**Module 1 Task Thursday, 16 June 2022**

**To, Sprocket Central Pty Ltd.**

**Subject: Data Quality Issues and Software platform used.**

This dataset contains “CustomerDemographic”, “CustomerAddress”, “Transactions” and “New Customer List”.

**Platform**: Microsoft Excel and Tableau.

**Technology used**: ERP Software and RFM analysis.

RFM analysis is a marketing technique for identifying the best customers and executing focused marketing campaigns by quantitatively ranking and grouping them based on the recency, frequency, and monetary amount of their recent transactions.

This dataset is Incomplete and Inconsistent data with more Null, blanks and N/A values. Additionally, that missing data is telling nothing as Some important variables are missing like it’s a customer segmentation we need Age groups, Profit, Customer types, Transaction Month etc.

**Problem Statement:**

Sprocket Central Pty Ltd. It’s a Bicycle company they want to increase their sale analysing these data.

**Exploratory Data Analysis:**

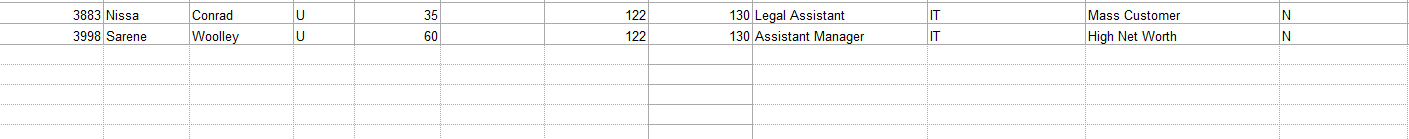
**Important Variables:**

* Deceased Number. (A deceased indicator is a term used to describe the code they use to say that a consumer is dead. This information gets inserted into your credit report. It simply says "deceased". That the end results the deceased indicator.).
* Age Group and Gender.
* Profit and customer buying ratio.

**Missing Vales:**

Customer Demographic:

* Age having a lot of missing value.
* Job title are blank (25%).
* Customer Id’s are missing.
* Deceased have some N/A Values, but I removed that.
* Some unwanted columns have been deleted.
* In DOB People having the age of 170 years which is practically Not possible.
* Gender having three subcategories that Male, Female and U. More than 85 customers indication the U Gender which we are clueless what it’s saying.
* Number of Rows after Cleaning is 3998.



Customer Address:

* Customer Id’s having more Null Values.
* States is not filled properly (In Tableau it’s showing error).
* Number of Rows after Cleaning is 4001.

Graphical user interface, application

Description automatically generated

Transactions:

* Created Transaction Month variable for graphical Representation, Profit variable, Recency variable (How many days it’s taken to purchase that product), Then we create Customer Title using the last 3 years purchase.
* Converted date into short date.
* Converted variables like profit etc. into Dollars.
* 360 customers have blank values in the online order field, and 197 customers have blank values in the product line, product size, and product class columns.
* Number of Rows after Cleaning is 19969.

Graphical user interface, application

Description automatically generated

Transaction data has more rows than customer demographic.

New Customer List:

* Number of Rows after Cleaning is 1002.

Graphical user interface, diagram, application, Word

Description automatically generated

Major Issue : Data Types are the major issue.