Predicting Pet Insurance Claims

Executive Summary

Project goal:
Identify 3-4 customer segments to improve customer marketing efforts and outcomes

Results:
4 clearly defined customer segments identified driven primarily by number of years with claims

Other Key Findings:

- Dogs tend to have higher claims totals on average
- Dogs also tend to have a higher number of claims
- Dogs are almost twice as likely to have claims in both policy years

Background

Why is this important?

Targeted marketing campaigns help to:

- Reduce customer shrinkage
- Increase customer satisfaction
- Increase customer LTV (lifetime value)

A better understanding of our customer base and segmentation, helps to drive revenue growth.

What does the data look like?

Data we have:

Policy details:

- Enrollment date, Premium & Deductible

Pet info:

- Species, Breed, Age

Claims data:

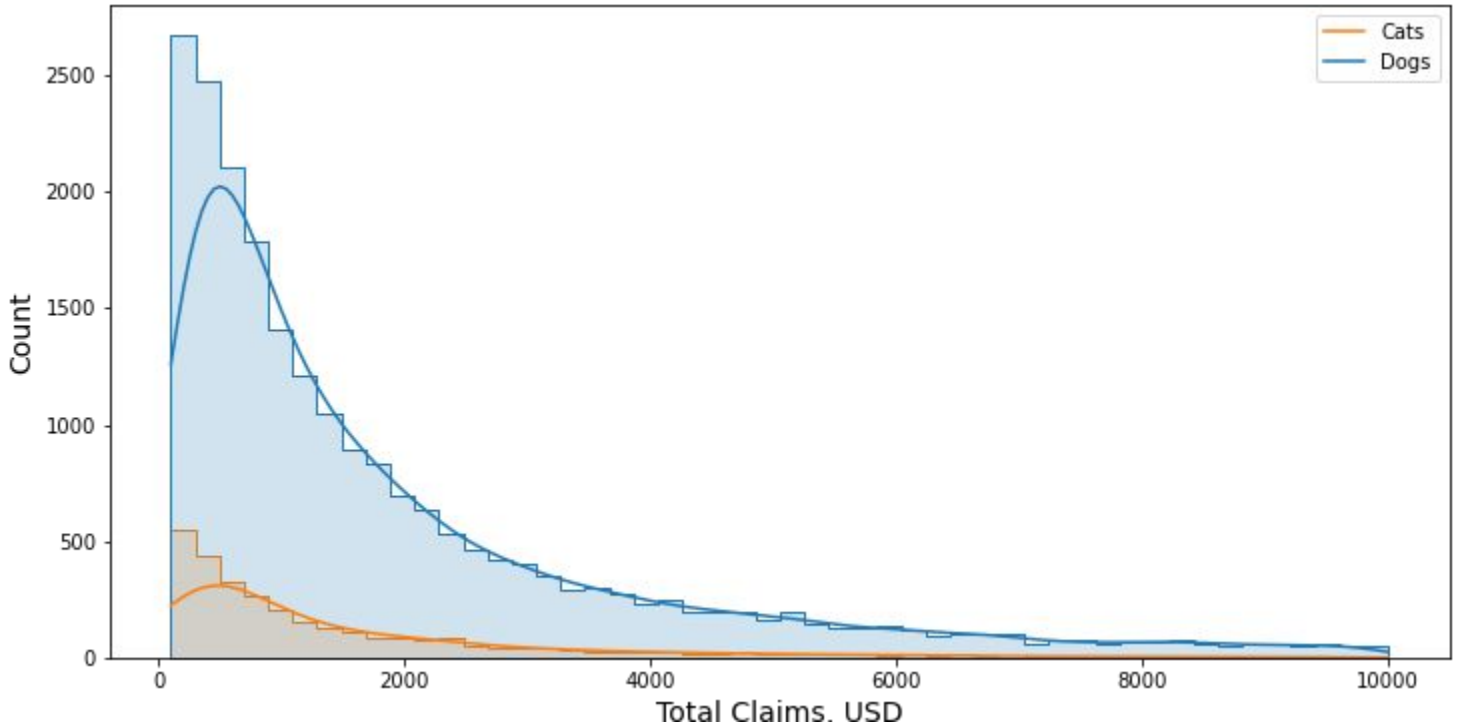
- Pet Id, claim date, claim amount

Data that might improve segmentation:

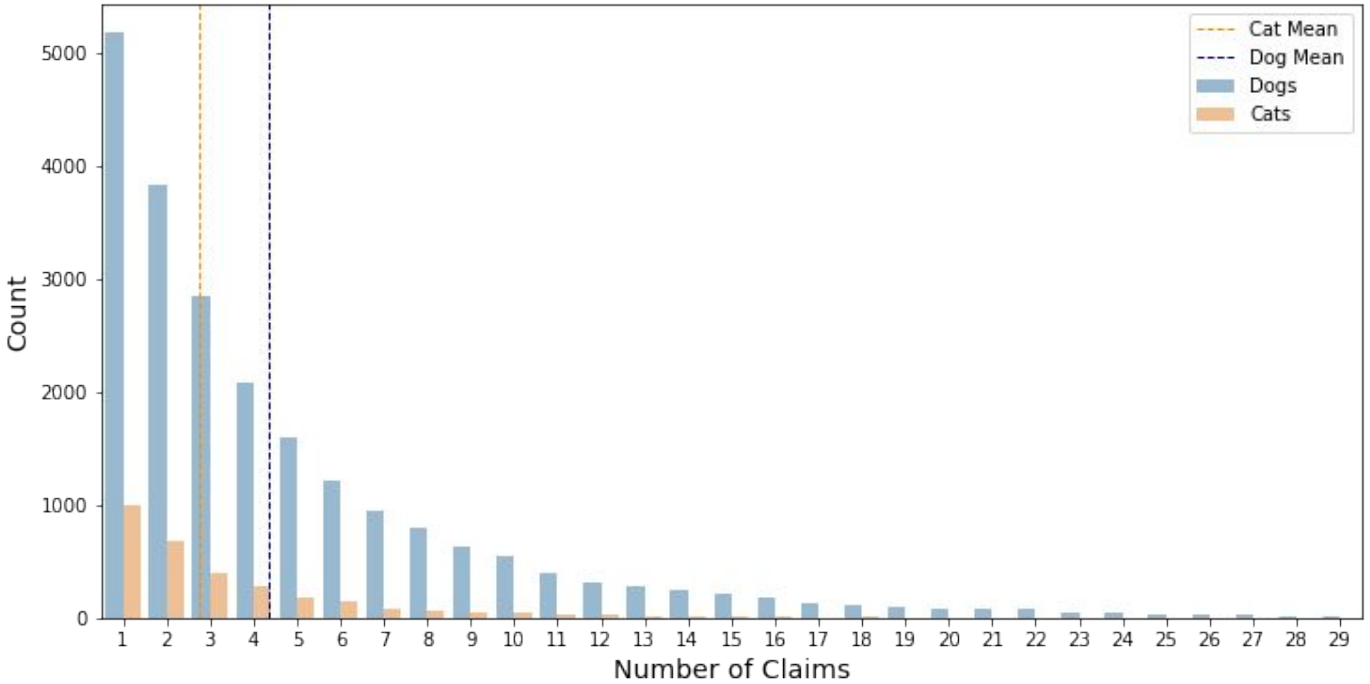
- Claim status (e.g., paid, rejected, etc.)
- Pet owner demographics of CLV
- Species or breed-related info:
  - Average lifespan
  - Average annual claims
  - Risk of significant health issues

