# Improving Diversified Pricing Strategy with Machine Learning

INTEGRATED PROJECT | SUMMER 2024

Team Arocha
Orange Ao | Luyang Feng | Ruizhe Qiu

August 9, 2024

# Table of Contents

# Executive Summary

By using a combination of techniques of machine learning and exposure rating, we are able to help Columbia Insurance Company (CIC), a fictitious insurance company, classify the risk levels in their product lines and create an improved pricing algorithm for its products. The business issue that CIC currently faces, the specifics of our solution and analysis, and recommendations and future considerations are provided in this report.

In the past, CIC has mostly employed aggregate loss modeling and Monte Carlo simulation for its pricing strategy, with satisfactory results. However, the same pricing strategy has seen a decline in performance metrics with the introduction of the new product lines in commercial properties, auto, marine cargo, and aviation hull since 2018. This decline can be attributed to the insufficient claim experience in the new product lines, as experience rating typically requires three years of loss data.

In cases where there is a want of claim experience to perform experience rating, exposure rating becomes an invaluable tool for us to look at the claim experience of another portfolio of the same kind. In using exposure rating, we can derive the missing claim experience from a reference portfolio whose claim experience is sufficiently supported statistically, and we can use the reference portfolio to derive a better loss distribution for the insurance products. In our analysis, the reference portfolio we employed comes from the Maxwell-Boltzmann, Bose-Einstein, and Fermi-Dirac distribution (MBBEFD). The MBBEFD is a well known distribution in statistical mechanics, and it's well adapted at modeling loss on the interval $[0,1]$ and on the interval $[0,\infty]$. When the parameters of 1.5, 2, 3, 4, respectively, are put into the MBBEFD distribution, the outputs are the exposure curves that approximate the four SwissRe curves used in practice by professionals to model loss. The four SwissRe curves will be referred to as Y1, Y2, Y3, and Y4, with Y1 being the least risky and Y4 being the most risky, and it is these four curves that will be used as the exposure curves to model the loss distribution of CIC's products.

In the industry, the standard practice is to look at the probable maximum loss (PML), namely, the maximum loss an insurer would expect to incur on a policy, and assign the exposure curves based on arbitrarily decided ranges (e.g. 0-4 million in Y1; 4-10 million in Y2, etc). It is our opinion that this method falls short of the standards of work we aim to achieve. As such, we have opted to include in the extra step of creating a clustering model to first classify the risk classes for each policy, and then use the aforementioned exposure curves for pricing. Our clustering model uses features such as property value, Probable Maximum Loss (PML), deductible, and the Coefficient of Variation (CV) of the claim amount of each property type. More details on the model are included in the details of work performed section.

Through our analysis, we have found that four risk classifications work best in accurately representing the risk levels of CIC's product lines. We have also developed a pricing model in python, wherein CIC can input new polices' property type, property value, PML, deductible, and receive an output of layer price (i.e. the price of claims between a lower and an upper bound).

To improve the quality of our risk classification model, we recommend CIC collect and provide more comprehensive claims data, including geographic location (high risk, hurricane, tornado, proximity to high risk area), age of property, in the case of cars and aviation, age of drivers, and in the case of house, construction type, in the case of buildings, number of floors, property age, etc. In general, the more granular the data provided, the better the clustering model and the pricing model downstream perform. In addition, we also recommend that CIC continuously monitor claim experience and update the model every two years using new claim data to keep up with any changes in the market.

Following the completion of our final presentation, we have conducted a thorough review of the Actuarial Standard of Practice No. 56 (ASOP 56) on Modeling. We have since verified that our procedures are in full compliance with these guidelines.

# Business Background

Columbia Insurance Company, which started its operation in 2004, has been a small insurance company that primarily offers homeowner's insurance. In the past, CIC has used aggregate loss modeling and Monte Carlo simulation as part of its pricing strategy. The company has seen significant growth in its homeowner's product. As such, CIC has decided to open up new product lines in commercial property, auto, aviation hull, and marine cargo in 2018. Since the introduction of the new product lines, however, the experience rating pricing strategy that CIC traditionally used has been performing suboptimally, and the company has seen a decrease in profitability.

