Toronto Police Facilities Service

David Cardenas Ochoa

Toronto Metropolitan University

SA 8902

Table of Contents

[Project Overview 3](#_Toc152184950)

[Spatial Data Types 3](#_Toc152184951)

[Geometry 3](#_Toc152184952)

[Geography 3](#_Toc152184953)

[Projection 4](#_Toc152184954)

[Data Modeling and Database Design 5](#_Toc152184955)

[Spatial Indexing 5](#_Toc152184956)

[Normalization Process 5](#_Toc152184957)

[ER Diagram 6](#_Toc152184958)

[Tables 6](#_Toc152184959)

[Table 1. ER Diagram 8](#_Toc152184960)

[Relations between entities 8](#_Toc152184961)

[Data Dictionary 9](#_Toc152184962)

[Table 2. Data Dictionary 10](#_Toc152184963)

[Query Files 13](#_Toc152184964)

[Query 1 14](#_Toc152184965)

[Query 2 14](#_Toc152184966)

[Query 3 15](#_Toc152184967)

[Query 4 15](#_Toc152184968)

[Query 5 16](#_Toc152184969)

[Sources 18](#_Toc152184970)

# Project Overview

The escalating concern over security in Toronto underscores the critical need to comprehend the geographical distribution and efficacy of Toronto police facilities (TPS facilities) in various neighbourhoods. Five specific queries have been devised to assess the city's security landscape. These queries aim to illuminate how the city is served from a geographic perspective. An SQL Server database was developed to facilitate spatial questions analysis, incorporating data from Open Data Toronto. Before implementing the final SQL Server, an Entity-Relationship (ER) diagram was crafted, accompanied by a data dictionary, ensuring a systematic and well-organized approach to the final database structure. This process provides a robust foundation for evaluating and addressing security concerns within the city.

# Spatial Data Types

## Geometry

Geometric data is stored utilizing a planar, Euclidean coordinate system. This data type is commonly employed to retain the X and Y coordinates, depicting lines, points, and polygons within two-dimensional spaces.

## Geography

Data is stored based on a spherical coordinate system representing the Earth. Geometric data types depict two-dimensional spatial objects, with each location on Earth represented by coordinates within coordinate reference systems. Data projection between diverse coordinate systems is necessary due to the Earth's ellipsoidal shape. To accurately pinpoint object locations on the planet, employing a coordinate system that accommodates this irregular shape is crucial. In a 3-D Earth space, data is organized according to coordinate components, requiring a suitable projection to portray the Earth's curved surface on a flat plane.

Latitude and longitude undergo mathematical transformation from a spherical to a planar (XY) plane, allowing the creation of maps with precise distance measurements. This transformation facilitates the determination of area and direction. The outcome is a flattened, two-dimensional representation of the Earth, commonly called a projected coordinate system. This system employs linear units for coordinates, streamlining distance and area calculations. Through this flat projection, the traditional latitude and longitude coordinates transform x and y coordinates, facilitating more straightforward and precise spatial measurements and analyses.

The conversion of a three-dimensional geographic coordinate system to a two-dimensional flat projected system involves a process known as "Map Projection."

# Projection

When constructing a spatial database, the choice of the coordinate system, such as NAD\_1983\_UTM\_ZONE\_17N, offers several advantages. The Universal Transverse Mercator (UTM) system, specifically Zone 17 North, provides localized accuracy by dividing the world into zones with optimized projection parameters. This enables higher precision for spatial data in a specific geographic area. UTM coordinates offer a consistent unit of measurement, simplifying distance calculations within the spatial database. The UTM projection minimizes distortion, ensuring a more accurate representation of distances, angles, and areas. UTM also supports distance measurements and area calculations, making it an easy choice to construct the database around this projection. Considering all these factors, the UTM projection was used as it provided many benefits and provided the best results.

# Data Modeling and Database Design

The initial step in constructing the database involved segmenting the data into entities, ensuring uniformity across all records and fields. Identification and listing of potential candidate keys were conducted at this stage. Subsequently, a comprehensive analysis of functional dependencies was performed, and determinants were addressed to enhance data integrity. The normalization process followed, scrutinizing any determinants not qualifying as candidate keys. As a result of this meticulous process, four distinct entities were established, each with its associated attributes, laying the foundation for a structured and well-organized database.

