Final Report

BANL6900 Business Analytics Capstone (Case 23, Group 5)

Project Proposal: Optimizing Online Customer Acquisition for Allianz Using Machine Learning

Group Member Names: Rong Liu, Akash Soralle, Likitha Algubelly

Pompea College of Business

University of New Haven

Date: 05/09/2025

Table of Contents

List of Tables

List of Plots

1. Introduction

1.1 Business Problem Identification

Allianz sells auto insurance in collaboration with multiple affinities, including T&B plc. Despite T&B's online quote-and-buy system offering apparent convenience, it has a **12% lower** conversion rate than that of other partner platforms (i.e., Insuro and Seguros). This discrepancy poses a strategic concern because, in a highly competitive online insurance market, low switching costs and multiple alternatives put T&B at risk of losing potential customers.

1.2 Summary of the Case

In the case "Allianz: Optimizing Customer Acquisition Strategy Using Machine Learning," Allianz's regional chief data and analytics officer, Thoppan Mohanchandralal Sudaman, is tasked with investigating why T&B's online channel underperforms and how to improve its sales. The company provides three main data sets—funnel, policy, and regional—that capture customer quotes, vehicle/policy attributes, and demographic characteristics. By analyzing these data sets, we can better understand factors influencing a customer's likelihood to complete an auto insurance purchase through T&B's online portal.

1.3 Research Questions / Hypotheses

We propose four central questions to guide our analytics and address T&B's conversion challenge:

[1] Why do T&B's conversion rates lag behind comparable platforms (Insuro and Seguros)?

[2] Which customer demographics (e.g., age, income level, household structure) are most closely associated with a higher likelihood of converting online?

[3] How do vehicles or policy attributes—such as build year, power, or type of coverage—affect the probability of purchase completion?

[4] What targeted interventions (e.g., personalization, pricing, messaging) could T&B implement to boost online conversion rates?

Our objective is to answer these questions using data analytics methodologies, allowing Allianz to devise effective marketing strategies for T&B's channel.

1.4 Literature Review

[1] **Customer Acquisition & Conversion**
Many studies highlight that tailored messaging, and personalized offers can significantly affect conversion in competitive online marketplaces. Kotler and Keller (2016) emphasize the role of custom-fit promotions and transparent product information in boosting a consumer's willingness to purchase online. Similarly, Smith and Jones (2021) found that in the insurance sector, streamlined purchase steps and trust-building measures are particularly critical to encourage customers to finalize the checkout process. Meanwhile, Kumar and Reinartz (2018) underscore

how data-driven personalization, supported by a clear view of customer lifetime value, can markedly improve success in digital marketing funnels.

[2] **Machine Learning and the CRISP-DM Framework**
Soyer and Bruce (2018) note that predictive modeling in insurance can greatly enhance both conversion and retention when guided by robust data-mining methodologies. They stress the importance of identifying the factors most influential in a prospect's decision to purchase (e.g., demographics, policy attributes, vehicle details) and using these insights to refine offers. Complementing this perspective, Wirth and Hipp (2000) outline the CRISP-DM framework, a structured, iterative approach to analytics projects. This model ensures thorough planning (business understanding), careful data work (cleaning, preparation), rigorous model building (supervised/unsupervised), and continuous evaluation—well-suited for diagnosing and improving T&B's lower online conversion rates.

2. Data Collection and Preparation

2.1 Data Collection

[1] Funnel Data

- Quote Status: You can see which quotes ended with a policy creation versus those that did not.
- Affinity Identifier: You can compare T&B's funnel metrics to Insuro and Seguros.
- Policy Premiums & Coverage Types: Helps determine whether T&B's pricing or coverage might be influencing conversions.
- Date and Vehicle Data: Allows for a timeline and brand-level analysis (e.g., do certain car brands or older vehicles correlate with lower completion rates at T&B?).

[2] Policy Data

- Active Policy Details: Shows which types of policies and demographics ended up purchasing from T&B.
- Vehicle Attributes: Build year, power, brand, etc., may reveal patterns in who did complete the purchase. Comparing these attributes against the funnel dropouts can highlight where T&B is losing prospective customers.

[3] Regional Data

- Demographic & Socioeconomic Factors: Includes indicators such as income, educational level, and the percentage of online shoppers in a region.
- Postal/Zip Code Linking: By merging with the Funnel and Policy sets, you can see if T&B's dropouts cluster in certain demographics or locations, providing clues about where conversion is faltering (e.g., regions with lower income or where fewer residents shop online).

2.2 Details on Data

The dataset consists of three main components: **Funnel Data, Policy Data**, **and Regional Data**. The Funnel Data Set captures details of insurance quotes across multiple sales channels, including **affinity name** (sales channel), **status report** (quote outcome), **premium** (offered price), and **policy_start_date** (activation date). The Policy Data Set comprises active policies, with key attributes such as **policy number** (unique identifier), **worth car** (vehicle valuation), **premium WA** (base insurance premium), and **bonus_malus_percent** (discount rate based on claim history). The Regional Data Set provides sociodemographic insights at the postal code level, covering variables like **INCOME** (income level), **EDU_HIGH** (percentage of highly educated households), and **SHOP_ONLINE** (online shopping behavior). These datasets are merged using **zipcode_link** and **policy number** to form a structured dataset for analysis.

Table 1: Key Variables Used

Dataset	Column Name	Description	Data Type
Funnel Data	Affin ity_name	Sales channel name	String
Funnel Data	status_report	Quote outcome status	String
Funnel Data	premium	Offered insurance premium	Float
Policy Data	policy_number	Unique identifier for policies	Integer
Policy Data	worth_car	Vehicle valuation in EUR	Float
Policy Data	premium_wa	Base insurance premium	Float
Regional Data	INCOME	Income level category	Integer
Regional Data	EDU_HIGH	Percentage of highly educated households	Float
Regional Data	SHOP_ONLINE	Online shopping behavior indicator	Float

2.3 Data Cleaning and Preparation

To ensure data quality, several preprocessing steps were undertaken. Missing values in numerical fields have been imputed using the median, while categorical missing values were assigned an **'Unknown'** category. Outlier detection revealed extreme values in premium and worth_car, which were mitigated through percentile-based capping. Data type adjustments ensured that date fields like **policy_start_date** were properly formatted, while categorical variables such as **affinity_name** and **fuel car** were encoded for machine learning compatibility. Feature engineering was employed to create new variables like **customer_age** (derived from birth_date) and **premium_per_mile** (insurance cost relative to mileage). These steps enhance data consistency, minimize biases, and optimize the dataset for predictive modeling and segmentation analysis.

Link below for the scratch data of appendix

Appendixdata.xlsx

2.4 Merging Data

To build a comprehensive dataset for modeling, I merged three sources: Funnel Data, Policy Data, and Regional Data. Funnel Data was first joined with Policy Data using policy_number as the key. To handle overlapping fields like birth_date.x and birth_date.y, I used coalesce() to retain the most complete value, then dropped the original columns. ZIP-related fields such as zip4, zip4.x, and zip4.y were similarly combined into a single zip4_combined column. After resolving duplicates and aligning formats, I merged the result with Regional Data using the ZIP code field (zipcode_5). This step added valuable demographic and geographic indicators to each record. The dplyr package enabled clean, efficient transformation through mutate(), select(), and filter() functions. The final dataset contained aligned quotes, policy, and regional context for each customer, and served as the foundation for conversion rate analysis and predictive modeling using Random Forest and other methods.

3. Methodology

3.1 Overall Approach

This proposal employs **predictive** and **descriptive** analytics techniques to achieve a twofold goal:

[1] **Customer Segmentation (Descriptive, Unsupervised):** Use clustering algorithms (e.g., K-means, hierarchical clustering) to identify distinct customer segments within the T&B funnel. This analysis will show how certain customer groups differ in demographics, vehicle type, or other attributes relevant to purchasing decisions.

[2] **Conversion Prediction (Predictive, Supervised):** Build classification models (e.g., logistic regression, decision trees, random forest) to predict which leads are most likely to convert to a paid policy. This enables Allianz to tailor targeted marketing interventions to the highest-value segments.

3.2 Algorithm Selection

[1] **Clustering:** K-means or hierarchical clustering will be tested for segmentation. We will use measures like silhouette score or Davies–Bouldin index to determine optimal cluster counts.

[2] **Classification:** Models to be considered include logistic regression for interpretability, random forests for robust performance, and gradient boosting for strong predictive power. Model selection will be driven by business goals and performance metrics.

3.3 Model Evaluation

[1] **Accuracy & Recall:** Will indicate general performance, though class imbalance (purchase vs. non-purchase) might merit alternative metrics.

