CAR INSURANCE

# ESTIMATE CAR INSURANCE PREMIUMS

LOW-RISK CUSTOMERS

JC

# BUSINESS PROBLEM

BACKGROUND

Competitive insurance companies provide customers with a reasonable price that is adjusted based on the level risk on the insurer. As an agent considers their time management, it would be beneficial to focus on medium-to-high risk policies.

A machine learning model will predict an economical price for a low-risk customer. The agent would still review the estimate based on the features known about the proposed insurance policy.

# DATASET

ETHIOPIAN INSURANCE COMPANY

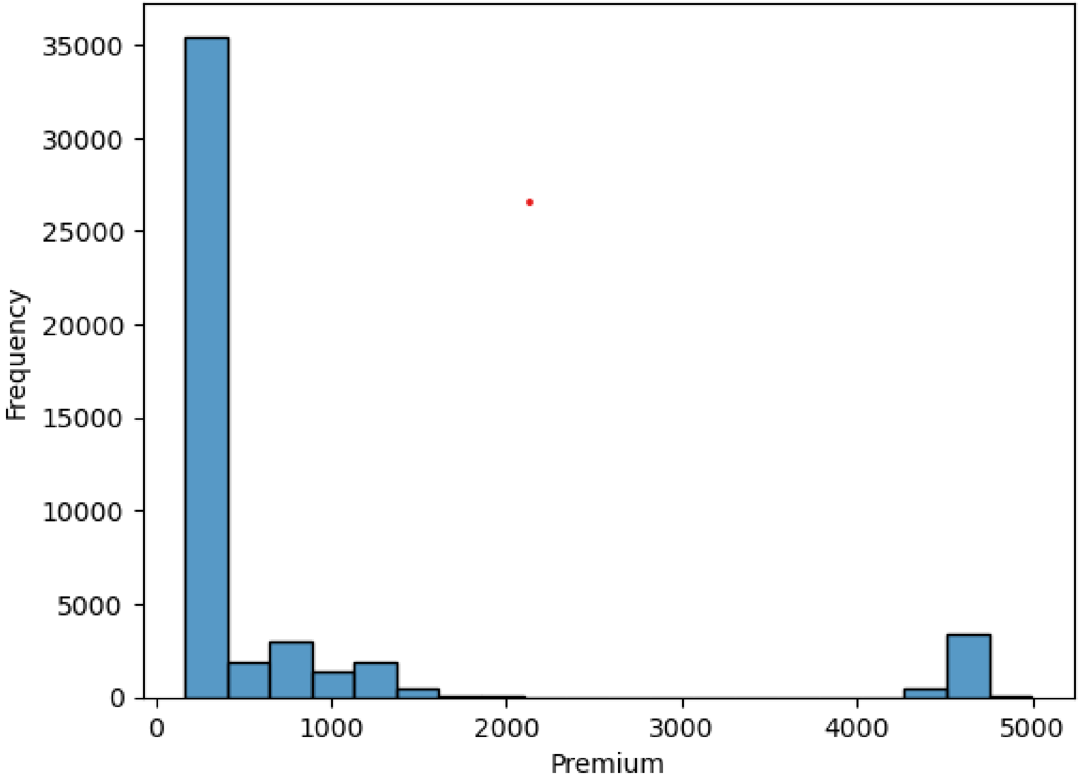
- A policy list from 2014-2018 was obtained from the Ethiopian Insurance Company posted on Kaggle.
- The policy features were defined as:
  - Liability (no insured value)
  - Starts January 2018 or later
  - No documented claims paid
  - Premium <= $6,000
- A low-risk customer's vehicle was defined as:
  - <= 10 years old
  - Small or Medium size
  - Vehicle Make >= 100 samples

# ANALYSIS

DATA PREPARATION

- Features required data type changes and consolidation of categories.
- Null values were found within Premium and were replaced with the mean.
- The following trends were noted:
    - Premium was primarily $2,000 or less
    - Motorcycles were the most common vehicle type for low risk-customers.
    - The most common usage was for taxi or similar type purposes.
    - Production year of vehicles were concentrated between 2012-2018.

