Motor Insurance Portfolio Analysis

Part 1. Data source and data model

A. Dataset

This dashboard project uses real motor insurance transaction records from a Spanish insurance company, spanning November 2015 to December 2018. The dataset contains 105,555 rows and 30 columns, representing annual policy transactions.

B. Data Model

The original data table stored a single row for each annual policy transaction, leading to redundant information as policy, insured, and vehicle details were repeated for every year of coverage. To address this issue, the data was restructured and normalized into a star schema, which is a multidimensional data model (see the Entity-Relationship Diagram in the Appendix). This allows for efficient querying and aggregation, making it ideal for creating interactive dashboards. In this model, the fact table holding details unique to each policy is linked to various dimension tables.

Type	Table
Fact table	• fact_Policy: This table stores policy details, including the first start date and date_lapse, which can be used to identify the policy status (i.e., Effective or Lapse). The Policy ID is the primary key, and this table serves as the central hub of the data model.
Dimension tables	• dim_Record: This table stores core insurance transaction data in each policy year, including policy premium, claim amount, payment method, etc. The Record ID is the primary key. • dim_Insured: This table contains information about insured individuals, including newly added attributes such as age, years of driving, and client scores. The Insured ID is the primary key. • dim_Vehicle: This table stores information about insured vehicles, including power, vehicle type, and value. The Vehicle Age column is derived from the year_matriculation field. The Vehicle ID is the primary key.
Mapping tables (mainly based on glossary, not from data source)	• map_distribution_channel: Maps the distribution channel codes to their descriptions. • map_vehicle_type: Maps vehicle type codes to their descriptions. • map_regular_driver: Maps codes for regular driver status to their descriptions. • map_payment_method: Maps codes for policy payment methods to their descriptions. • map_area: Maps area codes to their descriptions. • map_scoresum_segment: Maps client scores to designed segments. (to be discussed in Part 3C)
Variable-Score-Criteria table (not from data source)	• map_variable_score: This table defines the criteria used to calculate client segmentation scores for specific factors. It solely displays the criteria and is not linked to other tables in the data model.
Date table (not from data source)	• date_table: This table is used for trend analysis and year filtering.
Visual Parameter tables (not from data source)	• Parameter_avg_premium: This table contains parameters used for the average premium graphs. • Parameter_claim: This table contains parameters used for claim analysis graphs.

Part 2. Project Objective

This project aims to evaluate the performance and soundness of the motor insurance business in the insurance company through data visualization and analysis. To ensure a comprehensive analysis, the team decided to include all available transaction data from November 2015 to December 2018. Excluding data from an incomplete year, such as 2015, would have risked overlooking important trends and insights.

<u>**Part 3. Analytical framework and visual selection**</u>

This project utilizes Power BI to develop a three-page dashboard for analyzing the motor insurance portfolio.

A. Overview

The dashboard provides a high-level overview of the motor insurance portfolio from November 2015 to December 2018.

Visual	Usage
Card	Present 6 key figures (Business Metrics) relevant to the motor insurance business: (1) Premium, (2) Cost of Claims, (3) Cost of Claims : Premium ratio, (4) Number of Policies, (5) Number of Insured & (6) Number of Vehicles
Line and Clustered Column Chart	Visualize quarterly variations and the net gain percentage between premiums collected and claims paid out, as well as the similar balance of new policies versus lapsed policies.
Donut Chart	Visualize the channels for policy distributions (through agent or broker).
Area Chart	Provide general characteristics of the insured including age, driving experience and seniority.
Stacked Bar Chart	Visualize the number of insured vehicles based on the vehicle types and areas.
Matrix	Provide general characteristics of the insured vehicles including their average value, age and power.

B. By-Year Analysis

The dashboard presents an evaluation of premiums and costs of claims annually, taking into account various influencing factors.

Visual	Usage
Parameter Slicer	Allow users to evaluate trends and patterns of average premium and cost of claims with different dimensions.
Area Chart	Visualize average premium by different dimensions.
Column Chart	Visualize policy surplus (premium – cost of claims), no. of claims and cost of claims by different dimensions.

C. Client Scoring & Grading System

The dashboard utilizes a client segmentation approach to identify distinct customer groups based on their profiles and behaviors. This segmentation is achieved through a three-step process:

1. <u>Identifying scoring factors</u>

   The team identified three key scoring factors that reflect different aspects of customer behaviors and their relationship with the insurance company:

   ✧ Claim Ratio: This factor measures the frequency of claims filed by a customer during the whole coverage period.

