Dear John Doe,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The summary table below showcases key data findings within the three datasets. Please let us know if you have any question regarding the findings presented.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** | **Uniqueness** |
| **Transactions** |  | Missing values found:  online\_order: 360  brand: 197  product line: 197  product\_class: 197  product\_size: 197  standard\_cost: 197  product\_first\_sold\_date: 197 |  |  |  | There are some records with product\_id = 0 |  |
| **Customer Demographic** | DOB is inaccurate | last\_name: 125  DOB: 87  job\_title: 506  job\_industry\_category: 656  default: 302  tenure: 87 | Inconsistency found in the gender column:  [F, Female, U, Femal, M] |  | There are some records with the deceased customers  Deceased:02 |  |  |
| **Address** |  | customer\_ids are different when compare with demographics data.  min: 1,  max: 4000  total records: 3999 | state data has inconsistency  [New South Wales, QLD, VIC, NSW, Victoria] |  |  |  |  |

Below we have provided detail description of findings and mitigation methods and suggestions,

1. **Accuracy** 
   1. **Outliers are detected in the DOB column**
      1. *Mitigation: remove outliers*
      2. *Suggestions: create new feature age, age feature will help to easily locate the outliers and add information.*
2. **Completeness** 
   1. **There are some missing data**
      1. *Mitigation: apply missing value treatments*
   2. customer\_id is incomplete among customer\_demographic and customer\_address
      1. *Mitigation: use only records which are available on both sheets and filter out all other incomplete data or update spreadsheets with correct data*
3. **Consistency**
   1. **Inconsistency gender data in customer\_demographics**
   2. **Inconstancy state data in customer\_address** 
      1. *Mitigation: replace data to remove inconsistency*
4. **Validity**
   1. **There are some records with product\_id = 0** 
      1. *Mitigation: check database whether there is 0 product\_id*
5. **Relevancy**
   1. **There are some records with where the customer is deceased.** 
      1. *Mitigation: Remove the rows as the deceased records are no longer relevant.*

That summarises all the data quality issues discovered through the first stage of the data quality analysis. The mitigation strategies suggested are simple and effective ways of improving data quality for future analysis. They will not only improve the analysis output that one can perform within the company but will increase the level of analysis that can be performed by KPMG and other hired analysis teams.

Please let us know if you have questions regarding mitigation or any data quality issues identified.

Kind regards,

Mohd Shadab