26/06/2025

Vehicle Thefts Analysis for Insurance Optimization in New Zealand

Romain Bardou

Table of Content

1. Executive Summary

This report analyzes vehicle theft data in New Zealand to optimize insurance pricing, reduce claim costs, and enhance customer segmentation. Using datasets (stolen_vehicles.xlsx, make_details.xlsx, locations.xlsx) and a 2025 vehicle fleet sample, key findings include:

- **Vehicle Trends**: Saloons (18.77% of thefts vs. 11.22% of fleet) and motorcycles (10.97% vs. 3.74%) are overrepresented, as are silver vehicles (28.03% vs. 21.6%).

- **Geographic Insights**: Gisborne has the highest theft rate per 100,000 inhabitants (335.89, score 1.0), while Marlborough and Southland have minimal thefts.

- **High-Risk Profiles**: Nissan Saloons (risk score 0.81) and vehicles in Auckland (overrepresentation 1.0) are most theft-prone, while Toyota Caravans in Southland (risk score -0.28) are low-risk.

- **Pricing Model**: A risk-based premium model Annual Premium = Base_Premium* (1 + Risk_Score) + fees) adjusts premiums.

- **Recommendations**: Adjust vehicle premiums by using the pricing model developed (+20% for Nissan, -7.5% for a Hatchback), promote anti-theft devices, target urban customers with comprehensive plans and target rural customers with budget plans.

- **Limitations** include missing seasonal data, recovery and anti-theft device information.

2. Introduction

In the New Zealand automotive insurance sector, vehicle thefts pose a significant challenge, directly impacting claim costs and premium pricing. As an insurance company, our goal is to optimize our pricing strategy, reduce theft-related claims, and segment our customer base to offer tailored products while remaining competitive and compliant with local regulations. Data on stolen vehicles, including brands, types, years of manufacture, colors, and geographic locations, provide a critical foundation for identifying risk factors and refining our strategies.

This report analyzes the provided datasets (stolen_vehicles.csv, make_details.xlsx, locations.csv), enriched by a comparison with a standardized sample of 100,000 records from the 2025 New Zealand vehicle fleet. The primary objectives are:

- Identify the characteristics of the most frequently stolen and their overrepresentation relative to the vehicle fleet.
- Analyze geographic variations in thefts.
- Develop a predictive pricing model based on risk scores, incorporating negative coefficients (from -0.5 to 1) for underrepresented categories to adjust premiums for low-risk segments.
- Formulate recommendations for adjusting premiums, mitigating risks, and targeting specific customer segments.

The analysis relies on thorough data cleaning (handling missing values, standardizing categories, creating unique identifiers) performed in Excel and Python, with visualizations created in Power BI to illustrate key trends. A significant limitation is the dataset's insufficient temporal coverage, which prevents analysis of seasonal trends. Additionally, data on anti-theft devices, theft methods, or recovery rates are missing. Driver-related information was excluded, as the dataset focuses solely on stolen vehicle characteristics, making such data less relevant for this specific model. A pricing model, based on a weighted risk score by category (brand, type, color, year, region) with negative coefficients for low-risk segments and fixed agency fees, has been developed to translate findings into actionable insights.

The report is structured as follows: the methodology details the data cleaning, analysis, and modeling steps, including the use of Python for data filtering, Excel for complementary calculations, and Power BI for visualizations. The results present key trends, the discussion interprets these findings in relation to strategic objectives, and the recommendations propose premium adjustments, preventive measures, and segmentation strategies.

This work aims to provide the company with actionable tools to optimize operations, reduce risk exposure, and enhance customer offerings while addressing the constraints of the available data.

3. Methodology

This section outlines the data preparation, analysis, modeling, and visualization methods used to analyze vehicle thefts in New Zealand, aiming to optimize insurance premium pricing, reduce claim costs, and segment customer risk profiles. The approach leverages cleaned datasets, a sample of the 2025 vehicle fleet, and statistical techniques to derive actionable insights.

3.1. Data Sources

The analysis uses three original datasets:

- **stolen_vehicles.csv**: Details vehicle thefts, including brand, type, year of manufacture, color, region, and date of theft.
- **make_details.csv**: Classifies brands (standard vs. luxury), linked via a make_id.
- **locations.csv**: Provides geographic data (regions, population, density), linked via a location_id.

A reference dataset, fleet.csv, representing the 2025 New Zealand vehicle fleet, was sourced online, containing data on vehicle brands, types, years of manufacture, colors, and regions. After cleaning, these files were converted to .xlsx format (stolen_vehicles.xlsx, make_details.xlsx, locations.xlsx, fleet.xlsx) for better Power BI compatibility. A random sample of 100,000 records from fleet.xlsx was selected to compare with the stolen vehicles data, noting potential minor differences (e.g., more electric vehicles in 2025 vs. 2022).

3.2. Data Preparation

Data cleaning and standardization ensured consistency across datasets.

- **Handling Missing Values**: Blanks in stolen_vehicles.csv (e.g., color, year of manufacture) were addressed by cross-referencing with make_details.csv.
- **Format Standardization**: Date formats were unified (e.g., DD/MM/YYYY), and categorical variables (brands, types, colors) were normalized (e.g., “Red” vs. “red” consolidated).
- **Category Column Creation**: A category column was added to stolen_vehicles.csv to generalize vehicle types (e.g., “utility,” “trailer,” “caravan”), aiding aggregation.
- **Unique IDs and Tables**: Unique IDs were created for each category (brand, type, color, region) to facilitate linking, with corresponding tables established to integrate data across datasets.
- **File Conversion**: Post-cleaning, stolen_vehicles.csv, make_details.csv, locations.csv, and fleet.csv were converted to .xlsx for Power BI compatibility.
- **Brand Filtering**: The fleet.csv dataset was loaded using Python (pandas) and filtered to retain only brands present in stolen_vehicles.csv. Brand names were standardized to proper case (e.g., “TOYOTA” → “Toyota”).
- **Sampling**: A random sample of 100,000 records was extracted from the filtered fleet.csv, saved as a .csv, and converted to fleet.xlsx after cleaning.
- **Final Standardization**: In Excel, the sampled fleet.xlsx was reduced to relevant columns (brand, type, color, year of manufacture, region) and aligned with stolen_vehicles.xlsx categories using the established IDs.