# Spatial Indexing

The primary role of a spatial index lies in functioning as the principal filter for spatial queries, swiftly pinpointing a broad array of potential results. This initial filter significantly reduces the number of matched records that the subsequent, more meticulous secondary filter must assess. The secondary filter, responsible for delivering the precise query result, executes the specified spatial method. For example, in a spatial query such as STIntersects, the operation is performed solely on the subset of records identified by the primary filter, where each feature intersects with one or more cells in the grid. The grid cells are systematically arranged and ordered, and the spatial index entry for each feature contains a reference to the grid cells intersected by that geometry. In the context of this database, a spatial index is implemented across all tables to handle the demands of a sizable dataset efficiently.

# Normalization Process

When analyzing the files retrieved from Open Data Toronto, it was evident that they had to be deconstructed and normalized to be usable. The data was imported into ArcGIS Pro, where various functions were performed. This varied from employing procedures like “join and relate” to data to aggregate tables and “summarize within,” which merges point data with polygon data. Once the data had been organized and structured, an ER diagram was constructed relatively quickly, as the tables had already been built in ArcGIS Pro. The data dictionary was also relatively easy to make as ArcGIS had already given set data types, set null status and put primary keys. Due to the nature of the data being used, no functional dependencies were formed, but relations still existed. Going through candidate keys and creating entities helps normalize the data. The overall primary goal of the process was organizing and structuring data to minimize redundancy and enhance data integrity.

# ER Diagram

With the data provided, four entities were created: TPS, NEIGHBOURHOODS, SCHOOLS and SUBWAYSTATION. All these tables are individual but are joined through spatial relationships.

## Tables

Starting with the TPS table, there are five attributes, with one attribute being the primary key. The primary key is TPS\_ID, a unique identifier for each TPS facility within the system. It ensures that no two facilities share the same identifier. This primary key is vital for quickly accessing and referencing specific facilities when needed. The remaining entities are used to describe the facility from its name (FacilityName), the organization it belongs to (Organization), the address of the facility (address), and the facility's postal code (PostalCode). All the attributes mentioned in this section describe the TPS facilities, hence being grouped into one entity.

In the following table, NEIGHBOURHOODS, the primary key is Neighbourhood\_ID. It serves as a neighbourhood identifier distinguishing one neighbourhood from another. The four remaining attributes describe the neighbourhood’s name (NeighourhoodName), how many shootings occurred in that neighbourhood in the year 2022 (Shooting\_2022), the visible minority count in a neighbourhood (Minority), and the population density per square kilometre in a neighbourhood (Pop\_Den\_SQ\_Km).

In the second table, SCHOOLS, the primary key is School\_ID. It provides a unique number for each school in the City of Toronto. The attributes that describe this entity are Name – which is the name of the school; Board – identifying which school board the school may belong to; SchoolType – which tells if the school is a private school or a French emersion or any other variations, district – in which suburb of Toronto the school is within, address – the address of the school, and, PostalCode – the postal code of the school.

The last entity table is SUBWAYSTATION, with the primary key being Subway\_ID. The attributes in the entity are Name – which indicates the station's name, and Wheelchair – which indicates if the station is wheelchair accessible.

## Table 1. ER Diagram

A diagram of a neighborhood

Description automatically generated

## Relations between entities

As explained earlier, each entity has its primary keys and is spatially related to the other entities. The relationship can be demonstrated through the lines and symbols connecting different entities. Each relation can vary from one entity to another. There are three main symbols: ⎢, O, ∈. ⎢is meant to demonstrate one or mandatory. One way to know the difference is that if the symbol has a symbol on each side, it means compulsory, and if there is a symbol on one side and an entity on the other, it means one. O and ∈ each has one meaning, with the first symbol meaning many and the latter meaning many.

The relationship between TPS and Neighbourhood is “mandatory one to mandatory one.” Each TPS facility must be associated with one neighbourhood (mandatory) and can only be related to one neighbourhood. Still, one neighbourhood does not need to have a TPS facility; if it does, many can belong in one neighbourhood.

The relationship between SubwayStations and Neighbourhoods is “mandatory one to optional one.” This means that a subway station must be associated with a neighbourhood and can only belong to one neighbourhood, but a neighbourhood does not need to have a subway station. If it does, it can only be assigned to one neighbourhood.

The last relation is between School and Neighbourhood: "mandatory many to optional many.” This explains that each school must be connected to a neighbourhood, but there can be many schools in a neighbourhood. The other portion signifies that a neighbourhood does not have to be associated with a school; if it is, it can be related to many.

# Data Dictionary

Organizing the data, generating diverse entities, and forming relationships via primary keys pave the way for the subsequent phase of building a database. During this stage, the data dictionary acts as a centralized document, offering comprehensive details about the data employed in a database. It functions as a guide, providing insights into the data and elucidating its meaning, structure, and utilization within a database or system.

