

Motor Vehicle Theft Analysis using SQL

-Banuvathy Rajakumar-



About the dataset

- ❖ Stolen vehicle data from the New Zealand police department's vehicle of interest database containing 6 months of data.
- ❖ Each record represents a single stolen vehicle, with data on vehicle type, make, year, color, date stolen and region stolen.
- ❖ Data Structure : Multiple tables
- ❖ No. of records : 4,707
- ❖ No. of fields = 14
- ❖ Source:

<https://mavenanalytics.io/dataplayground?page=2&pageSize=5>

Tables

`SELECT * FROM stolen_vehicles;`

| | vehicle_id | vehicle_type | make_id | model_year | vehicle_desc | color | date_stolen | location_id | Year | Month | Week | Day Name | risk_status |
|---|------------|--------------|---------|------------|----------------------|--------|-------------|-------------|------|-------|------|----------|-------------|
| ▶ | 1 | Trailer | 623 | 2021 | BST2021D | Silver | 2021-11-05 | 102 | 2021 | 11 | 44 | Friday | Low Risk |
| | 2 | Boat Trailer | 623 | 2021 | OUTBACK BOATS FT470 | Silver | 2021-12-13 | 105 | 2021 | 12 | 50 | Monday | Low Risk |
| | 3 | Boat Trailer | 623 | 2021 | ASD JETSKI | Silver | 2022-02-13 | 102 | 2022 | 2 | 7 | Sunday | Low Risk |
| | 4 | Trailer | 623 | 2021 | MSC 7X4 | Silver | 2021-11-13 | 106 | 2021 | 11 | 45 | Saturday | Low Risk |
| | 5 | Trailer | 623 | 2018 | D-MAX 8X5 | Silver | 2022-01-10 | 102 | 2022 | 1 | 2 | Monday | Low Risk |
| | 6 | Roadbike | 636 | 2005 | YZF-R6T | Black | 2021-12-31 | 102 | 2021 | 12 | 52 | Friday | High Risk |
| | 7 | Trailer | 623 | 2021 | CAAR TRANSPORTER | Silver | 2021-11-12 | 114 | 2021 | 11 | 45 | Friday | Low Risk |
| | 8 | Boat Trailer | 623 | 2001 | BOAT | Silver | 2022-02-22 | 109 | 2022 | 2 | 8 | Tuesday | High Risk |
| | 9 | Trailer | 514 | 2021 | 7X4-6" 1000KG | Silver | 2022-02-25 | 115 | 2022 | 2 | 8 | Friday | Low Risk |
| | 10 | Trailer | 514 | 2020 | 8X4 TANDEM | Silver | 2022-01-03 | 114 | 2022 | 1 | 1 | Monday | Low Risk |
| | 11 | Trailer | 623 | 2018 | HOMEBUILT | Silver | 2021-10-15 | 108 | 2021 | 10 | 41 | Friday | Low Risk |
| | 12 | Trailer | 538 | 2018 | BRENT SMITH TRAILERS | Silver | 2022-02-28 | 102 | 2022 | 2 | 9 | Monday | Low Risk |
| | 13 | Moped | 629 | 2004 | ET2 | Black | 2021-12-19 | 102 | 2021 | 12 | 51 | Sunday | High Risk |
| | 14 | Moped | 550 | 2007 | TODAY | Blue | 2022-01-23 | 102 | 2022 | 1 | 4 | Sunday | High Risk |
| | 15 | Roadbike | 611 | 2007 | DL650 | Black | 2022-02-28 | 101 | 2022 | 2 | 9 | Monday | High Risk |
| | 16 | Trailer | 623 | 2021 | JOBMATE | Silver | 2022-02-08 | 109 | 2022 | 2 | 6 | Tuesday | Low Risk |
| | 17 | Trailer | 623 | 2021 | BSTB85B | Silver | 2021-10-10 | 102 | 2021 | 10 | 41 | Sunday | Low Risk |

`SELECT * FROM make_details;`

| | make_id | make_name | make_type |
|---|---------|---------------|-----------|
| ▶ | 501 | Aakron Xpress | Standard |
| | 502 | ADLY | Standard |
| | 503 | Alpha | Standard |
| | 504 | Anglo | Standard |
| | 505 | Aprilia | Standard |
| | 506 | Atlas | Standard |
| | 507 | Audi | Standard |
| | 508 | Bailey | Standard |
| | 509 | Bedford | Standard |
| | 510 | Benelli | Standard |
| | 511 | Bentley | Luxury |
| | 512 | BMW | Luxury |
| | 513 | Bricon | Standard |
| | 514 | Briford | Standard |
| | 515 | Buell | Standard |
| | 516 | Buffalo | Standard |

