

Predicting Pet Insurance Claims

Executive Summary

Project Goal:

Build a model to predict the amount of claims in the second policy year

Key Insights:

- Dog owners pay higher premiums on average
- Dog owners are more likely to have claims in one or both policy years
- Dogs tend to have more claims & a higher total claims amount

Project Results:

Model improved claims predictions **by nearly \$400** over baseline

Why is this important?

Accurate policy pricing is critical for business success

- Pricing too high results in fewer customers and higher risk
- Pricing too low results in lost revenue

How do we improve policy pricing accuracy to maximize revenue?

- Utilize historical data to better understand customer patterns
- Build a model to predict claims in the future

What does the data look like?

Data we have:

Policy details:

- Enrollment date, Premium & Deductible

Pet info:

- Species, Breed, Age

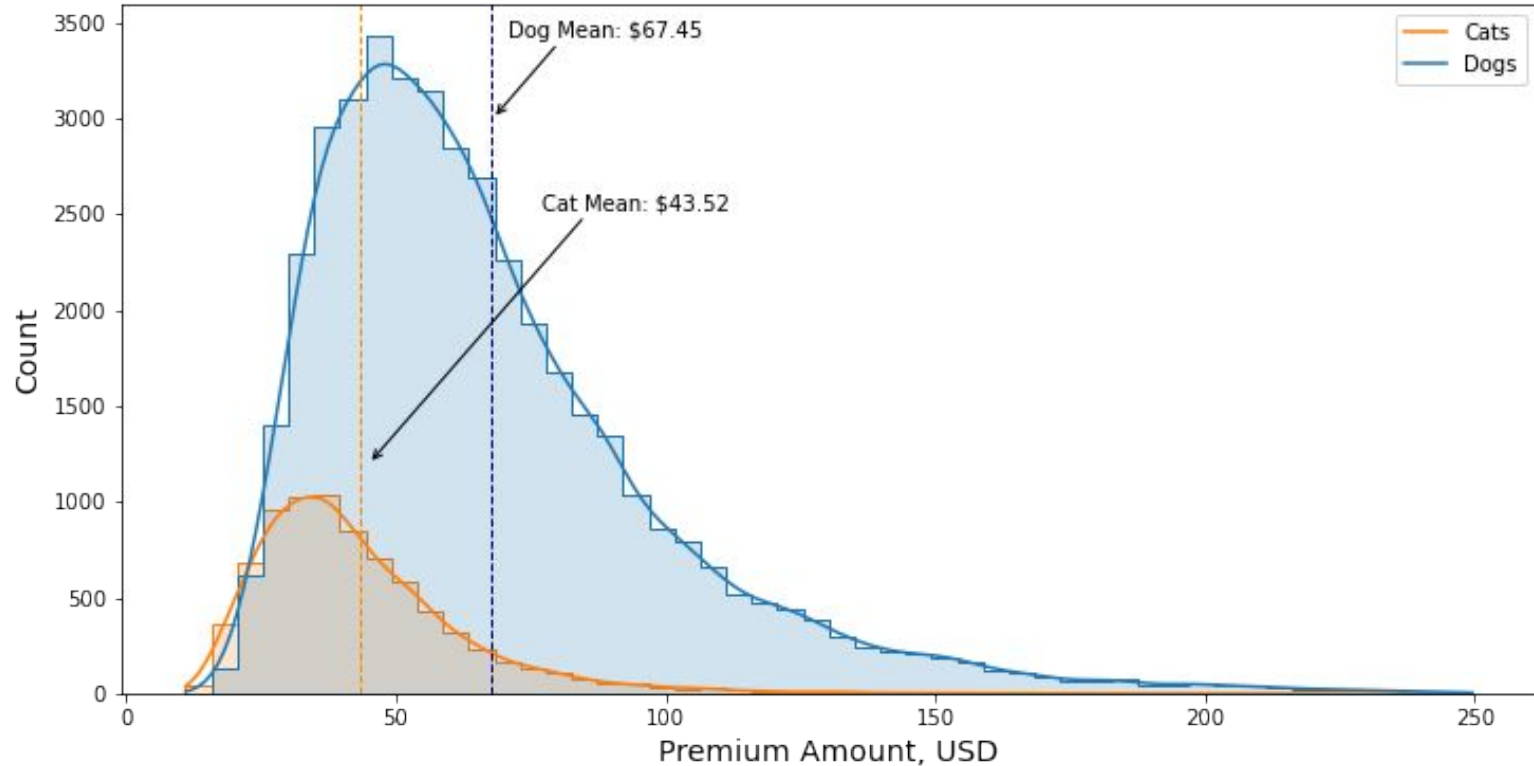
Claims data:

- Pet Id, claim date, claim amount

Data that might improve predictions:

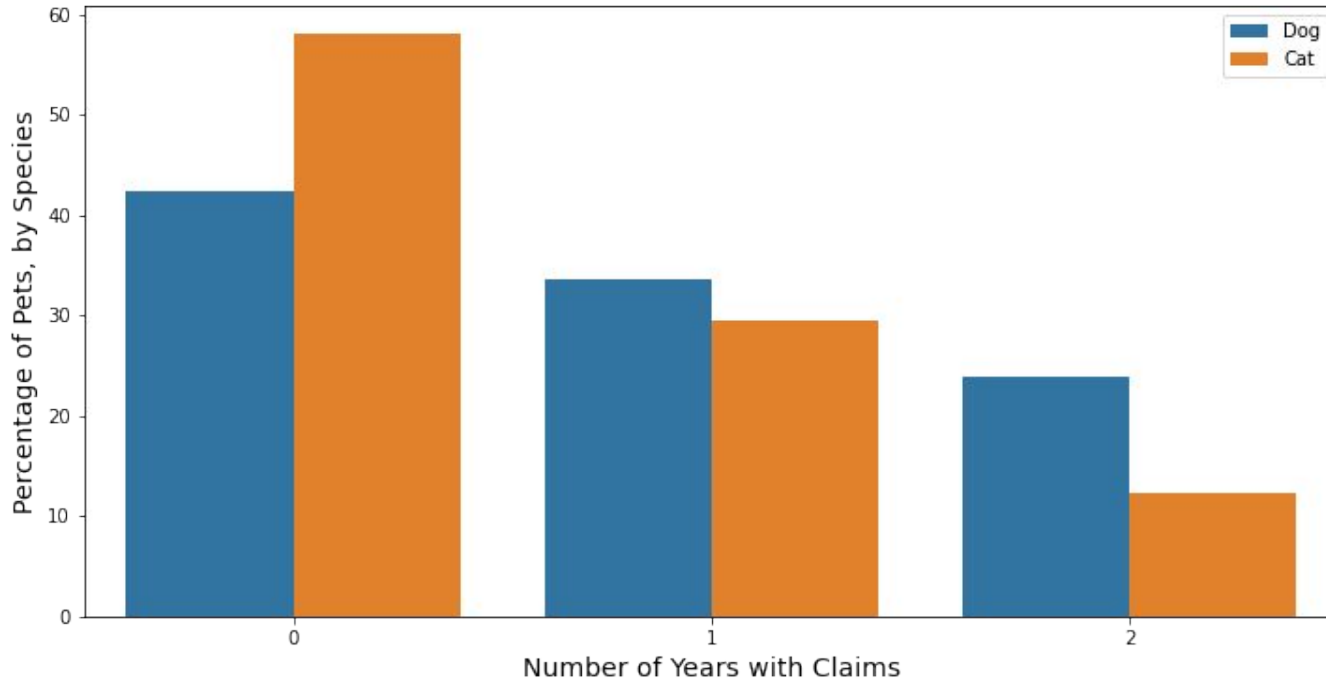
- Claim status (e.g., paid, rejected, etc.)
- Pet owner demographics of CLV
- Species or breed-related info:
 - Average lifespan
 - Average annual claims
 - Risk of significant health issues

