Case Study

Fast Rent a Car is interested in optimizing the utilization of their car fleet across various rental locations. The company has noticed some locations consistently run out of cars while others seem to have an excess. The aim is to minimize downtime and maximize utilization of the vehicles.

You have been given two datasets: Vehicle_Rentals.csv and Locations.csv.

Vehicle_Rentals.csv contains:

Rental_ID (unique ID for each rental)

Vehicle_ID (unique ID for each vehicle)

Location_ID (unique ID for each location)

Start_Date (Start date of the rental)

End_Date (End date of the rental)

Locations.csv contains:

Location_ID (unique ID for each location)

Location_Name (Name of the location)

City

Country

Consider that:

- The Vehicle_Rentals.csv dataset has one rental per row.
- The Start_Date is the pickup date and the End_Date is the dropoff date, consider both dates inclusive (i.e., if a certain vehicle has a start date of 01/01/2023 and end date of 07/01/2023, consider that it was rented for 7 days).
- The fleet size of a store can be determined by the number of different Vehicle_ID in that specific location.
- To calculate the average number of rental days, you must divide the total number of rental days by the number of bookings (different Rental_ID).
- A specific vehicle was being rented on 10/05/2023 if there's a row in the dataset where:
    - Start_Date is lower or equal 10/05/2023.
    - End_Date is higher or equal 10/05/2023.
    - If both conditions above are not met, consider that this specific vehicle was not being rented on this date.

- The utilization rate is the index used to determine whether a certain store is with the correct fleet size. To calculate this ratio, you must divide the number of days of rented fleet by the total days of available fleet. Consider that the ideal utilization rate is between 78% and 86%. If it's lower than 78%, the store has too many cars. If it's higher than 86%, it needs more fleet.

Tasks:

1. Load the datasets into a database (for the sake of the test, SQLite database can be used).
2. Inspect the data for any quality issues (missing data, incorrect data, outliers, etc.), perform necessary cleaning using Python.
3. Using SQL, generate a report that shows the number of rentals per location.
4. Using Python, determine the utilization rate of the fleet per location.
5. Using a data visualization tool of your choice (e.g., Matplotlib, Seaborn, Tableau), create a dashboard that shows:
    a. The number of rentals over time.
    b. The distribution of rentals across locations.
    c. The utilization rate per location.
    d. The average number of rental days per location.
6. Based on your analysis, identify any locations that have a car surplus or shortage and answer the questions below:
    a. What are the busiest and slowest locations in terms of vehicle rentals?
    b. Based on the current utilization of vehicles, do any locations frequently run out of cars or have an excess?
    c. What data-driven recommendations would you make to ensure an efficient distribution of cars across all locations?

Remember, the main goal of this exercise is to evaluate your technical skills in Python and SQL, your ability to clean and manipulate data, create meaningful visualizations, and derive insights that can drive decision-making. Thus, remember to provide the scripts, queries, and visual dashboards you've created during this process.