**REPORT ON DATA QUALITY ASSESSMENT ON THE CUSTOMER DEMOGRAPHIC DATA OF SPROCKET CENTRAL PTY LIMITED**

The data quality assessment report produced from the systematic survey of the customer demographic data and strategies to mitigate these issues. The statistics were computed using Excel

|  |  |  |  |
| --- | --- | --- | --- |
| **TABLE NAME** | **NUMBER OF RECORD** | **DISTINCT CUSTOMER ID** | **DATE DATA RECEIVED** |
| **TRANSACTION DATA** | **260,000** | **20,000** | **258,449** |
| **CUSTOMER DEMOGRAPHY** | **52,000** | **4,000** | **51,195** |
| **CUSTOMER ADDRESS** | **19,995** | **3,999** | **19,995** |

This was done following the standard data quality dimensions.

**Data accuracy:**

Accuracy in the general statistical sense denotes the closeness of computations or estimates to the (unknown) exact or true values. The customer transaction had about 0.596% inaccuracy while in the customer demographic data observed it can be said to be averagely accurate as some of the values provided seems to have correct details though with about 0.26 % incorrect values observed such as Inaccurate date of birth value, years of purchase for example Karee hymaa’s ,sher clevely and kippy conerding have customer ID but have 0 year of purchase, also some have job industry but no job title although the customers address had accurate data . Additional customer ids in the Transactions table and Customer Address table but not in Customer Master (Customer Demographic)

**Data completeness:**

Various columns, such as the brand of a purchase, or job title, have empty values in certain records. The values provided from the customer demographic data as observe had some values not complete with about 125 last names omitted (3.125 % of the last name data provided), 86 date of birth omitted (2.15 % of the data for the date of birth), 506 job titles were also omitted (making about 12.65% of the data provided for the job title ) and 87 tenure specification were omitted (2.175% of the tenure specification data). In all about 0.01547% data were omitted, the customer transaction data has about 1551 data omitted from the 260,000 total data aim to be recorded which results in ( 0.596 %), and the new customer data has 151 missing data from the 18000 expected data.

**Data consistency:**

Most of the data provided from the customer transaction, new customer list, customer address were free from contradiction though some of the data provided from the customer demography were found contradicting for example Mr Jephthah Bachmann, Cordy Readshall a structural engineer was identified with manufacturing industry, Mose Riddler a Liberian works in an entertainment industry some a high net worth customer like Shirline Yellowlee had 0 year purchase record and 87 gender undecided. Inconsistent values for the same attribute (e.g. Victoria being represented as “V”, “Vic” and “Victoria”, female was also replaced with F

**Data currency (Timely):**

From the data provided about 99.2% of the financial transaction have a valid time stamp.

**Data relevance**:

About 15% of the data provided can be said to be irrelevant ranging from the tenure, deceased record, default and date of birth in the customer demography data section.

**Data validity:**

Though most of the data provided are valid as most of the data would aid effective web analytics. The data was found to be about 99.2% valid. Only a few data such as the default data provided in the customer demography sector pose invalid characters and data values that can be of effect on the quality of the data provided. Inconsistent data type for the same attribute (e.g. numeric values for some fields and strings for others and float or where needed to aid the consistency of analyzing the data)

**Data uniqueness:**

The data provided were quite unique as no duplicate value was found in the data column.

Summarily, the data quality assessment has helped identified areas where correction should made and to rectify that the correction should be effected to enhance analytics activities.