PREDICTING CAR INSURANCE CLAIMS

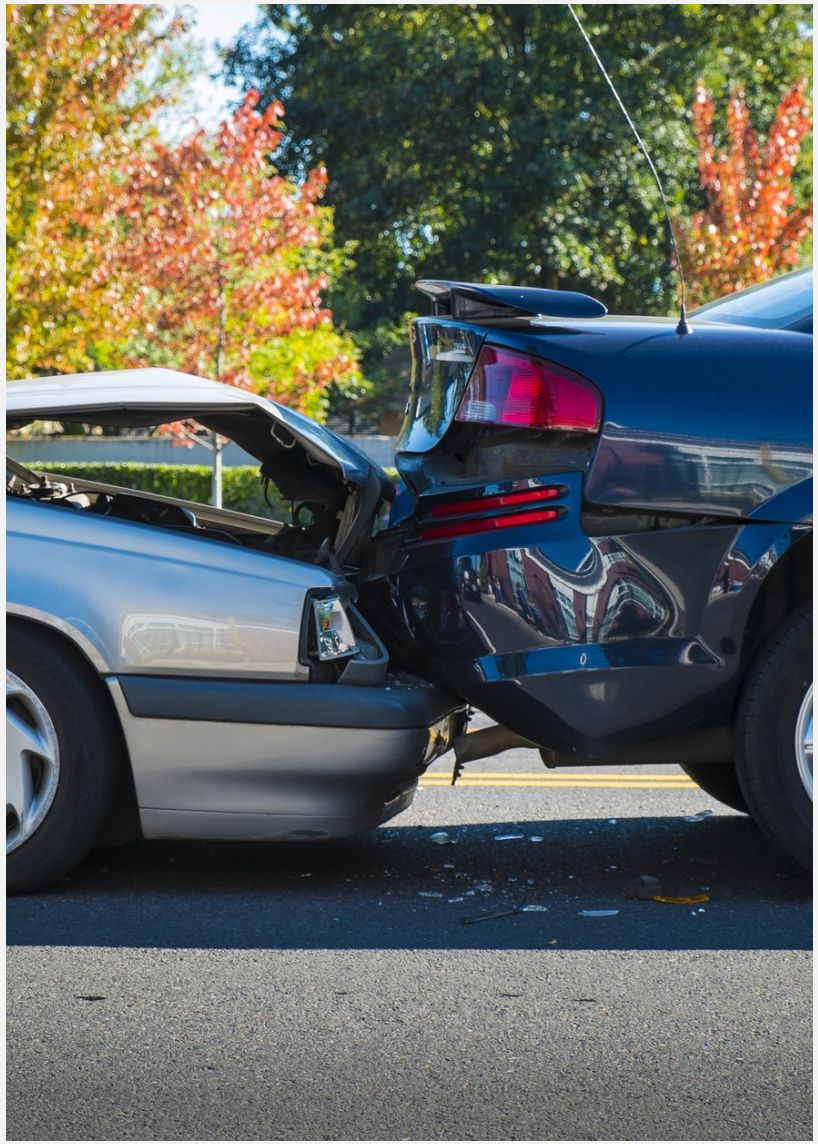
CLASSIFICATION MODULE PROJECT

INTRODUCTION

- Predicting car insurance claims
- Porto Seguro Modelling Competition on Kaggle
- Large, imbalanced dataset:
  - circa 600,000 datapoints
  - Each with 57 features
  - Only 3.6% datapoints show a claim

