



Data Analyst Technical Assignment

Background:

CoverTree is in an amazing growth phase where sales are growing by 30% MoM. We've sold policies via different distribution channels: communities, independent agents, and direct to consumer. Most new sales (55%) are driven by independent agents, followed by communities (35%), and direct to consumer (10%). Covertree is operating in 15 states currently. Texas and Arizona account for more than 50% of all new policies sold. On average, CoverTree is selling 1000 new policies/month with an average premium of \$1000/year.

Most independent agents policies (70%) are sold by 10 agents out of the 100 agents using the platform. 16 different communities in 3 states are actively contributing to the volume, and most of the community policies (70%) are concentrated in 8 out of 16 communities. Each community on average has 110 homes.

Problem Statement

For Covertree to be successful, we need to identify multiple key metrics and insights based on the customer and the policies associated with the customer. These key metrics and insights will help us pinpoint the areas that we would want to focus our growth efforts along with finding parts of the business that we need to improve on. The more areas we can find that are creating major profits/deficits in our company, the more efficient we can make each department of the company.

Also, converting quotes into policy purchases is crucial for revenue generation and business growth. However, accurately predicting whether a quote will be purchased by a customer remains a challenge, often resulting in inefficient allocation of resources and missed revenue opportunities. Therefore, another objective of this project is to develop a classification model that can accurately predict whether a quote will be purchased by a customer based on the details related to the quotes.

Requirements:

We've attached a sample dataset of what quotes and policies can look like in the business. With the dataset, we'd like you to:

1. Data Analysis:

- a. Extract the data, transform, and load it into a database of your choice.
- b. Create key metrics and insights in the visualization tool of your choice based on the data
- c. Report your findings

2. Machine Learning:

- a. Clean the test data for training model
- b. Split the data into a training dataset and a test dataset
- c. Train the model, based on the model of your choosing, on the training dataset and provide accuracy on the test dataset
- d. Report your findings