|  |  |  |  |
| --- | --- | --- | --- |
| **KPMG Virtual Internship** | | | |
| Sprocket Central Pty Ltd Data Set | | | |
|  |  |  |  |

Module Task-1

The data given and the quality of the data as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset Name | Total Columns | Valid Data Columns | Invalid Columns | Remarks |
| Customer Address | 6 | 6 | 0 | 1.The data is good; we can use for analysis.  2.The quality of the data is good, no missing values, has well structured. |
| Customer Demographic | 26 | 12 | 14 | 1.Data types are not accurate,  Has missing values,  2.the column name” default” has unidentified data. |
| New customer list | 23 | 18 | 5 unidentified | 1.Most of columns has data, the quality of the data is not good |
| Transactions | 13 | 13 |  | 1.The data is good; we can use for analysis.  2. has missing values, cleaning is required. |

From the whole data” Customer address” is good data, the left over requires cleaning after this we may perform some good visualisations.

For example ,gender column has different values, like “Femal”,”F”,”M”, to get clear view for visualisation ,I replaced the vales

Module 2 Task:

We need to clean the data, after the data cleaning we are able gather the 1000 records, from our data. For this we need concentrate on the 2 data sets

1.Customer Address

2.New Customer List

Module 3Task:

I am attaching the Power BI file to this submission, even though it is very small the data model permits to create more ,I did few I am submitting those beautiful insights.

Sample dashboard:

