Insurance Database Implementation and Analysis
Technical Documentation

[George Wright]

December 9, 2024

Contents

Introduction

This document describes how I built and analyzed a database for vehicle insurance data. I created this system to help better understand insurance risks and set fair prices for different types of vehicles.

Project Overview

I started with a large dataset containing information about different vehicle insurance policies. My goal was to turn this data into useful insights that could help make better business decisions. The project covered everything from setting up the initial database to performing detailed analysis of insurance patterns.

What I Did

My work included:

- Creating a database system using SQL to store all insurance information
- Building tools to load and check data accuracy
- Analyzing patterns in insurance claims across different vehicle types
- Studying how well different vehicle categories performed over time
- Making recommendations based on what I learned from the data

Purpose

The main purpose of this work was to:

- Better understand which vehicles have higher insurance risks
- Help set more accurate insurance prices
- Make better business decisions based on real data
- Find ways to improve insurance operations

1 Data Source and Database Setup

This project began with vehicle insurance data obtained from Kaggle (https://www.kaggle.com/datasets/imtkaggleteam/vehicle-insurance-data?resource=download). The dataset provides comprehensive information about vehicle insurance policies, including customer details, policy specifications, and claim information.

The implementation process started with the creation of a new database to house the insurance data. The first step involved accessing the MySQL command line interface and creating a dedicated database for our project:

```sql
CREATE DATABASE insurance_project;
USE insurance_project;
```

Listing 1: Creating the database

After establishing the database, the table structure was designed and implement the table structure that would store our insurance data. The table design required careful consideration of data types and constraints to ensure data integrity and efficient storage. The primary table was created using the following SQL query:

```sql
CREATE TABLE vehicle_insurance (
    id INT AUTO_INCREMENT PRIMARY KEY,
    sex VARCHAR(1),
    insr_begin DATE,
    insr_end DATE,
    effective_yr INT,
    insr_type INT,
    insured_value DECIMAL(12,2),
    premium DECIMAL(10,2),
    object_id BIGINT,
    prod_year INT,
    seats_num INT,
    carrying_capacity DECIMAL(10,2),
    type_vehicle VARCHAR(50),
    ccm_ton DECIMAL(10,2),
    make VARCHAR(50),
    usage VARCHAR(50),
    claim_paid DECIMAL(12,2),
    created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
);
```

Listing 2: Creating the vehicle insurance table

The table structure was designed with several important considerations. The id field serves as the primary key with an auto-increment feature, ensuring each record has a unique identifier. For monetary values such as insured_value, premium, and claim_paid, DECIMAL data types were chosen with sufficient precision to handle financial calculations accurately. Date fields were implemented to track insurance periods, while VARCHAR fields were sized appropriately for categorical data like vehicle type and make.

The created_at field, automatically populated with a timestamp when a record is inserted, provides valuable metadata for tracking when records enter the system. This timestamp helps maintain data lineage and supports audit requirements, as it can track exactly when each insurance record was added to the database.

Each field in the table was carefully selected to match the data requirements:

- Temporal data (insr_begin, insr_end) uses the DATE type for accurate period tracking
- Numerical identifiers (id, object_id) use appropriate integer types based on their expected ranges
- Monetary values use DECIMAL type with sufficient precision to avoid rounding errors
- Categorical data uses VARCHAR with lengths chosen to balance storage efficiency with data completeness
- Boolean-like data (sex) uses single-character VARCHAR to allow for potential additional categories

This database structure provides a solid foundation for storing and analyzing the vehicle insurance data, ensuring data integrity while maintaining efficient storage and retrieval capabilities.

2 Data Loading Implementation

After establishing the database structure, the next step involved developing a Python script to load the insurance data into the database. This process required careful consideration of data cleaning, error handling, and efficient data insertion techniques.

The implementation uses several Python libraries: pandas for data manipulation, pyodbc for database connectivity, and datetime for handling temporal data. Here's a detailed examination of the data loading process:

```python
def connect_to_db():
    """Establish database connection"""
    return pyodbc.connect(
        'Driver={SQL Server};'
        'Server=DESKTOP-451APUQ\SQLEXPRESS;'
        'Database=InsuranceDB;'
        'Trusted_Connection=yes;'
    )
```

Listing 3: Database connection function

The connection function establishes a secure link to the SQL Server database using Windows Authentication, which provides a more secure connection method compared to username/password authentication.

A particularly important aspect of the data loading process involves cleaning numeric values to ensure compatibility with SQL Server:

```python
def clean_numeric(val):
    try:
        cleaned = float(val)
        return round(cleaned, 2)
```

```
    except:
        return 0
```

Listing 4: Numeric data cleaning function

This cleaning function addresses several critical data quality issues:

- Converts string representations of numbers to actual numeric values
- Handles scientific notation in the input data
- Rounds monetary values to two decimal places for consistency
- Provides a fallback value (0) for invalid numeric data

The main data loading function implements a comprehensive approach to handling the data:

```
def load_insurance_data(file_path):
    # Read CSV file
    df = pd.read_csv(file_path)

    # Clean numeric columns
    numeric_columns = ['INSURED_VALUE', 'PREMIUM', 'OBJECT_ID',
                       'CCM_TON', 'CARRYING_CAPACITY']
    for col in numeric_columns:
        if col in df.columns:
            df[col] = df[col].apply(clean_numeric)

    # Convert date columns
    df['INSR_BEGIN'] = pd.to_datetime(df['INSR_BEGIN'],
                                      format='%d-%b-%y')
    df['INSR_END'] = pd.to_datetime(df['INSR_END'],
                                    format='%d-%b-%y')
```

Listing 5: Main data loading function excerpt

The data loading process incorporates several sophisticated features:

- Batch processing with progress tracking (updates every 1000 rows)
- Comprehensive error handling for each row insertion
- Data type validation and conversion for each field
- Automatic truncation of string fields to match database constraints
- Null value handling for optional fields

The script processes multiple CSV files sequentially, maintaining detailed logging of the process:

```
if __name__ == "__main__":
    file_paths = [
        r'C:\Users\georg\Desktop\Collingwood Project\Data\motor_data11
    -14lats.csv',
        r'C:\Users\georg\Desktop\Collingwood Project\Data\motor_data14
    -2018lats.csv'
    ]

```

```
    for file_path in file_paths:
        try:
            print(f"\nProcessing file: {file_path}")
            load_insurance_data(file_path)
        except Exception as e:
            print(f"Error processing file {file_path}: {e}")
```

Listing 6: Main execution block

The implementation includes robust error handling at multiple levels:

- File-level error catching for missing or corrupted files
- Row-level error handling to prevent single-row failures from stopping the entire process
- Data type conversion error handling with appropriate fallback values
- Database connection error management

This carefully designed data loading process ensures data integrity while providing detailed feedback about the loading progress and any issues encountered during the process. The script's modular design also allows for easy modifications and improvements as needed.