`SELECT * FROM locations;`

| | location_id | region | country | population | density |
|---|-------------|--------------------|-------------|------------|---------|
| ▶ | 101 | Northland | New Zealand | 201500 | 16.11 |
| | 102 | Auckland | New Zealand | 1695200 | 343.09 |
| | 103 | Waikato | New Zealand | 513800 | 21.50 |
| | 104 | Bay of Plenty | New Zealand | 347700 | 28.80 |
| | 105 | Gisborne | New Zealand | 52100 | 6.21 |
| | 106 | Hawke's Bay | New Zealand | 182700 | 12.92 |
| | 107 | Taranaki | New Zealand | 127300 | 17.55 |
| | 108 | Manawatū-Whanganui | New Zealand | 258200 | 11.62 |
| | 109 | Wellington | New Zealand | 543500 | 67.52 |
| | 110 | Tasman | New Zealand | 58700 | 6.10 |
| | 111 | Nelson | New Zealand | 54500 | 129.15 |
| | 112 | Marlborough | New Zealand | 51900 | 4.94 |
| | 113 | West Coast | New Zealand | 32700 | 1.41 |
| | 114 | Canterbury | New Zealand | 655000 | 14.72 |
| | 115 | Otago | New Zealand | 246000 | 7.89 |
| | 116 | Southland | New Zealand | 102400 | 3.28 |

Data cleaning

Used:
Row Number,
Over, Partition
by, Delete

1) Handling missing rows

```
SELECT
  *
FROM
  stolen_vehicles
WHERE
  vehicle_type = 'NULL';
```



```
DELETE FROM
  stolen_vehicles
WHERE
  vehicle_type IS NULL;
```

2) Check for duplicate rows

```
WITH CTE AS
(
  SELECT
    *,
    ROW_NUMBER() OVER(Partition by vehicle_id, vehicle_type, make_id, model_year, vehicle_desc, color, date_stolen, location_id ORDER BY
    vehicle_id, vehicle_type, make_id, model_year, vehicle_desc, color, date_stolen, location_id) AS row_num
  FROM stolen_vehicles
)
SELECT COUNT(*) FROM CTE WHERE row_num > 1;
```



| | COUNT(*) |
|---|----------|
| ▶ | 0 |

No duplicate rows

Data preparation

Extract Year, Month And Day Name from Column Date stolen and make separate columns for each

```
-- YEAR--  
SELECT EXTRACT(Year FROM date_stolen) FROM stolen_vehicles;  
  
ALTER TABLE stolen_vehicles  
ADD COLUMN Year int after location_id;  
  
UPDATE stolen_vehicles  
SET Year = EXTRACT(Year FROM date_stolen);
```

```
-- WEEK--  
SELECT EXTRACT(Week FROM date_stolen) FROM stolen_vehicles;  
  
ALTER TABLE stolen_vehicles  
ADD COLUMN Week int after Month;  
  
UPDATE stolen_vehicles  
SET Week = EXTRACT(Week FROM date_stolen);
```

```
-- MONTH--  
SELECT EXTRACT(Month FROM date_stolen) FROM stolen_vehicles;  
  
ALTER TABLE stolen_vehicles  
ADD COLUMN Month int after Year;  
  
UPDATE stolen_vehicles  
SET Month = EXTRACT(Month FROM date_stolen);
```

```
-- DAY NAME--  
SELECT Dayname(date_stolen) FROM stolen_vehicles;  
  
ALTER TABLE stolen_vehicles  
MODIFY COLUMN `Day Name` character(15) after Week;  
  
UPDATE stolen_vehicles  
SET `Day Name` = Dayname(date_stolen);
```

Alter,
Update,
Extract

Data Analysis



1) How many stolen vehicles were reported in each region?

```
SELECT
    loc.region,
    count(*) as No_of_stolen_vehicles
FROM
    stolen_vehicles sv
INNER JOIN
    locations loc
ON sv.location_id = loc.location_id
GROUP BY loc.region
ORDER BY No_of_stolen_vehicles desc;
```

| | region | No_of_stolen_vehicles |
|---|--------------------|-----------------------|
| ▶ | Auckland | 1626 |
| | Canterbury | 660 |
| | Bay of Plenty | 442 |
| | Wellington | 417 |
| | Waikato | 369 |
| | Northland | 233 |
| | Gisborne | 175 |
| | Manawatū-Whanganui | 138 |
| | Otago | 137 |
| | Taranaki | 112 |
| | Hawke's Bay | 100 |
| | Nelson | 92 |
| | Southland | 26 |

- Maximum number of thefts happened in Auckland, New Zealand. 1626 thefts were reported from Auckland.