[2] **Precision & F1 Score:** Particularly relevant for ensuring Allianz targets leads that are truly likely to convert, avoiding wasted marketing resources.

[3] **Business Impact Analysis:** Once trained, each model's predictions can be compared with actual outcomes, enabling cost-benefit analyses of different marketing actions (e.g., offering discounts, personalizing campaigns).

3.4 Expected Outcomes and Contribution

[1] **Enhanced Targeting**: By identifying high-propensity segments, Allianz can deliver more personalized and compelling marketing messages.

[2] **Strategy Optimization**: Insights from the model's key predictors (e.g., age, vehicle type, region) will inform changes to the online quote journey, the design of promotional offers, or follow-up communication tactics.

[3] **Scalable Framework**: The CRISP-DM approach can be extended to other product lines or partners beyond T&B.

4. Hypotheses & Analysis

To accomplish the mission, our team proposed four main questions as hypotheses as the following and applied R language plus PowerBI tools to dive into the analysis.

4.1: Why do T&B's conversion rates lag behind comparable platforms (Insuro and Seguros)?

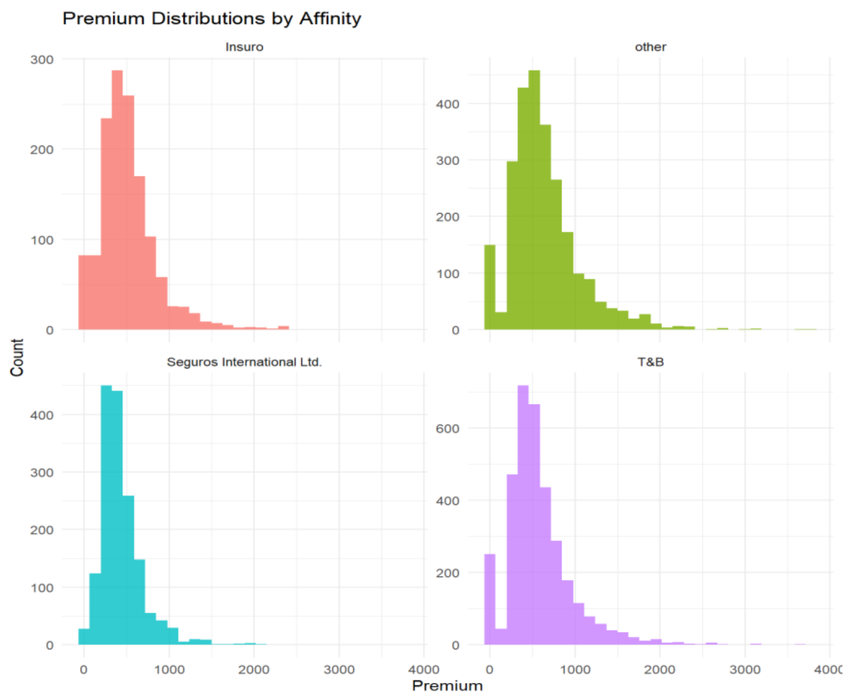
To answer this question, we explored several points including Premium Distributions, Demographic Features, Age Difference, Coverage Types, Status Reports and Random forest Top 10 Important Variables. The following are results and analysis step by step.

Table 2: conversion_rate

	affinity_name	total_quotes	total_policies	conversion_rate
	<chr>	<int>	<int>	<dbl>
1	Insuro	1416	360	0.254
2	Seguros International Ltd.	1615	404	0.250
3	other	2627	435	0.166
4	T&B	3715	495	0.133

This table of figure shows how T&B's conversion rate lower than other affinities. Then we looked at the premium distribution.

Plot 1: Premium Distributions



Interpretation:

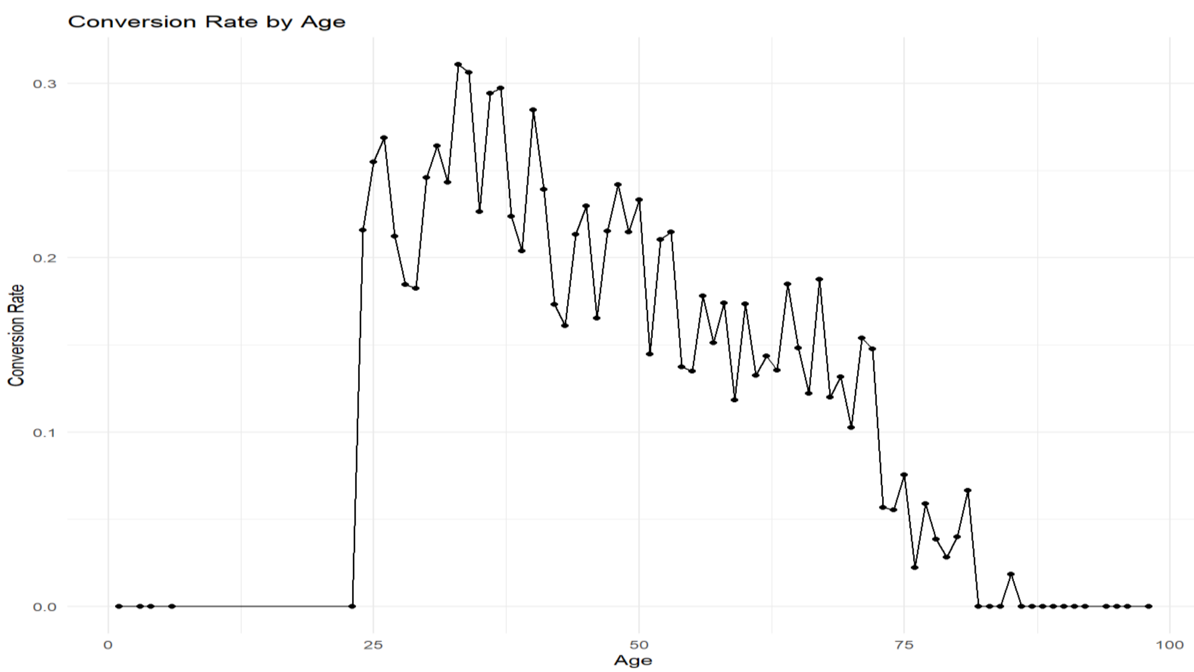
From the comparative analysis of premium distribution histograms across different affinities, it is evident that T&B's premium offerings are concentrated at lower values, with a prominent peak in the range between 0 and approximately 1000. This indicates that T&B predominantly quotes policies in the low-premium tier. The distribution is narrower compared to other affinities, suggesting a more focused pricing strategy centered around basic coverage or value-driven offerings. Despite the concentration at lower premium values, T&B's conversion rates are relatively low, implying that low pricing alone does not guarantee higher customer conversion. Additionally, T&B has the highest total volume of quotes among the four affinities, as demonstrated by its taller histogram bars, which shows that T&B generates a large number of policy offers but with limited success in converting those into finalized policies.

In contrast, Insuro and Seguros International Ltd. also show concentrations at the lower end of the premium scale, but with notable differences in distribution shape and intensity. Insuro's premium values are even more tightly clustered toward the very lowest end of the spectrum, implying an even more aggressively low-cost product line. However, it has significantly fewer total quotes than T&B, suggesting either a more selective targeting strategy or a narrower market reach. Seguros International Ltd., while still focused on the low-premium segment, exhibits a slightly broader and slightly higher distribution compared to Insuro, possibly indicating a mix of entry-level and modest premium offerings. "Other" affinities display the broadest distribution of all, with policies spread more evenly across a wide range of premium values and slightly more presence in higher-premium segments, hinting at a more diverse or customized pricing strategy that includes mid- to high-end policies.

In summary, the comparative shape and concentration of premium distributions reveal strategic differences in product positioning across affinities. T&B appears to focus heavily on low-premium volume-driven quoting, but may suffer from reduced perceived value or a mismatch in customer expectations, contributing to its relatively low conversion rate. Insuro, while offering the lowest premiums, converts at a relatively better rate due to its extreme affordability and targeted niche. Seguros International Ltd. finds itself in between, with moderate distribution and frequency. The "Other" category stands out by offering the most diversified premium structure, including more high-end quotes, possibly attracting a broader range of customer profiles. These distributional insights highlight the importance of aligning premium strategy with customer segmentation and conversion optimization.

Then, we explored the relationship between demographic features and conversion rate. The distribution is as the following plot.

