**Data Quality Issues and Solutions**

**1. Missing Values:**

* **Attrition\_Flag: 100 missing values.**
* **Gender: 200 missing values.**
* **Education\_Level: 200 missing values.**
* **Card\_Category: 202 missing values.**

**Solution: We handled missing values by filling all null values with the mode (most frequent value) of the respective columns.**

**2. Duplicated Values:**

* **Duplicated Records: 401 duplicated records were identified.**

**Solution: We dropped all duplicated records to remove redundancy.**

**3. Inconsistencies:**

* **Attrition\_Flag: Contains "Unknown" values.**
* **Date\_of\_birth: Recorded as an object (string) type.**

**Solution:**

* **We deleted the "Unknown" values from the target column, *Attrition\_Flag*.**
* **We converted the *Date\_of\_birth* column from object type to date format for consistency.**

**4. Outliers:**

* **Affected Columns: ['Age', 'Months\_on\_book', 'Months\_Inactive\_12\_mon', 'Contacts\_Count\_12\_mon', 'Credit\_Limit', 'Avg\_Open\_To\_Buy', 'Total\_Amt\_Chng\_Q4\_Q1', 'Total\_Trans\_Amt', 'Total\_Trans\_Ct', 'Total\_Ct\_Chng\_Q4\_Q1'].**

**Solution: We handled outliers by capping the extreme values to fall within acceptable thresholds.**

**5. Unclear Categorical Values:**

* **Affected Columns: *Card\_Category*, *Education\_Level*, and *Income\_Category* contain ambiguous or unclear categories.**

**Solution: We plan to group and clarify these categorical values, such as consolidating categories like "Unknown" into more meaningful groups.**

**6. Redundant Column:**

* **Date\_of\_birth: This column provides similar information as the *Age* column.**

**Solution: We dropped the *Date\_of\_birth* column as it is redundant given that the *Age* column is already present in the dataset.**