

## Data Quality Report

### **1. Introduction**

This report documents the data quality audit and cleaning actions performed on the three datasets provided for the outreach and application campaigns. The goal was to prepare analysis-ready datasets by identifying and correcting data quality issues such as missing values, duplicates, inconsistent formats, and invalid entries.

### **2. Datasets and high-level summary**

Outreach Data: originally 37881 rows  $\times$  8 columns; cleaned to 37435 rows  $\times$  8 columns.

Campaign Data: originally 23 rows  $\times$  7 columns; cleaned to 23 rows  $\times$  7 columns.

Applicant Data: originally 37882 rows  $\times$  4 columns; cleaned to 15417 rows  $\times$  4 columns.

### **3. Key issues identified (per dataset)**

#### **3.1 Outreach Data**

Missing values: notable columns in original data: Remark: 33804

Duplicates (entire row): 446 in original; 0 after cleaning

Data type issues: original date/time column stored as text; renamed to 'Received\_At' and converted to Date/Time where parable.

Free-text 'Remark' has many missing values and was left as optional text field.

#### **3.2 Campaign Data**

Missing values (original): None

Duplicates (entire row): 0 in original; 0 after cleaning

Data-type issues: 'Start\_Date' stored as text; converted to Date where passable. IDs kept as text.

#### **3.3 Applicant Data**

Missing values: App\_ID: 1

Duplicates (entire row): 16489 in original; 0 after cleaning

Phone numbers: inconsistent formatting; cleaned to 10 digits-only.

Missing App\_ID rows were removed as they could not be reliably recovered.