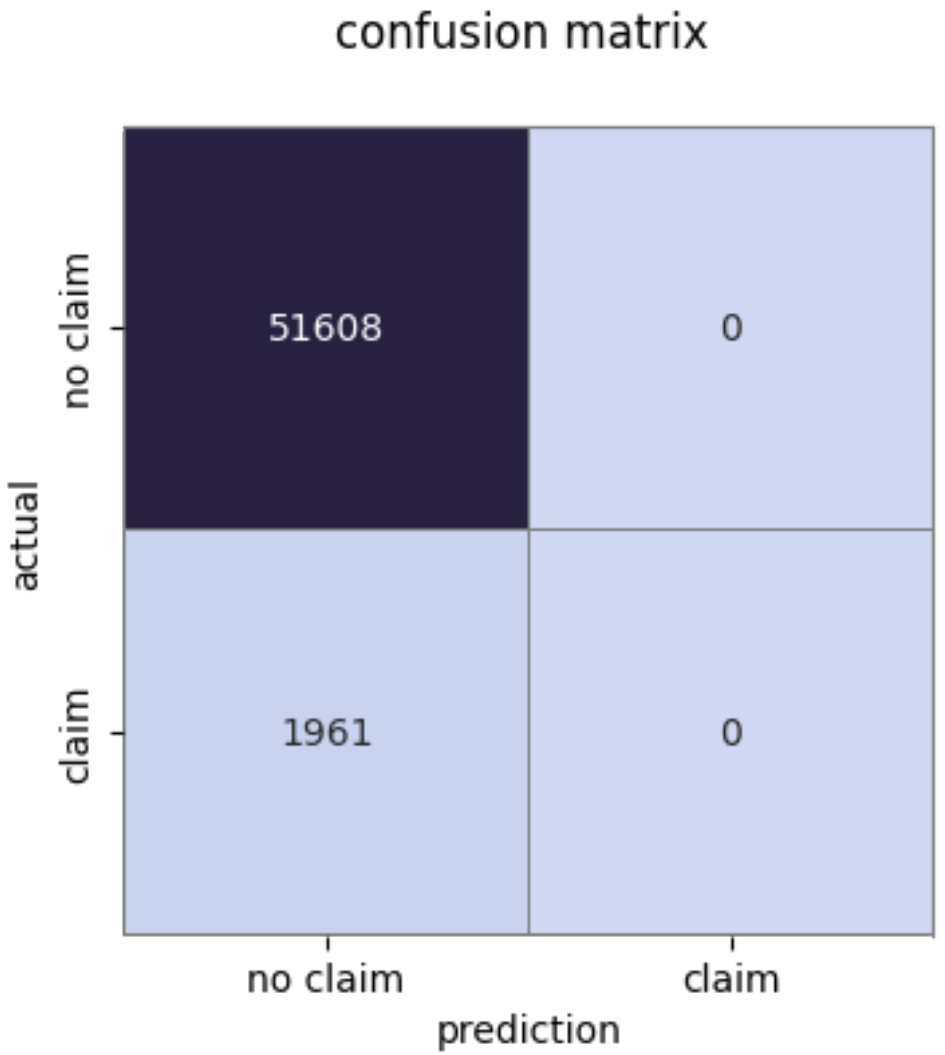
TOOLS

- Compute:
  - Google Cloud VM
  - Google Colab
- Modelling:
  - Sklearn
- Visualisations:
  - Matploltlib
  - Seaborn
  - Plotly