3 Data Validation and Quality Checks

After loading the data, it was crucial to verify the integrity and quality of our imported dataset. A series of SQL queries were implemented to perform comprehensive data validation checks. These queries helped ensure the data was loaded correctly and identified any potential quality issues that might need addressing.

3.1 Basic Data Verification

First, basic verification of the data import were performed by checking the total number of records and examining the most recently added entries:

```sql
-- Get total count and sample of latest records
SELECT COUNT(*) as total_records
FROM vehicle_insurance;

SELECT TOP 10 *
FROM vehicle_insurance
ORDER BY id DESC;
```

Listing 7: Basic record count and latest entries

This initial check made it possible to:

- Confirm the successful import of records
- Verify the most recent entries for completeness
- Ensure the auto-incrementing ID was functioning properly

3.2 Summary Statistics

To gain a broader understanding of the dataset,summary statistics were generated covering various key metrics:

```
SELECT
    COUNT(*) as total_records,
    COUNT(DISTINCT make) as unique_makes,
    AVG(premium) as avg_premium,
    MAX(insr_begin) as latest_start_date,
    MIN(insr_begin) as earliest_start_date
FROM vehicle_insurance;
```

Listing 8: Summary statistics query

This query provided crucial insights into:

- The total volume of records in the database
- The diversity of vehicle makes in our dataset
- Average premium values for business analysis
- The temporal range of our insurance data

3.3 Data Quality Assessment

To identify potential data quality issues, a comprehensive check was carried out for null values across critical fields:

```
SELECT
    COUNT(*) as null_count,
    'premium' as field_name
FROM vehicle_insurance
WHERE premium IS NULL
UNION
SELECT COUNT(*), 'make'
FROM vehicle_insurance
WHERE make IS NULL
UNION
SELECT COUNT(*), 'insr_begin'
FROM vehicle_insurance
WHERE insr_begin IS NULL
UNION
SELECT COUNT(*), 'object_id'
FROM vehicle_insurance
WHERE object_id IS NULL;
```

Listing 9: Null value check across critical fields

This query helps identify:

- Missing values in critical fields
- Potential data entry or import issues
- Fields that might require additional cleaning or processing
- Areas where data collection might need improvement

These validation checks form an essential part of the data quality assurance process, ensuring the reliability and completeness of the insurance dataset for subsequent analysis and reporting.

3.4 Data Quality Findings and Solutions

The null value analysis returned the following results:

Field Name	Null Count
premium	21
make	5
insr_begin	0
object_id	0

Table 1: Null Value Analysis Results

These results reveal several data quality issues that require attention:

3.4.1 Premium Field Issues

The presence of 21 null values in the premium field is particularly concerning as this is a critical financial metric. To address this:

```sql
-- First, examine the records with null premiums
SELECT *
FROM vehicle_insurance
WHERE premium IS NULL;

-- Update null premiums with average premium for similar vehicles
UPDATE v1
SET premium = (
    SELECT AVG(premium)
    FROM vehicle_insurance v2
    WHERE v2.type_vehicle = v1.type_vehicle
    AND premium IS NOT NULL
)
FROM vehicle_insurance v1
WHERE premium IS NULL;
```

Listing 10: Identifying records with null premiums

3.4.2 Make Field Issues

The 5 missing vehicle makes represent a smaller but still significant data quality issue. To resolve this:

```sql
-- Identify records with missing makes
SELECT *
FROM vehicle_insurance
WHERE make IS NULL;

-- Update null makes with 'Unknown'
UPDATE vehicle_insurance
SET make = 'Unknown'
```

```sql
WHERE make IS NULL;
```

Listing 11: Handling missing make values

The absence of null values in insr_begin and object_id fields confirms that the primary temporal and identification data is complete, which is crucial for maintaining data integrity and traceability.

3.4.3 Resolution Strategy

To address these data quality issues, the following solution strategy was implemented:

1. **Premium Nulls:**
   - Impute missing values using average premiums for similar vehicle types
   - Flag imputed records for future analysis
   - Implement validation checks in the data loading process to prevent future null premiums
2. **Missing Makes:**
   - Mark records with 'Unknown' make for easy identification
   - Create a process to review and update these records if information becomes available
   - Add data validation rules to the input process to reduce future missing makes

After implementing these fixes, the null check query was ran again to verify the corrections:

```sql
-- Verify fixes
SELECT
    COUNT(*) as null_count,
    'premium' as field_name
FROM vehicle_insurance
WHERE premium IS NULL
UNION
SELECT COUNT(*), 'make'
FROM vehicle_insurance
WHERE make IS NULL;
```

Listing 12: Verification of fixes

This comprehensive approach to handling missing data ensures the integrity of the dataset while maintaining transparency about any modifications made to the original data.

4 Macro-Level Data Analysis

To gain comprehensive insights into the insurance dataset,a systematic approach was implemented using various analytical views and tables. This macro-level analysis serves as a foundation for identifying key areas that warrant deeper investigation.

4.1 Primary Analytical Tables

To start the analysis, two fundamental analytical tables were created designed to provide structured views of the data across different dimensions. The first table, vehicle_analytics, consolidates metrics specific to vehicle categories and their performance characteristics. Through aggregation of vehicle-specific data, this table enables detailed examination of claims patterns, premium efficiency, and risk distribution across the portfolio's various segments.

```sql
-- Create vehicle analytics table
CREATE TABLE vehicle_analytics (
    make VARCHAR(50),
    type_vehicle VARCHAR(50),
    total_vehicles INT,
    avg_premium DECIMAL(18,2),
    avg_insured_value DECIMAL(18,2),
    total_claims DECIMAL(18,2),
    number_of_claims INT,
    avg_claim_amount DECIMAL(18,2)
);

INSERT INTO vehicle_analytics
SELECT
    make, type_vehicle,
    COUNT(*) as total_vehicles,
    AVG(premium) as avg_premium,
    AVG(insured_value) as avg_insured_value,
    SUM(claim_paid) as total_claims,
    COUNT(CASE WHEN claim_paid > 0 THEN 1 END) as number_of_claims,
    AVG(CASE WHEN claim_paid > 0 THEN claim_paid END) as
   avg_claim_amount
FROM vehicle_insurance
GROUP BY make, type_vehicle;
```

Listing 13: Vehicle Analytics Table Creation

The vehicle analytics table serves as a foundation for the risk assessment process. By aggregating data at the make and type level,the relative performance of different vehicle segments can be evaluated and potential areas of premium misalignment can be identified. The inclusion of both frequency and severity metrics provides a comprehensive view of risk patterns, while the premium and insured value fields enable assessment of pricing adequacy. This able will make it much easier to identify issues in pricing.