   ✧ Total Years of Association: This factor reflects customer loyalty and the strength of their long-term relationship with the insurance company.

   ✧ Maximum Number of Policies: This factor indicates the customer's willingness to purchase multiple insurance policies from the company, reflecting their overall satisfaction and trust in the company.

2. <u>Defining scoring criteria</u>

Visual	Usage
Box plot	Visualize the distribution of each factor for classifying client values. Establish scoring criteria based on the provided interquartile range and median, resulting in a three-tier scoring system (Score: 3-1, Grade: A-C) for each factor.

3. <u>Performing client segmentation</u>

   The individual scores for each factor are summed for each insured, resulting in a total score. This total score is then used to categorize insured into different segments, as outlined in the table below.

Total scores ▼	Level	Grade	No. of Insured
⊟ 9	⊟ Top	3A	2495
⊟ 8	⊟ Top	2A1B	6201
⊟ 7	⊟ High	1A2B	8495
⊟ 6	⊟ Mid	1A1B1C / 3B	9516
⊟ 5	⊟ Mid	1A2C/ 2B1C	15259
⊟ 4	⊟ Low	1B2C	4098
⊟ 3	⊟ Bottom	3C	2202

Based on this Client Scoring & Grading System, the ultimate outcome of client segmentation is visualized through:

Visual	Usage
Treemap	Display the distribution of customers across these segments.

D. Other Dashboard interaction:

Feature	Usage
Tooltip	Present valuable details to bolster user comprehension.
Filter	Allow users to focus on specific data (Year 2015-2018) for clearer analysis and understanding.
Page-navigator	Facilitate easy navigation between different report pages, enhancing user interaction and readability.

## Part 4. Highlights and Visualization Effectiveness
 A.   Trend analysis of key figures
Premium and Cost of Claims (Left)

It is a clustered column chart comparing the premium collected and cost of claims spent from November 2015 to December 2018, with a line highlighting variations in the net gain percentage in premium across different quarters. The net gain in premium was initially negative in Q4 2015 and Q1 2016, indicating that the cost of claims could not be covered by the premium collected. Despite fluctuations, the net gain in premium shows a significant increase from November 2015 to December 2018, reflecting an improvement in business performance in monetary terms over time.

Number of New and Lapsed Policies (Right)

It is a clustered column chart comparing the number of new and lapsed policies from November 2015 to December 2018, with a line highlighting variations in the net gain percentage in the number of policies across different quarters. The net gain percentage in Q4 2015 was 56.2%. This percentage dropped continuously until Q1 2017, a trend that the company's management might wish to note. From Q2 2017 to Q4 2018, the net gain percentage gradually increased, returning to the 40% - 50% range, indicating a recovery and improvement in business performance in terms of number of policies over time.
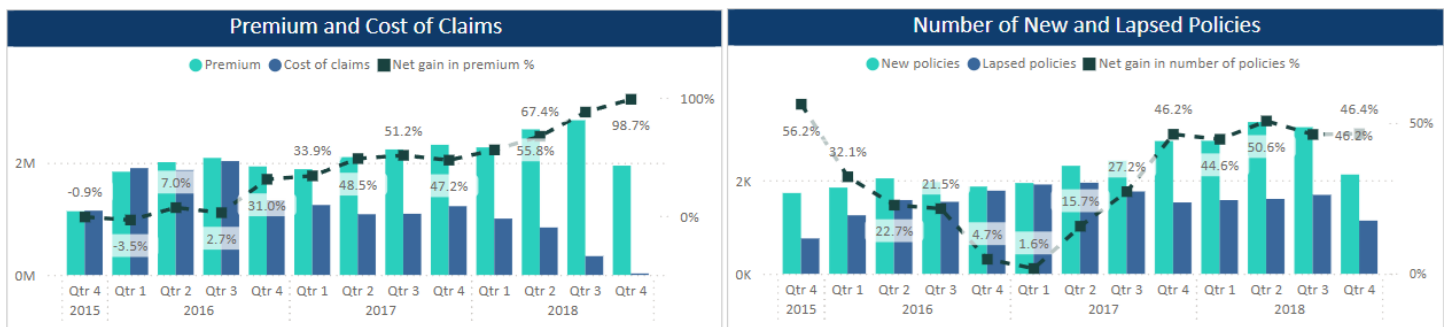

Ratio of Costs of claim : Premium

This ratio is prominently displayed at the top of the dashboard as a key performance indicator for the company. The industry benchmark for this ratio typically ranges from 0.6 to 0.7. To enhance performance relative to the industry standard, the team has set a target ratio of 0.6 or lower.

