Vehicle Insurance Interest Prediction
(using Data Analytics & Machine Learning)

A Hackathon Event

Aravindh C

Problem Statement

Our client, an insurance company, seeks to develop a **predictive model** to identify whether past health insurance policyholders are likely to express **interest in purchasing vehicle insurance**. This insight will enhance targeted marketing strategies and improve customer engagement.

Features

Variable	Definition
id	Unique ID for the customer
Gender	Gender of the customer
Age	Age of the customer
Driving_License	0 : Customer does not have DL, 1 : Customer already has DL
Region_Code	Unique code for the region of the customer
Previously_Insured	1 : Customer already has Vehicle Insurance, 0 : Customer doesn't have Vehicle Insurance
Vehicle_Age	Age of the Vehicle
Vehicle_Damage	1 : Customer got his/her vehicle damaged in the past. 0 :
Annual_Premium	The amount customer needs to pay as premium in the year
Policy_Sales_Channel	Anonymised Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.
Vintage	Number of Days, Customer has been associated with the company
Response	1 : Customer is interested, 0 : Customer is not interested

Data Overview

Data Features

Independent Features : 11

Dependent Feature : 1

Imbalanced Dataset

87% train data inclined towards Non-Interested Customers (Response=0).

Data Type

Numerical Features : 8

Categorical Features : 3

Missing Value

No missing value.

Statistical Analysis

	count	mean	std	min	25%	50%	75%	max
Annual_Premium	381109.0	30564.389581	17213.155057	2630.0	24405.0	31669.0	39400.0	540165.0

1. Data Variability in Premium :

A standard deviation of 17,213 indicates significant variability in the dataset, suggesting the presence of outliers. This diversity reflects different customer risk profiles.

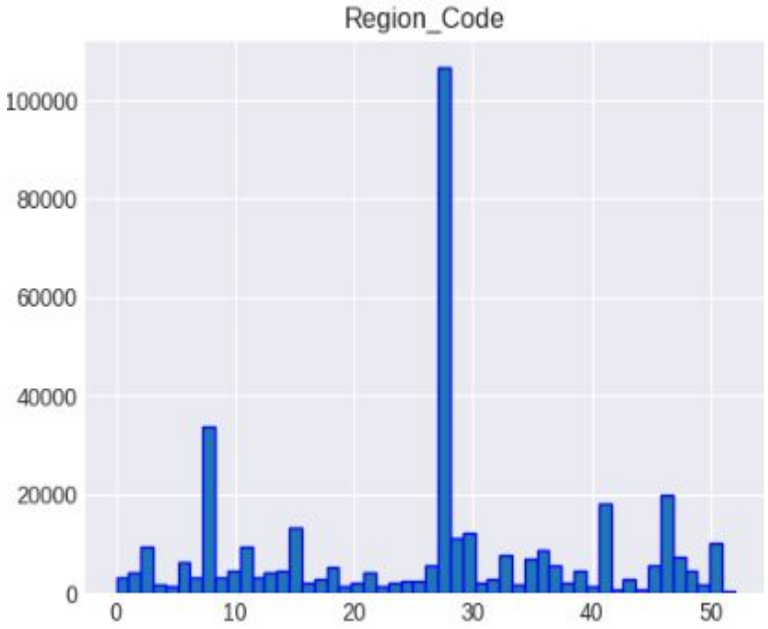
	count	mean	std	min	25%	50%	75%	max
Age	381109.0	38.822584	15.511611	20.0	25.0	36.0	49.0	85.0

2. Customer Age Insights

The average customer age is 38, with the highest average in the 21-27 age group, followed by 41-50. Insurance uptake declines after age 50, likely due to higher premiums and considerations of average life expectancy.

