Dear Client,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The summary table below highlights key data assessment findings we discovered 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 |  | There are some missing data,  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 |  |
| Demographics | DOB is inaccurate | last\_name: 125  DOB: 87  job\_title: 506  job\_industry\_category: 656  default: 302  tenure: 87 | gender data has inconsistency  [F,Female,U,Femal,M] |  |  |  |  |
| Address |  | customer\_ids are different when compare with demographics data.  min: 1,  max: 4003  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. Date of birth data is not accurate there are some outliers
      1. *Mitigation: remove outliers*
      2. *Suggestions: create new feature age, age feature will help to easily understand errors*
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*