**KPMG TASK -1**

**DATA QUALITY ASSESSMENT**

Dear Martin,

Thank you for providing us the datasets from the Sprocket Central Pty Ltd, The main key quality issues that we have faced are summarized in the below table. Please let us know any queries regarding this issues which has been presented.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Completeness | Consistency | Currency | Relevancy | Validity |
| Customer Demographic | Age: Missing  DOB: Inaccurate | job\_title, job\_industry\_category: Blanks  Customer\_Id:  Incomplete | Gender: Inconsistency | Deceased\_customers: filter out | Default feature: Delete |  |
| Transaction Data |  | online\_order, brand: Blanks  Customer\_Id: Incomplete |  |  | Cancel order\_online:  Filter out | Sold\_date:  Format |
| Customer Address |  | Customer\_Id, country:  Incomplete | States: Inconsistency |  |  |  |

**SUMMARY TABLE**

**Issues:**

1. **Accuracy: “DOB” was inaccurate in the customer Demographic dataset and missing an feature called “Age”;**

Recommendations: It was corrected by filtering out the missing values in excel and for the Age feature we recommend to create a new feature which would be easier to analyze the errors easily.

1. **Completeness: “Customer\_Id’s” were inconsistent for all the 3 datasets.**

Recommendations: Check whether the data is complete or else filter out those rows.

**Blanks in “job\_title” for the customer Demographic dataset and “online\_order, brand” features contains blanks**

Recommendations: Ensure you remove all the blanks before you proceed with the analysis.

1. **Consistency:** Inconsistency in “Gender” feature for customer Demographic dataset and “State” in Customer Address dataset.

Recommendations: Here you can filter all the ‘M’ into Males and ‘F’ into females and similarly we can replace New South Wales with “NSW”, Victoria with ‘”VIC”.

1. **Currency:** People that are ‘Y’ in deceased\_indicator are not current customers for Customer Demographic dataset.

Recommendations: You can filter out the ‘Y’ from the feature.

1. **Relevancy:** There was no such relation or relevancy in “Default” feature for customer demographic dataset and order\_status from transactions dataset.

Recommendations: Delete the Meta data from the dataset and for the order\_status you can remove the category “Cancelled” from the order\_status feature.

1. **Validity:** Format of Product\_sale\_date for transactions

Recommendations: Try to change the format to short date for the product\_sale\_date and

This summarizes all data issues discovered through the data quality assessment test. The strategy that has been followed is quite simple and can be used effectively for the future analysis.

Please let me know if there are some queries regarding the strategies and steps that have been followed during the analysis.

Kind regards,

Nandha Kishore