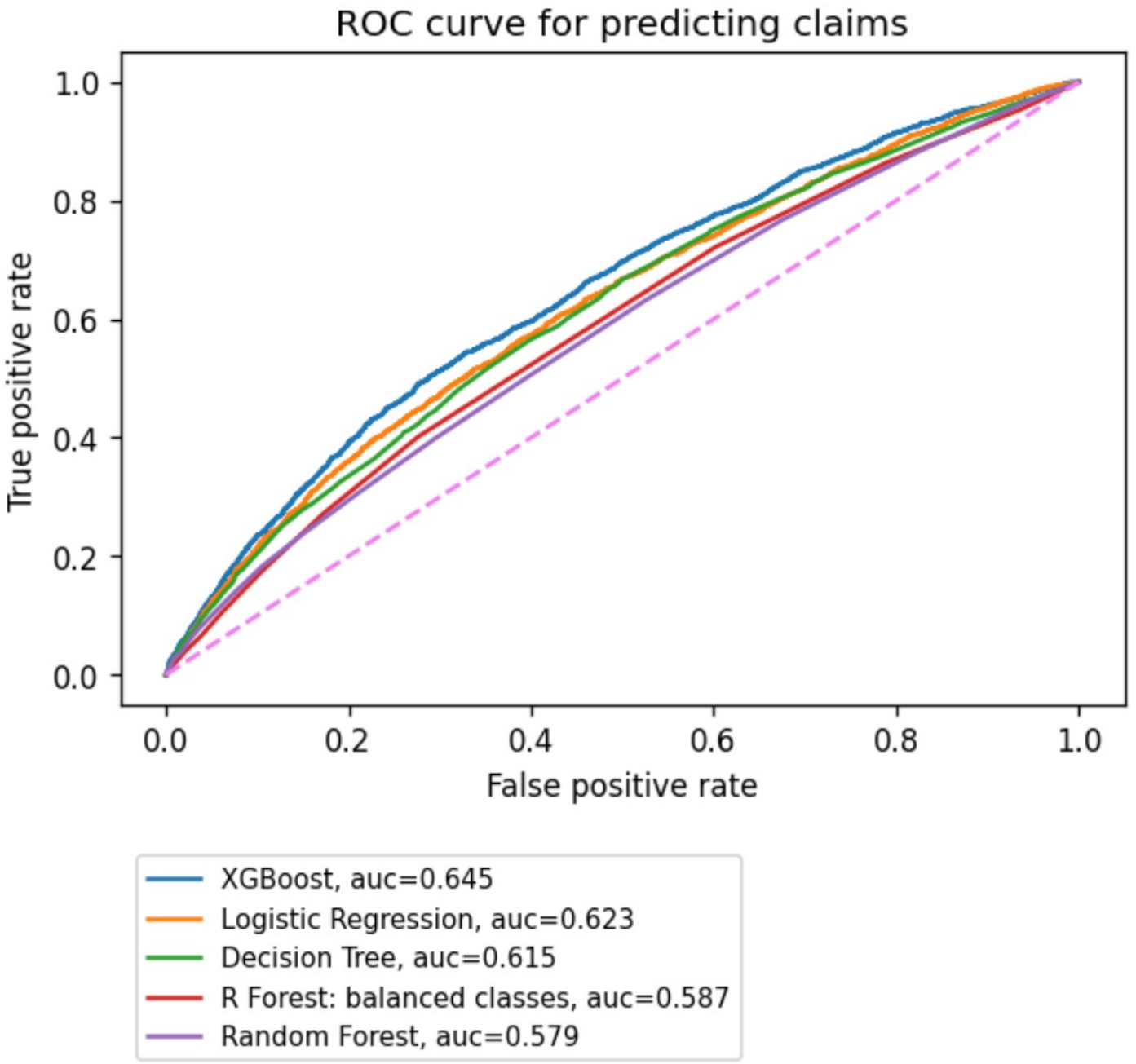
THE CHALLENGE

- Most models initially produced recall and precision values of zero
- Initial AUC scores near 0.5