Data Analysis - Key Findings

Dog Owners Tend to Have Higher Claims Totals

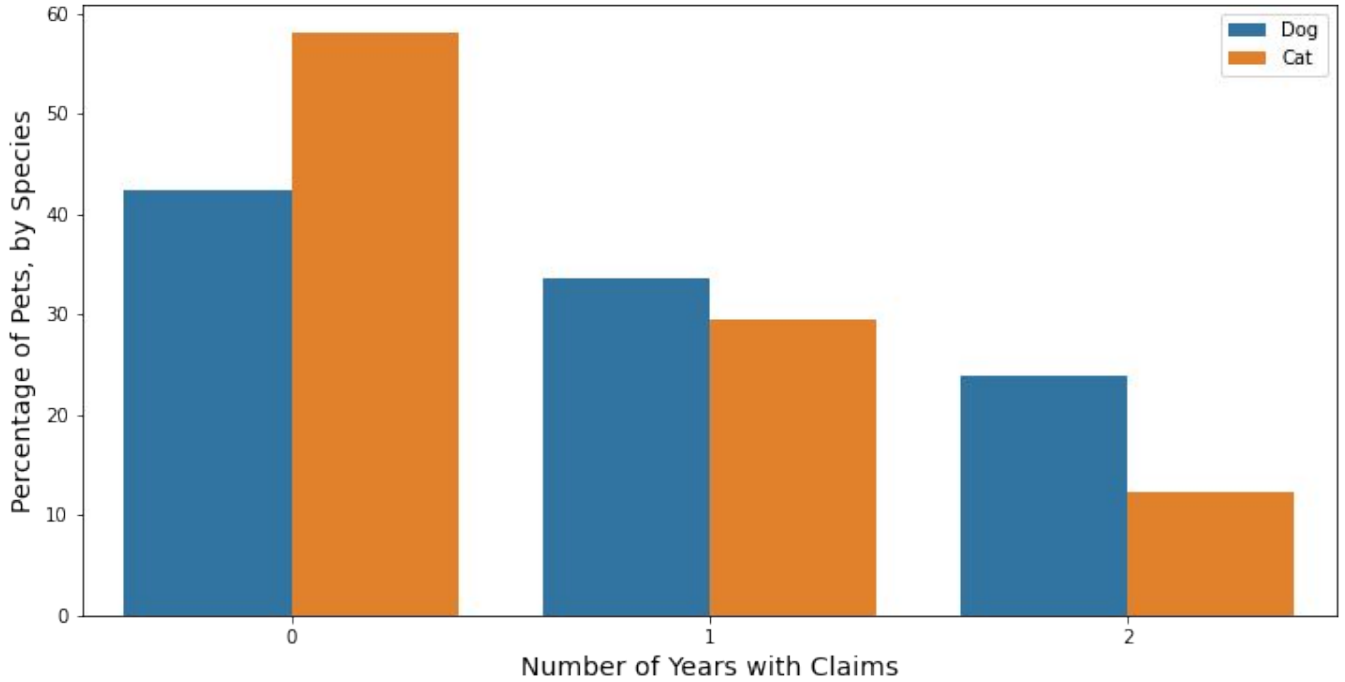


Dog Owners Tend Submit More Claims Overall



Dog Owners Are Also More Likely to Have Claims

And almost twice as likely to have claims in both of the first two policy years



Customer Segmentation

Overall Process

1. Performed Principal Component Analysis (PCA) to create new combinations of features that best explain the variance in the data

