research questions, develop optimization models, solve these models, and develop managerial insights. Of course, I do not expect you to fully demonstrate all of these abilities in this project, but I do expect you to demonstrate some competency in most of them.

Frequently, interesting questions start with a motivating situation. As a business leader, you will carefully consider the issues that arise in the motivating situation, develop models that address some of these issues, and analyze these models. You will then explain how the insight you develop in your analysis relates to the original situation, and also how these insights might be applicable to the real world.

Below, you will find a motivating context—bike-share in a smart city. Please use this setting to demonstrate your abilities. There is no one correct model or one correct answer. Instead, please define and solve a problem while maintaining some realism in your assumptions.

2 Motivating Context

The vision of "smart city" has been pursued worldwide to improve the efficiency, quality and sustainability of urban habitats. Boldly moving forward to a smart-city future, Montreal is undertaking major transformations in its mobility systems. In particular, Montreal is proud to embrace BIXI, a company providing bike-sharing services. With more than 500 bike docking stations deployed around the city, BIXI enables Montrealers to pick up a bike from a nearby station, ride it, and drop it off at another station to satisfy their trip demands. Being asset-light and carbon-free, this mobility mode has been rapidly gaining popularity among Montrealers.

However, a smart city cannot be built in one day, nor can a bike sharing system. BIXI is facing grand challenges at both strategical and operational levels.

At the strategical level, the company needs to decide where to locate docking stations. Given budget constraints, deploying a finite number of docking stations either too sparsely across the Montreal island or too densely concentrated in the downtown can hurt the overall accessibility of bikes. The company also needs to decide how many bikes to supply. Too few bikes in a docking station would hurt the availability of bikes to pick up, whereas too many bikes would hurt the availability of empty docks for drop-offs.

At the operational level, BIXI faces bike trip demands that are severely spatially and temporally asymmetric, driven by commuting patterns in the city. For example, before the Covid-19 pandemic, when McGill students flooded to the campus for classes, Professor Setareh Farajollahzadeh found no empty docks in the docking station outside Bronfman Building to drop off the bike. She had to rush to a farther station, only to witness that the only available dock there had just been occupied by Professor Rob Glew. However, on the late afternoon of that day, Professor Setareh Farajollahzadeh found no bikes in the nearby station, as people rode those bikes away after work.

To alleviate the supply and demand mismatch in its network, BIXI is trying to proactively reposition its bikes across locations. The company operates a fleet of trailers so that its workers can move a fraction of bikes from dock-constrained areas to bike-constrained areas. However, these repositioning operations can be costly, incurring workers wages, fuel cost, trailer depreciation, etc.

3 Task

After hearing that you'd taken MGSC 404, the board of directors of BIXI hired you as the Chief Data Scientist (with a salary of 2.5 million dollars per year) to help manage their bike-share network in Montreal. Many problems are awaiting you to tackle. For example, at the strategical level, where to deploy new docking stations? How many extra bikes to order from the bike manufacturers? At the operational level, how to reposition bikes in real time? How to set prices for customer membership? If introducing electrical bikes, when and where to charge them?

You are not expected to answer all these questions in three weeks! Instead, in this project, your primary task is to prepare a bike repositioning scheme to help BIXI alleviate the pain of supply-demand mismatch of their bikes across Montreal. For example, the CEO wonders, on a regular Wednesday in September, how many bikes should be repositioned from the docking stations in Downtown to the docking stations in Le Plateau during 8:00am - 10:00am? To complete the project follow the following steps:

3.1 Explore the data

The bike-share data sets of Divvy in Chicago contain richer information than the BIXI data sets, including station capacity and bike ID. Therefore, it is easier to instead work on those data sets and consider the same bike-repositioning question described above.

Download the data set of station information and the two data sets of origin-destination trip information from MyCourses. Import the data into a Jupyter Notebook, and display the data in the form of tables.

Meet with your group and start splitting the effort of processing data. Build intuition about how the trips are spatially and temporally distributed. Here are some points to think about to get you started, but this is not an exhaustive list (you may want to focus on other aspects of the data as well):

- Examine statistics of stations and trips, such as means, standard deviations, minimum, maximum, etc.
- Using the techniques learned in class, visualize your data in the form of tables, histograms, scatter plots, maps, etc.
- Are there any observations that display unusual behavior/outliers? Any other interesting findings?
- Use the results from your data exploration to motivate the problem you want to solve. Your results will not be assessed from the aesthetic perspective, but will be assessed in terms of clarity, richness, and the relevance to your problem. Avoid showing unclear, redundant or irrelevant results in your presentation and written report.