4.2 Multi-Dimensional Analysis Framework

The analytical framework encompasses five key dimensions that together provide a comprehensive view of the portfolio's performance and risk characteristics.

4.2.1 Vehicle Performance Analysis

The vehicle performance dimension examines the fundamental metrics that define each vehicle category's contribution to the portfolio. Through analysis of policy volumes, average premiums, loss ratios, and policy duration patterns,segments that may be under or overpriced relative to their risk profile can be identified.

4.2.2 Claims Pattern Analysis

The examination of claims patterns provides crucial insights into risk manifestation across different vehicle categories. By analyzing claim frequency by make, assessing average claim amounts, and identifying extreme claim events, a thorough understanding of how vehicle characteristics correlate with claim likelihood is developed. This analysis reveals particular make and model combinations that warrant special attention in the risk management framework.

4.2.3 Risk Assessment

To quantify and standardize the risk evaluation process, a risk scoring system was implemented using the following structure:

```sql
-- Create risk scores table
CREATE TABLE risk_scores (
    make VARCHAR(50),
    type_vehicle VARCHAR(50),
    avg_claim DECIMAL(18,2),
    claim_frequency DECIMAL(18,4),
    risk_score INT
);

INSERT INTO risk_scores
WITH claim_stats AS (
    SELECT make, type_vehicle,
           AVG(claim_paid) as avg_claim,
           COUNT(CASE WHEN claim_paid > 0 THEN 1 END) /
           CAST(COUNT(*) AS FLOAT) as claim_frequency
    FROM vehicle_insurance
    GROUP BY make, type_vehicle
)
SELECT make, type_vehicle, avg_claim, claim_frequency,
       NTILE(5) OVER (ORDER BY avg_claim * claim_frequency DESC)
       as risk_score
FROM claim_stats;
```

Listing 14: Risk Scoring Table Creation

4.2.4 Temporal Analysis

The temporal dimension of the analysis reveals crucial seasonal patterns in portfolio performance. Monthly policy acquisition trends were tracked alongside premium variations to understand market dynamics. This analysis also encompasses claim frequency patterns and revenue fluctuations, providing insights into cyclical risk factors that affect the portfolio. Understanding these temporal patterns enables more effective resource allocation and risk management throughout the year.

```sql
-- Create time-based analytics table
CREATE TABLE time_based_analytics (
    insurance_year INT,
    insurance_month INT,
    new_policies INT,
    total_premium_revenue DECIMAL(18,2),
    total_claims DECIMAL(18,2),
    avg_premium DECIMAL(18,2),
    loss_ratio DECIMAL(18,2)
);

INSERT INTO time_based_analytics
SELECT
    YEAR(insr_begin) as insurance_year,
    MONTH(insr_begin) as insurance_month,
    COUNT(*) as new_policies,
    SUM(premium) as total_premium_revenue,
    SUM(claim_paid) as total_claims,
    AVG(premium) as avg_premium,
    SUM(claim_paid)/NULLIF(SUM(premium), 0) * 100 as loss_ratio
FROM vehicle_insurance
GROUP BY YEAR(insr_begin), MONTH(insr_begin);
```

Listing 15: Time-Based Analytics Table Creation

4.3 Areas Identified for Deep-Dive Analysis

The macro analysis revealed several critical areas requiring detailed investigation. The findings can be categorized into three primary domains:

4.3.1 Vehicle Risk Findings

- **Age-Claim Correlation:** Strong relationship identified between vehicle age and increased claim frequency
- **Type Impact:** Vehicle type demonstrates significant influence on claim severity
- **Pricing Gaps:** Multiple opportunities identified for premium optimization across segments

4.3.2 Temporal Patterns

- **Seasonality:** Consistent and predictable patterns in claim frequency across annual cycles
- **Duration Effects:** Policy duration shows strong correlation with claim likelihood
- **Premium Dynamics:** Historical premium trends indicate need for strategic pricing adjustments

4.3.3 Risk Assessment Outcomes

- **High-Risk Combinations:** Several specific vehicle make/model combinations exhibit elevated risk profiles

- **Performance Issues:** Multiple segments identified operating below acceptable performance thresholds
- **Premium Adequacy:** Systematic premium inadequacy detected across several risk categories

This macro-level analysis establishes the foundation for targeted investigations into these areas of concern, enabling data-driven decisions in risk management and pricing strategies.

5 Analysis of Key Findings

The comprehensive analysis of the insurance portfolio reveals several significant patterns and areas of interest across vehicle types, risk profiles, and temporal trends. Let's examine the key aspects of the findings.

5.1 Portfolio Composition and Performance

The analysis of vehicle analytics reveals Toyota Pick-ups as the dominant vehicle type in the portfolio, with 98,600 vehicles insured. This segment demonstrates a relatively balanced risk-to-premium ratio, with an average premium of 6,408.47 and claim amounts averaging 166,176.92 per incident. Here are the top segments by volume:

Make	Type	Total Vehicles	Avg Premium	Avg Claim
TOYOTA	Pick-up	98,600	6,408.47	166,176.92
BAJAJI	Motor-cycle	67,342	483.56	29,323.69
TOYOTA	Automobile	66,955	3,289.88	91,788.56
TOYOTA	Bus	58,992	8,214.38	225,021.20

Table 2: Top Vehicle Segments by Volume

In terms of premium structure, the analysis shows significant variation across vehicle types. Iveco Trucks command the highest average premium at 23,808.35, reflecting their higher risk profile and potential for significant claims. This premium differentiation appears justified given the claims data, with Iveco Trucks showing an average claim amount of 966,409.35.

5.2 Risk Profile Analysis

The risk scoring system has identified several concerning patterns in the heavy vehicle segment. Sky Bus operations present particularly noteworthy risk characteristics, with a claim frequency of 40.63% and an average claim amount exceeding 11.9 million. This segment requires immediate attention for risk management strategies.

The data shows a clear correlation between vehicle size and claim frequency, with heavy vehicles dominating the highest risk category. Of particular concern is the Mercedes trailers segment, which shows a 100% claim frequency rate. The top risk-scored segments are:

Make	Type	Avg Claim	Claim Freq.
SKY BUS	Bus	11,940,404.47	0.4063
IVECO/CHINA	Bus	7,747,532.74	0.0652
TRACTOR	Special construction	6,984,260.37	0.0185

Table 3: Highest Risk Vehicle Segments

5.3 Temporal Analysis and Portfolio Evolution

Examination of the time-based analytics reveals significant portfolio development from 2011 to 2018. The data shows consistent growth in policy volumes and premium revenue, accompanied by improving loss ratios over time. July consistently emerges as a peak period for new policy acquisition, with the highest policy volumes and premium revenue across all years.