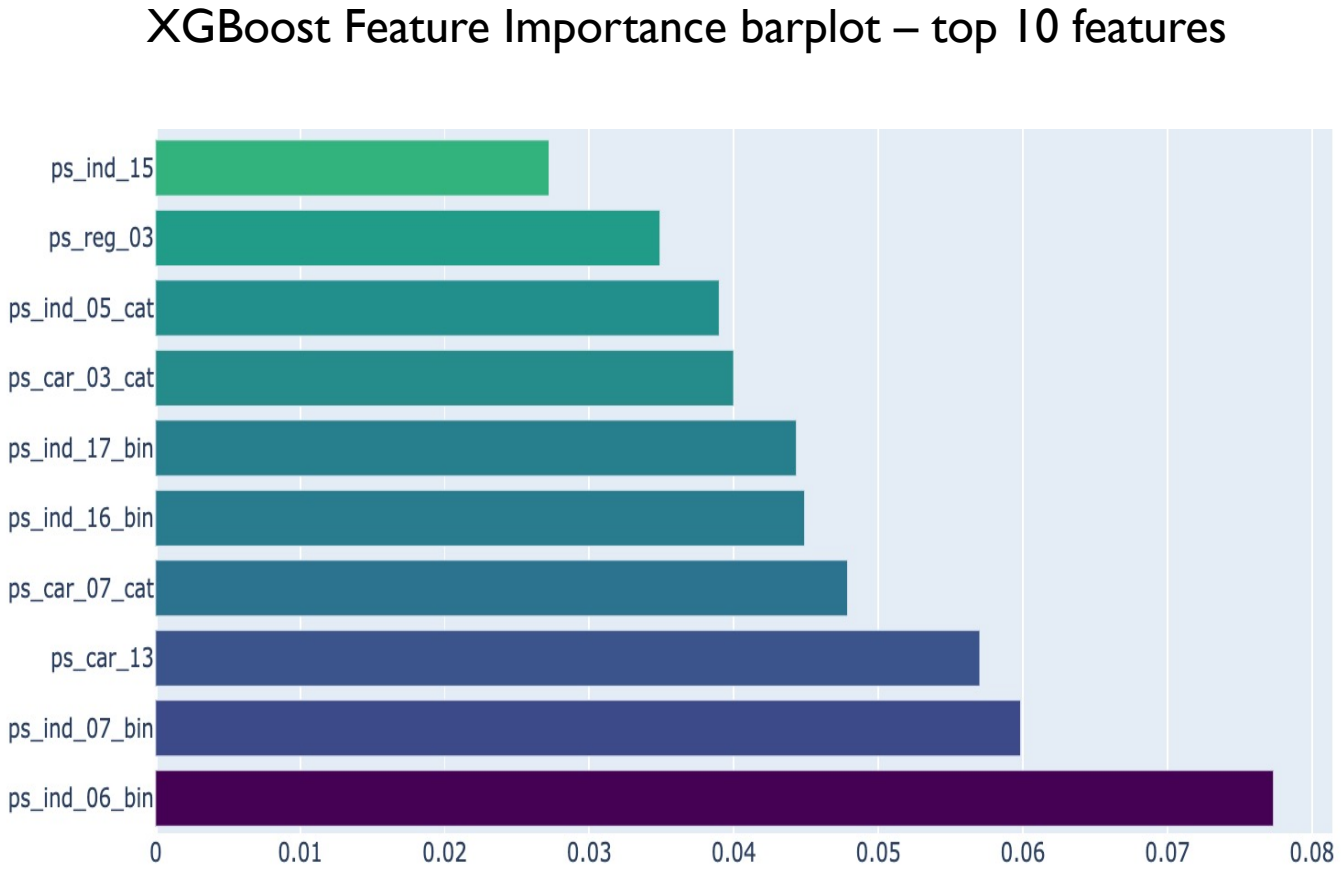
MODELLING RESULTS

- XGBoost was the winner
- Admittedly more tuned than other models
- But AUC of 0.638 with no tuning: - so winner anyway
- Model selection stage so AUC scores here are on validation data



