SUSEP
Seguro Auto

Alya Hussain Alsharakhat

DAB16

Agenda:

Introduction

This report examines the behavior of vehicle insurance claims in Brazil between 2017 and 2020, using aggregated data provided by SUSEP, the national insurance regulator. SUSEP publishes de-identified and aggregated data to protect customer privacy while allowing analysts to study market trends, assess risk, and evaluate insurer performance.

Before diving into the real data, the presentation introduces the basic idea of how insurance works using a simple illustration: a group of people contributing money into a shared “pool” so that those who experience losses can be compensated. In practice, insurers operate on the same principle. To keep the system financially sustainable, insurers must understand three core components of risk:

- How often claims occur (frequency)
- How severe or costly claims are (severity)
- Which groups or segments generate the highest loss ratio (profitability)

With these principles in mind, this study analyzes multiple dimensions of risk, such as vehicle brand, vehicle age, vehicle category, driver demographics, and geographic region, to understand which segments consistently drive higher losses.

Scenario

An insurance company in Brazil wants to Improve their underwriting decisions and understand what makes cars high-risk using vehicle insurance data from 2017-2020 published by SUSEP, the Superintendence of Private Insurance (Superintendência de Seguros Privados), which is the Brazilian federal regulatory agency for the insurance, private pension, and capitalization sectors.

Approaches

1. **Number of Claims Analysis:** Understand how often claims occur across vehicles, drivers, and regions.
2. **Amount Paid Analysis:** Evaluate how costly claims are when they occur.
3. **Profitability Analysis:** Examine where the insurer makes or loses money, using loss ratios.
4. **Risk Segment :** Combine results from **Number of Claims**, **Amount Paid** , and **profitability** to classify vehicle, driver, and region segments into high-risk groups.

Target Audience

- Actuaries at insurance companies
- Insurance analysts responsible for underwriting
- Risk management teams

Dataset(s)

Source: SUSEP (Brazil's federal insurance regulator)

Years Covered: 2017–2020

Rows / Columns: 20,651,636 rows × 22 columns (before normalization)

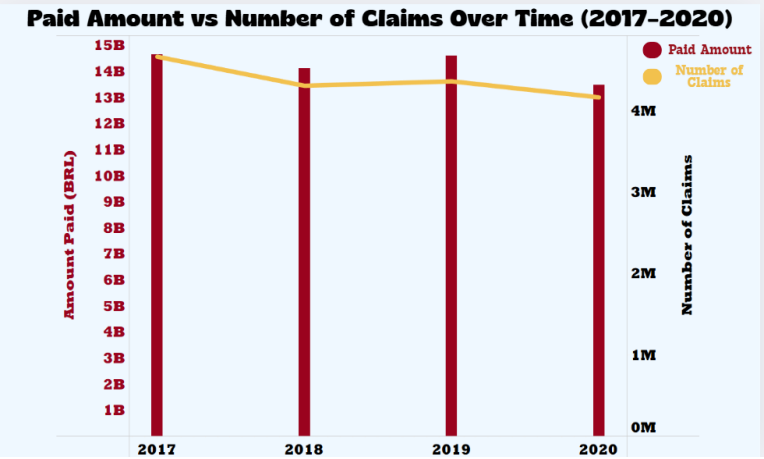
Old Column Name	New Column Name	Type	Description
COD_TARIF	Tariff Code	int	Insurance tariff category describing type/class of vehicle.
	Derived: Vehicle Category	object	Text label for vehicle category (from auto_cat.csv).
REGIAO	Region Code	int	Numeric code for Brazilian insurance zone.
	Derived: Region	object	Full regional name (from auto_reg.csv).
COD_MODELO	Vehicle Model Code	object	Unique ID for specific make/model (from auto2_vei.csv).
	Derived: Vehicle Model	object	Full model description.
	Derived: Vehicle Brand	object	Vehicle manufacturer.
ANO_MODELO	Model Year	int	Manufacturing year.
SEXO	Gender Code	object	Encoded gender or entity type.
	Derived: Gender	object	Clean gender label.
IDADE	Driver Age Code	int	Encoded driver age group.
	Derived: Driver Age Group	object	Readable driver age range.
EXPOSICAO1	Exposure1	float	Exposure time for claim cause 1.
PREMIO1	Premium1	float	Premium charged for cause 1.
EXPOSICAO2	Exposure2	int	Exposure for cause 2.
PREMIO2	Premium2	int	Premium for cause 2.
IS_MEDIA	Avg Insured Value	float	Average insured vehicle value.
FREQ_SIN1	Claim Freq1	int	Theft-related claim count.
INDENIZ1	Indemnity1	float	Theft claim payout.
FREQ_SIN2	ClaimFreq2	int	Partial collision claims.
INDENIZ2	Indemnity2	float	Partial collision payout.
FREQ_SIN3	Claim Freq3	int	Total loss collision claims.
INDENIZ3	Indemnity3	float	Total loss payout.
FREQ_SIN4	Claim Freq4	int	Fire claims.
INDENIZ4	Indemnity4	float	Fire damage payout.
FREQ_SIN9	Claim Freq9	int	Miscellaneous claims.
INDENIZ9	Indemnity9	float	Miscellaneous payouts.
ENVIO	Submission Code	object	Submission period (A/B semester).

