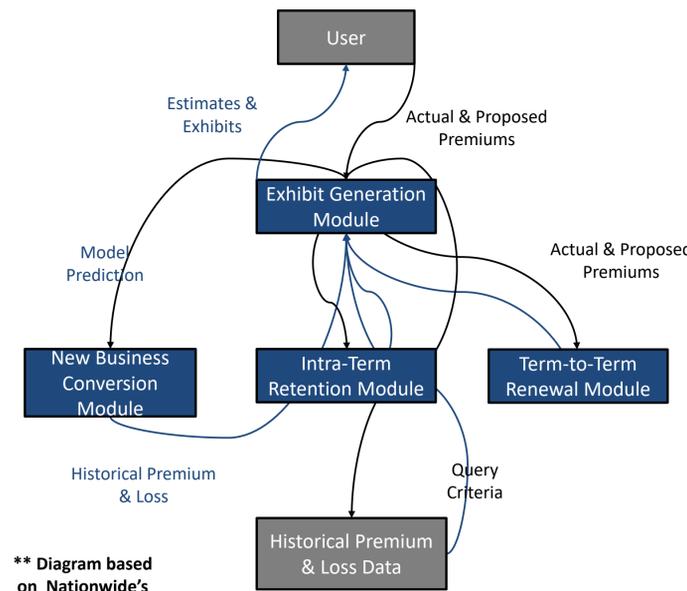


## PROJECT INTRODUCTION

*(Hypothetical exercise. Not actually used in ratemaking.)*

- The overall goal of this project is to develop a suite of predictive models focusing on different car insurance to create the Pricing Impact Estimator (PIE), used to reduce errors and estimate 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.



\*\* Diagram based on Nationwide's provided materials

Figure 1: Project Overview's Flowchart

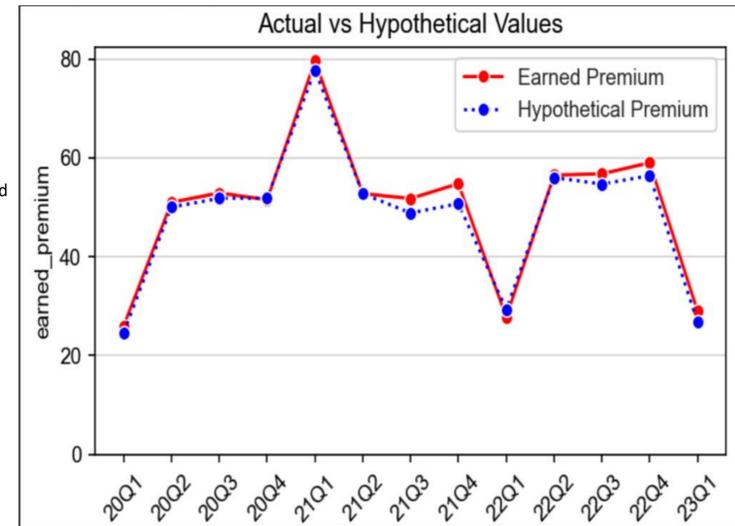


Figure 2: Example of visualization exhibit

## PIE-EXHIBIT GENERATION MODULE

- A significant portion of this module focused on data manipulation, which was carried out to exemplify the structural composition of the data. (Figure 3)
- Implement the calculation of various metrics, as well as visualization exhibits with PDF export functionality to illustrate relationship among factors of interest. (Figure 2)

## PIE-CONVERSION MODULE

Implement data preprocessing flow that utilizes as input the historical dataset detailing the customer purchase behaviors at specific premium levels

### Preprocessing tasks:

- Summarize the limit/deductible columns to the quote level and find the aggregated statistics.
- Calculate the quoted rate based on those statistics and find the average 6-month premium per vehicle.

### Output:

- Estimate and predict the quote-to-policy conversion rate
- Generate displays that depict the cumulative effects on pivotal strategic metrics.

## 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 the 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.

Given Data					Computed Data		Result		
Policy ID	Quarter	Earned Premium	Earned Car Year (Expo)	Gross Loss	Premium Rate	Adjusted Factor	Adjusted Premium	Adjusted Exposure	Adjusted Loss
A	2020Q1	25	0.125	0	200	0.978	24.45	0.12	0
A	2020Q2	50	0.25	100	200	0.978	48.91	0.24	97.81
A	2020Q3	25.63	0.125	0	205	0.978	25.06	0.12	0

Figure 3: Example of Data Manipulation's Structure

## ACKNOWLEDGEMENTS

We want to have special thanks to our CRP mentors Nick LeClaire, Andy Stevens and Josh Millis for this valuable opportunity to work on this project. The project was also conducted with the support and guidance from Professor Ward, our TA Quan Nguyen Hien and the Data Mine Staffs, including Nicholas Lenfestey, Josh Winchester, Maggie Betz and more.

## PROJECT SUMMARY

- Implemented functionalities and data flow in Exhibit generation module to generate estimates and exhibits illustrating the impact of proposed models.
- Implemented preprocessing 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