**Capstone Project Proposal**

**Susan Joseph  
 August 13, 2017**

The goal of this project is to predict the Banking products a suite of new customers are most likely to be interested in, based on their personal details, and the product favourability patterns of existing customers.

**The Problem**

An anonymous bank based in Spain has published 1.5 years of customer behavior data that consists of customer personal details (Customer Personal Details.csv) and the products of the bank they have subscribed to (Customer Products.csv). The different products that the customers subscribed to are Current Account, Particular Account, Direct Debit, E-account, Payroll Account, Taxes , Credit Cards, Pensions, Securities, Funds and Long Term Deposits. The major Customer attributes that are of interest are the Age, household income, the channel used by the customer to subscribe to a product and Segmentation 01 - VIP, 02 - Individuals 03 - college graduated. The goal of the project is to determine if the Customer attributes mentioned have any correlation to the different bank products so that the marketing department of the Bank could target new customers based on the analysis.

### **Client**

The Client is the anonymous bank in Spain. This project would help the bank’s marketing department to target potential new customers to subscribe to appropriate products of the bank based on Age, Channel and Customer Segmentation.

**Data**

The Bank has provided two .CSV files Customer Personal Details.csv and Customer Products.csv. The data starts at 2015-01-28 and has monthly records of products a customer has, such as "credit card", "savings account", etc. These products are the columns named: ind\_(xyz)\_ult1 etc. There are 1048576 rows and 24 observations in the Customer Peronal Details.csv file and 1048576 rows and 24 observations in the Customer Products.csv file. A sample of both the files will be selected for the project.

# Data fields

Customer Personal Details

| **Column Name** | **Description** |
| --- | --- |
| fecha\_dato | The table is partitioned for this column |
| ncodpers | Customer code |
| ind\_empleado | Employee index: A active, B ex employed, F filial, N not employee, P pasive |
| pais\_residencia | Customer's Country residence |
| sexo | Customer's sex |
| age | Age |
| fecha\_alta | The date in which the customer became as the first holder of a contract in the bank |
| ind\_nuevo | New customer Index. 1 if the customer registered in the last 6 months. |
| antiguedad | Customer seniority (in months) |
| indrel | 1 (First/Primary), 99 (Primary customer during the month but not at the end of the month) |
| ult\_fec\_cli\_1t | Last date as primary customer (if he isn't at the end of the month) |
| indrel\_1mes | Customer type at the beginning of the month ,1 (First/Primary customer), 2 (co-owner ),P (Potential),3 (former primary), 4(former co-owner) |
| tiprel\_1mes | Customer relation type at the beginning of the month, A (active), I (inactive), P (former customer),R (Potential) |
| indresi | Residence index (S (Yes) or N (No) if the residence country is the same than the bank country) |
| indext | Foreigner index (S (Yes) or N (No) if the customer's birth country is different than the bank country) |
| conyuemp | Spouse index. 1 if the customer is spouse of an employee |
| canal\_entrada | channel used by the customer to join |
| indfall | Deceased index. N/S |
| tipodom | Addres type. 1, primary address |
| cod\_prov | Province code (customer's address) |
| nomprov | Province name |
| ind\_actividad\_cliente | Activity index (1, active customer; 0, inactive customer) |
| renta | Gross income of the household |
| segmento | segmentation: 01 - VIP, 02 - Individuals 03 - college graduated |
| Customer Products  ind\_ahor\_fin\_ult1 | Saving Account |
| ind\_aval\_fin\_ult1 | Guarantees |
| ind\_cco\_fin\_ult1 | Current Accounts |
| ind\_cder\_fin\_ult1 | Derivada Account |
| ind\_cno\_fin\_ult1 | Payroll Account |
| ind\_ctju\_fin\_ult1 | Junior Account |
| ind\_ctma\_fin\_ult1 | Más particular Account |
| ind\_ctop\_fin\_ult1 | particular Account |
| ind\_ctpp\_fin\_ult1 | particular Plus Account |
| ind\_deco\_fin\_ult1 | Short-term deposits |
| ind\_deme\_fin\_ult1 | Medium-term deposits |
| ind\_dela\_fin\_ult1 | Long-term deposits |
| ind\_ecue\_fin\_ult1 | e-account |
| ind\_fond\_fin\_ult1 | Funds |
| ind\_hip\_fin\_ult1 | Mortgage |
| ind\_plan\_fin\_ult1 | Pensions |
| ind\_pres\_fin\_ult1 | Loans |
| ind\_reca\_fin\_ult1 | Taxes |
| ind\_tjcr\_fin\_ult1 | Credit Card |
| ind\_valo\_fin\_ult1 | Securities |
| ind\_viv\_fin\_ult1 | Home Account |
| ind\_nomina\_ult1 | Payroll |
| ind\_nom\_pens\_ult1 | Pensions |
| ind\_recibo\_ult1 | Direct Debit |

### **Approach**

Correlation and Regression Statistical techniques will be employed to determine if any of the three variables selected i.e Age, Channel and Segmentations had any correlation with subscribing to the different products.

### **Deliverables**

A full report detailing the process undertaken to complete this analysis will be uploaded to GitHub with R code included.