## Table 2. Data Dictionary

TPS (TPS\_ID, FacilityName, Organization, Address, PostalCode, LATITUDE, LONGITUDE, Shape)

* TPS is the location of existing Toronto Police Services facilities

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Column Name** | **Data Type (Length)** | **Key** | **Null Status** | **Default Value** | **Remark** | **Description** |
| TPS\_ID | Smallint | Primary Key | Not Null | None |  | Unique identifier for each TPS facility |
| FacilityName | nvarchar(55) | No | Not Null | None |  | The name of the facility |
| Organization | nvarchar(23) | No | Not Null | None |  | What organization the facility belongs to |
| Address | nvarchar(25) | No | Not Null | None |  | Address of the TPS facility |
| PostalCode | nvarchar(7) | No | Not Null | None | Format: A#A #A# | The postal code of the TPS facility |
| LATITUDE | numeric(23,15) | No | Not Null | None |  | The latitude coordinates of the TPS facility |
| LONGITUDE | numeric(23,15) | No | Not Null | None |  | The longitude coordinates of the TPS facility |
| Shape | Geometry | No | Not Null | None |  | Geometry file of the TPS facilities |

NEIGHBOURHOODS (Neighbourhood\_ID, NeighbourhoodsName, Shooting\_2022, Minority, Pop\_Den\_SQ\_Km, Shape)

* Neighbourhoods is a geographic boundary that can be any area within the City of Toronto

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Column Name** | **Data Type (Length)** | **Key** | **Null Status** | **Default Value** | **Remark** | **Description** |
| Neighbourhood\_ ID | numeric(38,8) | Primary Key | Not Null | None |  | Unique identifier for each neighbourhood |
| NeighbourhoodsName | nvarchar(80) | No | Not Null | None |  | The name for neighbourhoods |
| Shooting\_2022 | numeric(38,8) | No | Not Null | None |  | Shooting count in a neighbourhood in the year 2022 |
| Minority | numeric(38,8) | No | Not Null | None |  | Minority population per neighbourhood |
| Pop\_Den\_SQ\_ Km | numeric(38,8) | No | Not Null | None |  | Population density per kilometer square |
| Shape | geometry | No | Not Null | None |  | Geometry file of the neighbourhoods |

SCHOOLS (School\_ID, Name, Board, SchoolType, Address, PostalCode, District, Shape)

* Schools is the location of existing schools in the City of Toronto

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Column Name** | **Data Type (Length)** | **Key** | **Null Status** | **Default Value** | **Remark** | **Description** |
| School\_ID | numeric(18,0) | Primary Key | Not Null | None |  | Unique identifier for each school |
| Name | nvarchar(80) | No | Not Null | None |  | The name of each school |
| Board | nvarchar(80) | No | Null | None |  | The district school board the school belongs to |
| SchoolType | nvarchar(80) | No | Not Null | None |  | The type of school |
| Address | nvarchar(80) | No | Not Null | None |  | The address of the school |
| PostalCode | nvarchar(80) | No | Not Null | None | Format: A#A#A# | The postal code of the school |
| District | nvarchar(80) | No | Not Null | None |  | The inner suburb that the school belongs to |
| Shape | geometry | No | Not Null | None |  | Geometry file of the school’s location |

SUBWAY (Subway\_ID, Wheelchair, Name, Shape)

* Subway is the location of existing subway stations in the City of Toronto

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Column Name** | **Data Type (Length)** | **Key** | **Null Status** | **Default Value** | **Remark** | **Description** |
| Subway\_ID | numeric(38,8) | Primary Key | Not Null | None |  | Unique identifier for each subway station |
| Wheelchair | nvarchar(20) | No | Not Null | None |  | If the subway station is wheelchair accessible |
| Name | nvarchar(254) | No | Not Null | None |  | Name of the subway station |
| Shape | geometry | No | Not Null | None |  | Geometry file of the subway station |

The tables show that many variables were used and will be explained from the top tables to the bottom right to left, with duplicated values skipped in the explanation. Each entity shown in the ER diagram gets its table, with the table name written out in all capitals with the attributes used inside the brackets. Underneath the table name is a description of each entity and the requirements for an attribute to be included in the entity. All the attributes used are in the column name section with their respective tables. The first attribute is always the primary key, as that is the unique identifier for the table. There were various data types used. ‘Smallint’ represents that the value for that attribute must be a whole number. Nvarchar(#) uses Unicode characters to store multilingual and special characters, with the # sign indicating how many characters can fit in the field. Numeric(#,#) is used for numbers, with the first # showing how many number characters can be used and the latter how many decimals are to be used. Format ‘A#A #A#’ means that when the postal code field is being filled, it must follow the structure of letter, number, letter, number, and can only have six characters. The last data type used was geometry, which stores spatial data representing geometric shapes, such as points, lines, and polygons, within a two-dimensional or three-dimensional space. Null status helps identify if an attribute can be null in a record, and default values determine the default value if no existing variables are inputted; in this database, no data values were used.