2) Total number of vehicles stolen?

```
WITH CTE AS(
    SELECT
        loc.region,
        count(*) as No_of_stolen_vehicles
    FROM
        stolen_vehicles sv
    INNER JOIN
        locations loc
    ON sv.location_id = loc.location_id
    GROUP BY loc.region
    ORDER BY No_of_stolen_vehicles desc
)
SELECT SUM(No_of_stolen_vehicles) FROM CTE;
```



| | SUM(No_of_stolen_vehicles) |
|---|----------------------------|
| ▶ | 4527 |

- Totally 4527 vehicles were stolen

2) How many stolen vehicles have been reported in regions with a population greater than the average population?

```
WITH CTE3 AS(
  SELECT
    loc.region, COUNT(*) AS cnt
  FROM stolen_vehicles sv
  INNER JOIN
    locations loc
  ON sv.location_id= loc.location_id
  WHERE loc.population > (SELECT AVG(population) FROM locations)
  GROUP BY loc.region
) SELECT SUM(cnt) FROM CTE3;
```

| | SUM(cnt) |
|---|----------|
| ▶ | 3514 |



- Out of 4527 stolen vehicles, 3514 vehicles were reported in regions with a population greater than the average population

3) Classify stolen vehicles as 'High Risk' if their model year is before 2010, and 'Low Risk' otherwise.

```
ALTER TABLE stolen_vehicles
ADD COLUMN risk_status varchar(35) after `Day Name`;

UPDATE stolen_vehicles
SET risk_status =
CASE
    WHEN model_year < 2010 THEN 'High Risk'
    ELSE 'Low Risk'
END;
```

Alter,
Update
statements,
Case structure

| | vehicle_id | vehicle_type | make_id | model_year | vehicle_desc | color | date_stolen | location_id | Year | Month | Week | Day Name | risk_status |
|---|------------|--------------|---------|------------|----------------------|--------|-------------|-------------|------|-------|------|----------|-------------|
| ▶ | 1 | Trailer | 623 | 2021 | BST2021D | Silver | 2021-11-05 | 102 | 2021 | 11 | 44 | Friday | Low Risk |
| | 2 | Boat Trailer | 623 | 2021 | OUTBACK BOATS FT470 | Silver | 2021-12-13 | 105 | 2021 | 12 | 50 | Monday | Low Risk |
| | 3 | Boat Trailer | 623 | 2021 | ASD JETSKI | Silver | 2022-02-13 | 102 | 2022 | 2 | 7 | Sunday | Low Risk |
| | 4 | Trailer | 623 | 2021 | MSC 7X4 | Silver | 2021-11-13 | 106 | 2021 | 11 | 45 | Saturday | Low Risk |
| | 5 | Trailer | 623 | 2018 | D-MAX 8X5 | Silver | 2022-01-10 | 102 | 2022 | 1 | 2 | Monday | Low Risk |
| | 6 | Roadbike | 636 | 2005 | YZF-R6T | Black | 2021-12-31 | 102 | 2021 | 12 | 52 | Friday | High Risk |
| | 7 | Trailer | 623 | 2021 | CAAR TRANSPORTER | Silver | 2021-11-12 | 114 | 2021 | 11 | 45 | Friday | Low Risk |
| | 8 | Boat Trailer | 623 | 2001 | BOAT | Silver | 2022-02-22 | 109 | 2022 | 2 | 8 | Tuesday | High Risk |
| | 9 | Trailer | 514 | 2021 | 7X4-6" 1000KG | Silver | 2022-02-25 | 115 | 2022 | 2 | 8 | Friday | Low Risk |
| | 10 | Trailer | 514 | 2020 | 8X4 TANDEM | Silver | 2022-01-03 | 114 | 2022 | 1 | 1 | Monday | Low Risk |
| | 11 | Trailer | 623 | 2018 | HOMEBUILT | Silver | 2021-10-15 | 108 | 2021 | 10 | 41 | Friday | Low Risk |
| | 12 | Trailer | 538 | 2018 | BRENT SMITH TRAILERS | Silver | 2022-02-28 | 102 | 2022 | 2 | 9 | Monday | Low Risk |
| | 13 | Moped | 629 | 2004 | ET2 | Black | 2021-12-19 | 102 | 2021 | 12 | 51 | Sunday | High Risk |

