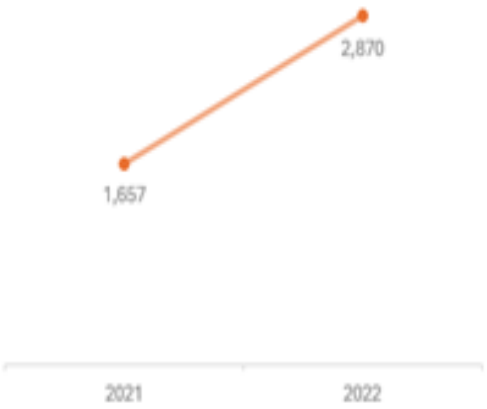
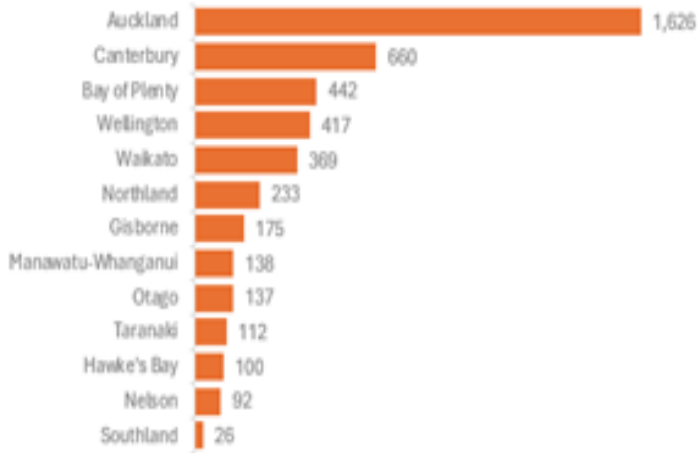
# There were 4,527 stolen vehicles in New Zealand from 2021-2022.