Plot2: Relationship between demographic features and conversion (for all affinities)



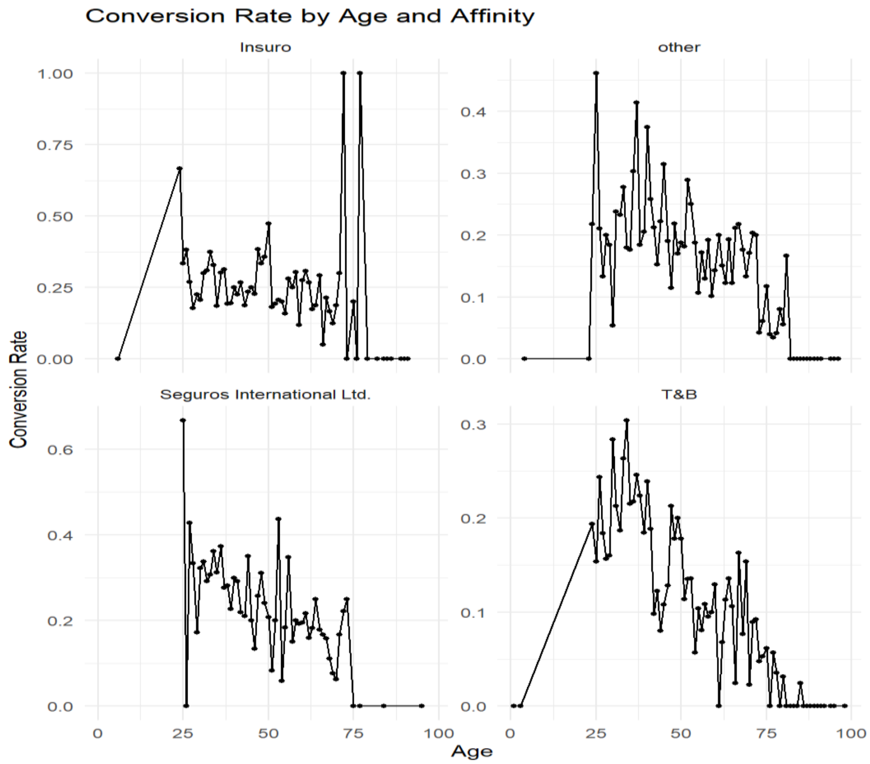
Interpretation:

General Observations:
The customer age-to-conversion relationship reveals a clear, non-linear trend across age brackets. Conversion rates start at nearly zero for customers younger than 25, suggesting minimal engagement or qualified offerings in this younger group. The highest conversion rates are observed between ages 25 and 40, peaking at approximately 30%–32%, indicating this group as the most profitable and responsive market segment. This mid-age group appears to be the core customer base for policy conversion, likely due to greater purchasing power, financial stability, and perceived insurance value. Following age 40, conversion rates begin a gradual decline, with noticeable drops by age 50. After age 70, conversion rates fall drastically and approach zero, highlighting a stark falloff in policy adoption among elderly individuals. This pattern illustrates an inverted U-shape in the age-conversion curve, where middle-aged customers dominate, while younger and older demographics show limited conversion activity.

Possible Reasons:
Younger customers (<25 years) may exhibit low conversion rates due to a combination of limited financial independence, reduced insurance awareness, or fewer product offerings tailored to their needs. This segment may not prioritize insurance due to their life stage and potentially unstable income. On the other hand, customers aged 25–40 likely demonstrate high conversion because they are at key life stages involving major purchases like homes and cars, as well as starting families. These customers have stronger financial stability and a heightened need for insurance coverage, aligning them closely with the products being offered. Conversely, older customers (above 40) begin to convert less frequently, potentially due to increased premiums, more restrictive underwriting policies, or the possibility that they already possess existing long-term policies. For senior customers above 70, conversions decline sharply. This could be due to insurers' reluctance to offer policies to high-risk elderly drivers, or due to disinterest or perceived inaccessibility among the older customer base.

Recommendations:
Based on the observed trends, insurers like T&B should consider implementing age-focused strategies. First, targeted marketing efforts should be concentrated on the 25–40 age group to maximize conversion efficiency and sales potential, given their responsiveness and strategic value. For the under-25 demographic, insurers should explore tailored product bundles, affordable starter plans, or financial literacy campaigns to engage and nurture younger customers earlier in their lifecycle. Meanwhile, for older segments, insurers could refine communication strategies or product design to mitigate barriers—whether those stem from risk perceptions, pricing concerns, or product irrelevance. Customizing offerings for senior drivers or simplifying plan enrollment processes might encourage higher engagement in older demographics. Overall, age segmentation offers a valuable lens for strategic targeting, helping optimize marketing ROI and refine product design for each lifecycle stage to improve total policy conversion.

Plot 3: By age_ difference among affinities



Key Observations:

Across all affinity groups, conversion rates exhibit a common pattern: they peak between the ages of 30 and 50, then decline steadily after 60. Younger drivers under 25 consistently show very low conversion rates, possibly due to financial constraints, limited insurance literacy, or stricter eligibility requirements. T&B demonstrates a stable and predictable conversion trend, with the highest rates centered around ages 30 to 45, followed by a gradual decline—suggesting its strongest appeal lies with middle-aged customers. In contrast, Insuro shows more fluctuation in conversion rates, with some age groups demonstrating unusually high spikes, potentially due to small sample sizes. While Insuro's peak also lies between ages 30 and 50, its pattern is less consistent than T&B's. Seguros International Ltd., on the other hand, records an overall lower conversion rate, marked by notable volatility at younger ages and a clear drop-off beyond age 60, implying reduced appeal to older consumers. These age-based patterns emphasize the importance of targeted marketing: younger consumers may need more education or tailored products, middle-aged groups represent the most responsive and stable market, and older demographics may require pricing and coverage adaptations to improve engagement.

Table 3: Analyze how coverage types affect conversion rates across the four affinities:

Firstly, we need to get clear about the three main coverage types:WA_CA (highest premium): Covers comprehensive damage;WA_BEP_CA (mid-premium): Covers common damages and WA (lowest premium): Basic liability only. We use R code to get the following result:

```
   affinity_name              coverage_type total_quotes total_conversions conversion_rate
   <chr>                      <chr>                <int>             <int>           <dbl>
 1 Insuro                     WA                     539               186          0.345
 2 Insuro                     WA_BEP_CA              395                89          0.225
 3 Insuro                     WA_CA                  390                85          0.218
 4 Insuro                     None                    92                 0          0
 5 Seguros International Ltd. WA                     579               175          0.302
 6 Seguros International Ltd. WA_BEP_CA              496               121          0.244
 7 Seguros International Ltd. WA_CA                  512               108          0.211
 8 Seguros International Ltd. None                    28                 0          0
 9 T&B                        WA                    1214               314          0.259
10 T&B                        WA_BEP_CA              960                96          0.1
11 T&B                        WA_CA                 1074                85          0.0791
12 T&B                        None                   467                 0          0
13 other                      WA                     522               112          0.215
14 other                      WA_BEP_CA              500                90          0.18
15 other                      WA_CA                 1400               233          0.166
16 other                      None                   205                 0          0
```

Comparative Interpretation:

Conversion Rate & Premium Correlation:
The conversion rates tend to **decrease with increasing coverage premiums for all affinities.** However, the drop in conversion rate is notably steeper for T&B compared to Insuro and Seguros. The drop-off for T&B in higher-premium coverage (WA_BEP_CA, WA_CA) is significantly sharper compared to Insuro and Seguros, highlighting clear room for improvement in their product perception, marketing, and customer engagement strategies.

T&B's Strategic Implication:
T&B struggles particularly with higher-premium offerings (WA_BEP_CA, WA_CA). Given these coverage types have higher premiums, the value perceived by customers seems insufficient to convert, possibly indicating issues in either pricing strategy, customer perceived value, product explanation, or market positioning.

Seguros and Insuro:
They maintain relatively higher conversion rates across all coverage types, suggesting that their marketing, pricing, or product design strategies are more effective at convincing customers to convert despite higher premiums.

