Analysis Summary: Predicting Automobile Insurance Claims

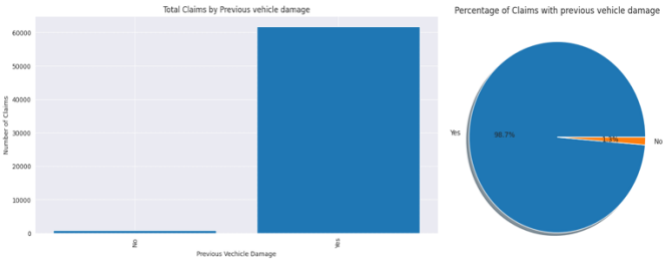
1. Objective

Northbridge sell automobile insurance policies where customers pay them an annual premium to insure their vehicles. The goal of this analysis is to understand the factors influencing whether a customer will submit an insurance claim. We will try to identify how various demographic and other related factors affect the likelihood of a customer being involved in an accident and subsequently filing a claim.

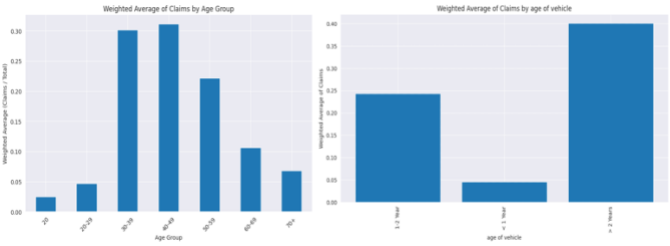
2. Key Findings:

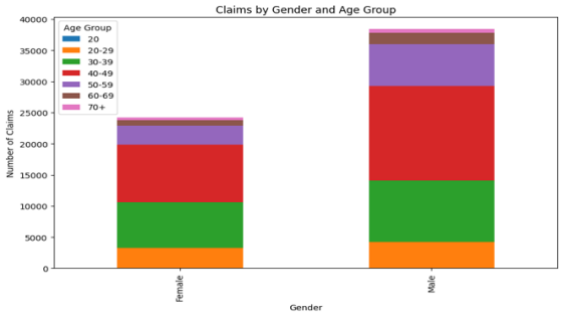
Customer Demographics & Behavior:

- Customers who have a **previous vehicle damage** are more likely to submit a claim, 99.8% of people who Made claims were having previous vehicle damage, suggesting that past incidents influence future claims behavior.



- The **age of the customer** and **vehicle age** are also important factors, with older customers and vehicles showing more claim patterns or met with more accidents than other demographic.



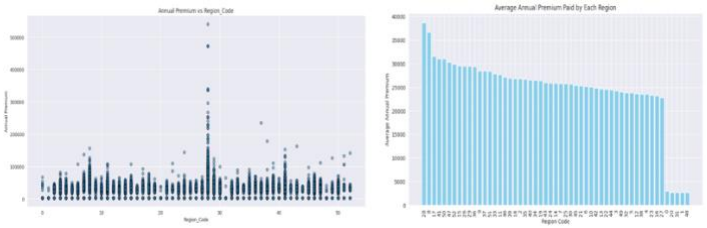


- **Previously insured customers** with Northbridge have a higher likelihood of submitting a claim, possibly due to familiarity with the claims process.

Previously Insured	Claim	No Claim
No	132745	62466
Yes	186808	135

Region & Annual premium:

- Customers from some regions were made to pay higher annual premium suggesting the risk of accidents is higher in these regions.



3. Model Insights:

Fig : Shap Plots for XGBoost Model

Feature Influence in making predictions: Previous Vehicle Damage, **Previously Insured** are the most influential features. **Policy Sales Channel**, and **Region Code** have moderate impacts, which may imply that certain sales channels or regional factors could influence driving behavior or risk exposure. **Vehicle Age** also tends to influence predictions greatly as older vehicles might lead to a higher chance of claims due to their potential for mechanical issues. **Age**, **Gender**, and **Annual Premium** show relatively lower influence on the prediction but might still provide some insight when combined with other factors.

4. Recommendations:

Find New Features: Exploring additional external features, such as historical claim patterns, driving history or geographical risk data to improve the model's ability to capture more nuanced behaviours.

Improve Class Balance: The no claim (16.4%) class is underrepresented which could create a high bias towards the overrepresented class (83.6%). We can either enhance the data collection or apply advanced sampling techniques like SMOTE or use weighted loss functions or oversampling of underrepresented class to improve model performance for underrepresented class.