A notable trend in the loss ratio indicates improving risk management practices:

- 2014 showed concerning loss ratios, peaking at 955.69% in January
- By 2017-2018, loss ratios showed marked improvement, generally staying below 250%
- The latest data from 2018 shows loss ratios below 100%, indicating significant improvement in risk selection and pricing

5.4 Areas for Further Investigation

Based on the macro-level analysis, several areas warrant deeper investigation:

1. The high claim frequency in the bus segment, particularly among Sky Bus and Iveco/China manufacturers, suggests a need for detailed analysis of operating conditions, routes, and driver patterns.

2. The correlation between vehicle size and claim frequency requires further study to determine if current premium structures adequately reflect the risk exposure of larger vehicles.

3. The seasonal variation in policy acquisition and claims, particularly the July peak, needs investigation to ensure appropriate resource allocation and risk assessment during high-volume periods.

4. The dramatic improvement in loss ratios from 2014 to 2018 should be studied to identify and potentially replicate successful risk management strategies.

These findings suggest that while overall portfolio performance has improved, specific segments still require attention for optimized risk management and pricing strategies.

6 Comprehensive Portfolio Analysis

6.1 Analytical Framework and Data Structure

To gain comprehensive insights into the insurance portfolio, we implemented a systematic analytical framework using specialized data structures. Two fundamental analytical tables form the foundation of the investigation:

```sql
CREATE TABLE vehicle_analytics (
    make VARCHAR(50),
    type_vehicle VARCHAR(50),
    total_vehicles INT,
    avg_premium DECIMAL(18,2),
    avg_insured_value DECIMAL(18,2),
    total_claims DECIMAL(18,2),
    number_of_claims INT,
    avg_claim_amount DECIMAL(18,2)
);

INSERT INTO vehicle_analytics
SELECT
    make, type_vehicle,
    COUNT(*) as total_vehicles,
    AVG(premium) as avg_premium,
    AVG(insured_value) as avg_insured_value,
    SUM(claim_paid) as total_claims,
    COUNT(CASE WHEN claim_paid > 0 THEN 1 END) as number_of_claims,
    AVG(CASE WHEN claim_paid > 0 THEN claim_paid END) as
   avg_claim_amount
FROM vehicle_insurance
GROUP BY make, type_vehicle;
```

Listing 16: Vehicle Analytics Table Creation

6.2 Temporal Analysis and Seasonal Patterns

The temporal analysis framework captures both seasonal patterns and long-term trends:

```sql
CREATE TABLE time_based_analytics (
    insurance_year INT,
    insurance_month INT,
    new_policies INT,
    total_premium_revenue DECIMAL(18,2),
    total_claims DECIMAL(18,2),
    avg_premium DECIMAL(18,2),
    loss_ratio DECIMAL(18,2)
);

-- Analyze seasonal patterns
SELECT
    insurance_month,
    AVG(new_policies) as avg_policies,
    AVG(total_premium_revenue) as avg_revenue,
    AVG(loss_ratio) as avg_loss_ratio
FROM time_based_analytics
GROUP BY insurance_month
```

```sql
ORDER BY insurance_month;
```

Listing 17: Time-Based Analytics and Seasonal Analysis

The analysis reveals consistent seasonal patterns, with July emerging as the peak period for new policy acquisition (25,000-33,000 new policies). This peak coincides with maximum premium revenue collection, though loss ratios show varying patterns year over year.

6.3 Risk Assessment Framework

To quantify risk across different vehicle segments,a comprehensive scoring system was implemented:

```sql
CREATE TABLE risk_scores (
    make VARCHAR(50),
    type_vehicle VARCHAR(50),
    avg_claim DECIMAL(18,2),
    claim_frequency DECIMAL(18,4),
    risk_score INT
);

INSERT INTO risk_scores
WITH claim_stats AS (
    SELECT make, type_vehicle,
           AVG(claim_paid) as avg_claim,
           COUNT(CASE WHEN claim_paid > 0 THEN 1 END) /
           CAST(COUNT(*) AS FLOAT) as claim_frequency
    FROM vehicle_insurance
    GROUP BY make, type_vehicle
)
SELECT make, type_vehicle, avg_claim, claim_frequency,
      NTILE(5) OVER (ORDER BY avg_claim * claim_frequency DESC)
      as risk_score
FROM claim_stats;
```

Listing 18: Risk Scoring Implementation

6.4 Key Findings

The analysis revealed several critical insights across multiple dimensions:

6.4.1 Vehicle Segment Performance

- Toyota Pick-ups dominate the portfolio (98,600 vehicles) with balanced risk-to-premium ratios
- Iveco Trucks command highest premiums (23,808) with corresponding high risk profiles
- Mercedes trailers show concerning 100% claim frequency

6.4.2 Risk Patterns

- Heavy vehicles, particularly buses and trailers, consistently appear in highest risk categories
- Sky Bus operations show 40.63% claim frequency with substantial average claims
- Significant variation in claims-per-vehicle ratios across vehicle types

6.5 Strategic Implications

Based on the comprehensive analysis, I recommend the following strategic initiatives:

1. **Resource Allocation:** Enhance operational capacity for July peak period
2. **Risk Management:** Implement specialized procedures for heavy vehicles
3. **Premium Optimization:** Adjust premium structures based on claims-per-vehicle analysis
4. **Portfolio Balance:** Carefully manage exposure across risk categories while maintaining growth

This analysis establishes the foundation for targeted interventions in specific segments while supporting overall portfolio optimization. Regular monitoring of these metrics will enable continuous refinement of the risk management and pricing strategies.

7 Portfolio Evolution Analysis: 2011-2018

A comprehensive analysis of the portfolio's development from 2011 to 2017 reveals remarkable growth coupled with significant operational improvements. The data demonstrates a transformation from a moderate-sized operation to a substantial insurance portfolio with enhanced risk management capabilities.

7.1 Policy Volume and Revenue Growth

The portfolio experienced consistent and substantial growth in policy volumes from 2011 to 2017. Starting with 71,047 policies in 2011, the portfolio expanded to 138,667 policies by 2017, representing nearly a doubling of the customer base with a 95% increase over this period. The most notable expansion occurred between 2011 and 2012, with a significant 32% year-over-year increase in policy volume.

The revenue growth tells an even more impressive story. Annual premium revenue grew from 376 million in 2011 to 1.16 billion in 2017, representing a substantial 208% increase. Notably, this revenue growth significantly outpaced the growth in policy numbers, indicating a strategic shift in the portfolio composition. The average revenue per policy increased from 5,299 in 2011 to 8,371 in 2017, suggesting either a shift toward higher-value policies or successful premium optimization strategies.

7.2 Risk Management and Loss Ratio Improvements

Perhaps the most significant achievement during this period was the dramatic improvement in loss ratio performance. The portfolio showed a consistent downward trend in loss ratios, improving from 365% in 2011 to 192% in 2017. This improvement was particularly pronounced during the 2015-2017 period, indicating the successful implementation of enhanced risk assessment procedures and more effective pricing strategies.