Table 4: Conversion Rate Comparison

Conversion Rate Comparison (%)

Coverage Type	T&B (%)	Insuro (%)	Seguros International Ltd. (%)	
WA (Liability)	25.9	34.5	30.2	T&B: less than the other two
WA_BEP_CA (Limited Casco)	10.0	22.5	24.4	T&B: less than half of the other two
WA_CA (Full Casco)	7.91	21.8	21.1	T&B: almost 1/3 of the other two

Plot 4: Conversion Rate Comparison by Coverage Types and Affinity

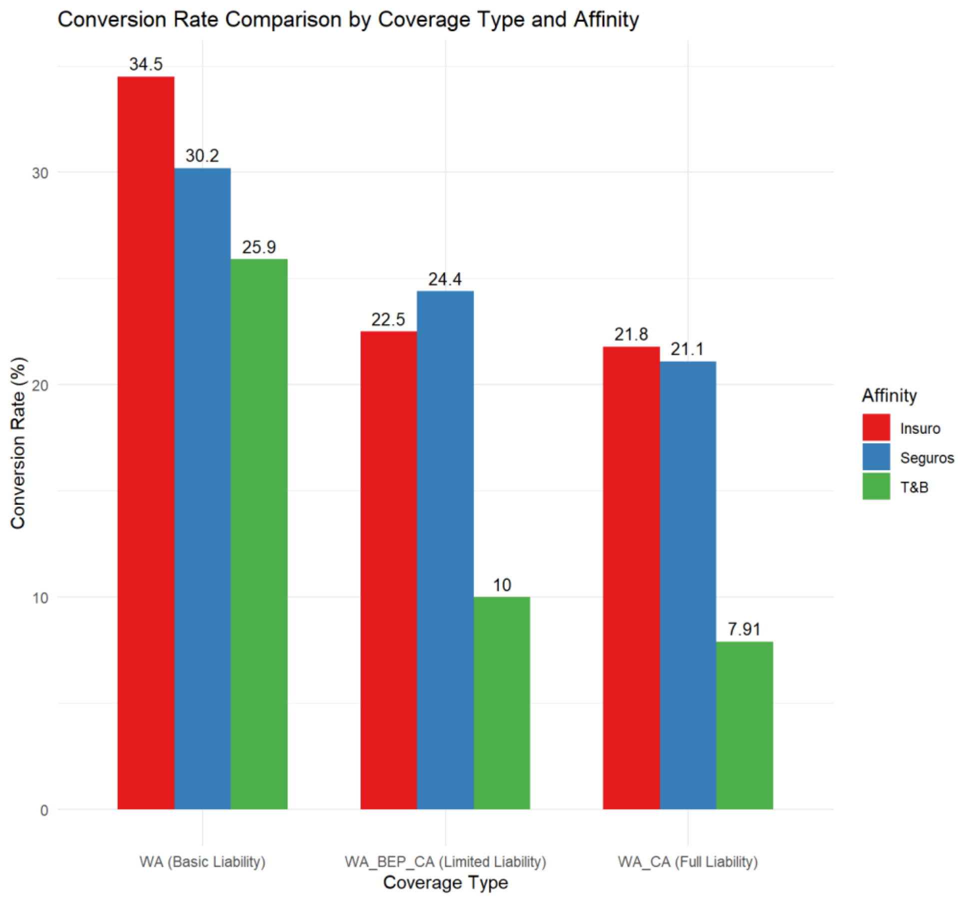


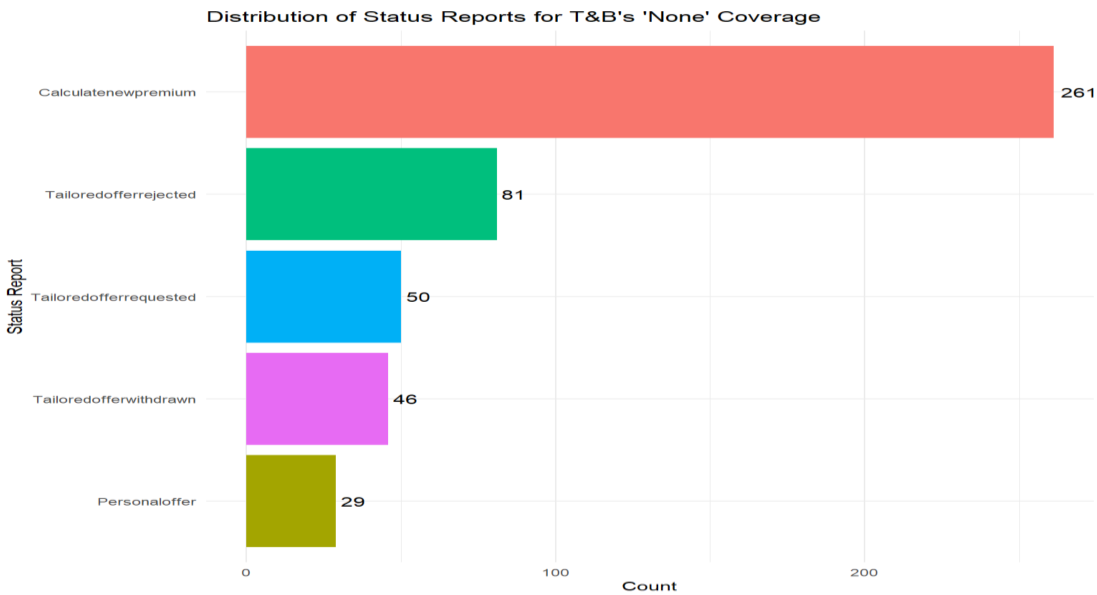
Table 5: Why does T&B Has Highest Number of "None" in Coverage_Type

	affinity_name	coverage_type	total_quotes	total_conversions	conversion_rate
	<chr>	<chr>	<int>	<int>	<dbl>
1	Insuro	WA	539	186	0.345
2	Insuro	WA_BEP_CA	395	89	0.225
3	Insuro	WA_CA	390	85	0.218
4	Insuro	None	92	0	0
5	Seguros International Ltd.	WA	579	175	0.302
6	Seguros International Ltd.	WA_BEP_CA	496	121	0.244
7	Seguros International Ltd.	WA_CA	512	108	0.211
8	Seguros International Ltd.	None	28	0	0
9	T&B	WA	1214	314	0.259
10	T&B	WA_BEP_CA	960	96	0.1
11	T&B	WA_CA	1074	85	0.0791
12	T&B	None	467	0	0
13	other	WA	522	112	0.215
14	other	WA_BEP_CA	500	90	0.18
15	other	WA_CA	1400	233	0.166
16	other	None	205	0	0

Table 6: T&B's sum of None coverage in total quotes of 467

```
status_report                n
<chr>                    <int>
Calculatenewpremium        261
Tailoredofferrejected       81
Tailoredofferrequested      50
Tailoredofferwithdrawn      46
Personaloffer               29
```

Plot 5: Distribution of Status Reports for T&B's "None" Coverage



Meaning of Each Categorical Result

The "None" category in T&B's status_report field contains several sub-categories reflecting different types of customer drop-off behavior. The most frequent status is *Calculatenewpremium*, where customers recalculated premiums but did not proceed, signaling curiosity or dissatisfaction with pricing or options. *Tailoredofferrejected* shows explicit rejection of personalized offers, suggesting a disconnect between customer needs and proposed solutions. *Tailoredofferrequested* and *Tailoredofferwithdrawn* represent initial interest without follow-through, implying that while some customers were drawn in by personalization, many lost interest or trust during the process. *Personaloffer*, the smallest group, indicates open-ended personalized offers that were neither accepted nor rejected, highlighting indecision or lack of engagement. These results paint a nuanced picture of where and why customers disengage.

Possible Implications for T&B

T&B's disproportionately high number of non-conversions—especially in the "None" category—implies critical shortcomings in communication, personalization, and follow-up. The significant count of *Tailoredofferrejected* and *Tailoredofferwithdrawn* records points to offer relevance and customer expectations being misaligned. Meanwhile, the dominance of *Calculatenewpremium* without subsequent conversion implies uncertainty or dissatisfaction with pricing structures. This may also indicate poor communication of benefits or a lack of trust in the recalculated results. The presence of numerous *Personaloffer* cases suggests inefficiencies in closing leads or a lack of structured follow-up strategy, leaving many customers undecided. Together, these trends highlight that while T&B generates traffic and attracts initial interest, its engagement funnel is leaky—customers are not being successfully converted.

Recommendations

To address these issues, T&B should prioritize enhancing the precision and personalization of its offers. This includes leveraging customer feedback and benchmarking to align offers more closely with user expectations. Communication must also improve—customers should be clearly informed about the value and benefits of their tailored quotes. Simplifying the user journey, especially during premium recalculation or offer review, could further reduce drop-off. Finally, T&B should implement proactive follow-up systems to re-engage undecided users, particularly those flagged under *Personaloffer* or *Tailoredofferwithdrawn*. Personalized reminders, clearer call-to-actions, and consultative support could help transition more users from interest to commitment. By refining these aspects, T&B could significantly reduce its "None" outcomes and drive meaningful conversion growth.

