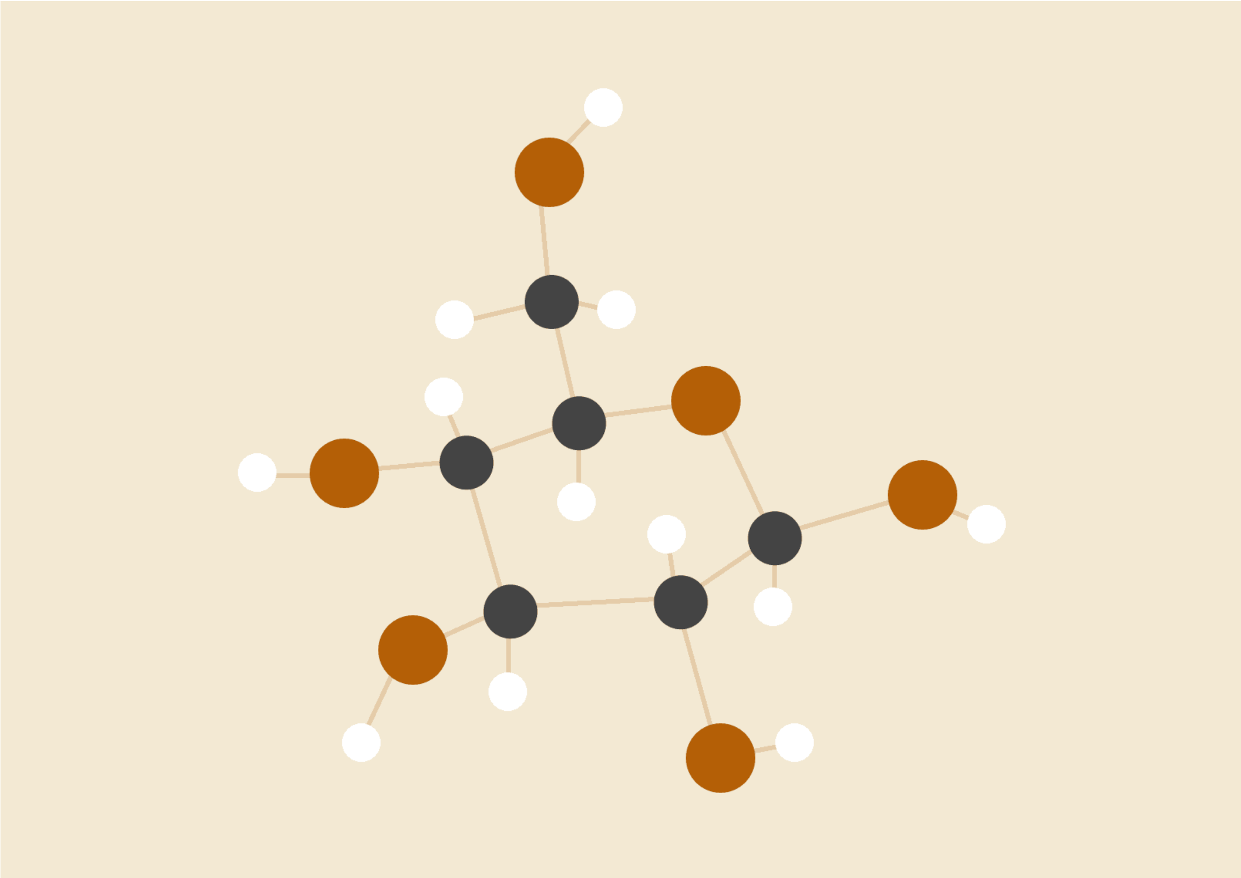
TECHNICAL ASSESSMENT REPORT

Analysis Report: Prioritising Firms based on Risk metrics



Anuj Suhag

04.12.2023

INTRODUCTION

In the dynamic and intricate world of insurance, assessing and managing risk is important for both the stability of the firms and the protection of their customers. This analysis aims to evaluate firms across three fundamental dimensions of risk: liquidity risk, operational risk, and strategic risk. Utilising data spanning the years 2016-2020, the analysis delves into various metrics to isolate firms that exhibit potential vulnerabilities or inefficiencies in these areas.