A ratio greater than 0.6 suggests that the company is spending a significant portion of the premium collected on claims, which could impact profitability (see the figures in 2015 and 2016).	A ratio of 0.6 or lower indicates that the company is managing its costs of claims effectively in relation to the premium. This is aligned with or better than the industry benchmark, suggesting a healthy financial position (see the figures in 2017 and 2018).
Costs of Claims : Premium **0.91** Goal: <= 0.60	**Costs of Claims : Premium** **0.54** Goal: <= 0.60

B. Variables related to Premium and Cost of Claims

Parameter for Average Premium: Vehicle Value, **Vehicle Age**, Vehicle Power, Driver Age, Years of Driving Experience Parameter for Cost of Claims: Vehicle Type, **Years Of Driving Experience**, Seniority, Area	The dashboard enables users to evaluate trends and patterns in average premiums and costs of claims across various dimensions. Dynamic interaction with a single graph is allowed by applying various dimension parameters. When a particular parameter is selected, the graph changes accordingly to reflect how that specific variable influences average premium or cost of claims, encouraging users to actively participate in data exploration. Overall, the factors align with expectations for premium and cost of claims in the insurance context, indicating that the current premium and cost of claims are reasonable.
 Surplus (Premium - Cost of Claims)	This section also helps users analyze the factors that affect surplus, calculated as premium minus cost of claims. If there is a surplus in a year under a particular factor, the positive value will be represented as an upward blue bar. Conversely, if there is a loss, the negative value will be represented as a downward red bar.

C. Client Segmentation

Distribution of Client Segments

Mid 51.33% | Top 18.02% | High 17.60% | Low 8.49% | Bottom 4.56%
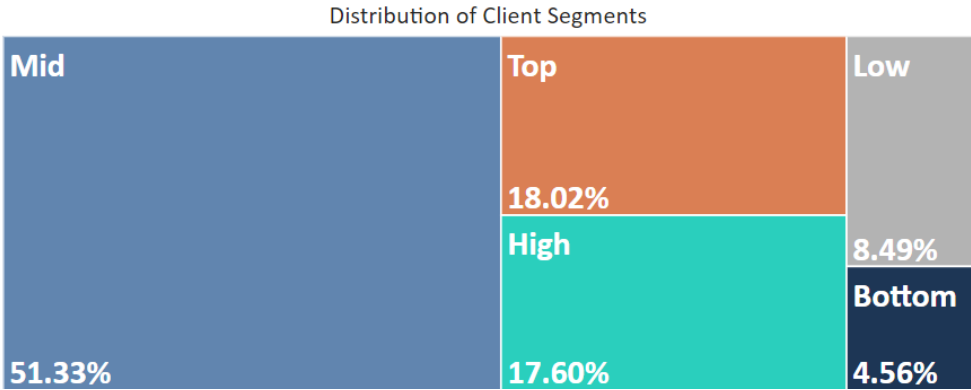
This treemap visually represents the distribution of clients across different segments in the motor insurance business, categorized based on their scores and assigned groups.

- Top Tier (18.02%): This segment represents the most valuable clients, indicating a strong potential for profitability and retention.
- High Tier (17.60%): These clients contribute significantly and demonstrate loyalty, making them essential for growth strategies.
- Mid Tier (51.33%): The largest segment, which may require targeted marketing efforts to elevate them to higher tiers.
- Low Tier (8.49%): This includes clients with lower engagement or value, suggesting areas for improvement in services or products.
- Bottom Tier (4.56%): This represents the least valuable customers, who may need reassessment regarding retention strategies.

The significant proportions of Top Tier and High Tier customers, which together account for 35.62% of the total client base, are essential for the long-term sustainability and growth of the motor insurance business.

Part 5. Insights

- The company continued to reduce its deficit as noticed from Q4 2015 to Q1 2016, and actively developed new customers.
- Since 2017, the company has achieved a lower ratio (cost of claims : premium) than the industry benchmark.
- Customers at Top Tier and High Tier are significant for supporting the business long-term development; therefore the company should concentrate resources on consolidating these customers, while also exploring opportunities to upgrade low and bottom-tier customers to a higher level.

