# Task 1:

**Name Shagufta Perveen**

# Subject: Data Quality Issues – Sprocket Central Pty Ltd

Dear *Uzair Sarfraz*,

Thank you for providing the three datasets from Sprocket Central Pty Ltd. Based on our preliminary review, we have identified several data quality issues that could impact the accuracy and effectiveness of our analysis. Please see the summary below:

## Identified Data Quality Issues

|  |  |
| --- | --- |
| **Dataset** | **Issues Identified** |
| Customer Demographic | - Missing values in job\_title, last\_name, DOB, tenure - Inconsistent gender entries ("F", "Female", "female") - DOB values indicating unrealistic ages (e.g., over 100 years) |
| Customer Address | - state values are inconsistent (e.g., mix of "NSW" and "New South Wales") |
| Transactions | - Missing values in online\_order, brand, product\_line, product\_class, product\_size, standard\_cost, product\_first\_sold\_date - Some customer\_id entries do not match records in the other two datasets |

## Recommendations to Improve Data Quality

* Fill in missing values for important fields or confirm if they can be excluded
* Standardize categorical values, such as:
* - Gender: Use consistent terms like "Male" and "Female"
* - State: Use standard abbreviations (e.g., "NSW", "VIC")
* Validate and clean DOB values to ensure realistic age ranges (e.g., 18–90)
* Ensure key integrity: All customer\_id values in transactions should match with customer demographic and address records
* Include additional columns if possible, like derived age from DOB or transaction month
* Implement validation rules in data entry to avoid future inconsistencies

We will be able to proceed with a more accurate analysis once these quality issues are addressed or clarified.  
Please let me know if you have any questions or if you'd like assistance in cleaning the data.

Kind regards,  
Shagufta Perveen