IMIx Data Analytics Certificate Fall 2020

Individual Final Capstone Project Due date: 2020-10-26 23:59 EDT

Deliverables (to be submitted by email to Gerhard.Trippen@utoronto.ca):

1. Case Report with your findings and all the graphs (and other appendices). *Maximum* 10 pages of written text (not including graphs and appendices).

2. All Python source code either in a Jupyter Notebook (*.ipynb) or a Python file (*.py). One file!

Include an Executive Summary (at the beginning) describing your most salient findings. Explain all steps and results clearly and cogently, so that a reasonably intelligent though statistically naïve manager could understand it. Your narrative should be clear and concise, accompanied by supporting evidence in the form of graphics and tables. All tables and graphics should be well formatted (e.g., tables should not run over from one page to another).

This project is open book, however, all parts (code and report) must be authored by yourself. You can use any reference you want (citation/link **required**), except for presenting somebody else's solution, of course.

You must use *Python* to perform the analysis and generate the graphs.

The Case:

We are presented with a sample of 4499 customers of a large Canadian bank that currently has a surplus of liability customers, i.e., depositors. The bank would like to grow their asset customer base, i.e., borrowers, through a targeted marketing campaign. Previous marketing efforts were able to garner a roughly 10% conversion rate, where a conversion was considered to be a sale of a personal loan product to an existing liability customer. From this campaign, which was held in Toronto, a data set was established and to be used to improve the efficiency of this year's marketing campaign. This efficiency boost will stem from designing an algorithm to predict which clients are most likely to convert on the personal loan, and then focus marketing on this group.
All sampled customers have a financial advisor/planner with this bank. The financial planners are mobile and can come to the customers if they wish so, however, listed in the data set are the branches that they most often serve. Financial advisors earn commission if they can advise customers to buy a personal loan product.

Any references to real people or real locales are used fictitiously. Other names, characters, places, and incidents are the product of the author's imagination, and any resemblance to actual events or locales or persons is entirely coincidental.

The data set includes the following variables:

Age	Customer's age in completed years
Experience	Number of years of professional experience
Income	Annual income of the customer ($000)
Branch Address	Address of customer's home branch
Family	Family size of the customer
CCAvg	Avg. spending on credit cards per month ($000)
Mortgage	Value of house mortgage if any ($000)
Personal Loan	Did this customer accept the personal loan offered in the last campaign?
Brokerage Account	Does the customer have a brokerage account with this bank?
GIC	Does the customer have a Guaranteed Investment Certificate (GIC) account with this bank?
Online	Does the customer use internet banking facilities?
CreditCard	Does the customer use a credit card issued by this bank?
Advisor Name	Customer's financial advisor/planner with this bank
Advisor Designation	Financial advisor's designation

The following columns (Branch Address *i*) list the branch addresses that a particular advisor most often serves.

Brief "Financial Glossary":

- Designations:
  PFP: Personal Financial Planner
  CFP: Certified Financial Planner
  For a discussion on the designations see, for example,
  https://www.theglobeandmail.com/globe-investor/personal-finance/financial-advisers-a-tale-of-two-certificates/article4083578/
  "The problem with the PFP is that it's redundant. The Certified Financial Planner, or CFP, is bigger, better and an ideal designation on which to build a single accreditation for the professional adviser."
  "To keep things super simple, look for the CFP if you're seeking true financial planning and not someone to sell you mutual funds. The PFP is okay, but it's the No. 2 brand."
- A brokerage account allows customers to access more complex investments, such as stocks and bonds
- A guaranteed investment certificate (GIC, French: certificat de placement garanti, CPG) is a Canadian investment that offers a guaranteed rate of return over a fixed period of time, most commonly issued by trust companies or banks.[1] Due to its low risk profile, the return is generally less than other investments such as stocks, bonds, or mutual funds. It is similar to a time or term deposit as known in other countries.[2]" (https://en.wikipedia.org/wiki/Guaranteed_investment_certificate)

A) **Data Cleaning and Basic Data Exploration**

Carefully inspect and clean the data. Use simple descriptive statistics and basic univariate plots. Report your findings and how you resolved any potential issues.

Use the cleaned data set for next questions.

B) **Exploratory Data Analysis**
Ignoring the variables containing text, i.e., names, addresses and designations, explore the data with respect to the goal of the bank. Present your four most important results also using visualizations. What are your insights?

C) **Business Analytics**
Who is or are the most successful financial advisor(s)? Discuss your results.
Hint: How could "successful" be defined? There might be multiple possible definitions.

D) **Machine Learning**
Partition the data into training (75%) and test (25%) sets. Use a number of different machine learning techniques and evaluate the algorithms. Show the classification matrix for the test data that results from using the best model. Clearly explain your results. How well did your model help you achieve the objective?

---------- EOF ----------