Dear Manager,

I am Cindy from KPMG Data Analytics (Virtual Internship) team. We confirm that I did receive an XLS file named “KPMG\_VI\_New\_raw\_data\_update\_final” including 05 sheets (Title Sheet, Transactions, NewCustomerList, CustomerDemographic, CustomerAddress) from you on 25/01/2021.

We reviewed the data quality. It looks well. However, they remain some issues we will list down below.

**1. Transactions Sheet**

Number records: 2000

**Table 1: Issues in Transactions sheet**

|  |  |  |
| --- | --- | --- |
| Field name | Issue | How to fix |
| customer\_id | Duplicate records exist (16506 cases) | Remove duplicate value, 3494 unique value remain |
| customer\_id | Whether is it a type error for ID 5034? | Check and fix |
| online\_order | 55 blank records | Check and fill or remove |
| brand | 26 blank records | All cases have the correspondent customer ID (listed in table 2). Therefore, we wonder whether lacking typing information.  => Check and fill or remove |
| product\_line | 26 blank records |
| product\_class | 26 blank records |
| product\_size | 26 blank records |
| standard\_cost | 26 blank records |
| product\_first\_sold\_date | 26 blank records |
| product\_first\_sold\_date | Format compliance | Change to date format |
| list\_price | Format compliance | Change to currency format |

**Table 2: Customer ID for 26 cases above**

|  |  |  |
| --- | --- | --- |
| 137 | 1753 | 5261 |
| 160 | 2187 | 5267 |
| 407 | 2948 | 6644 |
| 677 | 2991 | 7206 |
| 781 | 3260 | 9454 |
| 1004 | 4225 | 10541 |
| 1131 | 4228 |  |
| 1283 | 4339 |  |
| 1654 | 4653 |  |
| 1681 | 4896 |  |

**2. NewCustomerList Sheet**

Number records: 1000

**Table 3: Issues in NewCustomerList Sheet**

|  |  |  |
| --- | --- | --- |
| Field name | Issue | How to fix |
| past\_3\_years\_bike\_related\_purchases | Format Compliance | Change to number format with no decimal to easily statistic (if any) |
| postcode | Format compliance | Change to number format with no decimal to easily statistic (if any) |
| property\_valuation | Format compliance | Change to number format with no decimal to easily statistic (if any) |
| deceased\_indicator | Why are all “N” value? | Check carefully and fix |
| rank | Not continuous ranking | Check carefully and fix |

**3. CustomerDemographic Sheet**

Number records: 4000

**Table 4: Issues in CustomerDemographic Sheet**

|  |  |  |
| --- | --- | --- |
| Field name | Issue | How to fix |
| DOB | Whether customer ID 34 has the right DOB (1843-12-21) | Check carefully and fix |
| DOB | 87 blank records | Check carefully and fill or remove |
| gender | Error typing (M, F, Femal) | Replace to right values/ Create a drop-down list for faster and more precise |
| default | Nonsense information | Remove |
| customer\_id | Lack of information (9 cases listed in table 5) about DOB, job\_tittle, tenure | Check and fill or remove |
| job\_tittle | 506 blank records | Check and fill or remove |

**Table 5: Customer ID for 9 cases above**

|  |  |
| --- | --- |
| 267 | 2426 |
| 548 | 3779 |
| 680 | 3931 |
| 983 | 3935 |
| 1773 |  |

**4. CustomerAddress Sheet**

Number records: 3999

Issue: In “state” field, the data is not consistent.

How to fix: Identify only one to use (New South Wales or NSW, Victoria or VIC). However, we recommend that you should use NSW and VIC to have correspondence with other values in the state field. Another way is that you could create a drop-down list for faster and more precise filling.

Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Moving forward, the team will continue with the data cleaning, standar disation and transformation process for the purpose of model analysis. Questions will be raised along the way and assumptions documented. After we have completed this, it would be great to spend some time with your data SME to ensure that all assumptions are aligned with Sprocket Central’s understanding. Let me know if you have any questions or concerns. Thank you very much.

Best regards,

Cindy.

KPMG Data Analytics (Virtual Internship) team.