**Risk Analysis for Got Your Back Insurance**

Completed by Madison Costanza on 11/7/2022

Below were the steps taken to ensure clean data before completing the analysis:

1. Ensured that all data in the provided CSV file is relevant and there were no duplicate rows.
2. Ensured column entries contained expected values only.
   1. Replaced 4121 “Wood Frame” entries with “WD10” in construction code column.
   2. Replaced 4253 “Wood” entries with “WD00” in construction code column.
3. Used graphical analysis to identify any outliers in value variables (none identified).
4. Identified and replaced 3 missing data entries (two in ‘Stories’ column and one in ‘Year Built’ column). The missing data was filled with the most frequent value for each variable.
5. The variable, “Stories” contained entries “5+” and “Over 10” which would make our analysis difficult when we need to look at integer values to categorize the building height bands. To remedy this, we can replace these entries with randomly generated integers between 5 and 20 and between 11 and 20, respectively.

*Note: We are assuming a maximum building height of 20 stories here. This could easily be adjusted if upon speaking with the company, we find that they ensure buildings up to 50 stories, for example.*

Tasks:

1. For each state in the output file, find the total insured value (TIV) and number of risks.

Explanation: To complete this task, I first created new columns for “Risk Count” (populated with value “1”) and “Total Insured Value” (the sum of building value, other value, contents value, and time element value) for every location, or row, in the dataset. I then aggregated the sum of these columns by state.

1. Which 5 counties contain the largest total insured value (TIV) for construction code WD10?

Explanation: Cleanliness of the “Construction Code” variable was crucial for this task, as many Wood Frame entries would not have been counted if I did not replace “Wood Frame” with “WD10.” For this step, I created a filtered table for “WD10” entries, grouped this table by state *and* county (careful to control for the possibility of some states having the same county names), and aggregated the TIV totals for each county.

1. For each portfolio, find the 10 postal codes that are the most susceptible to damage from winter storms.

Explanation: Here, it makes sense to consider locations with the highest average annual losses as “most susceptible to damage from winter storms.” To complete this step, I first created filters for each portfolio, grouped each filtered table by postal code, then found the aggregate sum of AAL for each group.

1. For each Pennsylvania postal code, find the breakdown of total insured value (TIV), risk count, and average annual loss (AAL) by building height band.

Explanation: The first step of completing this task involved generating random integers to replace string variables in the Stories column, then categorizing each location entry to a building height band of Small, Medium, or Large. Next, I created a filtered table for PA locations, then aggregated TIV, risk counts, and AAL totals by building height bands and postal codes.

Questions:

1. Does one of the portfolios appear to be more vulnerable to damage from winter storm? If so, which one? Explain your reasoning and any analysis you conducted to support your conclusion.

To answer this question, we must first answer what it means for each portfolio to be “vulnerable” to damage from a winter storm. We cannot consider only total TIV of each portfolio because this does not tell us anything about historical losses at each location. On the other hand, we cannot consider only AAL because this does not tell us anything about the amount of total exposure we have at each location. Instead, it would make sense to consider the proportion of AAL to TIV at each location as this takes both factors into account.

Using this reasoning, we can calculate a “vulnerability score” at each location and aggregate the sum of these scores for each portfolio. Since the vulnerability score is calculated as AAL divided by TIV, the lower the vulnerability score, the less vulnerable the portfolio is to losses from the storm. We find Portfolio 1 to have a cumulative score of 190.48 and Portfolio 2 to have a cumulative score of 74.21, so Portfolio 1 is more vulnerable to damage from winter storms.

1. Which factors appear to have the greatest influence on AAL? Explain your reasoning and any analysis you conducted to support your conclusion.

It appears that the mobile-home and masonry policies are having a large influence on AAL in Portfolio 1. To reach this conclusion, we can look at the aggregated sum of AAL, grouped by each portfolio and construction code. We observe that only Portfolio 1 contains polices for mobile homes and that the AAL aggregations for these locations are quite high when compared to other construction types. In fact, these construction types are yielding significantly higher AALs as a proportion of risk counts than wood and concrete based structures.

It is also evident that new and average-aged buildings in Portfolio 1 are influencing a high AAL for the portfolio. We can see this by aggregating sums of AAL by portfolio and our building age band. By a similar method, we can also observe that small and medium sized buildings are contributing a higher AAL in Portfolio 1.

To conclude this analysis, to manage the risks in Portfolio 1, I would advise to reduce policies for mobile homes and masonry buildings and consider that relatively newer and smaller buildings in this Portfolio are experiencing higher AAL trends.