1. Liquidity Risk is defined by the inability of a firm to meet its financial obligations. This report uses the Solvency Capital Requirement Coverage Ratio (referred to as SCR cr in the analysis) as introduced in the problem sheet.

2. Operational Risk is defined by the presence of operation inefficiencies in a firm or losses noticed in the underwriting process. The key indicator for operation risk used in this analysis is Net combined ratio (referred to as NCR in the analysis). NCR is a function of both operating expenses (OPEX) and insurance claims; it therefore combines both efficiency of underwriting and operation costs aspects.

3. Strategic Risk is defined by the potential risks surrounding strategic decisions taken by the firm. The differentiating factor between strategic risk and operational risk lies in the fact that strategic risks can be affected by factors outside of the firm such as the market conditions. In this analysis, the GWP/NWP ratio is used as a proxy for strategy risk as this ratio indicates a firm's reliance on reinsurance.

METHODOLOGY

Data Preparation:

A lot of careful steps were taken to ensure the features (columns) are consistent across the analysis to make the analysis digestible. The steps included:

1. Renaming the columns to a common schema
2. Merging the two datasets (on left) to get representation from the underwriting dataset.
3. Removing outliers using median absolute deviation. The outliers were imputed with two different methods (mean and median) to have sensible proxies for them. Some interesting outliers were noticed in SCR cr feature and they were

processed using mean imputation. Similarly, some outliers were also seen in Net Combined Ratio (referred as NCR in the analysis) that were processed with median imputation to attempt to keep NCR above 1.0 (an assumption is made here that insurance companies optimise to keep NCR > 1). Similarly, simplifying assumptions have been made about the negative entries - they are not errors but excess liabilities on the part of the firm.

In terms of the structure of the project files, the following steps were carried out for the ease of reading and understanding:

1. The two programming tasks are shown in separate file - task_1.py and task_2.py
2. Functions used in the analysis are in utils.py

RESULTS - Task 1

As discussed before, the analysis was carried across three different dimensions of risk.

Figure 1 shows the top 5 firms that have the lowest SCR coverage ratio as seen in the dataset. The dataset contains some firms that have negative SCR coverage ratio implying that firm's liabilities exceed its assets reflecting adverse financial conditions or some changes in the regulatory framework on how SCR is calculated. In either scenario, this warrants a deeper look into the following five firms - Firm 109, Firm 141, Firm 183, Firm 225 and Firm 251.

Figure 2 shows the top 5 firms that have the highest average NCR. These firms have higher than average NCR seen across all firms over 6 years indicating higher than average operations risk. The five firms that have higher than average operational costs are - Firm 210, Firm 305, Firm 157, Firm 40, Firm 299.

Finally, Figure 3 shows the top 5 firms that have higher than average GWP/NWP ratio or reinsurance dependence. These firms are - Firm 282, Firm 248, Firm 142, Firm 157, Firm 143. Reinsurance is generally a hedging bet but with the rising frequency of events today because of climate change, this can have a trickle down effect.