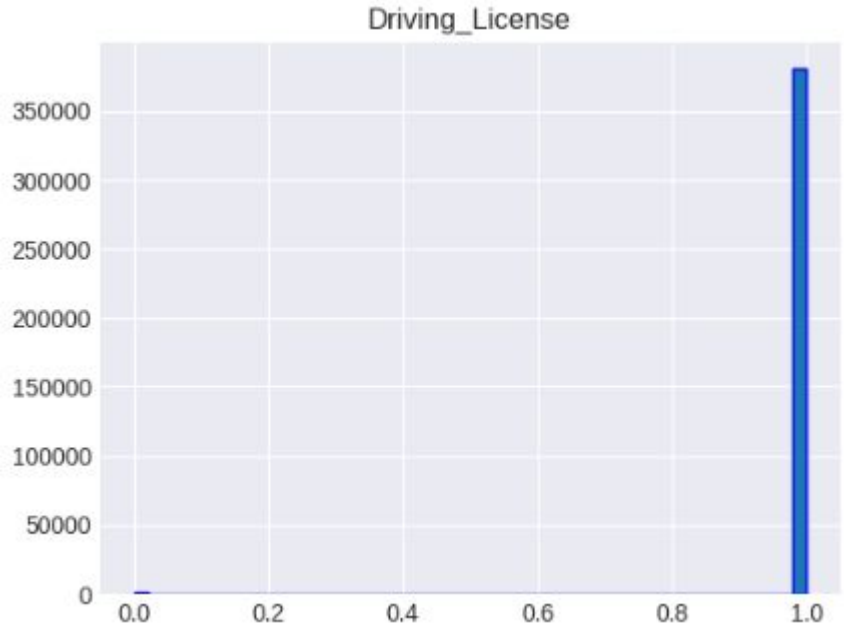
Feature Analysis

Regional Insights



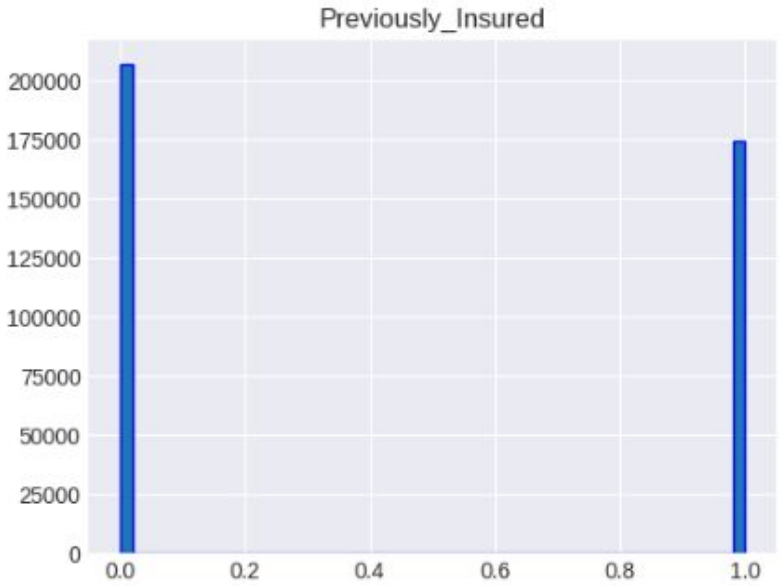
Region_28 has the **highest** customer base at 130,000, indicating potential **risk factors**. This presents a significant opportunity for selling vehicle insurance, as targeting this region could enhance market penetration and address specific customer needs.