7.3 Claims Management Evolution

The claims management data reveals a particularly interesting trend. Despite the portfolio doubling in size over the period, total claims actually peaked in 2014 at 2.98 billion and subsequently declined. By 2017, even with the highest number of policies in force, total claims had reduced to 1.97 billion. This pattern strongly suggests the implementation of more sophisticated risk assessment and claims management processes during this period.

Year	Policies	Revenue (M)	Loss Ratio	Claims (B)
2011	71,047	376.47	365.19%	1.56
2014	107,719	795.32	396.05%	2.98
2017	138,667	1,160.82	191.85%	1.97

Table 4: Key Performance Metrics: Critical Years

7.4 2018 Partial Year Analysis

While the 2018 data appears incomplete, showing only 57,265 policies and revenue of 383 million, it continues to demonstrate the positive trends established in previous years, particularly in loss ratio improvement. The average loss ratio of 88% for the available 2018 data suggests continued enhancement of risk management practices.

7.5 Strategic Implications

This analysis reveals several key strategic developments in the portfolio's evolution:

First, the successful expansion in both policy volume and revenue indicates strong market penetration and effective growth strategies. The fact that revenue growth outpaced policy growth suggests successful premium optimization and possibly a strategic shift toward higher-value segments.

Second, the dramatic improvement in loss ratios, particularly from 2015 onward, indicates the successful implementation of enhanced risk management practices. This improvement occurred while the portfolio was growing, making it even more noteworthy.

Finally, the reduction in total claims despite portfolio growth demonstrates the effectiveness of improved underwriting practices and risk assessment procedures. This suggests the development of more sophisticated approaches to risk selection and pricing during the period.

These trends collectively paint a picture of a portfolio that has not only grown substantially but has also matured in its operational sophistication, particularly in risk management and underwriting practices.

8 Vehicle Type Analysis and Risk Assessment

The analysis of vehicle type performance reveals significant variations in fleet composition, risk profiles, and premium adequacy across different vehicle categories.First, an comprehensive examination of data across all vehicle types will be performed:

Vehicle Type	**Total Count**	**Typical Premium**	**Total Claims (Millions)**	**Claims/ Vehicle**
Truck	152,639	12,142.54	7,901.60	51,766.57
Pick-up	146,969	3,635.26	2,376.04	16,166.92
Motor-cycle	143,129	516.55	27.43	191.66
Automobile	125,791	3,648.45	823.76	6,548.67
Bus	107,644	12,102.93	3,165.71	29,409.05
Station Wagones	60,800	7,807.30	1,046.85	17,217.86
Trailers and semitrailers	36,221	6,277.29	1,233.01	34,041.24
Special construction	12,132	17,418.04	182.50	15,042.48
Tractor	11,434	8,623.74	26.88	2,350.96
Tanker	10,740	20,013.64	556.44	51,809.68

Table 5: Vehicle Type Performance Analysis

8.1 Portfolio Composition and Risk Distribution

The portfolio demonstrates considerable diversity in both vehicle types and associated risk profiles. Trucks form the largest segment with 152,639 vehicles, generating the highest total claims at 7.9 billion. However, this substantial claims figure must be viewed in the context of the segment's size and operational characteristics.

Motorcycles present an intriguing contrast, representing the third-largest segment with 143,129 vehicles but demonstrating remarkably low risk with only 191 claims per vehicle. This suggests either highly effective risk management practices within this segment or inherently lower risk exposure.

The tanker segment, while comprising only 10,740 vehicles, emerges as one of the highest-risk categories with 51,809 claims per vehicle. This mirrors the risk level seen in trucks (51,766 claims per vehicle), suggesting similar risk factors despite different vehicle purposes.

8.2 Risk Stratification Analysis

The analysis reveals three distinct risk tiers among vehicle types:

8.2.1 High-Risk Tier (> 30,000 claims/vehicle)

Tankers, trucks, and trailers/semitrailers dominate this category, with claims per vehicle ranging from 34,041 to 51,809. These vehicles typically involve commercial operations with high-value cargo and significant exposure to road risks.

8.2.2 Medium-Risk Tier (10,000-30,000 claims/vehicle)

This category includes buses (29,409 claims/vehicle), station wagons (17,217), pick-ups (16,166), and special construction vehicles (15,042). These vehicles typically serve mixed commercial and personal use purposes.

8.2.3 Low-Risk Tier (< 10,000 claims/vehicle)

Automobiles (6,548 claims/vehicle), tractors (2,350), and motorcycles (191) comprise this category, showing significantly lower risk profiles despite varying usage patterns.

8.3 Premium Structure Analysis

Examination of premium structures reveals several noteworthy disparities between risk exposure and pricing:

Tankers command the highest premium at 20,013, appropriately reflecting their high-risk profile with 51,809 claims per vehicle. However, trucks, with nearly identical claims per vehicle (51,766), are priced significantly lower at 12,142, suggesting potential premium inadequacy.

Special construction vehicles present an inverse scenario, with high premiums (17,418) relative to their claims per vehicle (15,042), potentially indicating over-pricing in this segment.

The trailer segment shows concerning premium inadequacy, with moderate premiums (6,277) despite high claims per vehicle (34,041), suggesting immediate need for pricing review.

8.4 Strategic Recommendations

Based on this comprehensive analysis, several strategic initiatives warrant consideration.

8.4.1 Premium Structure Optimization

The trailer segment requires immediate premium adjustment to better align with its risk profile. Current pricing appears inadequate given the claims experience. Conversely, the special construction vehicle segment may benefit from premium review to ensure market competitiveness while maintaining profitability.

8.4.2 Risk Management Knowledge Transfer

The motorcycle segment's exceptional performance, with minimal claims per vehicle and efficient premium pricing, merits detailed study. Understanding the success factors in this segment could provide valuable insights for risk management across other vehicle categories.

8.4.3 Commercial Vehicle Risk Alignment

The similar risk profiles of tankers and trucks suggest potential benefits from shared risk mitigation strategies. Development of unified risk management protocols for these segments could improve overall portfolio performance.

These findings provide a foundation for strategic portfolio adjustments aimed at optimizing risk-adjusted returns across vehicle types while maintaining appropriate premium levels for each risk category.

9 Make and Type Risk Analysis

A detailed examination of vehicle make and type combinations reveals complex risk patterns that merit careful consideration for underwriting and pricing strategies. The analysis identifies several critical risk patterns and premium alignment issues that warrant immediate attention.

9.1 High-Risk Vehicle Combinations

The analysis reveals particularly concerning risk profiles among certain vehicle combinations. Most notably, SKY BUS emerges as the highest-risk manufacturer in the bus category, with an average claim amount of 11.94 million and a claim frequency of 40.63%. This represents an exceptionally high risk level, especially considering that these vehicles have a premium of only 68,085, suggesting significant underpricing relative to risk exposure.

