

Task 1: Data Quality Assessment

Here is the background information on your task

Sprocket Central Pty Ltd, a medium size bikes & cycling accessories organisation, has approached Tony Smith (Partner) in KPMG's Lighthouse & Innovation Team. Sprocket Central Pty Ltd is keen to learn more about KPMG's expertise in its Analytics, Information & Modelling team.

Smith discusses KPMG's expertise in this space (you can read more [here](#)). In particular, he speaks about how the team can effectively analyse the datasets to help Sprocket Central Pty Ltd grow its business.

Primarily, Sprocket Central Pty Ltd needs help with its customer and transactions data. The organisation has a large dataset relating to its customers, but their team is unsure how to effectively analyse it to help optimise its marketing strategy.

However, in order to support the analysis, you speak to the Associate Director for some ideas and she advised that *"the importance of optimising the quality of customer datasets cannot be underestimated. The better the quality of the dataset, the better chance you will be able to use it drive company growth."*

The client provided KPMG with 3 datasets:

- Customer Demographic
- Customer Addresses
- Transactions data in the past 3 months

You decide to start the preliminary data exploration and identify ways to improve the quality of Sprocket Central Pty Ltd's data.

Here is your task

You arrive at your desk after the initial client meeting. You have a voicemail on your phone which contains the following instructions.

"Hi there – Welcome again to the team! The client has asked our team to assess the quality of their data; as well as make recommendations on ways to clean the underlying data and mitigate these issues. Can you please take a look at the datasets we've received and draft an email to them identifying the data quality issues and how this may impact our analysis going forward?"

I will send through an example of a typical data quality framework that can be used as a guide. Remember to consider the join keys between the tables too. Thanks again for your help."

Hi there,

As per voicemail, please find the 3 datasets attached from Sprocket Central Pty Ltd:

- Customer Demographic
- Customer Addresses
- Transaction data in the past three months

Can you please review the data quality to ensure that it is ready for our analysis in phase two. Remember to take note of any assumptions or issues we need to go back to the client on. As well as recommendations going forward to mitigate current data quality concerns.

I've also attached a data quality framework as a guideline. Let me know if you have any questions.

Thanks for your help.

Kind Regards
Your Manager

Here is your task:

Draft an email to the client identifying the data quality issues and strategies to mitigate these issues. Refer to 'Data Quality Framework Table' and resources below for criteria and dimensions which you should consider.

You may use programs like Excel, [Google Sheets](#), [Tableau](#), [Power BI](#) to start. Feel free to use [Python](#), [R Programming Language](#), [MatLab](#) and other data analytics tools that you know of.

Data Quality Framework Table

Below is a list of the Data Quality dimensions our team may use to evaluate a dataset. Some of these terms are common to the whole industry, so you may find more information and clarity on these terms by searching online.

Standard Data Quality Dimensions	
Correct Values	Accuracy
Data Fields with Values	Completeness
Values Free from Contradiction	Consistency
Values up to Date	Currency
Data Items with Value Meta-data	Relevancy
Data Containing Allowable Values	Validity
Records that are Duplicated	Uniqueness

Submit your work

You may only credit this task to your CV if you make a genuine attempt at the work. You may resubmit work, iterate upon and reattempt work as many times as you want.