Part 6. Challenges and Limitation

A. Data Restructuring

The team faced a significant challenge in restructuring the original dataset, which contained redundant information due to its design. Normalizing the data into a structured schema was a time-consuming task, requiring meticulous planning and execution to ensure efficient data retrieval and analysis for use in this project.

B. Limitation of dataset

The dataset does not include customer details such as gender, occupation, income, or other private information. If this information were available, we could conduct a more in-depth analysis of customer characteristics.

Part 7. Future Work

A. Analysis of traits of the Top and High Tier segments

While it is possible to identify premium clients using their Insured ID and Policy ID in the box plots, the team is eager to delve deeper by analyzing specific traits of the Top and High-value segments, aiming to help elevate clients to more valuable groups.

B. Year-over-year comparisons

The team's primary goal is to prioritize the level of granularity by focusing on the characteristics of insured and vehicles within each independent year, while maintaining a clear and concise dashboard layout that highlights key insights. Looking ahead, the team considers exploring the inclusion of year-over-year comparisons in the dashboard. This addition could help identify trends and patterns that may not be immediately apparent and assess improvements or declines in key metrics.

Part 8. Key takeaways

A. Data Model Formulation

As discussed, the initial challenge was constructing the data model in the BI tool due to unexpected dataset issues. This process facilitated a deeper understanding of the data and led to the creation of new variables for analysis.

B. Business Domain Knowledge

- Initially, a very low Claim-to-Premium ratio benchmark was defined; however, after reviewing industry norms, it was raised to 0.6, emphasizing the importance of aligning analysis with industry benchmarks.
- Besides, the team also referenced real-world premium practices to determine the dimensions used in premium analysis.

C. "Result-oriented" Analysis

- The team initially developed the dashboard with exploratory analysis but found the results unsatisfactory and lacking business context.
- By shifting the focus to defining key objectives and related metrics, the team redesigned the dashboard to deliver more meaningful and actionable insights.
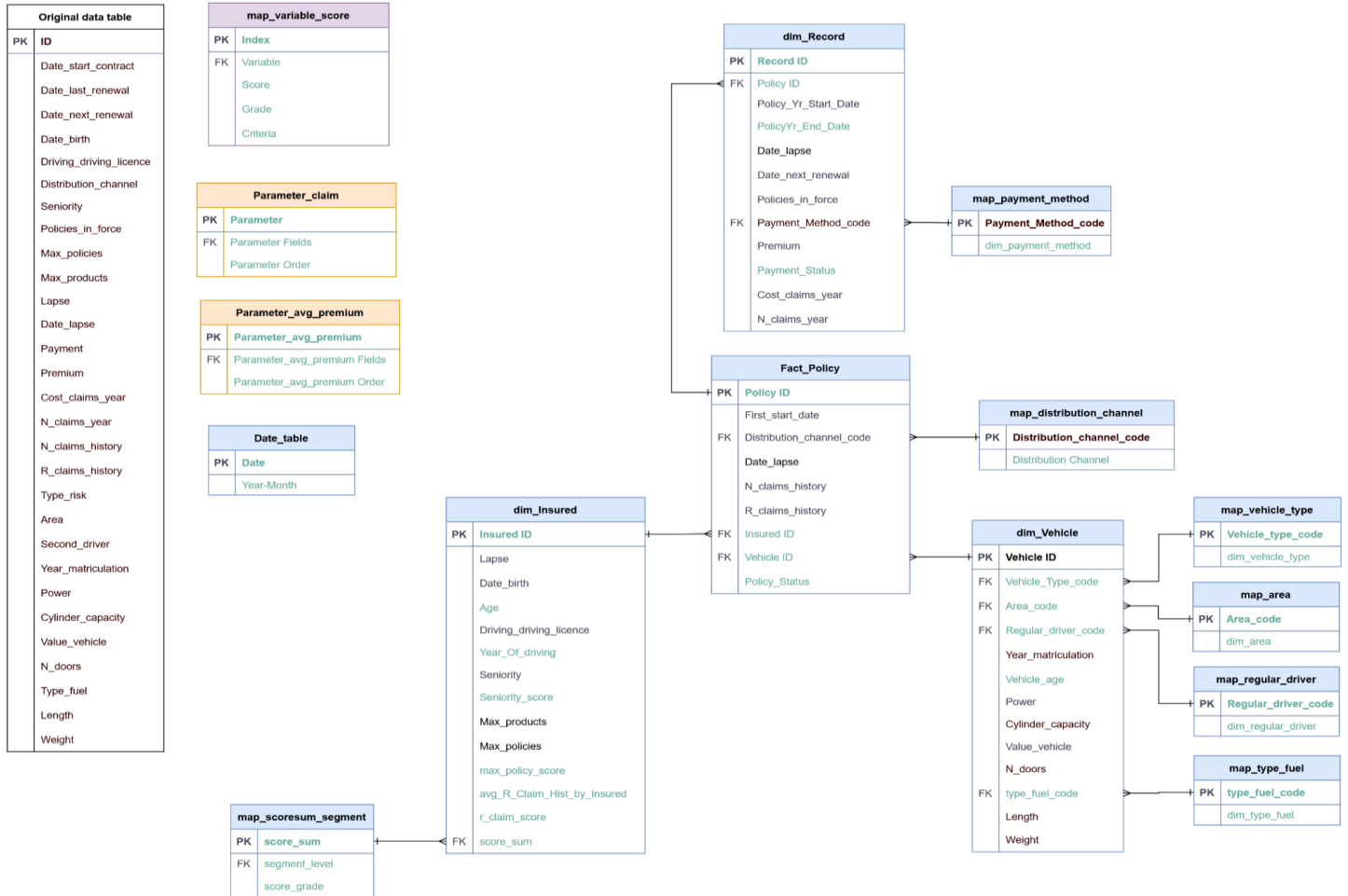
Part 9. Appendix