3.3. Analysis Methods

The analysis identified theft patterns, calculated risk scores, and estimated claim costs using descriptive statistics and normalized metrics.

<u>Descriptive Statistics</u>

Frequencies and Percentages calculated for brands, types, colors, years of manufacture, and regions in stolen_vehicles.xlsx and fleet.xlsx (e.g., percentage of Toyota thefts vs. Toyota's fleet share).

<u>Overrepresentation Analysis</u>

For each category (brand, type, color, year of manufacture), representation in stolen vehicles was compared to the fleet:

- **Overrepresentation**: If a category's share in thefts exceeded its fleet share (e.g., trailers = 60% of thefts but 20% of fleet), the difference was calculated as theft_share - fleet_share.

- **Normalization**: Positive differences were divided by the maximum difference (score: 0 to 1); negative differences by the minimum (score: 0 to -0.5).

- **Year of Manufacture**: Analyzed to assess whether newer or older vehicles were over- or underrepresented

<u>Geographic Analysis</u>

- **Thefts per 100,000 Inhabitants**: Theft counts per region were normalized by population (from locations.xlsx) to calculate theft rates per 100,000 inhabitants.
- **Regional Risk Score**: Differences between regional theft rates and the national average were normalized: positive by the maximum (score: 0 to 1), negative by the minimum (score: 0 to -0.5).
- **Dual Measures**: Two regional scores were derived: overrepresentation (theft share vs. fleet share) and thefts per 100,000 inhabitants.

3.4. Risk Scoring and Pricing Model

A predictive pricing model translated theft risk into premiums.

<u>Risk Score Calculation</u>

- **Category Scores**: Brand, type, color, and year of manufacture received scores from -0.5 to 1 based on normalized overrepresentation (Section 5.3.2).
- **Regional Scores**: Two scores per region (overrepresentation, thefts per 100,000 inhabitants), ranging from -0.5 to 1.
- **Total Risk Score**:

Risk_Score = 0.2 * Brand + 0.15 * Type + 0.05 * Color + 0.2 * Year + 0.2 * Region_Overrepresentation + 0.2 * Region_Risk

Weights: Brand (20%) and year of manufacture (20%) reflect predictive power; type (15%) for category significance; color (5%) for overrepresentation; each region (20%) for geographic risk.

Year: Refers to the vehicle's year of manufacture, assessing risk based on age

Range: Total score ranges from -0.5 to 1, enabling premium reductions for low-risk vehicles.

<u>Pricing Model</u>

- **Formula**:

Annual Premium = Base_Premium * (1 + Risk_Score) + Fees

- Base_Premium: Vehicle estimated value multiplied by a coefficient. (See part 6.4)
- **Fees**: Fixed agency fees of 300 NZD, based on a New Zealand insurer average.
- **Implementation**: Built in Excel with dropdowns for brand, type, color, year of manufacture, and region, with manual numeric inputs (e.g., vehicle estimated value, year).

3.5. Visualizations

In Power BI, using .xlsx datasets, created interactive visualizations:

- **Histograms**: Theft counts by brand, vehicle type, colour and location.
- **Heatmaps**: Theft intensity by region, based on thefts per 100,000 inhabitants.
- **Line Charts**: Theft counts by year of manufacture.
- **Tables**: Risk scores.

3.6. Tools

- **Excel**: Data cleaning, standardization, pricing model.
- **Python**: Filtering, sampling large dataset.
- **Power BI**: Visualization.
- **Online Research**: Supplemented missing data (e.g., fleet.csv).

3.7. Limitations

- **Temporal Coverage**: Insufficient for seasonality, with focus on year of manufacture.
- **Data Gaps**: No anti-theft device, theft method, or recovery rate data.
- **Fleet Sample**: 2025 fleet.xlsx may differ from 2022 (e.g., more electric vehicles).
- **Driver Data**: Excluded, as stolen_vehicles.xlsx focuses on vehicle characteristics and no data.

This methodology provides a robust framework for analyzing vehicle thefts, calculating risk scores, and developing a pricing model. Results and visualizations follow.

4. Results

This section presents the findings from the analysis of vehicle thefts in New Zealand, based on the methods outlined in the Methodology section. Results begin with visualizations of observed trends, followed by overrepresentation analysis, geographic insights, risk scores, and insurance premium estimates.

4.1. Visualizations

- **Brands**:

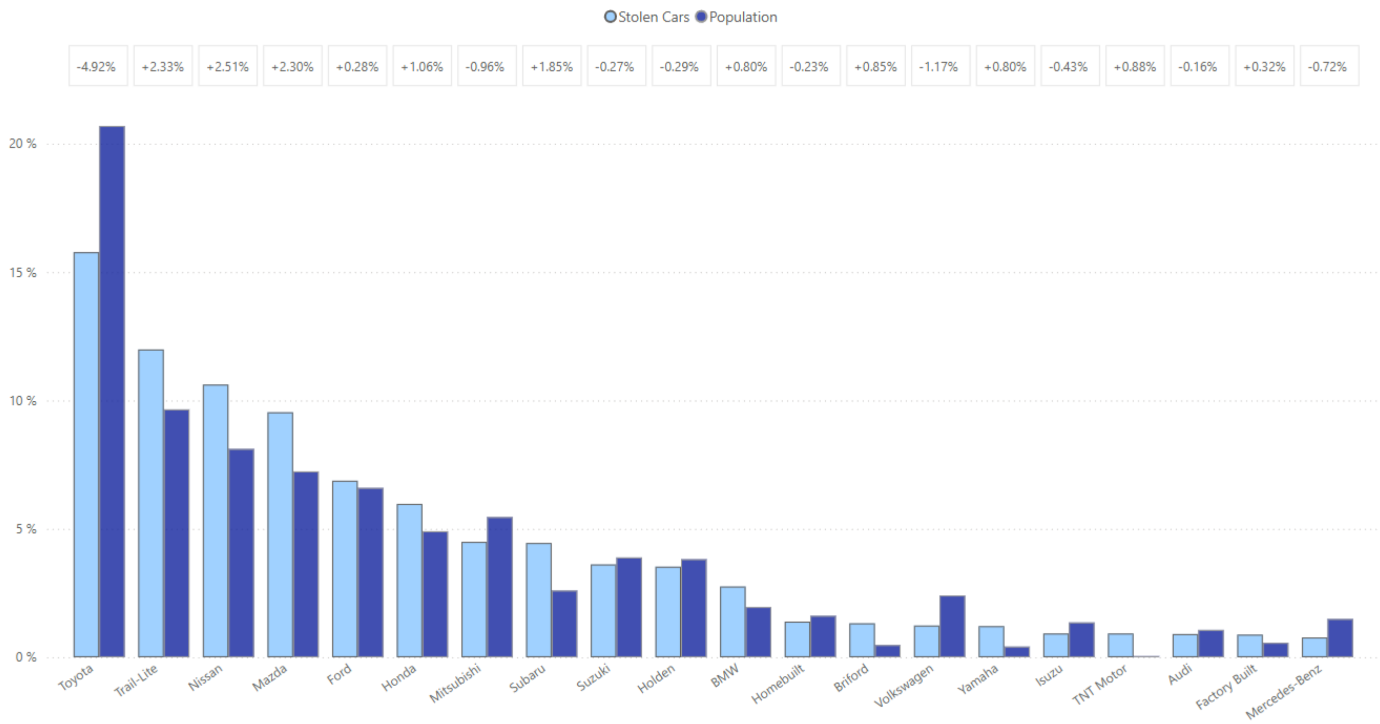

Figure 01: Representation of Brands in Stolen Cars and Car Population

This bar chart shows the distribution of thefts by brand, with brands like Toyota or Nissan dominating thefts but one is under-represented while the other one is over-represented, probably because Toyota is harder to re-sell or to steal, and this should be valuated. The percentages at the top show the difference between the representation of the brand in the stolen car population and the representation of the brand in the car population in New Zealand. This clear concentration suggests that brand data should be incorporated into the risk score calculation to account for high-risk manufacturers, including this representation’s insight.

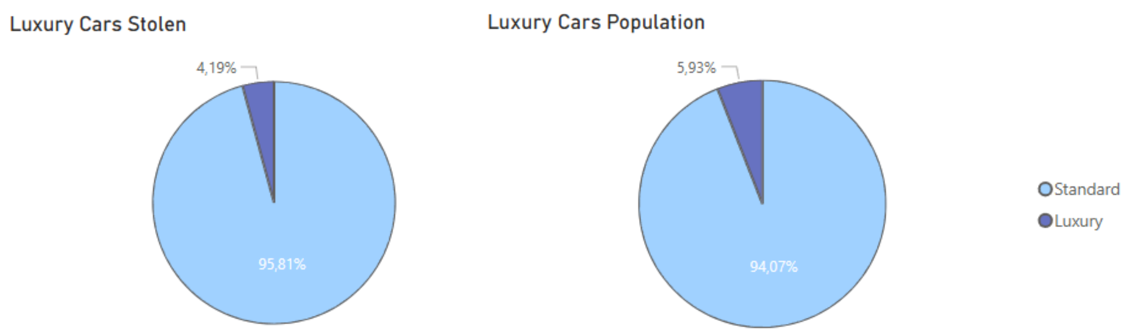

Figure 02: Representation of Luxury Brands in Stolen Cars and Car Population

These pie charts compare the proportion of luxury vs. non-luxury vehicles in thefts and the fleet. If the theft share of luxury vehicles aligns closely with their fleet share, there is no real impact of luxury vehicle status on theft risk, and it may not need adjustment in the risk score.

- **Vehicle Types**:

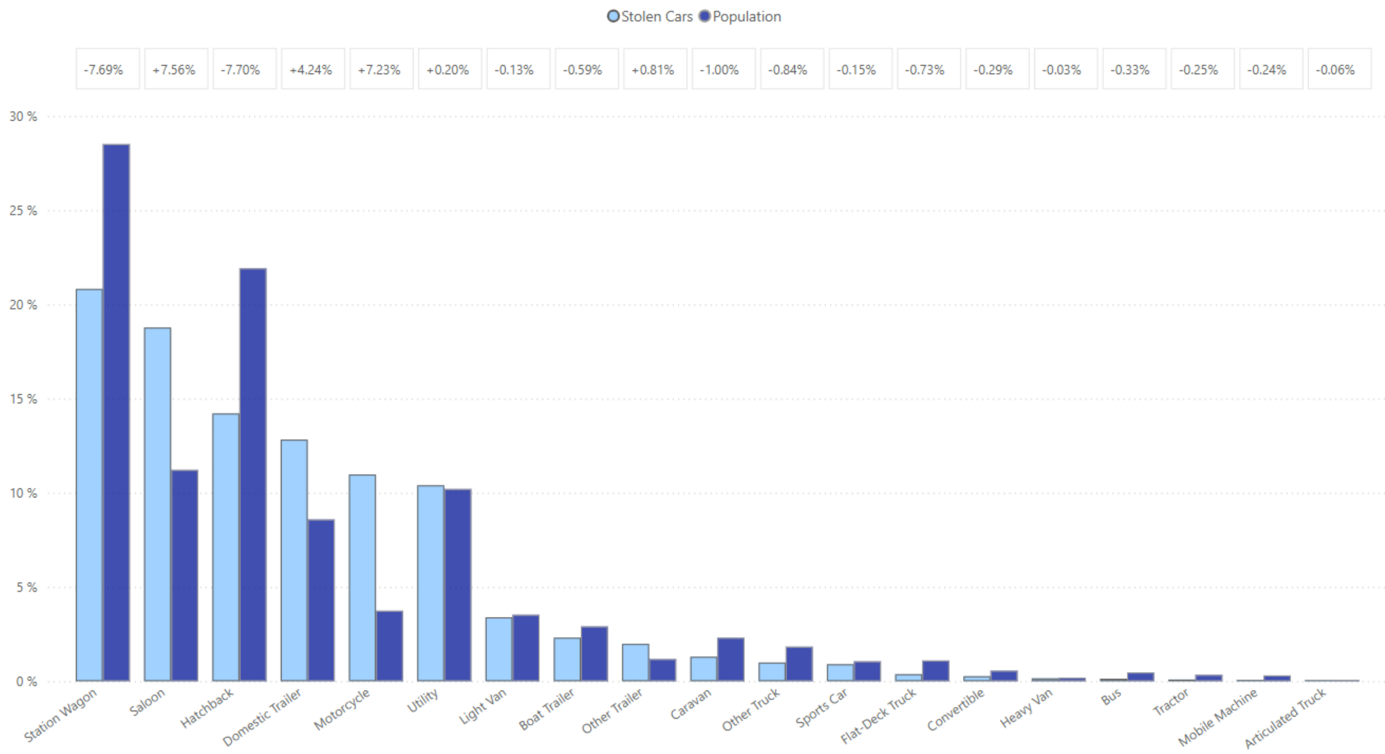

Figure 03: Representation of Vehicle Types in Stolen Cars and Car Population

This bar chart highlights vehicle type distribution, with Station Wagons and Saloons leading. The percentages at the top show the difference between the representation of the type in the stolen car population and the representation of the brand in the car population in New Zealand. Mostly, car types are leading, but only Saloons are over-represented, so this level of detail is more relevant to use in our model than just the type of vehicle. Given this, it would be valuable to include this data in the risk score design to reflect these high-risk categories.

- **Colors**:

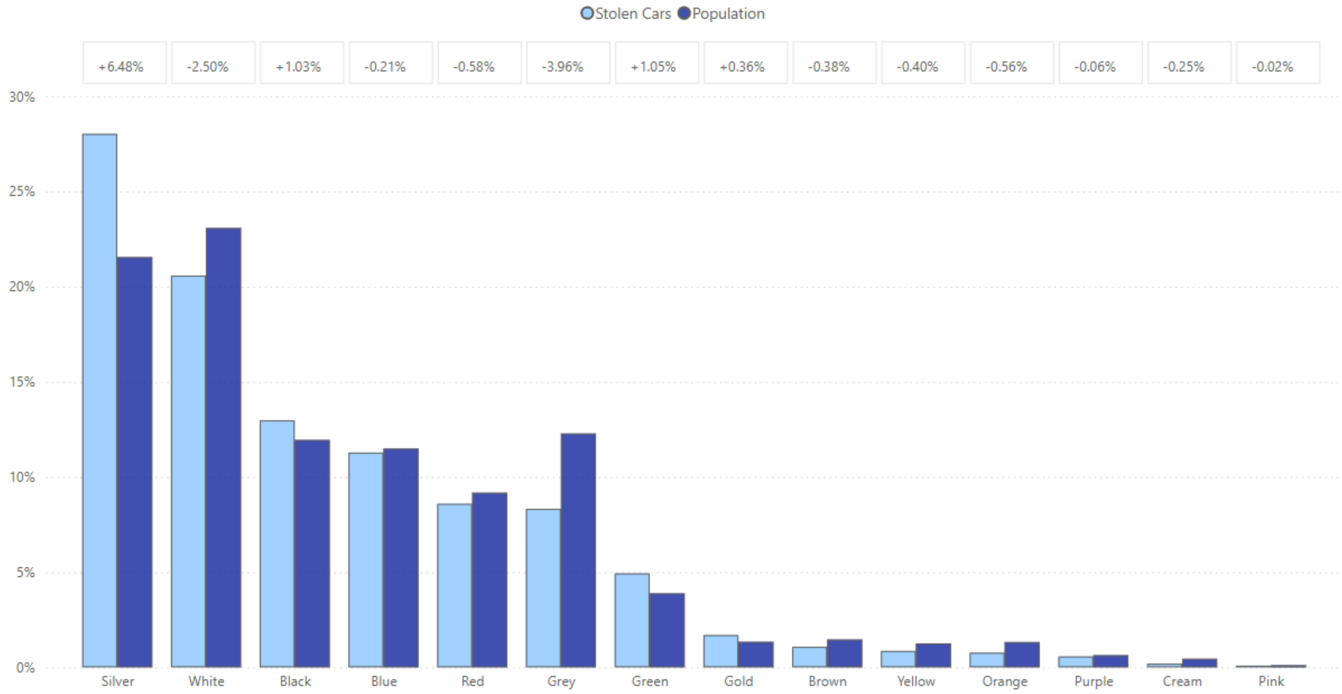

Figure 04: Representation of Colours in Stolen Cars and Car Population

This bar chart indicates color-based theft trends, with colors like silver or black overrepresented. The percentages at the top show the difference between the representation of the colour in the stolen car population and the representation of the brand in the car population in New Zealand. Based on this pattern, color data should be considered for the risk score at a low weight to capture this risk factor.

- **Manufacturing Years**:

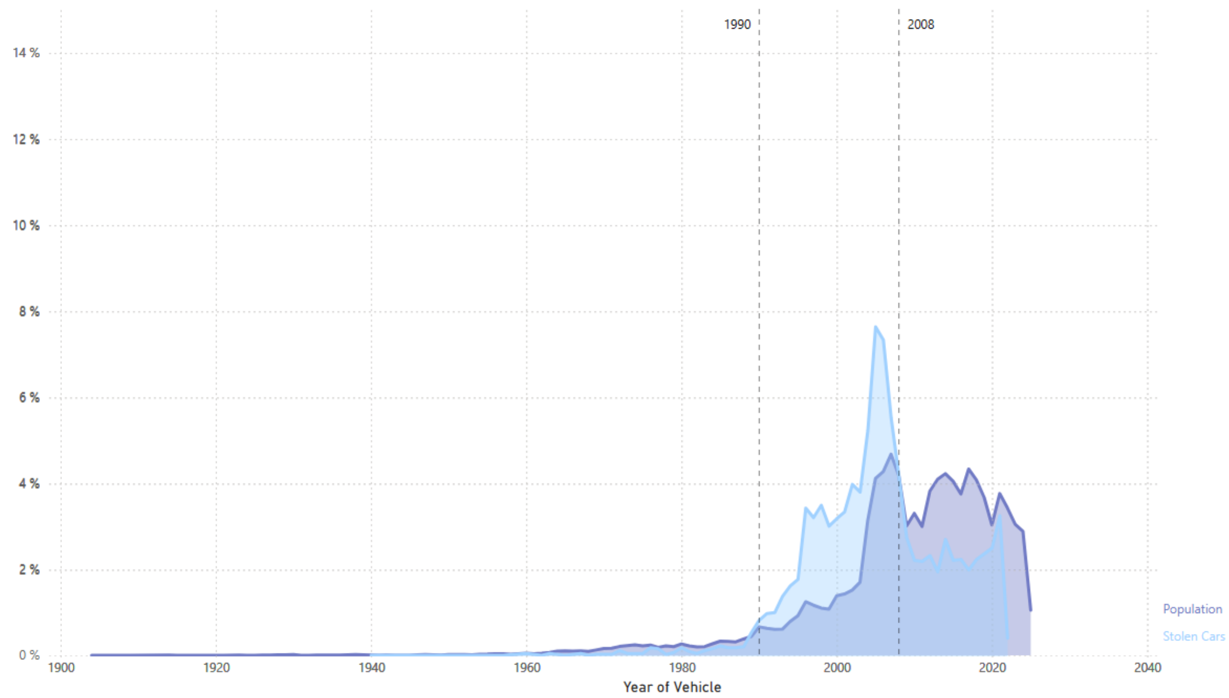

Figure 05: Repartition of Manufacture's Years within Stolen Cars and Population

This line chart likely shows theft trends by manufacturing year, with an upward trend for vehicles from 1990 and 2008, probably due to the new security technologies for newer vehicles and low amount of older ones. This trend supports the inclusion of manufacturing year as a key factor in the risk score design.

4.2. Over-Representation Trends

Based on the visualizations, the overrepresentation analysis reveals significant disparities across categories, comparing theft shares to fleet shares and normalizing scores from 0 to 1 (overrepresentation) or 0 to -0.5 (underrepresentation), as outlined in the methodology.

- **Brands**:

The analysis indicates varying theft prevalence among brands, with some overrepresented and others underrepresented. This disparity highlights the need to include brand data in the risk score, weighted at 20%. Nissan is the most over-represented, so the rate is 1, and the opposite, Toyota will have a rate of -0.5. (See 8.1 in Appendix for the full table)

- **Vehicle Types**:

Vehicle_Type	% of Stolen Cars	% of Car Population	Overrepresentation	Rate
Articulated Truck	0,02%	0,08 %	-0,06	0,00
Boat Trailer	2,31%	2,91 %	-0,59	-0,04
Bus	0,13%	0,46 %	-0,33	-0,02
Caravan	1,30%	2,30 %	-1,00	-0,07
Convertible	0,26%	0,56 %	-0,29	-0,02
Domestic Trailer	12,83%	8,59 %	4,24	0,56
Flat-Deck Truck	0,37%	1,10 %	-0,73	-0,05
Hatchback	14,21%	21,91 %	-7,70	-0,50
Heavy Van	0,15%	0,19 %	-0,03	0,00
Light Van	3,39%	3,53 %	-0,13	-0,01
Mobile Machine	0,07%	0,31 %	-0,24	-0,02
Motorcycle	10,97%	3,74 %	7,23	0,96
Other Trailer	1,98%	1,17 %	0,81	0,11
Other Truck	0,99%	1,83 %	-0,84	-0,05
Saloon	18,77%	11,22 %	7,56	1,00
Sports Car	0,90%	1,05 %	-0,15	-0,01
Station Wagon	20,82%	28,51 %	-7,69	-0,50
Tractor	0,09%	0,34 %	-0,25	-0,02
Utility	10,40%	10,20 %	0,20	0,03

Figure 06: Over and Under-Representation of Vehicle Type and Rate

Saloons (18.77% of thefts vs. 11.22% of fleet) and motorcycles (10.97% vs. 3.74%) are highly overrepresented (scores of 1.00 and 0.96), while caravans and station wagons are underrepresented (scores of -0.07 and -0.50). This supports a 15% weight for vehicle type in the risk score.

- **Colors**:

Colour	% of Stolen Cars	% of Car Population	Overrepresentation	Rate
Black	12,98 %	11,9 %	1,03	0,16
Blue	11,28 %	11,5 %	-0,21	-0,03
Brown	1,08 %	1,5 %	-0,38	-0,05
Cream	0,20 %	0,4 %	-0,25	-0,03
Gold	1,70 %	1,3 %	0,36	0,06
Green	4,94 %	3,9 %	1,05	0,16
Grey	8,33 %	12,3 %	-3,96	-0,50
Orange	0,77 %	1,3 %	-0,56	-0,07
Pink	0,09 %	0,1 %	-0,02	0,00
Purple	0,57 %	0,6 %	-0,06	-0,01
Red	8,59 %	9,2 %	-0,58	-0,07
Silver	28,03 %	21,6 %	6,48	1,00
White	20,58 %	23,1 %	-2,50	-0,32
Yellow	0,86 %	1,3 %	-0,40	-0,05

Figure 07: Over and Under-Representation of Colour and Rate

Silver (28.03% of thefts vs. 21.6% of fleet) and Black (12.98% vs. 11.9%) are overrepresented (scores of 1.00 and 0.16), while grey is underrepresented (-0.50). This will be weighted at 5% for color in the risk score.

- **Manufacturing Years**:

The data suggests varying theft risks by vehicle age. The trend of higher thefts in certain years (e.g., 1990-2008) supports a 20% weight for year in the risk score. We will apply the following coefficient: between 1990 and 1995, rate is 0.3, between 1995 and 2008 included, rate is 1, and for all other years, rate is 0.

4.3. Geographic Analysis:

The geographic distribution of thefts, shows varying rates across regions, analyzed through overrepresentation and theft rates per 100,000 inhabitants.