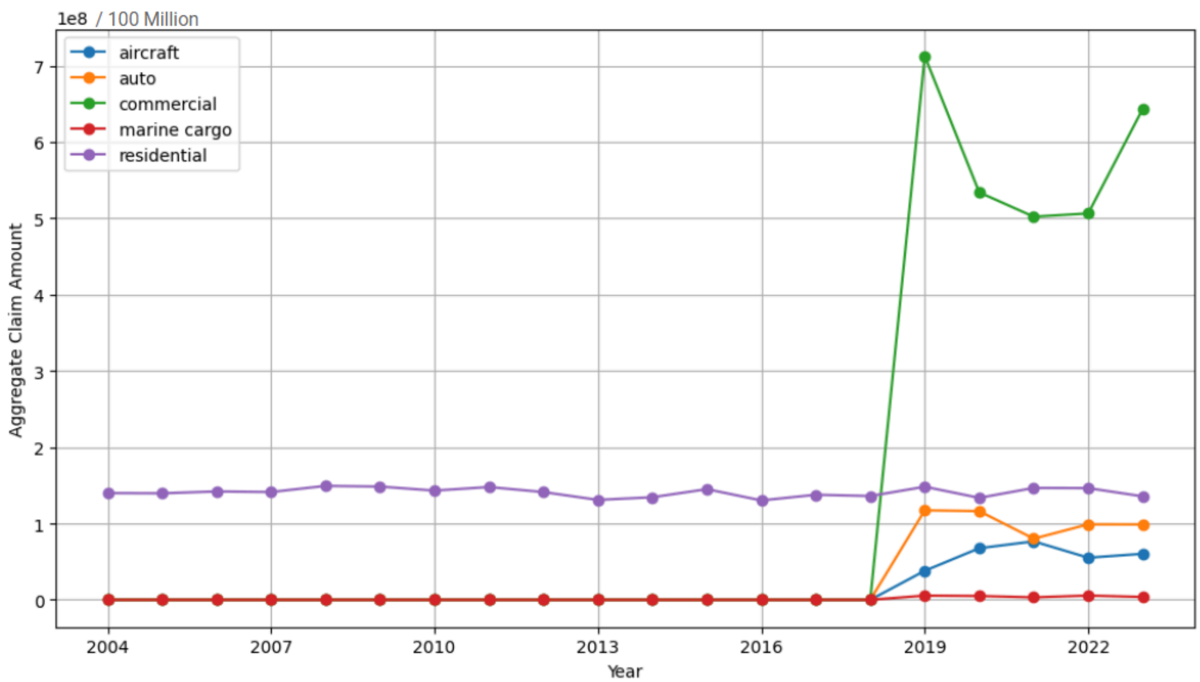


Figure 1: Aggregate claim amount for each product line by year

Figure 1 shows the aggregate claim amount for each product line from 2004 to 2023. Note that only residential (i.e. homeowner's insurance), as indicated by the purple line, has value above zero prior to 2018, as the other product lines were first introduced in 2018. We can see that the residential property line has been mostly steady and showing little growth over the past 20 years. There is no evidence to suggest that the new product line has grown much either.

This stagnant growth trajectory for these product lines are particularly problematic when compared to the market landscape. In contrast to CIC's performance, the forecasted growth rates for various insurance markets are looking notably optimistic. The global marine cargo insurance market is projected to grow at a compound annual growth rate (CAGR) of 5.86%, reaching US$34.0 billion by 2032. Similarly, the aircraft insurance market is expected to expand at a CAGR of 2.98%, increasing from USD 14,505 million in 2022 to USD 18,893 million by 2031. Meanwhile, the auto insurance market, valued at USD 652.5 billion in 2021, is anticipated to grow at a robust CAGR of 8.7%, reaching USD 1,383 billion by 2030. Furthermore, the commercial property insurance sector is projected to achieve a CAGR of 11.3%, with its market size soaring from US$254.9 billion in 2022 to US$724 billion by 2032.

In light of the challenging landscape faced by CIC, our team was hired to develop an alternative pricing strategy aimed at revitalizing CIC's profitability. The insurance market as a whole is experiencing significant growth, with sectors like marine cargo, aviation hull, auto, and commercial property insurance all forecasting impressive compound annual growth rates. These figures underscore a positive outlook for these sectors, highlighting the problematic nature of CIC's stagnant growth compared to their dynamic expansion.

# Goals of the Project

## Project Goal 1: Transition from Experience Rating to Exposure Rating

To help CIC find a solution to its business issue, it's essential to first identify the root causes of these challenges. Our analysis indicates that CIC's reliance on experience rating is a significant factor contributing to its current difficulties. Although experience rating has historically performed well for CIC's homeowner insurance products, it requires at least three to five years of loss data to be effective. This reliance on experience rating has become problematic for CIC's new property lines, where sufficient historical loss data is unavailable, leading to inaccurate pricing models and financial performance issues.

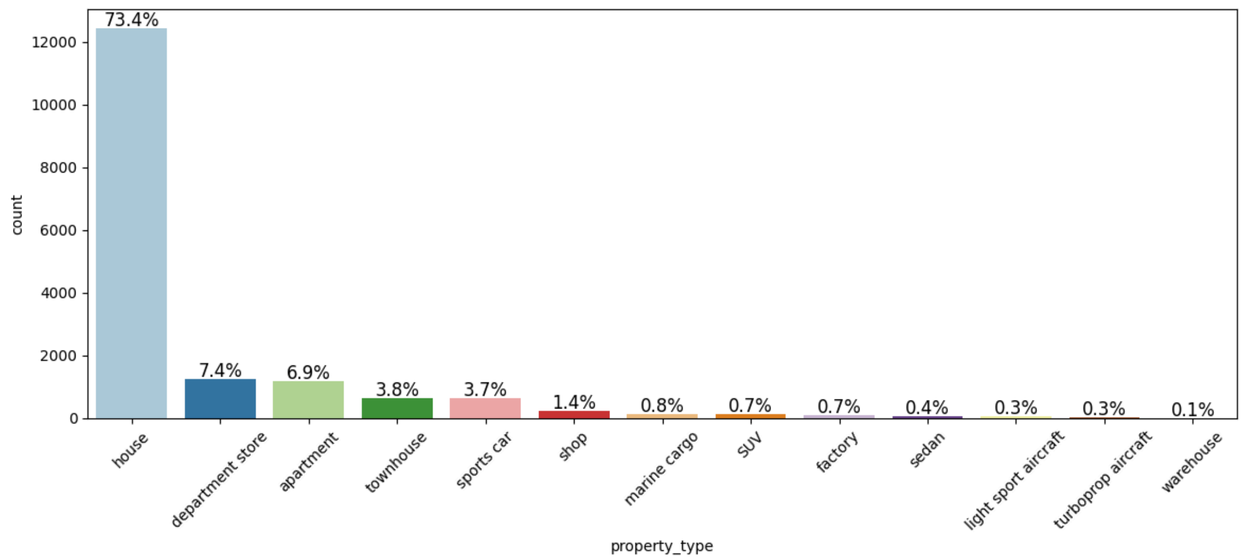


Figure 2: CIC's Proportion of Products

Figure 2 above shows that the majority of CIC's products are in homeowner's insurance. To address this lack of claim experience data in CIC's new products, our first project objective is to develop a method for utilizing exposure rating. Exposure rating serves as an excellent supplement to experience rating and becomes vital as an alternative in situations where there is a deficiency of claim data, as is the case with CIC's newer offerings. By using exposure rating, we aim to derive the missing claim experience from a reference portfolio that is statistically robust, thereby allowing CIC to make more informed pricing decisions.

For this purpose, we will employ industry loss experience developed by Swiss Re, a leading global reinsurer. The Swiss Re Exposure Curves will serve as our reference portfolio. These curves are derived from the Maxwell-Boltzmann, Bose-Einstein, and Fermi-Dirac distribution, collectively known as the MBBEFD distribution. Widely recognized in physics, the MBBEFD distribution is adept at modeling loss and provides a strong foundation for insurance pricing models.