4) Show the count of number of vehicles based on the risk status

```
SELECT  
    risk_status,  
    COUNT(*) AS No_of_stolen_vehicles  
FROM  
    stolen_vehicles  
GROUP BY risk_status;
```

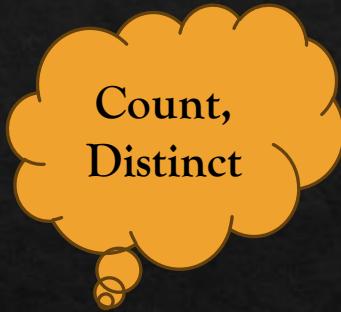
Group by,
Aggregate
function
(Count)

| risk_status | No_of_stolen_vehicles |
|-------------|-----------------------|
| Low Risk | 1297 |
| High Risk | 3230 |

- Older vehicles (before 2010) more susceptible to theft due to potentially outdated security features or easier tampering methods. Hence number of vehicles stolen also higher compared to the vehicles made after 2010.

5) What are the total number of regions in the dataset?

```
SELECT  
    COUNT(DISTINCT vehicle_type)  
FROM stolen_vehicles;
```



Count,
Distinct

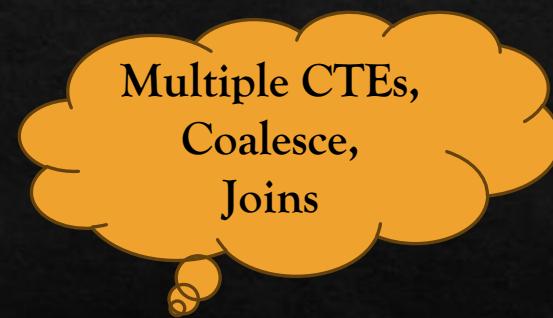
| | COUNT(DISTINCT region) |
|---|------------------------|
| ▶ | 16 |

- Totally there were 16 regions in New Zealand reported to have thefts

6) Number of vehicles stolen with respect to each region and their population density.

```
WITH StolenVehicleCounts AS (
    SELECT
        location_id,
        COUNT(*) as num_stolen_vehicles
    FROM
        stolen_vehicles
    GROUP BY
        location_id
    ),
RegionPopulation AS (
    SELECT
        l.location_id,
        l.region,
        l.population,
        l.density,
        COALESCE(svc.num_stolen_vehicles, 0) AS num_stolen_vehicles
    FROM
        locations l
    INNER JOIN
        StolenVehicleCounts svc
    ON
        l.location_id = svc.location_id
)
SELECT
    location_id,
    region,
    population, density,
    num_stolen_vehicles
FROM
    RegionPopulation;
```

| location_id | region | population | density | num_stolen_vehicles |
|-------------|--------------------|------------|---------|---------------------|
| 102 | Auckland | 1695200 | 343.09 | 1626 |
| 114 | Canterbury | 655000 | 14.72 | 660 |
| 104 | Bay of Plenty | 347700 | 28.80 | 442 |
| 109 | Wellington | 543500 | 67.52 | 417 |
| 103 | Waikato | 513800 | 21.50 | 369 |
| 101 | Northland | 201500 | 16.11 | 233 |
| 105 | Gisborne | 52100 | 6.21 | 175 |
| 108 | Manawatū-Whanganui | 258200 | 11.62 | 138 |
| 115 | Otago | 246000 | 7.89 | 137 |
| 107 | Taranaki | 127300 | 17.55 | 112 |
| 106 | Hawke's Bay | 182700 | 12.92 | 100 |
| 111 | Nelson | 54500 | 129.15 | 92 |
| 116 | Southland | 102400 | 3.28 | 26 |



- The region with highest population density (Auckland) is found to have maximum number of vehicles stolen compared to other region while the region with least density (Southland) has minimal number of thefts.

7) What day of the week are vehicles most often and least often stolen?

```
SELECT
    `Day Name`,
    COUNT(*) AS No_vehicles_stolen
FROM
    stolen_vehicles
GROUP BY
    `Day Name`
ORDER BY No_vehicles_stolen DESC;
```

Count,
Group by,
Order by

| Day Name | No_vehicles_stolen |
|-----------|--------------------|
| Monday | 760 |
| Tuesday | 704 |
| Friday | 653 |
| Wednesday | 624 |
| Thursday | 616 |
| Sunday | 594 |
| Saturday | 576 |

- Monday and Tuesday are the days when the vehicles are most stolen.
- Saturday and Sunday are the days when the vehicles are least stolen.