Dog Owners Pay Higher Premiums

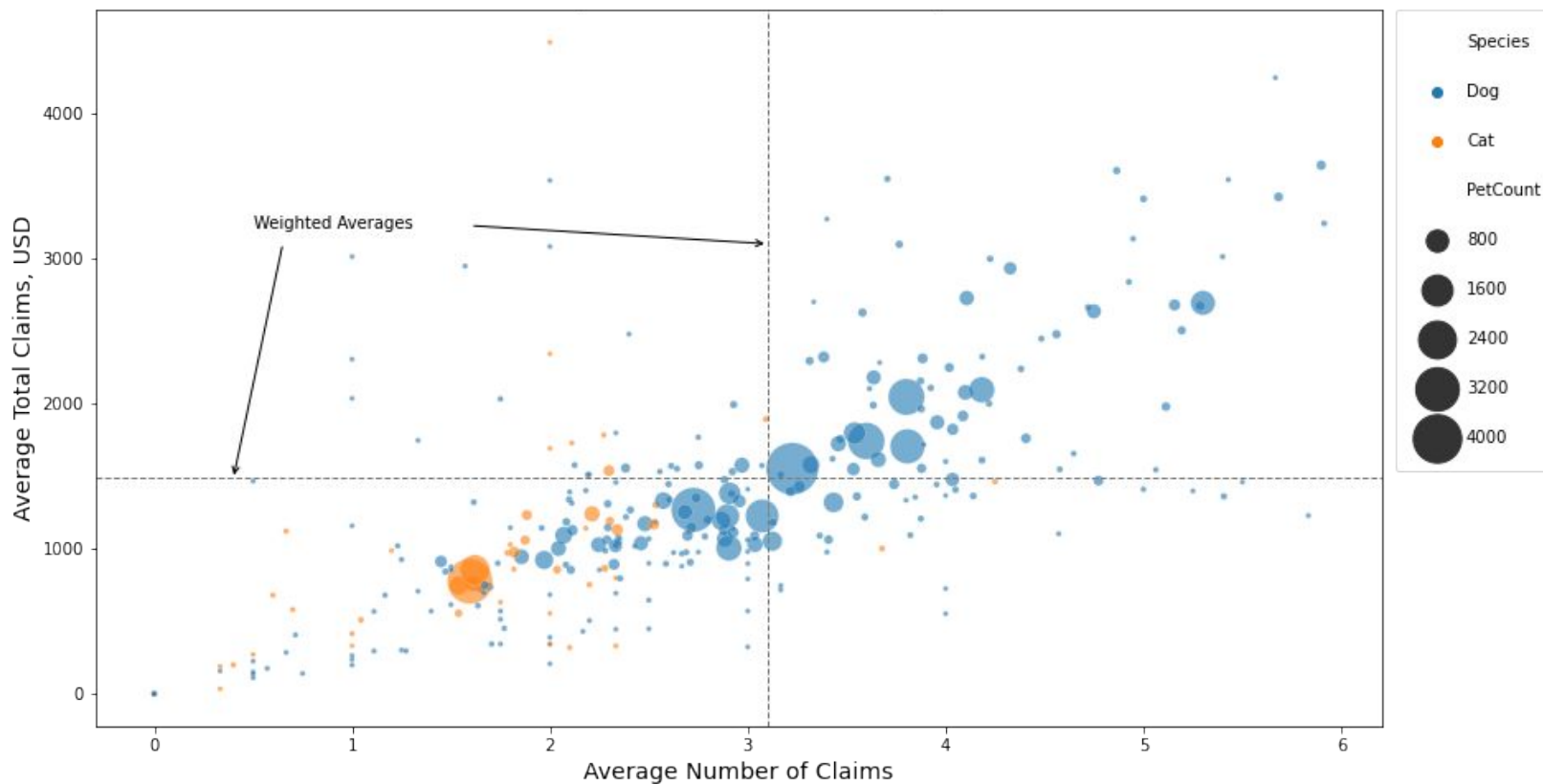


Dog Owners Are More Likely to Have Claims

In one or both policy years...



Cat Owners Have Lower Total Claims (on Avg.)



Predictive Modeling & Evaluation

Method

- Explored a range of predictive models
- Model evaluation based on Mean Absolute Error (MAE) in units of USD

Model Evaluation

Predictive Modeling Method	MAE (lower is better)
Baseline model (using average claims amount)	\$1020
Best linear regression model	\$930
Best gradient boosting model	\$675

Predictive Results & Conclusions

Model Results on New Data:

- Mean Absolute Error - \$640
 - \$35 improvement over model evaluation results (\$675)
 - \$400 improvement over baseline (\$1020)

Recommendations:

- Repeat predictions with new data to confirm result
- If successful, employ model to evaluate current and future policy pricing

Initial Results Show Promise, But Can We Do Better?

Opportunities for improvement:

- Start with more a balanced dataset
 - Species & Breeds
 - Pet age
- Add more supporting data
 - Owner demographics and CLV
 - Species / Breed info
- Engineer additional features
 - Claims timing
 - Breed-related 'risk index'
- Evaluate additional predictive methods