By inputting specific parameters into the MBBEFD distribution, we can approximate the Swiss Re exposure curves, labeled as Y1, Y2, Y3, and Y4. These curves represent different levels of risk. The Swiss Re Exposure Curves have been empirically validated to accurately model property losses based on real-world data. These four curves will serve as the cornerstone for modeling our loss distribution, ensuring that CIC's pricing strategy is both precise and competitive.

In summary, by transitioning to exposure rating and leveraging the Swiss Re Exposure Curves, CIC can develop a more reliable pricing model that reflects true market conditions and diverse risk profiles. This approach will enable CIC to improve pricing accuracy, enhance competitiveness, and restore profitability across its new property lines.

## Project Goal 2: Implementing Advanced Risk Classification through Machine Learning

To effectively utilize exposure rating, it is crucial to first undertake a comprehensive risk classification process. The core concept of risk classes is that policies with similar exposure to a particular risk form a risk class. Each risk class is then modeled by a corresponding exposure curve, enabling more accurate risk assessment and pricing strategies. In the insurance industry, the standard practice for risk classification involves examining the Probable Maximum Loss (PML) and assigning risk classes based on that analysis. For instance, in a product line where policy PML ranges from 0 to 10 million dollars, policies with a PML from 0 to 2 million might be assigned to risk class one and modeled using the Swiss Re Y1 curve. Similarly, policies with a PML between 2 to 5 million could be assigned to risk class two and modeled using the Swiss Re Y2 curve. This method relies heavily on extensive datasets to accurately determine risk thresholds, which can present challenges when historical data is scarce, as is the case with CIC.

At Team Arocha, we acknowledge this limitation and are committed to devising a solution that extends beyond conventional methods. The second part of our project objective focuses on implementing unsupervised machine learning in our risk classification process to overcome the constraints posed by limited data. Specifically, we will utilize the K-Means clustering model, incorporating features such as property value, PML, deductible, and the coefficient of variation of claim amounts for each property type. Figures 3, 4, 5, and 6 below show our exploratory data analysis on our features.

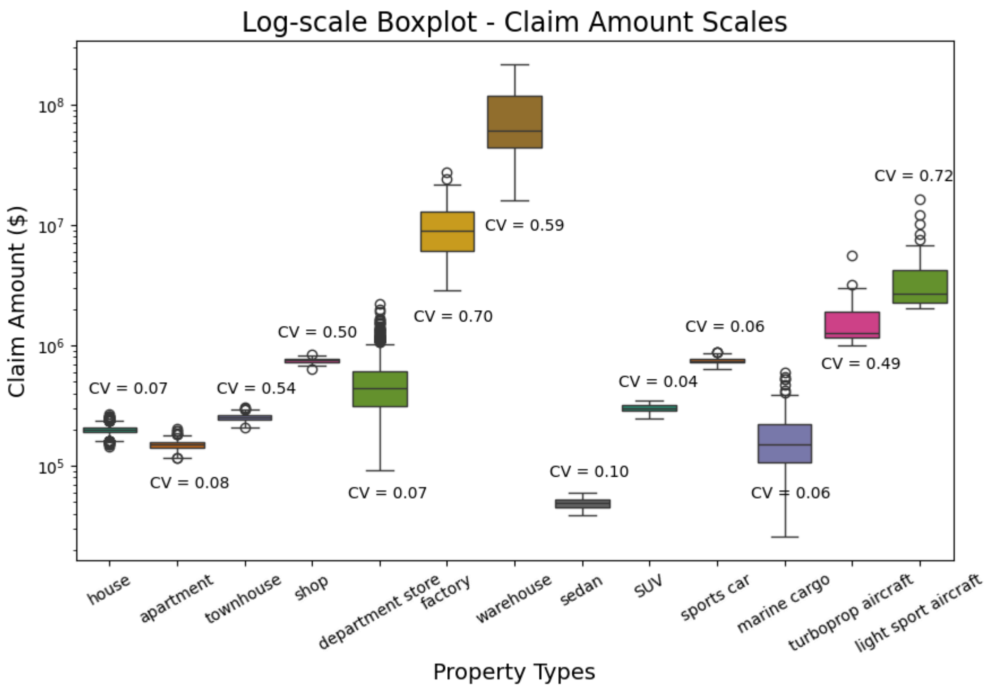


Figure 3: Coefficient of Variation (CV) of Claim Amount for Property Type

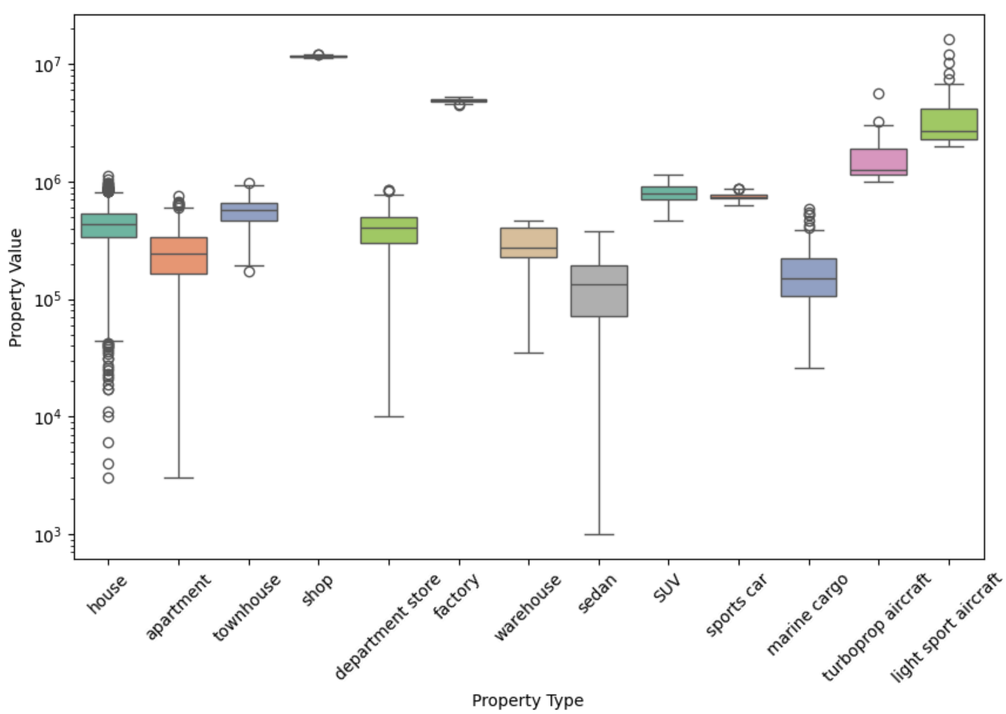


Figure 4: Boxplot of Property Value for each Property Type

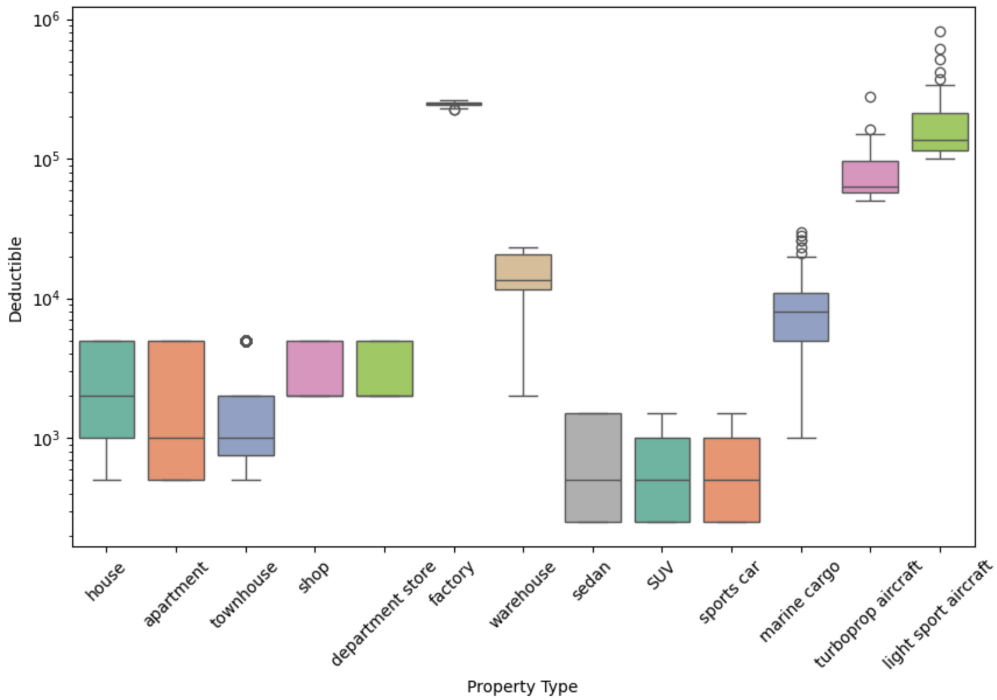

Figure 5: Boxplot of Deductible for each Property Type

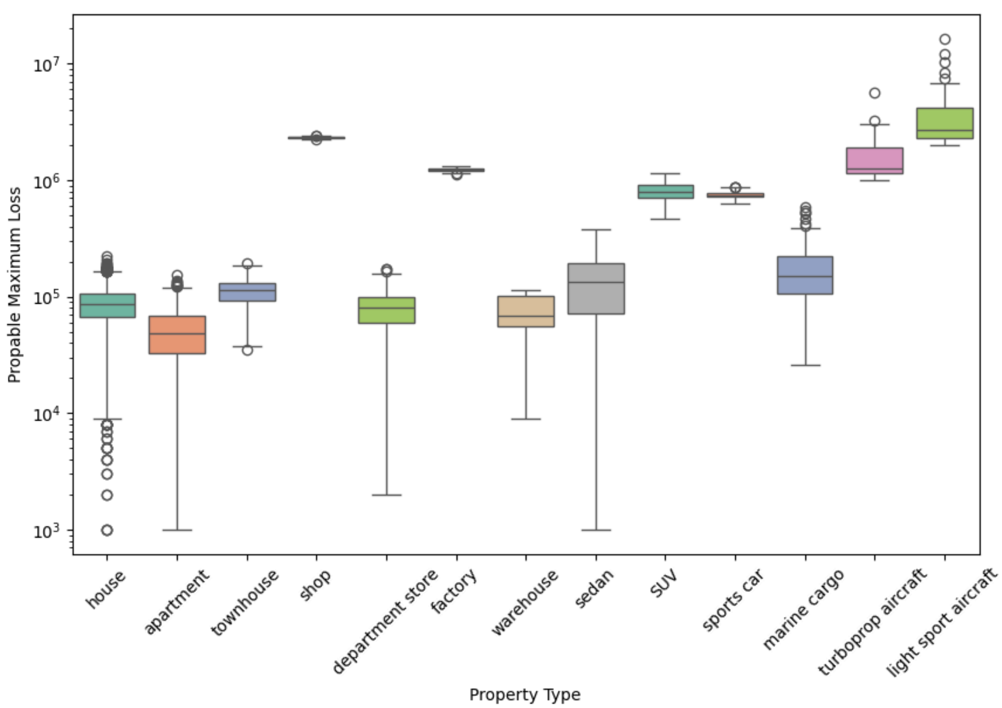

Figure 6: Boxplot of Probable Maximum Loss for each Property Type

This innovative approach allows us to leverage both the policy profile and claims profile comprehensively, circumventing CIC's data limitations and enabling more precise risk classification. By integrating these machine learning techniques, we aim to enhance the accuracy and effectiveness of CIC's exposure rating strategy, providing a robust framework for pricing that aligns with modern industry standards.

# Details of Work Performed

In this part, we will elaborate on the details of work performed, mainly focusing on the post-midterm work, i.e. the pricing processes. After Exploratory Data Analysis performed prior midterm, we started developing the pricing algorithm, which consists of 2 major steps: 1) *K-Means Clustering* for identifying the risk levels, and 2) *Exposure Rating* for calculating the prices according to risk levels.

