Motor Insurance Renewal Prediction

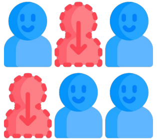
Shilpa

Business Scenario

- A global insurance firm is facing minor decline in its motor insurance portfolio sales due to heavy competition in different regions. So, the company wants to take all measures to retain its existing customer base.
- The Head of Marketing wants to know which customers are more likely to renew their existing policy and which ones who are likely to churn. The team has planned lucrative offers for both group of customers.

In order to plan and target specific groups of customers, the company is relying on us - <u>the data science team</u> to help predict the behavior of customers in advance.

Understanding Area of Focus



- ❑ Understanding if a customer might churn (without renewing) based on what we can glean about them during their lifetime value, could give us a leading indication to retain them before a churn could actually happen

- ❑ With such early indication, we could channel the budgeted lucrative discounts/offers to proactively engage with highest-risk-of-churning customers
    - ❑ *Example, LinkedIn sends discounts to continue being a paid Premium member when free trials are close to ending which drives retainment of premium members*

***Measure** of success: **Reduce** customer **churn** → **Increase** lifetime value → **Increase** revenue (cost of retaining existing customer is lesser than acquiring new!)*

Applying Data Science: Predicting Customer Renewal

- Our predictive model should be able to:
  - Learn by relating historic policy holders attributes information with their policy renewal status
  - Generalize well enough to predict for new insurance sign-ups to target actions for churn prevention

Features

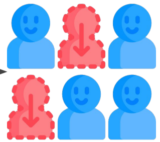
Descriptive attributes of historic policy holders

Response (Prediction Target)

Whether or not the historic policy holders renewed

Machine Learning Model (training)
Task: Binary classification
- renewed or churned

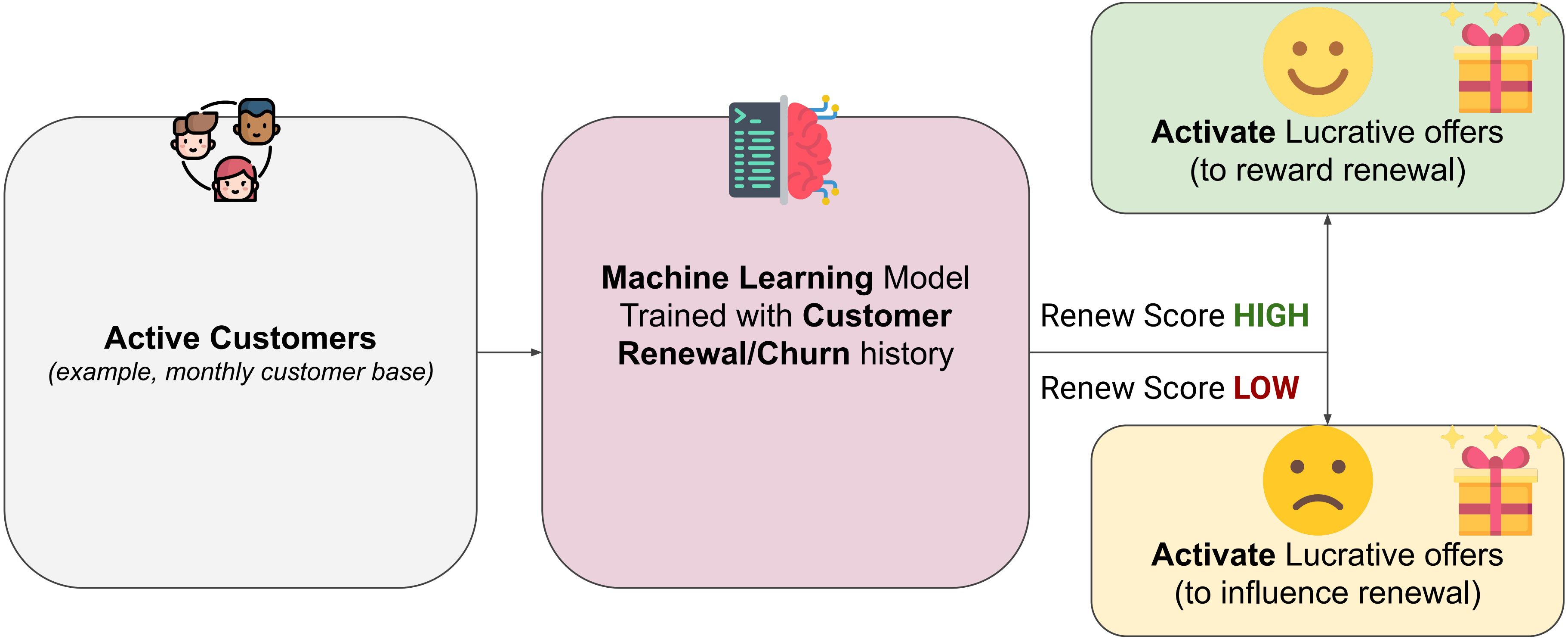
New customer attributes (testing)



