KPMG VI Task 1- Data Quality Issues

Task- Draft an email to the client identifying the data quality issues and strategies to mitigate these issues.

To: Sprocket Central Pty Ltd

Subject: Data Quality issues and strategies to mitigate these issues.

Respected Sir/Madam,

Below are the discovered issues in your data.

1. Completeness

The following are the columns in the various worksheets that have missing values.

Transactions sheet → *online_order, brand, product_line, product_class, product_size, standard cost, product first sold date* contain null values.

NewCustomerList → last_name, *DOB, job title, job industry category* contain null values.

CustomerDemographic → last name, *DOB, job title, job industry category, default, tenure* contain null values.

★ Your team can provide more old information and data records to fill the null values.

2. Consistency

Some records do not show consistency in the tables,

Customer Demographics → gender ⇒ use of *"F"* and *"Female"* to refer to *"Female" and "M"* and *"Male"* to refer to 'Male'

Customer Address → state ⇒ Use of both eg. *"NSW"* and *"New South Wales"*

★ The rest of the values can be formatted to be similar to the rest.

3. Validity

Some records contain unallowable values,

Customer Demographics → gender ⇒ presence of the gender *"U"*, cannot tell whether the said individuals were Male or Female.

★ The Data Handling team should tally the record of "U" in gender to find

out the correct values.

4. **Relevancy**

Customer Demographics → The column *"default"* has undecodable data

→ The column *"deceased indicator"* has *"Y" to indicate they are dead*

★ These columns will be deleted as will not help with data analysis