

**To: Sprocket Central Pvt Ltd**

**Subject: Data quality Evaluation of user database**

**Dear Sir or Madam,**

Below is the KPMG's Data quality evaluation report of the dataset. The recommendations have also been provided for the issues found in customer Demographics, customer Address and past transaction data

## Customer Demographics

### Accuracy

There are various first name which have been incorrectly stored, Its recommended to review them.

Customer ID	Name
630	D'arcy
249	D'arcy

A similar issue occurred for last names reading the customer IDs have been incorrectly entered.

Customer ID	Last Name
2590	Tewkesbury.

**Furthermore the date of birth has a significant outlier it should be reviewed**

Customer ID	DOB
34	1843-12-21

**Minor spelling mistake in gender**

Customer ID	Gender
54	Femal

### Consistency

Some of the last name have an apostrophe in them the company should decide weather to keep it or not to maintain consistency some are listed below

ID	LAST NAME
409	O'Kane
440	O'Duggan
473	O'Shirine
567	O'Leahy
613	O'Hallagan
632	O'Hegertie
640	O'Keenan

781	O'Leagham
881	O' Lone

**NOTE** since the majority doesn't have spaces in their last names it is recommended that the wones with spaces be replaced without spaces so as to follow the same trend

Further it is recommended that Sprocket Central Pvt Ltd categorises the gender as Male, Female, U therefor the initials M and F should be replaced with full

## Completeness

The customer demographics database further lacks a lot of data, therefore it is recommended that Sprocket Central Pvt Ltd collects the necessary information to complete the dataset. A software can be easily designed which can help ease this process during data collection.

Field	Entries
DOB	87 Missing entries
Tenure	87 Missing entries
Job title	506 Missing entries

## Validity

The default column is unreadable with gibberish data which leads to lack of accuracy and reliability of dataset. It should be deleted or reconfigured

## Customer Address

### Consistency

It is important to maintain similar trend and names of the states there the following entries need to be changed

The name of the states should be changed to NSW and VIC

State	Number of customers
New South Wales	86
Victoria	82

## Completeness

The following customer IDs are missing therefore it is recommended that the company reviews this

Customer ID
3
10
22
23

Also there were 4000 customers and there are records of only 3999 customers. There is a possibility that data of some customers is missing this should be reviewed

## Transactions Data

**Currency:** The data, therefore, the following transactions should be altered to maintain consistency is missing all transactions from 31 December.

**Consistency:** Since the customer id are given in chronological order the following ids are incorrect.

Cust ID(given )	Correct customer ID	Transaction number
5034	3501	8708,16701,17469

Furthermore, the standard cost maintains a consistent 2 decimal point rounding and therefore should be changed to maintain consistency.

**Completeness:** There are several values in each column that has some missing information. To account for this, it is recommended that Sprocket collect this information from the customers to maintain hg quality of data quality

Column Title	Information missing
Online Order	360
Brand of bike	197
Product line	197
Bike class	197
Bike Size	197
First Sold date	197
Standard cost price	197

Due to recurring value nature of brand, product line standard cost bike class, size and first sold date ,KPMG recommends that sprocket central research into the relation ship between the unknown variables and thereby solve the issue from either educating its staff debugging its program and implementing alert system

## Recommendations for future operations

### Accuracy

Implement double checking and tally mechanism between employee and customer to ensure that all information is accurate and relevant

### Consistency

Enforce rules and data recording guidelines so that all the new data which is entered in the database is consistent with the previous data

### Completeness

Enforcing a system which ensures that the data cannot be entered unless the data is complete i.e. a warning when there is some missing field in the dataset

### **Currency**

Strict guidelines to comply with currency standard. An auto currency converter script can be implemented so as to convert the amount in other currency considering the conversion value at that time.

**Kind Regards**

**Mukhar Jain**

**KPMG Data Analytics Virtual Intern**