**DATA QUALITY ASSESSMENT**

**Accuracy Issues:**

* DOB column in “Customer Demographics” had few inappropriate values.

*Mitigation: Filter out outliers in DOB.*

*Recommendations: Create an age\_column, allowing for more comprehensible data. Set a range for values of age column using Tableau Prep Builder.*

* Grammatical error in the ‘job\_industry\_category.’

*Mitigation: Change ‘Argiculture’ to ‘Agriculture.’*

*Recommendations: Create an age\_column, allowing for more comprehensible data. Set a range for values of age column using Tableau Prep Builder.*

**Completeness:**

* Null Values in many columns.

*Mitigation: Please ensure that all tables are free from Null values.*

*Recommendations: Filter out null values by using Keep Only non-null values filter in Tableau Prep Builder.*

* There were many customer\_ids in Transactions which were not present in Customer Demographics and Customer Address sheets.

*Mitigation: Please ensure that all tables are from the same period. Only customers in the Customer Master list will be used as a training set for our model.*

*Recommendations: Do an inner join of customer demographics and customer address on customer\_id. On this perform join with Transactions data on customer\_ids column.*

**Consistency:**

# Consistency should be maintained in all the values of each attribute.

*Mitigation: Change all ‘F’,’Femal’ to Female. All ‘M’ to ‘Male’ for ‘gender’ column. Change all ‘New South Wales’ to ‘NSW’ and ‘Victoria’ to ‘VIC’, in state column of Customer Demographics..*

*Recommendation: Edit values in tableau prep builder.*

# Inconsistent data type for the attributes

*Mitigation: Convert selected records in characters to numeric. Remove non-numeric characters from string.*

*Recommendation: Can follow Tableau Prep Builder Recommendations.*

**Currency:**

# Records with ‘Y’ in deceased\_indicator in “Customer Demographics” are irrelavent.

# *Mitigation: Retain only those customers who have a ‘F’ in deceased\_indicator.*

# *Recommendations: Keep only records with ‘F’ in deceased indicator in Tableau Prep Builder.*

**Relevancy:**

# Records with ‘Cancelled’ order\_status in “Transactions” data are irrelevant.

# *Mitigation: Retain only those customers who have a ‘Approved’ in order\_status.*

# *Recommendations: Keep only records with Transactions” in deceased indicator in Tableau Prep Builder.*

**Validity:**

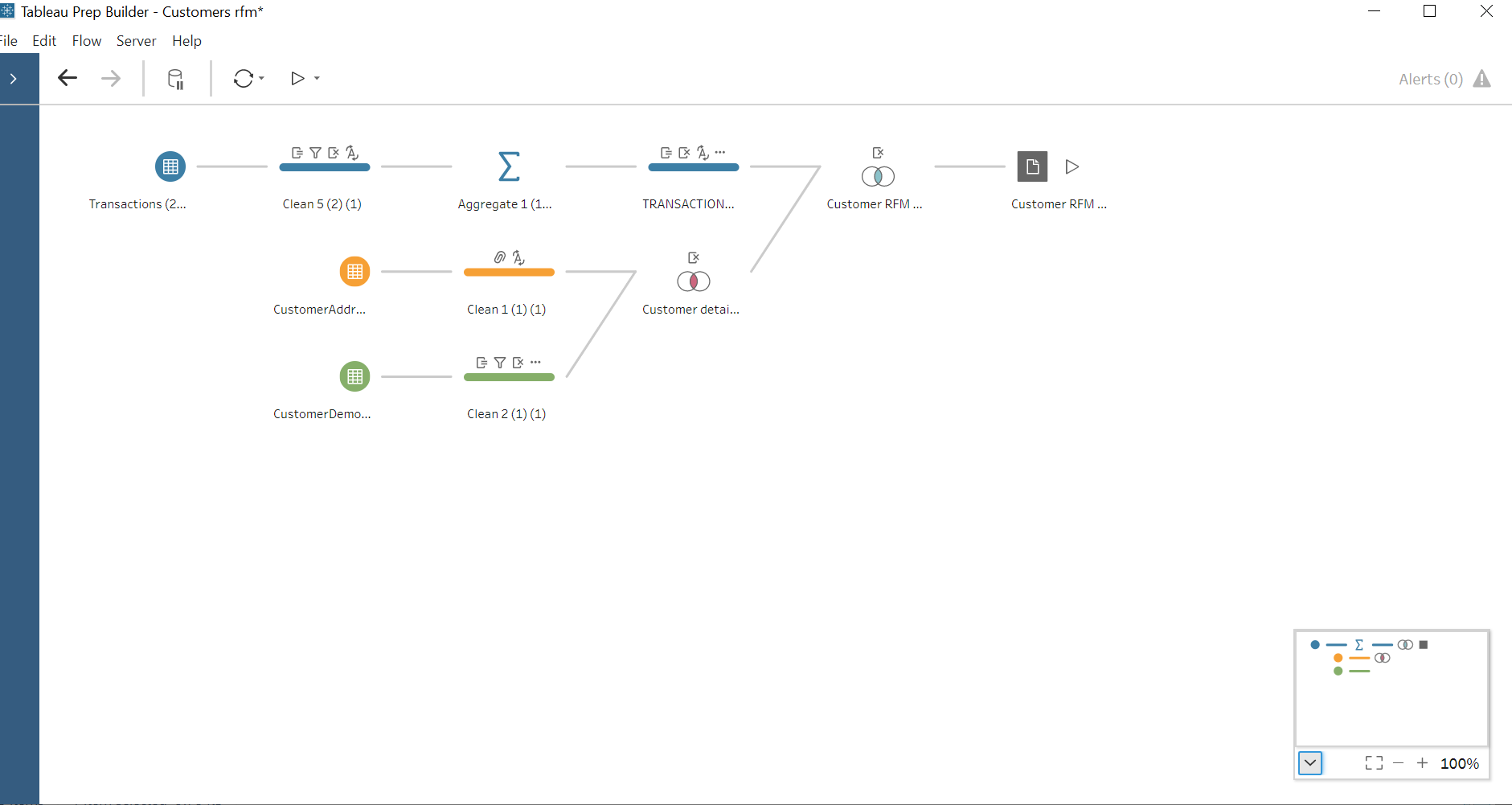
# Ensure all the columns are assigned proper data types.

# *Mitigation: Format columns to relevant data types.*

*Recommendations: Can follow Tableau Prep Builder Recommendations.*

**PREPARATION OF TRAINING DATA**

*Do an inner join of customer demographics and customer address on customer\_id. On this perform join with Transactions data on customer\_ids column.*



**Customer Demographics:**

Calculated fields:

Age: (TODAY()-[DOB])/365

**Transactions:**

5 Aggregate fields:

1. Group by customer\_id

2. Count transaction\_id

3. Maximum of transaction\_date

4.SUM list\_price

5.SUM standard\_cost

Calculated Fields:

1. Margin(Monetary): Profit is calculated.

[list\_price]-[standard\_cost]

2. Recency: Number of days between the latest transaction of the customer and the latest transaction in the company.

(DATE("2017-12-30")-[transaction\_date])

3.Frequency: Records are grouped by customer\_ids and count of transaction\_ids will give the frequency of transactions of each customer.

Rename count transaction\_id to Frequency.