## K-Means Clustering: Identify Risk Levels

Our project commenced with the application of K-Means clustering to identify distinct risk clusters within CIC's diverse insurance product lines, specifically, 4 clusters. Then we ranked the 4 clusters identified as Risk Level 1, 2, 3, and 4, which are matched to the 4 Swiss Re Curves Y1, Y2, Y3, and Y4, respectively. This process was carried out in the following 4 steps:

1. **Feature Extraction: The Selection of Risk Sensitive Features**

   The process of feature extraction is pivotal in any machine learning project, as it involves identifying and selecting the most relevant variables that contribute to the predictive power of the model. For our project with CIC, we chose four specific features from the data provided by our client, based on a thorough understanding of the nature of risk assessment in the property insurance: *Property Value*, *Probable Maximum Loss (PML)*, *Deductible*, and *Coefficient of Variation (CV)*[1] for each property type.

   1.1. ***Property Value.*** Since higher-value properties typically represent a higher risk and, consequently, higher potential claim amounts, this feature provides a direct measure of the insured value at risk, which is essential for calculating premiums and assessing risk.
   1.2. ***PML.*** As an estimate of the maximum potential loss that could arise from a single event, it considers all policy exposures.Thus by including *PML*, we can better differentiate between policies that have significantly different exposure levels.
   1.3. ***Deductible.*** Policies with higher deductibles generally transfer more risk to the policyholder, who agrees to pay more out-of-pocket before the insurance coverage kicks in, thus representing lower potential payouts for the insurer. This feature is essential for understanding the financial implications of a claim and how much risk the insurer carries.

[1] Coefficient of Variation, indicates how large the standard deviation is in relation to the mean. It is a measure of relative volatility. For more details, see: Wikipedia contributors, "Coefficient of variation," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Coefficient_of_variation (accessed August 7, 2024).

1.4. ***CV.*** In the context of insurance, a higher CV for claim amounts indicates greater variability and unpredictability in claims, which can be indicative of higher risk. By including the CV for each property type calculated from the historical data, we account for the inherent variability in claims across different types of properties, allowing for a more nuanced risk assessment.

## 2. Optimal K Determination: The Trade-Off Between Cluster Separation And Cohesion

Utilizing grid-searching, we explored values of K ranging from 2 to 9, to find the optimal K, i.e. number of clusters. For each K, we calculated the *Within-Cluster Sum of Squared Error (WCSSE)*[2], and the *Silhouette Score*[3] to evaluate the clustering goodness, visualized through an elbow plot and a Silhouette Score plot in Figure 7.

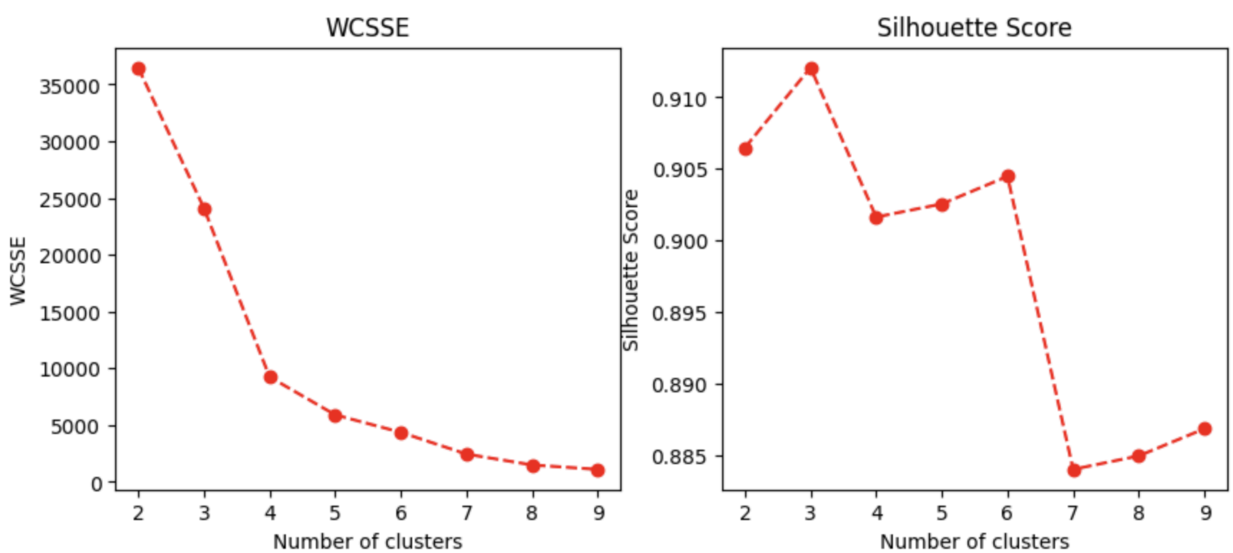


Figure 7: Trade-Off Between WCSSE And Silhouette Score

The *WCSSE* is a measure of the compactness of the clusters. It calculates the sum of the squared distances between each point in a cluster and its centroid. Conventionally, the goal in K-Means clustering is to minimize this value, as a lower WCSSE indicates that the points within a cluster are closer to each other, which implies better-defined and more distinct clusters. By minimizing the *WCSSE*, we aimed to ensure that each policy within a

[2] Within-Cluster Sum of Squared Error, i.e. variance, is the metric used by the "elbow" method for clustering model evaluation. Refer to "Elbow method (clustering)" Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Elbow_method_(clustering) (accessed August 7, 2024).

[3] Wikipedia contributors, "Silhouette (clustering)," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Silhouette_(clustering) (accessed August 7, 2024).

cluster is similar to others in the same group, reducing the variability within clusters and enhancing the model's ability to differentiate between different risk profiles.

The Silhouette Score is another measure used to evaluate the quality of clusters. A higher Silhouette Score indicates that the clusters are well-defined and well-separated. By maximizing this score, we aimed to ensure that each policy is not only close to the centroid of its own cluster but also sufficiently distant from the centroids of other clusters, reinforcing the distinctiveness of each risk class.