Car Insurance Premium Histogram



# CORRELATION

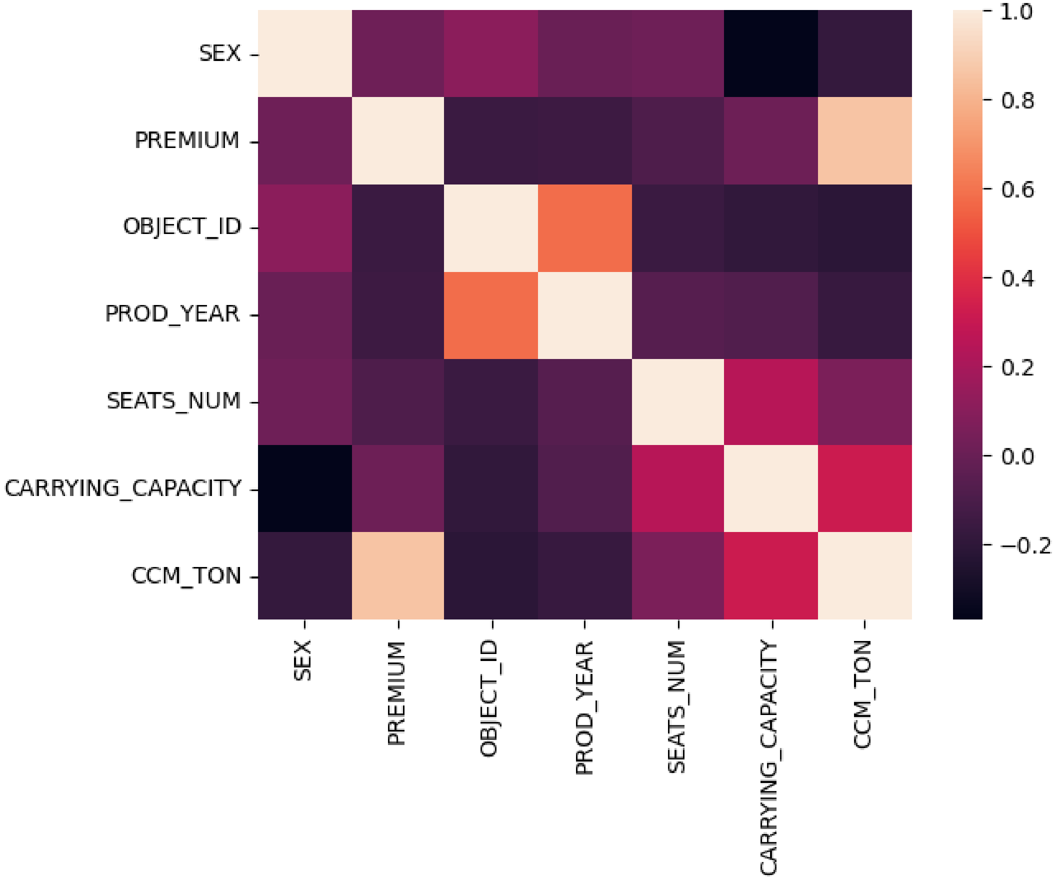
### Positive Correlation to Premium

| Types | Correlation |
|---|---|
| Types: Truck | 0.9711 |
| Make: Isuzu | 0.9572 |
| CCM Ton | 0.8624 |
| Usage: Commercial | 0.5561 |
| Type: Pick-up | 0.0744 |

### Negative Correlation to Premium

| Types | Correlation |
|---|---|
| Type: Motorcycle | -0.7149 |
| Usage: Taxi | -0.4016 |
| Make: Bajaji | -0.3714 |
| Usage: Private | -0.3714 |
| Production Year | -0.1597 |

### Feature Correlation Matrix



# MODEL EVALUATION

LINEAR REGRESSION

The dataset was split into 70/30 for model training with 12 features selected based on the p-value.

Three model types were explored within linear regression.