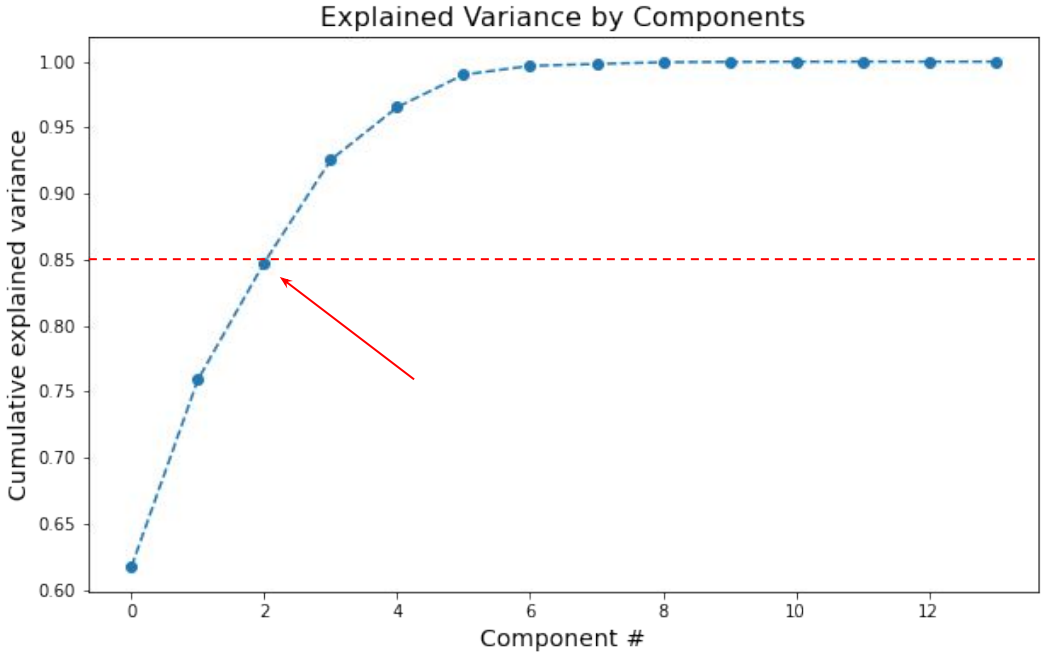
2. Applied two clustering algorithms to determine customer segments - KMeans and DBSCAN

3. Analyzed the clustering results to identify how the customer segments are different from one another

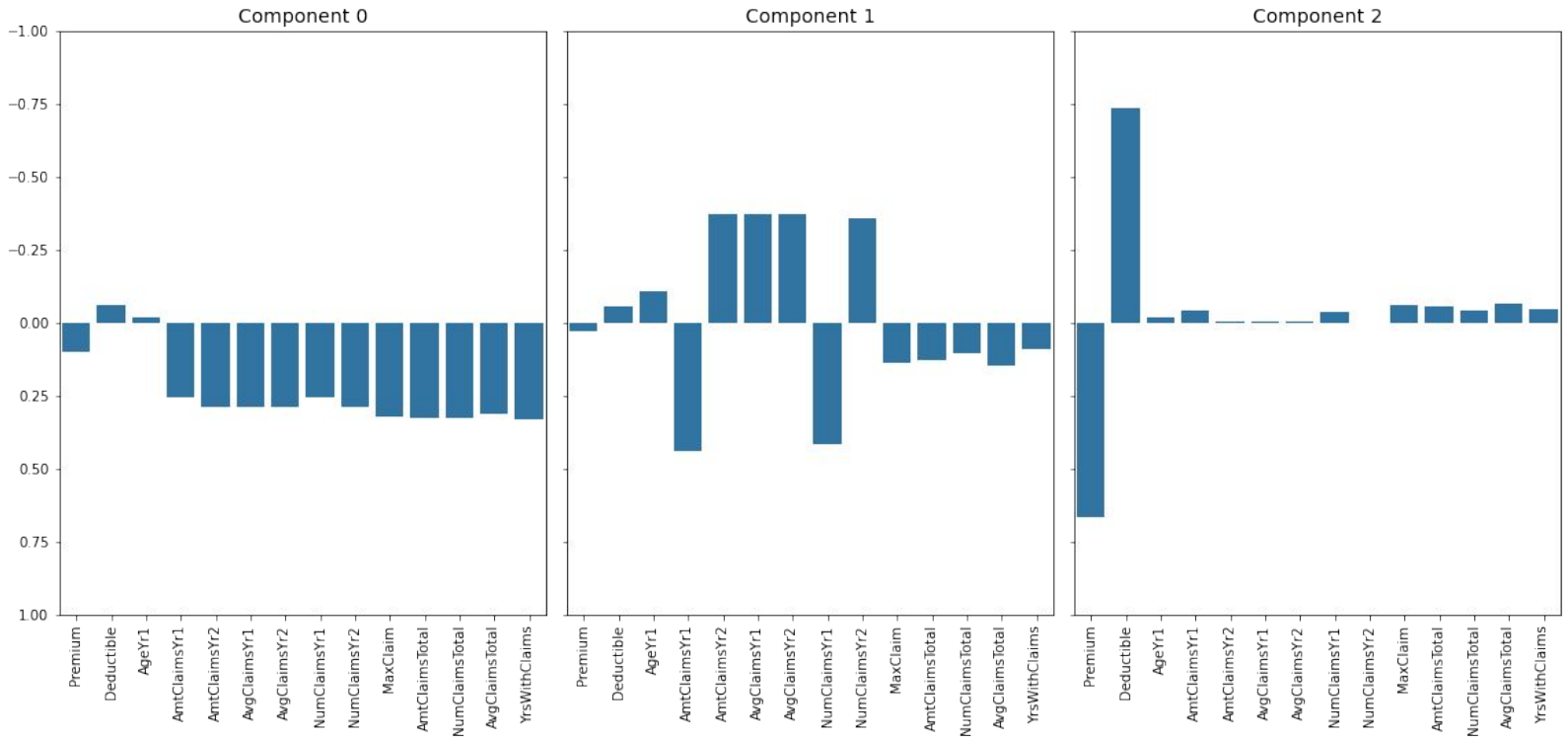
Principal Component Analysis (PCA)

