

**Integrated Project 2024**

**Actuarial Science Program**

**Columbia University**

**July 9, 2024**

STATEMENT OF WORK (FINAL)

Improving Diversified Pricing Strategy with Machine Learning

# Team members

|  | **Name** | **Tel** | **Email** |
| --- | --- | --- | --- |
| **Mentor** | Carlos Arocha | 41-78-607-24-92 | ca@ArochaAndAssociates.ch |
|
| **Instructors** | Tom Murphy | 201-797-2324 | tm2980@columbia.edu |
| Ira Kastrinsky | 201-390-1639 | [ik2379@columbia.edu](mailto:ik2379@columbia.edu) |
| **Group Members** | Luyang Feng | 240-330-5452 | lf2786@columbia.edu |
| Orange Ao | 917-518-4578 | ya2538@columbia.edu |
| Ruizhe Qiu | 857-777-9603 | rq2194@columbia.edu |
|

# Introduction

Over the past few years, Columbia Insurance Company (CIC), a small property insurer, has experienced significant growth. Initially, CIC focused solely on homeowners’ insurance, but it has now diversified its product palette to include commercial and industrial property, auto, marine cargo, aviation hull, and other property lines. Their pricing strategy relied on aggregate loss modeling using Monte Carlo simulation. But now, with multiple products and different risk profiles, CIC needs to update its pricing strategy to stay competitive in the property insurance market. With this goal in mind, CIC will redefine its risk classes and charge premiums commensurate with the different risk profiles.

However, due to a lack of actuarial resources, CIC is seeking external consultants to propose new pricing protocols. Following initial meetings, it has been agreed that a potential solution may be using exposure rating techniques. However, before this, unsupervised machine learning will be used to extract information from claims data.

# Objectives

Develop and implement a pricing strategy for Columbia Insurance Company (CIC) by creating a robust risk classification system and suitable premium calculation models. Leveraging unsupervised machine learning techniques and exposure rating methods, the objective is to enhance the pricing of property insurance products and ensure premiums are commensurate with the associated risks.

Specifically, we aim to:

1. Redefine Risk Classes: Use unsupervised machine learning such as K-Means clustering to analyze historical claims data and identify distinct risk classes for various types of property insurance, including homeowners’, commercial, industrial, auto, marine cargo, and aviation hull insurance.
2. Develop Exposure Ratings: Use exposure rating techniques to estimate the expected losses for each risk class. This involves building severity curves based on relevant data such as property value, location, type of property, and cause of loss.
3. Create More Refined Pricing Models: Design and validate pricing models that incorporate the new risk classes and exposure ratings. At the first stage, calculate premiums based on exposure curves, adjusting for administrative expenses, and profit margins.
4. Implement and Monitor: We aim to implement the new pricing models across CIC’s product lines and monitor their performance to ensure they remain accurate and reliable. We recommend updating the models every 2–3 years with new data to adapt to changing risk profiles and market conditions.

# Scope of services

The completion of the project includes the following steps:

Request data and ensure that data complies with our request June X – June Y

Analysis

* Exploratory Data Analysis (EDA)
  + Historical Claims Data:
    - Gather comprehensive historical claims data for the various types of property insurance: homeowners’, commercial, industrial, auto, marine cargo, and aviation hull insurance
  + Data Cleaning:
    - Handle missing values by omitting/ interpolation
    - Detect outliers and inconsistencies
    - Standardize and normalize the data as necessary
  + Feature Engineering:
    - Perform correlation analysis and create a correlation heat map
    - Extract relevant features such as property value, location, type of property, cause of loss, claim amounts, and other pertinent variables
    - Using Principal Component Analysis (PCA) to reduce dimension if there are too many features
* Unsupervised Machine Learning for Risk Classification
  + Clustering Techniques: Determine The Number of Risk Classes
    - K-Means Clustering: Determine the optimal K (number of clusters) using Cross-Validation and Grid Search
    - Gaussian Mixture Models (GMM): if the K-Means algorithm (hard assignment) does not yield satisfactory results, try a soft assigning classifier
  + Feature Analysis:
    - Identify key features that contribute to the differentiation of risk classes
* Exposure Ratings
  + Severity Curves:
    - Build severity curves for each identified risk class
    - Alternatively, use industry exposure curves
    - Estimate the expected losses based on property value, location, type of property, and trigger of loss
  + Loss Distributions:
    - Fit appropriate loss distribution models to the claims data: Gamma, Log-normal, Weibull, Pareto, Exponential, etc.
  + Exposure Rating Techniques:
    - Apply exposure rating methods to estimate expected losses for each risk class
    - Calculating loss costs
    - Adjust for various factors like deductibles, limits, and inflation.