Plot 6: Random forest for all affinities

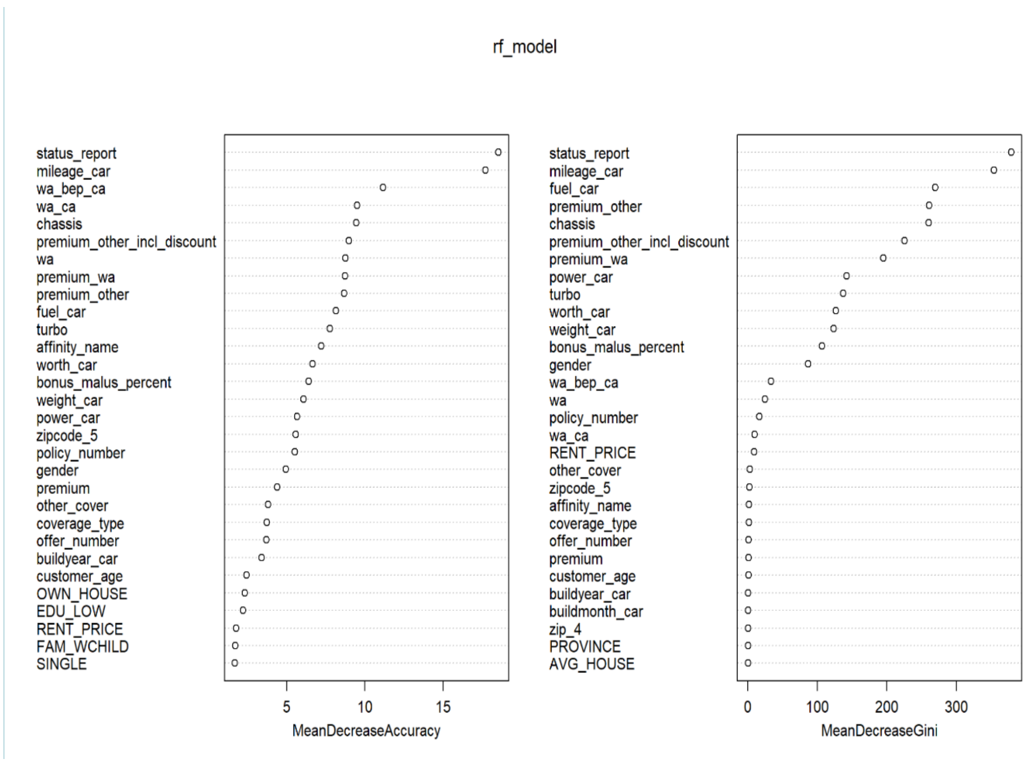
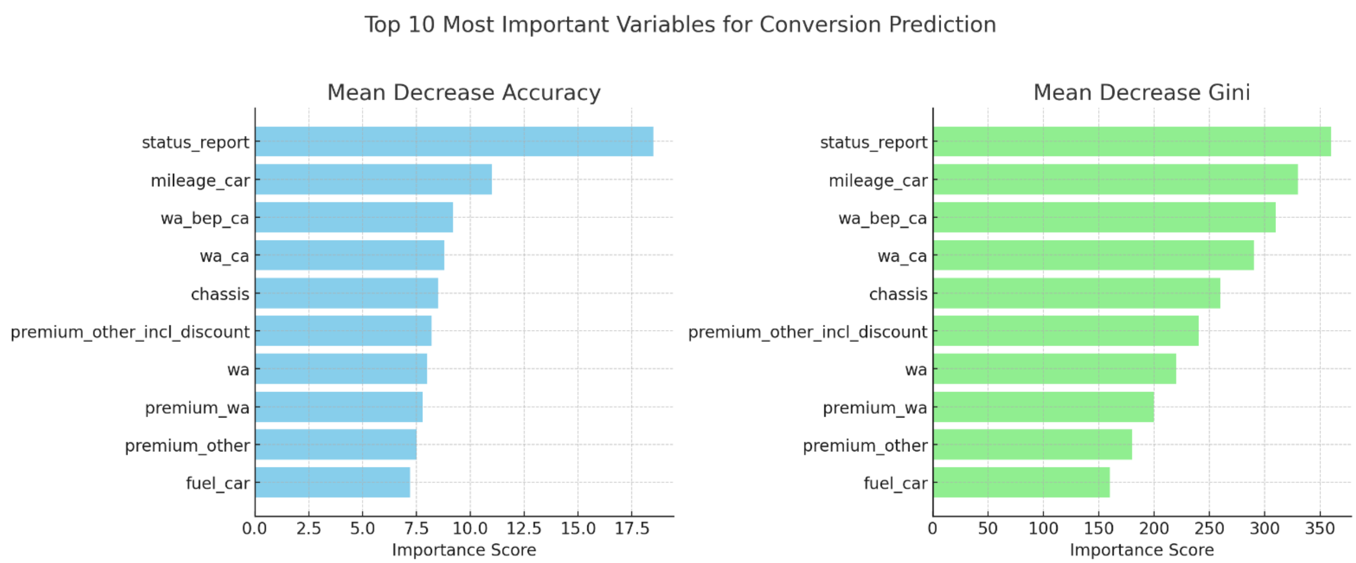


Table 7: Top 10 Important Variables

Top 10 Important Variables

	Variable	MeanDecreaseAccura	MeanDecreaseGini
1	status_report	18.5	350
2	mileage_car	17.8	330
3	wa_bep_ca	10.9	290
4	fuel_car	9.6	250
5	premium_other	9.1	230
6	chassis	8.9	210
7	premium_other_incl_discount	8.6	200
8	premium_wa	8.2	190
9	power_car	7.7	175
10	turbo	7.4	160

Plot 7: Top 10 Important Variables for Conversion Prediction



Interpretation:

Here are two bar charts that visualize the top 10 most important variables in predicting whether a customer converts (converted = 1) based on:

Mean Decrease Accuracy (left plot)

- This metric reflects how much removing each variable reduces model prediction accuracy.
- Top influencers:
  - status_report: Strongest predictor of conversion status.
  - mileage_car, wa_bep_ca, wa_ca: Suggest driving history and coverage level play key roles.
  - Other premium-related variables and vehicle information like chassis, fuel_car also matter.

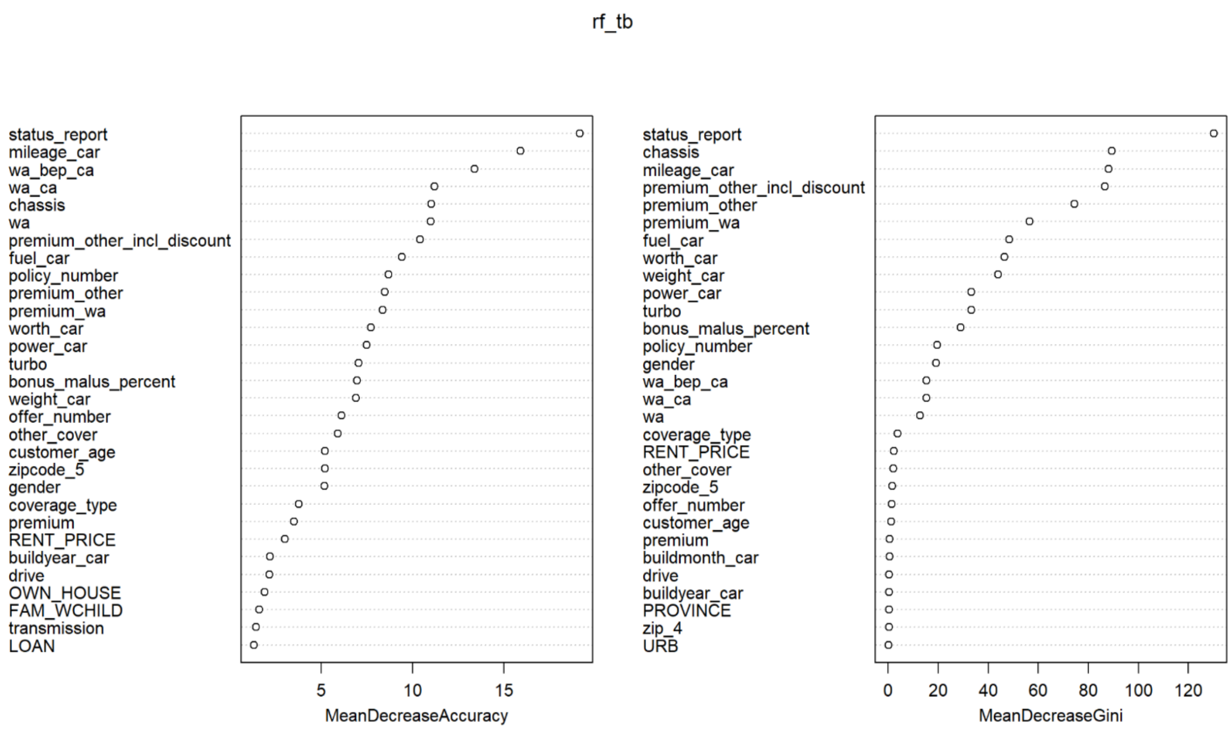
Mean Decrease Gini (right plot)

- Measures how much each variable contributes to reducing node impurity (better splits) in the decision trees.
- Top influencers are similar, reaffirming the importance of vehicle mileage, coverage types (wa_bep_ca, wa_ca), and policy-related variables (premium_other, premium_wa, etc.).

