

Data Cleaning Report

Client Name: [Insert Client Name]

Project Overview

The customer call list dataset was cleaned and standardized to ensure it is suitable for call agents. The process involved handling duplicates, missing values, and formatting inconsistencies while removing irrelevant data.

Initial Data Summary

- **Raw Data Shape:** 21 rows, 8 columns
- **Key Findings:**
 - Duplicate rows: 1
 - Null values:
 - Last_Name: 1
 - Phone_Number: 2
 - Do_Not_Contact: 4
 - Irrelevant column: Not_Useful_Column
 - Erroneous characters in Last_Name.
 - Inconsistent phone number formats and placeholders like "N/a".
 - Variations in categorical values for Paying Customer and Do_Not_Contact.

Cleaning Actions

1. Removed 1 duplicate row.
2. Removed the Not_Useful_Column.
3. Cleaned Last_Name column by removing invalid characters (e.g., '/', '.').
4. Standardized Phone_Number formats to a consistent format.
5. Split Address into Street_Address, State, and Zip_Code.
6. Standardized categorical values in Paying Customer and Do_Not_Contact to "Yes" and "No".
7. Removed records with empty Phone_Number or flagged as "Do_Not_Contact".
8. Identified and flagged duplicate phone numbers.

Results of Data Cleaning

- **Cleaned Dataset Shape:** 10 rows, 10 columns
- **Key Statistics:**
 - Rows removed due to Do_Not_Contact == 'Y': 4
 - Rows removed due to missing Phone_Number: 6

- Rows with duplicate `Phone_Number`: 7

Challenges and Notes

- Handling inconsistent phone number formats required manual verification for edge cases.
- Address splitting was successful, but some records had ambiguous data in the `Address` column.

Deliverables

1. Cleaned dataset exported as: **Cleaned_Customer_Call_List.xlsx**
2. Summary of findings and cleaning process in this report.

Recommendations

1. Implement data validation rules at the point of entry to reduce inconsistencies.
2. Regularly audit and clean datasets to ensure quality.
3. Use consistent formats for categorical data across systems.

Prepared by: Amedu Stephen
Date: [Insert Date]