To Concerned authority,

I am glad that you offered me this opportunity to work upon.Thanks for providing with three datasets from Sprocket Central Pty.Ltd. I went through the whole data set all sheets in excel. I would like put forward few points regarding those, which might help the concern authority working on it for better visualization. At the very start I would like bring your attending to the line which is provided at every sheet of this workbook which made different for me read excel sheet the way I wanted so I modified those by deleting them from all sheets.

Starting with the very first sheet namely ‘Transactions’ in the excel sheet provided to me for reference which basically involves details of transactions for certain period of time. Firstly I needed to drop unnamed columns. There where in all 2000 transaction details of 101 different products, 3494 different customer id’s. Same customers have arrived several times.

|  |  |
| --- | --- |
| ***Columns in ‘Transactions’*** | ***Observations*** |
| *online\_order* | Presence of few 0’s and 1’s and few non numeric values which I safely converted to zero. |
| *Brand* | 2897 entries where not specified of 6 different brands |
| *product\_line,*  *product\_class,*  *product\_size,*  *standard\_cost,*  *product\_first\_sold\_date* | 197 entries where not specified |

On starting with sheet 2 namely which includes 5 unnamed columns which were hide in excel. This sheet majorly included details of new customers including their first name some without last and few personal details.

|  |  |
| --- | --- |
| ***Columns in*** NewCustomerList’ | ***Observations*** |
| *gender* | Presense of few U’s I renamed them to unspecified. |
| *Dob* | 17 entries where not specified |
| *job\_title* | 106 entries where not specified with 167 different job title. |
| job\_industry\_category | 165 not specified 9 distinct job industry category |

In third sheet namely ’ CustomerDemographic’ highlighting experience of person in terms of job . The ‘default’ column of this sheet includes few characters which are not readable , they are of no meaning.

|  |  |
| --- | --- |
| ***Columns in*** CustomerDemographic’ | ***Observations*** |
| *DOB* | 67 entries were not specified |
| *job\_title* | 506 entries were not specified with 195 distinct entries |
| job\_industry\_category | 656 not specified and 9 distinct entries |
| tenure | 85 not specified and 22 distinct job tenture |

On looking at fourth sheet namely ‘CustomerAddress’, which says about location of customer. This sheet holds records of 4000 customers in Australia.

|  |  |
| --- | --- |
| ***Columns in*** ‘CustomerAddress’ | ***Observations*** |
| *post* | 873 different postal address |
| *states* | 5 different states |

All the not specified values can be changed to something like empty space or ‘Not defined’ at all those places. Hence, making the whole excel file readable with proper entries. Please do let me know if anything is not clear.

Regards,

Pranita Shinde, a KPMG virtual internship candidate.