Process:
Utilized PCA to determine which features contribute most to the variance in our data

Result:
~85% of the variance is explained in the first 3 principle components



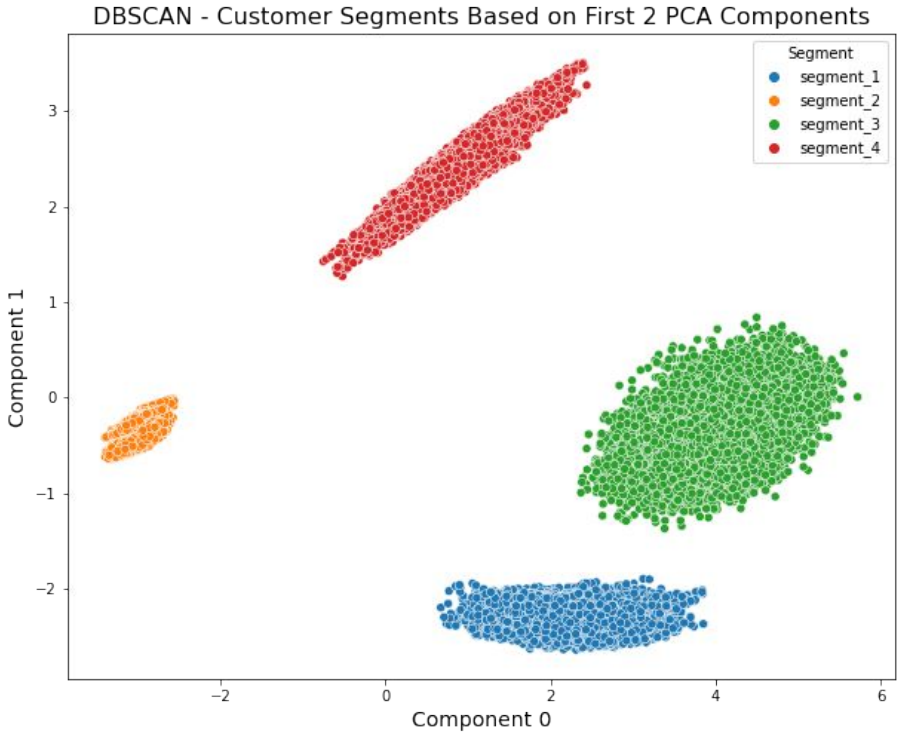
PCA Feature Importance - First 3 Components