## Yearly Trends

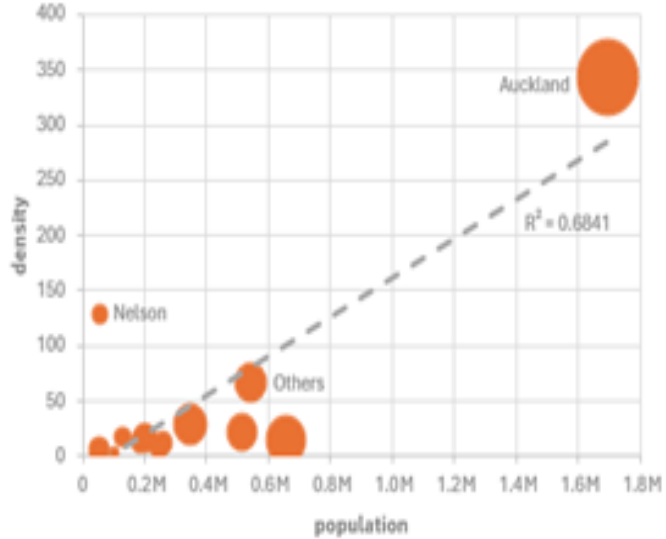



## Monthly Trends

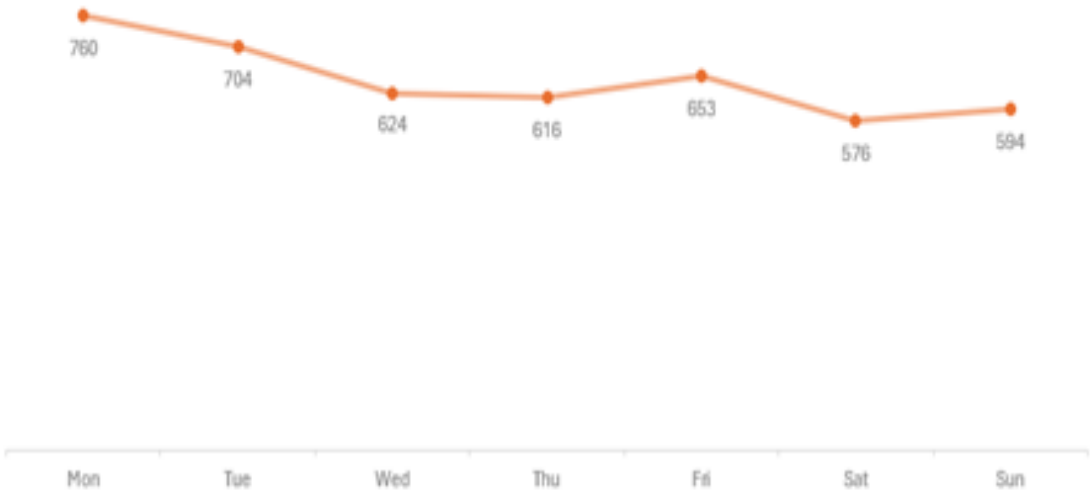
## Stolen Vehicles per Region



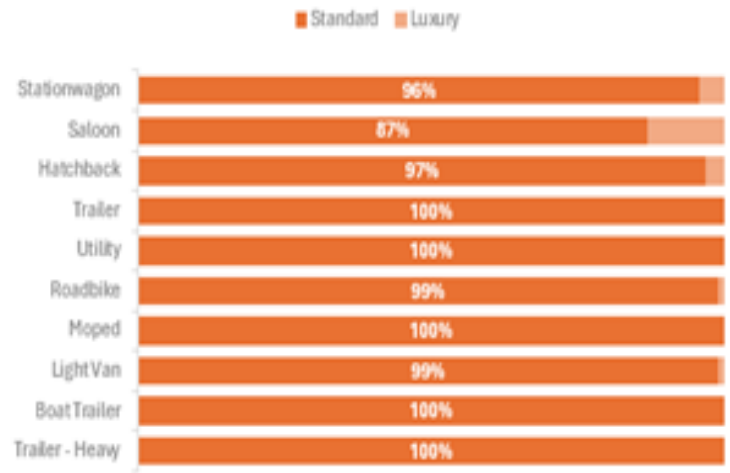

## Population vs. Density




## Weekly Trends




## Top 10 Stolen Vehicles




## Total Share

# Abstract

- I analyzed a relational database of stolen vehicles from New Zealand.
  - I had to help law enforcement raise awareness of stolen vehicles and improve safety.
- My analysis focused primarily on the seasonal trends and the characteristics of the stolen vehicles.
- I used Microsoft Excel and MySQL Workbench to conduct my analysis.
  - I created a relational data model using the EER diagram editor.
  - I used Microsoft Excel to create a dashboard that summarized the following information:
    - Yearly, monthly, and weekly trends in stolen vehicles.
    - Stolen vehicles per region and vehicle type.
    - The share of stolen vehicles per vehicle type.
    - Population and Density correlation for the stolen vehicles.
- I have uploaded all the files for this project onto my GitHub.

