Week 9

Team member's details

Group Name: Ctrl_C&Ctrl_V

Name:

1. Yihsuan Sun
2. Tiantian Xie

Email:

- Yihsuan S.: yihsuan.sun88@gmail.com
- Tiantian X.: tenxie0411@gmail.com

Country: United States

College/Company:

- Yihsuan S.: Washington State University
- Tiantian X.: Case Western Reserve University

Specialization: Data Analyst

Problem description & Business understanding

XYZ Credit Union, located in Latin America, does well in selling banking products such as: credit cards, deposit accounts, retirement accounts, safe deposit boxes, etc. However, after statistics, they found that their existing customers basically only buy one product, which means that the bank does not perform well in cross-selling. So XYZ Credit Union wants analysts to build models such as marketing models through machine learning to solve their problems**.**

Data cleansing and transformation done on the data

For most of the missing data in this dataframe, we noticed that it might not be a good choice to replace them with mean or any other way, the reason is because most of the columns containing the missing data don't do the calculation. In other words, those columns are used to do the classification, if we chose to replace all the missing values with mean value in that column, this doesn't help anything in the future. The second reason is that some missing values don't really mean 'missing'. The reason we say that is because 'missing' may be a choice and represent some information in that column. For instance, in column ind_nuevo, the missing value only means that the customer didn't register in the last 6 months. This is information we want to know, not the missing value we want to remove.

So the final decision we make is we choose not to remove NA for those columns in the above two situations.

EDA will be present next week.

https://github.com/Yihsuansun/Cross_selling_recommendation/tree/main/Week%209