Border Crossing Analysis | HubbleMind

Project Overview

This SQL internship project focuses on the analysis of border crossing data between the U.S.-Canada and U.S.-Mexico borders. Interns will engage in tasks that involve data exploration, aggregations, and advanced SQL queries, simulating real-world data analysis scenarios.

Dataset : Download | Data Source : USA Government (data.gov)

Dataset Description

The dataset contains information about inbound crossings at various ports, detailing the type of crossing (e.g., trucks, buses, pedestrians) and the number of vehicles or individuals recorded. The relevant columns are:

- **Port Name**: The name of the port where the crossing was recorded.
- **State**: The U.S. state where the port is located.
- **Port Code**: A numeric code associated with the port.
- **Border**: Indicates whether the crossing is on the US-Mexico or US-Canada border.
- **Date**: The date of the crossing (formatted as DATE).
- **Measure**: Describes the type of crossing (e.g., trucks, buses, pedestrians).
- **Value**: The number of vehicles, containers, passengers, or pedestrians recorded for the crossing.

Week 1: Data Exploration and Basic Queries

Tasks:

1. List all distinct Port Names and their corresponding States.
2. Count the total number of unique Borders and the total number of entries associated with each Border.
3. Retrieve the total number of entries (crossings) for each year, sorted from most recent to oldest year.
4. Find all ports that have recorded more than 5000 crossings for the Trucks measure type.
5. Identify the top 3 states with the highest total number of pedestrian crossings.
6. For the year 2023, extract the total number of crossings per month, categorized by measure type.
7. Find which measure type is the most frequently recorded for each state.
8. Generate a summary report showing the total number of crossings for each measure type, grouped by border.

Week 2: Intermediate Queries with Aggregations

Tasks:

1. For Texas, calculate the average number of crossings per month for each measure type.
2. Find the port on the U.S.-Canada border with the highest number of crossings. Include the measure type and total crossings.
3. Calculate the total number of crossings for the "Buses" measure type in each state, ordered by total crossings in descending order.
4. For the U.S.-Mexico border, calculate the average and total number of crossings for each port in the year 2022.
5. List all ports that reported pedestrians as a crossing measure in 2023 and show their total number of pedestrian crossings.
6. Extract the total number of crossings in each border for every year available in the dataset.
7. Identify the month in 2023 with the highest number of truck crossings. List the month and total truck crossings.
8. List the top 5 ports with the highest crossing activity (all measure types) in 2021, showing the measure type and total crossings for each port.

Week 3: Advanced Queries and Research-Like Tasks

Tasks:

1. Calculate the total number of crossings for each year from 2019 to 2023, grouped by border and measure type.
2. For Texas, find the most frequently recorded measure types for 2023. Rank the measure types by the number of entries without using RANK().
3. Compare the total number of container crossings over the last 3 years for each border.
4. Identify the busiest month of each year (2020-2023) in terms of pedestrian crossings. Show the year, month, and total pedestrian crossings.
5. Compare the total number of truck crossings in 2021 and 2022 at the top 5 busiest ports for trucks. Display both years' totals side by side.
6. Find the port with the lowest total crossings on the U.S.-Canada border for any measure type in 2023.
7. List the monthly total number of crossings for buses across all states in 2022, sorted in ascending order.
8. Display the sum and average number of crossings for each state, grouped by measure type and year. Only show entries where the average crossings exceed 500.

Week 4: Reporting, Documentation, and Portfolio Building

Tasks:

1. **Create a final summary report that includes:**
    - A description of each query written over the past three weeks.
    - Key findings and insights from the dataset, such as:
        - The busiest ports and states for crossings.
        - Patterns and trends by border (U.S.-Mexico vs U.S.-Canada).
        - How the number of crossings varies by measure type, year, and month.
2. **Document all the SQL queries used in a single file for review and portfolio building.**
3. **Optionally, create visualizations** using tools like Google Data Studio or Tableau to display insights such as:
    - Total crossings per year and border.
    - Most common measure types by state.
    - Monthly trends in truck and pedestrian crossings.

Submit your work

Submit the document via the provided **Google Form**

Importing Data into MySQL

1. Create a new schema and name it as "**Cross_Border_Crisis**".

2. Import the dataset:

- Download the dataset from Google Drive.
- Open MySQL Workbench.
- Select the Table Data Import Wizard under the Server menu.
- Choose the downloaded CSV file and follow the wizard steps to import the data into a table named "**Border_Crossing_Entry_Data**".