**HUBBLEMIND SQL Internship**

Project Name: border\_crossing\_entry\_data

All SQL Queries Document

-- Creating Schema and Table.

**CREATE DATABASE IF NOT EXISTS** Cross\_Border\_Crisis;

**USE** cross\_border\_crisis;

-- Creating border\_crossing\_entry\_datatable

**CREATE TABLE** border\_crossing\_entry\_data(

Port\_Name VARCHAR(50),,

State VARCHAR(50),,

Port\_Code int,

Border VARCHAR(50),

Date CHAR(10),

Measure VARCHAR(50),

Value int

);

**SELECT** \***FROM** border\_crossing\_entry\_data;

-- Then I imported data using Data Import Wizard in MySQL workbench

**SELECT** \* **FROM** border\_crossing\_entry\_data;

-- list all distinct port names and their corresponding states.

**SELECT DISTINCT** port\_name, state **FROM** border\_crossing\_entry\_data;

-- count the total number of unique Borders and the total number of entries associated with each Border.

**SELECT COUNT**(**DISTINCT** Border) **AS** Total\_Unique\_Borders, **COUNT**(\*) **AS** Total\_Entries

**FROM** border\_crossing\_entry\_data

**GROUP BY** Border;

-- Retrieve the total number of entries (crossings) for each year, sorted from most recent to oldest year.

**SELECT YEAR**(Date) **AS** year, **SUM**(Value) **AS** total\_entries **FROM** border\_crossing\_entry\_data

**GROUP BY** YEAR(Date)

**ORDER BY** year **DESC**;

-- find all ports that have recorded more than 5000 crossings for the Trucks measure type.

**SELECT** Port\_Name **FROM** border\_crossing\_entry\_Data

**WHERE** Measure = 'Trucks'

**GROUP BY** Port\_Name

**HAVING SUM**(Value) > 5000;

-- Identify the top 3 states with the highest total number of pedestrian crossings.

**SELECT** State, **SUM**(Value) **AS** total\_pedestrian\_crossings **FROM** border\_crossing\_entry\_data

**WHERE** Measure = 'Pedestrians'

**GROUP BY** State

**ORDER BY** total\_pedestrian\_crossings **DESC**

**LIMIT** 3;

-- For the year 2023, extract the total number of crossings per month, categorized by measuretype.

**SELECT DATE\_FORMAT** (Date, '%Y-%m') **AS** month, Measure, **SUM**(Value) **AS** total\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** YEAR(Date) = 2023

**GROUP BY** month, Measure

**ORDER BY** month, Measure;

-- Find which measure type is the most frequently recorded for each state.

**SELECT** State, Measure, **COUNT**(\*) **AS** Frequency

**FROM** border\_crossing\_entry\_data

**GROUP BY** State, Measure

**HAVING COUNT**(\*) = (

**SELECT MAX**(sub\_count)

**FROM** (

**SELECT** State, Measure, **COUNT**(\*) **AS** sub\_count

**FROM** border\_crossing\_entry\_data

**GROUP BY** State, Measure

) **AS** subquery

**WHERE** subquery.state = border\_crossing\_entry\_data.state

)

**ORDER BY** State;

-- Generate a summary report showing the total number of crossings for each measure type, grouped by border.

**SELECT** Border, Measure, **SUM**(Value) **AS** total\_crossings **FROM** border\_crossing\_entry\_data

**GROUP BY** Border, Measure

**ORDER BY** Border, total\_crossings **DESC**;

-- For Texas, calculate the average number of crossings per month for each measure type.

**SELECT** Measure, **AVG**(Value) **AS** average\_crossings\_per\_month **FROM** border\_crossing\_entry\_data

**WHERE** State = 'Texas'

**GROUP BY** Measure;

-- Find the port on the U.S.-Canada border with the highest number of crossings. Include the measure type and total crossings.

**SELECT** Port\_Name, Measure, **SUM**(Value) **AS** total\_crossings **FROM** border\_crossing\_entry\_data

**WHERE** Border = 'US-Canada border'

**GROUP BY** Port\_Name, Measure

**ORDER BY** total\_crossings DESC

**LIMIT** 1;

-- Calculate the total number of crossings for the "Buses" measure type in each state, ordered by total crossings in descending order.

**SELECT** State, **SUM(**Value) **AS** total\_bus\_crossings **FROM** border\_crossing\_entry\_data

**WHERE** Measure = 'Buses'

**GROUP BY** State

**ORDER BY** total\_bus\_crossings **DESC**;

-- For the U.S.-Mexico border, calculate the average and total number of crossings for each port in the year 2022.

**SELECT** Port\_Name, **AVG**(Value) **AS** average\_crossings, **SUM**(Value) **AS** total\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** Border = 'US-Mexico border' **AND** YEAR(Date) = 2022

**GROUP BY** Port\_Name

**ORDER BY** total\_crossings **DESC**;

-- List all ports that reported pedestrians as a crossing measure in 2023 and show their total number of pedestrian crossings.

**SELECT** Port\_Name, **SUM**(Value) **AS** total\_pedestrian\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** Measure = 'Pedestrians' **AND** Year(Date) = 2023

**GROUP BY** Port\_Name

**ORDER BY** total\_pedestrian\_crossings **DESC**;

-- Extract the total number of crossings in each border for every year available in the dataset.

**SELECT** Border, Year(Date) **AS** Year, **SUM**(Value) **AS** total\_crossings

**FROM** border\_crossing\_entry\_data

**GROUP BY** Border, Year(Date)

**ORDER BY** Border, Year;

-- Identify the month in 2023 with the highest number of truck crossings. List the month and total truck crossings.

