Agent Selling Behaviour Challenge

1. Introduction

Agents are a vital part of AIA. As a data scientist in AIA, we constantly work with the agency team to understand more about our agents. Being successful in this objective enables us to make more informed business decisions and craft our strategy in various areas, such as improving agents' productivity, which leads to attracting more customers as well as improve customer experience. One key area that we always keep an eye on is our agents' selling behaviour, which can inform us several things:

i. Do some agents lack the knowledge to sell certain products?
ii. Are some of our products priced sub-optimally, which lead to agents not selling them?
iii. Etc.

Note: In this exercise, we are only interested in the *Inforce* agents under the *agency distribution channel*.

2. Data

You are provided with 6 csv files, which consists of 2 main files and 4 supporting files.

i. AGENT.csv (**main**)
ii. MAIN.csv (**main)**
iii. B.csv, C.csv, D.csv, E.csv (**support**)

To understand what the field means, a data dictionary is provided. Also, to finish the exercise, you will need to utilize the supporting files to get the information. To do this, there are 2 SQL files, which provide the logic (in SQL) to derive product categories, which we are ultimately interested in.

3. Rules

You may use Python or R for the challenge. A notebook might be the best way to show your code and write your comments/thoughts for the reviewers to follow along. **DO NOT** use any help from other people or online forum, etc. Your submission should be solely your ideas and work. There is no hard limit on the amount of time you can spend on this challenge. Please also provide the amount of time you spend on this challenge once you've completed it.

4. Guidelines

By the end of the challenge, we hope to see:

i. Agent segments based on their selling behaviour and interpretation of these segments.
ii. A product category score assigned to agents. (Optional)

5. Questions

To answer the following questions, you will need to understand the concept of basic policy and rider and derive the product categories based on the provided logic. Answering these questions will help you understand more about the data.

When one purchases an insurance product, they will have a basic policy. This basic policy usually only covers certain aspects. However, different customers have different needs based on various factors, in order to accommodate most of the customers' needs, an insurance company designs various Add-Ons, aka Rider for customers to attach to their basic policy.

To derive the product categories we are interested in, the files in **sql_logic** will help you, ultimately, we are interested in 3 categories, known as **'Cat 1', 'Cat 2', 'Cat 3'**. To derive this at the policy level, a policy is considered **'Cat 1'** or '**Cat 2**' if the basic policy or any of the riders attached is **'Cat 1'** or **'Cat 2'**. The only time when a policy can be in 2 categories is when they are both **'Cat 1'** and **'Cat 2'**.

i. By product categories?
    a. What is the total number of policies sold?
    b. What is the total ANP?
ii. Number of agents who sold at least **5** Vitality attached policy.
iii. Number of policies with at least **3** riders attached.

6. Deliverables

A zip file containing the following folders:

1. notebooks:
    a. HTML file that contains the code and output from your notebook.
    b. Notebook, ideally we are able to execute from top to bottom.
2. artifacts: Any models and external data source you have created or used.
3. README.md file.
4. helper: Any utility or helper functions you use in this challenge.

7. Hints

i. We are interested in understanding **AGENT's selling behaviour**.
ii. It will be much easier to complete the challenge by looking at **Policy Level** instead of **Rider Level**.
iii. We are not only interested in the final deliverables mentioned in (4), but also your thought process leading up to the deliverables and any interesting insights you found in the dataset.
iv. Don't over-focus on wanting to show case the most complex and technical algorithms and skills you have. In most cases, simple is best. In AIA, data scientists deal with many business stakeholders who do not have the same level of technical knowledge as you.