**Motor Vehicle Thefts Analysis – Documentation**

**Introduction:**Motor Vehicle Thefts Analysis presents detailed data on vehicle theft in New Zealand, analyzed by vehicle type, region, day of the week, and other key variables. The report provides visualizations and metrics to help you better understand theft patterns and the factors that influence theft in New Zealand.

**Problem:**

Aim of this project is to dig into the stolen vehicles database to find details on when, which and where vehicles are most likely to be stolen.

The project answers key analytical questions:

1. What day of the week are vehicles most often and least often stolen?
2. What types of vehicles are most often and least often stolen? Does this vary by region?
3. What is the average age of the vehicles that are stolen? Does this vary based on the vehicle type?
4. Which regions have the most and least number of stolen vehicles? What are the characteristics of the regions?

The following sections of this documentation will provide detailed information about each page within the reporting solution:

Introduction - Provides an overview of the report and instructions for navigation:

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Visual Insights - Charts showing key metrics related to vehicle theft, with the ability to filter data by region and vehicle type:

Obraz zawierający tekst, zrzut ekranu, diagram, design

Opis wygenerowany automatycznie

Pivot Table Analysis - Tables with key indicators that provide detailed vehicle theft figures, grouped by different categories:

Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie

**Data source:**

The dataset comes from the Maven Analytics website. This dataset contains 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:**

The dataset used for the analysis is a MySQL database from Maven Analytics. The database has been imported into SQL Server Management Studio by the following queries:  
  
-- Table structure for table `stolen\_vehicles`

CREATE TABLE stolen\_vehicles (

vehicle\_id INT NOT NULL,

vehicle\_type VARCHAR(45),

make\_id INT,

model\_year INT,

vehicle\_desc VARCHAR(45),

color VARCHAR(45),

date\_stolen DATE,

location\_id VARCHAR(45),

PRIMARY KEY (vehicle\_id)

);

-- Table structure for table `locations`

CREATE TABLE locations (

location\_id INT NOT NULL,

region VARCHAR(45),

country VARCHAR(45),

population INT,

density DECIMAL(5,2),

PRIMARY KEY (location\_id)

);

-- Table structure for table `make\_details`

CREATE TABLE make\_details (

make\_id INT NOT NULL,

make\_name VARCHAR(45),

make\_type VARCHAR(45),

PRIMARY KEY (make\_id)

);

-- Insert data into table stolen\_vehicles

INSERT INTO stolen\_vehicles (vehicle\_id, vehicle\_type, make\_id, model\_year,

vehicle\_desc, color, date\_stolen, location\_id) VALUES

(1, 'Trailer', 623, 2021, 'BST2021D', 'Silver', '2021-11-05', 102),

(2, 'Boat Trailer', 623, 2021, 'OUTBACK BOATS FT470', 'Silver', '2021-12-13', 105),

(3, 'Boat Trailer', 623, 2021, 'ASD JETSKI', 'Silver', '2022-02-13', 102),

(4, 'Trailer', 623, 2021, 'MSC 7X4', 'Silver', '2021-11-13', 106),

(5, 'Trailer', 623, 2018, 'D-MAX 8X5', 'Silver', '2022-01-10', 102),

(6, 'Roadbike', 636, 2005, 'YZF-R6T', 'Black', '2021-12-31', 102),

(7, 'Trailer', 623, 2021, 'CAAR TRANSPORTER', 'Silver', '2021-11-12', 114),

(8, 'Boat Trailer', 623, 2001, 'BOAT', 'Silver', '2022-02-22', 109),

(9, 'Trailer', 514, 2021, '7X4-6" 1000KG', 'Silver', '2022-02-25', 115),

(10, 'Trailer', 514, 2020, '8X4 TANDEM', 'Silver', '2022-01-03', 114),  
(…)

-- Insert data into table locations

INSERT INTO locations VALUES (101, 'Northland', 'New Zealand', 201500, 16.11),

(102, 'Auckland', 'New Zealand', 1695200, 343.09),

(103, 'Waikato', 'New Zealand', 513800, 21.5),

(104, 'Bay of Plenty', 'New Zealand', 347700, 28.8),

(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, 6.1),

(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);

-- Insert data into table make\_details

INSERT INTO make\_details VALUES (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'),

(…)

In outcome the database contains 3 tables. The tables contains following columns:

|  |  |  |
| --- | --- | --- |
| Table | Column | Description |
| stolen\_vehicles | vehicle\_id | Unique ID of a stolen vehicle |
| stolen\_vehicles | vehicle\_type | Type of vehicle |
| stolen\_vehicles | make\_id | Matches make\_id in the make\_details table |
| stolen\_vehicles | model\_year | Model year of vehicle |
| stolen\_vehicles | vehicle\_desc | Description of vehicle |
| stolen\_vehicles | color | Color of vehicle |
| stolen\_vehicles | date\_stolen | Date the vehicle was stolen (MM/DD/YY) |
| stolen\_vehicles | location\_id | Matches location\_id in the locations table |
| make\_details | make\_id | Unique ID of the make |
| make\_details | make\_name | Name of the make |
| make\_details | make\_type | Type of make (Standard or Luxury) |
| locations | location\_id | Unique ID of the region |
| locations | region | Name of the region |
| locations | country | Country where the region is located |
| locations | population | Population of the region |
| locations | density | Density of the region (population / km2) |

**Data Preprocessing:**

Prior to building the Motor Theft Analysis, the raw data might have undergone certain preprocessing steps. These steps may include data cleaning, transforming data types, handling missing values, and aggregating data at the desired level of granularity. The data preparation process ensures the data is in a suitable format for analysis and reporting.

**Pivot Table Analysis:**

The Pivot Table Analysis worksheet contains the PivotTables shown below, which were created from the data extracted by the corresponding SQL queries.

Table:

* **Stolen cars by weekday**
  + SQL query

SELECT vehicle\_id

,vehicle\_type

,DATENAME( WEEKDAY, date\_stolen) AS [weekday]

FROM stolen\_vehicles

WHERE vehicle\_type IS NOT NULL

|  |  |  |
| --- | --- | --- |
| **vehicle\_id** | **vehicle\_type** | **weekday** |
| 1 | Trailer | Friday |
| 2 | Boat Trailer | Monday |
| 3 | Boat Trailer | Sunday |
| 4 | Trailer | Saturday |
| 5 | Trailer | Monday |

* + Table

|  |  |
| --- | --- |
| **Weekday** | **Number of cars** |
| Monday | 760 |
| Tuesday | 704 |
| Wednesday | 624 |
| Thursday | 616 |
| Friday | 653 |
| Saturday | 576 |
| Sunday | 594 |

* + Pivot Table (rows: weekday, values: Count of vehicle\_id)
* **Average age of the stolen type of vehicle**
  + SQL query

SELECT vehicle\_id

, vehicle\_type

, YEAR(date\_stolen) AS [date\_stolen]

, model\_year

, YEAR(date\_stolen) - model\_year AS age

FROM stolen\_vehicles

WHERE model\_year IS NOT NULL

AND vehicle\_type IS NOT NULL

* + Table

|  |  |  |  |
| --- | --- | --- | --- |
| **vehicle\_type** | **date\_stolen** | **model\_year** | **age** |
| Trailer | 2021 | 2021 | 0 |
| Boat Trailer | 2021 | 2021 | 0 |
| Boat Trailer | 2022 | 2021 | 1 |
| Trailer | 2021 | 2021 | 0 |
| Trailer | 2022 | 2018 | 4 |

* + Pivot Table (rows: vehicle\_type, values: Average of age)

|  |  |
| --- | --- |
| **Type** | **Avg age** |
| Special Purpose Vehicle | 64 |
| Mobile Home - Light | 35 |
| Flat Deck Truck | 28 |
| Caravan | 28 |
| Other Truck | 23 |
| Convertible | 23 |

* **Most popular type of stolen cars**
* **Most popular color of stolen cars**
* **Most popular color of stolen cars by type**
  + SQL query

SELECT vehicle\_type

,color

,COUNT(\*) AS [number\_cars]

FROM stolen\_vehicles

WHERE vehicle\_type IS NOT NULL

GROUP BY vehicle\_type, color

ORDER BY number\_cars DESC

* + Table

|  |  |  |
| --- | --- | --- |
| **vehicle\_type** | **color** | **number\_cars** |
| Trailer | Silver | 399 |
| Saloon | Silver | 226 |
| Stationwagon | Silver | 223 |
| Utility | White | 183 |
| Hatchback | Silver | 172 |
| Saloon | White | 160 |

* + Pivot Tables

(rows: color, values: Sum of number\_cars)

|  |  |
| --- | --- |
| **Color** | **SUM** |
| Silver | 1272 |
| White | 932 |
| Black | 586 |
| Blue | 511 |
| Red | 388 |

(rows: vehicle\_type, values: Sum of number\_cars)

|  |  |
| --- | --- |
| **Type** | **Number of cars** |
| Stationwagon | 945 |
| Saloon | 851 |
| Hatchback | 644 |
| Trailer | 582 |
| Utility | 466 |
| Roadbike | 297 |

(rows: vehicle\_type, columns: color, values: Sum of number\_cars)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **'** | **Color** |  |  |  |  |  |  |
|  | **Silver** | **White** | **Black** | **Blue** | **Red** | **Grey** | **Green** |
| **Type** |  |  |  |  |  |  |  |
| Stationwagon | 223 | 159 | 141 | 142 | 84 | 84 | 59 |
| Saloon | 226 | 160 | 99 | 125 | 75 | 71 | 52 |
| Hatchback | 172 | 114 | 76 | 104 | 58 | 46 | 24 |
| Trailer | 399 | 21 | 29 | 17 | 9 | 73 | 22 |
| Utility | 71 | 183 | 36 | 46 | 45 | 32 | 38 |
| Roadbike | 17 | 42 | 105 | 38 | 51 | 11 | 12 |
| Moped | 8 | 25 | 85 | 18 | 34 | 3 | 1 |

* **Stolen cars by the type of vehicle in region**
  + SQL query

SELECT region

,vehicle\_type

,COUNT(\*) AS [number\_cars]

FROM stolen\_vehicles st

LEFT JOIN locations loc ON st.location\_id = loc.location\_id

WHERE vehicle\_type IS NOT NULL

GROUP BY region, vehicle\_type

ORDER BY number\_cars DESC

* + Table

|  |  |  |
| --- | --- | --- |
| **region** | **vehicle\_type** | **number\_cars** |
| Auckland | Saloon | 327 |
| Auckland | Stationwagon | 306 |
| Auckland | Hatchback | 296 |
| Canterbury | Stationwagon | 165 |
| Auckland | Roadbike | 163 |
| Auckland | Trailer | 141 |

* + Pivot Table (rows: vehicle\_type, columns: region, values: Sum of number\_cars)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Region** |  |  |  |  |  |
| **Type** | **Auckland** | **Canterbury** | **Bay of Plenty** | **Wellington** | **Waikato** | **Northland** |
| Saloon | 327 | 99 | 86 | 69 | 80 | 37 |
| Stationwagon | 306 | 165 | 79 | 92 | 72 | 69 |
| Hatchback | 296 | 67 | 48 | 57 | 61 | 28 |
| Roadbike | 163 | 20 | 11 | 48 | 15 | 14 |
| Trailer | 141 | 128 | 54 | 46 | 56 | 23 |
| Moped | 88 | 33 | 14 | 27 | 7 | 4 |
| Utility | 83 | 61 | 100 | 39 | 35 | 34 |
| Light Van | 71 | 24 | 8 | 22 | 8 | 2 |