A. Dataset

B. Entity-Relationship Diagram

new column added

Original data table

PK	ID
	Date_start_contract
	Date_last_renewal
	Date_next_renewal
	Date_birth
	Driving_driving_licence
	Distribution_channel
	Seniority
	Policies_in_force
	Max_policies
	Max_products
	Lapse
	Date_lapse
	Payment
	Premium
	Cost_claims_year
	N_claims_year
	N_claims_history
	R_claims_history
	Type_risk
	Area
	Second_driver
	Year_matriculation
	Power
	Cylinder_capacity
	Value_vehicle
	N_doors
	Type_fuel
	Length
	Weight

map_variable_score

PK	Index
FK	Variable
	Score
	Grade
	Criteria

Parameter_claim

PK	Parameter
FK	Parameter Fields
	Parameter Order

Parameter_avg_premium

PK	Parameter_avg_premium
FK	Parameter_avg_premium Fields
	Parameter_avg_premium Order

Date_table

PK	Date
	Year-Month

dim_Record

PK	Record ID
FK	Policy ID
	Policy_Yr_Start_Date
	PolicyYr_End_Date
	Date_lapse
	Date_next_renewal
	Policies_in_force
FK	Payment_Method_code
	Premium
	Payment_Status
	Cost_claims_year
	N_claims_year

map_payment_method

PK	Payment_Method_code
	dim_payment_method

Fact_Policy

PK	Policy ID
	First_start_date
FK	Distribution_channel_code
	Date_lapse
	N_claims_history
	R_claims_history
FK	Insured ID
FK	Vehicle ID
	Policy_Status

map_distribution_channel

PK	Distribution_channel_code
	Distribution Channel

dim_Insured

PK	Insured ID
	Lapse
	Date_birth
	Age
	Driving_driving_licence
	Year_Of_driving
	Seniority
	Seniority_score
	Max_products
	Max_policies
	max_policy_score
	avg_R_Claim_Hist_by_Insured
	r_claim_score
	score_sum

dim_Vehicle

PK	Vehicle ID
FK	Vehicle_Type_code
FK	Area_code
FK	Regular_driver_code
	Year_matriculation
	Vehicle_age
	Power
	Cylinder_capacity
	Value_vehicle
	N_doors
FK	type_fuel_code
	Length
	Weight

map_vehicle_type

PK	Vehicle_type_code
	dim_vehicle_type

map_area

PK	Area_code
	dim_area

map_regular_driver

PK	Regular_driver_code
	dim_regular_driver

map_type_fuel

PK	type_fuel_code
	dim_type_fuel

map_scoresum_segment

PK	score_sum
FK	segment_level
	score_grade

C. DAX (Reference)

D. Common rating features for motor insurance (Reference)