From the plot, our choice of K was narrowed down between 3 and 4. With the trade-off analysis, by increasing K from 3 to 4, the *WCSSE* was reduced by 62%, only at the cost of a drop of 1.2% in Silhouette Score. So the optimal K was ultimately decided as 4.

## 3. Risk Level Ranking And Curve Matching

Once the optimal K was determined, we ranked the risk levels of the four clusters by calculating the CV of claim amounts for each cluster. As explained in 1.4., the CV represents the  unpredictable risk in claim amounts. In other words, the higher the CV, the higher the risk level.

Then we matched these risk levels to corresponding Swiss Re Exposure Curves in Figure 8. Specifically, since a higher Loss Elimination Ratio (LER) means a greater proportion of loss will be transferred to (i.e. eliminated by) other risk parties (e.g. in the form of deductibles which transfers risk to the policy holders), the Swiss Re Curve Y4 (red curve) indicates the lowest risk, which matches the cluster with the lowest CV, i.e. cluster 3.

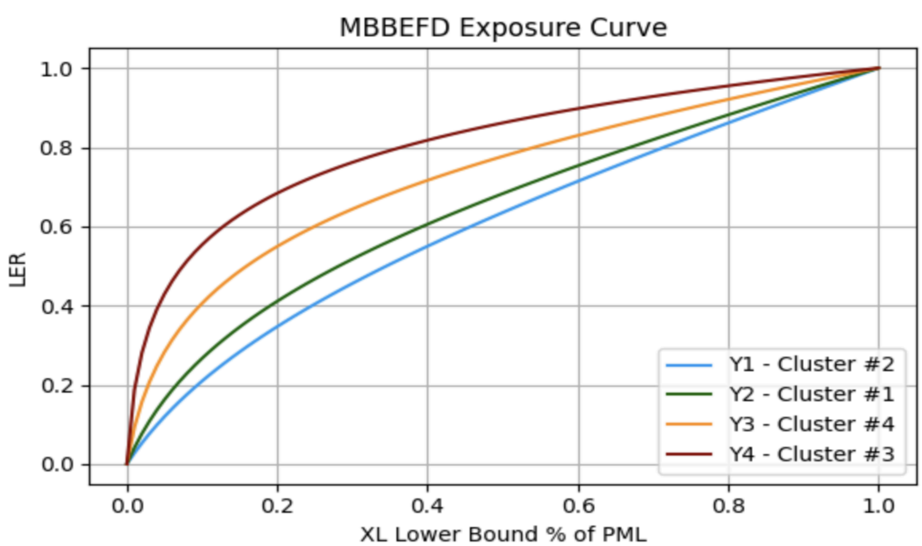


Figure 8: Swiss Re Curves

This matching process allowed us to categorize each policy into a risk class with a corresponding exposure curve for pricing. The results are shown in the following table.

| Cluster | Policy Count | CV For Claim Amounts | Swiss Re Curve |
|---|---|---|---|
| **1** | 15049 | 1.95 | Y2 |
| **2** | 240 | 23.44 | Y1 |
| **3** | 1432 | 0.13 | Y4 |
| **4** | 184 | 1.44 | Y3 |

Table 1: Cluster-Curve Matching

## Exposure Rating: Price New Policies

With risk clusters identified and exposure curves matched, we moved on to the *Exposure Rating* step, which involves pricing new policies based on their assigned risk levels and corresponding exposure curves. In this part, we will explain in detail: 1) How the pricing algorithm works; 2) From the client's perspective, how to use the algorithm to calculate reasonable prices for new policies.

The pricing algorithm requires 2 types of data inputs.

1) Features for identifying the risk level: *Property Value*, *Probable Maximum Loss (PML)*, *Deductible*, and *Property Type*[4].
2) The lower bound and upper bound of the *Excess of Loss (XL)*, as well as the preliminary premium (denoted as *premium*), which will be decided by the CIC itself for each new policy.

With all these inputs, the pricing algorithm works by the following steps and formulas.

### 1. Risk Level Identification

With the K-Means Clustering Model trained in the previous step and the first type of inputs, a risk level, and subsequently an exposure curve is assigned to the new policy.

[4] The algorithm will automatically calculate the *Coefficient of Variation (CV)* for a certain property type based on historical data.

2. **LER Determination**

The LER for a new policy is calculated from its PML and the lower bound of XL decided by CIC, following the formula below:

$$LER = SwissRe(\ min(1,\ \frac{XL\ Lower\ Bound}{PML})\ )$$

Here, $SwissRe(\bullet)$ represents the MBBEFD distribution function corresponding to a specific Swiss Re exposure curve[5].

3. **Layer Premium Calculation**

The price, or layer premium, for a new policy is then calculated by the formulas below:

$$Net\ Premium = min(1,\ \frac{XL\ Upper\ Bound}{PML}) \times premium$$

$$Layer\ Premium = Net\ Premium \times (1 - LER)$$

So briefly speaking, as the client, CIC only needs to input the *Property Value*, *Probable Maximum Loss (PML)*, *Deductible*, *Property Type, lower bound* and *upper bound of XL,* as well as *the premium* for a new policy. Then the pricing algorithm will automatically compute a price as the output. This process is illustrated by specific examples in the following flowchart (Figure 9). Note that since the second type of inputs will be actually decided by CIC itself, here we are using assumed numbers for the purpose of illustration.

**Input: New Policy** → **Output**

| New Policy ID | Property Type | Property Value | PML* | Deductible | XL Lower Bound | XL Upper Bound | Premium |
|---|---|---|---|---|---|---|---|
| 1 | factory | 5,083,000 | 1,271,000 | 254,000 | 300,000 | 1,000,000 | 20342 |
| 2 | SUV | 605,000 | 605,000 | 500 | 3,000 | 700,000 | 592 |
| 3 | sedan | 20,000 | 20,000 | 250 | 500 | 10,000 | 68 |
| ... | ... | ... | ... | ... | | | |

| Layer Premium |
|---|
| 643 |
| 567 |
| 56 |
| ... |

Figure 9: Flowchart of Illustration Examples Using Assumed Numbers

[5] https://www.casact.org/sites/default/files/2021-03/8_Bernegger.pdf

# Analysis of Results

## Project Goal 1: Transition from Experience Rating to Exposure Rating

The first goal, transitioning from experience rating to exposure rating, has been successfully achieved. Exposure rating allowed us to provide the layer premium to CIC. To ensure rigor, we developed a comprehensive strategy and algorithm to calculate the layer premium given the necessary variables. Specifically, we calculated the net premium based on the upper and lower bounds and the tentative premium, and then used the LER and net premium to determine the layer premium.