Similarly, IVECO/CHINA buses demonstrate concerning risk patterns, with average claims of 7.75 million, though their claim frequency is lower at 6.52%. While this manufacturer presents a lower frequency of claims, the severity of these claims suggests a need for careful risk management.

9.2 Large Fleet Performance

Analysis of large fleets (those exceeding 10,000 vehicles) provides valuable insights into sustained risk patterns. Toyota Pick-ups, the largest fleet with 98,600 vehicles, demonstrates relatively well-managed risk with an average claim of 166,102 and a claim frequency of 11.22%. The premium of 6,408 appears reasonably aligned with this risk profile.

Isuzu trucks, with 55,160 vehicles, show higher risk characteristics with an average claim of 431,127 and a 9.51% claim frequency. Their premium of 14,044 reflects this higher risk profile, suggesting appropriate risk assessment in this segment.

9.3 Manufacturer Risk Profiles

The data reveals significant variations in risk profiles across manufacturers. Toyota and Nissan consistently demonstrate lower risk profiles across multiple vehicle types, suggesting superior build quality or potentially better driver profiles. This pattern holds true particularly in their consumer vehicle segments.

The analysis also confirms the earlier findings regarding motorcycles, which show notably lower risk profiles across manufacturers. This consistency across brands suggests that the vehicle type, rather than manufacturer, may be the primary risk determinant in this category.

9.4 Premium Efficiency Patterns

Examination of premium efficiency reveals several notable examples of effective risk pricing. Iveco trucks demonstrate particularly efficient premium structuring, with a 23,808 premium against an average claim of 966,409, representing a sustainable risk-premium ratio. Similarly, Sino Howo trucks show effective premium alignment with a 19,352 premium supporting an average claim of 678,643.

9.5 Strategic Recommendations

Based on this comprehensive analysis, several strategic initiatives warrant immediate consideration:

First, an urgent review of the SKY BUS segment is essential. The combination of high claim frequency and severity, coupled with current premium levels, suggests an unsustainable risk exposure that requires immediate attention.

Second, the development of manufacturer-specific risk protocols appears justified. The significant variations in risk profiles across manufacturers, even within the same vehicle type, suggest that manufacturer-specific underwriting guidelines could improve risk management.

Finally, the success patterns observed in Toyota and Nissan fleets should be studied for potential insights that could be applied to higher-risk segments. Understanding the factors contributing to their superior risk profiles could inform broader risk management strategies across the portfolio.

10 Risk Analysis Methodology

The enhanced analytical framework employs multiple interconnected analyses to develop a comprehensive understanding of risk patterns and pricing optimization opportunities. This multi-dimensional approach examines seasonal variations, premium adequacy, temporal trends, age-based risk factors, and optimization potential through carefully structured queries.

10.1 Seasonality and Risk Correlation

The seasonal risk analysis explores the relationship between time of year and risk patterns across vehicle types:

```
WITH SeasonalRisk AS (
    SELECT
        v.type_vehicle,
        MONTH(v.insr_begin) as month_number,
        COUNT(*) as policies,
        AVG(v.premium) as avg_premium,
        SUM(v.claim_paid) as total_claims,
        SUM(v.claim_paid)/NULLIF(SUM(v.premium), 0) * 100 as loss_ratio
    FROM vehicle_insurance v
    GROUP BY v.type_vehicle, MONTH(v.insr_begin)
)
SELECT *
FROM SeasonalRisk
WHERE policies > 100
ORDER BY loss_ratio DESC;
```

Listing 19: Seasonal Risk Analysis Query

This analysis reveals critical seasonal patterns in risk exposure, particularly focusing on statistically significant samples with more than 100 policies. By examining loss ratios across months, periods of heightened risk for specific vehicle types can be identified, enabling more nuanced pricing strategies.

10.2 Premium Adequacy Assessment

To evaluate premium adequacy across different makes and models, a sophisticated risk metrics analysis was developed:

```
WITH RiskMetrics AS (
    SELECT
        make,
        type_vehicle,
        COUNT(*) as total_policies,
        AVG(premium) as avg_premium,
        SUM(claim_paid) as total_claims,
        COUNT(CASE WHEN claim_paid > 0 THEN 1 END) as claims_count,
        SUM(claim_paid)/NULLIF(SUM(premium), 0) * 100 as loss_ratio,
        AVG(CASE WHEN claim_paid > 0 THEN claim_paid END) as
   avg_claim_amount
    FROM vehicle_insurance
    GROUP BY make, type_vehicle
    HAVING COUNT(*) > 50
)
```

Listing 20: Premium vs Claims Analysis

This analysis introduces a structured categorization of pricing adequacy:

- Severe Underpricing: Loss ratios exceeding 300%
- Significant Underpricing: Loss ratios between 200% and 300%
- Moderate Underpricing: Loss ratios between 100% and 200%

- Balanced Pricing: Loss ratios between 50% and 100%
- Profitable Pricing: Loss ratios below 50%

10.3 Temporal Trend Analysis

The year-over-year trend analysis provides crucial insights into the evolution of risk patterns:

```
WITH YearlyTrends AS (
    SELECT
        type_vehicle,
        YEAR(insr_begin) as insurance_year,
        COUNT(*) as policies,
        AVG(premium) as avg_premium,
        SUM(claim_paid) as total_claims,
        SUM(claim_paid)/NULLIF(SUM(premium), 0) * 100 as loss_ratio
    FROM vehicle_insurance
    GROUP BY type_vehicle, YEAR(insr_begin)
)
```

Listing 21: Year-over-Year Trend Analysis

This analysis incorporates year-over-year comparisons using LAG functions to track changes in loss ratios, enabling the identification of emerging risk patterns and the evaluation of pricing strategy effectiveness.

10.4 Vehicle Age Risk Assessment

The relationship between vehicle age and risk profiles is examined through detailed cohort analysis:

```
SELECT
    type_vehicle,
    YEAR(insr_begin) - prod_year as vehicle_age,
    COUNT(*) as vehicle_count,
    AVG(premium) as avg_premium,
    SUM(claim_paid) as total_claims,
    COUNT(CASE WHEN claim_paid > 0 THEN 1 END) as claims_count,
    SUM(claim_paid)/NULLIF(SUM(premium), 0) * 100 as loss_ratio
FROM vehicle_insurance
WHERE prod_year > 1950
GROUP BY type_vehicle, YEAR(insr_begin) - prod_year
HAVING COUNT(*) > 20
```

Listing 22: Age-Based Risk Analysis

10.5 Premium Optimization Framework

The culmination of the analysis is a comprehensive premium optimization framework:

```
WITH PremiumAnalysis AS (
    SELECT
        make,
        type_vehicle,
        AVG(premium) as current_premium,
        AVG(claim_paid) as avg_claim,
```

```sql
        COUNT(*) as policy_count,
        SUM(claim_paid)/NULLIF(SUM(premium), 0) * 100 as loss_ratio,
        COUNT(CASE WHEN claim_paid > 0 THEN 1 END) * 100.0 / COUNT(*)
    as claim_frequency
    FROM vehicle_insurance
    GROUP BY make, type_vehicle
    HAVING COUNT(*) > 50
)
```

Listing 23: Premium Optimization Analysis

This framework provides actionable recommendations through a tiered approach:

- Immediate Review: Loss ratios exceeding 200%
- High Priority Review: Loss ratios between 150% and 200%
- Review Needed: Loss ratios between 100% and 150%
- Acceptable: Loss ratios below 100%

The suggested premium adjustments are calculated using a sliding scale based on loss ratio severity, with more aggressive adjustments recommended for higher loss ratios.