Driving License Ownership



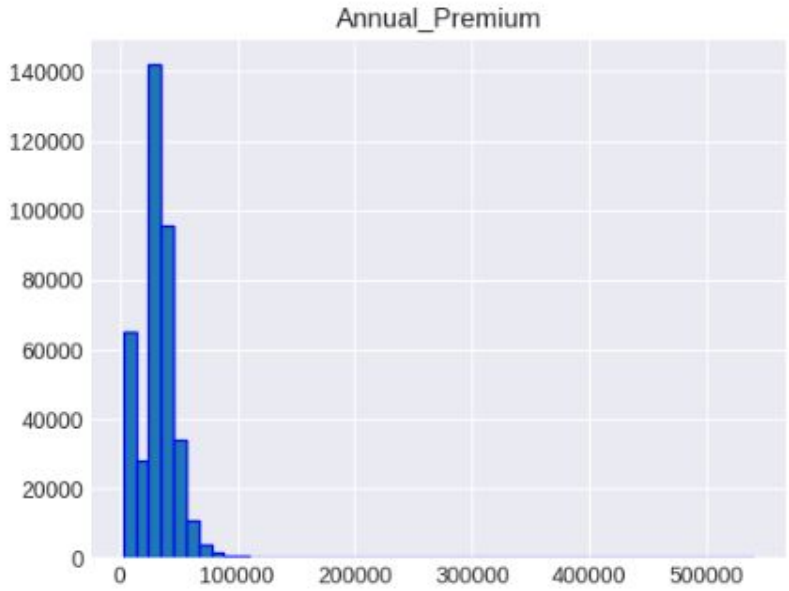
Nearly **99.7%** of customers possess a driver's license, presenting a strong opportunity for vehicle insurance sales. This high ownership rate indicates a ready market for insurance products tailored to **licensed drivers**, enhancing potential customer engagement and revenue growth.

Customer Base Overview



54% of the customer base comprises new clients, reflecting a **strong acquisition strategy and growth potential**. However, **retaining** the existing 46% is equally important. Balancing acquisition and retention efforts will be vital for sustaining long-term success and profitability.

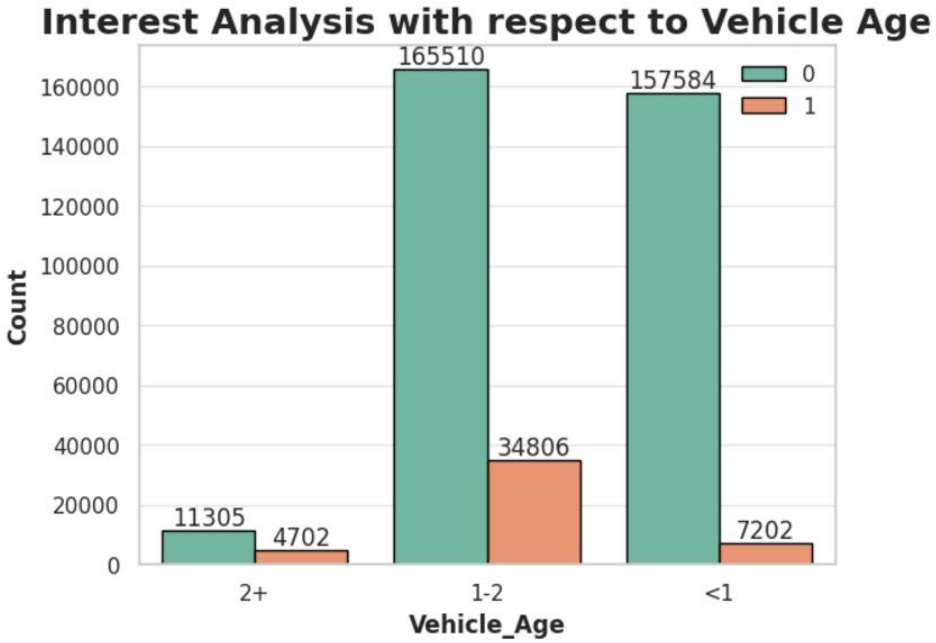
Annual Premium Insights



17% of the customer base pays the lowest premium of 2,630, with a high standard deviation. If these satisfied low-premium customers are engaged effectively, there are significant opportunities to cross-sell vehicle insurance, enhancing overall revenue potential.