Model prediction: churn or renew

Modeling Framework: Predicting Customer Policy Renewal

Renewal Model: Business Application



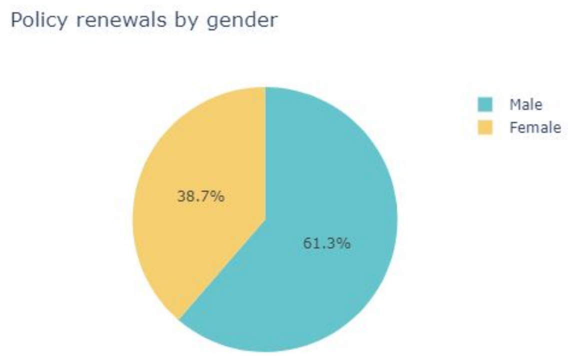
Data Understanding

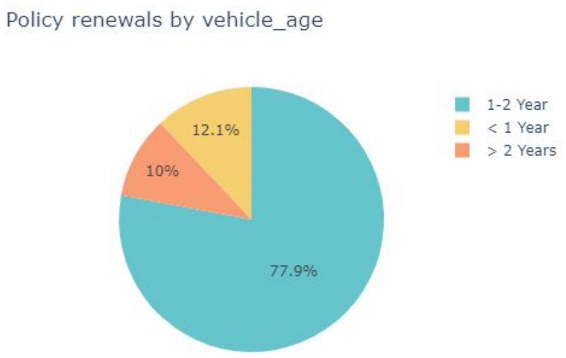
Description

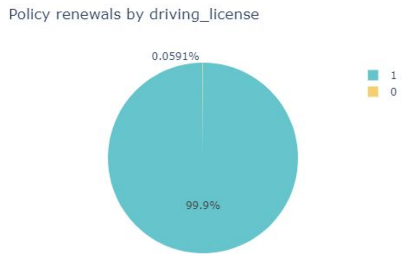
Field Name	Description	Machine Learning: Feature/Response
1. cust_id	Unique customer id	feature
2. gender	customer gender	feature
3. age	Current age of the customer	feature
4. driving_license	Whether the policy holder , holds a driving license or not	feature
5. region_code	The company is a global firm. Hence has business set up across different regions	feature
6. previously_insured	whether the customer has bought some policy previously with the company	feature
7. vehicle_age	Age of the vehicle	feature
8. vehicle_damage	Has vehicle got into any accident / damage before	feature
9. annual_premium($)	Annual vehicle insurance premium	feature
10. policy_sales_channel	Channel from which customer bought the policy. The company sells via online , offline and many third party brokers	feature
11. days_since_insured	days elapsed since purchase of policy	feature
12. Response	whether the customer renewed (1) or no (0)	response