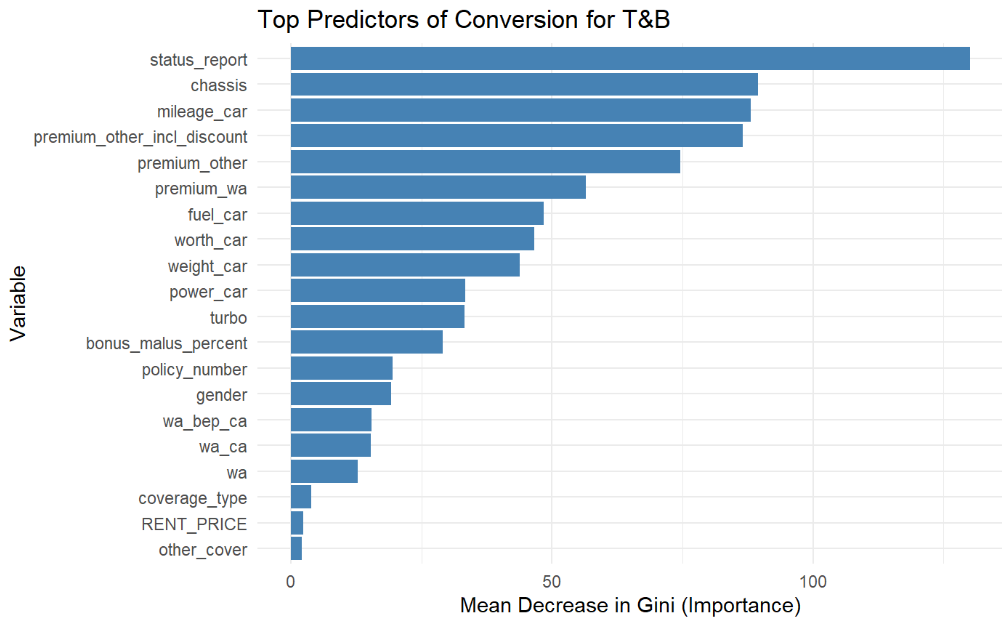
Summary

- Both plots show policy and car-related features dominate in influencing conversion, especially the status_report, car mileage, and the type of coverage (wa, wa_bep_ca, wa_ca).
- This aligns with earlier findings that coverage types and car features critically impact conversion rates, particularly for T&B.

Plot 8: Random forest for T&B



Plot 9: Top Predictors of Conversion for T&B



Interpretation:

Top Important Variables:
Random forest analysis identified the most influential predictors of customer conversion: status_report, chassis, and mileage_car. These variables consistently surfaced across importance metrics, indicating their strong predictive power. Additionally, a set of premium-related features—such as premium_other_incl_discount, premium_other, and premium_wa—were also found to significantly impact conversion likelihood. Car characteristics (fuel_car, worth_car, weight_car, power_car) and demographic information like gender and zipcode_5 further contributed to understanding customer behaviors. These insights provided a solid foundation for developing more refined targeting strategies.

Actionable Strategies for T&B:
To address underperformance, T&B should eliminate or repackage "None" coverage quotes, which were associated with zero conversions, and offer at least basic coverage like WA. Tailoring offerings to align with successful patterns seen in other affinities could raise conversion rates. The most responsive age group, 25–40, should be prioritized in marketing and loyalty campaigns. Pricing strategies for full and bundled plans such as WA_BEP_CA and WA_CA need revision, as T&B's conversion drops significantly for these levels. Launching re-engagement campaigns for leads stalled at the CalculateNewPremium stage and integrating quote calculators personalized by mileage and premium sensitivity could reduce abandonment. Additionally, collapsing high-cardinality variables like policy_number and chassis into meaningful groups may improve performance in segmentation and modeling tasks.

Key Learnings from Insuro & Seguros International:
Insuro and Seguros provide useful benchmarks. Seguros achieved a ~21% conversion rate for full coverage (WA_CA), far outperforming T&B's ~8%, indicating better customer alignment

and perceived value. Insuro succeeded with a more balanced distribution of premium types, notably achieving 34.5% conversion for basic WA coverage. These findings suggest that a well-tailored sales funnel and consistent value communication boost results. Seguros also outperformed T&B in follow-up execution, with fewer drop-offs at Tailoredofferwithdrawn and Tailoredofferrejected stages, implying stronger engagement and offer relevance. T&B could benefit from adopting similar sales process discipline and value framing to improve conversion across all stages.

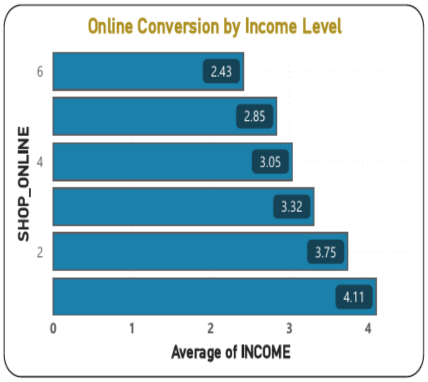
4.2 Which customer demographics (e.g., age, income level, household structure) are most closely associated with a higher likelihood of converting online?

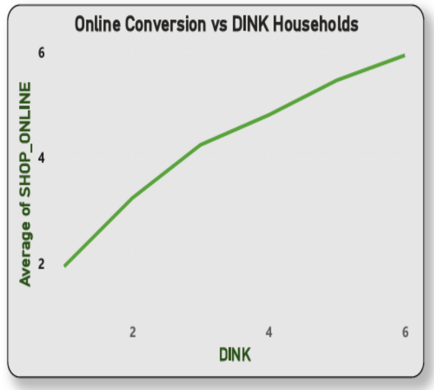
Key Demographics Associated with Higher Online Conversion

Demographic characteristics can have a significant impact on how consumers buy insurance online. Younger adults, specifically those aged 25–34 and 35–44, tend to have the highest conversion rates not just because of their affinity for buying on digital platforms, but also because they are simply more comfortable making online transactions. Income can be another significant characteristic. For instance, customers earning $75K or more are more likely to purchase online as they potentially have more disposable income available and are usually more comfortable buying digitally. Household type also matters because DINK (Dual Income, No Kids) households often have less responsibility with money so they could be more inclined to make a purchase quickly, while families with children tend to have good conversion as well but with less intention to spend, since the decision to buy likely comes from the convenience of shopping from home.

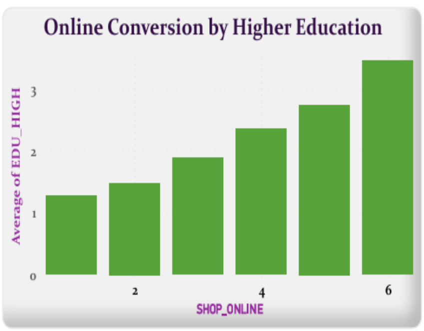
Education and location further influence online buying behavior. People with higher education levels often feel more confident and informed when making decisions online, which contributes to a greater likelihood of completing purchases. Additionally, urban households tend to convert more frequently than rural ones, mostly because they enjoy better internet access and more exposure to digital services. Together, these insights are incredibly useful for tailoring marketing efforts allowing businesses to better target and connect with the audiences most likely to engage and convert online.

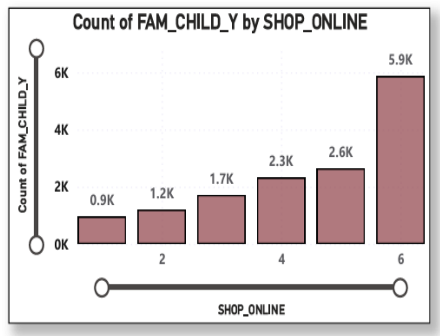
Plot 10~14: Customer Demographic & Conversions





DINK	Average of SHOP_ONLINE
1	1.95
2	3.24
3	4.24
4	4.81
5	5.46
6	5.94
Total	**4.51**





Interpretation:

Income level also affects online conversion rates. Individuals in higher-income groups are more likely to convert therefore, primarily since they possess greater disposable income. This provides them with the choice of making quicker buying decisions with greater confidence. Their propensity to spend online will usually result in quicker adoption of digital channels and services, especially when they perceive additional value or convenience.

DINK households two incomes, no kids also demonstrate a strong and enduring growth in online conversions. With dual income and reduced expenses, these households have more financial flexibility and flexible lifestyles. This is likely to translate into increased willingness to look for and spend on online products and services, particularly those that provide greater convenience or help in their hectic lifestyles.

