Task 1

Dear Sir/Madam,

As per your last e-mail, I went throught the 3 datasets attached from Sprocket Central Pty Ltd namely:

- Customer Demographic
- Customer Addresses
- Transaction data in the past three months

The following list includes notable data quality issues that were found and the techniques used to address the found data inconsistencies. In addition, suggestions have been made to prevent the recurrence of data quality problems and enhance the accuracy of the underlying data that serves as the basis for business decisions.

  1)Transactions sheet:

   a)Blanks in:

   - online_order
   - brand
   - product_line
   - product_class
   - product_size
   - standard_cost
   - product_first_sold_day

   Solution: delete the blank rows

   b)product_first_sold_day values aren't clear. The values dont show clearly what they want to express. I would like some to know more about it in some detail.

  2)NewCustomerList sheet:

   a)Blanks in:

   - Job title
   - DOB

   Solution: delete the blank rows

  3)CustomerDemographic sheet:

   a)Blanks in:

   - Customer_id
   - First_name
   - Gender
   - Past_3_years_bike_related_purchases
   - DOB
   - Job_title

- Job_industry
- Wealth_segment
- Decreased_indicator
- Default
- Owns_car
- Tenure

Solution: delete the blank rows

b)The gender column has Femal, Female, and Male duplicates of F and M causing data inconsistency.

Solution: replace the Female, Femal values with F and the Male values with M.

c)DOB has a date for 1843-12-21which is not valid.

Solution: this particular row can be deleted.

d)While importing data, the default column couldn't import the rows in the original form and hence has data which is difficult to be deciphered and put to use further.

4)CustomerAddress sheet:

a)Blanks in:
- Customer_id
- Address
- Post_code
- State
- Country
- property_valuation

Solution: delete the blank rows

b) in the state column, the New South Wales attribute is indicated by both NSW and the full form which causes discrepancies in data

Solution: replace 'New South Wales' by 'NSW'

The team will keep up with the process of standardising, cleansing, and transforming data in order to analyse models in the future. Along the journey, queries and presumptions will be recorded.

After we've finished, it would be excellent to meet with your data SMEs to make sure that all of the presumptions are in line with what Sprocket Central is aware of.

Kind Regards,
Sanika Ardekar.