**Models Evaluated**

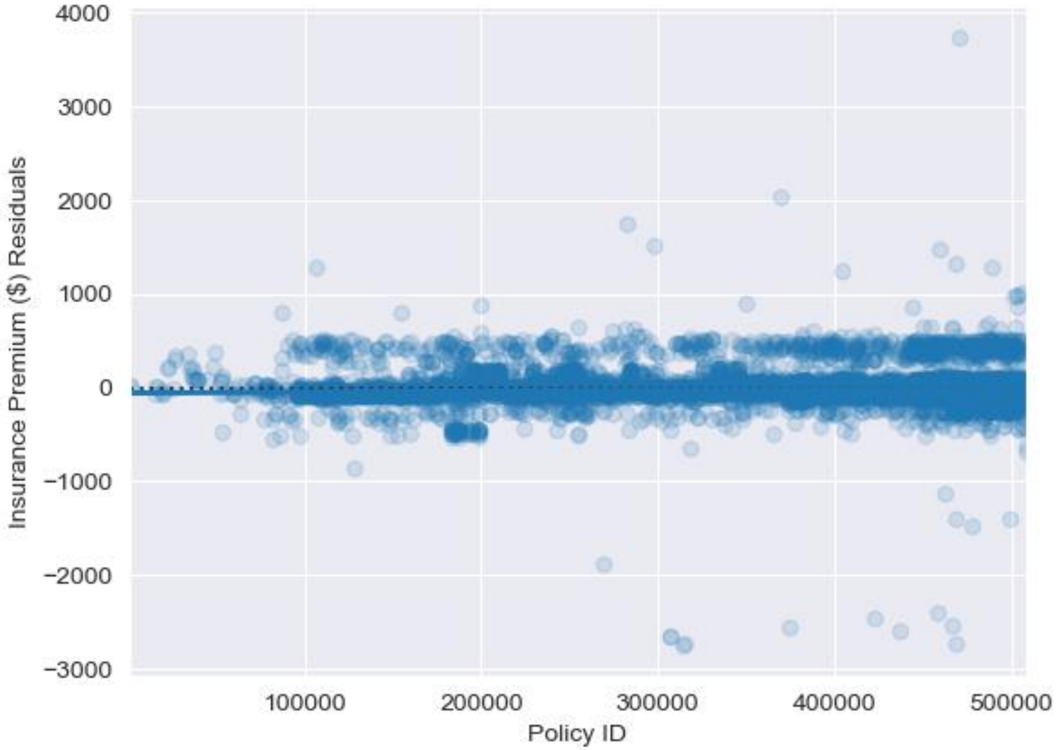
| Types | R2 Score | RMSE |
|---|---|---|
| Linear | 98.201% | $155.86 |
| Ridge | 98.201% | $155.87 |
| Lasso | 98.197% | $156.05 |

- Sex
- Carrying Capacity
- CCM Ton
- Seat Number
- Vehicle Type
  - Automobile
  - Motorcycle
  - Pick-up
  - Station Wagon
- Make
  - Bajaji
  - Nissan
  - Isuzu
- Usage
  - Taxi

# LINEAR REGRESSION

RESIDUALS

Low-Risk Premium Model Residuals



The linear regression model performed the best with a 98.201% R2 score and $155.86 RMSE. With the premium concentrated under $2,000, the RMSE is in acceptable range compared to the tradeoff of saving agents' time.

The residuals between the actual and predicted values were plotted to determine the success of the model. It appears that the predicted value varies more with the increase of the Policy ID (newer written policies).

# ETHICAL CONCERNS AND LIMITATIONS

For policies with insured values that equal zero were associated with a liability policy. A liability policy is defined as a product that provides protection against claims resulting from injuries and damage to other people or property (Kagan, 2022).

The low-risk premium limit was chosen based on the unfiltered histogram of premium after all other filters were applied.

The model should not replace agent validation time once it enters underwriting.

The dataset was over 6 years old and may or may not be accurate or approved for distribution outside of an educational setting from the Ethiopian Insurance Corporation.