3.2 Formulate and solve a model

In a parallel effort, discuss with your group how to build your model as an optimization problem. While the primary task is described above, you are responsible for clearly showing the following components:

- Assumptions: What are the main assumptions you use to simplify the setting?
- Objective: What is your objective function to optimize?
- Decisions: What are your decision variables? Are there any auxiliary variables (i.e., variables that are not of your primary interest but are needed for LP formulation)?
- Parameters: What are the parameters in your formulation? For example, you may include costs and trip demands. How do you extract parameter values from the data sets, or estimate parameter values based on reasonable assumptions and justifications?
- Constraints: You cannot achieve infinite profit, zero cost or zero gap between supply and demand, given constraints in the real world. Usually, some constraints are used to model resource limits. For example, you cannot move 50 bikes away from a station if there are only 40 bikes there. Meanwhile, some other constraints are used to model the system state transition. For example, suppose that x_{ijt} is your decision of moving x_{ijt} bikes from station i to station j in time period t, and w_{it} is the variable representing the number of bikes at station i at the end of period t. Then you would need to include constraints such as

$$w_{it} = w_{i,t-1} + \sum_{j \in J} x_{jit} - \sum_{j \in J} x_{ijt} + \sum_{j \in J} y_{jit} - \sum_{j \in J} y_{ijt}$$

  for each station i and time period t, where y_{ijt} is the number of bikes transported by customers from station i to station j in time period t (this would be a parameter estimated from the data). **Note:** you do not have to include all 585 stations in your model.

You are free to build your model as you like, but you may wish to check with the professor. Please start early because it takes time to massage your model (i.e., adjusting your LP model to be reasonable and tractable).

Once you have formulated a model, code it in a Jupyter Notebook and call Gurobi to solve the problem. Present the optimal solution clearly in non-technical terms. It is advisable to use tables, summaries and graphs when applicable. The solution should be presented in a way that is understandable to management personnel that have no training in analytics.

More often than not, the process of data exploration, model formulation and coding works in an iterative and parallel fashion. In other words, start modelling before you have processed all the data; start coding a basic version of your model before further improving the model. If your model involves many decision variables, try using for-loops and the quicksum() function to set up the objective function and constraints where applicable.

3.3 Model analysis and managerial insights

After finding the optimal solution, conduct meaningful sensitivity analysis. You don't have to conduct sensitivity analysis with respect to every parameter. Instead, propose two questions that interest you, and use sensitivity analysis to answer those questions. For example, what will happen to your objective function and bike repositioning scheme if the docking station capacity scales up?

Clearly motivate your questions and present the results with tables and/or figures. Moreover, as a business leader, you should prescribe some managerial insights in non-technical language based on your results, such as how BIXI should proceed given your findings.

4 Group

Work with your assigned team. Every member of a team is required to understand every part of the project. It is the responsibility of the members to ascertain everyone in the team does a fair share of the work and has a clear understanding of the project.

Collaboration across teams is prohibited. Although collaboration was encouraged for doing lab assignments, each team must independently conduct the project without communication to students outside the team. However, you are still encouraged to consult the professor and/or the internet if any questions arise.

5 Presenting your results (deliverables)

Oral Presentation: Presentations will take place in class on Tuesday April 8^{th} and Thursday April 10^{th}. The sequence of the presentations will be randomly generated and announced in the near future. Each group will present for about 12 minutes, with 3 additional minutes for questions. Everyone in the group should contribute to the entire project, but not everyone is required to do the actual presentation. **The participation in presentation sessions is mandatory for all team members.** Consider organizing your presentation as follows:

1. Demonstrate your data-processing ability. In particular, display meaningful summary statistics and visualization (in the form of figures and/or tables). There is no need to introduce the general background of BIXI in your presentation. Your data processing results should naturally motivate the specific problem that you want to tackle.

2. Clearly describe the problem you are trying to solve. Introduce the general setting of your model, your objective, decisions to make, and main assumptions.

3. Show your LP (or IP/MIP) formulation at the end of the presentation. You are not expected to have a final solution of your optimization problem by the time of presentation (although you should have at least begun to code and solve it). You can keep improving your formulation after presentation.

Written Report (due Friday April 11^{th} at 11:59pm, submitted electronically via MyCourses): Each team should submit one report with explanation, analysis, and Python code integrated into a coherent and readable document. Please submit both the Jupyter Notebook file and a PDF copy, and a PDF of your report. The page limit for the report is 5 pages. You can put all plots, figures and tables in appendix, and refer to them in your report. There is no page limit on your appendix. However, redundant results (such as repeating sensitivity analysis for many similar parameters) will work against the clarity of your report. Consider using the following framework to organize your report.

- Introduction: Here, you provide a brief summary of the project, goals, etc.

- Main Sections: Follow the aforementioned steps to describe your results.

- Conclusion: Provide a brief conclusion of your activities and findings.

- Appendices: Not required, but can be included if you feel that some contents would fit better in an appendix than in the main text.

6 Grading

This project is worth 30% of your course grade, divided as follows:

- Presentation is worth 10%. Grading of the presentation will be based on data processing and visualization (50%) and problem definition and formulation (50%).
- Written report is worth 20%. Grading of the report will be based on data processing and visualization (30%), problem definition and formulation (30%), analysis and insights (30%), and overall presentation clarity (10%). I won't be critical about the accuracy of parameter values that are not available from the data and you have to assume, as long as they are within a reasonable range. I am mainly interested in your of data processing, model building, and ensuing analysis. Grading will be based on the richness and effectiveness of these components.

7 Tips

How to distribute work? Decision analytics is a creative and collaborative activity. Group discussion is necessary for coming up with ideas about how to analyze and interpret the data, LP formulations, etc. (not to mention debugging!). On the other hand, not every group member needs to work on all aspects of the project, and tasks can be divided among team members.

How to coordinate work?

- Use a shared folder such as Dropbox to synchronize files
- Use social network apps to facilitate group discussion
- Set milestones to achieve with deadlines
- Create a spreadsheet to track progress
- Start early