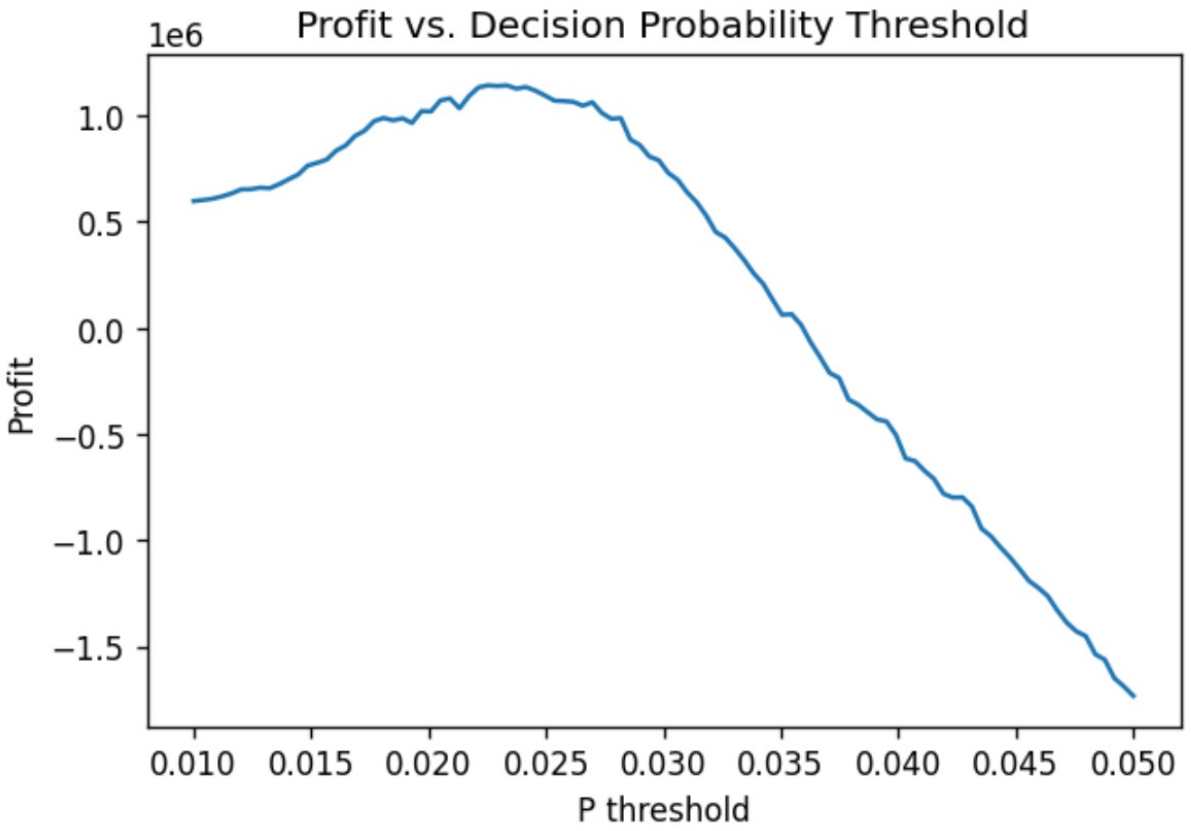
FEATURES

- Feature importance is obtainable from XGBoost
- But in this dataset features are disguised, so of limited use



PROFIT FUNCTION

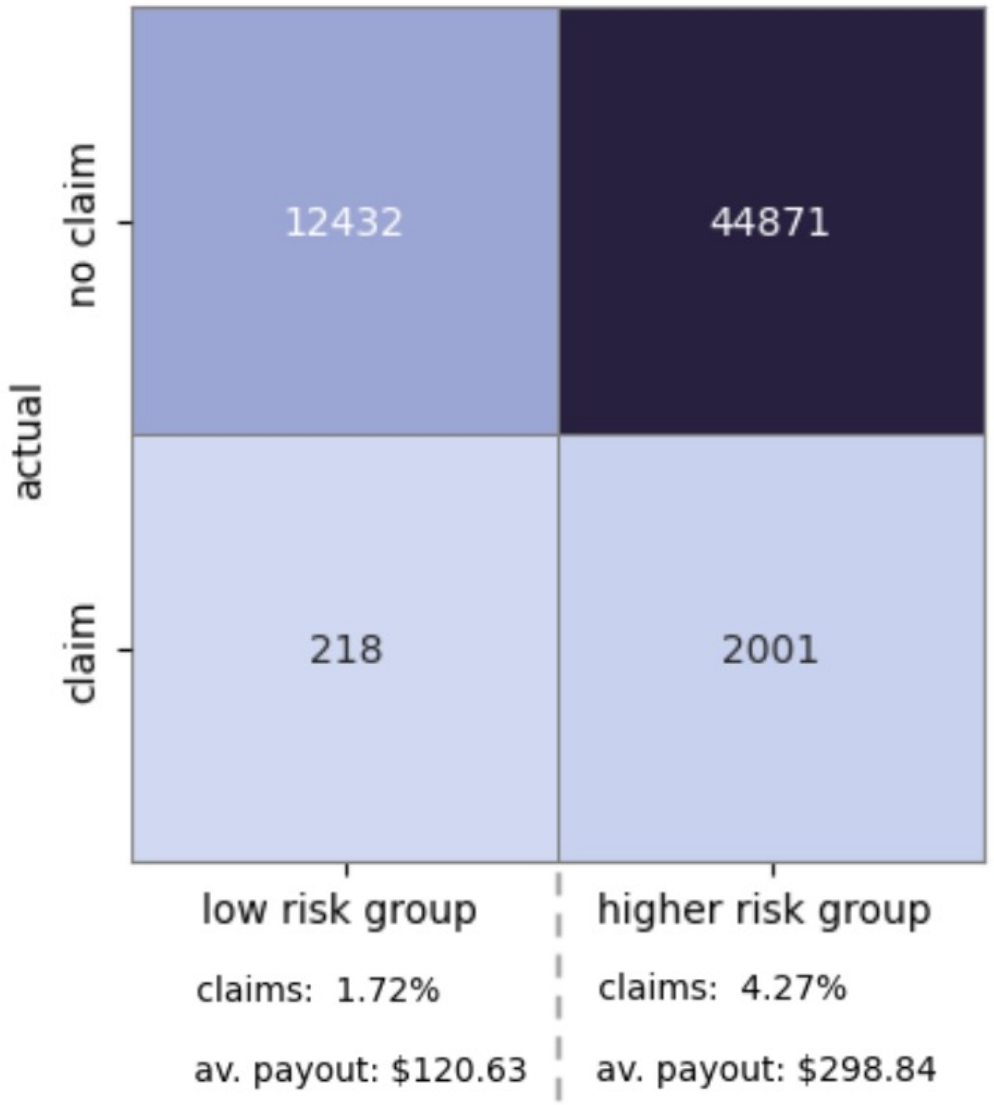
- Scenario:
  - Average payout per claim: $7000
  - Marketing department wants $199 premium for low risk drivers
  - Difference between $199 and average payout per low risk driver (plus fixed cost) is profit
  - Higher risk drivers are priced at $10 markup over cost




```
Highest total profit $1143820 at prob decision threshold >= 0.023
Profit per policy $19.22
```