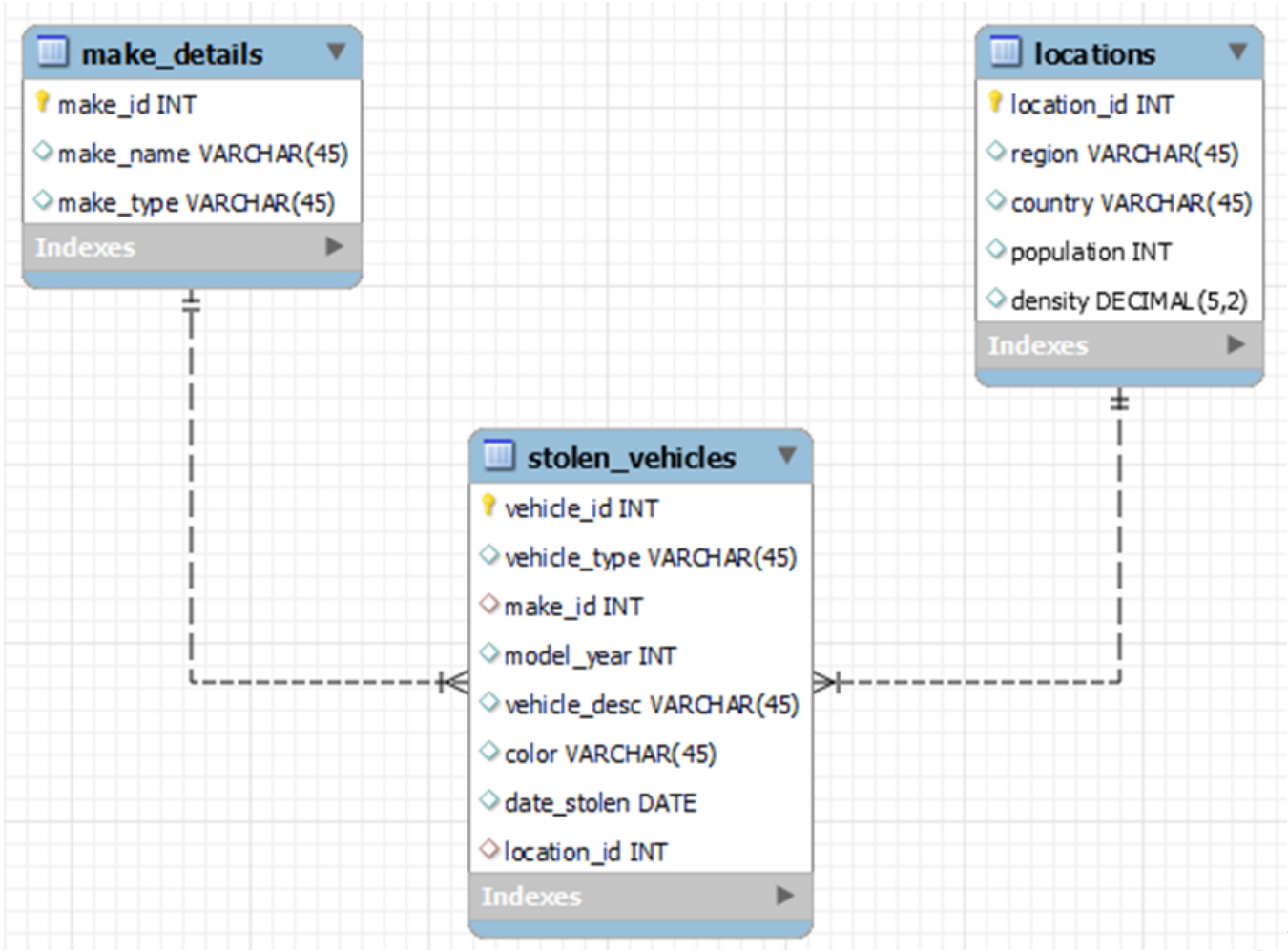
# Data-Cleaning Strategy

- I am working with structured data for this project.
- The data for this project consists of three tables.
  - locations
    - A dimension table with information on where vehicles were stolen.
  - make_details
    - A dimension table with information about the make of each stolen vehicle.
  - stolen_vehicles
    - A fact table with logs of each stolen vehicle and their characteristics.
- I made some slight modifications to the database file for this project.
  - I added foreign keys to the stolen_vehicles table.
  - I took away the NOT NULL constraints from all ID columns.
    - primary keys are always unique, so this constraint is redundant.

# Assumptions

- There were null values in the stolen_vehicles table.
  - I assumed there was a user error in recording these stolen vehicles.
  - The scope of my analysis did not focus on any uncertainties.
    - Therefore, I filtered out these vehicles from my analysis.