Reporting

* Mid-term presentation (July 1)
  + We will summarize the project’s introduction, our objectives, the initial EDA, as well as a report on our progress with the project and the next steps.
* Final presentation (August 6)
  + We will provide a summary of all the final report elements in the final presentation [PowerPoint Slides].
* Final report (August 9)
  + We will include in the final report the following elements:
    - Introduction and scope
    - Executive summary
    - Detailed model description
    - Development of assumptions
    - Lessons learned
    - Readings
    - Appendices (if applicable), which might contain detailed results
    - Footnotes

# 

# Timetable

* A midterm presentation showing our progress will be delivered on July 1, 2024.
* A final presentation to senior management of CIC will be delivered on August 6, 2024.
* A report including all the aforementioned elements will be delivered on August 9, 2024.

The final version of this Statement of Work will include a spreadsheet with a detailed timeline.

# Handling of proprietary information

Our consultancy undertakes to maintain stringent secrecy and not to disclose to any third parties, during the lifetime of the project and after the project has ended, any documents or information related to CIC.

Our actuaries adhere to the Actuarial Standard of Practice No. 23 (ASOP 23) of the Actuarial Standards Board.

Actuaries and other personnel involved in the project will adhere to the highest standards of data security and privacy, ensuring that all proprietary information remains secure. This includes implementing robust measures to prevent unauthorized access, sharing, or leakage of any information.

Our consultancy will only use the provided data for the intended purposes of the project and will not reproduce, distribute, or utilize the information in any manner that is not explicitly authorized by CIC.

This commitment to confidentiality extends indefinitely beyond the conclusion of the project, underscoring our dedication to maintaining the trust and integrity placed in us by CIC.

# Proposed learning outcomes

Upon the successful completion of this project, our team members will gain a comprehensive understanding of the workings of a real actuarial consulting project. Under the guidance of our mentor, Carlos, we will acquire hands-on experience in implementing actuarial consulting methodologies. This includes the application of actuarial science techniques such as clustering, Principal Component Analysis (PCA), and exploratory data analysis to real insurance consulting cases.

We will also deepen our business insights within the insurance industry. By participating in the formulation of diversified product pricing strategies, we will learn about the variations in insurance policies across different products and understand the pivotal role of clustering and machine learning in this domain. The application of exposure rating will equip team members with the skills to evaluate expected losses and make informed decisions, which is a critical aspect of the insurance sector. Additionally, the practice of future monitoring for models will help us anticipate and adapt to potential changes in insurance conditions. These experiences will provide us with a thorough understanding of insurance operations, enabling us to avoid common pitfalls in the future.

Furthermore, this project will enhance our soft skills. We will develop effective communication strategies to interact with clients efficiently, ensuring their confidence in our capabilities. Presenting technical concepts to non-experts and writing clear business reports will also be key areas of improvement. These skills, while not easily quantifiable, are essential for professional success and will significantly contribute to our ability to convey progress and outcomes in an understandable manner.

In summary, we aim to refine our technical expertise, business acumen, and soft skills through this project, preparing us to excel in our future roles.