Figure 07: Map Showing the repartition of Stolen Vehicles in New Zealand

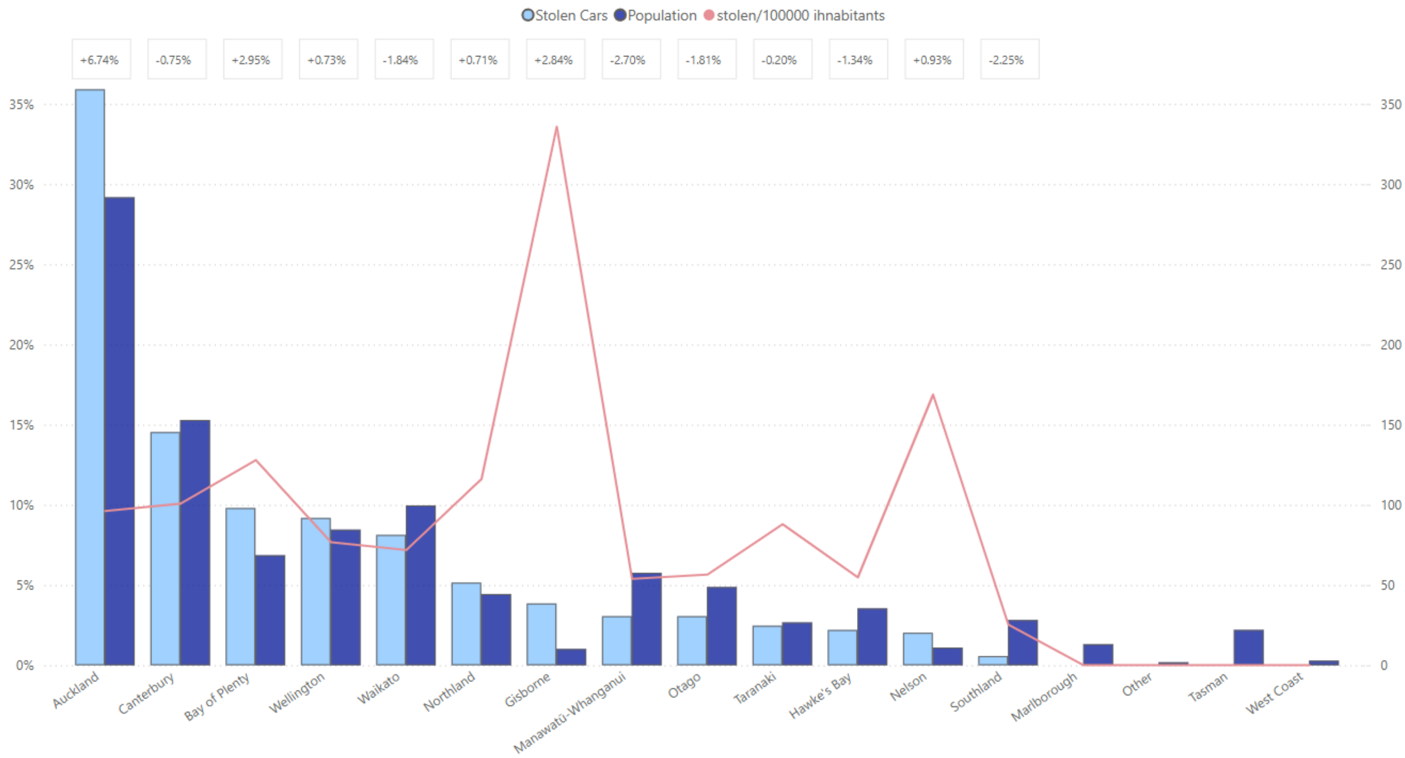

Figure 08: Number of Stolen Vehicles for 100,000 inhabitants for each Region and their Representation in Stolen Cars and Car Population

Region	% of Stolen Cars	% of Car Population	Overrepresentation	Rate
Auckland	35,92%	29,2 %	6,74	1,00
Bay of Plenty	9,81%	6,9 %	2,95	0,44
Canterbury	14,54%	15,3 %	-0,75	-0,14
Gisborne	3,86%	1,0 %	2,84	0,42
Hawke's Bay	2,20%	3,5 %	-1,34	-0,25
Manawatū-Whanganui	3,06%	5,8 %	-2,70	-0,50
Marlborough	0,00%	1,3 %	-1,32	-0,24
Nelson	2,03%	1,1 %	0,93	0,14
Northland	5,16%	4,4 %	0,71	0,11
Otago	3,06%	4,9 %	-1,81	-0,34
Other	0,00%	0,2 %	-0,19	-0,03
Southland	0,57%	2,8 %	-2,25	-0,42
Taranaki	2,47%	2,7 %	-0,20	-0,04
Tasman	0,00%	2,2 %	-2,21	-0,41
Waikato	8,13%	10,0 %	-1,84	-0,34
Wellington	9,19%	8,5 %	0,73	0,11
West Coast	0,00%	0,3 %	-0,29	-0,05

Figure 09: Over and Under-Representation of Region and Rate

Auckland which represents 35.92% of thefts versus 29.2% of fleet is the most overrepresented region (score 1.00), while Southland and Tasman are underrepresented (scores -0.42 and -0.41).

Region	Stolen Cars per 100,000 inhabitants	% from Average	Rate
Auckland	96,15	+9%	0,03
Bay of Plenty	127,98	+44%	0,16
Canterbury	100,76	+14%	0,05
Gisborne	335,89	+279%	1,00
Hawke's Bay	54,73	-38%	-0,19
Manawatū-Whanganui	53,83	-39%	-0,20
Marlborough	0,00	-100%	-0,50
Nelson	168,81	+91%	0,32
Northland	116,13	+31%	0,11
Otago	56,50	-36%	-0,18
Other	0,00	-100%	-0,50
Southland	25,39	-71%	-0,36
Taranaki	87,98	-1%	0,00
Tasman	0,00	-100%	-0,50
Waikato	71,82	-19%	-0,09
Wellington	76,72	-13%	-0,07
West Coast	0,00	-100%	-0,50

Figure 10: Rating for Stolen Cars per inhbitants

Gisborne (335.89 per 100,000) being the riskiest region in our dataset, it has the highest rate (1.00), while Southland and Marlborough have no stolen cars recorded, and so have the lowest score (-0.50).

The importance of having 2 scores for the population can be seen with Auckland, which is over-representated in terms of stolen car, but, with a high density of population, there are not that much stolen cars per 100,000 people, close to the average, while this statistic is very high for Gisborne.

This analysis, with each regional factor weighted at 20% in the risk score, underscores the importance of considering both overrepresentation and population-adjusted theft rates for accurate risk assessment.

4.4. Risk Scores and Pricing Model

Risk score, derived from the observed trends and over or under-representation, is calculated as shown earlier:

Risk_Score = 0.2 * Brand + 0.15 * Type + 0.05 * Color + 0.2 * Year + 0.2 * Region_Overrepresentation + 0.2 * Region_Risk

where Region_Risk is the thefts per 100,000 inhabitants score. Each category score ranges from -0.5 to 1, and the total score drives the pricing model.

