The overall goal of this project is to develop a suite of predictive tools for choosing the best model and for estimating the impact of the proposed model improvement on the overall ecosystem. The final product will be a tool that takes input of actual and proposed premium, considering different insurance policies along with historical premium and loss data, then returns expected estimates and exhibits. The main tech stack of the project would focus on using Polars in Python for extensive data processing as well as utilizing the AWS tools for implementing and developing the final product.

RESEARCH METHODOLOGY

**Prototype:**
- Pie-Exhibit Generation Module: Test data on a smaller scale to modularize it and using window functions to gather a better understanding of the dataset.
- Pie-Conversion Module: Summarize the limit/deductible columns to the quote level and find aggregated statistics. Calculate the quoted-rate based on those statistics and find average 6-month premium per vehicle.

**Model Building:**
- We build model with three levels of complexity:
  - Placeholder model: baseline predictive model (currently in use)
  - Logistic GLM model: predictive model with linear assumptions
  - Logistic GAM model: predictive model with non-linear assumptions
- Input: Pre-processing conversion data
- Model evaluation:
  - MSE: we use MSE as a criteria that measures the average of square of error, with less MSE illustrates better fitting.
  - Cross Validation: we use Cross Validation to assess how well a predictive model will generalize to new data, essentially partitioning the data set into subsets then evaluating it on “unseen” data. This is crucial for choosing the best model and to avoid issues with over fitting.

PROJECT SUMMARY

- Implemented functionalities and data flow in Exhibit generation module to generate estimates and exhibits illustrating the impact of proposed models.
- Implemented pre-processing data flow to be used for model training for New Business Conversion module, which predicts probability that a customer will make a purchase at a given price.

FUTURE WORKS

- Establish an automated data pipeline to ensure continuous data updates, maintaining the relevance and accuracy of the predictive models.
- Implement model to predict probability that a customer facing renewal under a given price change will renew at new price for Term-to-Term Renewal Module
- Implement model to predict probability that a customer that has purchased will remain until their next renewal for Intra-Term Retention Module