

## INTRODUCTION

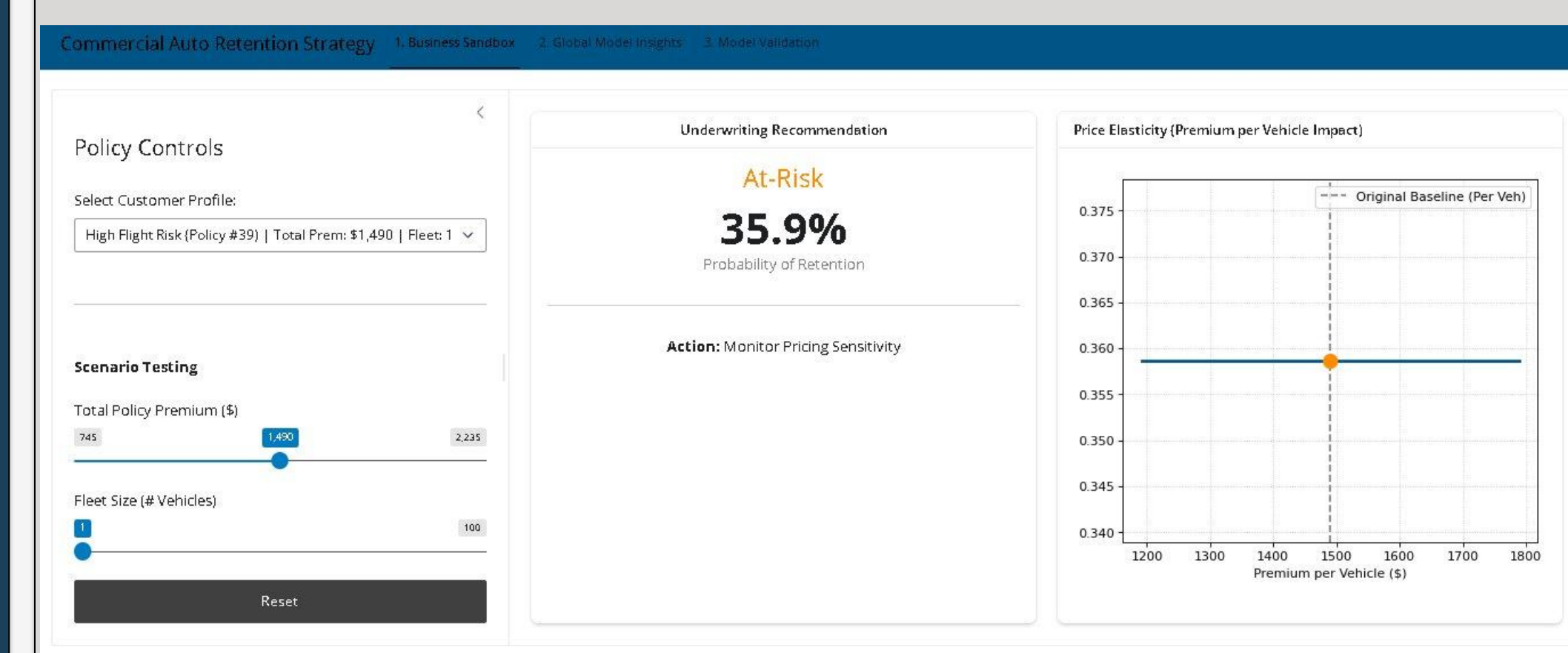