Additional Column	Type	Formula / Source	Description
Total Claims	int	SUM(FREQ_SIN1 + FREQ_SIN2 + FREQ_SIN3 + FREQ_SIN4 + FREQ_SIN9)	Total number of claims across all claim types for each policy.
Total Indemnity	float	SUM(INDENIZ1 + INDENIZ2 + INDENIZ3 + INDENIZ4 + INDENIZ9)	Total payout amount for all claim types (overall claim cost).
Loss Ratio	float	Total_Indemnity / Total_Premium	Measures profitability: how much the insurer paid vs. how much premium was collected.
Avg Claim Size	float	Total_Indemnity / Total_Claims	Average payout per claim (severity indicator).
Claim Freq per Expo	float	Total_Claims / Exposure1	Claims standardized by exposure (risk rate).
Premium per Exposure	float	Premium1 / Exposure1	Premium relative to exposure; risk-adjusted premium measure.
Vehicle Age	int	Submission_Year - Model_Year	Age of the vehicle at the time of submission (used for risk segmentation).
Profitability Flag	string	IF Loss_Ratio < 1 THEN "Profitable" ELSE "Loss" END	Quick categorical indicator of profitability for each policy.
Claim Status	string	IF Total_Claims = 0 THEN "No Claim" ELSE IF Total_Indemnity = 0 THEN "Unpaid Claim" ELSE "Paid Claim"	Identifies whether a policy had no claim, an unpaid claim, or a paid claim.
Has Claim	string	IF Total_Claims > 0 THEN "Claim" ELSE "No Claim"	Binary indicator used for filtering and segmentation.
State	string	TRIM(LEFT(Region, 2))	Two-letter UF code extracted from the region description.
State Name	string	CASE [State] WHEN "SP" THEN "São Paulo" WHEN "RJ" THEN "Rio de Janeiro" ... END	Full state name used in visualizations and geographic segmentation.

Data Handling

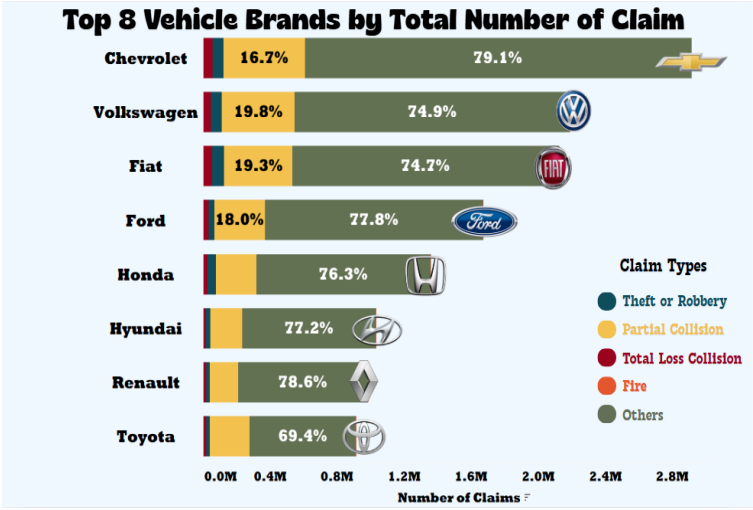
Data cleaning and preprocessing were conducted in Python (Jupyter Notebook), including merging files, deriving new variables, and preparing fields for analysis. Additional calculated fields created directly in Tableau (such as adjusted loss ratio and state name extraction) .

Analysis and Findings

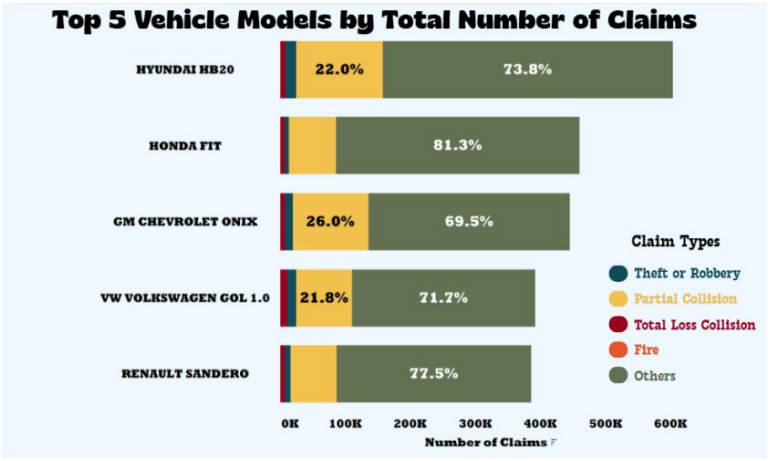


- The number of claims gradually decreased from 2017 to 2020, showing a downward trend. Meanwhile, the total amount paid remained relatively stable, with a small peak in 2019 before dropping slightly in 2020. This suggests that even with fewer claims, the overall cost to the insurer stayed high.

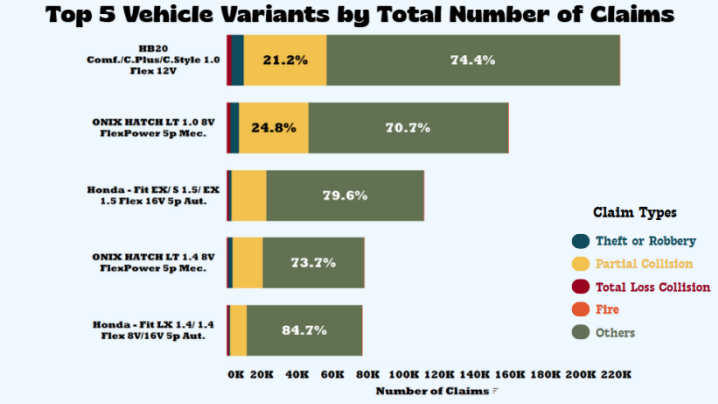
Number of Claims Analysis:



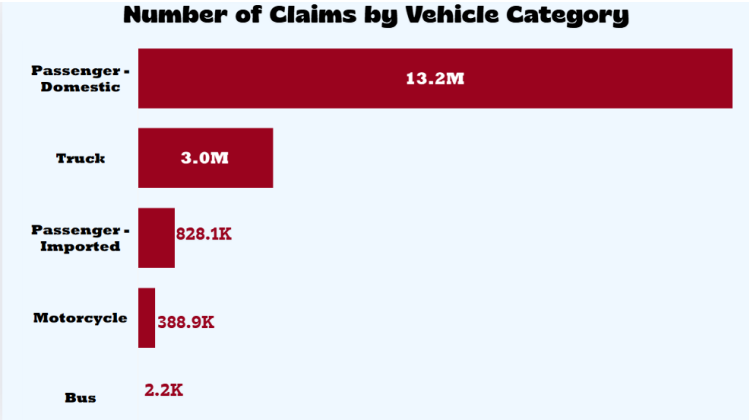
- Chevrolet, Volkswagen, and Fiat record the highest number of claims, with most claims falling under the “Other” category (non-collision or miscellaneous incidents). Toyota also appears in the top eight, but with a lower claim share compared to the leading brands. Across all brands, partial collisions and other incidents make up most claims.



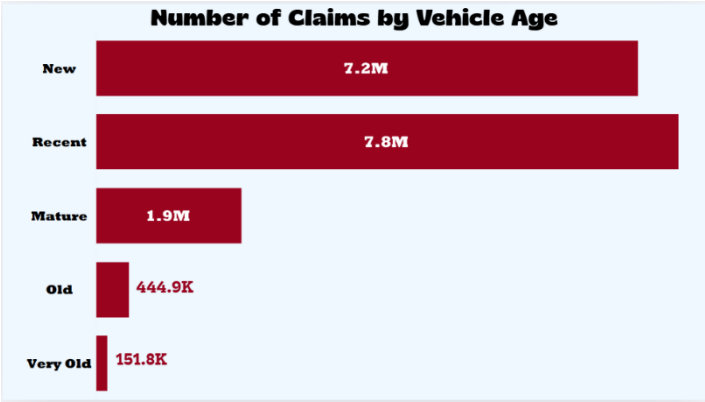
- The Hyundai HB20, Honda Fit, and Chevrolet Onix are the models with the highest number of claims. For all top models, most claims fall under the “Other” and “Partial Collision” categories, indicating frequent minor or miscellaneous incidents rather than severe accidents.



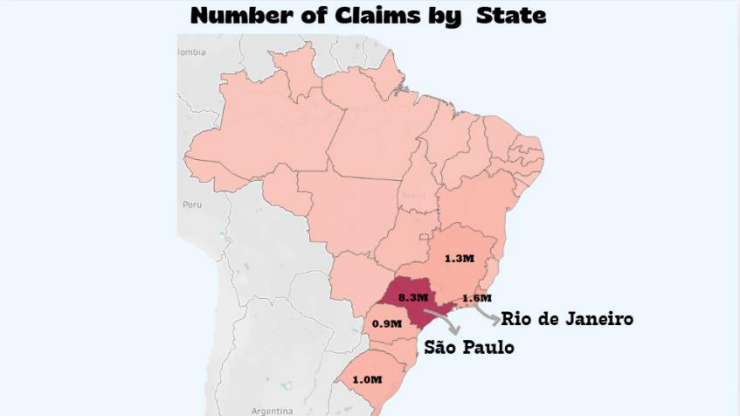
- The HB20 and Onix variants appear most frequently in claims, followed closely by Honda Fit variants. Across all five variants, most claims come from "Others" and "Partial Collision", showing that these vehicles tend to experience high volumes of minor or miscellaneous incidents rather than severe crashes.



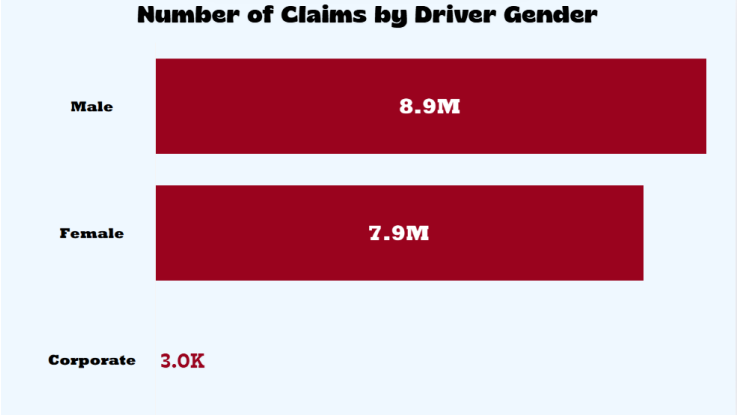
- Passenger–Domestic vehicles dominate the claims profile, with 13.2 million claims, far more than any other category. All remaining categories contribute only a small share. This shows that most claims are concentrated on domestic cars every day, which is expected because they represent the largest share of vehicles on the road.



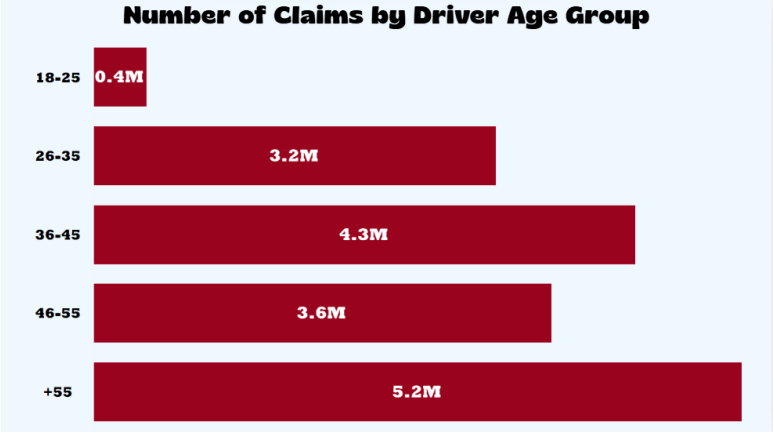
- Most claims come from New and Recent vehicles, with each category contributing over 7 million claims. As vehicles get older, the number of claims drops sharply " Mature, Old, and Very Old vehicles"
 represent a much smaller share. This suggests that newer cars are used more frequently and are more exposed, leading to higher claim volumes.



