Sagar Limbu

# KPMG Data Quality Assessment

1. Customer demographic:
   1. DOB:
      1. Earliest DOB is 1843-12-21 which is 177 years ago. It is not possible for such DOB of a customer to exist in the database
   2. Gender:
      1. gender: ['F' 'Male' 'Female' 'U' 'Femal' 'M']
      2. customer id 54 has gender ’Femal’
   3. default:
      1. column name ‘default’ has rubbish characters and strings.
   4. Total missing values(Nan ) from Each column:
      1. customer\_id : 0
      2. first\_name : 0
      3. last\_name : 125
      4. gender : 0
      5. past\_3\_years\_bike\_related\_purchases : 0
      6. DOB : 87
      7. job\_title : 506
      8. job\_industry\_category : 656
      9. wealth\_segment : 0
      10. deceased\_indicator : 0
      11. default : 302
      12. owns\_car : 0
      13. tenure : 87
2. Transaction:
   1. Total missing values in each column:
      1. transaction\_id : 0
      2. product\_id : 0
      3. customer\_id : 0
      4. transaction\_date : 0
      5. online\_order : 360
      6. order\_status : 0
      7. brand : 197
      8. product\_line : 197
      9. product\_class : 197
      10. product\_size : 197
      11. list\_price : 0
      12. standard\_cost : 197
      13. product\_first\_sold\_date : 197
   2. column ‘product\_first\_sold\_date’: This column ‘arbitrary’ values listed. No proper date format. EG: 41245.0, 41701.0, 36361.0, 36145.0, 42226.0
   3. Column ‘brand’: It has value ‘Giant Bicycles’ which can be contradicting with the actual size of the product
   4. Duplicate records on columns: address. EG:
      1. Table

         Description automatically generated
3. Customer Address:
   * 1. Invalid address: Column ‘Address’ has invalid address, starting with string. Total of 84 records that has invalid address. Address starting with ‘0 ‘

EG:

* + - 1. 0 Holy Cross Court,
      2. 0 David Terrace
      3. 0 Pearson Crossing
      4. 0 Golden Leaf Crossing
    1. Property Valuation: The valuation of property has the rating of 1 to 11.