KPMG Virtual Internship
Sprocket Central Pty Ltd Data Set

To view the datasets, click on the yellow tabs/sheets at the bottom of this file
Click on the sheets are named "CustomerDemographic", "CustomerAddress", "Transactions" and "NewCustomerList"

Module 1 Task

Your task is to take a look at the following datasets provided by Sprocket Central Pty Ltd and identify all data quality issues. Once you've had a look at these datasets, draft an email to the client identifying all data quality issues.

Module 2 Task

Create a PowerPoint presentation wihch outlines the approach we will be taking to identify which of the 1000 customers Sprocket Central Pty Ltd should target, based on this dataset. Explain the three phases: Data Exploration; Model Development and Interpretation.

Module 3 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). Specifically, your presentation should specify who Sprocket Central Pty Ltd' should be targeting out of the new 1000 customer list.
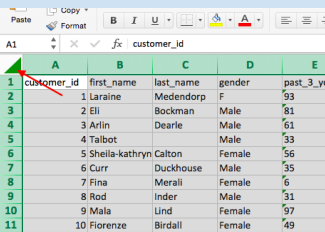
Return to the virtual internship here:
https://www.insidesherpa.com/virtual-internships/theme/m7W4GMqeT3bh9Nb2c/KPMG-Data-Analytics-Virtual-Internship

Need some help getting started, or new to excel/data?
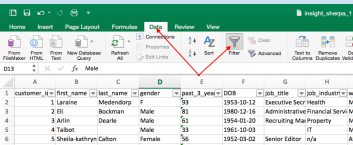Here is a warm up task with this dataset

Warm up task:
1) Look over the datasets provided by Sprocket Central Pty Ltd (you can navigate through sheets by clicking the tabs at the bottom).
2) Go to the 'CustomerDemographic' dataset
3) Select entire dataset by clicking cell in the uppermost top left hand corner (see picture below)

4) Once you've selected the entire dataset, click 'Data' in the top bar of the sheet and click 'Filter' (see picture below). Once you click this, you'll notice that arrows appear in each cell in the top row. If you click the arrows, this creates a pop-up which allows you to only select certain values. Make sure you understand how to use the 'Filter' tool properly as this will be your main tool for

5) Go to the 'DOB' column and click the filter arrow. This should bring up a pop-up which allows you to sort your data. Click 'Ascending' and this should sort your data-set from least to most recent from top to bottom (by date of birth). Once you do this, your data set should look exactly the same as the picture below

6) Now do you notice anything strange about the data? If you have a look at the first customer (Jephthah Bachmann), you'll notice that they were born in 1843, meaning that they are 175 years old! This is obviously an error in the data so it would be something you'd include in the email to the client.

7) For the rest of this Module 1 task, it is up to you to look through the client datasets and identify problems like in the example above with the DOB. We recommend you dot point everything you find in a Word Document before drafting your email. It is also really important that you use the 'Data Quality Framework Table' (you can find this back in the module page in the 'Resources' section). Note that the parameters and criteria provided to you is vague because this is reflective of what happens in the real world. We encourage you to use other sources like Google and Youtube to help you with this task. Even if you're unsure about the quality of your final email, just give it a go, submit the task and you'll get to see model work which is an actual email that a KPMG Data Analyst has submitted!