- claims are heavily concentrated in the Southeast region, with São Paulo and Rio de Janeiro showing the highest volumes. The top states in number of claims were heavily affected by floods and mudslides in recent years.
These natural disasters increase vehicle damage dramatically, especially in densely populated areas.

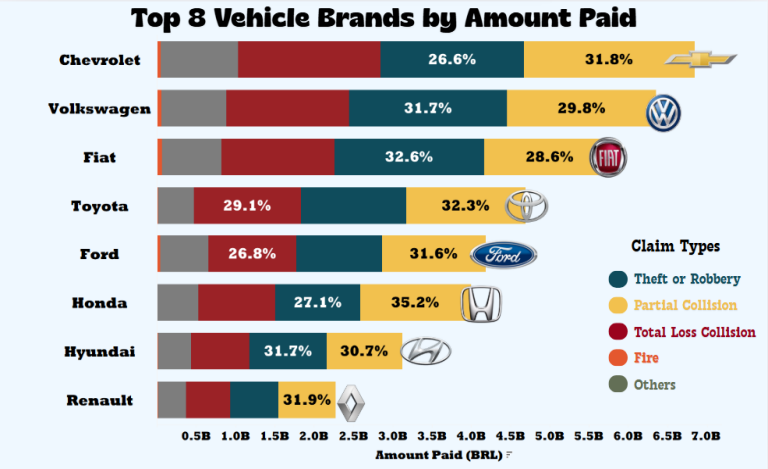


- Male drivers have the highest number of claims at 8.9M, followed closely by female drivers at 7.9M. Corporate policies show almost no claim activity (3K), indicating that most claims come from individual policyholders.

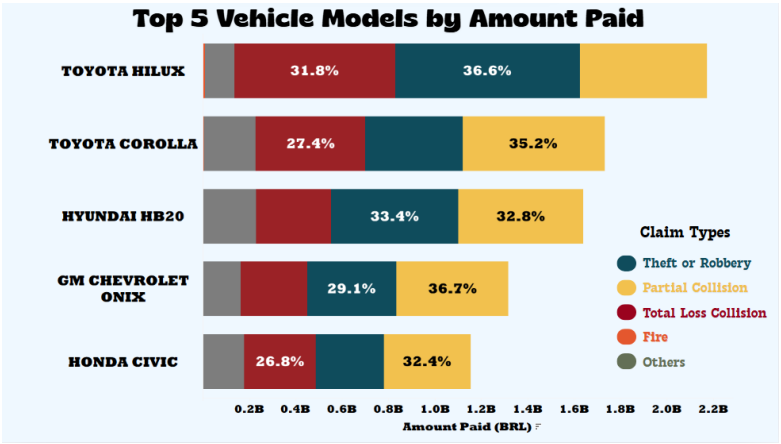


- The age group **55+** has the highest number of claims (**5.2M**), followed by **36–45** (**4.3M**) and **46–55** (**3.6M**). Younger drivers (18–25) generate the fewest claims (**0.4M**), likely due to lower vehicle ownership in this range.

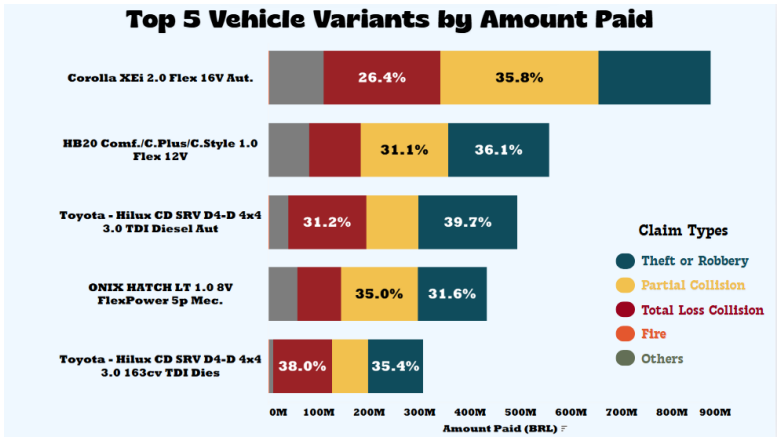
Amount Paid Analysis:



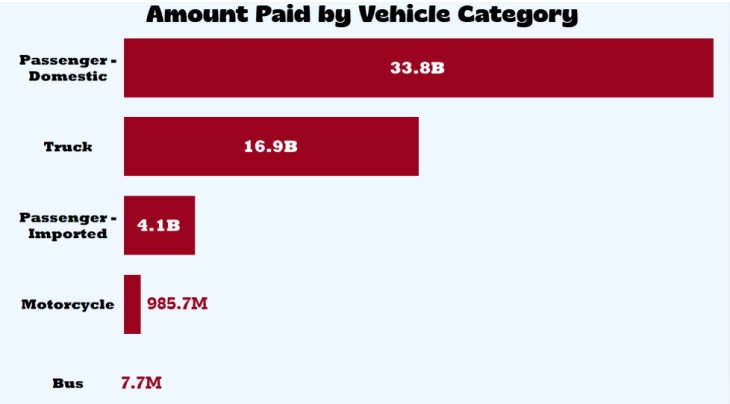
- This chart shows the top 8 vehicle brands by total amount paid in claims. Chevrolet, Volkswagen, and Fiat account for the highest payouts, largely because they have the largest presence in the market. Toyota also stands out with a high total payout, driven mainly by partial collision and total loss collision claims. The distribution of claim types shows most payments come from partial collisions, followed by theft/robbery and total loss events.



