

## Data Collection and Preprocessing Phase

|               |  |
|---------------|--|
| Date          | 20 June 2024   |
| Team ID       | 739986   |
| Project Title | Customer Acquisition cost estimation using machine Learning. |
| Maximum Marks | 2 Marks  |

### Data Quality Report:

The Data Quality Report will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

### Data Quality Report:

| Data Source    | Data Quality Issue  | Severity | Resolution Plan   |
|----------------|---|----------|---|
| Kaggle Dataset | Missing values in the following columns: 'Breakfast Foods', 'Bread', 'Canned Shrimp', 'Baking Goods', 'Vegetables', 'Frozen Desserts', 'Candy', 'Snack Foods', 'Dairy', 'Starchy Foods', 'Cleaning Supplies', 'Decongestants' | Moderate | Use mean/median imputation.   |
| Kaggle Dataset | Categorical data in the dataset   | Moderate | Apply appropriate encoding techniques (e.g., One-Hot Encoding, Label Encoding). |