YAT Logo	DEPI Logo	Project logo

Mobility in Cairo project

By

Abeer AbdelNaeem

MennaTullah Reda Mohamed Mahmoud Atta

Mohammad Hosny

Mostafa Nasr

Saeed Magdy

Project supervisor

Dr. Amal Mahmoud

2025

Abstract

Smart Cairo Mobility is a data analytics project focused on optimizing Cairo's urban transportation through integrated analysis of ride-hailing, driver, customer, fuel, and metro data. It applies advanced analytics to reveal mobility patterns, efficiency gaps, and sustainability trends. The project provides actionable insights for improving transport operations and forecasting demand. By linking private and public transport systems, it supports data-driven mobility planning. Its outcomes align with Egypt's vision for smart and sustainable urban development.

Table of Contents/Index

Introduction

Project rationale

Cairo's transportation network contains multitudes of complexity. Between public and private ownership, formal and informal modes of transportation, fluctuating energy prices, and other variables, the ability of a citizen to move from point A to point B within greater Cairo is determined by numerous variables beyond their control. This complexity was the main driver behind pursuing this project.

Project objectives and expected outcomes

- What we want to describe
- What we want to understand
- What we predict
- What we prescribe and to whom (recommendations to relevant stakeholders)

This project is a Transportation and Metro Data Analysis web application that focuses on studying public transportation usage, especially metro systems, before and after the fuel price increase. The project analyzes trip distances, user behavior, and traffic volume changes. The application provides visual insights and analytical dashboards based on real-world transportation datasets. The system is designed for data-driven decision-making and offers scalable insights for urban planning and transportation efficiency.

- Compare Usage Trends: Analyze changes in metro and transportation usage before and after the fuel price increase.
- Distance Analysis: Study how trip distances vary across time periods and different transport modes.

- Public Transport Insights: Provide detailed insights into metro usage patterns including peak hours, station popularity, and route preferences.
- Policy Impact Assessment: Evaluate how fuel price changes impacted user decisions and transportation behavior.
- Data Visualization: Offer clear and interactive visualizations for better understanding of complex data.

Project Background

- General description of Cairo's transportation infrastructure

Public transportation

Private transportation

Key challenges

- Review of similar research/analytical projects, highlighting the limitations of other projects, and explaining the contribution of our research project to the literature, as well as the provision of actionable insights

Project Methodology

- Dataset description

The dataset was created by combining XX dataset with XX dataset. The datasets were procured from the following websites:

The dataset contains 5 sheets, with the following number of rows, columns, and data types:

Sheet name	# of rows	# of columns	Types of data
Trips	1000	10	• Numeric • String • Date and time
Customers	200	6	
Drivers	100	6	
Metro ridership	4053	3	
Fuel prices	20	4	• Numeric • Date and time

- Dataset preprocessing

A. Data cleaning

<u>Data cleaning in Excel</u>

- Detecting duplicates. No duplicates were detected.
- Transforming inconsistent data types

<u>Data cleaning in SQL</u>

For the Customers table

- Detecting null values. No null values were detected.
- Detecting duplicates. No duplicates were detected.
- Detecting outliers using IQR and STD. No outliers were detected.

For the Drivers table

- Detecting null values. No null values were detected.
- Detecting duplicates. No duplicates were detected.
- Detecting outliers using IQR and empirical rule. No outliers were detected.

For the Fuel prices table

- Detecting null values. No null values were detected.
- Detecting duplicates. No duplicates were detected.
- Detecting outliers using IQR and empirical rule. No outliers were detected.

For the trips table

Additionally, copies were made for each table to be able to retrieve the data if necessary.

Data cleaning in Python

1. Data loading

- Libraries used: pandas, numpy, matplotlib, and seaborn
- Loaded dataset using pd.read_excel()
- Used head() to explore data structure and column names

2. Data cleaning

- Checked for missing data using df.isnull().sum()
- Handled missing data appropriately (filled or removed)
- Merged Excel sheets in Python

B. Data modelling

Data modelling in Excel

Data modelling in SQL

- Added ID in Fuel prices table to use as a foreign key in the Trip fact
- Filled null values of Fuel prices ID
- Joined the dates of the trips and fuel price

- Modified IDs in all of the remaining tables to be used as foreign keys in the Trip fact

Data modelling in Python

Data modelling in Tableau

Data modelling in Power BI

- Programs used [Excel - SQL - Python - Tableau - Power BI]

Program name	Why and how it was used
Excel	Preliminary cleanup, analysis, and visualization of the dataset
SQL	Advanced data analysis
Python	Advanced data analysis and rudimentary visualization
Tableau	Advanced data analysis and visualization
Power BI	Advanced data analysis and visualization

- Analysis process
    - Explaining the linkages made between the different data sheets and types across the 5 programs [Example: focusing on customer mobility patterns and preferences in Excel, linking the fuel prices to the fare in SQL, etc.]

        Examples: linking car models to fuel type to better understand the fares
    - Including screenshots that showcase these linkages in the different programs and the visualizations we're proud of
    - Another suggestion could be to structure this section according to the four types of data analysis, so we'd start with the descriptive linkages, followed by

the diagnostic and predictive linkages, and end with the prescriptive linkages. That being said, it's important to differentiate between this section and the following one. Here, we say that we did the linkages to learn/gain specific insights. The following section will explain the collected insights.

Results and Discussion

- Outline the descriptive, diagnostic, predictive, and prescriptive insights we gained because of our analytical efforts
- Link the insights - whenever applicable - to the literature and offer reasons for the insights when they can't be explained by the data. The example that comes to mind was when Dr. Amal asked us why certain types of cars were more common than others, and we couldn't explain why. In this instance, we'll need to do extra desk research to explain the insight.

Recommendations

- Rewrite the prescriptive analysis as recommendations for different stakeholders, such as ride-hailing companies, drivers, customers, the National Transportation Authority, etc.

References

- Include the bibliography used for the literature review and analysis

Appendices

- Showcase coding efforts from SQL and Python