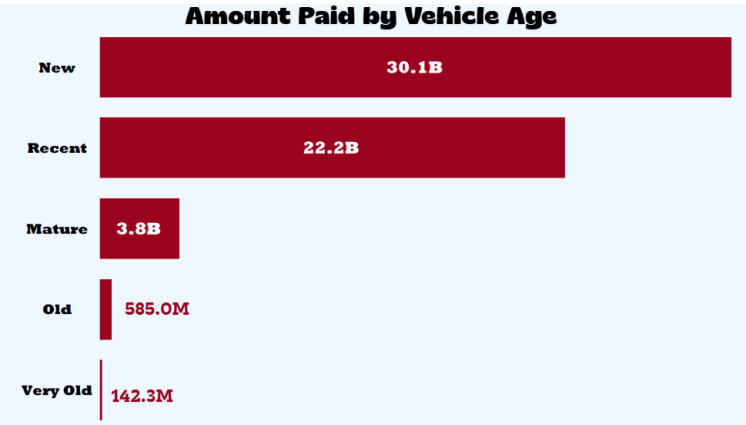
- This chart highlights the top 5 vehicle models with the highest total claim payouts. Toyota models (Hilux and Corolla) lead the ranking, together forming a significant portion of total costs. The Hilux stands out with high payouts driven mainly by total loss collisions and theft/robbery. Corolla, HB20, and Civic also show large payouts, mostly due to partial collisions.



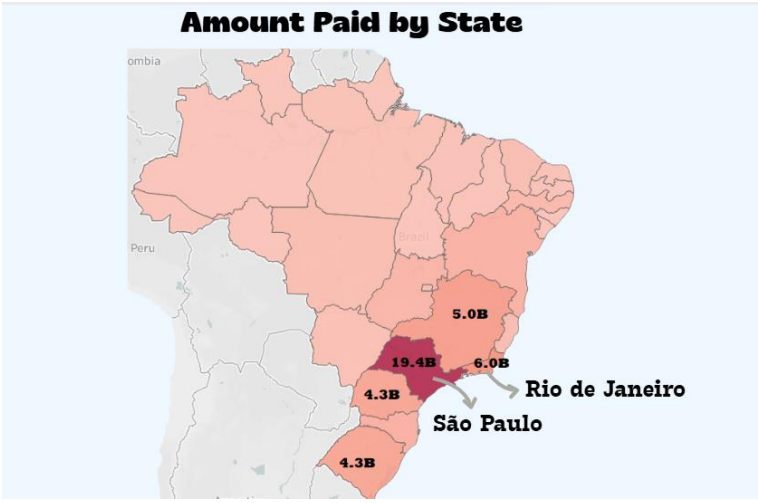
- This chart shows the top 5 vehicle variants with the highest total claim payouts. Toyota variants dominate the ranking, especially the Hilux CD SRV versions, which show large payouts mainly due to theft/robbery and partial collision claims. The Corolla XEi and HB20 variants also contribute significantly, with a high share of partial collision costs.



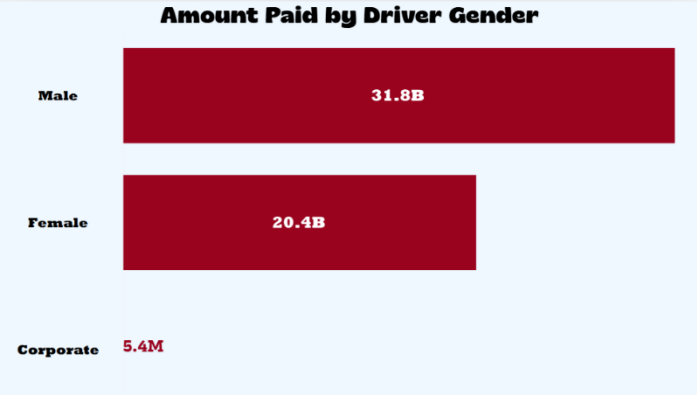
- This chart shows that Passenger–Domestic vehicles generate the highest total payout, reaching 33.8B BRL, followed by Trucks at 16.9B BRL. These two categories dominate the insurer's costs because they represent the largest share of insured vehicles and are more frequently involved in accidents and claims.



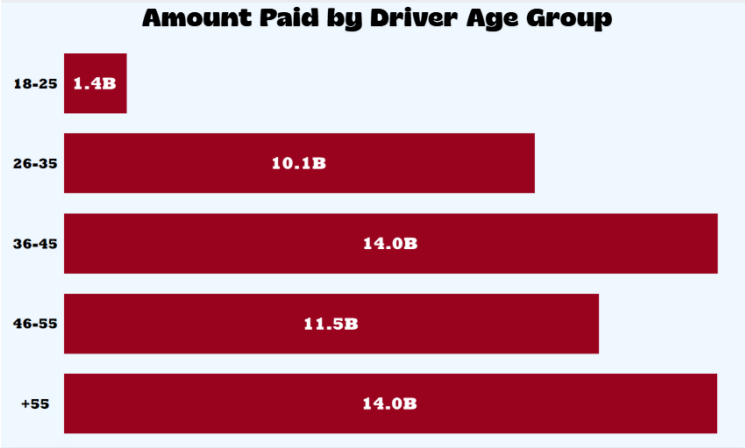
- This chart shows that **New and Recent vehicles generate the highest payout amounts**, with **30.1B BRL** and **22.2B BRL**, respectively. These vehicle ages dominate total costs because they are more valuable and therefore more expensive to repair or replace when claims occur.



- This map shows that São Paulo and Rio de Janeiro account for the highest total claim payouts, with 19.4B BRL and 6.0B BRL. These states have dense populations and high vehicle usage, which naturally leads to more costly claims.

