Forage(/dashboard)

Explore >   **My Dashboard**(/dashboard)   ›

Task One

1   2   3   4   5

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

Back

?

Next

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:

XLS

Required File: The dataset from Sprocket Central

This is the dataset to evaluate for data quality. Read the instructions on the title sheet. There are multiple sheets.
KPMG_Virtual_internship_dataset_insidesherpa.xlsx

Click to download file →

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) You may use programs like Excel, Google Sheets, (https://www.google.com.au/sheets/about/) Tableau, (https://www.tableau.com/) Power 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.

?

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**

Upload your email draft

⌂ **Select a File**   or drag it here.

?