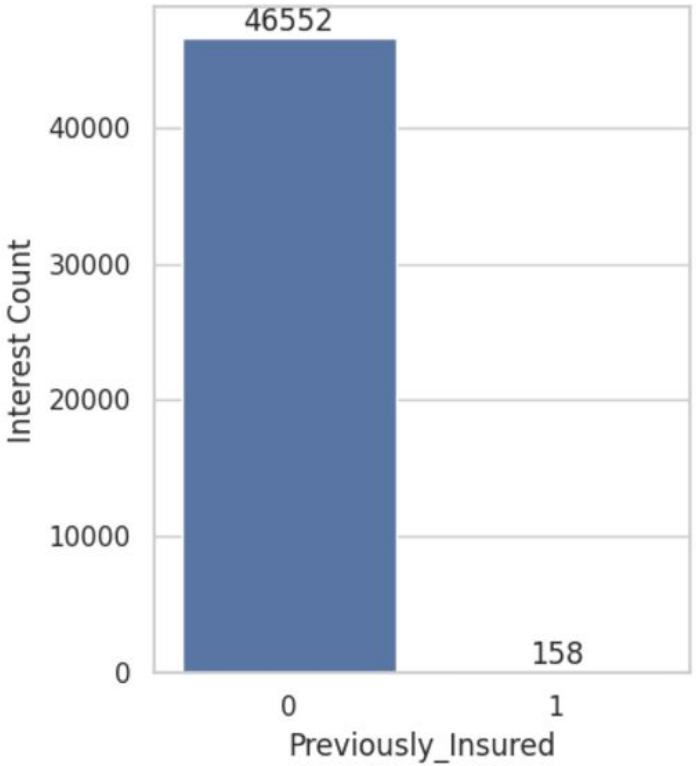
Impact of Features on Target

Vehicle Age Impact



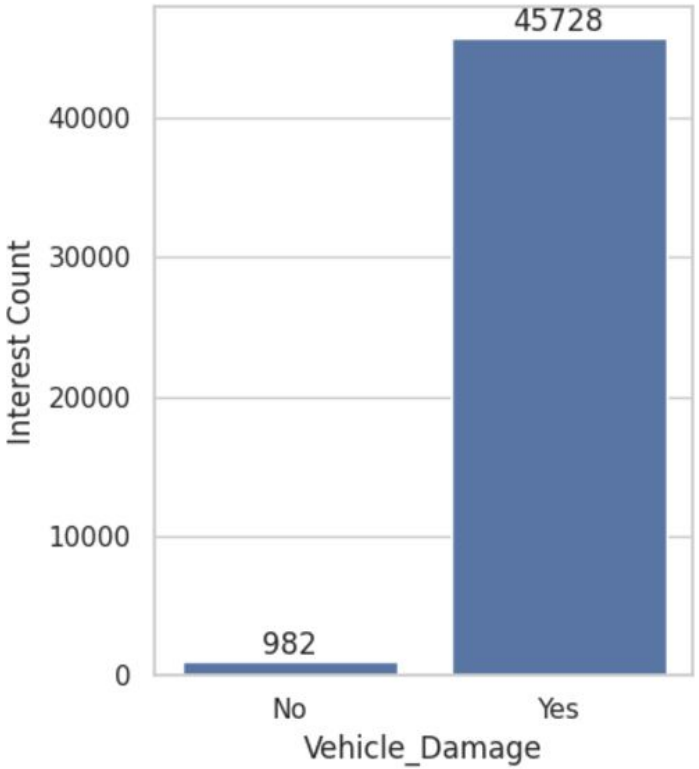
Customers show **interest** in vehicle insurance primarily for vehicles **aged two years or less**. Beyond this age, the likelihood of purchasing vehicle insurance from our company significantly decreases, highlighting the importance of targeting newer vehicles in marketing efforts.

Previous Insurance Impact



An impressive 99.6% of customers **without previous vehicle insurance** have expressed **interest** in purchasing it now. This indicates a strong market opportunity to convert these potential buyers into policyholders, enhancing overall sales prospects.