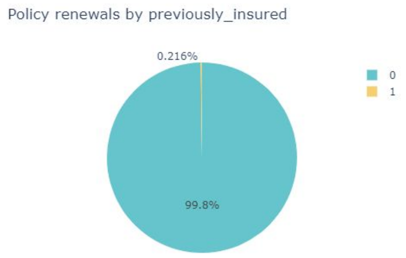
Exploratory Data Analysis

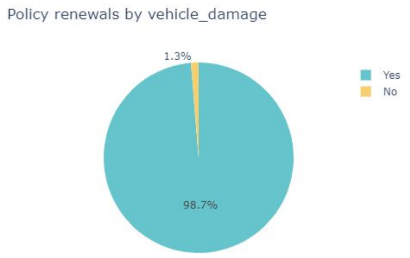
Renewals by Discrete Attributes







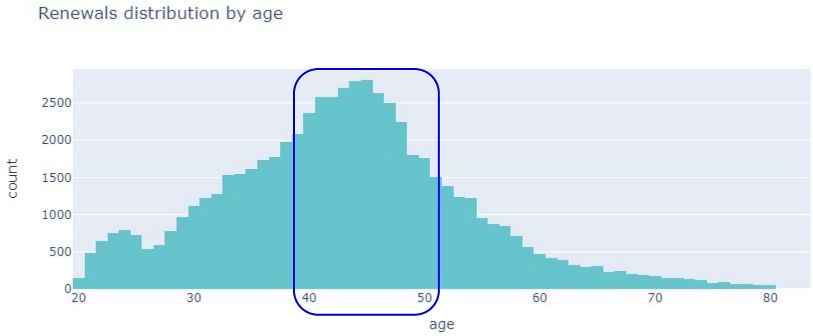


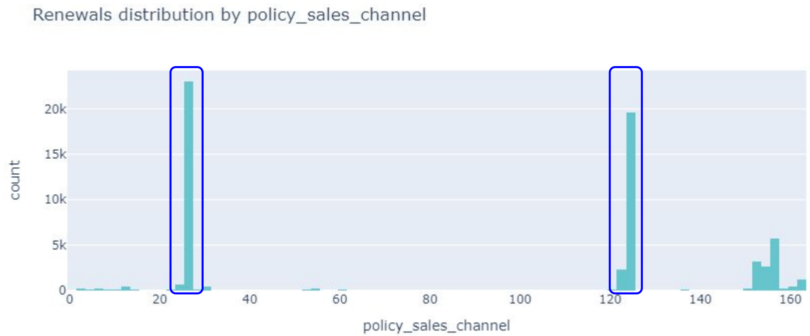


Zooming into the customer profiles who historically renewed our policy, here are some descriptive insights by **discrete** attributes:

- **~60%** customers that renewed were **Males**
- **~78%** renewals happened within the **1st 2 years of a vehicle's age**
- Almost all renewals were from vehicle owners holding a driving license, that weren't previously insured - that is, they are new customers, and have history of prior vehicle accident/damage