- **Risk Score Examples**:

  Nissan Saloon 2000 in Auckland:

  - Brand: Nissan = 1.0
  - Type: Saloon = 1.0
  - Color: [Assume Silver] = 1.0
  - Year: 2000 (1995-2008) = 1.0
  - Region's Representation: Auckland = 1.0
  - Region's Risk: Auckland = 0.03
  - **Total** = (0.2 * 1.0) + (0.15 * 1.0) + (0.05 * 1.0) + (0.2 * 1.0) + (0.2 * 1.0) + (0.2 * 0.03) = **0.806**

  Toyota Caravan 1985 in Southland:

  - Brand: Toyota = -0.5
  - Type: Caravan = -0.07
  - Color: [Assume White] = -0.32
  - Year: 1985 (outside 1990-2008) = 0
  - Region's Representation: Southland = -0.42
  - Region's Risk: Southland = -0.36
  - **Total** = (0.2 * -0.5) + (0.15 * -0.07) + (0.05 * -0.32) + (0.2 * 0) + (0.2 * -0.42) + (0.2 * -0.36) = **-0.2825**

The base insurance value is determined by multiplying the estimated vehicle price by a coefficient, which is then adjusted by the risk score and agency fees.

- **Price Coefficients**: (See Appendix 8.2)

Estimated Price	Coefficient
9000	0.0800
10000	0.0738
...	...
78000	0.0250
79000	0.0249
80000	0.0248

Estimated_Price * Coefficient = **Base_Premium**

- **Annual Premium Estimates**:

As shown earlier, the annual premium estimate is:

Annual Premium = Base_Premium * (1 + Risk_Score) + Fees

Where Fees are agency fees = 300 NZD.

Nissan Saloon 2000 in Auckland (estimated price 30,000 NZD):

- coefficient for 30,000 NZD = 0.0364
- Base value = 30,000 * 0.0364 = 1,092 NZD
- **Premium** = 1,092 * (1 + 0.806) + 300 = **2,272.35 NZD**

Toyota Caravan 1985 in Southland (estimated price 9,000 NZD):

- coefficient for 9,000 NZD = 0.0800
- Base value = 9,000 * 0.0800 = 720 NZD
- **Premium** = 720 * (1 - 0.2825) + 300 = **816.6 NZD**

4.5. Summary

The analysis reveals significant theft patterns in New Zealand, with Nissan Saloons (risk score 0.806) and vehicles in Auckland (overrepresentation 1.0) identified as high-risk profiles, driving premiums up to 2,272.35 NZD for a 30,000 NZD vehicle. Conversely, Toyota Caravans in Southland (risk score -0.2825) represent low-risk segments, reducing premiums to 816.6 NZD for a 9,000 NZD vehicle. Silver (1.0) and Saloons (1.0) are overrepresented, while Grey (-0.50) and Station Wagons (-0.50) are underrepresented. Moreover, geographic data highlights Gisborne's high theft/inhabitants rate (1.0) and Marlborough's low rate (-0.50). These findings support tailored pricing and segmentation, though limitations such as missing seasonal trends and anti-theft device data constrain broader insights.

5. Discussion

The results align with the objectives of optimizing pricing, reducing claims, and segmenting customers. Nissan's overrepresentation (1.0) suggests higher theft susceptibility, possibly due to demand in black markets, while Toyota's underrepresentation (-0.5) may reflect anti-theft features or lower resale value, warranting further investigation. Auckland's high risk (0.806 combined score) reflects urban theft pressures, contrasting with Southland's low risk (-0.2825), supporting geographic pricing adjustments. The 20% weight on years (1995-2008 = 1.0) underscores older vehicles' vulnerability, likely due to outdated security. However, data gaps (e.g., recovery rates, driver behavior) limit the model's precision, suggesting future data collection could refine risk scores.

6. Recommendations

- **Premium Adjustments**: Use the model developed to increase premiums for high-risk categories (e.g., Nissan, +20%) and reduce premiums for low-risk categories (e.g., Toyota, -10%). For example, a Nissan Saloon valued at 30,000 NZD would have a premium of 2,272.35 NZD (up from 1,872.35 NZD without adjustment) and a Toyota Caravan valued at 9,000 NZD would have a premium of 816.6 NZD (down from 1,020 NZD).

- **Risk Mitigation**: Offer 5–10% premium discounts for anti-theft devices (e.g., GPS trackers), following practices of New Zealand insurers. For example, a 10% discount on a 2,272.35 NZD premium saves 227.24 NZD annually.

- **Customer Segmentation**: Target urban customers (Auckland, Gisborne) with comprehensive coverage and rural customers (Southland, Marlborough) with budget plans. For example, the <u>comprehensive plan</u> might include 15% premium increase to cover theft and damage risks. Promote these via targeted social media ads (e.g., Facebook, Instagram) aimed at drivers aged 25–45. <u>Budget plans</u> for rural customers in Southland, Marlborough, and Tasman, might include 10–15% premium reductions and basic coverage (third-party, fire, and theft). Market these through local radio and agricultural fairs.

- **Data Enhancement**: Collect seasonal theft data and anti-theft device statistics to refine the model.

- **Marketing**: Launch a digital campaign advertising the customer segmentation recommendation, to target Auckland and Gisborne via social media and Google Ads, aiming to attract new clients.
  Host community events in rural regions (Southland, Marlborough) with partners like Federated Farmers to promote budget plans, offering free insurance consultations.

7. Sources

- **stolen_vehicles.csv, make_details.csv, location.csv**:
  Datasets provided by Kaplan Business School in the Data Visualization subject.