10.6 Low-Risk Opportunities

The analysis also identifies segments with consistently favorable risk profiles that present opportunities for market expansion:

10.6.1 Motorcycle Segment

Motorcycles demonstrate remarkably stable and profitable performance:

- Loss ratios consistently range between 20% and 53%
- July shows highest loss ratio but remains within profitable range
- Risk exposure shows minimal seasonal variation

10.6.2 Tractor Segment

Tractors show particularly strong performance in later months:

- Loss ratios decrease significantly in latter half of year
- December shows exceptional performance with 0.84% loss ratio
- Opportunity for competitive pricing during low-risk periods

10.7 Strategic Recommendations

Based on this comprehensive analysis, I recommend implementing a multi-faceted approach to address seasonal risk patterns:

10.7.1 Premium Adjustment Strategy

Implement a dynamic premium structure that reflects seasonal risk variations:

- Tankers: Establish three-tier pricing system with significant premium increases during May
- Trailers: Implement winter season premium adjustments
- Trucks: Introduce winter surcharge program

10.7.2 Risk Management Protocols

Develop specialized risk management protocols for high-risk periods:

- Enhanced underwriting guidelines during peak risk months
- Specialized risk assessment procedures for seasonal operations
- Modified deductible structures during high-risk periods

10.7.3 Market Opportunity Development

Capitalize on identified low-risk segments:

- Develop competitive pricing strategies for motorcycle segment
- Implement seasonal incentives for tractor insurance during low-risk months
- Create targeted marketing campaigns for well-performing segments

This analysis provides a foundation for implementing data-driven premium adjustments and risk management strategies that reflect the complex seasonal patterns observed across different vehicle types.

11 Seasonal Risk Pattern Analysis

The examination of seasonal risk patterns reveals significant variations in loss ratios across different vehicle types and months, suggesting opportunities for strategic premium adjustments and risk management interventions. The analysis uncovers patterns that require immediate attention and presents opportunities for portfolio optimization through targeted interventions.

11.1 High-Risk Seasonal Patterns

The analysis reveals several critical risk periods that demand immediate attention. Most notably, the tanker segment during May exhibits an unprecedented loss ratio of 1,608%, indicating a severe misalignment between premium pricing and actual risk exposure. The current premium of 22,233 falls significantly short of what risk patterns suggest is necessary, pointing to the need for a substantial adjustment to approximately 66,700 to achieve risk-premium equilibrium.

In the trailers and semi-trailers category, concerning performance patterns emerge during the late autumn period. November demonstrates a particularly high loss ratio of 882%, suggesting that the current premium structure significantly underestimates the risk exposure during this season. The existing premium of approximately 7,000 requires a substantial increase to around 17,500 during these high-risk months to ensure portfolio sustainability.

11.2 Vehicle Type Seasonal Analysis

11.2.1 Tanker Segment

The tanker category exhibits the most volatile seasonal risk pattern in the portfolio. May emerges as the critical risk period with its 1,608% loss ratio, followed by September with a 950% loss ratio. This dramatic variation in premium adequacy across seasons necessitates the implementation of a dynamic seasonal pricing structure that better reflects the fluctuating risk exposure throughout the year.

Based on these patterns, I recommend implementing a three-tier premium structure for tankers:

- Peak Risk (May): Premium adjustment to 66,700
- High Risk (September): Premium adjustment to 43,770
- Standard Risk (Other months): Baseline premium with moderate seasonal adjustments

11.2.2 Trailers and Semi-trailers

The trailers and semi-trailers segment demonstrates consistent high-risk patterns during the winter months, with November-December period showing loss ratios exceeding 800%. This persistent elevation in risk exposure throughout the winter months suggests a fundamental misalignment between current premium structures and seasonal risk variations. The pattern indicates a need for a comprehensive revision of the pricing strategy, particularly focusing on winter month operations.

11.2.3 Truck Segment

Within the truck category, distinct seasonal patterns emerge that warrant attention. January presents the highest risk with a 512% loss ratio, and winter months consistently show elevated risk levels. The current premium structure, ranging from 11,000 to 15,000, requires substantial adjustment during these peak risk periods, with recommended increases to 22,000-30,000 during the winter months to adequately cover the increased risk exposure.

11.3 Low-Risk Opportunities

The analysis identifies several segments with consistently favorable risk profiles that present opportunities for market expansion. The motorcycle segment demonstrates remarkably stable and profitable performance throughout the year, with loss ratios consistently ranging between 20% and 53%. Even during July, which shows the highest loss ratio for motorcycles, the segment remains well within profitable ranges. This stability in risk exposure presents an opportunity for competitive pricing strategies.

The tractor segment similarly shows strong performance, particularly in later months of the year. December stands out with an exceptional 0.84% loss ratio, suggesting potential for aggressive market expansion during these low-risk periods. This pattern of decreasing risk exposure in the latter half of the year opens opportunities for seasonal pricing strategies that could drive market share growth while maintaining profitability.

11.4 Strategic Recommendations

The findings support the implementation of a comprehensive strategy to address seasonal risk patterns. The premium adjustment strategy should incorporate a dynamic structure that reflects seasonal risk variations across all vehicle types. For tankers, this means establishing a three-tier pricing system with significant premium increases during May. Trailers require winter season premium adjustments, while trucks need a winter surcharge program to adequately cover increased risk exposure during colder months.

Risk management protocols should be enhanced to address seasonal variations in risk exposure. This includes:

- Development of specialized underwriting guidelines for high-risk months
- Implementation of modified deductible structures during peak risk periods
- Creation of seasonal risk assessment procedures that account for weather-related factors

Market opportunity development should focus on capitalizing on identified low-risk segments through competitive pricing strategies, particularly in the motorcycle and tractor segments during their respective low-risk periods. This approach should be supported by targeted marketing campaigns that highlight the value proposition during these favorable periods.