# Data Model


make_details
make_id INT
make_name VARCHAR(45)
make_type VARCHAR(45)
Indexes
locations
location_id INT
region VARCHAR(45)
country VARCHAR(45)
population INT
density DECIMAL(5,2)
Indexes
stolen_vehicles
vehicle_id INT
vehicle_type VARCHAR(45)
make_id INT
model_year INT
vehicle_desc VARCHAR(45)
color VARCHAR(45)
date_stolen DATE
location_id INT
Indexes

# SQL Query Objectives

1. Identify when vehicles were stolen.
    1. Total number of stolen vehicles per year
    2. Total number of stolen vehicles per month
    3. Total number of stolen vehicles per day of week
2. Identify what vehicles were stolen.
    1. Total number of stolen vehicles per make type
    2. Pivot table
3. Identify where vehicles were stolen.
    1. Total number of stolen vehicles per region

# Objective 1

```sql
-- Find the total number of stolen vehicles per year
SELECT
    YEAR(date_stolen) AS year,
    COUNT(*) AS num_vehicles
FROM stolen_vehicles
WHERE vehicle_type IS NOT NULL
GROUP BY year;
```

| year | num_vehicles |
|---|---|
| 2021 | 1657 |
| 2022 | 2870 |

Back to Query Objectives

# Objective 1

```sql
-- Find the total number of stolen vehicles per month
SELECT
    LEFT(MONTHNAME(date_stolen), 3) AS month,
    COUNT(*) AS num_vehicles
FROM stolen_vehicles
WHERE vehicle_type IS NOT NULL
GROUP BY MONTH(date_stolen)
ORDER BY MONTH(date_stolen);
```

Back to Query Objectives

| month | num_vehicles |
|---|---|
| Jan | 737 |
| Feb | 757 |
| Mar | 1049 |
| Apr | 327 |
| Oct | 461 |
| Nov | 556 |
| Dec | 640 |

# Objective 1

```sql
-- Find the total number of stolen vehicles per day of the week
SELECT
    CASE
        WHEN WEEKDAY(date_stolen) = 0 THEN 'Mon'
        WHEN WEEKDAY(date_stolen) = 1 THEN 'Tue'
        WHEN WEEKDAY(date_stolen) = 2 THEN 'Wed'
        WHEN WEEKDAY(date_stolen) = 3 THEN 'Thu'
        WHEN WEEKDAY(date_stolen) = 4 THEN 'Fri'
        WHEN WEEKDAY(date_stolen) = 5 THEN 'Sat'
        WHEN WEEKDAY(date_stolen) = 6 THEN 'Sun'
    END AS day_of_week,
    COUNT(*) AS num_vehicles
FROM stolen_vehicles
WHERE vehicle_type IS NOT NULL
GROUP BY WEEKDAY(date_stolen)
ORDER BY WEEKDAY(date_stolen);
```

| day_of_week | num_vehicles |
|---|---|
| Mon | 760 |
| Tue | 704 |
| Wed | 624 |
| Thu | 616 |
| Fri | 653 |
| Sat | 576 |
| Sun | 594 |

Back to Query Objectives

# Objective 2

```sql
-- Find the total number of stolen vehicles per make type
SELECT
    m.make_type,
    COUNT(*) AS num_vehicles
FROM stolen_vehicles s
LEFT JOIN make_details m ON s.make_id = m.make_id
WHERE s.vehicle_type IS NOT NULL
GROUP BY m.make_type
ORDER BY num_vehicles DESC;
```

| make_type | num_vehicles |
|---|---|
| Standard | 4338 |
| Luxury | 189 |

Back to Query Objectives

# Objective 2

```sql
-- Create a pivot table with the following dimensions:
-- rows representing the top 10 stolen vehicle types
-- columns representing the make type ratio and the total number of stolen vehicles
-- values representing the total number of stolen vehicles
SELECT
    s.vehicle_type,
    ROUND(COUNT(CASE WHEN m.make_type = 'Standard' THEN s.vehicle_id ELSE NULL END)/
          COUNT(*), 2) AS standard,
    ROUND(COUNT(CASE WHEN m.make_type = 'Luxury' THEN s.vehicle_id ELSE NULL END)/
          COUNT(*), 2) AS luxury,
    COUNT(*) AS total
FROM stolen_vehicles s
LEFT JOIN make_details m ON s.make_id = m.make_id
WHERE s.vehicle_type IS NOT NULL
GROUP BY s.vehicle_type
ORDER BY total DESC
LIMIT 10;
```

