**Submitted By:** Osama Mumtaz

KMPG Internship

Module 1

Data Quality Assessment

Hi there,

As per voice mail I have resolved some quality issue that is defined below.

This is Osama Mumtaz from KPMG Data Analytics (Virtual Internship) team. We have reviewed the data sets which were provided by your company and during the data quality analysis, we have found the some errors in the data sets.

The data quality analysis is the core phase and due to errors in the data set, we suggest the following mitigates in order to improve the data quality, which will eventually help us to driven the better analytics, results for your company.

1. In transaction table the column “online\_order” should be like “order\_type” and in that write online or physical order.
2. In last column of transaction sheet date is not accurate, I mean you should define correct format for date.
3. In Customer Demographic column “gender”, data accuracy issue write proper word like male or female instead of “F” or “M”.
4. In Customer Address sheet the column State has data accuracy issue where name of states are written in full format and some with short format like “NSW”,”VIC”,”QLD”.
5. We can take a mode year value for the missing records of customers DOB.
6. We can assign a uniform last name of customers, which values are missing.
7. For tenure values, we can take a mean of rest of the values and assign the mean value to the missing fields in order to maintain the consistency of data.
8. Eliminate the blank orders considering fake orders.

The following are the details of error encountered in the data set.

**Customer Demographics (Total Records 4000):**

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| DOB | 1843-12-21(not aligned properly to other records)  Blank record found |
| last\_name | Blank record found |
| Gender | Values are not consistence( Record consists of “U”, ”M”, ”Femal” and “F” |
| job\_title | Blank record found. |
| job\_industry | Found records with “n/a” |
| Default | Found records with value “special characters” including null and Blanks |
| Tenure | Blank records Found |

**Customer Address (Total Records 4003):**

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| State | Data accuracy issue (Some state has been written with full name or some with “VIC”,”NCW”,”QLD”). |

**Transaction (Total Records 20000-past 3 months):**

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| Online\_order | Blank records found |
| brand | Blank records found |
| product\_line | Blank records found |
| product\_class | Blank records found |
| product\_size | Blank records found |
| standard\_cost | Blank records found |
| product\_first\_sold\_date | Blank records found |