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 "GustomerDemographic". "GustomerAddress". "Transactions" and "NewCustomerDemographic".

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?

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)

| Pa | ste Sepp | B 1 | <u>u</u> + 🖽 + 🔇 | - <u>A</u> - | = - - | | | | | | |
|----|-------------|----------------|------------------|--------------|--|--|--|--|--|--|--|
| A1 | A1 | | | | | | | | | | |
| 4 | Α | В | С | D | E | | | | | | |
| 1 | customer_id | first_name | last_name | gender | past_3_ye | | | | | | |
| 2 | 1 | Laraine | Medendorp | F | 93 | | | | | | |
| 3 | 2 | Eli | Bockman | Male | 81 | | | | | | |
| 4 | 3 | Arlin | Dearle | Male | 61 | | | | | | |
| 5 | 4 | Talbot | | Male | 33 | | | | | | |
| 6 | 5 | Sheila-kathryn | Calton | Female | 56 | | | | | | |
| 7 | 6 | Curr | Duckhouse | Male | 35 | | | | | | |
| 8 | 7 | Fina | Merali | Female | 6 | | | | | | |
| 9 | 8 | Rod | Inder | Male | 31 | | | | | | |
| 10 | 9 | Mala | Lind | Female | 97 | | | | | | |
| 11 | 10 | Fiorenze | Birdall | Female | 49 | | | | | | |

4) Once you've selected the entire dataset, click 'Data' in the top bar of the sheet and click 'Filler' (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 'Filler' 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 up to bottom (by date of birth). Once you do this, your data set should look exactly the same as the picture below

| | A | В | C | D | Е | F | G | Н | 1 | J | K | L | |
|----|--------------|--------------|---------------|---------|----------------|-------------|------------------|-----------------|----------------|---------------|---------------------|----------|-----|
| | customer_ic= | first_name = | last_name = | gender | w past_3_yearw | DOB - | job_title w | job_industri = | wealth_segi= | deceased_it v | default v | owns_car | ¥ 1 |
| 2 | 34 | Jephthah | Bachmarn | U | 59 | 1843-12-21 | Legal Assistant | п | Affluent Custo | N | | No | |
| 3 | 1182 | Carolann | Ottiwill | Female | 84 | 1932-03-15 | Data Coordiate | п | Mass Custome | 2N | \$3964 N NAPSÚ | Yes | |
| 4 | 1099 | Jarrett | Maffia | Male | 12 | 1932-03-17 | Tax Accountan | Manufacturing | Affluent Custo | N | \$3¢ | Yes | |
| 5 | 2567 | Tobe | Spirit | Male | 97 | 1932-03-17 | Systems Admir | Financial Servi | Mass Custome | 2N | THE . | No | |
| 6 | 3621 | Umeko | Hearfield | Female | 13 | 1932-03-20 | Developer IV | Health | High Net Wor | t N | AxAcsezeză | No | |
| 7 | | Rickey | Gregolotti | Male | 98 | 1932-04-30 | Technical Writ | Manufacturing | Mass Customs | N | 1.00E+02 | Yes | |
| 8 | | Jozef | Frizzell | Male | 12 | 1932-05-15 | Graphic Design | | High Net Wor | | [0618020 8020 80 | No | |
| 9 | | Hamil | Britch | Male | 13 | 1932-05-28 | Media Manage | | | | | No | |
| 10 | | Kirsten | Slegg | Female | 45 | 1932-06-15 | | Financial Servi | | | PARTONNARE WITH | | |
| 11 | 3326 | | Crotch | Male | 83 | 1932-06-15 | | Manufacturing | | | 48 | Yes | |
| 12 | 766 | Nonna | Hun | Female | 16 | 1932-06-24 | | Retail | Mass Custome | :N | \$3etest\$3e | Yes | |
| 13 | 389 | Dorene | Scrivner | Female | 97 | 1932-06-29 | Help Desk Ope | Entertainment | Affluent Custo | N | 230 | Yes | |
| 14 | 901 | Letta | McCarter | Female | 13 | 1932-06-29 | | Retail | Mass Custome | 2N | (06)\$330 \$350 \$4 | Yes | |
| 15 | 3547 | Ganny | Dwerryhouse | Male | 19 | 1932-07-08 | Senior Financia | Financial Servi | Affluent Custo | N | 1-Feb | No | |
| 16 | 256 | Unell | Beadle | Female | 16 | 1932-07-09 | Director of Sal- | Health | High Net Wor | t N | NIL | Yes | |
| 17 | 170 | Jammal | Gever | Male | 80 | 1932-08-04 | | | High Net Wor | | \$x0_0mins | No | |
| 18 | | Enrichetta | Broadberry | Female | 12 | 1932-08-04 | Budget/Accou | | | | 6000 | No | |
| 19 | 3523 | | Tarver | Male | 33 | 1932-08-16 | | | Affluent Custo | | \$335eE%seXCe | | |
| 20 | | Idette | Mapam | Female | 51 | 1932-08-17 | Web Designer | | Mass Customs | | 83*83*test85* | Yes | |
| 21 | | Mercy | Donisi | Female | 19 | 1932-08-20 | Web Designer | | | | #300eE%seXCe | | |
| 22 | 3853 | Kerri | Marrington | Female | 91 | 1932-08-28 | Accounting As: | n/a | Mass Customs | i N | 83£ | Yes | |
| 23 | | Colas | Pumfrett | Male | 35 | 1932-09-10 | Business Syste | | High Not Wor | | \$1.00 | No | |
| 24 | 640 | Galvin | O'Keenan | Male | 12 | 1932-09-13 | Librarian | Entertainment | High Net Wor | i N | o . | Yes | |
| 25 | 1090 | Lorettalorna | | Female | 15 | 1932-09-16 | Computer Syst | Telecommunic | Affluent Custo | N | 1.00E+96 | No | |
| 26 | 1927 | Masha | Bibble | Female | 11 | 1932-09-16 | Software Test | Manufacturing | Affluent Custo | N | 800014 80000 80 | Yes | |
| 27 | 2452 | Sophronia | Yerrill | Female | 93 | 1932-09-25 | Chemical Engir | Manufacturing | Mass Custome | :N | \$32 | Yes | |
| 28 | | Cristobal | Whitrod | Male | 63 | 1932-10-09 | Speech Patholi | | Mass Customs | N | \$3"85" \$3µ\$35 | | |
| 29 | 3995 | Rosalia | Halgarth | Female | 8 | 1932-10-11 | VP Product Me | | Mass Customs | | -1.00E+02 | No | |
| 30 | | Gaye | Quickenden | Female | 48 | 1932-10-19 | Account Repre | | Affluent Custo | N | 092051(A1803) | | |
| 31 | 2115 | Shennon | Pflieger | Male | 17 | 1932-11-02 | Food Chemist | Health | Mass Customs | N | /dev/null; tou | Yes | |
| 7 | ► Title Pag | See Cus | lomerDemograp | No Onto | | reactions N | reCustomerList | + | | | | | |

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 Analysh has submitted!