Renewals by Continuous Attributes



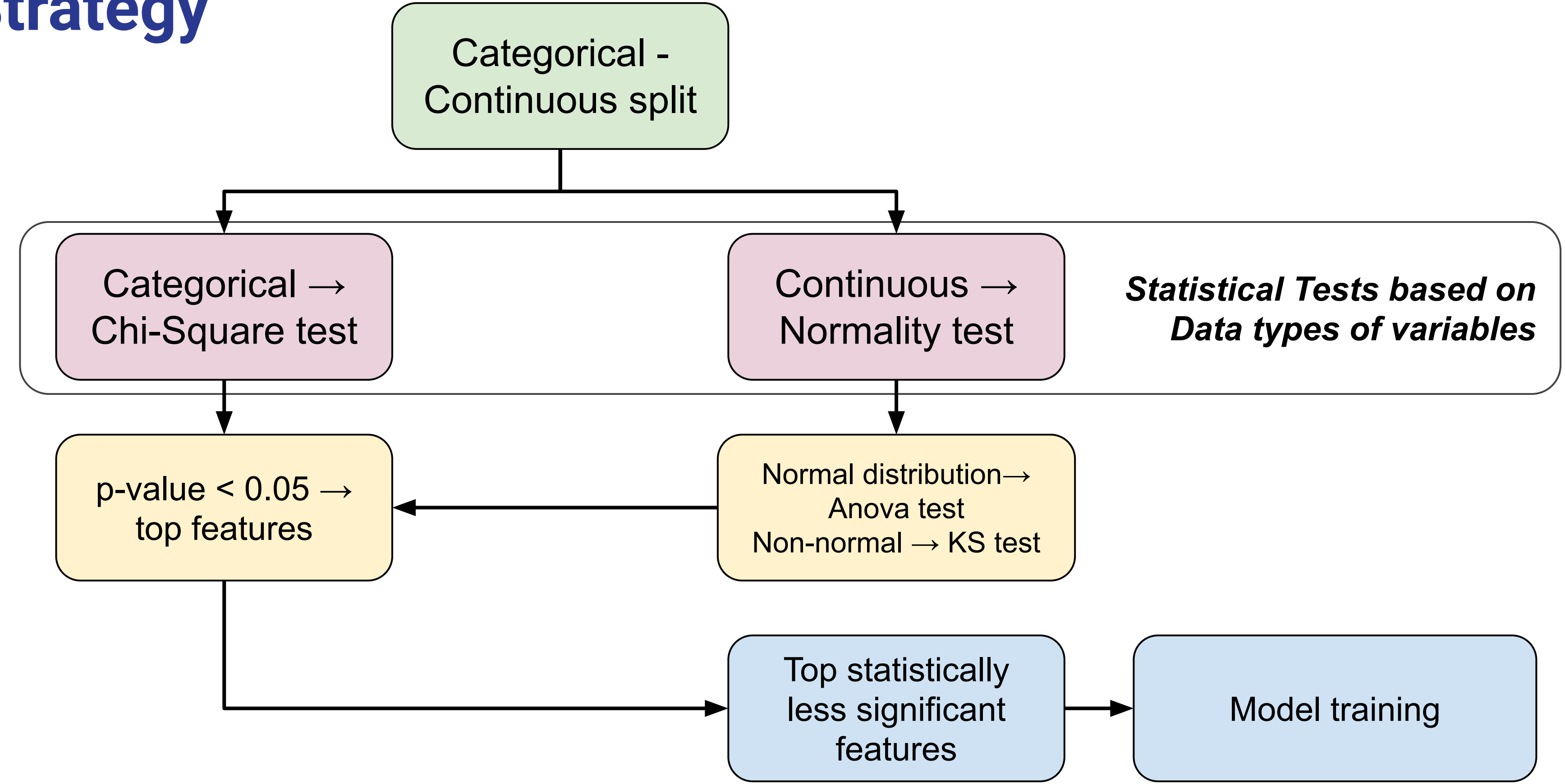


Zooming into the customer profiles who historically renewed our policy, here are some descriptive insights by **continuous** attributes:

- Renewals stream came mainly from customers aged **~40-50** years
- Region code **28** generated most policy renewals
- Sales channels **26-27**, followed by **124-125** generated most policy renewals

Features Selection

Strategy


Categorical - Continuous split
Categorical → Chi-Square test
Continuous → Normality test
Statistical Tests based on Data types of variables
p-value < 0.05 → top features
Normal distribution→ Anova test
Non-normal → KS test
Top statistically less significant features
Model training

Results

Discrete features with most statistically significant impact on “Renewal”

	p_value
vehicle_age	0.000000e+00
vehicle_damage	0.000000e+00
previously_insured	0.000000e+00
gender	7.879733e-174
driving_license	7.216373e-01