12 Vehicle Model Performance Analysis

The comprehensive analysis of make and model performance reveals significant variations in risk profiles and identifies critical areas requiring immediate premium adjustments. The findings indicate systematic underpricing across several major vehicle categories, particularly in the heavy commercial vehicle segment.

12.1 Critical Risk Areas

Among high-volume vehicle categories, several combinations demonstrate severe premium inadequacy. The Mitsubishi truck segment, with 2,616 policies in force, shows an alarming loss ratio of 832%, indicating significant underpricing relative to actual risk exposure. The current premium of 11,578 requires a substantial increase to approximately 35,000 to achieve risk-premium equilibacy.

Similarly, the Fiat truck category, comprising 8,500 policies, exhibits a concerning loss ratio of 781%. The existing premium structure of 7,266 needs significant revision, with our analysis suggesting an increase to approximately 28,000 to adequately reflect the risk profile. This adjustment would bring the premium structure more in line with observed claim patterns while maintaining market competitiveness.

Perhaps most significantly, the Toyota bus segment, representing a substantial portfolio of 58,992 policies, demonstrates persistent underpricing with a loss ratio of 312%. Despite its high volume, the current premium of 8,214 fails to adequately cover the risk exposure, necessitating an increase to approximately 26,000 to achieve sustainable performance.

12.2 Premium Adjustment Framework

The analysis supports the implementation of a structured premium adjustment strategy across vehicle categories. In the heavy commercial vehicle segment, several critical adjustments are necessary. The Iveco truck category, despite its relatively high current premium of 23,808, requires an increase to approximately 45,000 to reflect its actual risk profile. This adjustment accounts for both claim frequency and severity patterns observed in the data.

The bus segment requires particularly careful attention. The Sky Bus category, while maintaining a high premium of 68,085, shows severe underpricing relative to its risk profile. The analysis suggests an increase to approximately 145,000 to achieve premium adequacy. Similarly, the Daewoo bus segment requires an adjustment from 21,128 to approximately 42,000 to align with its observed risk patterns.

12.3 Success Models

The analysis also identifies several segments demonstrating excellent performance that can serve as models for risk management. The motorcycle segment, in particular, shows consistently strong results across multiple manufacturers. Yamaha motorcycles maintain a remarkably efficient 39% loss ratio, while Bajaji achieves 34%, and TVS demonstrates exceptional performance with a 24% loss ratio. These results suggest effective risk selection and pricing strategies that could be adapted for other segments.

12.4 Strategic Recommendations

Based on these findings, I recommend implementing a comprehensive risk-based pricing framework that incorporates multiple factors. The foundation of this framework should be vehicle type average loss ratios, modified by make and model-specific multipliers that reflect their individual risk characteristics. This base structure should then be further refined through volume-based adjustments that account for portfolio concentration risk.

For immediate implementation, I recommend prioritizing premium increases for severely underpriced categories, particularly in the heavy commercial vehicle segment. This should be accompanied by a thorough review of underwriting criteria for high-risk vehicle types, with potential implementation of more stringent acceptance standards for the most problematic combinations.

12.5 Looking Forward

The development of a dynamic pricing strategy that incorporates these findings will be crucial for portfolio sustainability. I recommend implementing a monitoring system to track the effectiveness of premium adjustments and allow for rapid response to emerging risk patterns. This approach should include regular reviews of make and model performance metrics to ensure pricing remains aligned with evolving risk profiles.

The success observed in the motorcycle segment provides valuable insights for improving performance across other categories. By studying the factors contributing to their superior loss ratios, a more effective risk selection and pricing strategies for the broader portfolio can be developed.

13 Recommended Future Analysis Priorities

While the current analysis provides substantial insights into portfolio performance and risk patterns, several critical dimensions warrant further investigation to develop a more comprehensive understanding of the risk landscape and enhance the organisations' strategic decision-making capabilities.

13.1 Geographical Risk Distribution

Understanding the spatial distribution of risk represents a crucial next step in the analysis. Different regions often present varying risk profiles due to factors such as road conditions, traffic patterns, and local driving behaviors. A geographical analysis would enable us to identify potential regional premium adjustments and develop location-specific underwriting guidelines. This analysis should examine claims frequency and severity patterns across different regions, potentially revealing opportunities for geographical portfolio optimization.

13.2 Integrated Seasonal Vehicle Analysis

The current analysis has examined seasonal patterns and vehicle types as separate dimensions. However, the interaction between these factors likely holds valuable insights. For instance, certain vehicle types might demonstrate significantly different risk profiles during specific seasons. Understanding these interactions would allow us to develop more sophisticated pricing strategies that account for both temporal and vehicle-specific risk factors. This analysis could prove particularly valuable for commercial vehicles, where seasonal usage patterns might significantly impact risk exposure.

13.3 Driver Demographics and Experience Patterns

The relationship between driver characteristics and risk profiles represents an untapped analytical opportunity. By examining how factors such as driver age, experience, and usage patterns correlate with claims experience, the organization could enhance their risk assessment capabilities substantially. This analysis would be particularly valuable for developing more nuanced underwriting guidelines and could inform targeted marketing strategies for lower-risk demographic segments.

13.4 Policy Tenure Impact Assessment

An examination of how policy duration affects risk profiles could provide valuable insights for retention strategies. Understanding whether longer-term policyholders demonstrate different risk characteristics compared to new policies could inform renewal pricing strategies and help optimize customer retention efforts while maintaining portfolio profitability.

13.5 Claims Severity Distribution Analysis

A more granular analysis of claims severity patterns could enhance the organizations' understanding of risk dynamics across different vehicle categories. By distinguishing between high-frequency, low-severity events and low-frequency, high-severity incidents, they could develop more targeted risk management strategies and optimize their reinsurance structure.

13.6 Economic Correlation Study

External economic factors often influence insurance risk profiles in complex ways. Understanding how variables such as fuel prices, economic growth rates, and vehicle market values correlate with claims patterns could provide valuable context for strategic planning and help anticipate future risk trends.

13.7 Competitive Position Analysis

While maintaining appropriate confidentiality, examining the organizations market position relative to competitors could provide crucial context for their pricing and product strategies. This analysis would help ensure that their risk management improvements maintain market competitiveness while achieving profitability targets.

13.8 Implementation Priority

I recommend prioritizing these analyses in the following order:

First, the geographical risk distribution analysis should be undertaken as it has the most immediate potential impact on pricing and underwriting decisions. This should be followed by the integrated seasonal vehicle analysis, as it builds upon the organizations existing findings and could inform near-term pricing adjustments.

The driver demographics analysis should form the third priority, as its findings could significantly influence their risk selection criteria. The remaining analyses could be conducted subsequently, with their precise ordering determined by resource availability and strategic priorities.

These additional analytical dimensions will provide a more complete understanding of their risk landscape and enable more refined, data-driven decision-making in their ongoing portfolio optimization efforts.