**SELECT DATE\_FORMAT**(Date, '%Y-%m') **AS** Month**, SUM**(Value) **AS** total\_truck\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** YEAR(Date) = 2023 **AND** Measure = 'Trucks'

**GROUP BY DATE\_FORMAT**(Date, '%Y-%m')

**ORDER BY** total\_truck\_crossings **DESC**

**LIMIT** 1;

-- 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.

**SELECT** Port\_Name, Measure, **SUM**(Value) **AS** total\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** YEAR(Date) = 2021

**GROUP BY** Port\_Name, Measure

**ORDER BY** total\_crossings DESC

**LIMIT** 5;

-- Calculate the total number of crossings for each year from 2019 to 2023, grouped by border and measure type.

**SELECT** YEAR(Date) AS year, Border, Measure, **SUM**(Value) **AS** total\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** YEAR(Date) **BETWEEN** 2019 **AND** 2023

**GROUP BY** year,Border, Measure

**ORDER BY** year, Border, Measure;

-- For Texas, find the most frequently recorded measure types for 2023. Rank the measure types by the number of entries without using RANK().

-- Since you can't use the RANK() function, we'll use COUNT() and ORDER BY to achieve the ranking.

**SELECT** Measure, **COUNT**(\*) **AS** entry\_count **FROM** border\_crossing\_entry\_data

**WHERE** State = 'Texas' **AND** YEAR(Date) = 2023

**GROUP BY** Measure

**ORDER BY** entry\_count **DESC**;

-- Compare the total number of container crossings over the last 3 years for each border.

**SELECT** Border, YEAR(Date) **AS** Year, **SUM**(Value) **AS** total\_container\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** Measure **LIKE** '% Containers %' **AND** YEAR(Date) **BETWEEN** YEAR(CURDATE()) - 2 **AND** YEAR(CURDATE()) - 0

**GROUP BY** Border, YEAR(Date)

**ORDER BY** Border, Year;

-- Identify the busiest month of each year (2020-2023) in terms of pedestrian crossings. Show the year, month, and total pedestrian crossings

**WITH** PedestriansMonthlyTotal **AS** (

**SELECT** YEAR(Date) **AS** year, MONTH(Date) **AS** month, **SUM**(Value) **AS** total\_pedestrian\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** Measure = 'Pedestrians' **AND** YEAR(Date) **BETWEEN** 2020 **AND** 2023

**GROUP BY** YEAR(Date), MONTH(Date))

**SELECT** year, month, total\_pedestrian\_crossings **FROM** PedestriansMonthlyTotal

**WHERE** total\_pedestrian\_crossings = (

**SELECT MAX**(total\_pedestrian\_crossings)

**FROM** PedestriansMonthlyTotal **AS** subquery

**WHERE** subquery.year = PedestriansMonthlyTotal.year)

**ORDER BY** year, month;

-- 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.

**WITH** TruckCrossings **AS** (

**SELECT** Port\_Name, YEAR(Date) **AS** year, **SUM**(Value) **AS** total\_truck\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** Measure = 'Trucks' **AND** YEAR(Date) **IN** (2021, 2022)

**GROUP BY** Port\_Name, YEAR(Date)),

Top5Ports **AS** (

**SELECT** Port\_Name, **SUM**(total\_truck\_crossings) **AS** total\_crossings

**FROM** TruckCrossings

**GROUP BY** Port\_Name

**ORDER BY** total\_crossings **DESC**

**LIMIT** 5)

**SELECT** t.Port\_Name,

**SUM**(**CASE WHEN** t.year = 2021 **THEN** t.total\_truck\_crossings **ELSE 0 END**) **AS** total\_truck\_crossings\_2021,

**SUM**(**CASE** **WHEN** t.year = 2022 **THE**N t.total\_truck\_crossings **ELSE 0 END**) **AS** total\_truck\_crossings\_2022

**FROM** TruckCrossings t

**JOIN** Top5Ports tp **ON** t.Port\_Name=tp.Port\_Name

**GROUP BY** t.Port\_Name

**ORDER BY** total\_truck\_crossings\_2021 + total\_truck\_crossings\_2022 **DESC**;

-- Find the port with the lowest total crossings on the U.S.-Canada border for any measure type in 2023.

**SELECT** Port\_Name, Measure, **SUM**(Value) **AS** lowest\_total\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** Border = 'US-Canada Border' **AND** YEAR(Date) = 2023

**GROUP BY** Port\_Name, Measure

**ORDER BY** lowest\_total\_crossings **ASC**

**LIMIT** 1;

-- List the monthly total number of crossings for buses across all states in 2022, sorted in ascending order.

**SELECT** YEAR(Date) **AS** Year, MONTH(Date) **AS** Month, **SUM**(Value) **AS** total\_bus\_crossings

**FROM** border\_crossing\_entry\_data

**WHERE** Measure = 'Buses' **AND** YEAR(Date) = 2022

**GROUP BY** YEAR(Date), MONTH(Date)

**ORDER BY** total\_bus\_crossings **ASC**;

-- 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.

**SELECT** State, Measure, YEAR(Date) **AS** Year, SUM(Value) **AS** total\_crossings, **AVG**(Value) **AS** average\_crossings

**FROM** border\_crossing\_entry\_data

**GROUP BY** State, Measure, YEAR(Date)

**HAVING AVG**(Value) > 500

**ORDER BY** State, Measure, Year;