Education is also an important factor. Conversion rates for online behavior will tend to rise steadily with higher levels of education. More educated individuals are more likely to be digitally literate and at ease with utilizing online resources, which enables them to make more informed and streamlined buying decisions. Their experience with researching and comparing products online also helps to drive higher conversion rates.

These children's households also illustrate strong online conversion behavior, most notably for convenience- and need-driven purchases such as grocery items, back-to-school supplies, and household needs. These families are time-starved, making the convenience of shopping online an effective motivator for managing everyday necessities. Their need component combined with the ease of digital availability results in strong conversion rates by this demographic.

Conclusion & Recommendations:

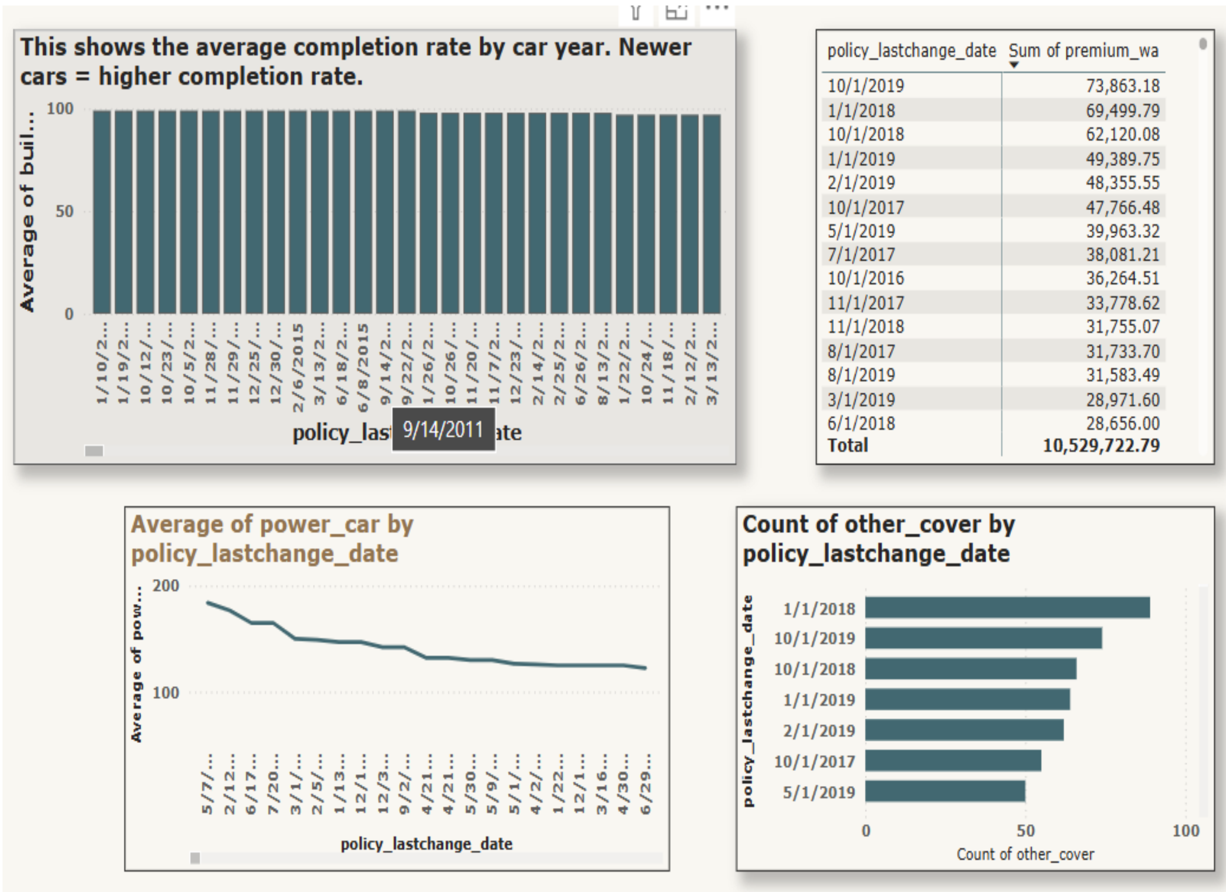
To drive higher online conversion rates, businesses should prioritize younger, affluent, and digitally engaged segments particularly DINK households, families with children, and highly educated individuals. These groups demonstrate a greater propensity for completing online insurance purchases due to their financial flexibility, digital comfort, and demand for convenience.

4.3 How do vehicles or policy attributes such as build year, power, or type of coverage affect the probability of purchase completion?

Vehicle & Policy Attributes Impacting Conversion

Vehicle and policy features play a crucial role in influencing online purchase completion rates, as they directly impact customer perceptions of value, affordability, and trust. By analyzing these attributes—such as vehicle age, engine power, coverage type, and premium size—businesses can gain valuable insights into customer behavior patterns and preferences. Features that convey reliability, protection, or cost-effectiveness tend to build trust and confidence, significantly increasing the likelihood of conversion. Optimizing these key factors through data-driven strategies and user-centric design is essential not only for boosting online sales but also for enhancing the overall customer journey and strengthening competitive positioning in the digital marketplace.

Plot 15~18: POWER BI DASHBOARDS



Interpretation:

Online policy completion rates are notably higher among newer vehicles, suggesting that recent car buyers are more inclined to finalize purchases through digital channels. Lower-powered vehicles tend to convert better, likely due to their lower insurance premiums and overall affordability. There is also a growing preference for full coverage (CAS) policies, indicating that consumers are increasingly valuing comprehensive protection. Additionally, higher premiums are predominantly associated with recent vehicle models, often reflecting newer or high-value cars paired with broader coverage plans. Collectively, factors such as vehicle age, engine power, coverage type, and premium size play a significant role in shaping online conversion behavior.

Conclusion & Recommendations:

Vehicle and policy attributes such as newer car models, moderate engine power, and affordable coverage types (e.g., WA) significantly influence purchase behavior. Customers increasingly favor streamlined offerings that balance value with cost.

4.4 What targeted interventions (e.g., personalization, pricing, messaging) could T&B implement to boost online conversion rates?

1. **Personalization:**

   Vehicle-Specific Recommendations: By suggesting products or services based on the customer's vehicle type, model, and year — like tailored maintenance packages or compatible accessories.

   Driving History-Based Offers: If the customer drives frequently or over long distances, offer service plans or discounts on high-usage parts like tires and oil changes.

   Profile-Based Customization: Use customer data like location, past purchases, and service history to personalize product suggestions and promotional offers.

   Dynamic Web Content: Change homepage banners, product categories, and offers based on the user's profile — like showing regional promotions or vehicle-specific deals.

2. Pricing:

Tailored Pricing Models: Adjust pricing based on customer loyalty, purchase history, and engagement — offering better rates to repeat customers or high-value buyers.

Vehicle Type-Based Pricing: Offer competitive pricing based on the vehicle's make and model — for example, different pricing for economy cars vs. luxury vehicles.

Usage-Driven Discounts: Provide discounts based on driving behavior — like offering service discounts for customers who drive longer distances or require frequent maintenance.

Bundled Offers: Create value packages by bundling related products or services, like discounted oil change and tire rotation packages tailored to the customer's vehicle type.

3. Messaging:

Personalized Service Reminders: Send messages based on vehicle data — like reminders for oil changes, tire replacements, or annual servicing based on driving history.

Driving Behavior Insights: Use data to craft messages like "Based on your driving history, it's time for a tire check!" or "Your [Car Model] is due for maintenance."

Targeted Promotions: Send special offers based on customer behavior — like discounts on parts or services the customer frequently purchases.

A/B Testing on Messaging: Test different subject lines, calls-to-action, and product descriptions to determine which language drives more conversions.

Power BI Workflow for Personalization Strategies.

Personalization: Use bar charts for top product recommendations, maps for regional driving patterns, and funnel charts for conversion rates.

