Greeting from KPMG Data Analytics Team,

This is the work I have done on the raw dataset I have received from your side and the following are the key points that I have noticed using my analysis using Excel and Python that could be rectified that would help us do the analysis and visualization better.

**Points To Work On:**

**For the sheet:** ‘CustomerDemographic’

* There are 1763 n/a values in the sheet “CustomerDemographic”
* Inconsistency of the Gender column found values containing “'F', 'Male', 'Female', 'U', 'Femal', 'M'”.
* The job\_industry\_category column had additional n/a values of 656.
* The job\_title column had titles following the roman numerals (which indicate job titles in different sectors) which could cause problems during later analysis:

Account Representative, Accountant , Accounting Assistant, Administrative Assistant, Automation Specialist , Biostatistician , Budget/Accounting Analyst , Computer Systems Analyst, Database Administrator, Developer, Engineer , Geologist , Health Coach , Human Resources Assistant , Media Manager, Office Assistant, Programmer Analyst, Programmer, Research Assistant , Safety Technician, Software Engineer, Software Test Engineer, Staff Accountant, Systems Administrator, Web Designer, Web Developer.

**For the sheet:** ‘Transactions’

* There are 1542 n/a values in the sheet “Transactions”.
* Column “customer\_id “has 5034 as an incorrect data as the maximum number is 3500.
* Transaction ID: 8708, 16701, 17469 has incorrectly rounded “standard\_cost” without the type of currency as 312.7350159, 270.2999878, 667.4000244 respectively.

**For the sheet:** ‘NewCustomerList’

* There are 317 n/a values in the sheet “NewCustomerList”
* New customer list dataset has 2 columns that has numerical data but was given the input as string, which are: past\_3\_years\_bike\_related\_purchases, Postcode & Property code
* The job\_industry\_category column had additional n/a values of 166.
* The job\_title column had titles following the roman numerals (which indicate job titles in different sectors) which could cause problems during later analysis as same as ‘CustomerDemographic’.
* The Values column has numbers rounded inconsistently.

**For the sheet:** ‘CustomerAddress’

* There are no n/a values.

I hope my analysis help you in resolving the issues and you will be in touch with us ASAP to continue our work together.