

POLYTECHNIQUE MONTRÉAL

LAB WORK 2

Exploration and analyses with a Data Base

*CIV8760E : Transport Data
Management*

Teaching Assistant :
Guillaume NEVEN

Fall 2024



1 Introduction

This practical work aims to make you manipulate, design, and explore transport data using database software. This assignment will allow you to practice using software for database manipulation (SQLite or similar) with a spatial extension (Spatialite).

2 Development of a Data Model [15]

The objective of this part is to design a data model for a bike-sharing agency such as Bixi, whose management is centralized. Every trip consists of a start time t_1 at a station s_1 , ending at a station s_2 at time t_2 with a bike v by a user u . The following entities must be represented:

- Trip
- User
- Bike (please include at least: District of origin)
- Station (please include at least: latitude, longitude, District)
- District

Your model must be coherent and not miss any important element necessary for the essential functioning of the system. Please add common attributes to the different entities (e.g.: age, tire type, etc.).

1. Propose a data model in the form of an Entity/Relationship diagram involving all the entities listed above. Add attributes (indicating the identifiers) and associations between entities with their cardinalities, both minimal and maximal, as well as their functionalities. [10]
2. Translate the Entity/Relationship model into a relational model. Clearly indicate the primary and foreign keys (and what the foreign keys refer to), and propose types for the attributes. [5]

3 Descriptive Analysis of a Database [25]

For this second part, you will need to improve a database and manipulate a database using SQL.

3.1 Database [15]

For this exercise, we will use the 2023-2024 Bixi trip data available on their website. Download and combine the 2023 and 2024 datasets.

1. How could you reduce redundancy in this database? [3]
2. Implement your changes. [5]
3. Bixi stations are identified by a *Name* and their location. By which attribute should we group them? Test both ways and compare them. Comment on the differences, how to resolve them, and which method you will use. [3]
4. Create a spatial column representing the position of the station point. This element must be readable by QGIS. [2]

Please provide all SQL commands used to answer these questions (copy-paste, no screenshots).

3.2 Exploration [10]

Please answer all the following questions using SQL. When a graph is requested, you can either use the DB Browser tool or transfer the query results to another program.

1. Create a graph showing the number of daily trips over time. [1]
2. Create a graph showing the average number of trips per day of the week. [1]
3. Find the station with the highest number of trips (departures and arrivals), and suggest an explanation. [1]
4. Using QGIS, create a heatmap that represents the number of trips departing from/arriving at each station. [1.5]
5. Select a departure station of your choice, and create a heatmap of the arrival stations. Comment. [1.5]
6. Look at the ten most common trip pairs: what do you notice? Explain and provide an explanation. [2]
7. Select a frequently occurring pair (top 10%), then create a graph showing the average travel time for the pair by time of day. Do one plot for weekdays and one for weekend days. [2]

4 Report Submission

This assignment is to be done individually. Please do not exceed 10 pages for the report, and keep in mind that this is an exercise, so be concise.

Please submit the report in PDF format by October 24th at 11:59 PM on Moodle. Ensure the report is free from grammatical errors and includes accurate graphs and clear explanations. Points will be deducted for writing errors and inaccuracies in data analysis. If a language model is used (e.g., ChatGPT), you must disclose its specific usage.