Data Immersion

Exercise 5.4

8/25/23

Micky Smith

**Intro to Data Mining**

**Step 2:**

|  |  |  |
| --- | --- | --- |
| **Column** | **Changes Made** | **Reasoning** |
| Row\_number | Removed Column | Unnecessary column |
| Customer\_ID | n/a | n/a |
| Last\_Name | Removed Column | Had errors (incomplete names), provided PII |
| Credit Score | Replaced ‘Blanks’ with N/A | Remove missing cells |
| Country | Changed ES, FR, and DE to Spain, France, Germany | To make country names consistent |
| Gender | Changed F and M to Female/Male. Replaced NULL with N/A. Replaced Femaleemale with N/A | To make the Gender column more uniform with data |
| Age | Changed values showing ‘2’ and NULL to N/A | To make the Age column more uniform with data |
| Tenure | n/a | n/a |
| Balance | n/a | n/a |
| NumOfProducst | n/a | n/a |
| HasCrCard? | n/a | n/a |
| IsActiveMember | n/a | n/a |
| Estimated Salary | Changed Blank and NULL value to N/A | Remove missing values |
| ExitedFromBank? | n/a | n/a |

**Step 3:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Customers** | | | | | | |
|  | **Credit Score** | **Age** | **Tenure** | **Balance** | **NumOfProducts** | **Estimated Salary** |
| Min | 411 | 18 | 0 | 0 | 1 | 371..05 |
| Max | 850 | 82 | 10 | $197,041.80 | 3 | 199,661.50 |
| Mean | 651.6 | 37.5 | 5.2 | $74,830.87 | 1.5 | $98,943.39 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ex-Customers** | | | | | | |
|  | **Credit Score** | **Age** | **Tenure** | **Balance** | **NumOfProducts** | **Estimated Salary** |
| Min | 376 | 22 | 0 | 0 | 1 | 417.41 |
| Max | 850 | 69 | 10 | $213,146.20 | 4 | 199725.39 |
| Mean | 636.5 | 45.3 | 4.7 | $90,239.22 | 1.5 | $97,155.20 |

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated A close-up of a computer screen

Description automatically generated

**Observation:**

From reviewing the data and creating pivot tables to find the average, max, and min of several columns within the data, as well as testing the influence of some of the columns of the data. I found that the biggest risk for Pig E. Banks clients leaving the bank started with active to inactive customers. Then it showed that customer’s age had the next largest effect on them leaving, and typically with a higher average balance, so average balance of over or under 90k plays a factor in customer’s leaving. And gender plays a role since more females leave Pig E. Bank then males, by a larger margin.

**Decision Tree**

A diagram of a bank risk management

Description automatically generated