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.

**Task 1: Data Quality Assessment**

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.

[Voicemail transcript below]

“*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*.”

[Read email below]

*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.

**Items Required to do this module:**

**[Required File: The dataset from Sprocket Central](https://cdn-assets.theforage.com/vinternship_modules/kpmg_data_analytics/KPMG_VI_New_raw_data_update_final.xlsx" \t "_blank)**

[This is the dataset to evaluate for data quality. Read the instructions on the title sheet. There are multiple sheets.](https://cdn-assets.theforage.com/vinternship_modules/kpmg_data_analytics/KPMG_VI_New_raw_data_update_final.xlsx" \t "_blank)

[KPMG\_Virtual\_internship\_dataset\_insidesherpa.xlsx](https://cdn-assets.theforage.com/vinternship_modules/kpmg_data_analytics/KPMG_VI_New_raw_data_update_final.xlsx" \t "_blank)

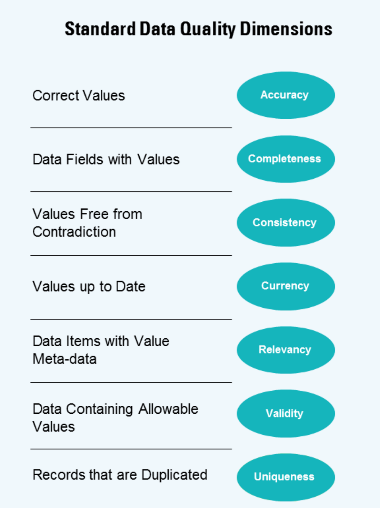
**[Click to download file →](https://cdn-assets.theforage.com/vinternship_modules/kpmg_data_analytics/KPMG_VI_New_raw_data_update_final.xlsx" \t "_blank)**

[File copyright KPMG. For use for the Virtual Internship Only.](https://cdn-assets.theforage.com/vinternship_modules/kpmg_data_analytics/KPMG_VI_New_raw_data_update_final.xlsx" \t "_blank)

You may use programs like Excel, [Google Sheets,](https://www.google.com.au/sheets/about/) [Tableau,](https://www.tableau.com/) P[ower BI to start. Feel free to use Python, R Programming Language, MatLab and other data analytics tools that you know of.](https://powerbi.microsoft.com/en-us/)

Here are some resources to help you

**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.



**Task 2: Data Insights**

Here is your task

For context, Sprocket Central Pty Ltd is a long-standing KPMG client whom specialises in high-quality bikes and accessible cycling accessories to riders. Their marketing team is looking to boost business by analysing their existing customer dataset to determine customer trends and behaviour.   
  
Using the existing 3 datasets (Customer demographic, customer address and transactions) as a labelled dataset, please recommend which of these 1000 new customers should be targeted to drive the most value for the organisation.   
  
In building this recommendation, we need to start with a PowerPoint presentation which outlines the approach which we will be taking. The client has agreed on a 3 week scope with the following 3 phases as follows - Data Exploration; Model Development and Interpretation.

Prepare a detailed approach for completing the analysis including activities – i.e. understanding the data distributions, feature engineering, data transformations, modelling, results interpretation and reporting. This detailed plan needs to be presented to the client to get a sign-off. Please advise what steps you would take.

Please ensure your PowerPoint presentation includes a detailed approach for our strategy behind each of the 3 phases including activities involved in each - i.e. understanding the data distributions, feature engineering, data transformations, modelling, results interpretation and reporting. This detailed plan needs to be presented to the client to get a sign-off.

Tips: Raw data fields may be transformed into other calculated fields for modelling purposes (i.e. converting D.O.B to age or age groups).  Tips: You may source external data from the ABS / Census to add additional variables that may help support your model.

**Task 3: Data Insights and Presentation**

Here is your task

Please develop a dashboard that we can present to the client at our next meeting. Display your data summary and results of the analysis in a dashboard (see tools/references for assistance).  Maximum of 3 dashboard views/tabs, creativity in layout and presentation is welcome.

As this is not a KPMG branded deliverable, please find attached a client logo for incorporation.

A black background with a black square

Description automatically generated with medium confidence

It is important to keep in mind the business context when presenting your findings:

* What are the trends in the underlying data?
* Which customer segment has the highest customer value?
* What do you propose should be Sprocket Central Pty Ltd ’s marketing and growth strategy?
* What additional external datasets may be useful to obtain greater insights into customer preferences and propensity to purchase the products?

Specifically, your presentation should specify who Sprocket Central Pty Ltd’s marketing team should be targeting out of the new 1000 customer list as well as the broader market segment to reach out to.