**Traffic Data:**

* Contains details about traffic patterns such as day, date, time, vehicle counts, vehicle type, payment method, location, lane, and weather conditions.
* **Primary Key:** ID

**Violations Data:**

* Contains details about toll violations, including violation ID, date, time, vehicle type, payment status, location, lane, and whether a violation occurred.
* **Primary Key:** Violation ID
* **Foreign Key:** Traffic ID (links to ID in traffic data)

**Passenger Flow Data:**

* Captures information about bus departures, passengers per bus, date, time, and location.
* **Primary Key:** Bus ID
* **Foreign Key:** Traffic ID (links to ID in traffic data)

**Joining the Tables:**

* The first INNER JOIN merges TrafficData with ViolationsData on the Date column. Only records with matching dates are included.
* The second INNER JOIN combines the result from the first join (CombinedTrafficViolations) is then joined with passenger\_data on the Date column. This final join ensures that only records with matching dates across all three tables are included.

**Handling Nulls and Duplicates:**

* **Null Handling:** Since we used INNER JOIN, there are no null values in the columns coming from the joined tables because only matching records are selected.
* **Removing Duplicates:** Ensure that any duplicate records are handled by using the DISTINCT keyword.

**Outcome**

The final combined dataset includes all traffic data, includes with details of violations and passenger flow metrics. This allows for comprehensive analysis, including identifying patterns, correlating traffic and violations, and understanding passenger trends.

**Code used for generating the final data set using mysql:**

**1. Join traffic\_data with violations\_data**

-- Join Traffic Data with Violations Data on Date

SELECT

traffic\_data.\*,

violations\_data.ViolationID,

violations\_data.Time AS ViolationTime,

violations\_data.VehicleType AS ViolationVehicleType,

violations\_data.PaymentMethod AS ViolationPaymentMethod,

violations\_data.Location AS ViolationLocation,

violations\_data.Lane AS ViolationLane,

violations\_data.Violation

FROM

traffic\_data

INNER JOIN

violations\_data

ON

traffic\_data.Date = violations\_data.Date; **2. Join the Result with passenger\_data**

-- Join the Result with Passenger Data on Date

SELECT

CombinedTrafficViolations.\*,

passenger\_data.BusID,

passenger\_data.Time AS PassengerTime,

passenger\_data.Departures,

passenger\_data.PassengersPerBus,

passenger\_data.Location AS PassengerLocation

FROM

(SELECT

traffic\_data.\*,

violations\_data.ViolationID,

violations\_data.Time AS ViolationTime,

violations\_data.VehicleType AS ViolationVehicleType,

violations\_data.PaymentMethod AS ViolationPaymentMethod,

violations\_data.Location AS ViolationLocation,

violations\_data.Lane AS ViolationLane,

violations\_data.Violation

FROM

traffic\_data

INNER JOIN

violations\_data

ON

traffic\_data.Date = violations\_data.Date

) AS CombinedTrafficViolations

INNER JOIN

passenger\_data

ON

CombinedTrafficViolations.Date = passenger\_data.Date;

**3. To Remove Duplicates:**

SELECT DISTINCT \* FROM FinalCombinedDataset;