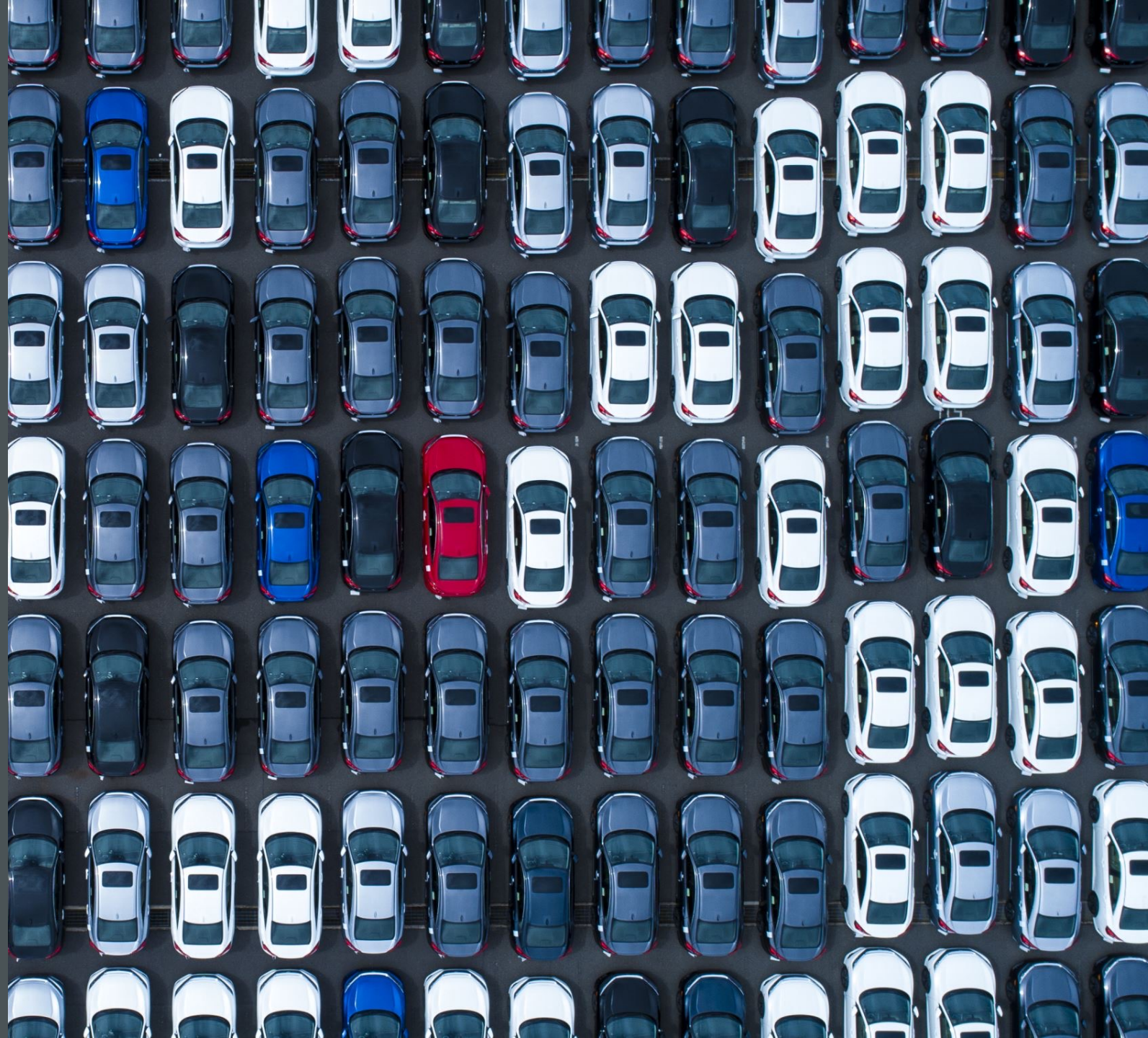
RESULTING GROUPS

- Low risk group members less than half as likely to submit a claim
- Average payout per driver therefore less than half that of higher risk group
- This enables profitable price discrimination

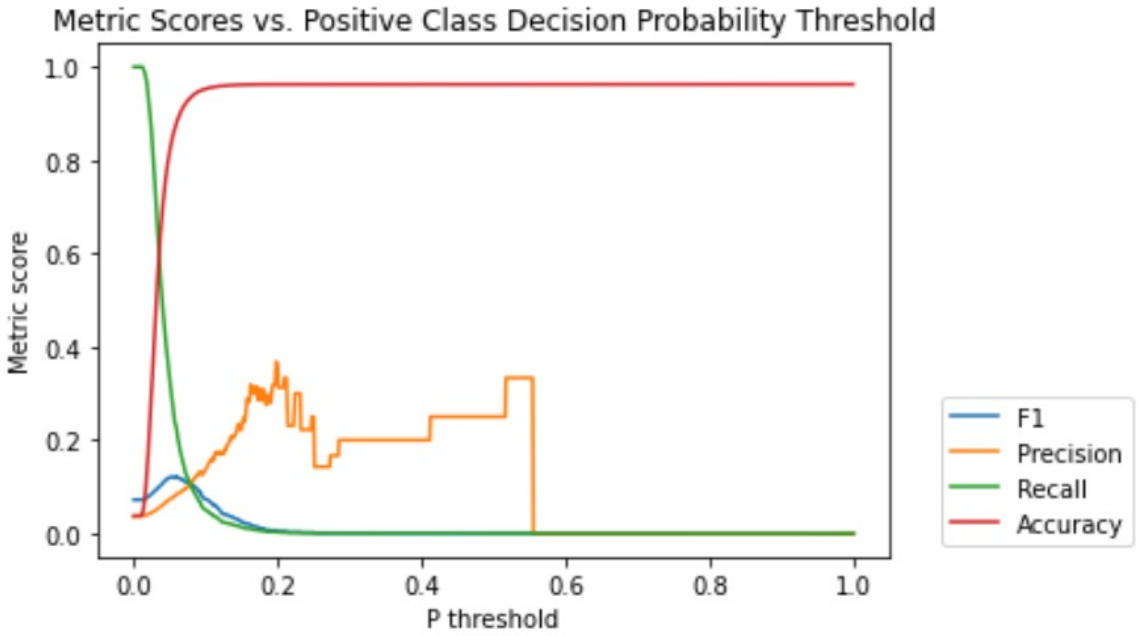


CONCLUSION

- Models here have relatively low predictive power for whether an individual driver will claim
- But probabilities assigned can be very useful at an aggregate level



APPENDIX I


Metric Scores vs. Positive Class Decision Probability Threshold
Metric score
P threshold
0.0
0.2
0.4
0.6
0.8
1.0
F1
Precision
Recall
Accuracy

APPENDIX 11

https://www.iii.org/fact-statistic/facts-statistics-auto-insurance

Passenger Vehicle Collision Coverage Insurance Losses, 2018-2020 Model Years

The chart below shows the claim frequency and average loss payment per claim and average loss payment per insured vehicle year under collision coverage for recent model vehicles.

	Claim frequency (1)	Claim severity	Overall loss (2)
Passenger cars (3)	7.2	$6,831	$495
Pickups	5.5	6,925	381
SUVs	5.5	6,816	377
All passenger vehicles (4)	**6.1**	**$6,839**	**$419**

Average payout per claim is roughly $7000 in US