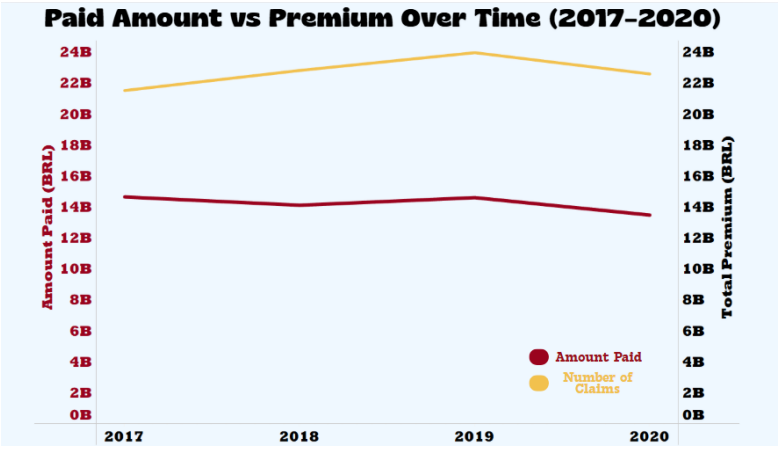


- This chart shows that male drivers account for the highest payout (31.8B BRL), significantly more than female drivers (20.4B BRL). The pattern aligns with the higher number of claims filed by male drivers, leading to higher total indemnity costs.

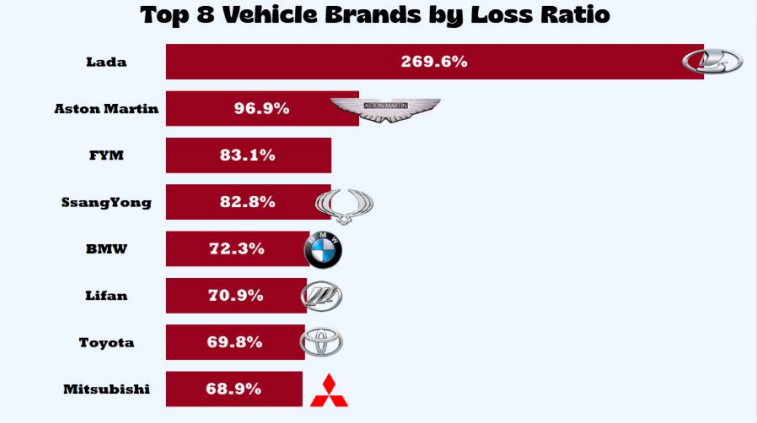


- This chart shows that the 36–45 and 55+ age groups have the highest total claim costs, each around 14.0B BRL, followed by 46–55 at 11.5B BRL. Younger drivers (18–25) have the lowest payout (1.4B), matching their much lower claim frequency.

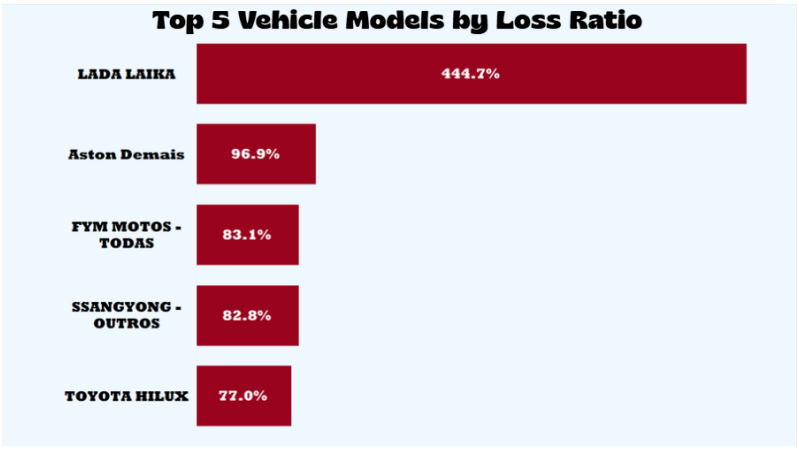
Profitability Analysis:



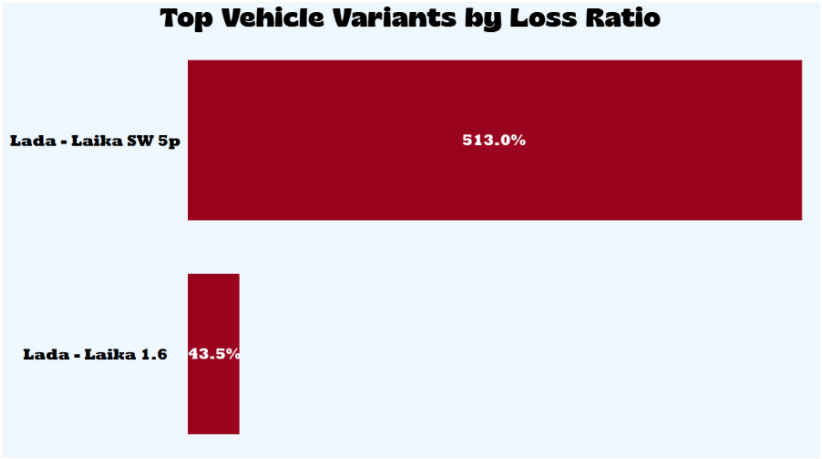
- This chart compares total premiums collected with total claim costs from 2017 to 2020. Premiums increased from 2017 to 2019 and then slightly dropped in 2020, while the paid amount stayed relatively stable but trended downward in 2020. Overall, premiums consistently remained higher than claims, indicating maintained profitability.