* **Stolen cars in region**
  + SQL query

SELECT region

,country

,population

,density

,COUNT(\*) AS number\_cars

FROM locations loc

JOIN stolen\_vehicles st ON loc.location\_id = st.location\_id

WHERE st.vehicle\_type IS NOT NULL

GROUP BY region

,country

,population

,density

ORDER BY number\_cars DESC

* + Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **region** | **country** | **population** | **density** | **number\_cars** |
| Auckland | New Zealand | 1695200 | 343.09 | 1626 |
| Canterbury Region | New Zealand | 655000 | 14.72 | 660 |
| Bay of Plenty | New Zealand | 347700 | 28.8 | 442 |
| Wellington | New Zealand | 543500 | 67.52 | 417 |
| Waikato | New Zealand | 513800 | 21.5 | 369 |
| Northland Region | New Zealand | 201500 | 16.11 | 233 |

* + Pivot Table (rows: region, columns: Sum of population, Sum of density, Sum of number\_cars)

|  |  |  |  |
| --- | --- | --- | --- |
| **Region** | **'Population** | **'Density** | **'Number** |
| Auckland | 1695200 | 343.09 | 1626 |
| Canterbury | 655000 | 14.72 | 660 |
| Bay of Plenty | 347700 | 28.8 | 442 |
| Wellington | 543500 | 67.52 | 417 |
| Waikato | 513800 | 21.5 | 369 |
| Northland | 201500 | 16.11 | 233 |

**Visual Insights:**

The Visual Insights sheet shows analysis in charts that have been created from PivotTables, the process of which has been described above.  
  
Charts:

* Bar chart *Stolen cars by type in region*

Obraz zawierający tekst, zrzut ekranu, numer, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, numer

Opis wygenerowany automatycznie

* + Insight: Represents the number of stolen different types of vehicles by region. You can use slicers to see the number of different types of vehicles by region, or the selected type by several regions.
* Column chart *Stolen cars by weekday*

Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie

* + Insight: Represents the number of stolen cars by day of the week.
* Bar chart *Most popular stolen type*

Obraz zawierający tekst, zrzut ekranu, Czcionka, Równolegle

Opis wygenerowany automatycznie

* + Insight: Represents stolen cars by type of the vehicle.
* Choropleth map *Stolen cars in region*

Obraz zawierający tekst, zrzut ekranu, mapa

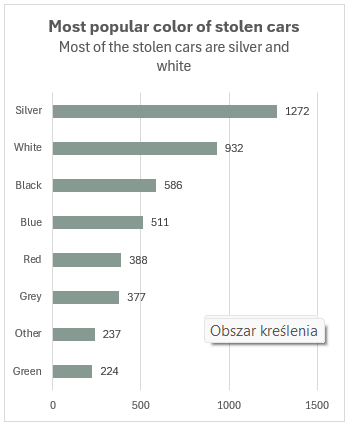
Opis wygenerowany automatycznie

* + Insight: Represents stolen cars in region on cartogram.
* Column chart *Average age of stolen cars*

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

* + Insight: Represents average age of stolen cars by type of the vehicle.
* Bar chart *Most popular color of stolen cars*



* + Insight: Represents most frequently color of stolen cars.
* Bubble chart *Stolen cars by population and density in region*

Obraz zawierający tekst, linia, zrzut ekranu, Wykres

Opis wygenerowany automatycznie

* + Insight: Represents number of stolen cars compare to the characteristic of a region: population and density.

**Conclusion:**

This project concerns the analysis of vehicle theft. The aim is to identify trends related to theft in different regions, vehicle types and other characteristics such as color, day of the week or average age of the vehicle. This report could be an announcement to encourage citiziens to be aware of thefts.