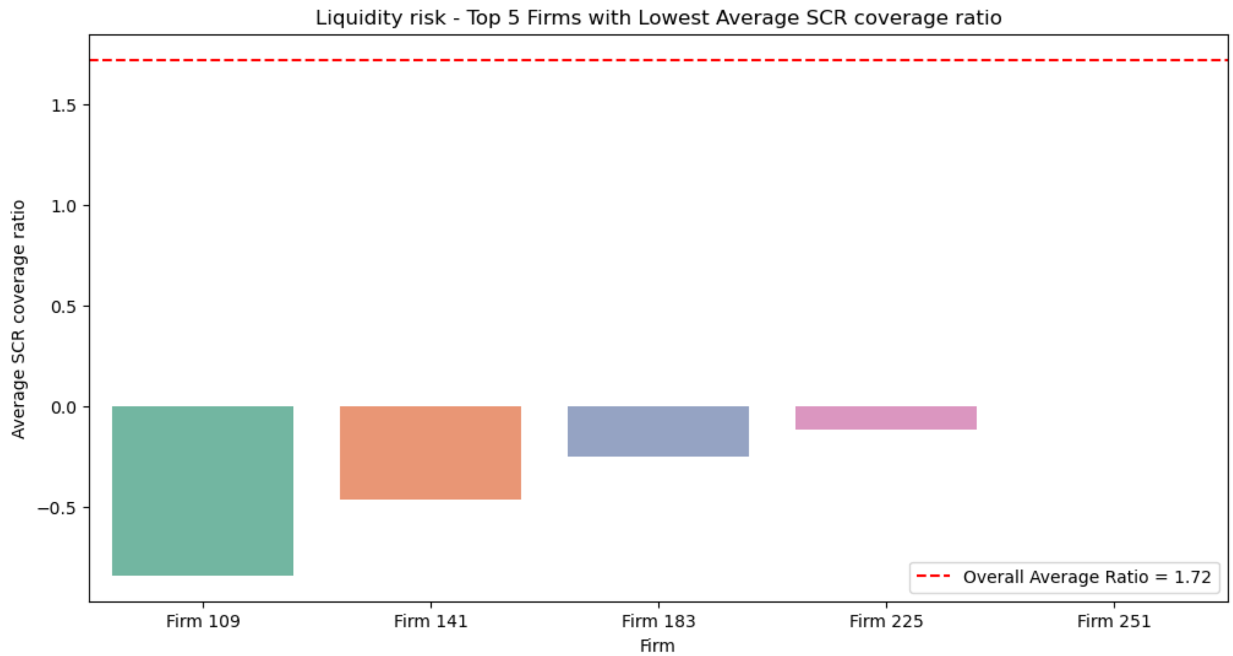


Figure 1. Average SCR cr vs Firms

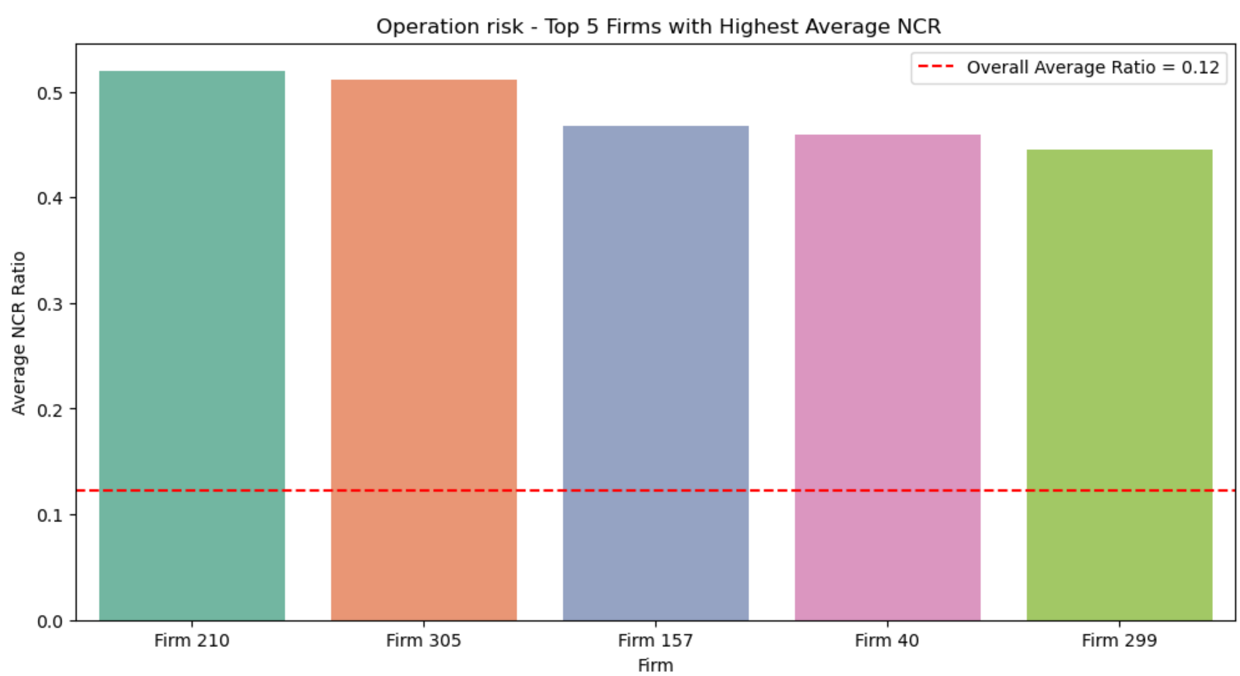


Figure 2. Average NCR vs Firms

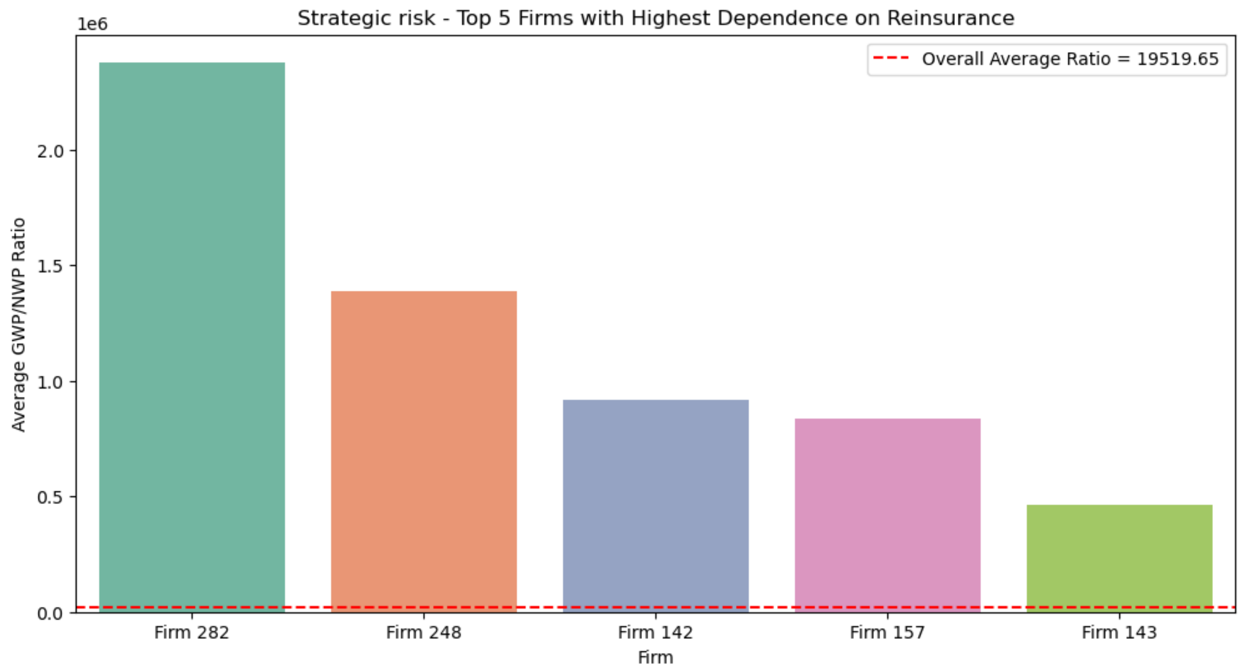


Figure 3. Average reinsurance dependence vs Firm

RESULTS - Task 2

The mission with this task is to cluster the processed data and see how many groups naturally exist. The key idea is to gain insights into other firms that exhibit similar risk compared to the ones we have identified from task 1, in a first pass autonomous manner.

To do this, the Kmeans unsupervised learning model is selected for its simplicity. Figure 4 shows the elbow plot that is used to decipher the number of clusters that exist in the data. As can be seen, a small elbow exists at k = 3

