Dear Mr./Ms. Client,

How are you Doing?

First of all, I want to thank you for placing your trust in KPMG with your data and to assure you that our team will offer the best services and advice.   
In regard of the data queries we have spoken about:

**Data sets received:**

* Transactions
* CustomerDemographic
* CustomerAddress
* NewCustomerList

We spoked previously about three data sets, but I mas inform you that four were found.

**Data Quality:**

1. **Data Normalization:** NewCustomerList has no Primary Key. This makes the information inside the dataset to be very difficult to link to the rest of the data.
2. **Resume:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Completeness | Consistency | Currency | Relevancy | Validity |
| CustomerDemographic: | **DOB:**  Age information not provided | **Job\_title** missing 506 rows (*13% of total records*)  **Last\_name** missing 125 rows (*6% of total records)*  **Job\_industry** with 656 ‘n/a’ rows. *(16.4% of total records)* | **Gender** is not consistent | **DOB** record with an 1843 year of birth | **Deceased\_indicator** Information about deceased clients | **Default** containing non- recognizable values  **Tenure** information not clear or relevant |
| CustomerAddress: | **Address** information all in one record | **-** | **State**  is not consistent | **-** | **-** | **Property\_valuation** no clearreference of meaning for the value |
| Transactions: | **Revenue** information not provided | **Brand, product\_line, product\_class, product\_size** missing 197 rows(*1% of total transactions*) | **-** | **-** | **-** | **Transaction\_date** is for all 2017. Not only 3 months.  **Product\_first\_sold\_date** format is not in date format  **List\_price** format not in currency |

**3.Measure to be taken:  
  
Data normalization:** Apply the same logic of CustomerID column from Transactions dataset to the NewCustomerList. This way this data set could be linked to the rest of the data and work with it.   
**Then:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Accuracy:** | | |
|  | **Issue** | | **Mitigation** | **Recommendation** |
| CustomerDemographic: | DOB information is not accurate. Age information is needed | | Creation of an **Age** column in the data set | Beyond gathering the data of birth information, also, at the same instance, collect the age information of the customers |
| CustomerAddress: | Address column has all the information in one field | | The column could be divided into two. One with the number and the other with the text | See this type of information as two separated instances.  One is the number of the door and the other is the name of the street.  The first is unique, the second could be repeated. |
| Transactions: | No Revenue or Profit column provided | | Calculation of a **Revenue** column, using the List\_price and Standard\_cost columns | Configure in the system a calculation that gives the profit of every transaction. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Completeness:** | | |
|  | **Issue** | | **Mitigation** | **Recommendation** |
| CustomerDemographic: | **Job\_title** missing 506 rows (*13% of total records*)  **Last\_name** missing 125 rows (*6% of total records)*  **Job\_industry** with 656 ‘n/a’ rows. *(16.4% of total records)* | | Keep the analysis going forward taking note of the missing values | In the process of creating a new client, generate a lock in the system that does not allow to finish the procedure with empty fields. |
| CustomerAddress: | **-** | | **-** | **-** |
| Transactions: | **Brand, product\_line, product\_class, product\_size** missing 197 rows(*1% of total transactions*) | | Keep the analysis going forward taking note of the missing values | In the process of creating a new client, generate a lock in the system that does not allow to finish the procedure with empty fields. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Consistency:** | | |
|  | **Issue** | | **Mitigation** | **Recommendation** |
| CustomerDemographic: | **Gender** is not consistent | | Unification of the classification | Set up pre-defined values and lock them. So this ones are the only ones to be used |
| CustomerAddress: | **State**  is not consistent | | Unification of the classification | Set up pre-defined values and lock them. So this ones are the only ones to be used |
| Transactions: | **-** | | **-** | **-** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Currency:** | | |
|  | **Issue** | | **Mitigation** | **Recommendation** |
| CustomerDemographic: | **DOB** record with an 1843 year of birth | | Not taking in consideration the row. Is only one data record | Configure the year field with a logical minimum from where to choose |
| CustomerAddress: | **-** | | **-** | **-** |
| Transactions: | **-** | | **-** | **-** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Relevancy:** | | |
|  | **Issue** | | **Mitigation** | **Recommendation** |
| CustomerDemographic: | **Deceased\_indicator** Information about deceased clients | | Not taking in consideration the row. There are only two data records | If possible, separate the data from de Y and the N of the field.  Don’t see possible use of deceased information. |
| CustomerAddress: | **-** | | **-** | **-** |
| Transactions: | **-** | | **-** | **-** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Validity:** | | |
|  | **Issue** | | **Mitigation** | **Recommendation** |
| CustomerDemographic: | **Default** containing non- recognizable values  **Tenure** information not clear or relevant | | Default column deleted  Not taking in consideration for future calculations | If possible, check for columns with random characters  Summit a standard scale of the values. |
| CustomerAddress: | **Property\_valuation** no clearreference of meaning for the value | | Not taking in consideration for future calculations | Summit a standard scale of the values. |
| Transactions: | **Transaction\_date** is for all 2017. Not only 3 months.  **Product\_first\_sold\_date** format is not in date format  **List\_price** format not in currency | | Make the analysis with all the data available.  Change the format to date    Change the format to currency | Clarify the range needed to use for the analysis  In the data entry process, fix the format of the columns. To avoid confusions  In the data entry process, fix the format of the columns. To avoid confusions |

A final comment on the data. The most important aspect that will improve the quality of this dataset is to create the connection between the ‘**NewCustomerList’** and the rest of the data. This can be accomplished by generating a **CustomerID** that follows the specificity of the actual datasets. If you consider that out assistance is needed, will be more than pleased to help you to solve this matter.

This analysis has the objective to improve the data quality so the future analysis could deliver solid outcomes that help improve de marketing strategy of the company.

Please don’t hesitate to question our assumptions or to make any enquiry that you consider regarding this report. We are here to **help you drive through** data driven implementations.

Looking forward to hearing from you,  
Martin Fernandez Cean  
**KPMG**