| vehicle_type | standard | luxury | total |
|---|---|---|---|
| Stationwagon | 0.96 | 0.04 | 945 |
| Saloon | 0.87 | 0.13 | 851 |
| Hatchback | 0.97 | 0.03 | 644 |
| Trailer | 1.00 | 0.00 | 582 |
| Utility | 1.00 | 0.00 | 466 |
| Roadbike | 0.99 | 0.01 | 297 |
| Moped | 1.00 | 0.00 | 187 |
| Light Van | 0.99 | 0.01 | 154 |
| Boat Trailer | 1.00 | 0.00 | 105 |
| Trailer - Heavy | 1.00 | 0.00 | 90 |

Back to Query Objectives

# Objective 3

```sql
-- Find the total number of stolen vehicles per region, population, and density
SELECT
    l.region,
    l.population,
    l.density,
    COUNT(*) AS num_vehicles
FROM stolen_vehicles s
LEFT JOIN locations l ON s.location_id = l.location_id
WHERE s.vehicle_type IS NOT NULL
GROUP BY
    l.region,
    l.population,
    l.density
ORDER BY num_vehicles DESC;
```

| region | population | density | num_vehicles |
|---|---|---|---|
| Auckland | 1695200 | 343.09 | 1626 |
| Canterbury | 655000 | 14.72 | 660 |
| Bay of Plenty | 347700 | 28.80 | 442 |
| Wellington | 543500 | 67.52 | 417 |
| Waikato | 513800 | 21.50 | 369 |
| Northland | 201500 | 16.11 | 233 |
| Gisborne | 52100 | 6.21 | 175 |
| Manawatu-Whanganui | 258200 | 11.62 | 138 |
| Otago | 246000 | 7.89 | 137 |
| Taranaki | 127300 | 17.55 | 112 |
| Hawke's Bay | 182700 | 12.92 | 100 |
| Nelson | 54500 | 129.15 | 92 |
| Southland | 102400 | 3.28 | 26 |

Back to Query Objectives

# Seasonal Insights

- Yearly Trends
  - The total number of stolen vehicles increased by 73% from 2021-2022.
- Monthly Trends
  - The total number of stolen vehicles peaked during the first quarter.
  - After the first quarter, the total number of stolen vehicles declined significantly.
    - The total was at its lowest value of 327 in April.
    - Then the total began to increase gradually again after April.
- Weekly Trends
  - More vehicles were stolen on weekdays than on weekends.

# Vehicle Insights

- Standard vehicles made up 96% of all stolen vehicles.
  - They were also the most frequently stolen vehicles across all vehicle types.
    - Stationwagans were the most frequently stolen vehicle type.
    - Luxury vehicles weren’t even stolen for some vehicle types.

# Regional Insights

- Auckland and Nelson are two regions that show some interesting insights.
  - Auckland has the largest population and density and was the region with the maximum number of stolen vehicles.
  - Nelson has one of the lowest populations, but it is denser than other regions outside of Auckland, and not many vehicles were stolen there.
  - There were more stolen vehicles in more populated and less dense regions outside of Auckland and Nelson.
- There is a direct correlation between population and density of the regions.
  - However, the correlation is weak since the coefficient of determination was 68%.

# Recommendations

- To prevent more vehicles from being stolen in the future and to increase public safety, I recommend doing the following:
  - Prioritize law enforcement patrolling during the first quarter of the year and weekdays.
    - Since most vehicles were stolen during these times, this strategy will help law enforcement proactively prevent stolen vehicles.
  - Prioritize funding for law enforcement in Auckland.
    - Since most vehicles were stolen here, this strategy will allow for improved training for law enforcement to prevent stolen vehicles.
  - Prioritize standard vehicles when trying to search for stolen vehicles.
    - Luxury vehicles seem to be stolen less frequently because they tend to have better theft protection systems.
      - This might be because luxury vehicles are more expensive compared to standard vehicles.