This methodology provides several benefits to CIC. First, it enables efficient calculation of premiums by standardizing the process, which reduces the time and effort required for pricing. Second, it offers flexibility by allowing CIC to adjust premiums according to specific situations, such as varying upper and lower bounds. This adaptability ensures that CIC can tailor its pricing strategies to meet the unique needs of different clients and scenarios, enhancing overall competitiveness and customer satisfaction.

Furthermore, the use of exposure rating aligns premiums more closely with the actual risk exposure, leading to more accurate and fair pricing. This transition from experience rating to exposure rating represents a significant improvement in pricing strategies, as it mitigates the limitations of relying solely on historical claims data. By focusing on current risk factors, CIC can better anticipate future claims and adjust premiums accordingly, resulting in more stable and predictable financial outcomes.

In conclusion, by implementing exposure rating and developing a robust framework for calculating layer premiums, we have effectively transitioned from experience rating to exposure rating. This achievement marks a significant milestone in improving CIC's pricing strategies, ensuring that they are more precise, flexible, and aligned with current risk exposures. This accomplishment not only enhances CIC's operational efficiency but also strengthens its ability to respond to market changes and meet client needs.

## Project Goal 2: Implementing Advanced Risk Classification through Machine Learning

The second goal, defining risk classes for both current claims and future potential policies, has also been successfully achieved. We utilized the K-Means clustering algorithm to classify claims into four distinct risk classes. This method involved grouping claims based on similarities in their characteristics, allowing us to identify patterns and differentiate between various levels of risk. To quantify the risk within each cluster, we calculated the coefficient of variation (CV) of the claim amounts. A higher CV indicates greater variability and thus a higher risk level within that cluster.

Moreover, we developed a robust clustering model designed to accommodate future policies. When CIC receives a new policy, the relevant information can be fed into this model, which will then predict the risk class of the new policy. This predictive capability ensures that CIC can continuously apply the same rigorous classification criteria to both existing and new policies, maintaining consistency in risk assessment.

This approach offers several advantages. First, it provides a clear and systematic method for defining risk classes based on historical data. By using K-Means clustering, we can objectively identify groups with similar risk profiles, reducing subjective bias in risk classification. Second, the calculation of CV for each cluster allows for a nuanced understanding of the risk levels, helping CIC to manage policies more effectively.

In conclusion, by employing K-Means clustering and developing a predictive model for future policies, we have successfully met the goal of defining risk classes. This achievement enhances CIC's ability to assess and manage risk with greater accuracy and efficiency. The systematic classification of risk and the predictive capabilities of the model ensure that CIC can continuously refine their risk management strategies, providing a robust foundation for future growth and stability.

## ASOP 56 (Modeling)

Following the completion of our final presentation, we embarked on an extensive review of Actuarial Standard of Practice No. 56 (ASOP 56) on Modeling. This involved a detailed examination of all relevant sections and guidelines to ensure a comprehensive understanding. After meticulously cross-referencing our procedures with the standards outlined in ASOP 56, we have confirmed that our modeling processes and methodologies are fully compliant with these rigorous guidelines. This verification process reinforces our commitment to adhering to the highest professional standards in actuarial practice.

# Recommendations & Lessons Learned

## Recommendations

To enhance CIC's business and pricing strategies, we recommend that CIC collect and provide more comprehensive data. This data should include the driver's age for auto and aviation policies, the number of floors and property ages for real estate policies, and geographical information for most policies. With these detailed features, we can perform a thorough rating factor analysis, such as using regression models to identify significant factors for various property types. This will enable us to offer deeper insights that can significantly improve CIC's operations. Additionally, the new data and the results from our rating factor analysis can be utilized to refine and enhance our risk clustering model, leading to more accurate risk assessments and better-informed pricing decisions.

Additionally, we also recommend that CIC continuously monitor claim experience and update the model every two years using new claim data to keep up with any changes in the market. Regularly updating the model will ensure that it remains accurate and reflective of current trends, allowing CIC to adapt to shifts in the industry and emerging risks. By incorporating the latest claim data, the model can provide more precise risk assessments and pricing strategies, enhancing CIC's ability to manage risk and maintain competitive advantage. This proactive approach to monitoring and updating will also help identify new patterns and anomalies early, enabling timely adjustments to policies and practices. Ultimately, a biennial update cycle will support CIC in maintaining a robust, data-driven strategy that is responsive to the evolving insurance landscape.

## Lessons Learned

Throughout this project, we learned several valuable lessons that can inform and improve our future work. Firstly, we recognized the limitation posed by low dimensionality. While a small number of features may sometimes provide sufficient information, they often restrict the scope of analysis and the insights that can be derived. To address this, one solution is to gather more comprehensive data. Expanding the dataset can unveil additional patterns and relationships that were previously hidden.

Secondly, we learned the importance of utilizing data more efficiently. One effective approach is to extract new features from existing ones. For example, in this project, we derived the coefficient of variation (CV) from the existing data. By incorporating the CV into our clustering model, we enhanced its ability to identify similarities between claims, leading to more accurate groupings. Additionally, the CV played a crucial role in determining the risk levels of different clusters, providing a clearer picture of risk distribution.

Thirdly, we discovered the significance of specialized tools. Often, there are tools developed specifically for certain problems, which can significantly streamline the process. Although it is possible to solve these problems using general methods, specialized tools offer efficiency and precision that general approaches may lack. Therefore, it is essential to continuously expand our knowledge and stay updated on the latest tools and technologies relevant to our field. This proactive approach enables us to tackle specialized problems more effectively and efficiently.