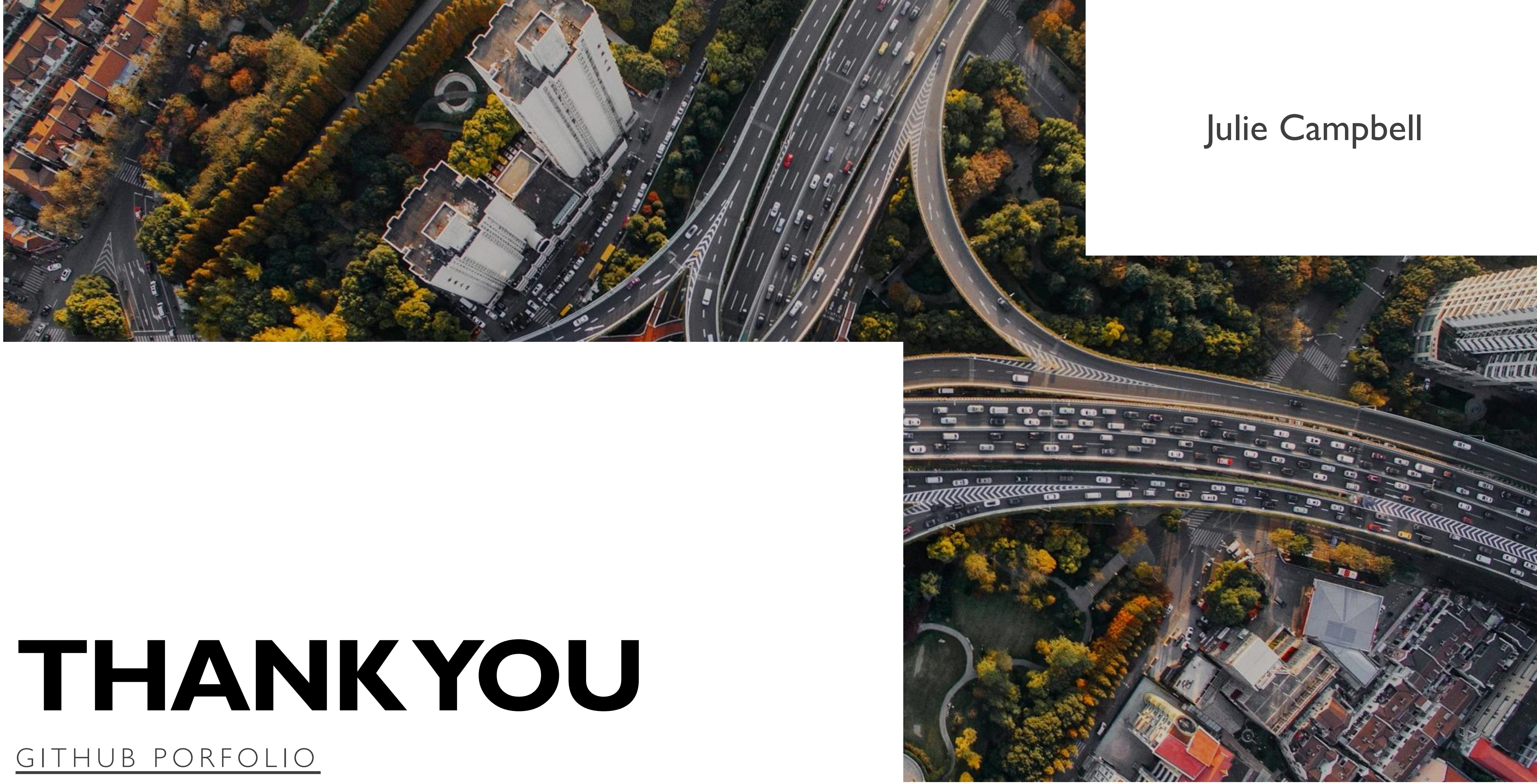
# FUTURE APPLICATIONS

RECOMMENDATIONS

The low-risk customer premium model could be used as a basis for insurance companies to reduce the workload on agents by creating an initial premium recommendation.

The model and associated report will be posted on GitHub for open-source use for further development. It is recommended to validate the model against newer data.

CAR INSURANCE



Julie Campbell

# THANKYOU

GITHUB PORFOLIO

JC