Customer Segmentation Results - DBSCAN

Process:
Utilized DBSCAN to cluster data and determine customer segments

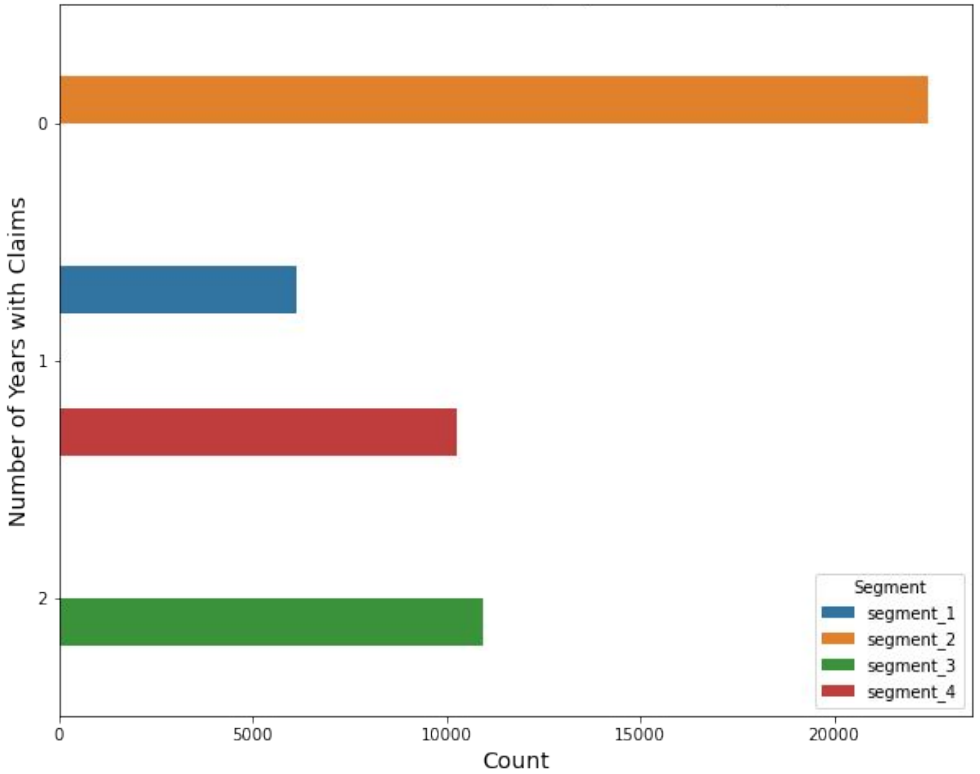
Result:
Plotting first 2 PCA Components shows 4 distinct clusters each mapping to its own segment



Customer Segments and Number of Years with Claims

Key Findings:
Clustering results map to a customer's number of years with claims in the first two policy years

Customers with claims in only one year are further segmented based on the year in which the claims took place



Clustering Results - How do the segments differ?

Customer Segment	Claims in Year 1	Claims in Year 2
Segment 1	No	Yes
Segment 2	No	No
Segment 3	Yes	Yes
Segment 4	Yes	No

Clustering Results - Marketing Opportunities

Marketing Goal #1:
Demonstrate value to customers with no claims, or no recent claims

Opportunity:
Target customers in Segments 2 and 4

Marketing Goal #2:
Support customers with consistent claims or whose claims are on the rise

Opportunity:
Target customers in Segments 1 and 3

Future Research

Opportunities for improvement:

- Consider splitting into two models (by Species)
- Add more supporting data
  - Owner demographics and CLV
  - Species / Breed info
- Engineer additional features
  - Claims timing
  - Breed-related ‘risk index’
- Evaluate additional clustering methods