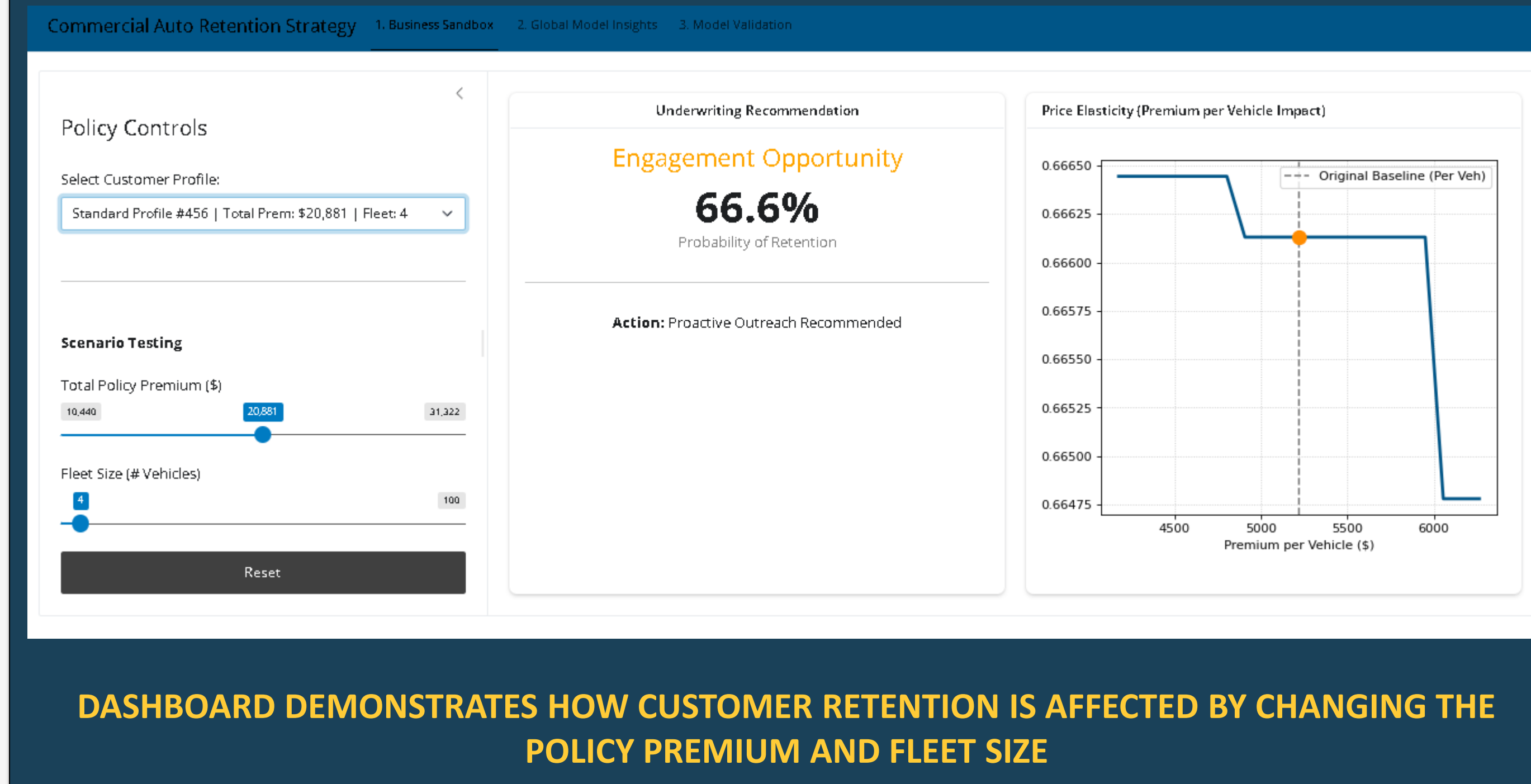
### The Core Problem

