**TASK 1**

**Data Quality Assessment**

Assessment of data quality and completeness in preparation for analysis

Hi there,

I’m writing a Data Quality Assessment report for the KPMG Internship by forage. Please note points that I have mentioned below while analyzing the given dataset ( Sprocket Central Pty Ltd Data Set ).

Table - **Transactions**: **20000 Rows, 13 Columns**

* Columns of great importance ( i.e., “brand”, “product\_line”, “product\_class”, “product\_size” ) have 197 blanks in common.
* Column “online\_order” has an importance in accounting which has 360 blanks which could not be ignored or adjusted.\*

*\*concern of high importance*

Table- **NewCustomerList** : **1000 Rows, 18 Columns**

* **Missing join keys**\* ( No primary key found ), Expected key can be

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Issue** | **Remark** |
| DOB | Format | Format should be changed from **Custom to Short Date** |
| DOB | Age Error\* | Blanks have default values as 1900 or less, which implies to Age to be 123 or more. **Blank DOB rows is a concern.** |
| Wealth\_segment | Blanks | 165 |
| Last\_name | Blanks | 29 |
| DOB | Blanks | 17\* |
| Job\_industry\_category | Blanks | 106 |

*\*concern of high importance*

Table- **CustomerDemographic**: **4000 Rows, 13 Columns**

* A join key should be mentioned to refer NewCustomerList table; **Recommended solution is to provide a join key like CustomerID.**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Issue** | **Remark** |
| default | nonsense | Have senseless data and so many blanks |
| gender | Multiple categories | **‘F’, ‘Femal’, ‘Female’** instead of Female  **‘M’ and ‘Male’** for Male |
| Job\_industry | Blanks | 656 |
| Job\_title | Blanks | 506 |
| tenure | Blanks | 87 |
| DOB | Format | Custom to Short Date |
| DOB | 1 Illogical Data\* | Found “**1843-12-21**” that implies to an age of **180** years; which is illogical. |

*\*concern of high importance*

Table- **CustomerAddress**: : **4000 Rows, 6 Columns**

* No blanks found.
* No join keys found\*: it is near to impossible to create a perfect connection between other tables like ‘CustomerDemographic’ or ‘Transactions’ table because ‘CustomerAddress’ doesnot have any join key.
* Recommended solution is to create a join key like ‘CustomerID’.

*\*concern of high importance*

Please refer to the details to ensure company’s data quality assurance and to help them take better decisions.

Thank You

Mayank Dwivedi ( Data Analyst Intern )