**Subject line: Information on the data quality issues and strategies to mitigate these issues**.

To Sprocket Central Pty Ltd,

Thank you for providing KPMG with three datasets. After data quality assessment of these three datasets our team found some important key points mention in table given below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Accuracy | Completeness | Consistency | Currency | Relevancy | Validity | Uniqueness |
| Customer  Demographic | DOB is inaccurate as customer with (1843-12-21) DOB is not possible. | DOB,  Job titles,  Job industry category,  Default,  Tenure,  for some customers values are missing or n/a | Gender data has inconsistency  Like: (F, Femal, Female, M, Male, U) |  |  |  |  |
| Customer address |  | Customer  demographic data:  total records:  4000 (Id 1 to 4000) customer address data:  total records:  3999 (Id 1 to 4003) | State data has inconsistency  (New South Wales, VIC, QLD, NSW, Victoria) |  |  |  |  |
| Transaction data of three months |  | For customer id 5034, does not found any record in other two datasets.  Incomplete values in  online order, brand, product line, product class, product size, standard cost, product first sold date |  |  |  | There are many records with product id ‘0’ |  |
| Recommendations to mitigate current data quality concerns. | There are some outliers that can affect the whole analysis process, try to remove them | Missing values should be treated well accordingly and for mismatched values try to filter them from records or match them from other datasets | Replace data to remove inconsistency |  |  | Check Product information dataset for confirmation is product id ‘0’ exists? |  |