- Client retention is essential for understanding and managing customer lifetime value. Central Insurance needs a reliable way to determine whether a customer will renew their automotive insurance policy.
- Renewal decisions depend on many factors: claims, policy characteristics, demographics etc.

### Project Objective

- Leverage machine learning algorithms to predict policy renewals, boost profits, and tailor marketing
- Develop a dashboard to demonstrate the model to stakeholders
- Project focuses on the **Commercial Auto** line of business dataset, spanning 2022-2023

## RETENTION DASHBOARD



The dashboard sorts policies into 5 different categories of retention probability. The relationship between premium per vehicle and retention is shown of the graph.

## CONCLUSIONS

- Retention is highly affected by policy premium per fleet size
- Select a Random Forest model for its explainability and performance
- Improve allocation of retention resources toward at-risk customers to improve profitability
- Provide insights about agency partners and churn behavior

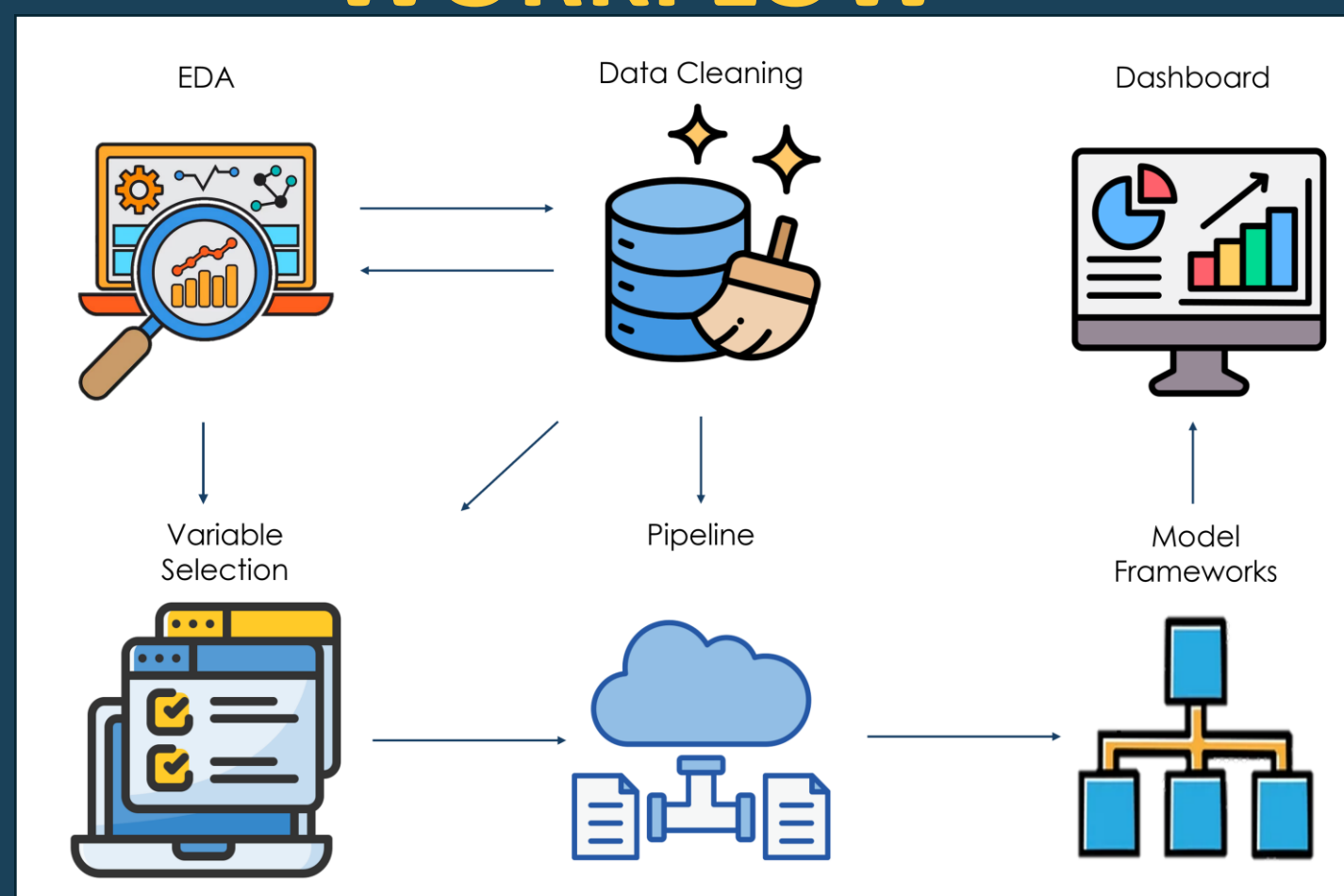
## FUTURE WORK

- Operationalize the model internally
- Establish Monitoring and Maintenance processes
- Embed Insights into Decision-Making

## ACKNOWLEDGEMENTS

- Thank you to Central Insurance, especially mentors Danny Marchese, Saviour Avorgah, and Jamie Walters for guidance, feedback, and technical direction
- Thank you to The Data Mine, for infrastructure, tools, and collaborative learning environment

## WORKFLOW



### Data Pipeline & Preprocessing

- Feature selection (36 → 26)
- Data transformation & missing value imputation
- Function-based pipeline implementation