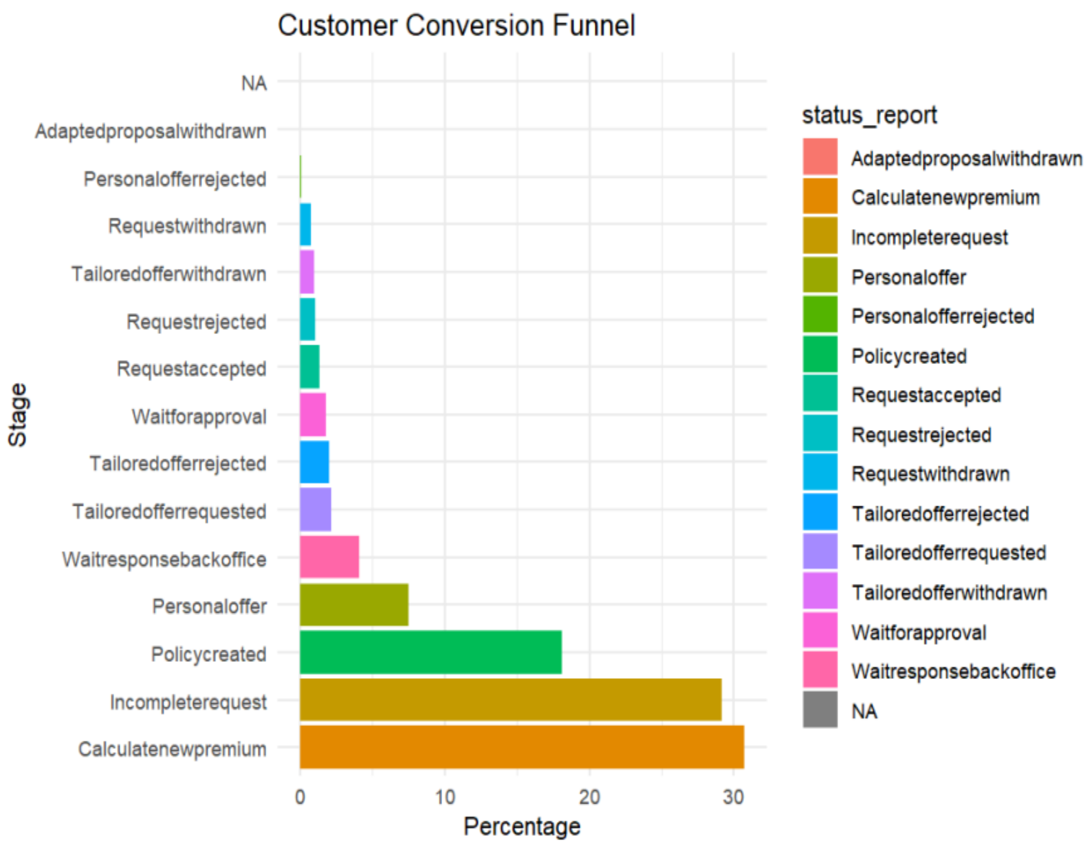
Pricing: Show pricing tiers with bar charts, discount utilization with scatter plots, and bundled offer revenue with stacked bar charts.

Messaging: Track reminder performance with line charts, promotional success with Bar charts, and A/B test results with clustered columns.

Filters: Add slicers for vehicle type, customer segments, and time periods for interactivity.

Dashboards: Create separate dashboards for personalization, pricing, and messaging for clear insights

Plot 19: Customer Conversion Funnel for T&B



Interpretation:

This Customer Conversion Funnel visualization represents the different phases a user follows in their journey while engaging with the system. Age display shows relative percentage measure of users in each step vertically while the steps of a funnel are depicted horizontally,

Key Insights from the Graph:

Critical Drop-Off Areas:

"Calculatenewpremium" and "Incomplete request" leads with 30% and 25% respectively.

This indicates that a huge portion of users compute premium but do not take any other steps or abandon the process halfway.

Potential issue: Users face doubt related to pricing, tough factors, or lack drive to continue.

Successful Conversions

"Policy created" shows significant percentage which means that a portion of users do in fact transform into paying users.

Though remarkable, this value is lower than what is anticipated at the completion stage.

Opportunity: More users are required to encourage movement towards tangible completion.

Rejection and Withdrawal Points:

"Tailored offer withdrawn," "Request withdrawn," and "Personal offer rejected" identify where users drop of the are other clear cut measures of rejection.

Actionable Recommendations to Boost Conversions:

To improve user conversion and engagement, it's crucial to reduce drop-offs at the calculation and request stages by streamlining the premium calculation process—this can be achieved by minimizing form fields or incorporating AI-based suggestions to simplify data input. Enhancing support through live chat assistance can further encourage users to complete their requests, while the use of discounts or time-sensitive offers can prompt quicker action. Addressing rejections and withdrawals involves understanding user concerns through feedback from those who declined offers or abandoned requests, and using this insight to craft tailored outreach, such as personalized discounts or flexible pricing strategies that appeal to a wider audience. Additionally, speeding up approval and response times is essential; refining internal workflows and automating approvals for less complex cases can help alleviate bottlenecks. To maintain user interest during waiting periods, proactive status updates via email or SMS can be employed, ensuring users stay informed and engaged throughout the process.

Final Thoughts:

This graph highlights the specific stages in the conversion journey where customers are most likely to disengage, offering valuable insight into potential drop-off points. By closely examining these moments, T&B can identify where improvements will have the greatest impact. Fine-tuning pricing strategies to ensure competitiveness and clarity can help maintain interest, while implementing more effective personalization tactics—such as tailoring product recommendations or messaging based on user behavior—can enhance relevance and engagement. Streamlining the approval process, particularly for straightforward cases, will reduce delays that often lead to abandonment. Additionally, reducing friction at key phases, such as simplifying forms, improving page load times, and offering real-time assistance, can smooth the overall user experience. Together, these targeted optimizations can significantly boost conversion rates by addressing the precise pain points where customers currently choose to exit.

5. Conclusion, Limitation & Recommendations

5.1 Conclusion

This project investigated the conversion gap faced by T&B in comparison to peer affinities Insuro and Seguros, focusing on identifying key drivers of conversion behavior through data integration and predictive modeling.

T&B exhibited the highest quote volume but the lowest conversion rate—especially in "None" coverage and high-premium tiers (WA_BEP_CA and WA_CA). These insights were reinforced through funnel and coverage-type analysis, which highlighted substantial drop-offs during the quote and tailored offer stages. Moreover, segmentation by age revealed T&B performs best among users aged 25–40, but suffers steep declines after age 45, in contrast to the more stable age performance of Seguros.

To address these issues, multiple modeling approaches were explored. Logistic regression was initially applied but ultimately rejected due to signs of overfitting—such as an AUC of 1 and abnormally low AIC—alongside zero statistical significance across all coefficients. These anomalies indicated likely data leakage (e.g., post-outcome variables like `status_report`) and instability caused by multicollinearity and high-cardinality categorical variables. In contrast, random forest modeling was adopted as a more robust, non-parametric alternative. It successfully handled non-linear interactions and multicollinearity and provided interpretable variable importance rankings. Top predictors included `status_report`, `chassis`, `mileage_car`, `premium_other_incl_discount`, and `worth_car`, emphasizing the role of pricing, vehicle traits, and offer progression status in conversion behavior.

Unsupervised clustering using k-means (with optimal k=3) further revealed three customer segments: one with high mileage and premiums, another price-sensitive group with low coverage and fuel cost, and a third with moderate values across all features. These insights can be used to tailor product strategies and communication for each customer group.

5.2 Limitations

Several constraints affect this study's generalizability and interpretability. First, high-cardinality variables (e.g., policy_number, place_residence) were excluded from modeling due to random forest's limitation with too many factor levels, potentially omitting geographic or behavioral nuances. Second, missing values required imputation, which, while managed via random forest, may introduce modeling uncertainty. Third, limited behavioral tracking (e.g., no session length or click data) restricts insights into customer hesitation during the quote funnel. Finally, the analysis relied on observational data without A/B testing, limiting causal inference.

5.3 Recommendations

To improve conversion rates, T&B should (1) repackage or eliminate "None" coverage quotes by defaulting to a basic plan; (2) optimize offers for high-converting groups (ages 25–40); (3) redesign pricing for WA_BEP_CA and WA_CA packages, emphasizing value rather than price; (4) re-engage "CalculateNewPremium" leads through email or chatbot campaigns; and (5) streamline product and funnel design using top predictors identified via random forest. For long-term improvements, collecting richer behavioral data and enabling controlled experiments will support more accurate, actionable modeling.

References

1. Kotler, P., & Keller, K. (2016). *Marketing Management* (15th ed.). Pearson.
2. Kumar, V., & Reinartz, W. (2018). *Customer Relationship Management: Concept, Strategy, and Tools* (3rd ed.). Springer.
3. Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29–39.
4. AI Revolutionizes Auto Insurance Via Real-Time Data
5. ChatGPT: https://chatgpt.com/c/67b7e962-6950-800a-89b3-6ef8fa95a5bf
6. https://www.statista.com/topics/3087/car-insurance-in-the-united-states/#topicOverview
7. Business Analytics and Car Insurance Mediation: a data driven model proposal（path: UNH-my charger-database A_Z-IEEE database）
8. Professor Ahmet Ozkul’ s module materials