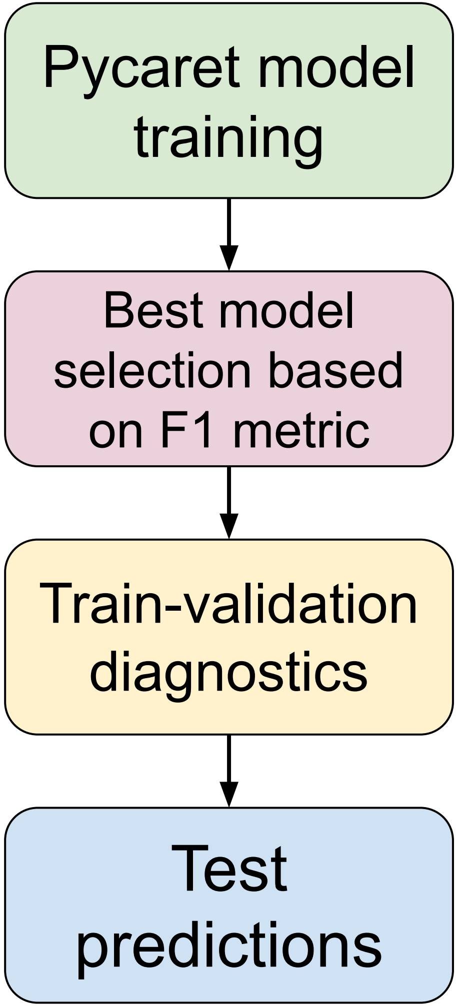
< 0.05

Continuous features with most statistically significant impact on “Renewal”

	p_value
age	0.000000e+00
region_code	0.000000e+00
policy_sales_channel	0.000000e+00
annual_premium	2.259331e-236
days_since_insured	4.118022e-01

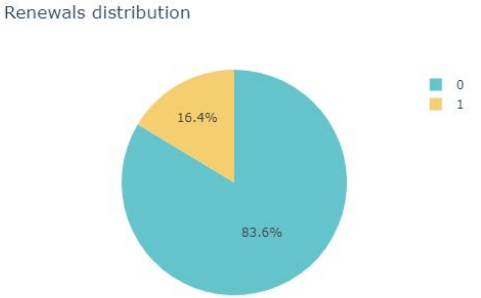
- Generally, all features except “driving_license” and “days_since_insured” had a statistically significant impact on policy renewals.
    - This means we could potentially drop these variables as features for model training as it does not have any statistical impact on our response variable
- Additionally, no multicollinearity was observed between features (that is, none of them correlated to each other)

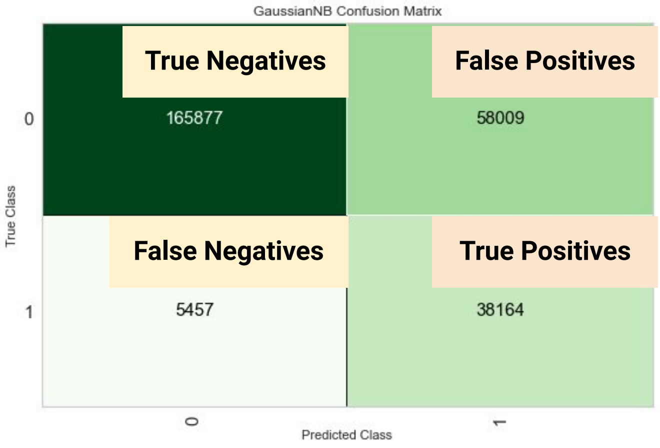
Modeling



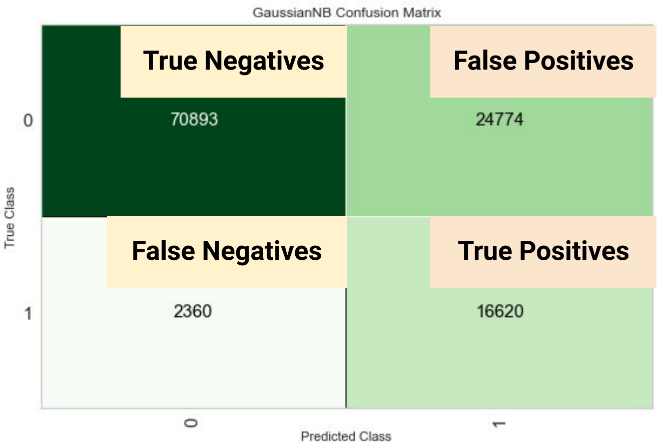
Results

- **F1 score metric** (balance between Precision and Recall) is used as target model evaluation metric, given the imbalanced distribution of the response variable
- Best model - Gaussian Naive Bayes, performed at **~55%** F1 score on both train validation datasets (balancing ~40% Precision and ~87% Recall calculated from confusion matrices below)



