



SISTEMA DE TRANSPORTE  
COLECTIVO

**CDMX**  
CIUDAD DE MÉXICO



# **Influx of passengers in *Sistema de Transporte Colectivo Metro* of Mexico City**

Fabrizio Alejandro Sánchez García

January 2024

## **Summary**

This work was carried out as an academic project under the Professional Certificate in *Data Analysis* offered by *Google*. The data used throughout this project is public and open. The main objective of this project is to distribute information to the public and in particular to Metro users, through visualizations and statistics that allow them to understand how their access to the STC network generates history over the years.

Last, but not least, the contents themselves made through the Data Analysis process allowed the knowledge acquired by the author to be put into practice throughout the Professional Certificate.

## **Data collection**

The data used in this project was obtained from the Mexico City Open Data Portal.

These are organized in long format, because the seasons appear in multiple rows for each day and year

The data is reliable because it comes from a first-hand government source that, in the first instance, should not have imprecise or incomplete numbers. They are original because they come from the Mexico City Open Data Portal, so the metro passenger flow counts are well accounted for through the turnstiles. They are comprehensive because they contain all the critical information needed to answer the passenger flow question and find a solution to system oversaturation. They are current because they cover from the year 2010 to the year 2023. They are cited because they come from SEMOVI, as of January 24, 2024, the last update of the data set for simple and broken down daily traffic was on January 19 of 2024.

The data is for public use and is open so the authorization comes from the data portal itself. The data does not contain privileged data or explicit information about people or entities, which is why the data can be processed without harming the privacy of third parties. They are accessible because this open data portal provides not only mobility information, but also data on the environment, security, health, tourism, among others.

## **Details to consider in the data set**

There is a particular situation for this data analysis. On July 11, 2022, Metro Line 1, which included the section from Pantitlán to Salto del Agua, was partially closed, reopening on October 28, 2022.

The STC counts the entries to the stations through a system in the turnstiles, so clearly during the closure of the first section of line 1, in the data set all the stations involved have recorded zero entries in total, since 11 July 2022 until October 28, 2023. Resuming the accounting of entries on October 29, 2023.

On Monday, May 3, 2021, an accident occurred on one of the elevated sections, near the Olivos station. One of the STC trains collapsed along with one of the parts of the elevated section, leaving the service for Line 12 inoperative. In particular, the stations that were inoperative were those between Tláhuac and Tezonco, leaving the Periférico Oriente section open to Mixcoac.

The data set for Line 12 counts zero revenue for all stations (even those not affected by the closure) from May 4, 2021 to January 14, 2023. As of January 15, 2023, it is already They account for the income to the stations from the stations from Periférico Oriente to Mixcoac.

The closure of both sections of the respective metro lines required the incorporation of a support service provided by the Passenger Transport Network (RTP). The RTP passenger influx data set is available, but in this set, in the Type of service field, in the “Metro L1 Support” category, only records from September 2023 are counted. (A on the date of January 25, 2024, clarification on this missing data was requested from the original data collectors).

And to account for the influx of passengers on metro line 12 in the closed section, it is not explicitly clarified what type of service is assigned for this part. (Likewise, as of January 25, 2024, clarification was requested from the original compilers).

The existing problem with the data in summary is the following: there are some stations in the entire Collective Transportation System network that record zero income due to different problems that occurred in the metro. To account for these revenues, we need to use the RTP data set, although it does not account for how

many passengers boarded at any particular station. Therefore, we will have to be careful when we want to count influx by stations rather than by the Line itself. It will be necessary to combine the data from both sets, from the STC and the RTP.

We must clarify that the data set on broken down traffic from the STC and the RTP account for income from cards, tickets and free tickets, both. The STC starts on January 1, 2021 and the RTP starts in January 2022.

There is another data set called Afluencia Simple STC that only counts passenger flow. This counts from January 1, 2010.

This set shows a somewhat larger problem because there are stations that record zero income to the stations, for reasons that we can investigate later. For example, the Metro offered free access temporarily after the Earthquake that occurred on September 19, 2017, one-day closures due to technical problems, etc.