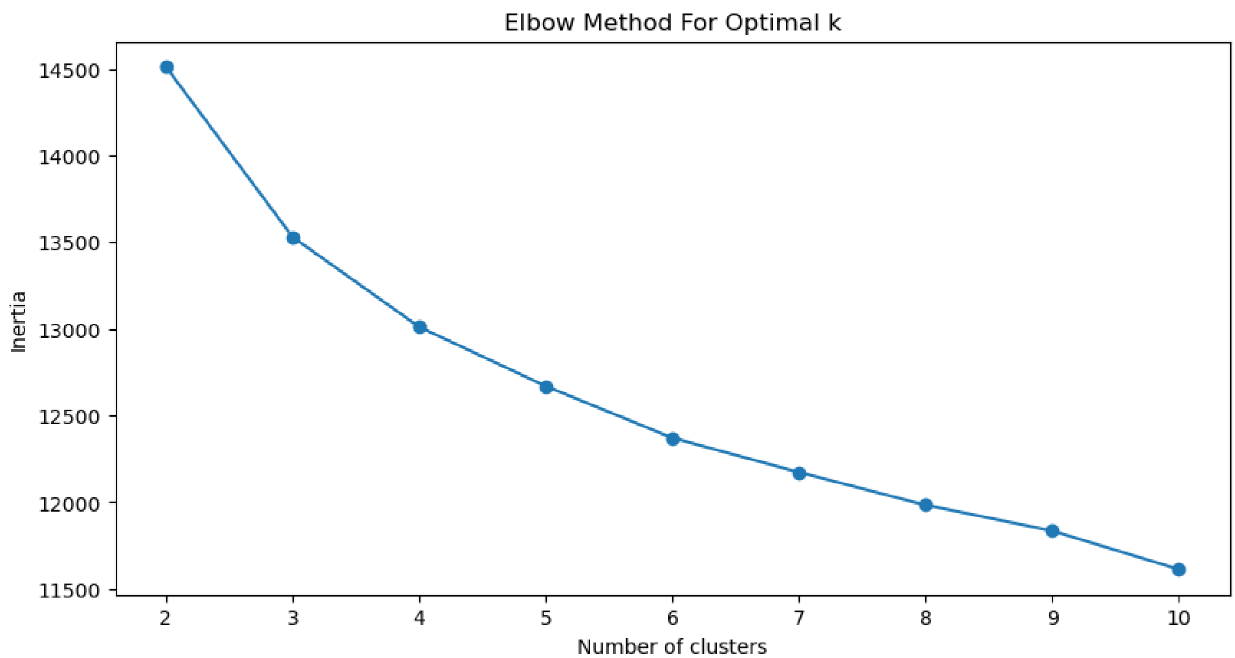


Figure 4. Elbow plot showing a small elbow a #cluster or k = 3

Further analysis shows (as can be verified by running the programs in the analysis) that all firms found to be at operational risk with high NCR fall in cluster 0 and the firms found to have low SCR cr and at liquidity risk fall into cluster 1.

A further deep dive into the latest data for each clusters is presented in the table below :

	NWP 2020	GWP 2020	EOAOL 2020	NCR 2020	SCR coverage ratio 2020
	mean	mean	mean	mean	mean
Cluster					
0	0.694596	4.301626	33.214831	0.026234	1.836768
1	0.106024	0.574373	4.460738	0.016622	0.675842
2	0.754370	4.608124	27.911071	0.002478	1.852212

Although it cannot be seen and concluded from the aggregated table above (the clusters aggregate don’t easily discriminate along these KPIs), a shallow dive does show some interesting trends and encourages a deeper analysis along with a subject matter expert.

RESULTS - Task 3

Creating an end-to-end data processing and analytics pipeline using Microsoft Azure for batch processing and analysis involves several steps and Azure services. Here's an outline of the process along with explanations for the choice of tools and services:

1. Data Storage and Ingestion using Azure Blob Storage: This permits storing of raw data in a cost-effective and scalable cloud storage solution. This is ideal to store CSV files from various firms.

2. Data Processing and Transformation using Azure Databricks: Excellent for data engineering, data exploration, and machine learning tasks. Use Databricks for data preprocessing (like handling missing values, outliers), feature engineering, and scaling the data before applying K-means clustering. Databricks also integrates perfectly with Azure Blob Storage and provides support to several programming languages.

3. Machine Learning and Analytics with Azure Machine Learning Service: Azure Machine Learning Service for building and training K-means clustering model or other ML models. Also has MLOps capabilities to support good model life cycle hygiene

4. Data Visualization and Reporting with Power BI: Power BI can connect to various data sources, including Azure Blob Storage and Azure SQL Database. Great for interactive reports and dashboards.