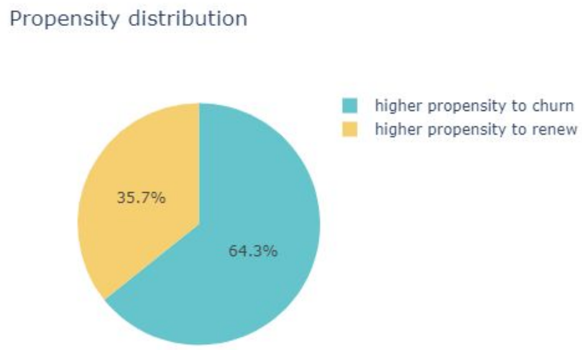


Training

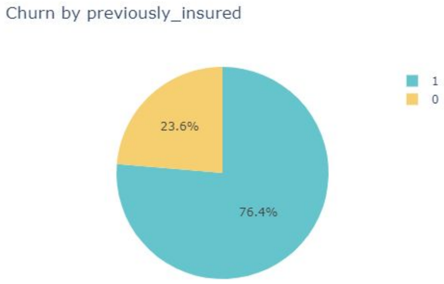


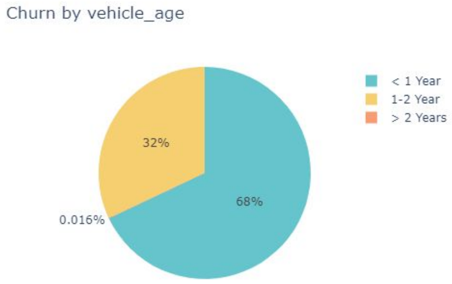
Validation

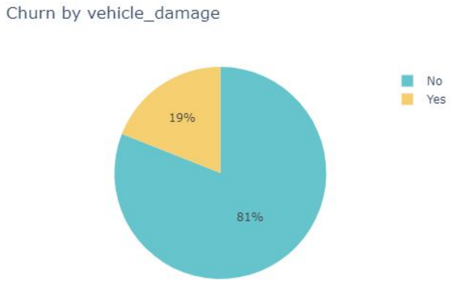
Prediction Insights: Churns

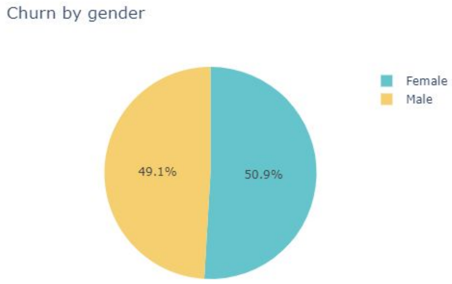


- Out of the 78,273 existing policy holders to be scored:
  - **~64%** were predicted by our machine learning model to churn, while **~36%** likely to renew *(assuming a propensity cut-off of 50%)*
- Given the higher weightage on churns, we could target the following customer population with **higher propensity to churn** based on descriptive insights captured below:
  - Customers who have previously bought policy with us, owning brand new vehicle within its 1st year, with no vehicle damage/accident history
  - No major toggle by Gender, thus both Males and Females can be targeted equally

