## Information on the Dataset

Name of Dataset: Churn\_Modelling.csv

Dataset URL: <https://www.kaggle.com/aakash50897/churn-modellingcsv>

Number of Rows: 10,000

Legend: Original Variables, Derived Variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable Name** | **(Assumed) Description** | **Default Data Type** | **Final Data Type** | **Example / Levels** |
| RowNumber | Row number in dataset | integer | integer | 0 |
| CustomerID | ID to uniquely identify customers | integer | integer | 15634602 |
| Surname | Last name of customers | character | character | “Hargrave” |
| CreditScore | FICO credit system score | integer | integer | 619 |
| Geography | Possibly where the customer opened the account | character | Factor with 3 levels | “France”, “Germany”, “Spain” |
| Gender | Gender of customer | character | Factor with 2 levels | “Female”, “Male” |
| Age | Age of customer | integer | integer | 42 |
| Tenure | How many years ago the account was opened | integer | integer | 2 |
| Balance | Account balance | numeric | numeric | 83808 |
| NumOfProducts | Number of the bank’s products the customer uses | integer | Factor with 4 levels | “1”, “2”, “3”, “4” |
| HasCrCard | Whether they own at least 1 of the bank’s credit card | integer | Factor with 2 levels | “0”, “1” |
| IsActiveMember | Whether the bank considers the customer to be actively using their services | integer | Factor with 2 levels | “0”, “1” |
| EstimatedSalary | The customer’s self declared annual salary in USD | numeric | numeric | 101349 |
| Exited | Whether the customer has churned | integer | Factor with 2 levels | “0”, “1” |
| AgeAtStartofTenure | The age of customer when they first opened the account | N.A | integer | 40 |
| hasManyProd | Whether the customer has more than 2 products | N.A | logical | TRUE |
| has4Prod | Whether the customer has 4 products | N.A | logical | TRUE |
| has3Prod | Whether the customer has 3 products | N.A | logical | TRUE |
| depositRate | The mean balance over the length of tenure. Obtained by dividing the customer’s balance by their tenure | N.A | numeric | 41904 |
| savingsToSalaryRatio | The ratio between a customer’s mean balance per tenure years and their estimated salary | N.A. | numeric | 0.372 |
| salaryToAgeSq | The ratio between a customer’s estimated salary and the squared of their age | N.A | numeric | 57.2 |
| salaryQuartile | The quartile that a customer’s salary falls in.  Q1: EstimatedSalary < 51002  Q2: 51002 < EstimatedSalary < 100193  Q3: 100193 < EstimatedSalary < 149388  Q4: EstimatedSalary > 149388 | N.A | Factor with 4 levels | "1Q", "2Q", "3Q", “4Q” |

## 