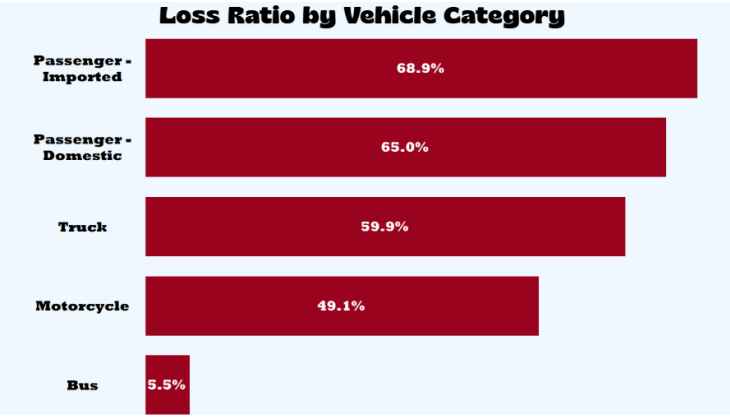
- This chart shows the vehicle brands with the highest loss ratios. Lada stands out with an extremely high loss ratio (269.6%), meaning claim costs are more than double the premiums collected. Other brands such as Aston Martin, FYM, and SsangYong also show elevated loss ratios, indicating underpriced premiums or higher-than-expected risk.



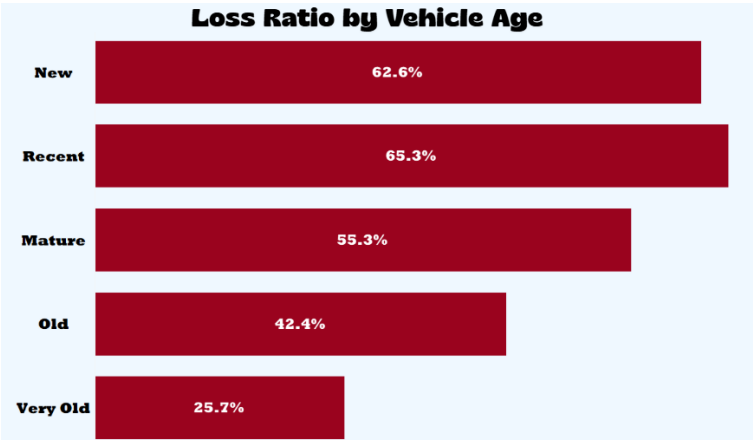
- This chart highlights the models with the highest loss ratios. The Lada Laika shows an extremely elevated loss ratio (444.7%), meaning claim payouts are more than four times the premiums collected — a major profitability concern. Other models like Aston Demais, FYM Motos, SsangYong, and Toyota Hilux also show loss ratios above typical thresholds.



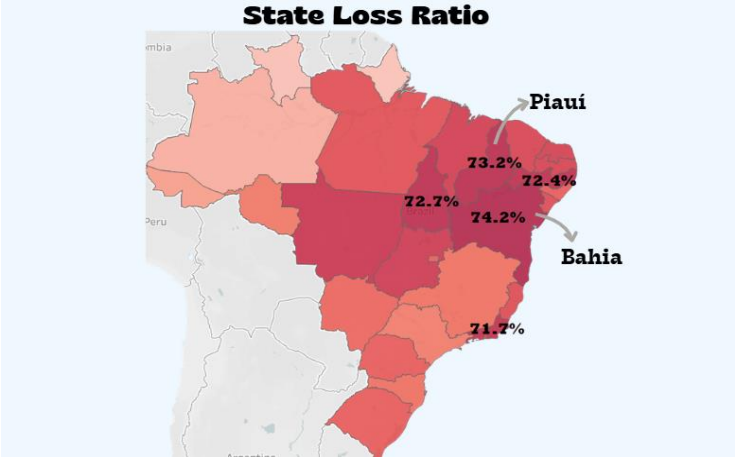
- The Lada Laika SW 5p variant shows an extremely high loss ratio of **513%**, meaning the insurer paid out more than five times the premium collected — a clear indication of severe underpricing or concentrated high-risk exposure. The second variant, Lada Laika 1.6, also has an elevated loss ratio (43.5%), reinforcing that this model family is significantly unprofitable.



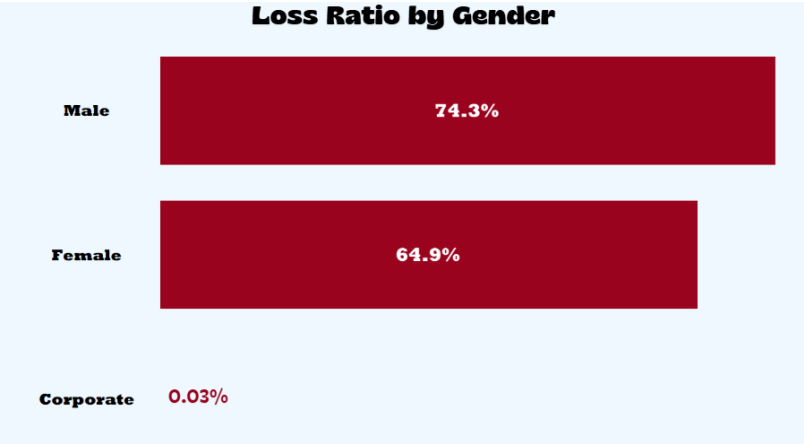
- Passenger-Imported vehicles show the highest loss ratio at 68.9%, indicating they are the least profitable category. Passenger-Domestic and Trucks also have moderately high loss ratios (65.0% and 59.9%), meaning these segments still cost the insurer a significant portion of the premium collected. Buses have the lowest loss ratio (5.5%), making them the most profitable and lowest-risk category.