Past Vehicle Damage Impact



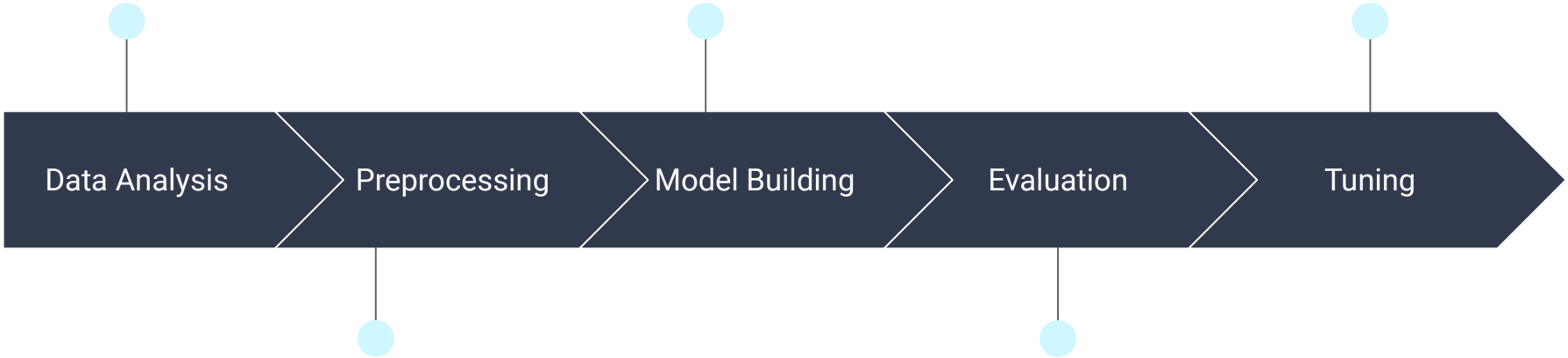
A remarkable **97.8%** of customers with a **history of vehicle damage** have expressed interest in purchasing vehicle insurance. This presents a significant opportunity to target this segment effectively and enhance insurance sales.

Model Flow:

Data understanding and Insight Interpretation

Model building with different parameters and estimators.

Tune and re-build model based on evaluation.



Data preparation by handling missing values, Outliers, Scale & Encoding.

Evaluate model with different metrics

Technical Approaches

Models Applied :

- LogisticRegression
- RandomForestClassifier
- DecisionTreeClassifier
- StackingClassifier
- XGBClassifier

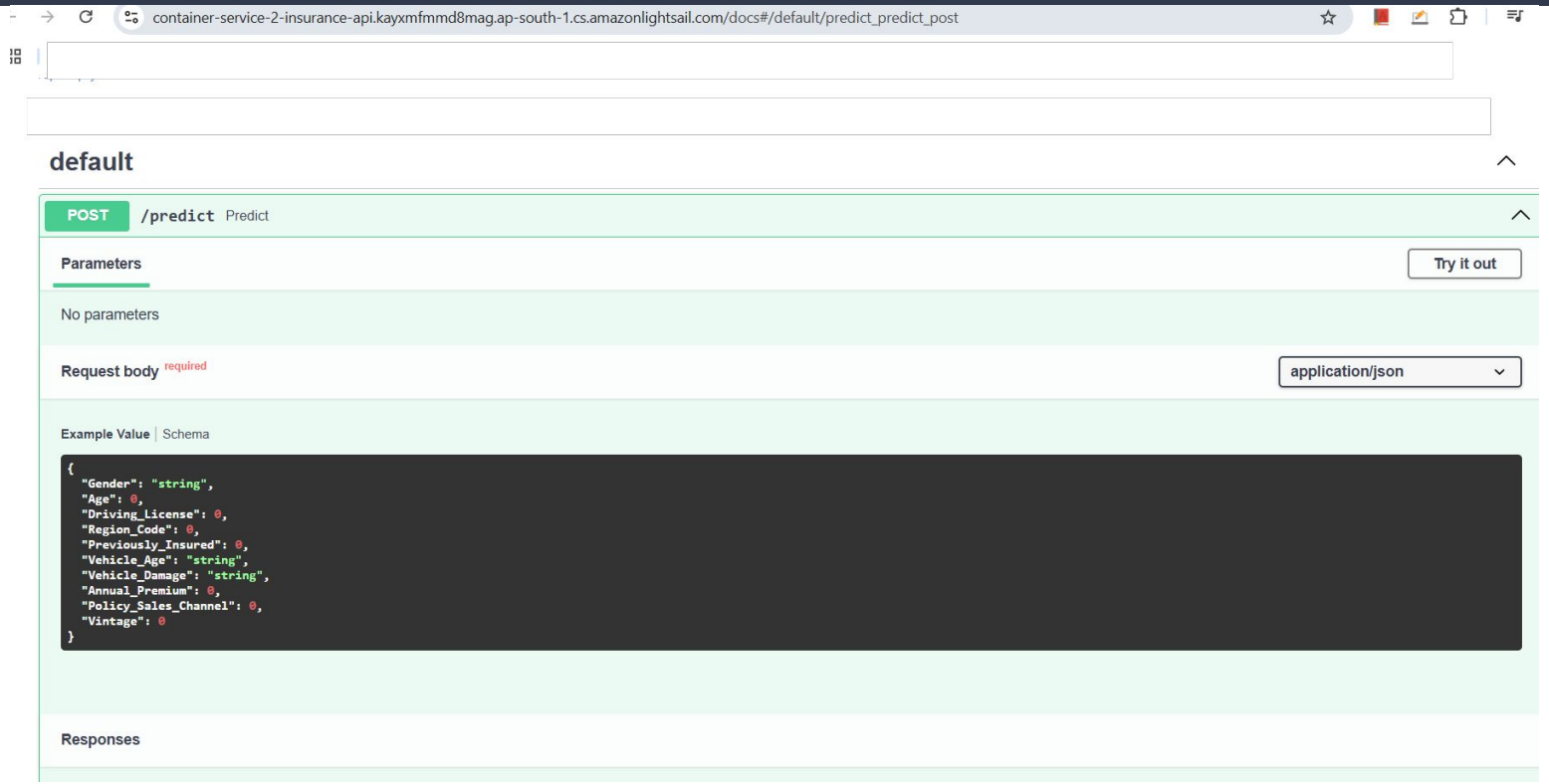
Transformation Applied :

