**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**](https://home.kpmg/au/en/home/services/advisory/management-consulting/digital/data-analytics-modelling.html)). 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.

*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.*

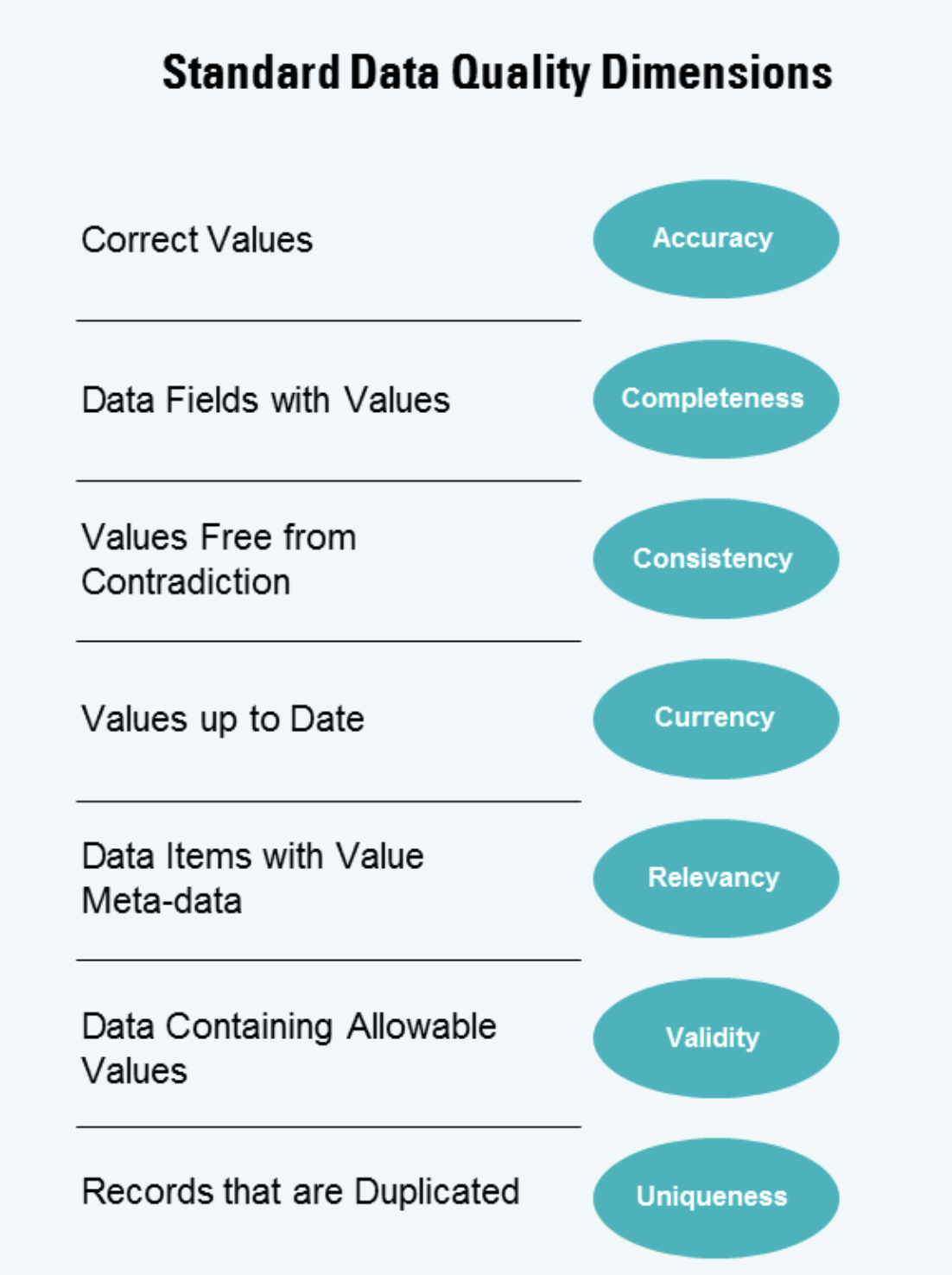
*I’ve also attached a data quality framework as a guideline. Let me know if you have any questions.*

*Thanks for your help.*

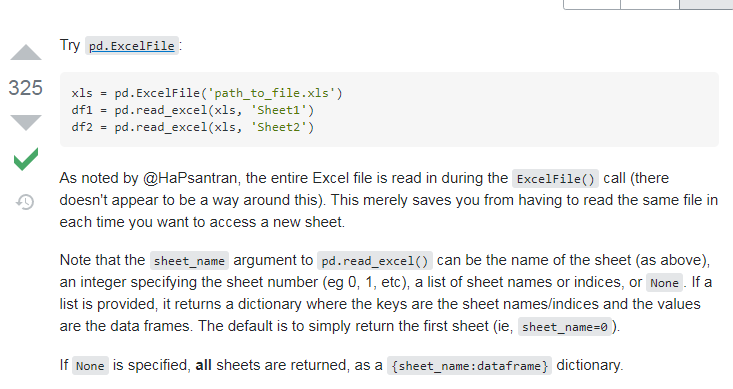
*Kind Regards  
Your Manager*

**Resources to help you with the task**

**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.



Reading Multiple Sheets with Pandas



1. Transactions:

* online\_order, brand, product\_line, product\_class, product\_size, standard\_cost, product\_first\_sold\_date- has null values.
* product\_first\_sold\_date- Datatype need to be converted into datetime,but is in float at present.

1. NewCustomerList:

* Column no.16,17,18,19,20 are missing.
* 'last\_name', 'DOB', 'job\_title', 'job\_industry\_category'- has null values
* ‘deceased\_indicator’ is a constant feature.
* property\_valuation- Unit not mentioned
* Value- Unit not mentioned

1. CustomerDemographic:

* Gender- has typing mistakes
* Default- has many inconsistent/garbage values

1. CustomerAddress:

* Country- is a constant feature
* property\_valuation- Unit not mentioned