### Model Development

- Built RF, Logistic Regression, and SVM models
- Applied monotonic constraints

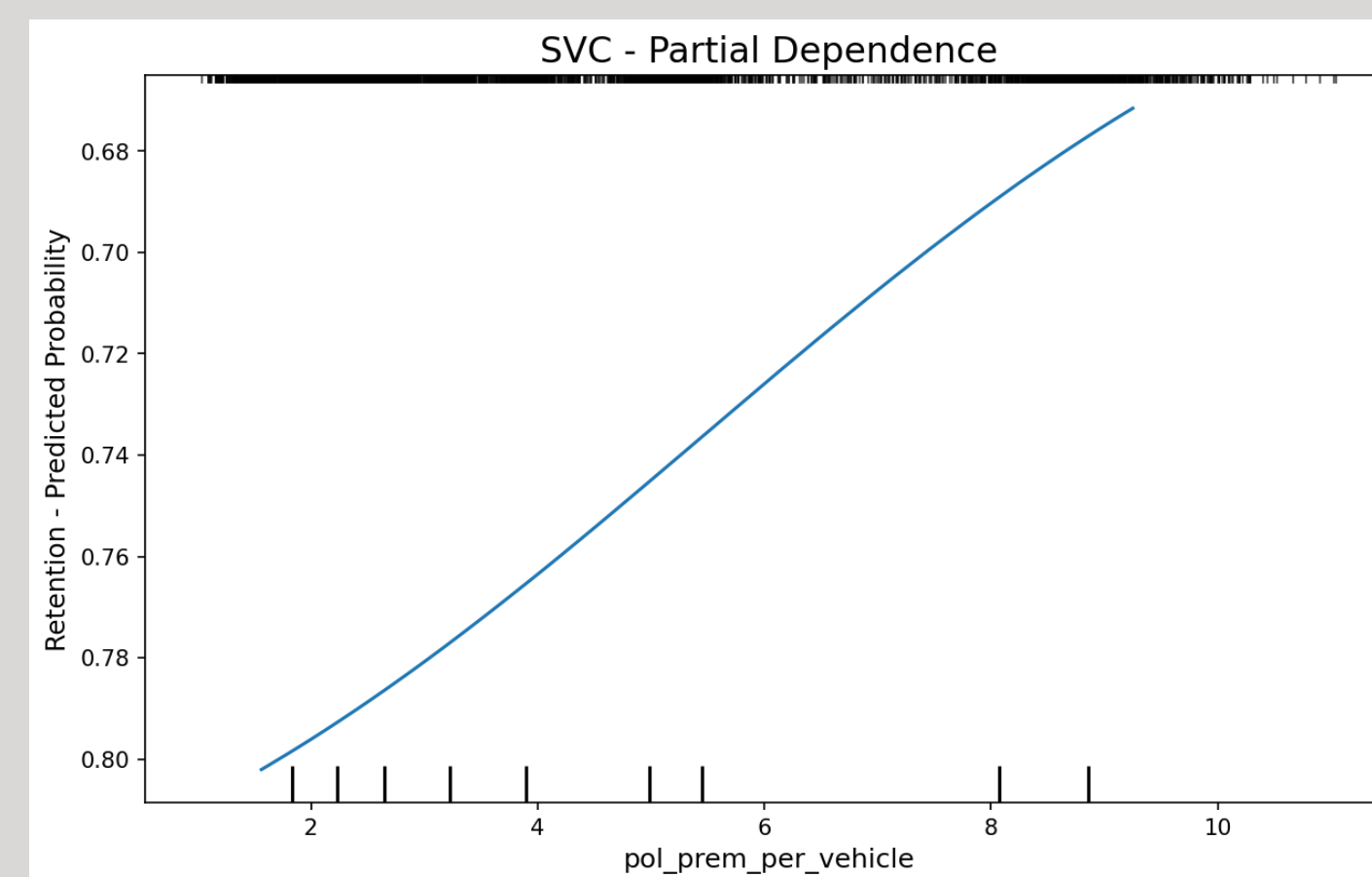
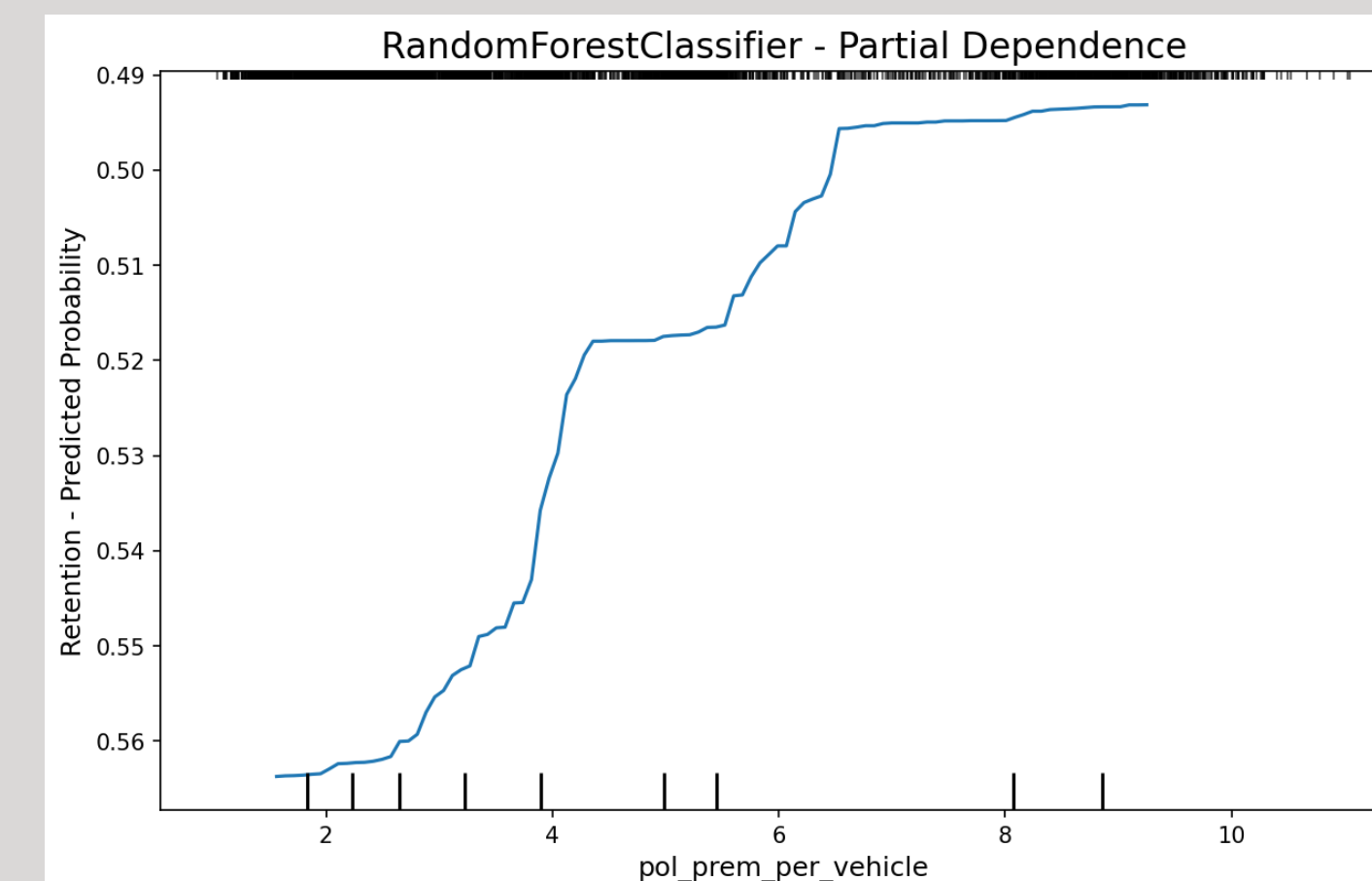
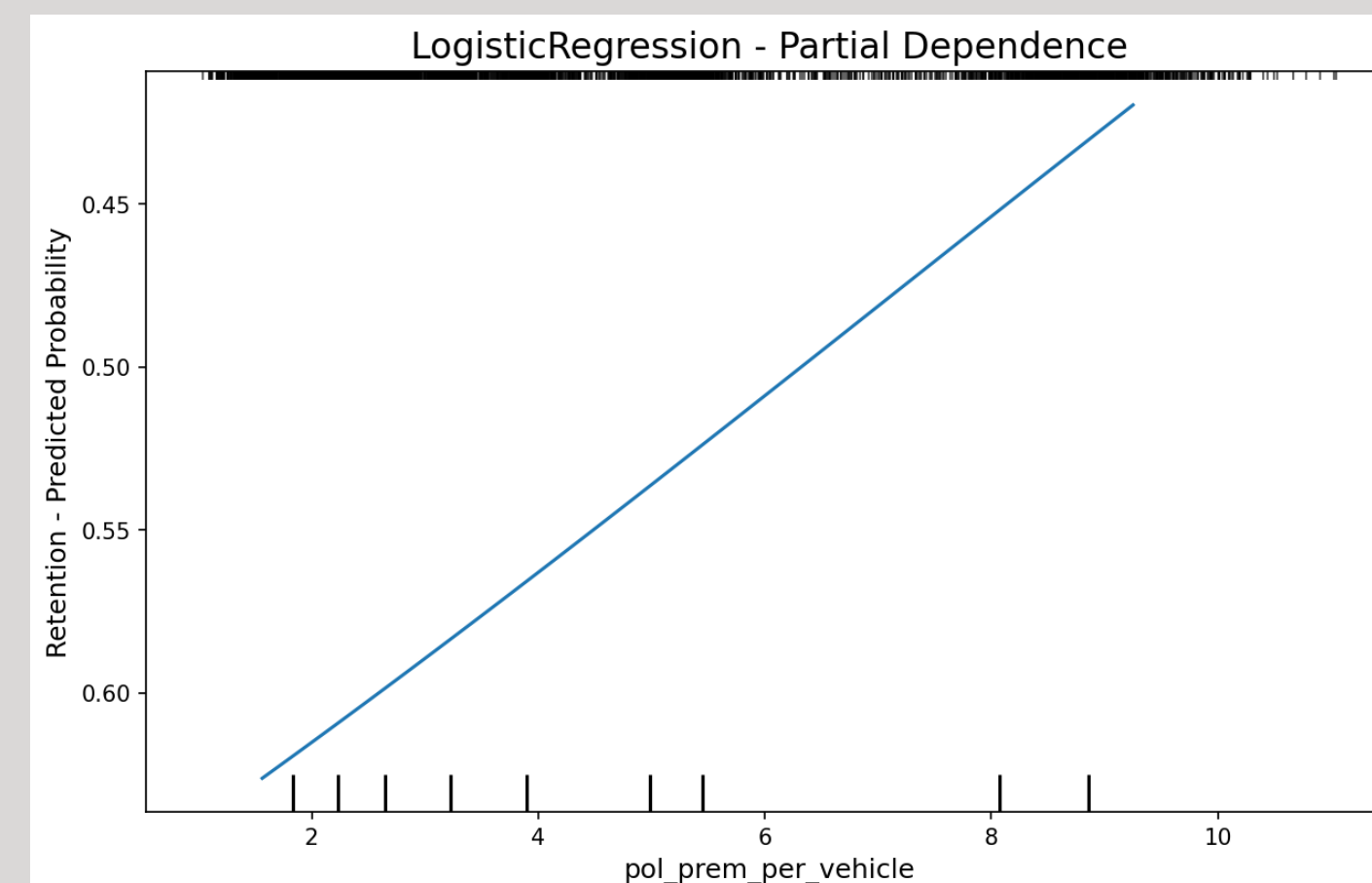
### Model Evaluation & Selection

- Tracked metrics using MLflow
- Interpreted with Partial Dependence Plots (PDPs)

### Dashboard Development

- Built Shiny dashboard using the selected model (RF)
- Demonstrated insights using example policies

## PDPs depict the effect of predictors on retained probability



Model	Accuracy	Sensitivity	# of Variables	F1 Score
Random Forest	71.0%	76.0%	26	79.8%
Tuned Random Forest	73.0%	79.6%	26	81.7%
Logistic Regression	81.3%	98.6%	26	88.9%
Support Vector Machine	80.3%	95.4%	26	88.0%