- Robust Scalar
- Simple Imputer
- One-hot encoder

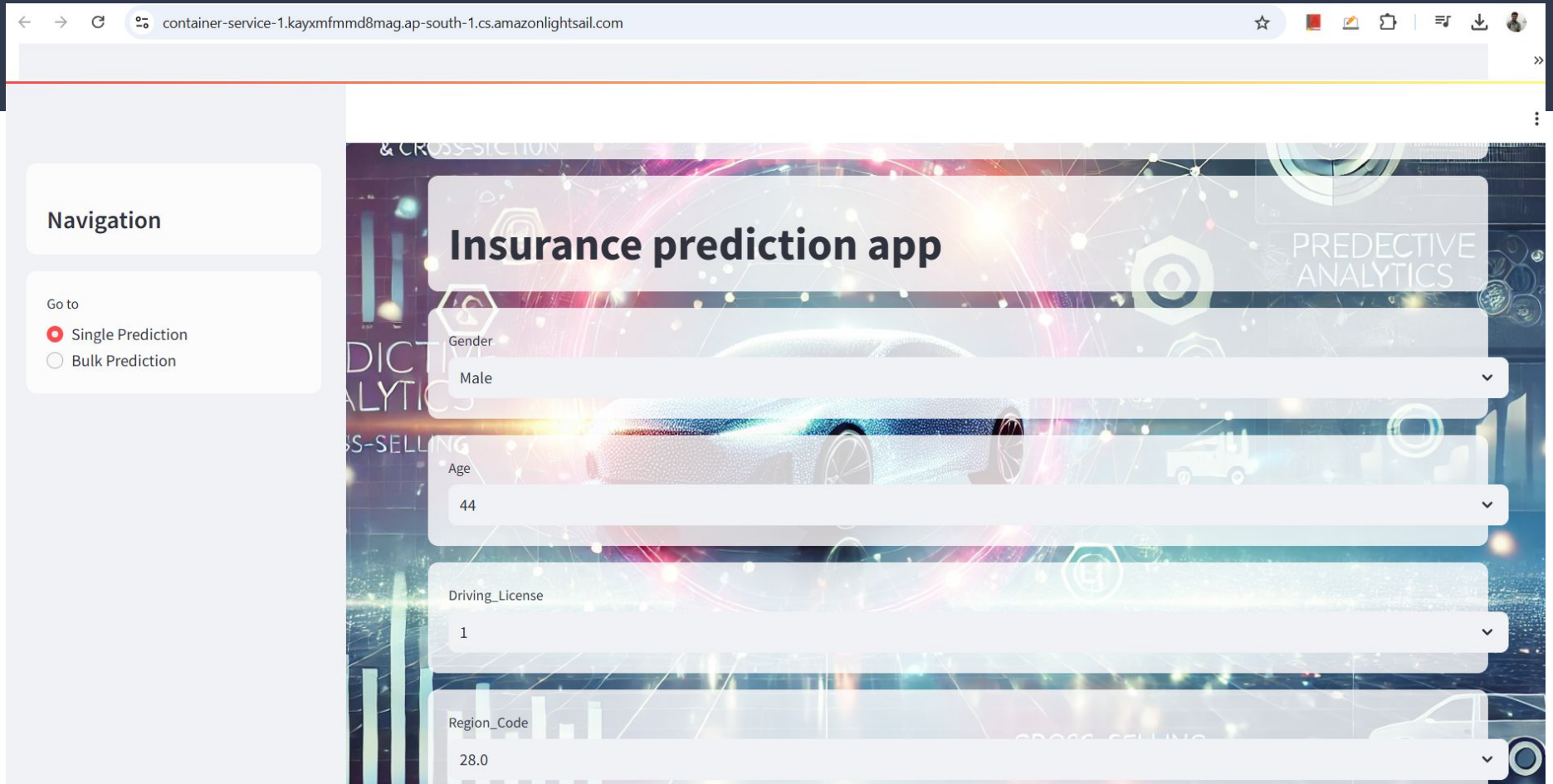
Top 5 Model Performances

PARAMS	Mean_Test_Score
{'model': XGBClassifier(),'model__learning_rate': 0.2, 'model__max_depth': 5, 'model__n_estimators': 200, 'model__subsample': 0.8}	0.8681675872
{'model': XGBClassifier(), 'model__learning_rate': 0.2, 'model__max_depth': 5, 'model__n_estimators': 200, 'model__subsample': 1.0}	0.8678343443
{'model': XGBClassifier(), 'model__learning_rate': 0.1, 'model__max_depth': 5, 'model__n_estimators': 200, 'model__subsample': 0.8}	0.8636026852
{'model': XGBClassifier(), 'model__learning_rate': 0.2, 'model__max_depth': 4, 'model__n_estimators': 200, 'model__subsample': 0.8}	0.8630350163
{'model': XGBClassifier(), 'model__learning_rate': 0.2, 'model__max_depth': 5, 'model__n_estimators': 100, 'model__subsample': 0.8}	0.8629874829

Application Program Interface (API)



Web Application Interface :



Q & A ... ??