COMMERCIAL BANKING, CORP

REQUEST FOR PROPOSAL
RFP #: IP – F1.H1

TITLE: BANKING INSURANCE PRODUCT – PHASE 1
CLOSING DATE AND TIME: SEPTEMBER 1. 2023 @ 5:00 PM

Banking Insurance Product – Phase 1: IP – F1.H1

Purpose

By responding to this Request for Proposal (RFP), the Proposer agrees that s/he has read and understood all documents within this RFP package.

Submission Details

Responders to this RFP should supply:

- A business report **up to 4 pages** (not including cover page, table of contents, or any needed appendix), including any supporting plots and tables.
- The commented code used to produce the results.

The report should address **all points described in the "Objective" section** below.

The report should be returned in the following way:

- Electronic (submit via Moodle)

Background

The Commercial Banking Corporation (hereafter the "Bank"), acting by and through its department of *Customer Services and New Products* is seeking proposals for banking services. The Bank ultimately wants to predict which customers will buy a variable rate annuity product.

A variable annuity is a contract between you and an insurance company / bank, under which the insurer agrees to make periodic payments to you, beginning either immediately or at some future date. You purchase a variable annuity contract by making either a single purchase payment or a series of purchase payments.

A variable annuity offers a range of investment options. The value of your investment as a variable annuity owner will vary depending on the performance of the investment options you choose. The investment options for a variable annuity are typically mutual funds that invest in stocks, bonds, money market instruments, or some combination of the three. If you are interested in more information, see: http://www.sec.gov/investor/pubs/varannty.htm

The project will be broken down into 3 phases:

- Phase 1 – Variable Understanding and Assumptions
- Phase 2 – Variable Selection and Modeling Building
- Phase 3 – Model Assessment and Prediction

Objective – Phase 1

The scope of services in this phase includes the following:

- For this phase use **only** the training data set.

- Explore the predictor variables **individually** with the target variable of whether the customer bought the insurance product.
    - Summarize **only the significant variables** in a table ranking from most significant to least significant – the Bank currently uses $\alpha = 0.002$, but is open to another if you defend your reason.
        - This table should separate out the four possible classes of variables – binary, ordinal, nominal, continuous.
        - (HINT: Explore the predictor variables individually for now since you have not yet accounted for missing values.)
        - (HINT: The downside to software sometimes is displaying a full p-value for ranking. That doesn't mean you cannot get them through the right commands. As long as you have the same degrees of freedom you can rank on test statistic as well.)
    - In an appendix, include a table with **all** of the variables ranked by significance.
- Provide a table of odds ratios for **only binary predictor variables** in relation to the target variable.
    - Rank these odds ratios by magnitude.
    - Interpret **only** the highest magnitude odds ratio.
    - Report on any interesting findings.
        - (HINT: This is open-ended and has no correct answer. However, you should get use to keeping an eye out for what you might deem important or interesting when exploring data to report in an executive summary.)
- Provide a summary of results around the linearity assumption of continuous variables.
    - List both which variables meet and do not meet the needed assumption for continuous variables.
    - (HINT: Do not get overly mathematical here. Just report what you find; do not teach.)
- Provide a summary of important data considerations as follows:
    - Visual representation of which variables have the highest (defined by you for now) amount of missing values.
    - List any combinations of variables that you feel have redundant information so the Bank might consider removing them in the future.
        - (HINT: This is open-ended and has no correct answer. For example, presence of a money market account and money market balance.)
    - Report on any interesting findings.
        - (HINT: This is open-ended and has no correct answer. However, you should get use to keeping an eye out for what you might deem important or interesting when exploring data to report in an executive summary. For example, teller visits as well as other variables might represent human contact with the bank as compared to only online contact.)

Data Provided

The following two sets of data are provided for the proposal:

- The training data set **insurance_t** contains 8,495 observations and 48 variables.
    - All of these customers have been offered the product in the data set under the variable **INS**, which takes a value of 1 if they bought and 0 if they did not buy.
    - There are 47 variables describing the customer's attributes **before** they were offered the new insurance product.
- The validation data set **insurance_v** contains 2,124 observations and 48 variables.
- The table below describes the Roles and Description of the variables found in both data sets.
    - **Except for Branch of Bank**, consider anything with more than 10 distinct values as continuous.

Name	*Model Role*	*Description*
ACCTAGE	Input	Age of oldest account
DDA	Input	Indicator for checking account
DDABAL	Input	Checking account balance
DEP	Input	Checking deposits
DEPAMT	Input	Total amount deposited
CASHBK	Input	Number of cash back requests
CHECKS	Input	Number of checks written
DIRDEP	Input	Indicator for direct deposit
NSF	Input	Number of insufficient fund issues
NSFAMT	Input	Amount of NSF
PHONE	Input	Number of telephone banking interactions
TELLER	Input	Number of teller visit interactions
SAV	Input	Indicator for savings account
SAVBAL	Input	Savings account balance
ATM	Input	Indicator for ATM interaction
ATMAMT	Input	Total ATM withdrawal amount
POS	Input	Number of point of sale interactions
POSAMT	Input	Total amount for point of sale interactions
CD	Input	Indicator for certificate of deposit account
CDBAL	Input	CD balance
IRA	Input	Indicator for retirement account
IRABAL	Input	IRA balance
LOC	Input	Indicator for line of credit
LOCBAL	Input	LOC balance
INV	Input	Indicator for investment account
INVBAL	Input	INV balance
ILS	Input	Indicator for installment loan
ILSBAL	Input	ILS balance
MM	Input	Indicator for money market account
MMBAL	Input	MM balance
MMCRED	Input	Number of money market credits
MTG	Input	Indicator for mortgage
MTGBAL	Input	MTG balance
CC	Input	Indicator for credit card
CCBAL	Input	CC balance
CCPURC	Input	Number of credit card purchases
SDB	Input	Indicator for safety deposit box
INCOME	Input	Income
HMOWN	Input	Indicator for home ownership
LORES	Input	Length of residence in years
HMVAL	Input	Value of home

AGE	Input	Age
CRSCORE	Input	Credit score
MOVED	Input	Recent address change
INAREA	Input	Indicator for local address
INS	Target	Indicator for purchase of insurance product
BRANCH	Input	Branch of bank
RES	Input	Area classification