

Project — Part 1

Due date: October 2, 2025 at 11:55 PM

BIXI Montréal is a non-profit organization that manages a bike sharing system in the greater Montreal area, see <https://bixi.com/en/> for details. The first part of the project will consist in an analysis of open access data on BIXI bike rentals from the 2024 season.

Data:

The [raw data](#) consist of the records of every BIXI rental for the 2024 season. In particular, each observation consists in an individual trip and includes the following information: the start date and time, the start station and location, the end date and time, the end station and location, and the total trip duration. Only a subset of the records are retained for this analysis. Specifically, only trips under 2 hours, departing from the most popular boroughs, in the months extending from May to October, inclusively, are considered for the statistical analysis here. *It is strongly recommended to explore the raw data available on the BIXI website in order to fully understand the manipulations carried out to create the dataset used in the project.*

In addition to BIXI usage, [weather information](#) was merged with the BIXI data to provide the daily average temperature (in °C) and the daily cumulated amount of rainfall (in mm). The data you will be working with includes the following variables on each individual BIXI trip:

station	departure station name
arrondissement	borough where the station is located
lat	latitude position of station location
long	longitude position of station location
dur	trip duration, in minutes
mm	departure month, ranging from 5 (May) to 10 (October)
jj	departure day, ranging from 1 (Monday) to 7 (Sunday)
temp	average daily temperature (in °C)
prec	total amount of daily rainfall (in mm)

Important: Each team will be assigned a distinct stratified sample of the data. Be sure to work with the specific dataset assigned to your team.

Mandate:

The goal of this first part of the project is to explore the factors which affect trip duration by addressing the questions given below. Throughout, be sure that your analyses allow you to answer the business questions in an appropriate and adequate manner. Comment on findings and discuss the main takeaways from these analyses from a business perspective, providing interesting and relevant insights. Whenever a statistical model is used, be sure to

- report estimated coefficients (with units), along with uncertainty measures.
- provide relevant parameter interpretations pertaining to the business question,
- provide relevant conclusions that reflect the context,
- discuss the validity of the analysis carried out,
- discuss any shortcomings or limitations of the analysis carried out.

While you may explore several models in analyzing the data, your report should only include the most appropriate model for answering the questions below. You may use a single model to answer all of the questions. Be sure to justify your choice of model, showing only relevant output, and perform model diagnostics to assess the adequacy of the model considered.

Before beginning, carry out an exploratory data analysis. A **maximum of 2 pages** is allotted for the exploratory analysis in your report. Be sure to include only **relevant** output and findings.

Business questions:

1. Do weekend BIXI trips tend to be longer?
2. How do weather conditions affect trip duration? Is the impact of weather conditions further accentuated on weekends?
3. Are there any seasonal effects on trip duration?

Evaluation:

Each part of the project will be graded according to the following criteria:

- (a) Quality and clarity of the report :
 - the structure and presentation of the report,
 - the syntax and grammar of the writing,
 - the clarity and conciseness of the writing.
- (b) Relevance of the discussions :
 - the appropriateness of the interpretations and insights discussed in the report,
 - the relevance of the conclusions given in the report.
- (c) Correctness of the analysis :
 - the appropriateness and adequacy of the model(s) considered,
 - the validity of the interpretations and conclusions,
 - the completeness of the analyses in addressing the questions.

Submission instructions:

- This project is teamwork (minimum three, maximum four students).
- A single student should submit online through ZoneCours
- Deliverables include
 - your PDF report, **at most 7 pages** (you are encouraged to create your report using Quarto or R Markdown)
 - the **R** code (use `knitr::purl` to extract the code if necessary)
- Use the naming convention `MATH60604A_P1_id.extension`, where `id` is the HEC identifier of the student submitting the report and `extension` is one of `pdf`, `R`.
- The assignment report must include a **coverage** including the names of team members and a brief description of each team members' contribution to the work (note that the coverage does not count toward the 7 page limit).
- In carrying out the analyses, you may create new variables (e.g., variable transformations) based on your team's assigned dataset, but you cannot merge in complementary data.
- Your analyses should be **reproducible**: I should be able to run your code to obtain the same output provided in your report.
- Please be sure to follow the instructions regarding the use of generative AI detailed in the course outline.

Important remarks:

- Policy on late submissions:
 - 24 hours or less late: –15%
 - 24 – 48 hours late: –30%
 - over 48 hours late: not accepted (grade of 0)
- Any part of your report that is copied verbatim from course material, or other sources, will be considered plagiarism and given a grade of zero. Provide proper attribution of sources and citations for any reference.