8) What is the average population density in regions where stolen trailers were reported?

```
SELECT vehicle_type
FROM stolen_vehicles
WHERE vehicle_type LIKE '%trailer';
```

```
WITH CTE2 as
(
  SELECT
    sv.vehicle_type, ROUND(AVG(loc.density),2) as Avg_pop_density_of_region_stolen, COUNT(*) as No_of_trailers_stolen
  FROM
    stolen_vehicles sv
  INNER JOIN
    locations loc
  ON
    sv.location_id = loc.location_id
  GROUP BY
    sv.vehicle_type
)
SELECT vehicle_type, Avg_pop_density_of_region_stolen, No_of_trailers_stolen
FROM CTE2
WHERE vehicle_type LIKE '%Trailer';
```

CTE,
Round, Count
Joins, Like

| vehicle_type | Avg_pop_density_of_region_stolen | No_of_trailers_stolen |
|--------------|----------------------------------|-----------------------|
| Trailer | 103.51 | 582 |
| Boat Trailer | 137.68 | 105 |

9) Rank the boat trailer thefts based on the date stolen within each region

```
SELECT
    loc.region, sv.date_stolen,
    RANK() OVER(Partition by loc.region ORDER BY sv.date_stolen asc) as rankn
FROM
    stolen_vehicles sv
INNER JOIN
    locations loc
ON
    sv.location_id = loc.location_id
WHERE vehicle_type = 'Boat trailer'
;
```

Rank, Over
clause, Joins

| | region | date_stolen | rankn |
|---|----------|-------------|-------|
| ▶ | Auckland | 2021-10-08 | 1 |
| | Auckland | 2021-10-22 | 2 |
| | Auckland | 2021-10-29 | 3 |
| | Auckland | 2021-11-21 | 4 |
| | Auckland | 2021-11-28 | 5 |
| | Auckland | 2021-12-06 | 6 |
| | Auckland | 2021-12-14 | 7 |
| | Auckland | 2021-12-15 | 8 |
| | Auckland | 2021-12-15 | 8 |
| | Auckland | 2021-12-16 | 10 |
| | Auckland | 2021-12-17 | 11 |
| | Auckland | 2021-12-22 | 12 |

10) What is the average time interval between the successive thefts?

```
● WITH CTE2 AS (
    SELECT
        date_stolen,
        LEAD(date_stolen) OVER (ORDER BY date_stolen) AS next_stolen_date
    FROM
        stolen_vehicles sv
    INNER JOIN
        locations loc ON sv.location_id = loc.location_id
),
CTE3 AS (
    SELECT
        (DATEDIFF(next_stolen_date, date_stolen)) AS inter
    FROM
        CTE2
)
SELECT AVG(inter) AS avg_interval FROM CTE3;
```

Multiple CTEs,
Lead,
Joins

| | avg_interval |
|---|--------------|
| ▶ | 0.0400 |

Summary of the Insights

- ❖ Auckland, New Zealand's most populous region, emerges as the epicenter of vehicle theft. With a total of 4527 reported stolen vehicles, Auckland tops the list, indicating a significant concentration of theft activity in this region.
- ❖ Regions with higher-than-average population densities tend to experience more vehicle theft incidents. Of the total stolen vehicles, 3514 were reported in regions with populations exceeding the average, highlighting a correlation between population density and theft occurrences.
- ❖ Older vehicles, manufactured before 2010, are found to be more susceptible to theft due to potentially outdated security features. Consequently, the number of stolen vehicles is higher among older models compared to those made after 2010, emphasizing the importance of vehicle security enhancements over time.

- ❖ A total of 16 regions in New Zealand reported vehicle theft incidents. While Auckland experiences the highest number of thefts, regions with lower population densities, such as Southland, record minimal theft activity.
- ❖ Vehicle theft exhibits distinct day-wise patterns. Mondays and Tuesdays emerge as the days with the highest incidence of theft, suggesting potential factors such as reduced surveillance or increased opportunity for theft during the start of the workweek.
- ❖ Conversely, Saturdays and Sundays witness the least number of thefts, possibly due to heightened security measures or increased public activity during weekends.

Thank You