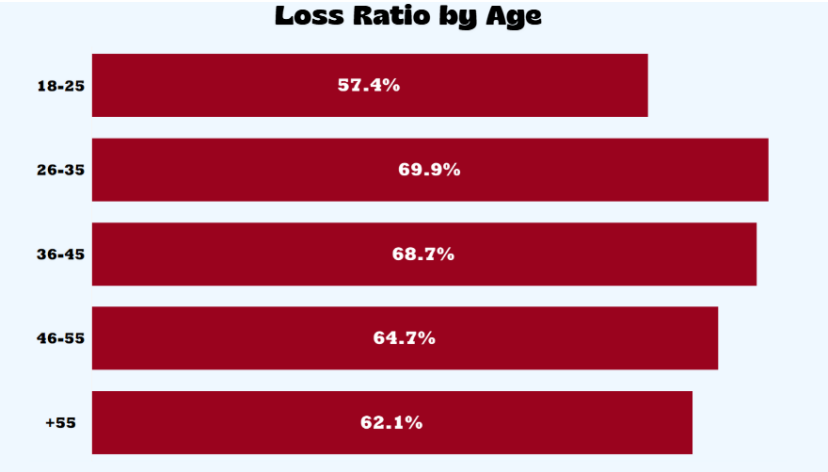
- Recent and New vehicles have the highest loss ratios (65.3% and 62.6%), indicating that newer cars are more expensive to repair and therefore less profitable for the insurer. Loss ratios decline steadily as vehicles age, with Very Old cars showing the lowest ratio at 25.7%. This means older vehicles generate less claim cost relative to the premium collected.



- Bahia, Piauí, Pernambuco, and neighboring Northeast states show the highest loss ratios, all above 70%. This pattern suggests that these regions experience higher claim costs relative to the premiums collected. Rio de Janeiro also stands out with a high loss ratio, indicating additional regional risk factors.



- Male drivers show the highest loss ratio at 74.3%, meaning their claim costs consume more of the premium collected. Female drivers follow at 64.9%. Corporate policies have almost no loss, as they represent a very small portion of the dataset.



- Loss ratios vary across age groups. The highest ratios are among 26–35 (69.9%) and 36–45 (68.7%), indicating these groups generate the most claim cost relative to premium. Younger drivers (18–25) show lower loss at 57.4%, and older groups (+55) also have more moderate ratios.

Risk Segments:

Across the full analysis, several segments consistently appeared at the top of all three risk dimensions: Number of Claims, Amount Paid, and Loss Ratio. These segments represent the areas where the insurer faces the highest concentration of financial exposure.

1. Vehicle Brand – Toyota
Toyota was the only brand appearing within the top ranks in all measures, indicating a combination of high claim frequency, high payout amounts, and a relatively high loss ratio.

2. Vehicle Age – Recent Vehicles
Recently manufactured vehicles (mid-life age) showed the highest claim counts and strong loss ratios, suggesting high usage intensity or insufficient pricing differentiation.

3. Vehicle Category – Passenger (Domestic)
This category recorded the highest number of claims and the largest total indemnity amounts, making it the main driver of portfolio losses.

4. State – Rio de Janeiro
Rio de Janeiro was the only state that consistently appeared among the top positions across all metrics, making it a leading geographical risk segment.

5. Driver Gender – Male
Male drivers reported the highest claim frequency and total payouts and also had a higher loss ratio than females, indicating elevated overall risk.

6. Driver Age Group – 36–45
This age group showed elevated claim frequency and payout amounts, making it one of the most active and costly segments within the portfolio.

Summary:
These segments represent the insurer's core high-risk areas and should be prioritized for pricing refinement, underwriting adjustments, and ongoing monitoring.

Extreme Loss Ratio Outlier:

One segment within the portfolio demonstrated exceptionally abnormal performance:
Lada – Laika SW 5p, with an extremely high loss ratio of 513%.

Recommendations

1. Adjust Premiums for High-Risk Vehicle Segments

Vehicle segments such as Toyota, Passenger–Domestic, and Recent vehicles consistently showed elevated claims and payout levels. Premium adjustments should be applied to ensure pricing properly reflects their risk exposure.

2. Strengthen Underwriting for High-Risk Drivers

Male drivers and drivers aged 36–45 represent the highest-risk groups across all measures. Underwriting rules—such as higher deductibles, usage verification, or additional documentation—should be reinforced for these segments.

3. Apply Territorial Pricing Adjustments

States such as Rio de Janeiro, which consistently appear at the top across all risk metrics, should receive specific territorial surcharges or refined pricing structures to reflect localized risk factors.

4. Re-Price Outlier Vehicles with Extremely High Loss Ratios

The Lada Laika SW 5p, with a loss ratio exceeding 500%, indicates severe mispricing. This vehicle should be re-priced, restricted, or excluded from standard underwriting to prevent disproportionate losses.

Limitations and Assumptions

- **unknown policy coverage type**: cannot determine comprehensive vs. partial coverage
- **Limited driver information**: only age and gender available
- **Safety features unknown**: no data on airbags, ESC, cameras, etc.
- **Claim type ambiguity**: Claim_Freq9 (“Others”) aggregates multiple rare claims
- **Aggregated data**: unclear if SUSEP totals represent sums or averages for indemnities

References

- The **official source** for the data: https://www2.susep.gov.br/menuestatistica/autoseg/principal.aspx
- São Paulo demographic reference: https://en.wikipedia.org/wiki/S%C3%A3o_Paulo#Demographics
- 2020 Brazil floods and mudslides: **https://en.wikipedia.org/wiki/2020_Brazilian_floods_and_mudslides?utm_source=chatgpt.com**
- Amazon rainforest wildfires (2019–2020): **https://en.wikipedia.org/wiki/2020_Brazil_rainforest_wildfires?utm_source=chatgpt.com**
  https://en.wikipedia.org/wiki/2019_Amazon_rainforest_wildfires