Assessment Task

Please use R or Python to answer the below questions concisely using statements, graphs and tables where applicable, making use of any packages of your choice. Jupyter Notebook or R script(s), R Markdown are all acceptable.

You have 1 week to complete this test so please do not feel rushed and it is “open book”, so feel free to use any online or offline resources during this time.

After completion, please email back an archived folder with all your answers (data, scripts, graphs etc.) and your initials in the file name.

If anything is unclear or you are unsure how to answer a question, please note this down in the email and do not worry.

This is an exchange of ideas exercise; we will be paying more attention to how you put your thoughts forward and how clear they are other than the results.

**Objective:**

Domestic and General is interested in using Data Science to lead their pricing strategy on their **new business book**. In this exercise, you will be tasked with producing a set of pricing analysis to support business decision in maintaining the current strategy or suggesting a new arrangement.

**Data:**

The dataset provided, **pricing\_analyst\_dataset.csv**, contains a sample of offers with information being collected at point of quote.

There are a few pricing strategies being used (column **pricing point**):

* ASIS: standard base rate
* @22%: demand constrained 1
* @23%: demand constrained 1

A description of each variable is provided on the second page. Data at Domestic and General is generally collected at the contact centre and as such the data may have missing values. You may choose to impute or omit these errors in your analysis.

**Important note:**

The premium offered to the customer (offered fee) is the result of a mathematical optimisation done whilst the customer is on the line by taking the appliance level price (base rate) and adding an increase or decrease price variation (price diff). This price variation is one the main parameters we use in the optimisation.

**Tasks:**

1. Explore the dataset and conduct any EDA that you feel will help you understand the data.
2. Compare the **pricing** performance of all strategies in a summary table with:
   1. No of offers;
   2. Average base, offered and sold premium
   3. Average price increase
   4. Conversion
   5. How would you test (statistically) if they are different?
3. Compare the price elasticity of the strategies in the dataset:
   1. Price elasticity

Note: this will give you the elasticity of one group versus other, would you be able to calculate the elasticity within a group?

1. Assuming each pricing point is generated by a model at a different global constraint, would you be able to indicate:
   1. which one is the best model from the statistical point of view?
   2. and from pricing point of view (revenue)?
2. Claims frequency and value are provided, is there any direct relationship between conversion and claims?
   1. Would you say these models are biased?
3. What would be your final recommendation based on your findings?

**DS\_Train.csv variables are as follows:**

1. **Claim\_amount (float):** sum of total value of claims in the last period
2. **Claim\_count (int):** number of claims in the last period
3. **Purchase\_Date (date):** date when the appliance was purchased.
4. **Purchase\_Price (float):** value paid for the appliance.
5. **Sale\_Flag (binary):** thisindicates that the offer was accepted (this is the target)
6. **Base rate (float):** appliance level premium, used as starting point for the optimisation.
7. **Sold premium (float):** fee when the offer was accepted.
8. **Offered premium (float):** Premium presented to the customer.
9. **Plan\_Flag (binary):** flag whether the customer has a live plan with D+G.
10. **Age (int):** Age of appliance in days.
11. **Price\_Diff (float):** price difference between base rate and offered premium.
12. **Plan\_Count (int):** number of plans the customer has.
13. **plansactive\_lastyear\_count:** totalnumber of plans that the customer had in the last year.
14. **planscancelled\_lastyear\_count:** total number of plans the customer cancelled in the last year.
15. **Pricing point(cat):** indicates the pricing strategy (not a modelling feature)
16. **Manufacturer brand (cat):** coded indicator of the manufacturer
17. **Item category code (cat):** coded indicator of the type of the appliance, such as, fridge, cooker, etc.
18. **Item super category code (cat):** coded 2nd level indicator of the type of the appliance.