# Query Files

As mentioned earlier, five spatial queries were constructed to answer security questions for the city of Toronto. The most used commands in the query are select, from, order by, group by, sum, asc, desc, STIntersect and STDistance. The select statement retrieves data from one or more database tables. It specifies which columns you want to retrieve from. The ‘from’ clause is used in SQL to identify the source table or tables from which data should be retrieved. ‘Order By’ is used to sort by ascending (asc) and descending (desc). ‘Group By’ aggregates and summarizes data. ‘STIntersect’ determines if two spatial objects intersect, while ‘STDistance’ calculates the shortest or longest distance between two spatial objects, depending on the parameters set.

## Query 1

How many TPS facilities are within a 1km buffer to a subway station for quick response time?

SELECT Count(FacilityName)

FROM TPS, Subway

ORDER BY tps.shape.STIntersects(subway.shape.stbuffer(1000)) = 1

This SQL query will illustrate the presence of 19 police stations located within a 1-kilometre buffer distance from every subway station.

## Query 2

Which 5 TPS facilities have the least population density within a 1.5km buffer?

SELECT TOP 5 sum(Pop\_Den\_SQ\_Km), FacilityName

FROM TPS, Neighbourhoods

WHERE tps.shape.stintersects(neighbourhoods.shape.stbuffer(1500)) = 1

GROUP BY FacilityName

ORDER BY sum(Pop\_Den\_SQ\_Km) asc

The following SQL query displays TPS facilities with the lowest population density within a 1.5-kilometre buffer. Arranged from least dense to most dense, the facilities include 43 Division, North Collision Reporting Centre, East Collision Reporting Centre, 33 Division, and Toronto Police College.

## Query 3

Which TPS facility has the highest count of visible minorities within a 2km buffer?

SELECT TOP 1 TPS\_ID, sum(minority), FacilityName

FROM TPS, Neighbourhoods

WHERE tps.shape.stintersects(neighbourhoods.shape.stbuffer(2000)) = 1

GROUP BY TPS\_ID, TPS.FacilityName

ORDER BY sum(minority) desc

This SQL query reveals that 52 Division is the TPS facility with the most minorities.

## Query 4

What are the top 5 TPS facilities with the highest shooting count in 2022 within a 3km buffer?

SELECT TOP 5 sum(Shooting\_2022), FacilityName

FROM TPS, Neighbourhoods

WHERE tps.shape.stintersects(neighbourhoods.shape.stbuffer(3000)) = 1

GROUP BY FacilityName

ORDER BY sum(Shooting\_2022) desc

The presented SQL query indicates the top 5 TPS facilities with the highest number of shootings in 2022 within a 3-kilometre buffer, listed from the most to the least: 31 Division recorded 67 shootings, followed by 52 Division with 65 shootings, Headquarters with 64 shootings, 12 Division with 61 shootings, and 51 Division with 59 shootings.

## Query 5

Which five schools are the furthest away from a TPS Facility?

WITH SchoolDistances AS (

SELECT s.School\_ID, s.Name AS SchoolName, t.TPS\_ID, t.FacilityName, t.LATITUDE AS TPS\_LATITUDE, t.LONGITUDE AS TPS\_LONGITUDE, s.Shape.STDistance(t.Shape) AS Distance

FROM SCHOOLS s CROSS JOIN TPS t )

SELECT TOP 5 sd.School\_ID, sd.SchoolName, sd.TPS\_ID, sd.FacilityName,

sd.TPS\_LATITUDE, sd.TPS\_LONGITUDE, sd.Distance

FROM SchoolDistances sd  
ORDER BY sd.Distance DESC;

This SQL query is designed to identify the top 5 schools farthest from a TPS Facility, ordered from the farthest to the closest. Queen’s Collegiate Secondary School is located 3.69 kilometres away from the 43 Division, West Rouge Junior Public School is situated 3.66 kilometres away from the 23 Division, West Rouge Junior Public School is positioned 3.65 kilometres away from Toronto Police College, William G Davis Junior Public School is situated 3.64 kilometres away from the 23 Division. Joseph Howe Senior Public School is located 3.62 kilometres away from the 23 Division.

# Sources

Open Data Toronto: <https://open.toronto.ca/>