Data and Information Management for Business Analytics

Individual Final Capstone Project – Sample

Deliverables (to be submitted on *Quercus*):

1. Case Report with your findings and all the graphs (and other appendices).
***Maximum* 4 pages of written text (not including graphs and appendices).**

2. All Python source code in a Jupyter Notebook (*.ipynb).
Add your code to the provided starter file.

This final project has 4 question parts, for a total of 40 points, and includes a bonus question for one additional bonus points, and 3 pages, please check it. You are not required to solve the bonus question! The points per question part, shown below, do not necessarily reflect the relative amounts of time that it will take to answer each question part.
If you need to make any assumptions, please state them clearly!

HONESTY PLEDGE:

In submitting this final deliverable, I confirm that my conduct during this final deliverable adheres to the Code of Behaviour on Academic Matters. I confirm that I did NOT act in such a way that would constitute academic dishonesty, misrepresentation, or unfairness, including but not limited to, using unauthorized aids and assistance, personating another person, and committing plagiarism.

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. Please see the paragraph above for more details.

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.

Data and Information Management for Business Analytics

Individual Final Capstone Project – Sample

A preliminary analysis of the data set will have to be performed now in preparation of the application of machine learning algorithms, which will be done by other members of your team.

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
Branch Manager	Name of the manager of the above 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 and Branch Manager i) list the branch addresses and branch managers 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)

Data and Information Management for Business Analytics

Individual Final Capstone Project – Sample

A) (14 points) Database Design and SQL
The answers to coding questions must be given in *SQLite*.

i)   What are the entities of the given scenario?

ii)   This data set is not in Boyce-Codd Normal Form (BCNF). Explain why not.

iii)   Transform the given data set into a database in BCNF.
        Draw an Entity-Relationship diagram of your resulting database.

iv)   Write SQL code that creates the necessary table(s) within the database.
        You do **NOT** need to populate the database with the given data!
        Bonus (**not required, *1 bonus point***): *Do* populate the new database.

Using the original data set, without pre-processing and not the new database in BCNF, write SQL code that lists

v)   … all financial advisors.

vi)   … Age, Experience, Income and Mortgage of customers who have a mortgage.

vii)   … the number of all customers for each branch.

viii)   … the maximum of the average credit card spending over all customers that took a personal loan.

ix)   Only consider advisors whose number of customers is more than half of the overall average number of customers per advisor. List these advisors and the mean income of their customers combined in descending order (by mean income).

For the following questions, do *not* use the database in BCNF. Your starting point should be the original data set. Modify the data set in any way needed from there on.

B) (8 points) Data Cleaning and Basic Data Exploration
You must use *Python* to perform the analysis and generate the graphs.
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.

C) (8 points) Exploratory Data Analysis
You must use *Python* to perform the analysis and generate the graphs.
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?

D) (10 points) Business Analytics
You must use *Python* to perform the analysis and generate the graphs.
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.

-- EOF --