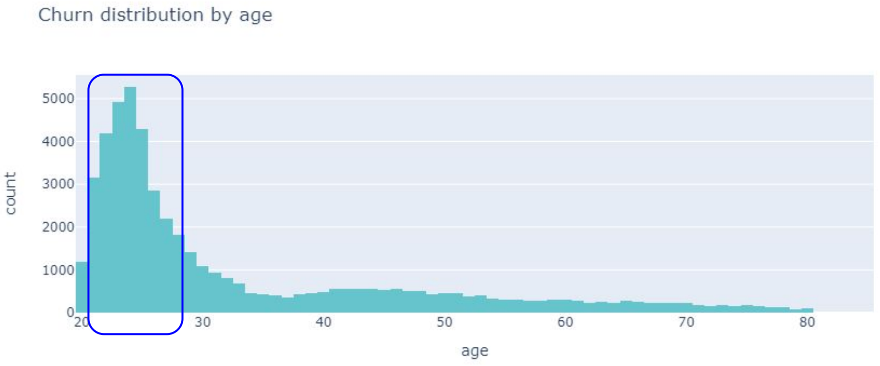


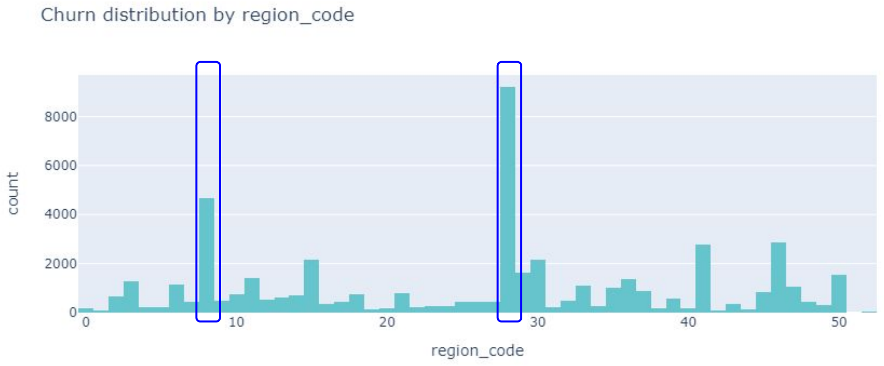


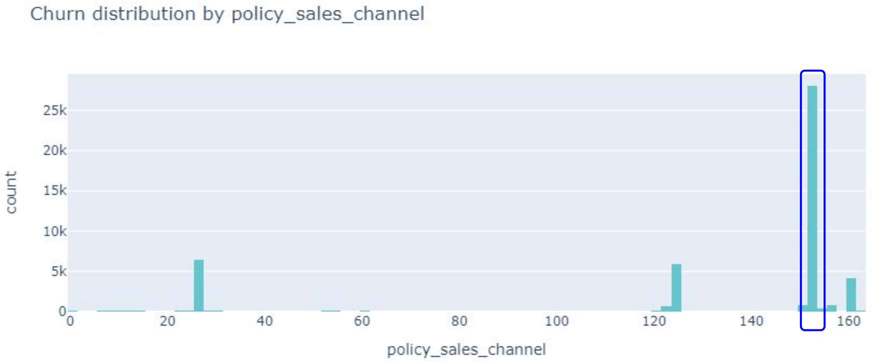




Prediction Insights: Churns







- ❏ Additionally, we could narrow our target customers for churn prevention being those in the age groups of >20 - <30 years, from region codes 28 and 8, as well as sales from channels 152 - 153

Inferences

Recommendations & Future Work

Recommendations

- ~64% of existing policy holders were predicted to have high propensity to churn by our machine learning model based on historic policy holders churn/renewal understanding → this signals immediate prioritization to salvage potential churns
- We could narrow to these customer profiles for churn prevention:
    - Customers who have previously bought policy with us, owning brand new vehicle within its 1st year, with no vehicle damage/accident history, those in the age groups of >20 - <30 years, from region codes 28 and 8, as well as sales from channels 152 - 153

Future Work

- Improve False Positives on model to improve Precision
- NLP sentiment, topic prediction on churning customer text feedback if available from exit surveys

References

- Icons: https://www.flaticon.com/
- Link to code notebook: https://nbviewer.org/github/shilpaleo/insurance_renewal_prediction/blob/main/notebooks/customer_policy_renewals_prediction.ipynb
- Auto ML Package built on Scikit learn foundations: Pycaret

Thank you!

Photo by Mateusz Dach:
https://www.pexels.com/photo/man-in-black-suit-inside-car-1135379/