

Bank Marketing Analytics Dataset

Name: Nishika Abeytunge, Cecilia Li, Venudhar Ravishankar

Introduction

A time deposit or term deposit is a deposit in a financial institution with a specific maturity date. In other words, subscribers lock away an amount of money for an agreed length of time during which they may not access the money until the term is up. The infamous Wells Fargo cross-selling scandal in 2019 involved opening millions of savings and checking accounts on behalf of their clients without consent. While its course of actions is abysmal, many banks engage in marketing their financial products, especially encouraging clients to deposit large sums of money into term deposits. With data science capabilities growing stronger, marketing analytics teams at financial institutions are now able to better predict the likelihood of a client subscribing to a term deposit. Various attributes ranging from age, demographics, income, etc. of different borrowers may be indicators of whether they subscribe.

The purpose of this project is to analyze a banking marketing campaign dataset in order to understand more about the term deposit subscription trends amongst the customers. It will focus on exploratory data analysis; statistical analysis and machine learning predictive analytics are out of scope.

Sample questions

- What's the relationship between income and whether the users had subscribed or not
- Are there any patterns with demographics to whether the users had subscribed or not
- What is the distribution breakdown of gender in this dataset
- What's the breakdown of age in this dataset

Dataset

Source: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>

The data is related to direct marketing campaigns of a Portuguese banking institution from 2008 to 2010. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to assess if the product (bank term deposit) would be (or

not) subscribed. There are two datasets: 1) bank-additional-full.csv with over 40K examples 2) bank-additional.csv with 10% of the examples (around 4K), randomly selected from 1).. For the purposes of this project, we will be using the file 'bank-additional'.

Process

1. Import data: Import 'bank additional' dataset from data source and perform initial high-level analysis
2. Gauge the data: Look at the shape of the file, attributes, missing value, columns and their values respective to the outcome.
3. Clean the data: Remove irrelevant columns, deal with missing and incorrect values, turn categorical columns into dummy variables.
4. Analyze the data: Using Numpy packages to analyze relationships between variables and plot figures of trends and garner insights and findings.

Illustrative Example (Source: jianwenwu):