Lastly, we realized the importance of meticulous attention to detail in presentations. Effective communication with customers hinges on the clarity and precision of our presentations. Every detail, from the organization of content to the visual appeal of our slides, contributes to the overall impact of our message. A well-structured presentation, combined with engaging delivery, ensures that our audience comprehends and retains the key points. This not only determines the immediate success of a project but also influences the potential for long-term collaboration and trust with clients.

By incorporating these lessons into our future projects, we can enhance our analytical capabilities, improve our efficiency, and strengthen our communication strategies, ultimately leading to better outcomes and sustained success.

# Future Considerations

## Model Improvement

This part is about our future considerations for the project and CIC. The first aspect to focus on is model improvement. As mentioned earlier, if CIC can collect and provide more detailed data, we can enhance the clustering model. Detailed data, such as the driver's age for auto and aviation policies or the number of floors and property ages for real estate policies, will enable us to implement more nuanced and accurate clustering models.

Another action is testing various machine learning models for risk classes. By exploring different algorithms, we can identify the most effective techniques for accurately categorizing risk and predicting claims. This experimentation will help us understand which models perform best under different conditions and for various types of insurance policies.

Additionally, we can explore alternative pricing methods. This includes not only traditional actuarial approaches but also more innovative techniques such as dynamic pricing and usage-based pricing models. These methods can provide more flexible and customer-specific pricing structures, enhancing customer satisfaction and retention.

Furthermore, with an expanded dataset, we can implement comprehensive factor analysis to determine pricing with greater precision and deliver tailored insights for different insurance products. For example, we can build sophisticated regression models to identify loss-sensitive features. This means analyzing the data to pinpoint which factors most significantly impact the likelihood and magnitude of claims. Understanding these relationships can help in better risk management and more effective pricing strategies. Additionally, we can develop regression models to estimate tentative premiums, providing a preliminary indication of pricing based on the identified loss-sensitive features. This will streamline the underwriting process and offer quicker, data-driven pricing estimates.

Lastly, we recommend ongoing monitoring of the CIC portfolio performance to ensure its continued alignment with our strategic objectives. This involves regularly reviewing key performance indicators, conducting periodic assessments, and making necessary adjustments based on the latest data and market conditions. By maintaining a proactive approach, we can swiftly identify any emerging trends or anomalies and update our model accordingly. This will help us to stay ahead of potential risks and capitalize on new opportunities, thereby optimizing the performance of the portfolio and model over time.

Overall, these future considerations aim to leverage detailed data and advanced analytical techniques to continuously refine and enhance CIC's risk assessment and pricing strategies. By adopting these approaches, CIC can maintain its competitive edge, improve operational efficiency, and offer more accurate and personalized insurance solutions to its customers.

## Market Monitoring

The second part is market monitoring. The market and environment are in a constant state of flux, and it is crucial for CIC to stay vigilant to maintain its competitive edge. For instance, environmental and climate factors play a significant role in the property and casualty (P&C) insurance industry. Changes in weather patterns, frequency of natural disasters, and long-term climate shifts can have profound impacts on insurance claims and risk assessment. Therefore, it is essential for CIC to continuously monitor environmental data and integrate it into their risk models and pricing strategies.

Similarly, keeping a close watch on market trends is vital. Market trends can be influenced by a multitude of factors, including economic conditions, technological advancements, and consumer behavior shifts. By being attuned to these changes, CIC can anticipate emerging opportunities and threats, allowing them to adjust their strategies proactively. For example, the rise of autonomous vehicles may alter the landscape of auto insurance, and staying ahead of such trends will enable CIC to develop innovative products and services that meet new market demands.

Competitor analysis is another critical aspect of market monitoring. Conducting thorough competitor analysis will help CIC identify areas for improvement and innovation. By understanding the strengths and weaknesses of competitors, CIC can learn from their successes and avoid their pitfalls. This analysis can provide insights into market positioning, pricing strategies, and customer service practices that can be leveraged to enhance CIC's offerings. For instance, if a competitor introduces a successful new product or feature, CIC can evaluate the potential benefits of adopting a similar approach or developing a unique alternative that better serves their customers.

Lastly, staying abreast of insurance regulations and laws is essential to ensure compliance and anticipate any changes that could affect operations. The regulatory landscape for insurance is complex and constantly evolving. Keeping informed about new regulations, compliance requirements, and legal changes will help CIC avoid penalties and maintain operational integrity. Additionally, understanding regulatory trends can provide strategic advantages, such as identifying new market opportunities created by regulatory shifts or preparing for potential challenges posed by stricter regulations.

In conclusion, by continuously monitoring environmental factors, market trends, competitor activities, and regulatory changes, CIC can remain agile and responsive to the dynamic nature of the insurance industry. This proactive approach will not only help CIC stay competitive but also enable them to innovate, improve customer satisfaction, and ensure long-term success.

# References


Actuarial Standards of Practice
https://www.actuarialstandardsboard.org/standards-of-practice/

Exposure rating
https://www.swissre.com/dam/jcr:7137dac0-83a6-4cfa-80a4-93d33c35562f/exposure-rating-brochure.pdf

Number of Years of Claim Experience Needed for Experience Ratings
https://www.ncci.com/Articles/Documents/UW_ABC_Exp_Rating.pdf

MBBEFD
https://www.casact.org/sites/default/files/2021-03/8_Bernegger.pdf

Loss sensitive features
https://www.actuaries.org.uk/system/files/documents/pdf/mata_0.pdf

Marine Cargo Growth
https://finance.yahoo.com/news/marine-cargo-insurance-market-set-133000167.html

Commercial Property Growth
https://www.insurancebusinessmag.com/us/news/breaking-news/commercial-property-insurance-market-to-hit-us724bn-by-2032-458055.aspx#:~:text=The%20commercial%20property%20insurance%20industry,report%20from%20Allied%20Market%20Research.

Auto Growth
https://straitsresearch.com/report/auto-insurance-market#:~:text=The%20global%20auto%20insurance%20market,of%20an%20accident%20or%20theft.

Aviation Growth
https://finance.yahoo.com/news/aircraft-insurance-market-size-expected-143500313.html