- **fleet.csv**: https://www.nzta.govt.nz/resources/new-zealand-motor-vehicle-register-statistics/new-zealand-vehicle-fleet-open-data-sets/

- **To understand and build the model**: https://www.icnz.org.nz/
  https://www.nzta.govt.nz/resources/new-zealand-motor-vehicle-register-statistics/
  https://www.aainsurance.co.nz/
  https://www.fma.govt.nz/

8. Appendix

8.1 Brands' Rates

Brand	Luxury	Rate
Aakron Xpress	Standard	0.00
ADLY	Standard	0.00
Alpha	Standard	0.00
Anglo	Standard	0.00
Aprilia	Standard	0.00
Atlas	Standard	0.08
Audi	Standard	0.00
Bailey	Standard	-0.02
Bedford	Standard	0.00
Benelli	Standard	0.00
Bentley	Luxury	0.00
BMW	Luxury	0.00
Bricon	Standard	0.32
Briford	Standard	0.00
Buell	Standard	0.34
Buffalo	Standard	0.00
Cadillac	Luxury	0.00
Can-Am	Standard	0.00
Caravan	Standard	0.00
Caterpillar	Standard	-0.03
Chery	Standard	0.00
Chevrolet	Standard	0.00
Chrysler	Standard	-0.02

Citroen	Standard	0.00
Classic	Standard	0.00
Crusader	Standard	0.00
Custombuilt	Standard	0.00
Dacia	Standard	0.00
Daewoo	Standard	0.00
DAF	Standard	0.00
Daihatsu	Standard	0.00
Diamond	Standard	0.00
DMW	Standard	0.00
Dodge	Standard	0.00
Domett	Standard	0.00
Ducati	Standard	0.00
Elddis	Standard	0.00
Factory Built	Standard	0.00
Ferrari	Luxury	0.13
Ford	Standard	0.00
Forza	Standard	0.11
FOTON	Standard	0.11
Fuso	Standard	0.00
Great Wall	Standard	-0.01
Harley Davidson	Standard	0.00
Hino	Standard	0.00
Hitachi	Standard	-0.04
Holden	Standard	0.00
Homebuilt	Standard	-0.03
Honda	Standard	-0.02
Hoskings	Standard	0.42
Husaberg	Standard	0.00
Husqvarna	Standard	0.00
Hyosung	Standard	0.00
Hyundai	Standard	0.07
Isuzu	Standard	-0.20
Jaguar	Luxury	-0.04
Jayco	Standard	-0.02
Jeep	Standard	0.00
John Deere	Standard	-0.03
Kawasaki	Standard	-0.02
Kea	Standard	0.12
Keeway	Standard	0.15
Kia	Standard	0.00
KTM	Standard	-0.17
Kymco	Standard	0.21
Lambretta	Standard	0.00
Land Rover	Luxury	0.00

Landrover	Luxury	-0.03
Lexus	Luxury	-0.02
Liteweight	Standard	-0.05
Lochiel	Standard	0.00
Mahindra	Standard	0.00
Maserati	Luxury	0.00
Massey	Standard	0.00
Mazda	Standard	0.00
Mercedes-Benz	Luxury	0.92
MG	Standard	-0.07
Mini	Luxury	-0.04
Mitsubishi	Standard	-0.02
Mitsubishio Fuso	Standard	-0.10
Mobile Machine	Standard	0.00
Moden	Standard	0.00
Mono - Way	Standard	0.00
Moped	Standard	0.00
Morris	Standard	0.18
Nissan	Standard	-0.01
Nissan Diesel	Standard	1.00
Niu	Standard	0.00
Over	Standard	0.00
Oxford	Standard	0.00
Peugeot	Standard	0.00
PGO	Standard	-0.01
Piaggio	Standard	0.04
Pinto	Standard	0.13
Porsche	Luxury	0.00
Reid	Standard	-0.02
Renault	Standard	0.00
Rhino	Standard	-0.01
Rover	Luxury	0.00
Royal Enfield	Standard	0.00
Saab	Standard	0.00
Scomadi	Standard	0.00
Seat	Standard	0.00
Skoda	Standard	0.00
Sprite	Standard	-0.03
Ssangyong	Standard	0.00
Steelbro	Standard	-0.03
Sterling	Standard	0.00
Subaru	Standard	0.00
Suzuki	Standard	0.74
Swift	Standard	-0.03
SYM	Standard	0.00

Takeuchi	Standard	0.00
TGB	Standard	0.00
Titan	Standard	0.00
TNT Motor	Standard	0.08
Toko	Standard	0.35
Toyota	Standard	0.00
Toyota Lexus	Luxury	-0.50
Tractor	Standard	0.00
Trailer	Standard	0.00
Trail-Lite	Standard	0.00
Trike	Standard	0.93
Triumph	Standard	0.00
Trojan	Standard	-0.01
Ubco	Standard	0.00
Universal	Standard	0.00
Vespa	Standard	0.00
Veteran	Standard	0.08
Victory	Standard	0.00
Vmoto	Standard	0.00
Volkswagen	Standard	0.00
Volvo	Luxury	-0.12
Voyager	Standard	-0.03
Yamaha	Standard	0.00
Zephyr	Standard	0.32
Znen	Standard	0.00

8.2 Premium Coefficients

Estimated Price	Coefficient
0.00	0.0800
9000	0.0800
10000	0.0738
11000	0.0687
12000	0.0644
13000	0.0609
14000	0.0578
15000	0.0551
16000	0.0528
17000	0.0507
18000	0.0489
19000	0.0473
20000	0.0458
21000	0.0444
22000	0.0432
23000	0.0421
24000	0.0411
25000	0.0402
26000	0.0393
27000	0.0385
28000	0.0378
29000	0.0371
30000	0.0364
31000	0.0358
32000	0.0353
33000	0.0347
34000	0.0342
35000	0.0338
36000	0.0333

Estimated Price	Coefficient
37000	0.0329
38000	0.0325
39000	0.0321
40000	0.0318
41000	0.0314
42000	0.0311
43000	0.0308
44000	0.0305
45000	0.0302
46000	0.0300
47000	0.0297
48000	0.0294
49000	0.0292
50000	0.0290
51000	0.0288
52000	0.0285
53000	0.0283
54000	0.0281
55000	0.0280
56000	0.0278
57000	0.0276
58000	0.0274
59000	0.0273
60000	0.0271
61000	0.0270
62000	0.0268
63000	0.0267
64000	0.0265
65000	0.0264

Estimated Price	Coefficient
66000	0.0263
67000	0.0261
68000	0.0260
69000	0.0259
70000	0.0258
71000	0.0257
72000	0.0256
73000	0.0254
74000	0.0253
75000	0.0252
76000	0.0251
77000	0.0251
78000	0.0250
79000	0.0249
80000	0.0248